

**uc3m** | Universidad **Carlos III** de Madrid

GRADO EN INGENIERÍA INFORMÁTICA  
CURSO 2019-2020

TRABAJO DE FIN DE GRADO

**ANÁLISIS DE DATOS PARA LA  
CONSERVACIÓN DEL PATRIMONIO  
HISTÓRICO ARQUITECTÓNICO**

Luis Monge Sanz de Bremond

Tutores

Antonio Berlanga de Jesús

Ana Carro Rossell

# 1. AGRADECIMIENTOS

*Quiero dedicar este trabajo a mis padres y agradecerles la educación y los valores que me han inculcado desde pequeño, valores sin los cuales no habría desarrollado el interés y la curiosidad que me ha empujado a emprender este y tantos otros proyectos. Gracias también a mis tutores por impulsarme a abordar este proyecto; al Dr. Don Antonio Berlanga de Jesús, por guiarme en la investigación en el campo de la ingeniería informática, y Doña Ana Carro Rossell por asesorarme desde el punto de vista del patrimonio cultural. Agradezco también a Hispania Nostra la cesión de sus datos de la Lista Roja de Patrimonio para la realización del estudio, y finalmente a todos mis profesores y docentes que han intervenido en mi formación académica.*

## 2. ÍNDICE DE CONTENIDO

<b>1. AGRADECIMIENTOS</b>	<b>1</b>
<b>2. ÍNDICE DE CONTENIDO</b>	<b>2</b>
<b>3. ÍNDICE DE ILUSTRACIONES</b>	<b>7</b>
<b>4. ÍNDICE DE TABLAS</b>	<b>13</b>
<b>5. DEFINICIONES Y ACRÓNIMOS</b>	<b>22</b>
5.1. DEFINICIONES	22
5.2. ACRÓNIMOS	25
<b>6. ABSTRACT</b>	<b>27</b>
6.1. INTRODUCTION	27
6.2. PROJECT PLANNING	28
6.3. PHASE 1: OBJECTIVES	28
6.4. PHASE 2: ENVIRONMENTAL SCANNING	28
6.5. PHASE 3: DATA PROCESSING AND INFORMATION EXTRACTION	29
6.5.1. INITIAL TREATMENT	29
6.5.2. INFORMATION EXTRACTION	30
6.5.3. INCLUSION OF EXTERNAL DATA	32
6.6. PHASE 4: DATA ANALYSIS	33
6.6.1. EXPLORATORY ANALYSIS	33
6.6.2. LINEAR CORRELATIONS	34
6.6.3. HYPOTHESIS CONTRAST	34
6.7. PHASE 5: RESULT ANALYSIS AND CONCLUSIONS	34
6.7.1. ANALYSIS OF THE RESULTS	34
6.7.2. CONCLUSIONS	35
6.8. PHASE 6: IMPROVEMENTS AND FUTURE RESEARCH LINES	36
<b>7. INTRODUCCIÓN</b>	<b>37</b>
7.1. MOTIVACIÓN Y ENFOQUE	39
7.2. OBJETIVOS	40
7.3. HISPANIA NOSTRA	40
7.4. HERRAMIENTAS	41

7.5.	ESTRUCTURA DEL DOCUMENTO	42
<b>8.</b>	<b>ESTUDIO DE VIABILIDAD DEL PROYECTO</b>	<b>43</b>
8.1.	ALCANCE DEL PROYECTO	43
8.2.	ESTUDIO DE LA SITUACIÓN ACTUAL	43
8.3.	ENTORNO SOCIOECONÓMICO	45
8.4.	MARCO REGULATORIO	46
8.5.	IDENTIFICACIÓN DE LAS PARTES INTERESADAS	47
<b>9.</b>	<b>GESTIÓN DEL PROYECTO</b>	<b>48</b>
9.1.	ORGANIZACIÓN DE TAREAS	48
9.2.	PLANIFICACIÓN	51
9.3.	ELABORACIÓN DE UN PRESUPUESTO	52
9.3.1.	RECURSOS HUMANOS	52
9.3.2.	RECURSOS MATERIALES	52
9.3.3.	DESPLAZAMIENTOS Y DIETAS	53
9.3.4.	COSTE TOTAL DEL PROYECTO	53
<b>10.</b>	<b>TRATAMIENTO DE LOS DATOS</b>	<b>54</b>
10.1.	EVALUACIÓN DE LA LISTA ROJA	54
10.2.	ADECUACIÓN DEL FICHERO DE DATOS	57
10.3.	EXTRACCIÓN DE INFORMACIÓN	58
10.3.1.	ATRIBUTO “ÉPOCA DE EDIFICACIÓN”	59
10.3.1.1.	UNIFICACIÓN DE FORMATOS	61
10.3.1.2.	TRATAMIENTO DE AÑOS	61
10.3.1.3.	TRATAMIENTO DE LA ERA CRONOLÓGICA	62
10.3.1.4.	TRATAMIENTO DE INTERVALOS	63
10.3.1.5.	TRATAMIENTO DE SIGLOS INDIVIDUALES	63
10.3.1.6.	TRATAMIENTO DE PERIODOS HISTÓRICOS	64
10.3.1.7.	BÚSQUEDA DE FECHAS DE EDIFICACIÓN EN EL ATRIBUTO “HISTORIA”	71
10.3.1.8.	UNIFICAR SIGLOS EN UNA ÚNICA VARIABLE	71
10.3.1.9.	INICIO Y FIN DEL PERIODO DE EDIFICACIÓN	72

10.3.1.10.	IDENTIFICACIÓN DE REFORMAS	72
10.3.1.11.	BATERÍA DE PRUEBAS	72
10.3.2.	ATRIBUTO “INTERVENCIONES”	85
10.3.2.1.	BATERÍA DE PRUEBAS	87
10.3.3.	ATRIBUTO “MOTIVO DE LA RETIRADA”	89
10.3.3.1.	BATERÍA DE PRUEBAS	93
10.3.4.	ATRIBUTO “PROPIEDAD”	98
10.3.4.1.	BATERÍA DE PRUEBAS	102
10.3.5.	ATRIBUTO “GRADO DE PROTECCIÓN LEGAL”	107
10.3.5.1.	BATERÍA DE PRUEBAS	110
10.3.6.	ATRIBUTO “ESTADO DE CONSERVACIÓN”	113
10.3.6.1.	BATERÍA DE PRUEBAS	119
10.3.7.	ATRIBUTO “RIESGO”	126
10.3.7.1.	BATERÍA DE PRUEBAS	134
10.3.8.	ATRIBUTO “DESCRIPCIÓN”	137
10.3.8.1.	MATERIALES	137
10.3.8.2.	ALTURAS	138
10.3.8.3.	SUPERFICIE	139
10.3.8.4.	BATERÍA DE PRUEBAS	139
10.3.9.	ATRIBUTO “HISTORIA”	141
10.3.9.1.	USO	141
10.3.9.2.	ACONTECIMIENTOS RELEVANTES	142
10.3.9.3.	BATERÍA DE PRUEBAS	142
10.4.	SELECCIÓN DE OTROS ATRIBUTOS	144
10.4.1.	ATRIBUTOS SOBRE LA LOCALIZACIÓN	144
10.4.2.	ATRIBUTO “FECHA DE INCLUSIÓN EN LA LISTA ROJA”	144
10.4.3.	ATRIBUTO “FECHA DE SALIDA DE LA LISTA ROJA”	144
10.5.	REDUCCIÓN DE LA DIMENSIONALIDAD	145

<b>11. INCLUSIÓN DE DATOS EXTERNOS</b>	<b>146</b>
11.1. DATOS CLIMATOLÓGICOS	146
11.2. VARIABLES SOCIOECONÓMICAS	147
11.3. INDICADORES DE LA CALIDAD DEL AIRE	147
11.4. TASA SISMOGRÁFICA	149
11.5. PROXIMIDAD A LA COSTA	150
11.6. NÚMERO DE SOCIOS DE HISPANIA NOSTRA	151
11.7. EXTENSIÓN GEOGRÁFICA	151
<b>12. ANÁLISIS DE LOS DATOS</b>	<b>152</b>
12.1. ANÁLISIS EXPLORATORIOS	152
12.1.1. DISTRIBUCIÓN GEOGRÁFICA DE LOS BIENES PATRIMONIALES DE LA LISTA ROJA	152
12.1.2. ESTADO DE CONSERVACIÓN DE LOS BIENES PATRIMONIALES	159
12.1.2.1. ANÁLISIS EN EL ÁMBITO NACIONAL	159
12.1.2.2. ANÁLISIS EN EL ÁMBITO AUTONÓMICO	162
12.1.2.3. ANÁLISIS EN EL ÁMBITO PROVINCIAL	164
12.1.3. ANÁLISIS EXPLORATORIO DE OTROS ATRIBUTOS NUMÉRICOS	167
12.1.4. ANÁLISIS EXPLORATORIO DE ATRIBUTOS BINARIOS	172
12.1.5. ANÁLISIS EXPLORATORIO DE ATRIBUTOS NOMINALES	175
12.2. CORRELACIONES LINEALES	178
12.2.1. NÚMERO DE FICHAS	179
12.2.2. ESTADO DE CONSERVACIÓN	182
12.3. CONTRASTE DE HIPÓTESIS	192
<b>13. ANÁLISIS DE LOS RESULTADOS</b>	<b>204</b>
<b>14. ANÁLISIS DE LA SOLUCIÓN</b>	<b>205</b>
<b>15. CONCLUSIONES</b>	<b>210</b>
15.1. POSIBLES MEJORAS Y PLANES DE ACTUACIÓN FUTUROS	211
<b>16. REFERENCIAS</b>	<b>213</b>
<b>17. ANEXO A: MEDIDAS DESCRIPTIVAS E HISTOGRAMAS DEL ESTADO DE CONSERVACIÓN A NIVEL AUTONÓMICO</b>	<b>216</b>
<b>18. ANEXO B: MEDIDAS DESCRIPTIVAS E HISTOGRAMAS DEL ESTADO DE CONSERVACIÓN A NIVEL PROVINCIAL</b>	<b>230</b>

<b>19. ANEXO C: PROPUESTAS PARA LA MEJORA DE LA LISTA ROJA DE PATRIMONIO</b>	<b>263</b>
19.1. ELIMINACIÓN DE LA REDUNDANCIA Y AGRUPACIÓN DE LOS ATRIBUTOS	263
19.2. UNIFORMIDAD DEL FICHERO DE DATOS	263
19.3. ERRORES EN EL FORMULARIO DE ENVÍO DE FICHAS	264
19.4. SISTEMATIZACIÓN DEL ESTADO DE CONSERVACIÓN	265
<b>20. ANEXO D: FICHA EMPLEADA PARA ANALIZAR EL FORMULARIO DE ENVÍO DE BIENES DE LA LISTA ROJA</b>	<b>266</b>

### 3. ÍNDICE DE ILUSTRACIONES

Figura 8.1. Principales grupos de interés del proyecto	47
Figura 9.1. Esquema de organización de tareas del proyecto	50
Figura 9.2. Diagrama de Gantt (parte 1)	51
Figura 9.3. Diagrama de Gantt (parte 2)	51
Figura 9.4. Diagrama de Gantt (parte 3)	51
Figura 10.1. Extracto del fichero de datos en su forma inicial	54
Figura 10.2. Extracto del fichero de datos tras su tratamiento	54
Figura 10.3. Muestras originales del atributo <i>Época</i>	59
Figura 10.4. Representación de los términos a detectar por el sistema	59
Figura 10.5. Resultado de la ejecución	59
Figura 10.6. Términos más frecuentes hallados en <i>Época</i>	60
Figura 10.7. Ejemplos de la operación de intersección aplicada a periodos históricos	68
Figura 10.8. Ejemplos de la operación de unión aplicada a periodos históricos	68
Figura 10.9. Muestras originales del atributo <i>Intervenciones</i>	85
Figura 10.10. Representación de los términos a detectar por el sistema	85
Figura 10.11. Resultado de la ejecución	85
Figura 10.12. Términos más frecuentes hallados en <i>Intervenciones</i>	86
Figura 10.13. Muestras originales del atributo <i>Motivo.de.la.retirada</i>	89
Figura 10.14. Representación de los términos a detectar por el sistema	89
Figura 10.15. Resultado de la ejecución	89
Figura 10.16. Términos más frecuentes hallados en <i>Motivo.de.la.retirada</i>	90
Figura 10.17. Muestras originales del atributo <i>Propiedad</i>	98
Figura 10.18. Representación de los términos a detectar por el sistema	98
Figura 10.19. Resultado de la ejecución	98



Figura 10.20. Términos más frecuentes hallados en <i>Propiedad</i>	100
Figura 10.21. Emparejamientos más frecuentes de términos hallados en <i>Propiedad</i>	100
Figura 10.22. Muestras originales del atributo <i>Grado.de.protección.legal</i>	107
Figura 10.23. Representación de los términos a detectar por el sistema	107
Figura 10.24. Resultado de la ejecución	107
Figura 10.25. Términos más frecuentes hallados en <i>Grado.de.proteccion.legal</i>	108
Figura 10.26. Emparejamientos más frecuentes de términos hallados en <i>Grado.de.proteccion.legal</i>	108
Figura 10.27. Muestras originales del atributo <i>Estado.de.conservación</i>	113
Figura 10.28. Representación de los términos a detectar por el sistema	113
Figura 10.29. Representación de la cuantización de las variables	113
Figura 10.30. Resultado de la ejecución	113
Figura 10.31. Términos más frecuentes hallados en <i>Estado.de.conservación</i>	114
Figura 10.32. Emparejamientos más frecuentes de términos hallados en <i>Estado.de.conservación</i>	115
Figura 10.33. Muestras originales del atributo <i>Riesgo.por.el.que.entró</i>	126
Figura 10.34. Representación de los términos a detectar por el sistema	126
Figura 10.35. Resultado de la ejecución	126
Figura 10.36. Emparejamientos más frecuentes de términos hallados en <i>Riesgo.por.el.que.entró</i>	127
Figura 10.37. Emparejamientos más frecuentes de términos hallados en <i>Riesgo.por.el.que.entró</i>	128
Figura 10.38. Emparejamientos más frecuentes de <i>caída</i> en <i>Riesgo.por.el.que.entró</i>	130
Figura 10.39. Emparejamientos más frecuentes de <i>pérdida</i> en <i>Riesgo.por.el.que.entró</i>	130
Figura 10.40. Emparejamientos más frecuentes de <i>desaparición</i> en <i>Riesgo.por.el.que.entró</i>	131

Figura 10.41. Emparejamientos más frecuentes de <i>desprendimiento</i> en <i>Riesgo.por.el.que.entró</i>	131
Figura 10.42. Emparejamientos más frecuentes de <i>destrucción</i> en <i>Riesgo.por.el.que.entró</i>	133
Figura 10.43. Emparejamientos más frecuentes hallados en <i>Descripción</i>	137
Figura 10.44. Emparejamientos más frecuentes hallados en <i>Historia</i>	141
Figura 11.1. Regiones sismográficas de la península ibérica	149
Figura 11.2. Mapa de municipios costeros	150
Figura 12.1. Fichas totales por comunidad autónoma	153
Figura 12.2. Número de fichas totales por comunidad autónoma	153
Figura 12.3. Número de fichas de la Lista Roja por comunidad autónoma	154
Figura 12.4. Número de fichas de la Lista Roja por comunidad autónoma	154
Figura 12.5. Número de fichas de la Lista Negra por comunidad autónoma	155
Figura 12.6. Fichas totales por provincia	156
Figura 12.7. Número de fichas totales por provincia	157
Figura 12.8. Número de fichas de la Lista Roja por provincia	157
Figura 12.9. Número de fichas de la Lista Verde por provincia	158
Figura 12.10. Número de fichas de la Lista Negra por provincia	158
Figura 12.11. Histogramas del estado de conservación	160
Figura 12.12. Diagramas de cajas del estado de conservación	160
Figura 12.13. Diagramas de densidad del estado de conservación	161
Figura 12.14. Diagramas de cajas de Valoración.estado por comunidades autónomas	163
Figura 12.15. Diagramas de cajas de Estado.deterioro por comunidades autónomas	163
Figura 12.16. Diagramas de cajas de Valoración.estado por provincias	165
Figura 12.17. Diagramas de cajas de Estado.deterioro por provincias	166
Figura 12.18. Distribución geográfica del número de socios	171

Figura 12.19. Distribuciones geográficas de bienes abandonados	172
Figura 12.20. Distribuciones geográficas de bienes en riesgo de abandono	172
Figura 12.21. Distribuciones geográficas de bienes expoliados	173
Figura 12.22. Distribuciones geográficas de bienes en riesgo de expolio	173
Figura 12.23. Distribuciones geográficas de bienes vandalizados	173
Figura 12.24. Distribuciones geográficas de bienes afectados por la vegetación	174
Figura 12.25. Distribuciones geográficas de bienes en riesgo de deterioro	174
Figura 12.26. Distribuciones geográficas de bienes con riesgos estructurales	174
Figura 12.27. Distribuciones geográficas de bienes en riesgo de ruina	175
Figura 12.28. Distribuciones geográficas de bienes en riesgo de demolición	175
Figura 12.29. Diagramas de dispersión del número de socios y el número de fichas	179
Figura 12.30. Diagramas de dispersión de la extensión geográfica y el número de fichas	180
Figura 12.31. Diagramas de dispersión del número de socios y la población	180
Figura 12.32. Matriz de correlaciones del número de fichas y las variables climáticas	181
Figura 12.33. Matriz de correlaciones del estado de conservación y la calidad del aire	182
Figura 12.34. Diagramas de dispersión del estado de conservación y la calidad del aire	183
Figura 12.35. Matriz de correlaciones de la media del estado de conservación y la calidad del aire por provincias	184
Figura 12.36. Matriz de correlaciones de la media del estado de conservación y la calidad del aire por comunidades autónomas	184
Figura 12.37. Diagramas de dispersión del estado de conservación y la calidad del aire por provincias	185
Figura 12.38. Diagramas de dispersión del estado de conservación y la calidad del aire por comunidades autónomas	185

Figura 12.39. Matriz de correlaciones del estado de conservación y las variables climáticas	186
Figura 12.40. Diagramas de dispersión del estado de conservación y las variables climáticas (parte 1)	186
Figura 12.41. Diagramas de dispersión del estado de conservación y las variables climáticas (parte 2)	187
Figura 12.42. Matriz de correlaciones de la media del estado de conservación y las variables climáticas por provincias	188
Figura 12.43. Matriz de correlaciones de la media del estado de conservación y las variables climáticas por comunidades autónomas	188
Figura 12.44. Matriz de correlaciones del estado de conservación y los indicadores socioeconómicos	189
Figura 12.45. Diagramas de dispersión del estado de conservación con respecto al PIB per cápita	189
Figura 12.46. Diagramas de dispersión del estado de conservación con respecto a la tasa de paro	190
Figura 12.48. Diagramas de dispersión del número de fichas intervenidas con respecto al PIB per cápita y la tasa de paro	191
Figura 12.49. Matriz de correlaciones del estado de conservación y la época de edificación	191
Figura A.1. Histogramas del estado por comunidades autónomas (parte 1)	225
Figura A.2. Histogramas del estado por comunidades autónomas (parte 2)	226
Figura A.3. Histogramas del estado por comunidades autónomas (parte 3)	227
Figura A.4. Histogramas del estado por comunidades autónomas (parte 4)	228
Figura A.5. Histogramas del estado por comunidades autónomas (parte 5)	229
Figura B.1. Histogramas del estado por provincias en Galicia	250
Figura B.2. Histogramas del estado por provincias en País Vasco.	251
Figura B.3. Histogramas del estado por provincias en Aragón.	252
Figura B.4. Histogramas del estado por provincias en Cataluña.	253
Figura B.5. Histogramas del estado por provincias en Castilla y León (parte 1).	254

Figura B.6. Histogramas del estado por provincias en Castilla y León (parte 2).	255
Figura B.7. Histogramas del estado por provincias en Castilla y León (parte 3).	256
Figura B.8. Histogramas del estado por provincias en Castilla-La Mancha (Parte 1)	256
Figura B.9. Histogramas del estado por provincias en Castilla-La Mancha (Parte 2)	257
Figura B.10. Histogramas del estado por provincias en Valencia	258
Figura B.11. Histogramas del estado por provincias en Extremadura	259
Figura B.12. Histogramas del estado por provincias en Andalucía (parte 1)	259
Figura B.13. Histogramas del estado por provincias en Andalucía (parte 2)	260
Figura B.14. Histogramas del estado por provincias en Andalucía (parte 3)	261
Figura B.15. Histogramas del estado por provincias en Islas Canarias	262
Figura C.1. Desplegable de Provincia en el formulario.	264
Figura C.2. Posibles combinaciones erróneas permitidas en el formulario.	264
Figura C.3. Desplegable de Tipología en el formulario.	265

*Nota: Con el objetivo de evitar fallos de codificación con caracteres especiales del castellano, se han suprimido las tildes y virgulillas de nombres de atributos y contenidos y por tanto aparecen así escritos en gran cantidad de gráficos generados por R.*

## 4. ÍNDICE DE TABLAS

Tabla 10.1. Clasificación del patrimonio	56
Tabla 10.3. Caso de prueba ÉPOCA_P1	73
Tabla 10.4. Caso de prueba ÉPOCA_P2	73
Tabla 10.5. Caso de prueba ÉPOCA_P3	74
Tabla 10.6. Caso de prueba ÉPOCA_P4	74
Tabla 10.7. Caso de prueba ÉPOCA_P5	75
Tabla 10.8. Caso de prueba ÉPOCA_P6	75
Tabla 10.9. Caso de prueba ÉPOCA_P7	76
Tabla 10.10. Caso de prueba ÉPOCA_P8	76
Tabla 10.11. Caso de prueba ÉPOCA_P9	77
Tabla 10.12. Caso de prueba ÉPOCA_P10	77
Tabla 10.13. Caso de prueba ÉPOCA_P11	78
Tabla 10.14. Caso de prueba ÉPOCA_P12	78
Tabla 10.15. Caso de prueba ÉPOCA_P13	79
Tabla 10.16. Caso de prueba ÉPOCA_P14	79
Tabla 10.17. Caso de prueba ÉPOCA_P15	80
Tabla 10.18. Caso de prueba ÉPOCA_P16	80
Tabla 10.19. Caso de prueba ÉPOCA_P17	81
Tabla 10.20. Caso de prueba ÉPOCA_P18	81
Tabla 10.21. Caso de prueba ÉPOCA_P19	82
Tabla 10.22. Caso de prueba ÉPOCA_P20	82
Tabla 10.23. Caso de prueba ÉPOCA_P21	83
Tabla 10.24. Caso de prueba ÉPOCA_P22	83
Tabla 10.25. Caso de prueba ÉPOCA_P23	84
Tabla 10.26. Caso de prueba ÉPOCA_P24	84

Tabla 10.27. Caso de prueba INTERVENCIÓN_P1	87
Tabla 10.28. Caso de prueba INTERVENCIÓN_P2	87
Tabla 10.29. Caso de prueba INTERVENCIÓN_P3	88
Tabla 10.30. Caso de prueba INTERVENCIÓN_P4	88
Tabla 10.31. Caso de prueba INTERVENCIÓN_P5	88
Tabla 10.32. Motivos de la retirada: términos y acciones	91
Tabla 10.33. Anexo a Motivos de la retirada: términos y acciones	92
Tabla 10.34. Caso de prueba MOTIVO_P1	93
Tabla 10.35. Caso de prueba MOTIVO_P2	93
Tabla 10.36. Caso de prueba MOTIVO_P3	94
Tabla 10.37. Caso de prueba MOTIVO_P4	94
Tabla 10.38. Caso de prueba MOTIVO_P5	95
Tabla 10.39. Caso de prueba MOTIVO_P6	95
Tabla 10.40. Caso de prueba MOTIVO_P7	96
Tabla 10.41. Caso de prueba MOTIVO_P8	96
Tabla 10.42. Caso de prueba MOTIVO_P9	97
Tabla 10.43. Caso de prueba MOTIVO_P10	97
Tabla 10.44. Clasificación de propietarios	101
Tabla 10.45. Caso de prueba PROPIEDAD_P1	102
Tabla 10.46. Caso de prueba PROPIEDAD_P2	102
Tabla 10.47. Caso de prueba PROPIEDAD_P3	103
Tabla 10.48. Caso de prueba PROPIEDAD_P4	103
Tabla 10.49. Caso de prueba PROPIEDAD_P5	103
Tabla 10.50. Caso de prueba PROPIEDAD_P6	104
Tabla 10.51. Caso de prueba PROPIEDAD_P7	104
Tabla 10.52. Caso de prueba PROPIEDAD_P8	104

Tabla 10.53. Caso de prueba PROPIEDAD_P9	105
Tabla 10.54. Caso de prueba PROPIEDAD_P10	105
Tabla 10.55. Caso de prueba PROPIEDAD_P11	105
Tabla 10.56. Caso de prueba PROPIEDAD_P12	106
Tabla 10.57. Caso de prueba PROTECCIÓN_P1	110
Tabla 10.58. Caso de prueba PROTECCIÓN_P2	111
Tabla 10.59. Caso de prueba PROTECCIÓN_P3	111
Tabla 10.60. Caso de prueba PROTECCIÓN_P4	111
Tabla 10.61. Caso de prueba PROTECCIÓN_P5	112
Tabla 10.62. Caso de prueba PROTECCIÓN_P6	112
Tabla 10.63. Lista de valoraciones ordenadas por categorías (parte 1)	116
Tabla 10.64. Lista de valoraciones ordenadas por categorías (parte 2)	117
Tabla 10.65. Lista de cuantificadores	118
Tabla 10.66. Caso de prueba ESTADO_P1	119
Tabla 10.67. Caso de prueba ESTADO_P2	120
Tabla 10.68. Caso de prueba ESTADO_P3	120
Tabla 10.69. Caso de prueba ESTADO_P4	120
Tabla 10.70. Caso de prueba ESTADO_P5	121
Tabla 10.71. Caso de prueba ESTADO_P6	121
Tabla 10.72. Caso de prueba ESTADO_P7	121
Tabla 10.73. Caso de prueba ESTADO_P8	122
Tabla 10.74. Caso de prueba ESTADO_P9	122
Tabla 10.75. Caso de prueba ESTADO_P10	123
Tabla 10.76. Caso de prueba ESTADO_P11	124
Tabla 10.77. Caso de prueba ESTADO_P12	124
Tabla 10.78. Caso de prueba ESTADO_P13	125



Tabla 10.79. Caso de prueba ESTADO_P14	125
Tabla 10.80. Caso de prueba ESTADO_P15	125
Tabla 10.81. Términos clave del atributo <i>Riesgo.por.el.que.entró</i>	129
Tabla 10.82. Elementos estructurales	132
Tabla 10.83. Elementos no estructurales	132
Tabla 10.84. Caso de prueba RIESGO_P1	134
Tabla 10.85. Caso de prueba RIESGO_P2	134
Tabla 10.86. Caso de prueba RIESGO_P3	134
Tabla 10.87. Caso de prueba RIESGO_P4	135
Tabla 10.88. Caso de prueba RIESGO_P5	135
Tabla 10.89. Caso de prueba RIESGO_P6	135
Tabla 10.90. Caso de prueba RIESGO_P7	136
Tabla 10.91. Caso de prueba RIESGO_P8	136
Tabla 10.92. Caso de prueba RIESGO_P9	136
Tabla 10.93. Caso de prueba RIESGO_P10	136
Tabla 10.94. Materiales más frecuentes empleados en la edificación	138
Tabla 10.95. Caso de prueba DESCRIPCIÓN_P1	139
Tabla 10.96. Caso de prueba DESCRIPCIÓN_P2	140
Tabla 10.97. Caso de prueba DESCRIPCIÓN_P3	140
Tabla 10.98. Caso de prueba DESCRIPCIÓN_P4	140
Tabla 10.99. Caso de prueba HISTORIA_P1	142
Tabla 10.100. Caso de prueba HISTORIA_P2	143
Tabla 10.101. Caso de prueba HISTORIA_P3	143
Tabla 10.102. Caso de prueba HISTORIA_P4	143
Tabla 11.1. Estructura de los ficheros de mediciones de la calidad del aire	148
Tabla 12.1. Medidas descriptivas del estado de bienes incluidos en la Lista Roja en el ámbito nacional	159

Tabla 12.2. Medidas descriptivas de los valores de la temperatura máxima	167
Tabla 12.3. Medidas descriptivas de los valores de la temperatura mínima	167
Tabla 12.4. Medidas descriptivas de los valores de la variación de temperatura	168
Tabla 12.5. Medidas descriptivas de los valores de la humedad máxima	168
Tabla 12.6. Medidas descriptivas de los valores de la velocidad máxima del viento	169
Tabla 12.7. Medidas descriptivas de los valores de la precipitación máxima	169
Tabla 12.8. Medidas descriptivas de los valores de la probabilidad de precipitación máxima	169
Tabla 12.9. Medidas descriptivas de los valores de NOx	170
Tabla 12.10. Medidas descriptivas de los valores de NO2	170
Tabla 12.11. Medidas descriptivas de los valores de SO2	170
Tabla 12.12. Medidas descriptivas de los valores del PIB per cápita	171
Tabla 12.13. Medidas descriptivas de los valores del paro	171
Tabla 12.14. Número de fichas según los tipos de protección más frecuentes	176
Tabla 12.15. Número de fichas según los tipos de propietario más frecuentes	177
Tabla 12.16. Número de fichas según los tipos de motivos de la retirada más frecuentes	178
Tabla 12.17. Tests Chi Cuadrado de Pearson TEST_01 y TEST_02	193
Tabla 12.18. Tests Chi Cuadrado de Pearson TEST_03 y TEST_04	193
Tabla 12.19. Tests Chi Cuadrado de Pearson TEST_05 y TEST_06	194
Tabla 12.20. Tests Chi Cuadrado de Pearson TEST_07 y TEST_08	194
Tabla 12.21. Tests Chi Cuadrado de Pearson TEST_09 y TEST_10	195
Tabla 12.22. Tests Chi Cuadrado de Pearson TEST_11 y TEST_12	195
Tabla 12.23. Tests Chi Cuadrado de Pearson TEST_13 y TEST_14	196
Tabla 12.24. Tests Chi Cuadrado de Pearson TEST_15 y TEST_16	196
Tabla 12.25. Tests Chi Cuadrado de Pearson TEST_17 y TEST_18	197

Tabla 12.28. Tests Chi Cuadrado de Pearson TEST_19 y TEST_20	197
Tabla 12.26. Tests Chi Cuadrado de Pearson TEST_21 y TEST_22	198
Tabla 12.27. Tests Chi Cuadrado de Pearson TEST_21 y TEST_22	198
Tabla 12.29. Tests Chi Cuadrado de Pearson TEST_25 y TEST_26	199
Tabla 12.30. Tests Chi Cuadrado de Pearson TEST_27 y TEST_28	199
Tabla 12.31. Tests Chi Cuadrado de Pearson TEST_29 y TEST_30	200
Tabla 12.32. Tests Chi Cuadrado de Pearson TEST_31 y TEST_32	200
Tabla 12.33. Tests Chi Cuadrado de Pearson TEST_33 y TEST_34	201
Tabla 12.34. Tests Chi Cuadrado de Pearson TEST_35 y TEST_36	201
Tabla 12.35. Test Chi Cuadrado de Pearson TEST_37	202
Tabla 12.36. Tests Chi Cuadrado de Pearson TEST_38 y TEST_39	203
Tabla A.1. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Galicia	216
Tabla A.2. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en el Principado de Asturias.	216
Tabla A.3. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Cantabria	217
Tabla A.4. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Cantabria.	217
Tabla A.5. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en la C.F. de Navarra.	218
Tabla A.6. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en País Vasco.	218
Tabla A.7. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Cataluña.	219
Tabla A.8. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Castilla y León.	219
Tabla A.9. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en La Rioja.	220
Tabla A.10. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en la Comunidad de Madrid.	220

Tabla A.11. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Castilla-La Mancha.	221
Tabla A.12. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en la Comunidad Valenciana.	221
Tabla A.13. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Extremadura.	222
Tabla A.14. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Andalucía.	222
Tabla A.15. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en la Región de Murcia.	223
Tabla A.16. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en las Islas Canarias.	223
Tabla A.17. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en las Islas Baleares.	224
Tabla A.18. Valores del estado del único bien incluido en la Lista Roja en Ceuta.	224
Tabla B.1. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en La Coruña	230
Tabla B.2. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Lugo	230
Tabla B.3. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Orense	231
Tabla B.4. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Pontevedra	231
Tabla B.5. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Guipuzcoa	232
Tabla B.6. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Vizcaya	232
Tabla B.7. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Huesca	233
Tabla B.8. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Teruel	233
Tabla B.9. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Zaragoza	234

Tabla B.10. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Barcelona	234
Tabla B.11. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Gerona	235
Tabla B.12. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Lérida	235
Tabla B.13. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Tarragona	235
Tabla B.14. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Ávila	236
Tabla B.15. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Burgos	236
Tabla B.16. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en León	237
Tabla B.17. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Ávila	237
Tabla B.18. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Salamanca	238
Tabla B.19. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Segovia	238
Tabla B.20. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Soria	239
Tabla B.21. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Valladolid	239
Tabla B.22. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Zamora	240
Tabla B.23. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Albacete	240
Tabla B.24. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Ciudad Real	241
Tabla B.25. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Cuenca	241
Tabla B.26. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Guadalajara	242

Tabla B.27. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Toledo	242
Tabla B.28. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Alicante	243
Tabla B.29. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Castellón	243
Tabla B.30. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Valencia	244
Tabla B.31. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Cáceres	244
Tabla B.32. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Badajoz	245
Tabla B.33. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Almería	245
Tabla B.34. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Cádiz	246
Tabla B.35. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Córdoba	246
Tabla B.36. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Granada	247
Tabla B.37. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Huelva	247
Tabla B.38. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Jaén	247
Tabla B.39. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Málaga	248
Tabla B.40. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Sevilla	248
Tabla B.41. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Sta. Cruz de Tenerife	249
Tabla B.42. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Las Palmas	249
Tabla D.1. Ficha de prueba (parte 1)	266
Tabla D.2. Ficha de prueba (parte 2)	267

## 5. DEFINICIONES Y ACRÓNIMOS

### 5.1. DEFINICIONES

#### A.

<b>ArcGIS</b>	Softwares de Sistemas de Información Geográfica.
<b>Análisis de sentimientos</b>	Análisis automático de documentos o textos capaz de clasificarlos según la connotación del lenguaje contenido.

#### B.

<b>Bien de Interés Cultural</b>	Figura jurídica de protección del patrimonio histórico.
<b>Biblioteca (informática)</b>	Conjunto de funcionalidades codificadas en cierto lenguaje.
<b>Bigrama</b>	Véase <b>N-grama</b> .
<b>Big data</b>	Conjuntos de datos masivos.
<b>Booleano</b>	Tipo de dato lógico binario (verdadero o falso, sí o no, etc.).

#### C.

<b>Centroide</b>	Centro geométrico de una figura.
<b>Chi-cuadrado (<math>\chi^2</math>)</b>	Tipo de distribución de probabilidad continua generalmente utilizada para la estimación de varianzas.
<b>Conservación</b>	Mantener la integridad o permanencia de algo.
<b>Consolidación</b>	Reunir o juntar algo quebrado y darle solidez.
<b>Contraste de hipótesis</b>	Prueba que permite determinar si puede ser tomada como válida o no una afirmación basándose en la discrepancia entre datos observados y datos esperados en una muestra.
<b>Cuartiles</b>	Valores que divide un conjunto de datos en cuatro partes iguales.

#### D.

<b>Data frame</b>	Estructura de datos bidimensional heterogénea.
<b>Data science</b>	Ciencia centrada en el estudio de los datos.
<b>Data wrangling</b>	Limpieza y transformación de los datos en bruto.

#### E.

<b>Exact match</b>	Coincidencia exacta de una cadena de caracteres.
<b>Expolio</b>	Delito de incautación del patrimonio histórico.
<b>Expresión regular</b>	Patrones para detectar combinaciones de caracteres.

## **F.**

### **Fuzzificación**

Transformar una variable real en borrosa, es decir, dotarla de grados de pertenencia a diversos valores.

## **G.**

### **Georreferenciación**

Posicionamiento de un punto en una localización geográfica.

## **H.**

### **Histograma**

Gráfica que representa datos agrupados por intervalos en forma de columnas.

## **I.**

### **Incoar**

Iniciar un proceso o expediente.

### **Information retrieval**

Actividad de obtención de información en documentos electrónicos.

### **Iterativo**

Que se repite varias veces.

## **M.**

### **Minería de datos**

Proceso de detección de patrones en conjuntos de datos.

### **Modelo relacional**

Modelo de datos agrupados por una propiedad común.

### **Motor de inferencia**

Sistema capaz de tomar decisiones en base a premisas.

## **N.**

### **N-grama**

Subconjunto de N elementos de un conjunto dado (por ejemplo, un conjunto de palabras).

## **O.**

### **OpenRefine**

Herramienta para el tratamiento de bases de datos.

### **Outlier**

Véase **Valor atípico**.

## **P.**

### **Palabra vacía**

Se denomina a los artículos, preposiciones, conjunciones, pronombres, etc.

### **Patrimonio Nacional**

Organismo dedicado al cuidado de los bienes vinculados a la Corona Española.

### **P.G.O.U.**

Herramienta para la ordenación urbanística de un territorio.

### **Pearson, Karl**

(1857-1936). Matemático británico iniciador de la disciplina de la estadística matemática.

## **R.**

### **R**

Lenguaje de programación para el análisis estadístico.



<b>Raw data</b>	Datos sin procesar.
<b>Red neuronal artificial</b>	Modelo computacional inspirado en el funcionamiento del cerebro humano.
<b>Registro</b>	Conjunto de campos de un fichero o base de datos. Se corresponde con una fila. El término “ficha” usado en este documento equivale a “registro” dentro del fichero de datos.
<b>Rehabilitación</b>	Restituir la funcionalidad de algo, especialmente de un espacio.
<b>Restauración</b>	Reparar algo que se ha deteriorado y ponerlo en el estado que antes tenía.
<b>RStudio</b>	Entorno de desarrollo para R.
<b>S.</b>	
<b>Sillar</b>	Piedra labrada que forma parte de un muro.
<b>T.</b>	
<b>Tesaurus</b>	Diccionario o lista de términos controlados que se usan para representar conceptos establecidos.
<b>Text summarization</b>	Resumen de textos automático.
<b>Trigrama</b>	Véase <b>N-grama</b> .
<b>U.</b>	
<b>Unicode</b>	Estándar universal de codificación de caracteres para su procesamiento.
<b>V.</b>	
<b>Valor atípico</b>	Observación alejada del resto de datos.
<b>W.</b>	
<b>Weather Spark</b>	Página web de informes meteorológicos.

## 5.2. ACRÓNIMOS

<b>AECT</b>	Agrupación Europea de Cooperación Territorial
<b>AEMeT</b>	Agencia Estatal de Meteorología
<b>a. C.</b>	Antes de Cristo
<b>BIC</b>	Bien de Interés Cultural
<b>BOE</b>	Boletín Oficial del Estado
<b>BRL</b>	Bien de Relevancia Local
<b>CC.AA.</b>	Comunidades Autónomas
<b>CE</b>	Constitución Española
<b>d. C.</b>	Después de Cristo
<b>df</b>	Data frame
<b>ETL</b>	Extract, Transform and Load [Extraer, Transformar y Cargar]
<b>EPA</b>	Encuesta de Población Activa
<b>GDPR</b>	General Data Protection Regulation
<b>GIS</b>	Geographic Information System
<b>IA</b>	Inteligencia Artificial
<b>IGME</b>	Instituto Geológico y Minero de España
<b>IPCE</b>	Instituto del Patrimonio Cultural de España
<b>LPHE</b>	Ley del Patrimonio Histórico Español
<b>LODP</b>	Ley Orgánica de Protección de Datos de Carácter Personal
<b>MHS</b>	Monitoring Heritage System [Sistema de Monitorización del Patrimonio]
<b>NO<sub>2</sub></b>	Dióxido de nitrógeno
<b>NO<sub>x</sub></b>	Óxidos de nitrógeno
<b>PGE</b>	Presupuestos Generales del Estado
<b>PGOU</b>	Plan General de Ordenación Urbana
<b>PIB</b>	Producto Interior Bruto
<b>RAE</b>	Real Academia Española
<b>S. A.</b>	Sociedad Anónima
<b>SEPG</b>	Secretaría de Estado de Presupuestos y Gastos
<b>S. L.</b>	Sociedad Limitada
<b>SO<sub>2</sub></b>	Dióxido de azufre

<b>UNESCO</b>	United Nations Educational, Scientific and Cultural Organization [Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura]
<b>UTF-8</b>	8-bit Unicode Transformation Format [Formato de Transformación Unicode de 8-bit]

## 6. ABSTRACT

### 6.1. INTRODUCTION

The economic and social progress that mankind has experienced since the First Industrial Revolution is the root of technological change. For many years, companies have made huge efforts and have invested heavily in technologies for data analysis and the exploitation of big data, thereby achieving a range of products and services better designed to respond to customer needs.

However, data science tools and the possibilities that they offer have not been frequently used by managers and institutions responsible for heritage conservation. Heritage is the human legacy: the one that our ancestors have left us and the one that we leave to our descendants.

Cultural heritage is an important asset that contributes to socioeconomic development and job creation in Spain. In 2019, the Spanish cultural industry employed more than 700,000 people (3.6% of total workers) representing around 3.2% of GDP.

The challenge of preserving, disseminating and bequeathing our heritage to future generations, in which the presence of digital tools is increasing, should encourage us to persevere in the study, to continue investigating the possibilities that these new technologies offer to tasks related to the preservation of the historical, artistic and cultural heritage. This work has been carried out aiming to study the possibilities offered by data analysis in the field of the conservation of historical, cultural and artistic heritage.

*Keywords:* cultural heritage, heritage preservation, data analysis, information retrieval.

## 6.2. PROJECT PLANNING

The project, whose duration has been twelve months, has been divided into six phases:

Phase 1: Objectives

Phase 2: Environmental scanning

Phase 3: Data processing and information extraction

Phase 4: Data analysis

Phase 5: Result analysis and conclusions.

Phase 6: Improvements and future research directions

The different phases are described in detail below.

## 6.3. PHASE 1: OBJECTIVES

It is essential to respond to a real need when embarking on a project of this nature which is intended to obtain the benefits of data science available to heritage conservation.

However, objectives may be conditioned by the data to which we have access for analysis, given the vast variety of possibilities offered by data analysis. Therefore, it is difficult to establish them *a priori* until the content of the database is compiled. Final objectives will be reconsidered after obtaining a more accurate view of the data.

The main objective has been to investigate "what can be done with what information". So it has been necessary to obtain a data model that offers us enough information about each patrimonial asset to be able to carry out significant analyses with them. The first thing we needed to create was a database of heritage assets. The second thing we needed was to investigate what elements can provide us with crucial information about the assets and extract them from the database we obtained. Finally we tried to find out what we can do with this information.

## 6.4. PHASE 2: ENVIRONMENTAL SCANNING

The second phase consists of a study of the environment in which the project is framed. Cultural heritage has already benefited from some applications of data science, two examples include data monitoring for preventive conservation of heritage assets and the use of big data to boost the reach of museum entities to visitors. On the other hand, data science is increasingly used in the efficient management and maintenance of civil infrastructure.

The main digitized databases of Spanish cultural heritage were sought, and after finding that important digitization work is still to be done, it was decided to work with Hispania Nostra.

Hispania Nostra is a non-profit association based in Madrid, Spain, dedicated to the defense and promotion of Spanish Cultural Heritage. Among many other lines of action there is the “*Lista Roja del Patrimonio*” (Red Heritage List), an initiative dedicated to promote abandoned heritage and one of the main databases for Spanish cultural assets. This will be our main source of data.

It works as follows: any user can provide information about new rows through a web form. These rows are assessed by a scientific committee who decides whether or not these rows should be included in the “Red List”. There are also two subcategories within the “Red List”: the “Green List” for rescued heritage, and the “Black List” for irreversibly lost heritage.

Regarding the political and legislative environment, the competences of public heritage assets are concurrent, as indicated in Articles 148 and 149 of the Spanish Constitution, which means that they are divided between the Spanish State, the Autonomous Administrations, and the Local Administrations.

In terms of financing, the General State Budgets allocate various budget items for the conservation, restoration and protection of historical heritage each year. These projects include intervention, maintenance and training plans. In the region, it is established that the financing will come straight from the budgets of several localities. Private entities can resort to patronage strategies as well as financial aid from either the State or the European Commission.

## 6.5. PHASE 3: DATA PROCESSING AND INFORMATION EXTRACTION

### 6.5.1. INITIAL TREATMENT

The third phase has consisted of processing the data to prepare it for following analyses.

For the treatment of raw data, cleaning the dataset of redundant attributes, splitting columns into multiple attributes, unifying date formats and cleaning up special characters in Spanish to avoid coding failures has been vital. The open source tool OpenRefine has been used for this.

### 6.5.2. INFORMATION EXTRACTION

Information retrieval techniques have been used to extract information. Most attributes contain texts of varying lengths in which the information rarely appears in order. Algorithms have been implemented that search more complex terms or constructions in these texts. It has been able to identify the information of interest in each attribute. This part of the project, as well as the subsequent analyses, have been developed in the R programming language due to the advantages it offers in terms of graphic representation and its large number of libraries for data mining.

The first attribute treated has been **Time** (“*Época*”), which contains information on the origin of the heritage asset. However, due to the ambiguity of the name that the attribute receives, the information contained is very uneven. Not only does this attribute contain years and centuries that can appear accompanied by BC or AD, but also historical periods or artistic movements. In addition, some rows include intervals of years and centuries, several historical periods concatenated or additional information on subsequent reforms and modifications of the heritage asset. Therefore, to treat this attribute it has been essential to evaluate all these characteristics.

To begin with, writing formats have been unified, evaluating all the possible ways in which BC and AD have been written to transform them into one and they have been stored in a new attribute called **Era** (“*Era*”). Then every year has been converted to centuries to unify the chronological unit of measurement. After that, all the centuries referring to reforms have been removed. The order in which is done is essential to avoid treatment failures.

Subsequently, the regular expressions of the centuries and century intervals have been detected and stored in chronological order in a new attribute. Next, a dictionary of historical periods and artistic movements has been implemented, associating them with the centuries they comprise in the form of numerical vectors. An algorithm has also been developed that uses set theory to transform a succession of historical terms into an interval or set of intervals, applying either union or intersection of intervals depending on whether the periods have overlapped or not.

Additionally, information on the construction date was examined in another attribute called **History** (“*Historia*”) for those rows that covered historical periods, in search of more accurate data for **Time**. With all this, three new attributes are generated: **Building.Century** (“*Siglo.Edificación*”) (full building dates), **Beginning.Century** (“*Siglo.Inicio*”) (when the construction process started) and **Ending.Century** (“*Siglo.Fin*”) (when the construction process ended).

The second attribute treated has been **Interventions** (“*Intervenciones*”), which contains information on whether a property has been altered without involving its removal from the “Red List”. For this attribute, we created a binary attribute that inserts the word “Yes” if it has been intervened, and a second attribute that stores the year of intervention if indicated.

The third attribute treated has been **Removal.Reason** (“*Motivo.de.la.Retirada*”), which contains information on the reasons why an asset has left off the “Red List”, either to enter the “Green List” or the “Black List”. The removal reasons included in the attribute have been identified, some of which are registered as restorations, rehabilitations, consolidation works, disappearance of external threats or demolitions. The assets that remain on the “Red List” will have this attribute marked as empty.

The fourth attribute treated was **Property** (“*Propiedad*”), which contains information about the owners of the asset. The most frequent types of owners have been identified within this attribute, as well as several others considered for future updates (City Council, Company, Bishopric, Archbishopric, Private, etc.). It has also been considered whether these assets are public or private. Accordingly, the owner category has been stored in a new attribute while a second new attribute contains whether it is a public or private asset.

The fifth attribute treated was **Degree.of.Protection** (“*Grado.de.Protección*”), which contains information regarding the protection of the asset, if any. The main types of protection have been identified, including the BIC (Property of Cultural Interest), BRL (Property of Local Relevance), PGOU (General Urban Planning Plan), diverse inventories and unprotected, among others. The different types of protection have been extracted and stored.

The sixth attribute treated was the **State.of.Conservation** (“*Estado.de.Conservación*”), which contains a text of variable length describing the current condition of the heritage asset. The goal has been to quantify this variable: to transform the texts into bounded numerical values that define the “level of bad condition”. A dictionary of terms has been implemented that includes “ruin”, “deteriorated”, “destroyed” or “damaged” among others, with values associated according to their importance and magnitude. Besides, a hierarchy has been established so that a text with redundant information does not add more points than a more concise one just because it has more adjectives indicating the same idea. In addition, the appearance of quantifiers such as “much”, “great” or “very” that highlight the bad condition points have also been evaluated. Lastly, binary variables have been generated to indicate whether the assets are plundered, abandoned, covered in vegetation or any other factor that, although they are important to the description of the asset condition, this does not determine their level of deterioration.

The eighth attribute treated was **Description** (“*Descripción*”). It consists of a text of variable length that contains several features of the heritage asset. From this attribute it has been possible to extract information about the construction materials and the number of floors of certain heritage assets. It was also attempted to compile data regarding the square metres of constructed area of the heritage assets, but it could not be carried out due to lack of information.

The ninth and last attribute treated was **History** (“*Historia*”) which contains a text of variable length, obtained in many cases from online bibliographic sources. In addition to the century of construction (as we have already seen in the **Time** attribute), this attribute has



been used to extract information about the heritage assets' use, as well as information about important historical events that it may have experienced. Knowing the use may be of interest as it can affect the state of conservation of the heritage asset (for instance, many ruins were and are used as stables). In the end, little information could be extracted about their current use, so it was finally discarded. Important events such as Spanish confiscations, the Spanish War of Independence, Spanish Civil War, earthquakes, floods or fires have been assessed and stored as binary attributes.

### 6.5.3. INCLUSION OF EXTERNAL DATA

Several attributes have been added to the dataset that were not treated since their information was already arranged. These attributes are the ones referring to location, date of inclusion on the "Red List", and in some cases, date of removal from the "Red List". Unfortunately, the coordinate values of the heritage assets were not available, so certain external data (such as weather) will have a more generalized level of accuracy.

Finally, a reduction on the dimension of the dataset has been carried out, removing empty rows, character tests or those rows referring to natural heritage, which cannot be classified with the same criteria as architectural heritage and will therefore be excluded from this draft. This has left the database with a total of 993 rows.

At this point, we returned to phase two and included external data that could be pertinent to the paper at hand.

**Climatological data.** Temperature, humidity and average rainfall can be decisive factors in the wear of a building. The data has been obtained from the Weather Spark website. Single values have been added for each province.

**Air quality measurements.** NO<sub>2</sub>, NO<sub>x</sub> and SO<sub>2</sub> particles are the most frequent. These pollutants produced by the combustion of hydrocarbons leave black crusts on building façades. The data has been obtained from the INE (National Statistics Institute). Single values have been added for each province.

**Socioeconomic variables.** CPI and the unemployment rate may be factors that determine the budget the public administrations can invest in the restoration of their heritage assets. The socioeconomic data has also been obtained from the INE (National Statistics Institute). Single values have been added for each province.

**Seismographic rate.** This data was used to study how the proximity to seismographic zones may affect the state of conservation. The provinces' seismographic rates were obtained by calculating the centroid of each

Spanish province and allocating them the value of the nearest seismographic region.

**Proximity to the sea.** The proximity to the sea affects the state of conservation of nearby properties, due to the corrosion characteristics of the marine breeze. For this, all the coastal municipalities coinciding with a 200 metre offset of the Spanish shores have been added to the dataset.

**Number of Hispania Nostra members by province.** This data was used to determine whether the number of rows for each province is related to the number of province members. In other words, if all cases where there are fewer partners, then there are also fewer registered rows.

**Geographical extent.** This data was used to determine the number of rows for each region is associated with its extension area.

After processing and including all the external data, the dataset consists of 993 rows and 72 columns, of which only 171 rows are included in the “Green List” and 8 rows are in the “Black List”. This amount of elements will be insufficient to perform tests such as artificial neural network classification or cross validation.

## 6.6. PHASE 4: DATA ANALYSIS

### 6.6.1. EXPLORATORY ANALYSIS

The data analysis phase begins with an exploratory analysis of the attributes. This consisted of analyzing the geographical distribution of the concentration of “Red List”, “Green List” and “Black List” rows at both provincial and regional scales. The first results show a high concentration of cultural assets in the autonomous community of Castilla y León for both the “Red List” and the “Green List”. The province of León stands out as the one with the most heritage assets on the “Red List”, although Zaragoza, Aragón, ranks first on the “Green List”. The properties of the “Black List” are few and far between, located throughout Spain.

Following that, the main descriptive measures of the numerical attributes related to the heritage assets state of conservation have been obtained. These have been studied at provincial, regional and national levels. Other numerical attributes have also been studied, such as those related to weather, pollution, socioeconomic factors or the number of Hispania Nostra members. Finally, a study of the main categories of the nominal attributes has been made: the **Degree.of.Protection**, **Property** and **Removal.Reason**.

### 6.6.2. LINEAR CORRELATIONS

The second section of the analysis consisted of carrying out several studies of Pearson's linear correlations to determine the relationship between various numerical attributes. We studied whether there is a relation between the number of "Red List" rows and the number of Hispania Nostra members or the geographical extent of their location. The link between the number of rows and the weather data has also been studied. Subsequently, the correlation between the weather data and the state of conservation has been studied. The state of conservation has also been studied in correlation to the air quality data, the provincial and regional levels and the socioeconomic data. The results of the correlations have been generally very poor due to the high dispersion of the data and the sampling error of the same. Nevertheless, the result of the analysis between **State.of.conservation** and socioeconomic data stands out, confirming that with a higher unemployment rate the heritage assets tend to be in a worse condition, and with a higher CPI, they tend to be in a better condition..

### 6.6.3. HYPOTHESIS CONTRAST

Finally, multiple Pearson chi-squared tests have been carried out to make hypothesis testing to determine the dependency between several variables. For this it has been required to transform some numerical attributes to nominal ones, dividing their range of values into several categories. The tests with the most significant results have been: **Building.Century** with the existence of structural risk, **Degree.of.Protection** with **State.of.Conservation**, **Property** with **State.of.Conservation**, **Risk.of.demolition** with **State.of.Conservation**, **CPI** with **State.of.Conservation** and the seismographic rate with **State.of.Conservation**.

## 6.7. PHASE 5: RESULT ANALYSIS AND CONCLUSIONS

### 6.7.1. ANALYSIS OF THE RESULTS

Although all the data analysis has not reached the desired results, due to the various limitations identified in the data sample, we obtained some clear ideas.

Thanks to the correlation analysis, it has been possible to determine that the number of rows included in the dataset is dependent on the geographical extent of the region in which they are located and, to a lesser extent, on the number of Hispania Nostra members in that region, and the number of members is dependent on the population of the area. This may lead us to suppose that there is a greater quantity of heritage assets at risk of disappearance of which we are not aware since it is not on the "Red List", in those regions

where the number of members of Hispania Nostra is lower. On the other hand, it has been verified that unemployment is slightly related to the state of conservation, in such a way that those regions with a higher unemployment rate tend to present heritage assets in a worse state of conservation. However, the number of rows is not related to this, which makes one wonder if it is just a coincidence of the data sample.

Subsequently, the correlations have revealed that heritage assets located in areas with a higher CPI tend to be on the "Green List". This indicates that there is indeed a relation between the socioeconomic factors of a region and how neglected the heritage assets are. It has also been verified that the state of conservation of a heritage asset is dependent on the maximum temperature, the maximum temperature variation and the probability of precipitation and humidity, revealing that a drier, hotter weather or one with sudden variations of temperature contributes to the deterioration of the heritage assets.

Regarding ownership, the probability of public property being included on the "Green List" is higher than private and have more chances to be BIC tagged, again confirming a higher likelihood of being green. Public property assets are also more likely to have the BIC label, confirming that they are more likely to be moving to the "Green List". For its part, privately owned assets are more likely to lack some type of protection assets that lack protection are more unlikely to be on the "Green List".

In short, it can be said that the results obtained are lower than our initial expectations. In this study many variables have given completely opposite results than the ones from other studies, demonstrating that they indeed have a crucial effect on the deterioration and conservation of the heritage asset. However, the analysis does not represent the total project objectives. The treatment of the data continues to be a part of vital importance.

### **6.7.2. CONCLUSIONS**

The objectives initially set have been partially met. On one hand, the treatment of the dataset and the transformation of its content to obtain the essential information for its analysis has been quite satisfactory. Many improvable aspects of the original dataset have been corrected and quite a lot of information has been categorized with substantial success. However, the quantification part has been the least satisfactory of the process in view of the results obtained. A priori, there is no precise and reliable way to quantify the condition state of a heritage asset, at least based on descriptions written by users.

On the other hand, the results show us that it is not possible to carry out an exhaustive analysis with the available data. A considerable part of the results obtained denies the existence of relation between the state of conservation of heritage assets and external factors, when we know from other previous studies that those factors are decisive. This, as has been glimpsed during the different tests carried out, may be due primarily to three reasons:

The first is the one mentioned in the previous paragraph, which refers to the method of quantifying the state of deterioration, which is one of the key factors of a heritage asset.

The second reason is the lack of precision while obtaining external data and merging them with the heritage assets datarow. Taking a single weather or air quality data value for a whole province leads to a high imprecision that leads to the alteration of the results.

The third reason is the data shortage and heterogeneity of the data sample. The dataset is 993 rows by 70 columns. Of those assets, around 17% are part of the Green List and less than 1% are included in the Black List. The data sample also shows little homogeneity at the geographical level. All this has introduced a high sampling error.

Special emphasis needs to be put on the urgency to digitize heritage databases. Not only because it is one of the main ways of preserving and promoting it, but also because it would allow more in-depth studies. The work of Hispania Nostra with the “Red List” is admirable and necessary. However, there is still a lot of undetected heritage that remains off the list. This study highlights the importance of establishing a systematization in the digitization process of the heritage assets, and the implementation of a thesaurus that allows one to control the ambiguity of the terms used in studies and descriptions; and the parameterization of properly agreed factors that allow us to determine the state of conservation of the assets.

## 6.8. PHASE 6: IMPROVEMENTS AND FUTURE RESEARCH LINES

Once the project is completed, the results of the study will be available to the Hispania Nostra association. The project should be open to the inclusion of new “Red List” data. The data storage method used by the “Red List” must be modified to improve it and avoid some failures found during the analysis. Subsequently, an attempt will be made to improve the system so that the analysis can give better results, for example obtaining coordinate values on a specific geographic coordinate system for a more accurate inclusion of external data.

Possible lines of research for the future are also proposed. For instance, the implementation of an expert system for the inclusion of new rows in the Red List or the use of neural networks for classification with additional data.

## 7. INTRODUCCIÓN

El cambio tecnológico se sitúa en la raíz del progreso económico y social que la humanidad ha experimentado desde la Primera Revolución Industrial hasta nuestros días. Desde hace ya muchos años las empresas han puesto una parte importante de sus esfuerzos y han realizado fuertes inversiones en tecnologías para el análisis de datos y en la explotación del *big data*, logrando con ello una oferta de productos y servicios mejor diseñados para responder a las necesidades de sus clientes.

Sin embargo, las herramientas de la ciencia de datos y las posibilidades que pueden ofrecer a los gestores e instituciones responsables de la conservación del patrimonio han sido muy poco utilizadas. El patrimonio es el legado humano: el que nos han dejado nuestros antecesores y el que dejamos a nuestros sucesores.

España es el tercer país del mundo con mayor patrimonio según la UNESCO, solo por detrás de China y de Italia [1]. Además de su evidente valor histórico, estos bienes culturales son un importante activo que contribuye al desarrollo socioeconómico y a la generación de empleo. La industria cultural española emplea a más de 700.000 personas, un 3,6% del total de trabajadores según la EPA del año 2019 [2], y representaba en torno al 3,2% del PIB en el año 2017 (últimos datos publicados) [3].

El reto de conservar, difundir y legar nuestro patrimonio a las generaciones futuras, en las que la presencia de lo digital será cada vez mayor, nos debe impulsar a perseverar y seguir investigando las posibilidades que estas nuevas tecnologías pueden ofrecer para la preservación del patrimonio histórico, artístico y cultural.

*Palabras clave:* patrimonio cultural, conservación del patrimonio, análisis de datos, búsqueda y extracción de la información.

*“Cuidad vuestros monumentos y no tendréis que restaurarlos”*

- John Ruskin (1819-1900) [4]

## 7.1. MOTIVACIÓN Y ENFOQUE

Si bien este trabajo podría servir como un manifiesto para la defensa de un futuro en que las técnicas de *data science* y *big data* sean empleadas en cualquier ámbito del saber, camino en el que ya vamos, deseamos que sea, en cierto modo, una reivindicación de la importancia del compromiso de todos los ciudadanos con nuestro patrimonio histórico, cultural y artístico, independientemente de la formación de cada individuo.

El mundo del patrimonio cultural, histórico y artístico y el mundo de la Ciencia de Datos no se encuentran desligados. Una parte esencial del fomento de la cultura es la difusión de la misma, y es frecuente que se recurra al análisis de datos, entre otras herramientas, para impulsar y optimizar la forma en que las entidades llegan al público. Este, por ejemplo, es el caso de algunos museos. En 2017, coincidiendo con la exposición "*Piedad y terror en Picasso. El Camino a Guernica*" albergada en el Museo Reina Sofía de Madrid, y en colaboración con Telefónica, se llevó a cabo el primer estudio de *big data* desarrollado por un museo español. Visitantes, redes sociales, meteorología e incluso la monitorización del movimiento de personas en las proximidades del Reina Sofía permitieron conocer el impacto que tuvo la exposición. Un aspecto interesante fue identificar patrones de comportamiento en los visitantes según su nacionalidad [5]. Todo esto persigue el fomento del turismo cultural.

En lo referente a su preservación, el patrimonio cultural podría beneficiarse de los avances de la ciencia de datos en otros ámbitos, y tratar de ir más allá de la monitorización, ya extendida para la conservación preventiva de los bienes. La amplitud de las posibilidades que ofrece el análisis de datos constituye un horizonte en el que queda mucho por explorar, especialmente en el campo de la conservación.

La pregunta que hay que hacerse a la hora de sacar adelante un proyecto de este tipo es si existe una necesidad que cubrir o si hay un interés real con respecto al beneficio que pueda reportar. Como pronto veremos tenemos una fuerte legislación en pos de la protección del patrimonio Histórico y hay un considerable número de entidades dedicadas a la defensa, mantenimiento y promoción de dicho patrimonio. No obstante, la conciencia social es un factor clave para dar respuesta a dicha pregunta, ya que, aunque el patrimonio es algo de todos, en ocasiones el compromiso y la implicación con el mismo parece cosa de unos pocos. Concienciar sobre la importancia que tiene el patrimonio para nuestra identidad, y su preservación, es esencial a todos los niveles. Este es el punto fuerte sobre el cual se sostiene este proyecto.

La metodología a seguir será la propia de un proceso ETL, es decir, de extracción, transformación y carga de los datos. Será necesario obtener acceso a diversas fuentes de datos que contengan información de interés para el estudio, y estos datos deberán ser unificados, tratados, transformados para obtener la información necesaria para su posterior análisis.



## 7.2. OBJETIVOS

A la hora de afrontar un proyecto de investigación es imprescindible plantearse qué preguntas buscamos responder. Así pues, si lo que se pretende con este proyecto es poner todos los beneficios de la ciencia de datos a disposición del mundo de la conservación del patrimonio, el objetivo final de todo este trabajo no debería ser otro que contribuir a la propia conservación.

Se pueden establecer unos objetivos generales, tales como clasificación de los bienes patrimoniales según su estado de deterioro y su riesgo de desaparición. También se podría tratar de desarrollar un sistema de predicción que permitiese estimar en cuánto tiempo podrían desaparecer los bienes de no ser intervenidos, teniendo en cuenta el estado, los factores intrínsecos (como su antigüedad, o los materiales de los que está compuesto), y las condiciones a las cuales se ven sometidos, logrando así determinar la urgencia con la que deben ser intervenidos. Sin embargo, este proyecto se presenta tan ambicioso como poco factible.

El principal objetivo será investigar “qué se puede hacer con qué información”. Llegar a un conjunto de atributos que nos ofrezca la suficiente información de cada bien patrimonial como para poder realizar análisis significativos con ellos. Para conseguirlo lo primero que necesitamos es una base de datos de bienes patrimoniales. Lo segundo que necesitamos es investigar qué elementos pueden ofrecernos información crucial sobre los bienes y extraerlos de la base de datos que obtengamos. Lo tercero será averiguar qué podemos hacer con dicha información. Así pues, para empezar, se debe obtener una base de datos. Ahí es donde entra Hispania Nostra.

## 7.3. HISPANIA NOSTRA

Hispania Nostra es una asociación sin ánimo de lucro, con sede central en Madrid, dedicada a la defensa y promoción del patrimonio cultural español, fundada en 1976 y declarada de utilidad pública [6]. Entre sus muchas líneas de actuación existe la denominada Lista Roja de Patrimonio, una iniciativa dedicada a la difusión del patrimonio abandonado. En la Lista Roja se recogen aquellos elementos del patrimonio cultural que se encuentran en riesgo de desaparición con el objetivo de darlos a conocer y lograr su restauración [7]. La lista está a disposición de toda la ciudadanía a través de su página web. Es además una de las principales bases de datos públicas para la difusión de bienes patrimoniales inmuebles españoles.

Cualquier ciudadano puede enviar una ficha (un registro) rellenando un formulario en el que se recoge información básica sobre el elemento patrimonial en cuestión: nombre, localidad, provincia, comunidad autónoma, época, etc., así como una breve descripción del bien patrimonial, su historia y su estado de conservación entre otros. Esta información se almacena en un fichero de datos. Posteriormente, un comité científico, conformado por diferentes expertos en diferentes áreas, estudia e investiga cada una de las fichas y concluye si el bien patrimonial debe o no ser incluido en la Lista Roja. Sin

embargo, y puesto que la Lista Roja persigue la difusión y defensa del patrimonio con el fin de salvar estos elementos en riesgo, existe una segunda lista denominada Lista Verde. En ella se almacenan aquellos bienes patrimoniales que, habiendo estado en la Lista Roja, han sido intervenidos y debidamente salvados. Así pues, un bien de la Lista Verde podría volver a entrar en la Lista Roja si se considerase nuevamente en riesgo. Finalmente, existe la posibilidad de que los bienes patrimoniales, al no ser intervenidos, terminen desapareciendo o alcancen a un estado irreversible. Estos bienes, entonces, salen de la Lista Roja para almacenarse en una tercera lista, denominada Lista Negra. Es importante tener en cuenta que la Lista Roja no tiene un seguimiento o mantenimiento exhaustivo en cuanto a la actualización de sus datos, dados los medios de que dispone Hispania Nostra en la actualidad. Es decir, un bien podría haber sido restaurado, que si la noticia no fuese difundida y no llegase a oídos de la asociación, seguiría en la Lista Roja. Aunque las fichas están aprobadas por el comité científico, muchos de los contenidos del fichero de datos carecen de uniformidad al estar cumplimentados los distintos campos por gente de a pie. Todo esto representó un gran desafío desde un primer momento.

Hay que mencionar que el patrimonio natural tiene también cabida en la Lista Roja. Sin embargo, no es posible valorar con los mismos criterios los mundos tan dispares del patrimonio arquitectónico y el patrimonio natural, pese a que ambos pertenecen al grupo del patrimonio inmueble. Por esa razón, esta es una parte de la Lista Roja en la que no entraremos en este proyecto. No obstante, sí que se incluyen aquellos bienes asociados al patrimonio natural que son jardines o huertas.

Es importante aclarar que, en el presente documento, el término “Lista Roja” se usa indistintamente para referirse al propio fichero de datos proporcionado por Hispania Nostra en su totalidad; así como también al subconjunto de datos que poseen la categoría de “Lista Roja”, diferenciándose de aquellos que se encuentran en las categorías de “Lista Verde” o “Lista Negra”.

## 7.4. HERRAMIENTAS

Para el desarrollo de este proyecto se ha utilizado mayoritariamente el lenguaje de programación R y el entorno de desarrollo RStudio. El motivo de la elección de este lenguaje de programación frente a otros como Python, cuyo uso está tan extendido en la actualidad, es por el enfoque estadístico y su gran cantidad de bibliotecas (también denominadas librerías) para la minería y el análisis de datos. Además, R posee herramientas de representación gráfica muy frontales, mientras que Python tiene un enfoque más dirigido al desarrollo como lenguaje de uso general. Ciertamente una gran parte del tratamiento de los datos podría haberse realizado con Python o R indistintamente. Sin embargo, como el objetivo del TFG ha sido más la investigación que el desarrollo, se ha preferido usar un mismo entorno de trabajo para todo el proceso. Esto convierte a R en el lenguaje idóneo para desarrollar el proyecto.

Además, de manera puntual se ha recurrido al uso de OpenRefine, una herramienta *online* de código abierto diseñada para realizar tareas de *data wrangling*. Se ha usado para explorar el *raw data*, identificar errores en la estructura y los contenidos del fichero de datos y preparar estos datos para ser correctamente leídos por RStudio [8].

Se ha hecho uso del software ArcGIS para trabajar con datos geográficos. Los software GIS son sistemas de información geográfica que trabajan con datos asociados a referencias espaciales. Esto permite la manipulación, el análisis y la organización de grandes cantidades de datos ligados a regiones o puntos geográficos, entre otras funcionalidades.

Finalmente ha sido imprescindible disponer de un pack de Microsoft Office para trabajar con Microsoft Excel, ya que la mayoría de los datos han sido importados desde ficheros .csv o .xlsx o .xml.

## 7.5. ESTRUCTURA DEL DOCUMENTO

Parte 1: Comprende el análisis del entorno en que se sitúa el proyecto, establece el alcance del mismo y expone la planificación detallada por fases, tiempo de duración y un desglose de los costes.

Parte 2: Constituye el bloque de mayor extensión del proyecto y comprende el tratamiento inicial del *raw data*, la extracción de información útil basándose en los contenidos de los datos cedidos, los procesos seguidos para su obtención y sus correspondientes baterías de pruebas.

Parte 3: Este apartado comprende la inclusión de datos externos que han sido valorados como información de interés para el motivo de la investigación.

Parte 4: En esta sección se exponen los primeros análisis exploratorios de los datos, medidas descriptivas, así como distribuciones geográficas de los diferentes atributos.

Parte 5: En esta sección se encuentran los análisis más avanzados, los cuales han consistido en el estudio de correlaciones lineales entre los diferentes atributos y posteriormente test de hipótesis chi-cuadrado de Pearson.

Parte 6: Incluye las conclusiones del proyecto, así como una valoración de la solución elegida. Propone mejoras y próximas líneas de investigación de interés

Parte 7: Referencias.

Parte 8: Anexos.

## 8. ESTUDIO DE VIABILIDAD DEL PROYECTO

En las siguientes páginas se detalla el establecimiento del alcance del proyecto, un estudio del entorno actual en los ámbitos de la conservación del patrimonio histórico y el mantenimiento de diversas infraestructuras e instalaciones, un análisis del entorno socioeconómico de los últimos años y del marco legislativo que ampara al patrimonio histórico, el marco normativo y técnico en el cual se desarrollará el propio proyecto y, finalmente, una identificación de las partes interesadas del mismo.

### 8.1. ALCANCE DEL PROYECTO

Recapitulando, el objetivo central de este proyecto será desarrollar un sistema que nos permita transformar datos de bienes patrimoniales históricos para su posterior análisis. Esto se puede ver como una primera parte de un futuro proyecto para la predicción del riesgo de desaparición de dichos bienes como consecuencia de los parámetros intrínsecos al propio bien cultural o factores externos a los cuales se ve sometido. Se establece como meta primordial la obtención de un conjunto de datos que documente las características y permita evaluar las condiciones de dichos bienes patrimoniales para su análisis. Se dispone de un periodo de un año para la elaboración del proyecto, durante el cual deberá no solo implementarse el sistema que mejor trate los datos, sino realizarse una serie de análisis que permitan obtener resultados concluyentes y determinar la viabilidad de futuras fases del proyecto. Además, este proyecto debe suponer ante todo un beneficio para Hispania Nostra que les permita mejorar la gestión de la Lista Roja.

### 8.2. ESTUDIO DE LA SITUACIÓN ACTUAL

Para comenzar se llevó a cabo una búsqueda sobre proyectos ya existentes que puedan asemejarse al que me propongo desarrollar. Es importante conocer a fondo los antecedentes y analizar la situación a nivel nacional e internacional. Podemos clasificar estos proyectos en dos grandes grupos: el uso del análisis de datos para la conservación de infraestructuras (no patrimoniales) y el uso de análisis de datos en el campo del patrimonio histórico.

El análisis de datos, entre sus crecientes aplicaciones, es frecuentemente aplicado al mantenimiento preventivo de elementos arquitectónicos o estructurales. Un ejemplo de un ámbito en que podemos encontrar estas técnicas aplicadas es en sistemas de gestión y mantenimiento de carreteras. Sobre esto podrán encontrarse en la bibliografía varios proyectos en los que se ha recurrido al análisis de datos georreferenciados o captación de datos del entorno mediante sensores situados en vehículos [9]. Además, existen también numerosas entidades que invierten gran parte de sus esfuerzos en la monitorización de parámetros ambientales en edificios u otro tipo de

instalaciones con el objetivo de llevar a cabo un mantenimiento predictivo de los mismos. Estas técnicas, como veremos a continuación, son unas de las más frecuentemente utilizadas en el área de investigación de datos sobre el patrimonio histórico. La parametrización de datos, además de la conservación y el mantenimiento preventivo, puede perseguir también fines distintos, tales como la eficiencia energética o la determinación de patrones de consumo. Algunos ejemplos de empresas dedicadas a ello en nuestro país son ENVIRA IoT, Iternova, Tyris AI, QUBO Arquitectura o Gerencia Energética S.L..

En el campo del patrimonio histórico y cultural, las principales aplicaciones del análisis de datos giran en torno a la monitorización de datos en tiempo real. Estos suelen ser parámetros ambientales vinculados con el riesgo de deterioro de los bienes patrimoniales. Factores tales como la humedad, la temperatura, la cantidad de radiación solar a la que se ven expuestos o la presencia de organismos, tales como xilófagos, dañinos para los materiales empleados. Esta información se recopila mediante la instalación de detectores cuidadosamente repartidos por diferentes puntos del bien patrimonial que suelen pasar inadvertidos a los ojos de los visitantes [10].

De entre las diversas entidades dedicadas a la investigación de estas técnicas aplicadas al Patrimonio, destaca la labor la Fundación Santa María la Real, entre cuyas líneas de actuación destacan proyectos como MHS-EnerCon2 (*Monitoring Heritage System - Energía, Conservación y Confort*), SHBuildings (*Smart Heritage Buildings*) o SHCity (*Smart Heritage City*), en los cuales se aplican las técnicas de monitorización de parámetros ambientales para la conservación de bienes o conjuntos patrimoniales inmuebles.

La AECT, o Agrupación Europea de Cooperación Territorial, es un organismo con personalidad jurídica propia, establecido por el Parlamento y el Consejo Europeo, dedicado a proyectos de cooperación transfronteriza. A través de programas como el Interreg Sudoe (*Programa de Cooperación Territorial del Espacio Sudoeste Europeo*), se han realizado actividades vinculadas con la monitorización de datos en bienes patrimoniales [11]. Precisamente el proyecto Smart Heritage City, que llevó a cabo la instalación de sensores en la muralla de Ávila, se enmarca dentro de dicho programa. Otro proyecto de cooperación transfronteriza para la conservación del patrimonio histórico enmarcado en nuestro país es Románico Atlántico, que promueve la conservación, restauración y difusión de templos románicos, y que se extiende a lo largo de la zona Noroeste de España y el Norte de Portugal.

Hallar una base de datos y conseguir tener acceso a ella constituyó el primer gran reto durante este proyecto. Para ello fue imprescindible investigar sobre la existencia de bases de datos sobre patrimonio histórico, digitalizadas y de cualquier institución española, que pudiesen servir al objeto de esta investigación. La gran mayoría de las administraciones e instituciones, entre cuyas competencias se encuentra el cuidado o la gestión de bienes patrimoniales, disponen de bases de datos [12]; no obstante, el acceso a la totalidad de sus contenidos está restringido en algunas de ellas. Además, su tamaño sería inabarcable para un proyecto de esta índole. Por estas razones, se buscó trabajar

conjuntamente con una entidad que pudiera verse beneficiada con este proyecto, que, como ya hemos comentado, será Hispania Nostra.

En el ámbito internacional se han realizado algunos estudios en torno al uso de analítica para la preservación, digitalización y virtualización del patrimonio histórico [13], así como existe un Máster en Ciencia de Datos para el Patrimonio Histórico ofertado por la University College de Londres.

### 8.3. ENTORNO SOCIOECONÓMICO

La gestión y el reparto de competencias del patrimonio no es un tema sencillo de tratar. El Artículo 149 de la Constitución Española (en adelante CE) dictamina que el Estado posee competencia exclusiva sobre múltiples materias, entre las que se encuentra la defensa y protección del “patrimonio cultural, artístico y monumental español”, y añade “sin perjuicio de su gestión por parte de las Comunidades Autónomas” (149.1.28.<sup>a</sup> CE) [14]. Por otro lado, el Artículo 148 añade que las CC.AA, podrán asumir competencias en el “patrimonio monumental de interés para la Comunidad Autónoma” (148.1.16.<sup>a</sup> CE) [15]. Todo esto queda reflejado también en los Artículos 2 y 7 de la Ley del Patrimonio Histórico [16]. Así pues, las competencias de los bienes patrimoniales públicos quedan repartidas entre el Estado y las Comunidades Autónomas.

La Ley Básica de Régimen Local establece en su Artículo 27 que tanto la Administración del Estado como las CC.AA podrán delegar competencias tales como la “gestión de instalaciones culturales de titularidad de la CC.AA o del Estado”. Por otro lado, el Artículo 25 establece que los municipios “ejercerán en todo caso como competencias propias” en materias, entre las que se encuentra la protección y gestión del patrimonio histórico [17]. De esta manera, la competencia es concurrente, es decir, participan tanto la Administración del Estado, las administraciones autonómicas y las administraciones locales.

Patrimonio Nacional, por su parte, funciona como un ente público con personalidad jurídica propia, adscrito al Ministerio de Presidencia, dedicado al mantenimiento de los bienes de la Corona cedidos al Estado, y cuyo presupuesto depende de los PGE [18].

En lo que respecta al Patrimonio bajo propiedad privada, este es gestionado por las entidades propietarias de los bienes patrimoniales, ya sean particulares, empresas o instituciones como la Iglesia. No obstante, los bienes de la Iglesia y todos aquellos que se vean sometidos por la LPHE, forman parte de los bienes patrimoniales del Estado y se ven sujetos a lo dispuesto en el 149.1.28.<sup>a</sup> CE [19], por lo que nunca quedan exentos de ciertas obligaciones recogidas en la LPHE.

En los informes económicos y financieros de Presupuestos Generales del Estado de los últimos años, los cuales se pueden consultar en la sede electrónica del SEPG, podemos observar las fluctuaciones que ha ido experimentando la partida destinada a Cultura y a sus múltiples programas. Nosotros nos centraremos en tres de esos

programas, “conservación y restauración de bienes culturales”, “protección del Patrimonio Histórico” y “administración del Patrimonio Histórico-Cultural”.

El programa de “conservación y restauración de bienes culturales” da financiación al IPCE y tiene la responsabilidad, entre otras, de desarrollar y ejecutar una serie de planes de actuación hacia los bienes inmuebles del patrimonio histórico [20]. Entre ellos encontramos el Plan Nacional de Conservación Preventiva, el Plan Nacional de Investigación en Patrimonio Cultural, el Plan Nacional de Educación y Patrimonio, y un largo etcétera de planes de formación, investigación y protección. En cuanto al programa de “protección del Patrimonio Histórico” está gestionado por la Dirección General de Bellas Artes a través de la Subdirección General del Patrimonio Histórico y su objetivo es el Patrimonio Histórico Español, para lo que mediante diversas líneas de actuación [21]. El programa de “administración del Patrimonio Histórico-Cultural” da financiación íntegramente a Patrimonio Nacional. En los últimos años, las partidas destinadas a Cultura han experimentado un crecimiento, situándose en torno a los 953 millones en 2019.

La financiación de las Comunidades Autónomas queda recogida en los Artículos 156, 157 y 158 de la CE [22], mientras la Ley Reguladora de las Haciendas Locales establece los fundamentos de la financiación local [23]. Dentro los presupuestos de cada uno se encuentran las correspondientes partidas para Cultura de la región, sobre las cuales, tal y como hemos visto, recae una parte importante de las competencias de protección y mantenimiento del patrimonio histórico. Por su parte, las entidades privadas recurren frecuentemente a estrategias como el patrocinio o campañas de mecenazgo para financiar muchas de las actividades necesarias para la conservación del patrimonio. Adicionalmente existen planes de ayudas y subvenciones para el patrimonio cultural por parte del Estado, así como ayudas europeas al Patrimonio [24].

## 8.4. MARCO REGULATORIO

El marco normativo legal en que se encuadra este proyecto no aplica restricciones de ningún tipo. Los datos de los cuales se dispondrá son de carácter público. No se trabajará con ningún dato de carácter personal, por lo que el proyecto no se ve sujeto a las regulaciones de la normativa Europea de protección de datos (GDPR) ni a la Ley Orgánica de Protección de Datos (LODP).

El marco normativo técnico no es más estricto que el caso anterior. Durante el desarrollo, las únicas restricciones con las que podremos toparnos son las inherentes a los datos o a la capacidad de cómputo del equipo informático con el que se trabaje. El rigor con el que la información ha sido plasmada en los datos, la capacidad de extracción de información, la precisión con la que se puedan incluir datos externos serán determinantes para la fiabilidad de los resultados. A la hora de plantear la solución del problema se han valorado las decisiones tomadas en otros proyectos por las diversas entidades que hemos visto anteriormente, las cuales cuentan con el respaldo de sus propios expertos en materia de patrimonio y conservación. Nuestro conjunto de atributos tratará de valorar

todos aquellos factores que estos proyectos consideran relevantes para la catalogación y conservación de los bienes patrimoniales. La elección de los datos y el tratamiento de los mismos se basará en ello. La solución elegida además valora en todo momento la precisión y fiabilidad de los datos con los que se trabaja, el porcentaje de datos extraídos mediante las técnicas empleadas, y el código requerido para ello.

## 8.5. IDENTIFICACIÓN DE LAS PARTES INTERESADAS

Constituyen las partes interesadas o *stakeholders* de un proyecto aquellos grupos de personas o entidades vinculadas al proyecto. Podemos discernir entre dos tipos: los *stakeholders* internos y los *stakeholder* externos.

INTERNOS	EXTERNOS
<b>AUTOR DEL PROYECTO</b> LUIS MONGE SANZ DE BREMOND	<b>ENTIDADES DEDICADAS A LA GESTIÓN DEL PATRIMONIO</b>
<b>DIRECTORES DEL PROYECTO</b> DR. D. ANTONIO BERLANGA DE JESÚS DÑA. ANA CARRO ROSSELL	<b>PROPIETARIOS DE LOS BIENES CULTURALES PRIVADOS</b>
<b>PROPIETARIOS DE LOS DATOS</b> HISPANIA NOSTRA	<b>ENTIDADES DEDICADAS A LA RESTAURACIÓN Y CONSERVACIÓN DEL PATRIMONIO</b>

Figura 8.1. Principales grupos de interés del proyecto



## 9. GESTIÓN DEL PROYECTO

### 9.1. ORGANIZACIÓN DE TAREAS

Con el objetivo de establecer una organización coherente y flexible se ha dividido el proyecto en seis fases. Varias de estas fases serán iterativas, permitiendo así regresar a ellas si fuese necesario sin frenar con ello el progreso del proyecto.

**Fase 1:** Establecimiento de los objetivos generales del proyecto.

Se debe determinar cuál es el objeto de esta investigación. Esto va a permitir encaminar la dirección a tomar. No obstante, en esta primera fase se desconoce cómo serán los datos a los que finalmente tendremos acceso. Como ya hemos visto, no existen muchos ejemplos de inventarios patrimoniales públicos, así que se ha asumido que dispondremos de información básica, como la época, la localización, su propiedad, su catalogación, una descripción del estado de conservación y sus riesgos. Por tanto, se establecen unos objetivos generales que, no obstante, podrán verse alterados según avance el proyecto.

**Fase 2:** Estudio del entorno y recopilación de datos.

Una vez fijados los objetivos es imprescindible investigar en profundidad la situación actual del campo en el cual va a desarrollarse el proyecto: documentarse al respecto del área del patrimonio histórico y la conservación, conocer sus bases, analizar el entorno socioeconómico en el campo de la conservación artística e histórica e investigar qué otros trabajos se han hecho que compartan similitudes con el nuestro.

Por otro lado, la recopilación de datos comprende tanto la obtención del acceso a la base o fichero de datos proporcionado por la entidad encargada de su gestión, objeto principal de estudio, como la recopilación de datos adicionales que puedan resultar de interés para la investigación, tales como variables climáticas, demográficas o económicas entre otras. Estos análisis quedan siempre abiertos a la inserción de información adicional, incluso encontrándose en un estado avanzado, razón por la cual se podrá regresar a esta fase de forma cíclica si fuera necesario.

**Fase 3:** Tratamiento de los datos.

Se asume que los datos no tendrán adecuación a las necesidades del análisis que se pretende hacer en el momento en que se obtenga acceso a ellos. Esto implica el tratamiento del *raw data* para que la información sea legible, la estructuración de los datos, la extracción de información mediante técnicas de *information retrieval* y la transformación de los datos para su adecuado análisis. Todo este proceso es el que más tiempo consume en un proyecto de este tipo.

**Fase 4:** Fase analítica.

Análisis exploratorio: Para comenzar esta fase se va a realizar un análisis de medidas descriptivas de todos los atributos. También trataremos de identificar las primeras correlaciones lineales entre las variables más representativas. Estos primeros análisis nos permiten conocer a grandes rasgos las posibilidades así como las limitaciones de los datos de que disponemos.

Valoración de nuevos objetivos: Llegado este punto será necesario hacer una revaloración de los objetivos fijados inicialmente. Usando los resultados de los primeros análisis, con una visión más exacta de los datos que finalmente dispongamos. Así el amplio abanico de posibilidades quedará reducido, ya sea por limitaciones técnicas, limitaciones de la propia información o porque se haya vislumbrado una meta destacada de entre las posibles. Es por esto que las fases 2 y 3 son iterativas.

Análisis avanzados: Concluidos los análisis descriptivos y las correlaciones lineales, entramos en materia más específica. Aquí se abre un amplísimo abanico de posibilidades que abarcan desde los test de hipótesis, las técnicas de clasificación, predicción, empleo de redes neuronales, etc. La elección de cuáles de estas se aplicarán finalmente para este proyecto depende de cómo sean los datos, la cantidad de fichas de las que dispongamos o de las limitaciones técnicas que puedan surgir.

**Fase 5:** Fase final y conclusiones.

Como cierre de la fase de análisis es imprescindible llevar a cabo una interpretación de los resultados finales obtenidos en cada una de las pruebas realizadas y la formulación de las conclusiones pertinentes sobre el resultado de la investigación. Deberá elaborarse un informe final para presentar a la entidad como resultado de todo el proyecto y de manera adicional se podrán proponer mejoras, correcciones o posibles líneas de ampliación para investigación futura.

**Fase 6:** Mantenimiento y actualizaciones.

Una vez finalizado el proyecto, los resultados del estudio quedarán a disposición de la asociación Hispania Nostra. La inclusión de nuevos datos de la Lista Roja debe ser siempre una posibilidad a la que el proyecto estará abierto. El método de almacenamiento de los datos empleado por la asociación deberá ser modificado en caso de detectarse errores o de surgir recomendaciones para su mejora. Posteriormente, se tratará de perfeccionar el sistema para que pueda dar mejores resultados.

**Fase “0”:** Elaboración de un informe de seguimiento.

Finalmente, se establece una tarea que debe ser llevada a cabo a lo largo de todo el proceso de desarrollo, elaborándose un informe de seguimiento del mismo. Informe que debe reflejar el progreso de la investigación de manera regular y que posteriormente conformará la memoria práctica del trabajo, es decir, el presente documento. Esta fase incluye además otro informe de propuestas para la mejora del fichero de datos de la Lista Roja, el cual será incluido como *Anexo C* al final del presente documento.

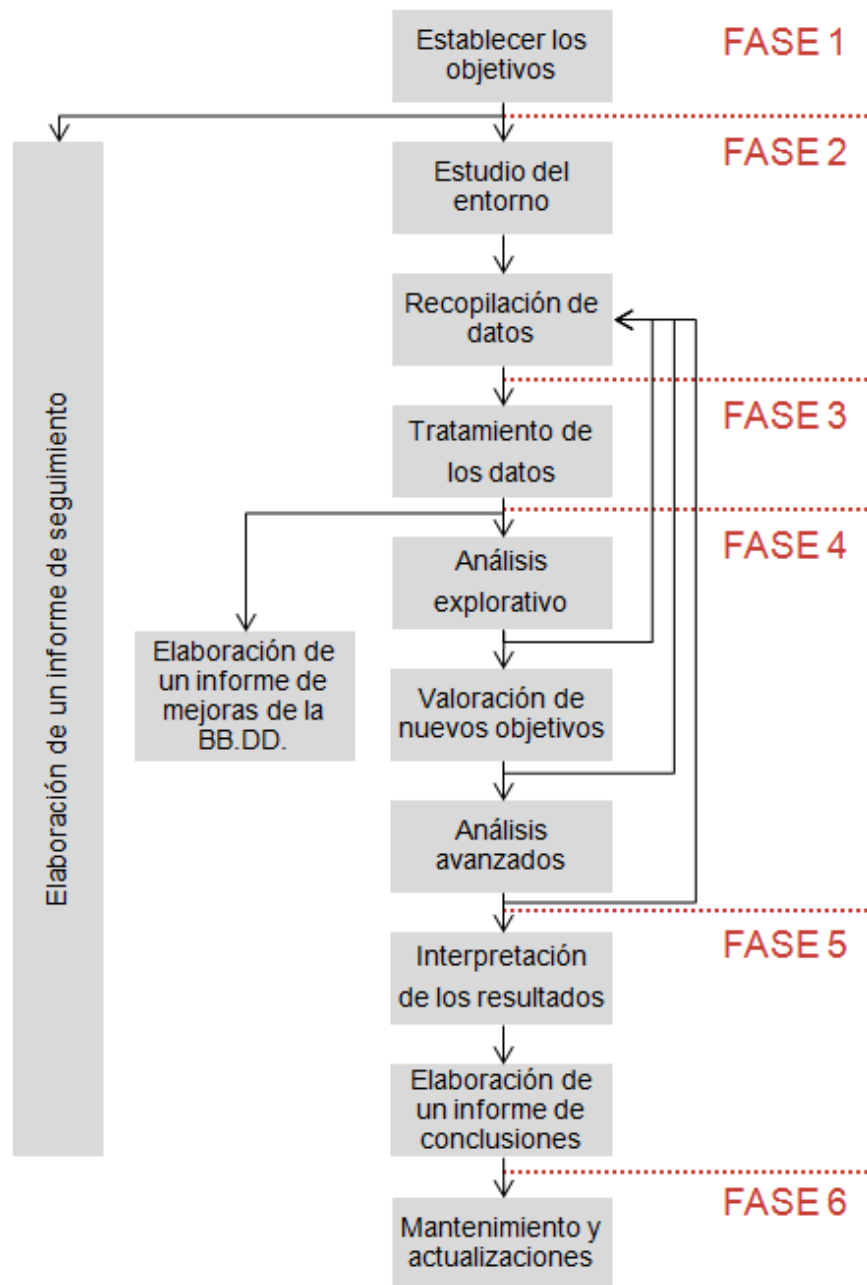


Figura 9.1. Esquema de organización de tareas del proyecto

## 9.2. PLANIFICACIÓN

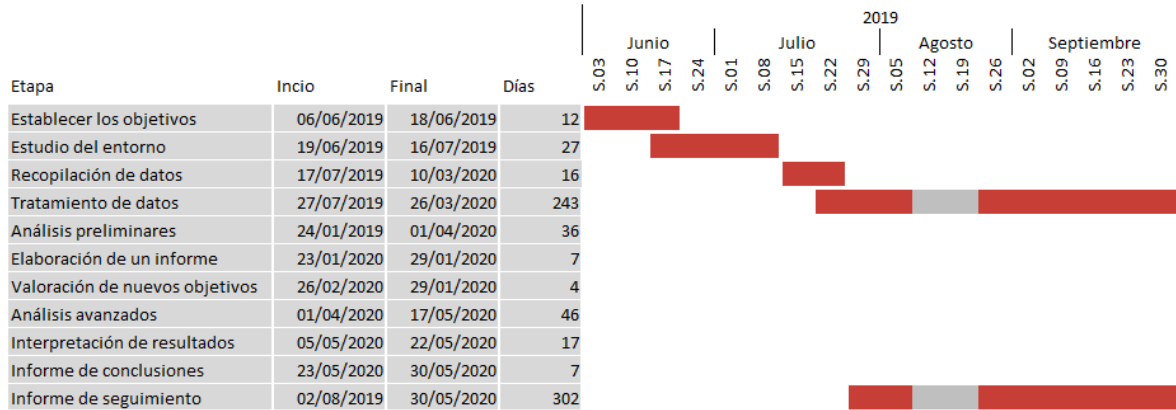


Figura 9.2. Diagrama de Gantt (parte 1)

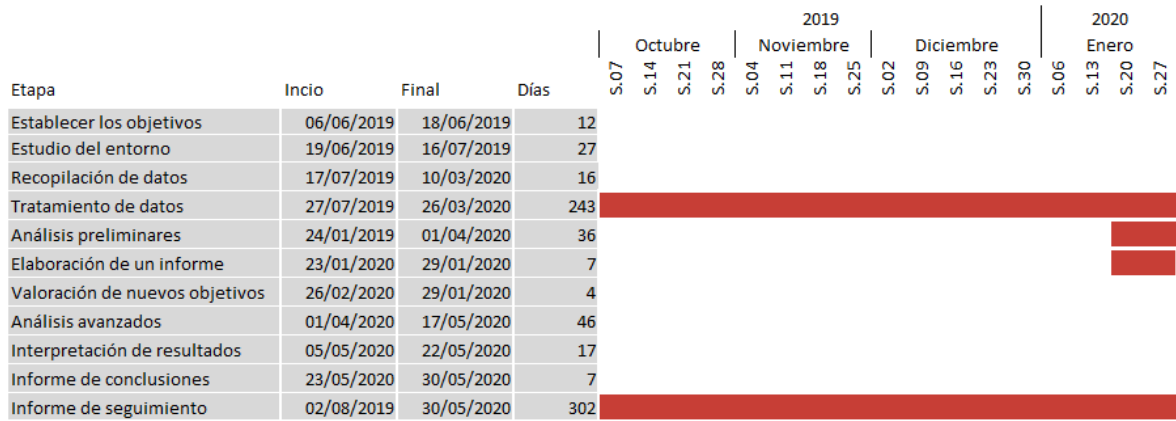


Figura 9.3. Diagrama de Gantt (parte 2)

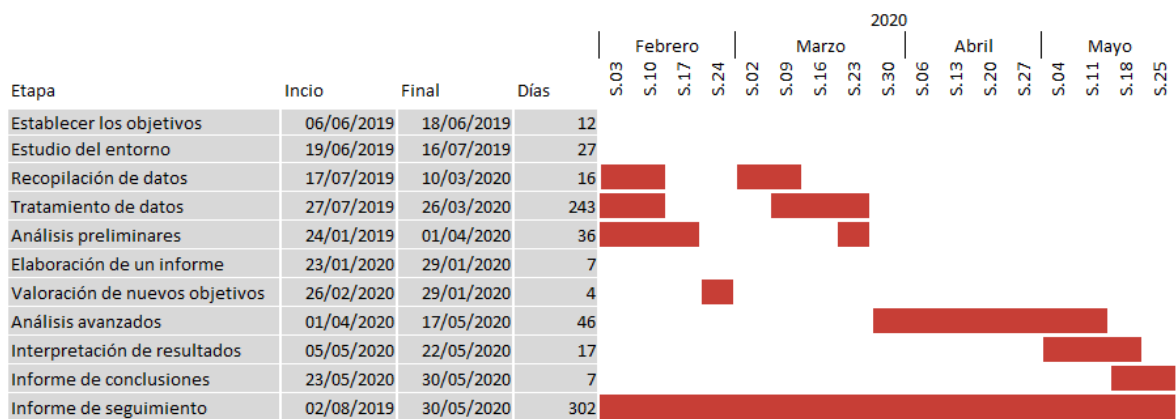


Figura 9.4. Diagrama de Gantt (parte 3)

### 9.3. ELABORACIÓN DE UN PRESUPUESTO

A continuación, se detalla el presupuesto de la elaboración del proyecto, detallando cada coste según su naturaleza.

#### 9.3.1. RECURSOS HUMANOS

Duración del proyecto: 12 meses (junio de 2019 – mayo 2020).

Días de trabajo efectivo (descontados días de vacaciones y fines de semana): 230 días.

Horas de trabajo al día (promedio): 3 horas.

Total de horas de trabajo: 690 horas.

Precio la hora: 11€

Por consiguiente, el total destinado a las horas de trabajo invertidas en el proyecto es de  $690 \times 12 = 7590\text{€}$ .

#### 9.3.2. RECURSOS MATERIALES

Material	Coste estimado	Vida útil promedio
Intel Core I5-6400 con DDR4	300€	16000 horas

Tabla 9.1. Recursos materiales

La duración del proyecto supone el 2.62% de la vida útil del equipo informático. Asumiendo que el rendimiento de la CPU es de aproximadamente un 4% en condiciones normales, el rendimiento percibido durante la mayor parte del desarrollo del proyecto ha sido del 8%. El rendimiento ha ascendido hasta un 50% durante las ejecuciones del código, estimando que esto se corresponde con un 10% del tiempo total empleado durante el desarrollo

Horas de máximo rendimiento del equipo informático:  $420 \times 0.1 = 42$  horas. Las horas de máximo rendimiento suponen un 0.26% del tiempo de vida útil del equipo.

Teniendo en cuenta todo lo anterior podemos asumir que 1 hora supone el 0.00625% de la vida útil del equipo al 4% de rendimiento. Por tanto, 1 hora se corresponde con el 0.0125% de la vida útil a un 8% de rendimiento y 1 hora supone el 0.078% de la vida útil a un 50% de rendimiento. Con esto:

378 horas se corresponden con un 4.725% de la vida útil al 8% de rendimiento.

42 horas se corresponden con un 3.28% de la vida útil al 50% de rendimiento.

En definitiva, el desgaste total del equipo durante el proyecto ha sido de aproximadamente el 8%, con lo que el coste en material del proyecto es  $300 \times 0.08 = 24\text{€}$ .

### 9.3.3. DESPLAZAMIENTOS Y DIETAS

No han sido requeridas dietas de ningún tipo durante el desarrollo del proyecto. El número total de desplazamientos han sido dos. El primero, el 22 de julio de 2019, para una reunión con miembros de la junta directiva de Hispania Nostra. El segundo tuvo lugar el 29 de enero de 2020 para dar parte del progreso del proyecto. Se pretendía realizar una tercera reunión en junio de 2020 para hacer una demostración final y entregar el informe completo, pero fue suspendida debido a las medidas de confinamiento tomadas por el Gobierno Español desde el 14 de marzo con motivo de la expansión del virus COVID-19.

Cálculo del coste por cada desplazamiento:

Autobús (Galapagar – Madrid) Tarifa estándar: 3.60€

Metro de Madrid (Moncloa – Ventura Rodríguez) Tarifa estándar: 1.50€

Metro de Madrid (Ventura Rodríguez – Moncloa) Tarifa estándar: 1.50€

Autobús (Madrid – Galapagar) Tarifa estándar: 3.60€

Gasto total por cada desplazamiento: 10.2€

Los dos desplazamientos han sido idénticos, por lo que el total ha sido de **20.4€**.

### 9.3.4. COSTE TOTAL DEL PROYECTO

Sumatorio de los costes:  $7590 + 24 + 20.4 = 7634.40\text{€}$

IVA (21%) = 1603.22€

Coste total + IVA = **9237.62€**

# 10. TRATAMIENTO DE LOS DATOS

Como hemos visto, y puesto que la Fase 1 y la Fase 2 del proyecto ya han sido debidamente desarrollados en los apartados 7 y 8 respectivamente, entramos en la Fase 3: el tratamiento del fichero de datos de la Lista Roja de Patrimonio y la extracción de la información de mayor interés comprendida en el mismo. Esta fase ha comprendido ocho de los doce meses que ha durado el proyecto, es decir, dos tercios del total.

## 10.1. EVALUACIÓN DE LA LISTA ROJA

Figura 10.1. Extracto del fichero de datos en su forma inicial

Figura 10.2. Extracto del fichero de datos tras su tratamiento

La Lista Roja se obtuvo en un primer momento en un archivo de excel corriente. Se trata básicamente de un dataset o fichero de datos de treinta y cinco columnas por

1196 filas. Este presentaba el aspecto que se aprecia en la *figura 10.1*. Tras un primer análisis superficial de los diferentes atributos, así como el contenido de los mismos, se extrajeron las siguientes anotaciones:

Existe una considerable cantidad de atributos redundantes que contienen u ofrecen información idéntica. En algunos casos parece que se deben a actualizaciones y modificaciones que ha sufrido el fichero de datos y el formulario de inclusión de bienes a lo largo de los años. Además, la información correspondiente a cada uno de estos campos se encuentra en uno u otro, indistintamente. De esto se ha deducido que en cierto momento se modificó el formulario, creándose nuevos campos incluso para algunos que ya debían estar contemplados en la anterior versión. Estos son los que han sido detectados:

*Nota: los nombres de los atributos entrecomillados están reflejados tal y como aparecían escritos en el fichero de datos.*

**“Localización”, “Provincia”, “Comunidad”**: la primera contiene una concatenación de las dos siguientes.

**“Localidad”, “localización”**: ambas contienen la localidad o el término municipal.

**“Location”, “Position”**: ambas contienen las coordenadas geográficas.

**“Historia/Descripción”, “Historia”, “Descripción”**: La primera contiene información de las dos siguientes, si bien cuando hay información dentro de la primera las dos siguientes quedan vacías y viceversa.

**“Propiedad:”, “Propiedad”**: ambas contienen el propietario.

**“Época:”, “Época”**: ambas contienen la época de origen.

**“Grado de Protección”, “Grado de protección legal”**: ambas contienen el nivel y tipo de protección legal.

**“Intervenciones”, “Intervención :”**: ambas recogen si han sido intervenidos en algún momento.

**“Estado”, “Estado de conservación”**: ambas contienen el estado de conservación en que se encuentra el bien.

**“Razones de inclusión”, “Riesgo por el que entró”, “carácter”**: ambas contienen las razones por las cuales fue incluido en la Lista Roja.

Existen campos que no han sido debidamente cumplimentados, como por ejemplo es el caso de **“Provincia”**, para aquellas fichas pertenecientes a bienes ubicados en comunidades autónomas uniprovinciales. Aunque se da por sentado que la provincia es la única perteneciente a dicha comunidad autónoma, el dato no debe quedar vacío.



<b>Tipología del patrimonio</b>	<b>Subtipos</b>
Patrimonio arqueológico	Elementos arqueológicos singulares
	Yacimientos arqueológicos
	Arte rupestre
	Paisajes o entornos de yacimientos de interés arqueológico
Patrimonio religioso	Iglesias y capillas
	Conventos y monasterios
	Eremitas y oratorios
	Cementerios y enterramientos
Patrimonio civil	Arquitectura vernácula
	Edificios singulares
	Palacios
	Conjuntos históricos
	Vías de comunicación históricas
	Plazas o espacios urbanos
Patrimonio industrial	-
Patrimonio militar	Castillos y conjuntos arquitectónicos fortificados
	Murallas
	Torres exentas
	Fuertes o construcciones militares históricas
Patrimonio natural	Elementos o conjuntos de interés medioambiental o paisajístico
	Fauna y flora
	Jardines o huertas históricas

Tabla 10.1. Clasificación del patrimonio

El atributo “**Categoría**”, que nos indica la tipología del patrimonio, clasificándose en función de si es militar, religioso, civil, natural, industrial o arqueológico, vuelve a ser un ejemplo de atributo que ofrece más de un dato, porque también contiene las subcategorías del mismo, tal y como se refleja en la *tabla 10.1*. Esto debería separarse para poder valorar la tipología y el subtipo de cada elemento patrimonial.

## 10.2. ADECUACIÓN DEL FICHERO DE DATOS

Para acelerar el proceso de preparación del fichero de datos para su tratamiento y evitar tener que desarrollar un script que lo hiciese, se ha recurrido a la herramienta open source OpenRefine. Con esta herramienta ha sido posible efectuar de forma fácil y rápida operaciones de transformación elementales sobre los datos. La interfaz gráfica de la herramienta permite además visualizar el estado del fichero de datos en todo momento durante cada una de las transformaciones. Esto además ha permitido localizar nuevos fallos que no habían sido identificados anteriormente.

### 1. Eliminación de tildes y de otros caracteres especiales del castellano.

Durante la lectura de ficheros en formato .csv con RStudio se han tenido frecuentes problemas con las codificaciones UTF-8. El Formato de Transformación Unicode de 8 bits permite la codificación de caracteres romances, signos diacríticos y símbolos de alfabetos como el griego, el árabe o el hebreo entre otros. Esto debería permitir leer cualquier tipo de fichero que contenga este tipo de caracteres. No obstante, en la mayoría de casos se han producido errores de codificación que reemplazaban estos caracteres por otros símbolos. Para evitar este problema se ha transformado el fichero de datos, convirtiendo todas las vocales tildadas en la misma vocal sin tilde y la eñes en enes.

### 2. Separación de un atributo en varios por un delimitador.

Como ya se ha comentado, los atributos “**Localización**” y “**Categoría**” contienen una concatenación de información que debería separarse en diferentes variables para poder realizar un análisis más riguroso. Por ello, comprobando la estructura de ambos atributos, se ha determinado un carácter delimitador que sirva de separación en la división del atributo. En el caso de localización debemos usar tanto la coma como el paréntesis. En el caso de Categoría se usa el carácter “>” para separar la tipología del patrimonio de su subtipo.

### 3. Fusión de atributos.

Se han unificado todos los atributos redundantes anteriormente descritos. Para ello se elige el atributo con mayor cantidad de campos cumplimentados de entre los distintos que contienen el mismo tipo de datos. A esta se le copiará el contenido de los redundantes a ella, siempre y cuando los campos de los redundantes no estén vacíos y

el contenido no sea idéntico al de la primera elegida. Ejemplo: “**Localidad**”, “**localización**”. Elegimos “**Localidad**” por ser la que tiene menos campos vacíos.

#### 4. Cumplimentación de campos incompletos.

Como también hemos visto, el atributo “**Provincia**” contiene campos vacíos en aquellas fichas pertenecientes a comunidades autónomas uniprovinciales. Por ello se han cumplimentado estos casos con el nombre de su respectiva provincia. Pero “**Provincia**” no es el único caso, porque la tipología del patrimonio “patrimonio industrial” carece de subcategoría, al contrario que el resto. Por ello, se debe cumplimentar los campos correspondientes del atributo “**Subcategoría**” con la frase “patrimonio industrial”.

Además, durante este análisis se han encontrado fragmentos de código HTML y CSS repartidos por los atributos “**Historia**” y “**Descripción**”, seguramente procedentes de las páginas web de las que se ha copiado la información, y que deberán ser limpiados para el correcto tratamiento de los atributos. También se han identificado fichas completas con campos vacíos o cumplimentadas con información aleatoria, posiblemente insertados como pruebas para comprobar el funcionamiento del formulario. Estos datos deberán ser eliminados para no introducir ruido en los análisis.

### 10.3. EXTRACCIÓN DE INFORMACIÓN

La tercera fase del proyecto ha consistido en la aplicación de técnicas de búsqueda y extracción de información en el fichero de datos, también conocido como *information retrieval*.

Se han implementado una serie de fragmentos de código que buscan los términos más frecuentes dentro de cada uno de los atributos. Estos devuelven un histograma que representa las N palabras más frecuentes, lo que permite conocer la información contenida en cada atributo, como si se tratase de una lista de términos clave comprendidos en un texto. También se ha implementado un código que permite detectar la vinculación de cada término con las palabras inmediatamente anexas a lo largo de cada texto. De esta forma nos permite visualizar mediante gráficos cuáles son los términos que aparecen más frecuentemente vinculados entre sí, lo que muchas veces nos permite conocer el contexto en los que aparecen los diferentes términos clave.

Basándonos la información que nos vayan mostrando estas gráficas, se irán implementando toda una serie de códigos diseñados para extraer la información clave de cada uno de los atributos del fichero de datos.

Los atributos tratados han sido **Época**, **Intervenciones**, **Motivo.de.la.retirada**, **Propiedad**, **Grado.de.protección.legal**, **Estado.de.conservación**, **Riesgo**, **Descripción** e **Historia**. A continuación, se explica detalladamente el procedimiento seguido para cada uno de ellos.

### 10.3.1. ATRIBUTO “ÉPOCA DE EDIFICACIÓN”

1905-1990
1836
Siglo XVIII
XII - XIII
XI, algunos restos datados en el VIII
Romana y medieval
VI a. C.-I a. C.
Románico con reformas del XVII y posteriores.

Figura 10.3. Muestras originales del atributo *Época*

<b>1905-1990</b>
<b>1836</b>
<b>Siglo XVIII</b>
<b>XII - XIII</b>
<b>XI, algunos restos datados en el VIII</b>
<b>Romana y medieval</b>
<b>VI a. C.-I a. C.</b>
<b>Románico con reformas del XVII y posteriores.</b>

Figura 10.4. Representación de los términos a detectar por el sistema

XX	D. C.	
XIX	D. C.	
XVIII	D. C.	
XII-XIII	D. C.	
XI VIII	D. C.	
V	D. C.	
I-VI	A. C.	
XI-XIII	D. C.	Reformas: XVII

Figura 10.5. Resultado de la ejecución

El atributo **Época** ha supuesto el primero de los retos para tratar el fichero de datos: el contenido de este atributo carece de formato consensuado alguno. La ambigüedad del nombre que se le ha dado a **Época** ha llevado a los usuarios a aportar no solo años o siglos, sino movimientos artísticos, épocas históricas y hasta descripciones breves, tal y como se puede observar en la *figura 10.3*. De igual manera encontramos intervalos de siglos o años o las abreviaturas *a. C.* y *d. C.*, así como intervalos de siglos y años separados mediante guiones e incluso información adicional como la datación de reformas posteriores a la época de edificación del edificio original. Todo esto puede apreciarse en la *figura 10.6* que recoge los treinta términos más frecuentes encontrados en el atributo.

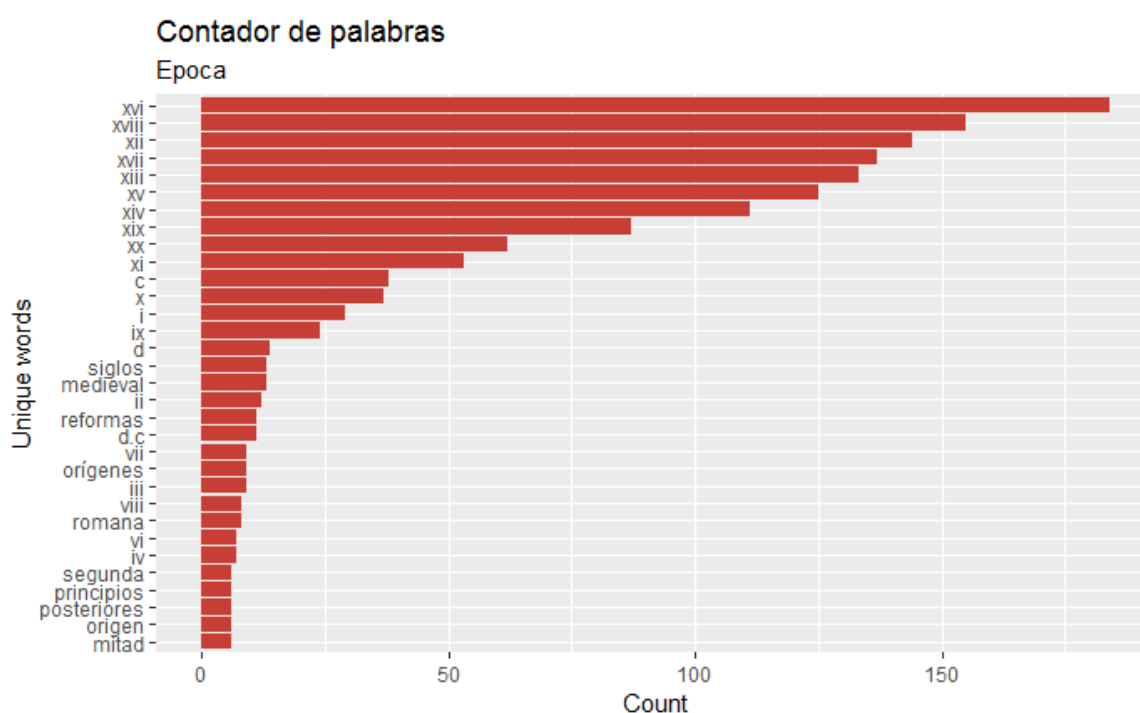


Figura 10.6. Términos más frecuentes hallados en *Época*

Lo primero y más esencial para unificar todos estos datos es tomar una unidad de referencia. Se ha decidido trabajar en todo momento con siglos. La gran disparidad de los datos almacenados acarrea cierta imprecisión y un exceso de precisión con tan pocos datos como tiene la Lista Roja no nos permitiría hacer agrupación con respecto a la época. El siglo o intervalos de siglos es un dato relativamente sencillo de obtener partiendo de años o periodos históricos.

Para encontrar información clave dentro de cadenas de caracteres se ha recurrido mayoritariamente a la función *grep()* (*pattern, data*), que devuelve un booleano o una lista de ellos. Este booleano será positivo si encuentra el patrón *pattern* dentro de la cadena de caracteres *data*. Otro recurso que se ha utilizado muy frecuentemente a lo

largo de estos análisis es ir implementado contadores de palabras que generan histogramas con las palabras más frecuentes encontradas a lo largo de un texto. Con una debida limpieza de palabras vacías (artículos, preposiciones y otras palabras que tienden a ser las más frecuentes en cualquier forma de expresión verbal) nos permite identificar información esencial dentro de cada atributo. A continuación, se describe de manera detallada la implementación del sistema que se ocupa del tratamiento de la variable **Época**.

#### 10.3.1.1. UNIFICACIÓN DE FORMATOS

Se han eliminado términos tales como “siglo”, “siglos” o “s.”. Son términos que aparecen de forma frecuente acompañando al dato y que no nos aportan información útil, al contrario, funcionan como “palabras vacías”. También se han tratado los guiones, de tal forma que todos aparezcan sin espacio antes o después del carácter ASCII del guion. Es decir, expresiones como “X - X”, “X- X” o “X -X”, quedarán convertidas “X-X”. De esta forma suprimimos expresiones carentes de uniformidad, fruto de una redacción rápida por parte del usuario, lo que posteriormente nos va a permitir tratar con mayor facilidad la aparición de intervalos de fechas. Es importante tener en cuenta que por el momento no se van a eliminar las palabras vacías. Esto es debido a que el fonema “a” sería identificada como palabra vacía, indistintamente de que aparezca como preposición (caso en el que efectivamente sería una palabra vacía) o que forme parte de la abreviatura de “antes de Cristo”. Por esta razón, no trataremos las palabras vacías hasta que se haya tratado la era de cada época.

#### 10.3.1.2. TRATAMIENTO DE AÑOS

El primer paso ha consistido en convertir todos los años en siglos. Esto se ha hecho mediante la sustitución de caracteres numéricos en números romanos según su estructura, tal y como queda reflejado a continuación.

$[num][num] \rightarrow I$

$1[num][num] \rightarrow II$

$2[num][num] \rightarrow III$

...

$10[num][num] \rightarrow XI$

$11[num][num] \rightarrow XII$

...

Para la sustitución de fragmentos de cadenas de caracteres se hace uso de la función *gsub(pattern, replacement, data)*, que sustituye por *replacement* cualquier coincidencia con *pattern* que haya dentro de *data*. Es importante valorar que todo número

natural mayor que 9 está conformado por una concatenación de números. Por ello, dado que trabajamos con ocurrencia de caracteres, si empezamos a sustituir los años en orden ascendente, el sistema encontrará más de los que realmente hay. Por ejemplo, la construcción “1[num][num]”, concebida para encontrar valores entre el 100 y el 199, aparece también dentro de grupos de cuatro números, y aceptaría valores como por ejemplo 1421 o 1186 ([142]1, 1[186]), que serían erróneamente etiquetados como “siglo II”. Por ello se establece un orden decreciente, dando prioridad a los años cuya longitud numérica sea mayor. Al reemplazarse el conjunto de dígitos identificado, el sistema no detectará posteriormente cifras en los que ya hayan sido tratados.

Tras la transformación de los años en siglos muchos intervalos de años que estuviesen comprendidos en el mismo siglo han quedado con la forma “SIGLO-SIGLO”, donde SIGLO toma el mismo valor a ambos lados. Por tanto, el código busca ocurrencias de todos los casos posibles que se hayan podido generar mediante la combinación de todos los siglos desde el I hasta el XXI, y se reemplazan por el correspondiente siglo único. En esta ocasión es importante valorar que en la nomenclatura de los números romanos, la mayoría de los números están conformados a su vez por otros. Por ejemplo, dentro del siglo XII podemos identificar las expresiones de los siglos X, XI, I e II. Por ello se establece un orden especial, dando prioridad a los siglos cuya expresión es más compleja y dejando para el final aquellos que sean más simples.

Se ha conseguido identificar todos estos intervalos sin riesgo a confusiones, mediante el uso de expresiones regulares con la estructura “(\\W|^)SIGLO-SIGLO(\\W|\$)”, donde (\\W|^) se corresponde con cualquier carácter no alfabético o inicio de línea, y (\\W|\$) se corresponde con cualquier carácter no alfabético o final de línea. Un ejemplo de confusión sería encontrar el intervalo “I-I” en “II-IV” o “II-I”, que logramos evitar mediante el uso de las expresiones de inicio y fin de línea y otros delimitadores que separen el intervalo de más texto.

### 10.3.1.3. TRATAMIENTO DE LA ERA CRONOLÓGICA

Como ya se ha mencionado anteriormente, se ha valorado también las fechas que contienen las abreviaturas “antes de Cristo” y “después de Cristo”, aunque no siempre estas aparecen escritas de la misma manera. “a. C.”, “a.c.” o “a c” son solo algunos ejemplos. Por consiguiente, se han recopilado las distintas formas de escribir estas abreviaturas para ser unificadas como “A. C.” o “D. C.”. Llegados a este punto, se puede hacer una limpieza de palabras vacías de tal forma que no detecten las “A” mayúscula.

Todas aquellas ocurrencias de la abreviatura A. C. se almacenan como cadena de caracteres en un nuevo atributo que llamamos **Era**, asumiendo que, si no aparece, siempre será D. C. Un caso especial serán aquellas instancias que contengan siglos A. C. y D. C., es decir, que posean partes fechadas en ambas eras o cuyo periodo de edificación pudiese extenderse a lo largo de ellas. En este caso se almacena la cadena “A. C. D. C.”.

#### 10.3.1.4. TRATAMIENTO DE INTERVALOS

Se hace distinción de los datos que se encuentran separados por guiones para ser tratados como intervalos. No se pueden considerar de igual manera los casos “XIV, XVI” y “XIV-XVI”, ya que asumimos que el primero nos indica que el bien cultural ha tenido dos fases de edificación aisladas (presumiblemente en diferentes partes de la construcción) y el segundo nos indica que fue edificado en un periodo que se extiende a lo largo de más de varios siglos. Esto se ha realizado mediante un código que identifica construcciones de la forma “Siglo-Siglo”. Como patrón para la detección de los siglos se ha utilizado la expresión regular “[X\*I\*X\*V\*I]\*+”. Esta expresión permite la representación de los siglos I al al XXXIX. Esta decisión ha sido tomada para simplificar el problema, pero podrían llegar a contemplarse posteriores ampliaciones mediante el uso de la expresión regular “[L\*X\*L\*I\*V\*I\*X]\*+”, lo cual permitiría representar hasta el valor 89 (LXXXVIII). Una vez identificados son almacenados en un nuevo atributo auxiliar provisional llamado **SigloIntervalos** y posteriormente eliminados del atributo **Época**.

#### 10.3.1.5. TRATAMIENTO DE SIGLOS INDIVIDUALES

Una vez concluido el tratamiento de los intervalos, y eliminados estos del contenido del atributo **Época**, se buscan los siglos de forma individual. Nuevamente y para evitar que el sistema detecte siglos dentro de otros que contengan la expresión de los primeros, se disponen los siglos del I al XXI según un orden determinado. Como además de los siglos hay otra información en forma de textos almacenada en muchos de los campos de este atributo, se ha tenido que determinar un protocolo que evite que las letras de las palabras de dicha información puedan ser confundidas con las letras que conforman los números romanos de los siglos. Los siglos serán almacenados en un nuevo atributo, **Siglo**, en forma de lista separada por espacios. El orden del procedimiento es el siguiente:

Primero: crear una lista con los siglos numerados en orden decreciente desde el XXI hasta el XI a la que añadimos el siglo IX. Los siglos conformados por varias letras no suponen un riesgo ya que la función empleada busca *exact matches*, y los textos están escritos en minúscula, por lo que, además de las iniciales y los siglos, no habrá más mayúsculas. Así pues no encontraremos dos o tres mayúsculas seguidas que no hagan referencia a un siglo. La búsqueda se hace individualmente y en el orden de la lista, de tal forma que sus expresiones no sean detectadas en las de otros siglos que pudiesen contenerlas (por ejemplo, para no detectar “XV” en “XVI”, buscamos “XVI” previamente). Conforme son detectados, los siglos son almacenadas en el atributo **Siglo**.

Tratamiento del siglo X: Los siglos cuya expresión está conformada por una única letra sí conllevan un riesgo de ser confundidos con iniciales. De las tres con las que trabajamos, X es precisamente la menos arriesgada, pero vamos a tratarla de igual forma que al resto. Para ello, convertiremos todas las iniciales de palabras en minúsculas. Esta es la razón por la que hemos tratado IX en el paso anterior. Mediante la operación `gsub("(X)([a-z])", "x\2", data)` se nos permite reemplazar la inicial “X” de cualquier



palabra por una “x” minúscula. Posteriormente se buscan todas las coincidencias con una una “X” mayúscula, que son almacenadas en el atributo **Siglo**.

Segundo: crear una lista con los siglos numerados en orden decreciente desde el VIII hasta el VI a la que añadimos el siglo IV. Nuevamente, los siglos formados por varias letras no suponen un riesgo. La búsqueda se hace individualmente y en el orden de la lista, de tal forma que sus expresiones no sean detectadas en las de otros siglos. Los siglos detectados son almacenados en el atributo **Siglo** conforme son detectados.

Tratamiento del siglo V: Los siglos cuya expresión está conformada por una única letra sí conllevan un riesgo de que puedan ser confundidos con iniciales. Convertiremos todas las iniciales de palabras en minúsculas. Esta es la razón por la que hemos tratado IV en el paso anterior. Mediante la operación `gsub("[V]([a-z])", "v\2", data)` se nos permite reemplazar la inicial “V” de cualquier palabra por una “v” minúscula. Posteriormente se buscan todas las coincidencias con una “V” mayúscula, que son almacenadas en el atributo **Siglo**.

Tercero: crear una lista con los siglos III e II en este orden. Nuevamente, los siglos conformados por varias letras no suponen un riesgo. Los siglos son buscados de uno en uno, de tal forma que sus expresiones no sean detectadas en las de otros siglos que pudiesen contenerlas. Los siglos detectados son almacenados en el atributo **Siglo** conforme son detectados.

Tratamiento del siglo I: Los siglos cuya expresión está conformada por una única letra sí conllevan un riesgo de ser confundidos con iniciales. Convertiremos todas las iniciales de palabras en minúsculas. Mediante la operación `gsub("[I]([a-z])", "i\2", data)` se nos permite reemplazar la inicial “I” de cualquier palabra por una “i” minúscula. Posteriormente se buscan todas las coincidencias con una “I” mayúscula, que son almacenadas en el atributo **Siglo**.

Cuarto: ordenar IX con respecto a X, e IV con respecto a V dentro del atributo **Siglo**. Como antes hemos invertido su orden para detectarlos para no dar lugar a errores, ahora debemos invertirlos para que los datos queden ordenados. De aparecer ambos siglos será uno a continuación del otro, basta con reemplazar “X IX” por “IX X” o “V IV” por “IV V”.

#### 10.3.1.6. TRATAMIENTO DE PERIODOS HISTÓRICOS

Como ya se ha dicho, los periodos históricos y movimientos artísticos también aparecen de forma frecuente en el atributo **Época**. Esto nos puede permitir obtener una estimación de los siglos de edificación cuando no tengamos nada más preciso. No obstante, se tendrá en cuenta la poca fiabilidad de los resultados de este procedimiento. La información extraída será almacenada en un atributo denominado **Época.estimada**.

Se han contemplado gran variedad de diferentes formas de referenciar cada un periodo histórico. Por ejemplo, la *Edad Media* puede referenciarse como *Edad Media* o *medieval*, y podemos diferenciar entre *Alta Edad Media* y *Baja Edad Media*, así como

*Medievo Tardío o Temprana Edad Media.* De igual manera se contempla el movimiento artístico como un adjetivo, masculino o femenino, que acompaña al bien patrimonial (*edificio románico, iglesia románica*) o un adjetivo sustantivado (*el románico*).

Se ha conformado una lista con todos estos términos [figura 10.2], y se ha asociado a cada uno de ellos una variable que contiene un vector numérico [25]. Dichos vectores representan el intervalo de siglos que comprenden cada uno de estos periodos, y son los valores por los que serán reemplazados los términos conforme aparezcan. Los intervalos de números enteros en R se representan como “N:M”, donde “N” se corresponde al primer valor y “M” al último, y toman todos los valores comprendidos entre ellos, ambos incluidos. Los siglos antes de Cristo serán representados en este caso como negativos. Aunque el siglo cero no exista no nos afecta que se encuentre dentro del vector porque nunca llegará a ser mostrado.

Periodo	Intervalo
Prehistoria	-13000:-32
Edad de Piedra / Etapa Lítica	-13000:-35
Paleolítico	-13000:-100
Paleolítico Inferior	-13000:-1500
Paleolítico Medio	-4000:-300
Paleolítico Superior	-400:-100
Epipaleolítico	-100:-55
Neolítico	-55:-35
Edad de los Metales	-35:-3
Edad del Cobre / Calcolítico	-32:-22
Edad del Bronce	-23:-8
Edad del Hierro	-9:-3
Antigüedad Clásica	-5:2
Celtíberos	-13:-2
Tartesos	-13:-6
Griegos	-7:-3
Fenicios	-9:-6

Íberos	-5:-1
Cartagineses	-6:-3
Púnico	-6:-3
Romano	-3:5
Paleocristiano	2:4
Visigodo	5:8
Edad Media	5:15
Alta Edad Media	5:12
Bizantino	6:7
Mozárabe	8:12
Hispanomusulmán \ Árabe	8:15
Taifa	11:13
Románico	11:13
Baja Edad Media	13:15
Mudéjar	12:15
Gótico	13:15
Nazarí	13:15
Renacimiento	15:16
Manierismo	16
Barroco	17:18
Rococó	18
Neoclasicismo	18:19
Contemporáneo	19:20
Actual	20:21

Tabla 10.2. Periodos históricos y movimientos artísticos

En el fichero de datos podemos encontrar fichas que son descritas con dos o más periodos históricos. Estas épocas pueden haber coexistido en algún intervalo de tiempo o pueden estar distantes entre sí. Tenemos dos posibles casos principalmente:

Un bien que pertenezca a dos épocas con siglos en común (por ejemplo, *romano* y *medieval*, que finaliza y termina en el siglo V respectivamente) puede significar como poco tres posibilidades: la primera es que sea un bien edificado en una época en que ambos periodos históricos hayan coexistido. La segunda es que sea un bien que comenzó a ser edificado en un momento perteneciente a uno de dichos periodos históricos y se finalizase en otro periodo posterior. La tercera es que sea un bien edificado durante la transición entre ambas épocas. Como la información de la que disponemos es insuficiente, vamos a tratar todos como el primer caso, acotando su época mediante la intersección ( $\cap$ ) de ambas.

Un bien que pertenezca a dos épocas que no han coexistido y que, por lo tanto, carecen de siglos en común (por ejemplo, *renacentista* y *neoclásico*), generalmente significa que el bien debió de ser edificado en diferentes fases o con posteriores añadidos, casos que contemplaremos mediante la unión ( $\cup$ ) de ambos periodos. No obstante, será imprescindible valorar el periodo de separación que haya entre las épocas de un bien. Puede suceder que la unión de dos épocas genere un intervalo de siglos sin interrupción, pero podría darse el caso de haber una separación de siglos entre ellos. En este segundo caso, tomar el año de inicio y el de finalización y unirlos mediante un intervalo sería muy poco conveniente.

Un bien que pertenezca a más de dos épocas se corresponderá a la combinación de estos dos casos.

La obtención de los siglos se ha resuelto mediante la implementación de un código que aplica teoría de conjuntos sobre listas o vectores numéricos ordenados. Permite trabajar con hasta cuatro listas de números. Para resolver la cuestión de las épocas separadas entre sí en el tiempo, se ha diseñado un algoritmo que se detalla a continuación:

Partiendo del vector resultante "A" de las operaciones de unión e intersección de los siglos de los distintos periodos históricos, se genera un nuevo vector "B" que contendrá el intervalo de números enteros comprendidos entre el primer valor y el último del vector "A". Si el número cero se encuentra entre los números de este vector "A", será eliminado, puesto que no existe el siglo cero. Con este vector "B" generamos un tercer vector "C", que será la diferencia entre el vector con el intervalo completo "B" y el vector de siglos "A". A este vector lo llamaremos *siglos\_intermedios*.

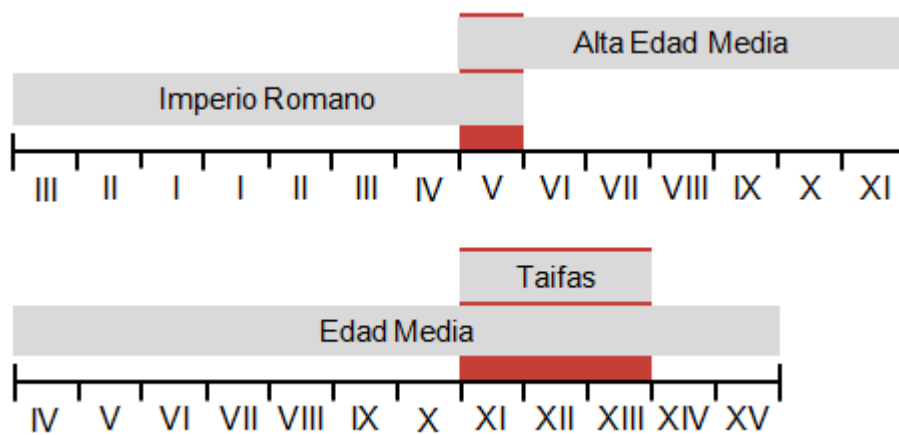


Figura 10.7. Ejemplos de la operación de intersección aplicada a periodos históricos

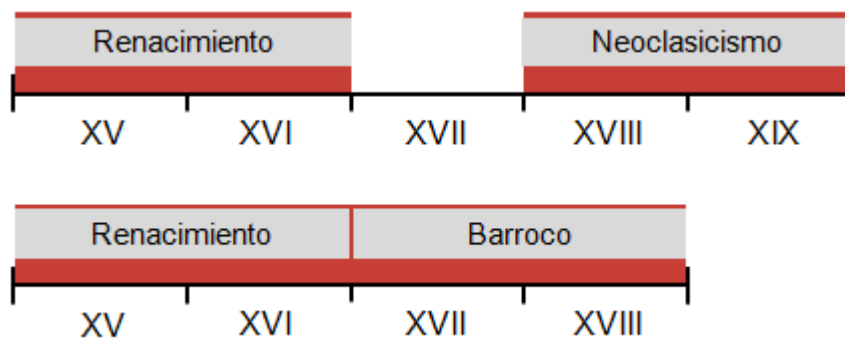


Figura 10.8. Ejemplos de la operación de unión aplicada a periodos históricos

Si *siglos\_intermedios* está vacío será porque los siglos no están separados y está conformado por un único periodo ininterrumpido, y procederemos a almacenarlo en un atributo **Época.estimada**. Para ello, extraemos el valor de la primera posición del vector y convertimos en número romano su valor absoluto mediante las operaciones *as.roman(abs(vector[1]))*. A continuación hacemos lo mismo con el valor de la última posición del vector, usando *as.roman(abs(vector[length(vector)]))*. Finalmente tomamos ambos números romanos y los concatenamos en orden separados por un guión. Este intervalo de siglos será el que almacenemos en **Época.estimada**. La razón de usar el valor absoluto es porque no existen los números romanos negativos. Los valores negativos nos indican que son siglos de la era antes de Cristo. La era la trataremos más adelante.

Si *siglos\_intermedios* no está vacío será porque efectivamente hay uno o varios conjuntos de siglos intermedios que separan diferentes fases de edificación. He querido acompañar esta sección del algoritmo de un ejemplo desarrollado paso por paso para facilitar su comprensión. La primicia del ejemplo será una ficha ficticia con tres épocas: “visigótico”, “renacimiento” y “neoclasicismo”.

$A1$  (vector de “Visigótico”): [5, 6, 7, 8]

$A2$  (vector de “Renacimiento”): [15, 16]

$A3$  (vector de “Neoclasicismo”): [18, 19]

$A$  (vector resultante de la operación de unión) = [5, 6, 7, 8, 15, 16, 18, 19]

#### GENERAR EL VECTOR “SIGLOS\_INTERMEDIOS”

$A = [5, 6, 7, 8, 15, 16, 18, 19] \rightarrow B = [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]$

$B - A = C$

$C = [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19] - [5, 6, 7, 8, 15, 16, 18, 19]$

$C$  (*siglos\_intermedios*) = [9, 10, 11, 12, 13, 14, 17]

En este caso, iremos almacenando los “fragmentos” de periodo de edificación en forma de lista ordenada. Para ello será necesario aplicar una función iterativa mediante un bucle *while* cuya condición de parada será que el vector *siglos\_intermedios* quede vacío.

El primer paso consiste en obtener el primer elemento del vector de siglos “*A*” y el primer elemento del vector *siglos\_intermedios*. Al restar 1 al valor del primer elemento de *siglos\_intermedios* se obtiene el último elemento del primer intervalo de los siglos de edificación. La excepción sería que el primer elemento fuese “1” (siglo I d. C.), que el elemento previo deberá ser “-1” (siglo I a. C.) en lugar de “0”, ya porque no existe el siglo cero, por lo que en dicho caso habrá que restar 2. Con ello hemos obtenido el primer “fragmento” en forma de intervalo o siglo individual.

#### PRIMERA ITERACIÓN:

$A = [5, 6, 7, 8, 15, 16, 18, 19]$

$C$  (*siglos\_intermedios*) = [9, 10, 11, 12, 13, 14, 17]

$9 - 1 = 8$

*Fragmento1* = [5, 6, 7, 8]

A continuación, se restará el intervalo de valores comprendido en este “fragmento” al vector de siglos “*A*”, se volverá a calcular el vector *siglos\_intermedios* con la primera y última posición del vector de siglos “*A*” y se repetirá toda esta sección hasta que *siglos\_intermedios* se quede vacío, tal y como establece la condición de parada del *while*.

$$A = A - \text{Fragmento1}$$

$$A = [5, 6, 7, 8, 15, 16, 18, 19] - [5, 6, 7, 8] = [15, 16, 18, 19]$$

$$A = [\underline{15}, 16, 18, \underline{19}] \rightarrow B = [15, 16, 17, 18, 19]$$

$$B - A = C$$

$$C = [15, 16, 17, 18, 19] - [15, 16, 18, 19] = [17]$$

$$C (\text{siglos\_intermedios}): [17]$$

#### SEGUNDA ITERACIÓN:

$$A = [\underline{15}, 16, 18, 19]$$

$$C (\text{siglos\_intermedios}) = [\underline{17}]$$

$$17 - 1 = 16$$

$$\text{Fragmento2} = [15, 16]$$

$$A = A - \text{Fragmento2}$$

$$A = [15, 16, 18, 19] - [15, 16] = [18, 19]$$

$$A = [\underline{18}, \underline{19}] \rightarrow B = [18, 19]$$

$$B - A = C$$

$$C = \emptyset$$

Llegado este punto, se comprobará si los siglos que quedan en el vector "A" son varios o solo uno. En caso de ser uno se almacenará como último "fragmento". Si son varios se almacenará el intervalo conformado por el primer y último término del vector "A".

$$A = [\underline{18}, \underline{19}]$$

$$\text{Fragmento3} = [18, 19]$$

Finalmente, solo nos resta almacenar la era de los intervalos. Para ello bastará comprobar el vector inicial, del que debemos haber guardado una copia previamente. Si el valor contenido en su primera posición es negativo, almacenamos "A. C." en la posición correspondiente del atributo **Era**. Y si la posición final es positiva, concatenamos "D. C." con lo que ya hubiese en la posición correspondiente del atributo **Era** y lo almacenamos en la misma.

### 10.3.1.7. BÚSQUEDA DE FECHAS DE EDIFICACIÓN EN EL ATRIBUTO “HISTORIA”

Supongamos que un usuario que rellena el formulario de inclusión de un bien introduce un periodo histórico en el campo **Época** debido a la ambigüedad del título de este y no por desconocimiento de la fecha o el siglo de edificación específico. Entonces, siguiendo este razonamiento, el dato, de ser conocido, podría encontrarse en el atributo **Historia**.

El atributo **Historia** consiste en un texto libre de extensión variable, en muchos casos extraído de manera literal de sitios web y, en menor medida, redactado por los usuarios. Constituye una variable carente de todo tipo de consenso. No se hace hincapié en ningún momento en qué aspectos deben ser incluidos. No obstante, se ha tratado de aprovechar su contenido para extraer información útil. Esto nos lleva a deducir que el hecho de que un dato no esté contemplado en **Historia** no es suficiente como para negarlo.

Se ha implementado por tanto un método que se ejecuta únicamente para aquellas filas que contuviesen en la variable **Época** un string que referencie a un periodo histórico. En caso de obtener una fecha o siglo de edificación, esta reemplazará posteriormente los periodos históricos previamente extraídos de la ficha correspondiente.

El método para extraer los siglos de edificación es sencillo: reutilizamos el código previamente implementado para tratar los intervalos, la era y las palabras vacías. La era será almacenada en un nuevo atributo llamado **Era.Historia**. Finalmente se buscan las expresiones edificado (“*constr*[:alpha:]”<sup>\*</sup>), edificado (“*edific*[:alpha:]”<sup>\*</sup>), erigido (“*erig*[:alpha:]”<sup>\*</sup>) o levantado (“*levant*[:alpha:]”<sup>\*</sup>) acompañados de la expresión regular de un intervalo de siglos (“*[X\*!X\*V\*!]\*+*”<sup>\*</sup>) o un único siglo. La información extraída será almacenada en un nuevo atributo que denominaremos **Siglo.Historia**.

### 10.3.1.8. UNIFICAR SIGLOS EN UNA ÚNICA VARIABLE

En este momento, con respecto a siglos, tenemos las variables **Siglo**, **Siglo.Intervalos**, **Época.Estimada** y **Siglo.Historia**. El siguiente paso es por tanto unificarlas. Para ello se concatena el contenido de **Siglo.Intervalos** con el de **Siglo**. El resultado es almacenado en un atributo que denominaremos **Siglo.Edificacion**. Si el resultado de la concatenación de los atributos fuese una cadena vacía (porque ambos estuviesen vacíos) es entonces cuando recurrimos a **Época.estimada**, cuyo contenido será copiado en **Siglo.Edificación**. Sin embargo, en aquellas fichas en las que **Siglo.Historia** no esté vacío, su contenido reemplazará al de **Época.Estimada**.

A continuación, se ordena la información recopilada, de tal forma que primero aparezcan los siglos más antiguos. Para ello, se ordenan los intervalos entre sí con respecto al siglo de inicio de cada uno, y de igual forma los intervalos con respecto a los siglos aislados. Si se diese la situación de tener un siglo aislado y un intervalo, y el siglo aislado se encontrase comprendido dentro del intervalo, el siglo sería eliminado porque se considera redundante.



En el siguiente paso se vuelve a consultar el atributo inicial **Época**. Si el atributo contiene alguno de los siguientes periodos históricos: “*prehistoria / prehistórico*”, “*paleolítico/a*”, “*paleolítico inferior*”, “*paleolítico medio*”, “*edad antigua*” o “*edad de piedra / etapa lítica*” y no ha sido reemplazado tras el análisis del atributo **Historia**, se incluirá en el atributo **Siglo.Edificación** la cadena de caracteres “*Anterior al siglo MMMDCCCXCIX*”. Esto es debido a que R solo trabaja con números romanos desde el 1 hasta el 3899, por lo que no es posible representar cifras mayores. Como todos esos periodos históricos se inician antes del siglo MMMDCCCXCIX antes de Cristo, usamos la frase indicada para simplificar el inicio de su periodo de edificación.

Para terminar, se unifican los atributos **Era** y **Era.Historia**, reemplazando el contenido de Era en aquellas fichas en que **Era.Historia** tenga información. Los demás atributos que hemos ido generado en los pasos intermedios se eliminan llegados a este punto.

#### 10.3.1.9. INICIO Y FIN DEL PERIODO DE EDIFICACIÓN

Con todo esto se ha decidido generar dos nuevas variables, que almacenan únicamente el siglo de inicio y el siglo de finalización respectivamente. Puesto que los siglos se encuentran debidamente ordenados, bastará con buscar la expresión regular de los siglos al inicio y al final de la cadena de caracteres almacenada en cada posición de **Siglo.Edificación**. Los atributos generados se han llamado **Siglo.Inicio** y **Siglo.Fin**. Esto se ha hecho en vista a los posteriores análisis estadísticos.

#### 10.3.1.10. IDENTIFICACIÓN DE REFORMAS

Se ha tratado de discernir también cuando se aportan fechas o siglos en los que se hicieron reformas. Para ello, se buscan patrones en los que se encuentre un siglo precedido por el término “*reforma*”, y se almacena en una nueva variable llamada **Reformas**. De forma similar se busca en el atributo **Historia**, porque como hemos podido comprobar muchas veces, las fechas son almacenadas en dicho atributo. No debe confundirse este tipo de intervenciones, que forman parte del proceso de edificación del bien patrimonial, de las que han sido hechas recientemente con fines de conservación, tales como las restauraciones.

#### 10.3.1.11. BATERÍA DE PRUEBAS

Las cadenas de entrada proporcionadas para los diferentes casos de prueba no se corresponden con ejemplos concretos del fichero de datos. Son casos creados para poner a prueba el sistema, en ocasiones más complejos de los que realmente aparecen en dicho fichero.

<b>Identificador de prueba</b>	<b>ÉPOCA_P1</b>
<b>Entrada</b>	1823
<b>Resultado esperado</b>	<b>Siglo.Edificación: XIX</b> <b>Era: D. C.</b> <b>Siglo.Inicio: XIX</b> <b>Siglo.Fin: XIX</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: XIX</b> <b>Era: D. C.</b> <b>Siglo.Inicio: XIX</b> <b>Siglo.Fin: XIX</b>

Tabla 10.3. Caso de prueba ÉPOCA\_P1

<b>Identificador de prueba</b>	<b>ÉPOCA_P2</b>
<b>Entrada</b>	18
<b>Resultado esperado</b>	<b>Siglo.Edificación: I</b> <b>Era: D. C.</b> <b>Siglo.Inicio: I</b> <b>Siglo.Fin: I</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: I</b> <b>Era: D. C.</b> <b>Siglo.Inicio: I</b> <b>Siglo.Fin: I</b>

Tabla 10.4. Caso de prueba ÉPOCA\_P2

<b>Identificador de prueba</b>	<b>ÉPOCA_P3</b>
<b>Entrada</b>	110
<b>Resultado esperado</b>	<b>Siglo.Edificación: II</b> <b>Era: D. C.</b> <b>Siglo.Inicio: II</b> <b>Siglo.Fin: II</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: II</b> <b>Era: D. C.</b> <b>Siglo.Inicio: II</b> <b>Siglo.Fin: II</b>

Tabla 10.5. Caso de prueba ÉPOCA\_P3

<b>Identificador de prueba</b>	<b>ÉPOCA_P4</b>
<b>Entrada</b>	1670-1780
<b>Resultado esperado</b>	<b>Siglo.Edificación: XVII-XVIII</b> <b>Era: D. C.</b> <b>Siglo.Inicio: XVII</b> <b>Siglo.Fin: XVIII</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: XVII-XVIII</b> <b>Era: D. C.</b> <b>Siglo.Inicio: XVII</b> <b>Siglo.Fin: XVIII</b>

Tabla 10.6. Caso de prueba ÉPOCA\_P4

<b>Identificador de prueba</b>	<b>ÉPOCA_P5</b>
<b>Entrada</b>	1670-1680
<b>Resultado esperado</b>	<b>Siglo.Edificación: XVII</b> <b>Era: D. C.</b> <b>Siglo.Inicio: XVII</b> <b>Siglo.Fin: XVII</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: XVII</b> <b>Era: D. C.</b> <b>Siglo.Inicio: XVII</b> <b>Siglo.Fin: XVII</b>

Tabla 10.7. Caso de prueba ÉPOCA\_P5

<b>Identificador de prueba</b>	<b>ÉPOCA_P6</b>
<b>Entrada</b>	190- 1100
<b>Resultado esperado</b>	<b>Siglo.Edificación: II-XII</b> <b>Era: D. C.</b> <b>Siglo.Inicio: II</b> <b>Siglo.Fin: XII</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: II-XII</b> <b>Era: D. C.</b> <b>Siglo.Inicio: II</b> <b>Siglo.Fin: XII</b>

Tabla 10.8. Caso de prueba ÉPOCA\_P6

<b>Identificador de prueba</b>	<b>ÉPOCA_P7</b>
<b>Entrada</b>	X - XI
<b>Resultado esperado</b>	<b>Siglo.Edificación: X-XI</b> <b>Era: D. C.</b> <b>Siglo.Inicio: X</b> <b>Siglo.Fin: XI</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: X-XI</b> <b>Era: D. C.</b> <b>Siglo.Inicio: X</b> <b>Siglo.Fin: XI</b>

Tabla 10.9. Caso de prueba ÉPOCA\_P7

<b>Identificador de prueba</b>	<b>ÉPOCA_P8</b>
<b>Entrada</b>	II d.c
<b>Resultado esperado</b>	<b>Siglo.Edificación: II</b> <b>Era: D. C.</b> <b>Siglo.Inicio: II</b> <b>Siglo.Fin: II</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: II</b> <b>Era: D. C.</b> <b>Siglo.Inicio: II</b> <b>Siglo.Fin: II</b>

Tabla 10.10. Caso de prueba ÉPOCA\_P8

<b>Identificador de prueba</b>	<b>ÉPOCA_P9</b>
<b>Entrada</b>	III a. C
<b>Resultado esperado</b>	<b>Siglo.Edificación: III</b> <b>Era: A. C.</b> <b>Siglo.Inicio: III</b> <b>Siglo.Fin: III</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: III</b> <b>Era: A. C.</b> <b>Siglo.Inicio: III</b> <b>Siglo.Fin: III</b>

Tabla 10.11. Caso de prueba ÉPOCA\_P9

<b>Identificador de prueba</b>	<b>ÉPOCA_P10</b>
<b>Entrada</b>	III dC. - IV dC
<b>Resultado esperado</b>	<b>Siglo.Edificación: III-IV</b> <b>Era: D. C.</b> <b>Siglo.Inicio: III</b> <b>Siglo.Fin: IV</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: III-IV</b> <b>Era: D. C.</b> <b>Siglo.Inicio: III</b> <b>Siglo.Fin: IV</b>

Tabla 10.12. Caso de prueba ÉPOCA\_P10

<b>Identificador de prueba</b>	<b>ÉPOCA_P11</b>
<b>Entrada</b>	III aC.-II a c
<b>Resultado esperado</b>	<b>Siglo.Edificación: III-II</b> <b>Era: A. C.</b> <b>Siglo.Inicio: III</b> <b>Siglo.Fin: II</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: III-II</b> <b>Era: A. C.</b> <b>Siglo.Inicio: III</b> <b>Siglo.Fin: II</b>

Tabla 10.13. Caso de prueba ÉPOCA\_P11

<b>Identificador de prueba</b>	<b>ÉPOCA_P12</b>
<b>Entrada</b>	III a C.- IV dC
<b>Resultado esperado</b>	<b>Siglo.Edificación: III-IV</b> <b>Era: A. C. D. C.</b> <b>Siglo.Inicio: III</b> <b>Siglo.Fin: IV</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: III-IV</b> <b>Era: A. C. D. C.</b> <b>Siglo.Inicio: III</b> <b>Siglo.Fin: IV</b>

Tabla 10.14. Caso de prueba ÉPOCA\_P12

<b>Identificador de prueba</b>	<b>ÉPOCA_P13</b>
<b>Entrada</b>	VI a C. -II dC
<b>Resultado esperado</b>	<b>Siglo.Edificación:</b> VI-II <b>Era:</b> A. C. D. C. <b>Siglo.Inicio:</b> VI <b>Siglo.Fin:</b> II
<b>Resultado obtenido</b>	<b>Siglo.Edificación:</b> VI-II <b>Era:</b> A. C. D. C. <b>Siglo.Inicio:</b> VI <b>Siglo.Fin:</b> II

Tabla 10.15. Caso de prueba ÉPOCA\_P13

<b>Identificador de prueba</b>	<b>ÉPOCA_P14</b>
<b>Entrada</b>	Romano
<b>Resultado esperado</b>	<b>Siglo.Edificación:</b> III-V <b>Era:</b> A. C. D. C. <b>Siglo.Inicio:</b> III <b>Siglo.Fin:</b> V
<b>Resultado obtenido</b>	<b>Siglo.Edificación:</b> III-V <b>Era:</b> A. C. D. C. <b>Siglo.Inicio:</b> III <b>Siglo.Fin:</b> V

Tabla 10.16. Caso de prueba ÉPOCA\_P14



<b>Identificador de prueba</b>	<b>ÉPOCA_P15</b>
<b>Entrada</b>	Romano y medieval
<b>Resultado esperado</b>	<b>Siglo.Edificación: V</b> <b>Era: D. C.</b> <b>Siglo.Inicio: V</b> <b>Siglo.Fin: V</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: V</b> <b>Era: D. C.</b> <b>Siglo.Inicio: V</b> <b>Siglo.Fin: V</b>

Tabla 10.17. Caso de prueba ÉPOCA\_P15

<b>Identificador de prueba</b>	<b>ÉPOCA_P16</b>
<b>Entrada</b>	Edad Media, hispanomusulmán
<b>Resultado esperado</b>	<b>Siglo.Edificación: VIII-XV</b> <b>Era: D. C.</b> <b>Siglo.Inicio: VIII</b> <b>Siglo.Fin: XV</b>
<b>Resultado obtenido</b>	<b>Siglo.Edificación: VIII-XV</b> <b>Era: D. C.</b> <b>Siglo.Inicio: VIII</b> <b>Siglo.Fin: XV</b>

Tabla 10.18. Caso de prueba ÉPOCA\_P16

<b>Identificador de prueba</b>	<b>ÉPOCA_P17</b>
<b>Entrada</b>	Medieval y bizantino
<b>Resultado esperado</b>	<b>Siglo.Edificación:</b> VI-VII <b>Era:</b> D. C. <b>Siglo.Inicio:</b> VI <b>Siglo.Fin:</b> VII
<b>Resultado obtenido</b>	<b>Siglo.Edificación:</b> VI-VII <b>Era:</b> D. C. <b>Siglo.Inicio:</b> VI <b>Siglo.Fin:</b> VII

Tabla 10.19. Caso de prueba ÉPOCA\_P17

<b>Identificador de prueba</b>	<b>ÉPOCA_P18</b>
<b>Entrada</b>	Romano, griego y gótica
<b>Resultado esperado</b>	<b>Siglo.Edificación:</b> III XIII-XV <b>Era:</b> A. C. D. C. <b>Siglo.Inicio:</b> III <b>Siglo.Fin:</b> XV
<b>Resultado obtenido</b>	<b>Siglo.Edificación:</b> III XIII-XV <b>Era:</b> A. C. D. C. <b>Siglo.Inicio:</b> III <b>Siglo.Fin:</b> XV

Tabla 10.20. Caso de prueba ÉPOCA\_P18

<b>Identificador de prueba</b>	<b>ÉPOCA_P19</b>
<b>Entrada</b>	Visigodo, taifa, barroco
<b>Resultado esperado</b>	<b>Siglo.Edificación:</b> V-VIII XI-XIII XVII-XVIII <b>Era:</b> D. C. <b>Siglo.Inicio:</b> V <b>Siglo.Fin:</b> XVIII
<b>Resultado obtenido</b>	<b>Siglo.Edificación:</b> V-VIII XI-XIII XVII-XVIII <b>Era:</b> D. C. <b>Siglo.Inicio:</b> V <b>Siglo.Fin:</b> XVIII

Tabla 10.21. Caso de prueba ÉPOCA\_P19

<b>Identificador de prueba</b>	<b>ÉPOCA_P20</b>
<b>Entrada</b>	X, reformas del siglo XV, iii
<b>Resultado esperado</b>	<b>Siglo.Edificación:</b> X <b>Era:</b> D. C. <b>Siglo.Inicio:</b> X <b>Siglo.Fin:</b> X <b>Reformas:</b> XV
<b>Resultado obtenido</b>	<b>Siglo.Edificación:</b> X <b>Era:</b> D. C. <b>Siglo.Inicio:</b> X <b>Siglo.Fin:</b> X <b>Reformas:</b> XV

Tabla 10.22. Caso de prueba ÉPOCA\_P20

<b>Identificador de prueba</b>	<b>ÉPOCA_P21</b>
<b>Entrada</b>	Paleolítico, edad del cobre
<b>Resultado esperado</b>	<b>Siglo.Edificación:</b> Anterior al s. MMMCMXCIX a. C. XXXVI-XXIII <b>Era:</b> A. C. <b>Siglo.Inicio:</b> Anterior al s. MMMCMXCIX a. C. <b>Siglo.Fin:</b> XXXII-XXIII
<b>Resultado obtenido</b>	<b>Siglo.Edificación:</b> Anterior al s. MMMCMXCIX a. C. XXXVI-XXIII <b>Era:</b> A. C. <b>Siglo.Inicio:</b> Anterior al s. MMMCMXCIX a. C. <b>Siglo.Fin:</b> XXXII-XXIII

Tabla 10.23. Caso de prueba ÉPOCA\_P21

<b>Identificador de prueba</b>	<b>ÉPOCA_P22</b>
<b>Entrada</b>	X-XX, XV
<b>Resultado esperado</b>	<b>Siglo.Edificación:</b> X-XX <b>Era:</b> D. C. <b>Siglo.Inicio:</b> X-XX <b>Siglo.Fin:</b> X-XX
<b>Resultado obtenido</b>	<b>Siglo.Edificación:</b> X-XX <b>Era:</b> D. C. <b>Siglo.Inicio:</b> X-XX <b>Siglo.Fin:</b> X-XX

Tabla 10.24. Caso de prueba ÉPOCA\_P22

<b>Identificador de prueba</b>	<b>ÉPOCA_P23</b>
<b>Entrada</b>	Renacimiento
<b>Historia</b>	Edificado entre 1560 y 1585. Fue desamortizado durante la época de Mendizábal y sufrió un incendio en 1860.
<b>Resultado esperado</b>	<b>SigloHistoria:</b> XVI <b>EraHistoria:</b> D. C.
<b>Resultado obtenido</b>	<b>SigloHistoria:</b> XVI <b>EraHistoria:</b> D. C.

Tabla 10.25. Caso de prueba ÉPOCA\_P23

<b>Identificador de prueba</b>	<b>ÉPOCA_P24</b>
<b>Entrada</b>	Renacimiento
<b>Historia</b>	Construido en el siglo XVII, fue usado como centro de operaciones durante la guerra civil española
<b>Resultado esperado</b>	<b>SigloHistoria:</b> XVII <b>EraHistoria:</b> D. C.
<b>Resultado obtenido</b>	<b>SigloHistoria:</b> XVII <b>EraHistoria:</b> D. C.

Tabla 10.26. Caso de prueba ÉPOCA\_P24

### 10.3.2. ATRIBUTO “INTERVENCIONES”

2017 El Ayuntamiento asume la reparación de la cubierta: "Estaba en muy mal estado y suponía un peligro para el mantenimiento del palacio"
Recientemente se han realizado algunas mejoras, acondicionamiento para la visita, limpieza, etc.
2009 La firma de ropa Pull and Bear elige el antiguo edificio como escenario para grabar la campana publicitaria de su nueva colección de ropa.
2004: Se llevaron a cabo reparaciones de urgencia insuficientes y poco respetuosas, así como de desescombros con maquinaria pesada, además de colocación de estructuras metálicas y tapiado de los huecos y lienzos derruidos, para impedir el desplome de muros, con un presupuesto de 235.778,73€ adjudicados por la Junta de Castilla y León, a la empresa Sopsa.

Figura 10.9. Muestras originales del atributo *Intervenciones*

<b>2017</b> El Ayuntamiento asume la <b>reparación</b> de la cubierta: "Estaba en muy mal estado y suponía un peligro para el mantenimiento del palacio"
Recientemente se han realizado algunas mejoras, acondicionamiento para la visita, <b>limpieza</b> , etc.
2009 La firma de ropa Pull and Bear elige el antiguo edificio como escenario para grabar la campana publicitaria de su nueva colección de ropa.
<b>2004</b> : Se llevaron a cabo <b>reparaciones</b> de urgencia insuficientes y poco respetuosas, así como de desescombros con maquinaria pesada, además de colocación de estructuras metálicas y tapiado de los huecos y lienzos derruidos, para impedir el desplome de muros, con un presupuesto de 235.778,73€ adjudicados por la Junta de Castilla y León, a la empresa Sopsa.

Figura 10.10. Representación de los términos a detectar por el sistema

Sí	2017
No	
No	
Sí	2004

Figura 10.11. Resultado de la ejecución

El atributo **Intervenciones** contiene información sobre distintos tipos de intervenciones de interés que hayan podido suceder a lo largo de la historia de un bien patrimonial. Su contenido, como se puede ver en la *figura 10.9*, se presenta nuevamente en forma de textos descriptivos de longitud variable. De esta variable nos interesa extraer, no solo si se ha producido una intervención previa, sino también el año en que fue realizada (que en muchos casos está documentado y queda reflejado). Cabe mencionar que estas intervenciones no siempre implican la salida del bien de la Lista Roja, únicamente referencian un dato de interés histórico.

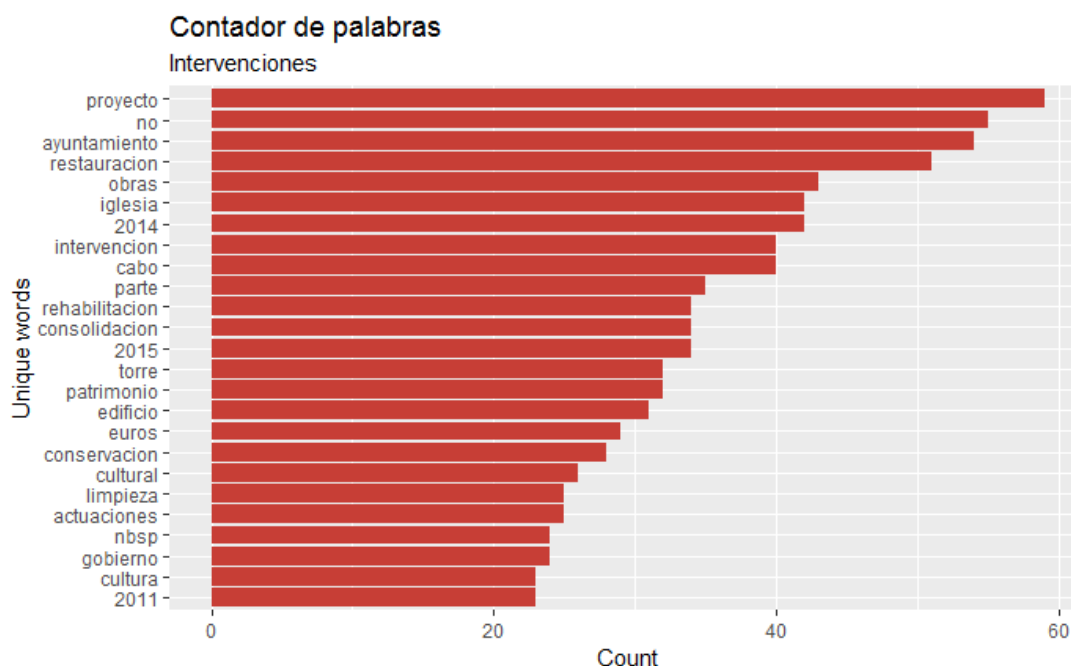


Figura 10.12. Términos más frecuentes hallados en *Intervenciones*

Para terminar de diferenciar este atributo con respecto a **Motivo.de.la.retirada**, y para que no haya redundancia, vamos a eliminar todas aquellas fichas en las que albergue información sobre las intervenciones que hayan supuesto la retirada del bien de la **Lista Roja**. Para ello, cuando se detecte una ficha que está incluida en la Lista Verde o en la Lista Negra, el contenido de Intervención será suprimido y copiado en la variable **Motivo.de.la.retirada** para que cualquier tipo de información de interés aquí reflejado pueda ser posteriormente analizado en dicha variable.

Una vez se haya limpiado el atributo de esta clase de intervenciones, se procede a extraer la información útil de ellos. En este caso se ha decidido hacerlo de forma muy sencilla. Nos basta con conocer si el bien ha sido intervenido o no y en qué año ha sido. Por lo que buscaremos la aparición de diferentes tipos de intervenciones, tales como “restauración”, “recuperación”, “consolidación”, “limpieza”, etc., y de aparecer almacenaremos un “S” en una nueva variable que llamaremos **Otras.intervenciones**. De manera similar, buscaremos apariciones de años en el atributo usando la expresión regular "[[:digit:]]+[[:digit:]]+[[:digit:]]+[[:digit:]]", y estos serán almacenados en otra variable que llamaremos **Fecha.de.intervención**.

Para posibles mejoras del sistema podríamos entrar a valorar el tipo de intervención y que ésta quedase reflejada en el contenido del nuevo atributo. Pero en este caso se ha decidido que no sea así para disponer de más atributos binarios en la fase de análisis y para diferenciar más esta variable de la ya mencionada **Motivo.de.la.retirada**, que sí contendrá los diversos tipos de intervención efectuados sobre los bienes patrimoniales. En dicha variable es crucial conocer la clase intervención para saber el motivo por el cual el bien pasa a la Lista Roja o la Lista Negra.

### 10.3.2.1. BATERÍA DE PRUEBAS

Las cadenas de entrada proporcionadas para los diferentes casos de prueba no se corresponden con ejemplos concretos del fichero de datos. Son casos creados para poner a prueba el sistema, en ocasiones más complejos de los que realmente aparecen en dicho fichero.

<b>Identificador de prueba</b>	<b>INTERVENCIÓN_P1</b>
<b>Entrada</b>	En el año 2002 se inicia un proyecto para la reparación del techo.
<b>Tipo</b>	Lista Verde
<b>Resultado esperado</b>	<b>Intervenciones:</b> <i>NULL</i> <b>Fecha.de.intervención:</b> <i>NULL</i>
<b>Resultado obtenido</b>	<b>Intervenciones:</b> <i>NULL</i> <b>Fecha.de.intervención:</b> <i>NULL</i>

Tabla 10.27. Caso de prueba INTERVENCIÓN\_P1

<b>Identificador de prueba</b>	<b>INTERVENCIÓN_P2</b>
<b>Entrada</b>	El complejo fue demolido en 2016.
<b>Tipo</b>	Lista Negra
<b>Resultado esperado</b>	<b>Intervenciones:</b> <i>NULL</i> <b>Fecha.de.intervención:</b> <i>NULL</i>
<b>Resultado obtenido</b>	<b>Intervenciones:</b> <i>NULL</i> <b>Fecha.de.intervención:</b> <i>NULL</i>

Tabla 10.28. Caso de prueba INTERVENCIÓN\_P2



<b>Identificador de prueba</b>	<b>INTERVENCIÓN_P3</b>
<b>Entrada</b>	En el año 2002 se inicia un proyecto para la reparación del techo.
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Intervenciones:</b> Sí <b>Fecha.de.intervención:</b> 2002
<b>Resultado obtenido</b>	<b>Intervenciones:</b> Sí <b>Fecha.de.intervención:</b> 2002

Tabla 10.29. Caso de prueba INTERVENCIÓN\_P3

<b>Identificador de prueba</b>	<b>INTERVENCIÓN_P4</b>
<b>Entrada</b>	Se llevan a cabo labores de limpieza y acondicionamiento del edificio.
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Intervenciones:</b> Sí <b>Fecha.de.intervención:</b> <i>NULL</i>
<b>Resultado obtenido</b>	<b>Intervenciones:</b> Sí <b>Fecha.de.intervención:</b> <i>NULL</i>

Tabla 10.30. Caso de prueba INTERVENCIÓN\_P4

<b>Identificador de prueba</b>	<b>INTERVENCIÓN_P5</b>
<b>Entrada</b>	El edificio está a punto de ser demolido.
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Intervenciones:</b> <i>NULL</i> <b>Fecha.de.intervención:</b> <i>NULL</i>
<b>Resultado obtenido</b>	<b>Intervenciones:</b> <i>NULL</i> <b>Fecha.de.intervención:</b> <i>NULL</i>

Tabla 10.31. Caso de prueba INTERVENCIÓN\_P5

### 10.3.3. ATRIBUTO “MOTIVO DE LA RETIRADA”

El Ministerio de Cultura comienza los trabajos de restauración y consolidación de la muralla. Un proyecto programado por el Instituto de Patrimonio Cultural de España (IPCE).
Se realizan obras de restauración en la cubierta y la fachada.
El Ayuntamiento comienza su consolidación.
Tanto la iglesia como el monasterio has sido rehabilitados quedando completamente consolidados

Figura 10.13. Muestras originales del atributo *Motivo.de.la.retirada*

El Ministerio de Cultura comienza los trabajos de <b>restauración</b> y <b>consolidación</b> de la muralla. Un proyecto programado por el Instituto de Patrimonio Cultural de España (IPCE).
Se realizan obras de <b>restauración</b> en la cubierta y la fachada.
El Ayuntamiento comienza su <b>consolidación</b> .
Tanto la iglesia como el monasterio has sido <b>rehabilitados</b> quedando completamente <b>consolidados</b>

Figura 10.14. Representación de los términos a detectar por el sistema

Restauración
Restauración
Consolidación
Consolidación

Figura 10.15. Resultado de la ejecución

El atributo **Motivo.de.la.retirada** contiene una breve descripción de las razones por las cuales un bien patrimonial ha salido de la Lista Roja, ya sea para pasar a la Lista Verde o para entrar a formar parte de la Lista Negra. En aquellos bienes que permanecen en la Lista Roja, información que queda reflejada en la variable **Tipo**, como hemos visto anteriormente, el contenido de la variable está vacío. No obstante, hay algunas excepciones que nos muestran bienes incluidos hoy en día en la Lista Roja que contienen información en **Motivo.de.la.retirada**, esto nos indica que estos bienes salieron de la Lista Roja y posteriormente volvieron a entrar. Podemos pensar que los bienes pasaron a la Lista Verde porque para incluirse en la Lista Negra su mal estado debe considerarse irreversible. Estas excepciones serán tratadas de manera idéntica al resto, pero de manera opcional crearemos una nueva variable que indique que un bien es reincidente en la Lista Roja.

El contenido de este atributo se presenta en forma de textos de longitud variable. El objetivo ha sido simplificar estos textos reduciéndolos a un único término que resuma las acciones realizadas sobre el bien inmueble que han desencadenado su salida de la

Lista Roja. Para extraer la información clave de este atributo se ha implementado un contador de palabras y generado un gráfico de barras que recoge las treinta palabras más frecuentes encontradas a lo largo de la variable, tal y como se aprecia en la *Figura 10.16*.

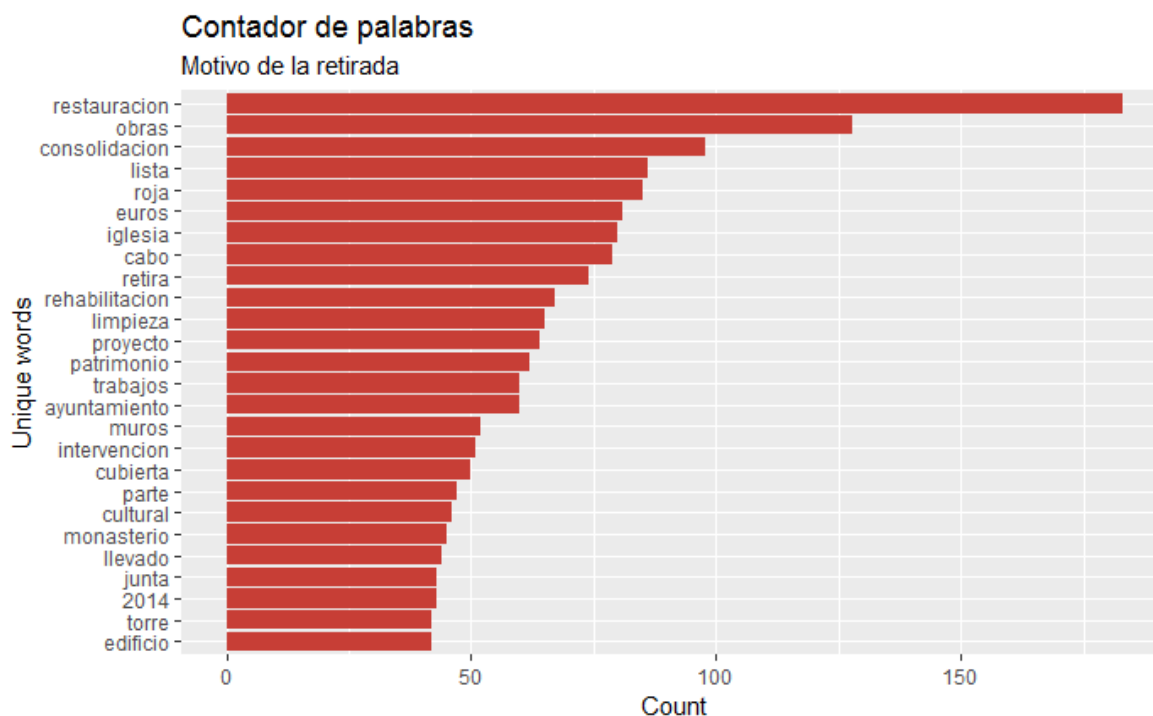


Figura 10.16. Términos más frecuentes hallados en *Motivo.de.la.retirada*

De este gráfico podemos extraer los motivos más frecuentes, tales como “restauración”, “consolidación”, “limpieza”, “recuperación”, “rehabilitación”, “intervención”. Esto conduce a pensar que se tratan de las más importantes, no tanto así las únicas que debemos valorar. La reducida cantidad de fichas pertenecientes a la Lista Negra implica que términos como “demolición” o “derrumbe” no se vean reflejados, por lo que debemos ampliar este diccionario.

Para contemplar todas las posibles, vamos a recurrir de forma frecuente a una técnica de lematización básica como es buscar únicamente la raíz esencial de cada término. Así pues, buscando por *“restaura[:alpha:]\*”*, donde *[:alpha:]* puede ser cualquier valor del alfabeto y \* nos indica que puede aparecer 0 o más veces, podemos detectar todas sus posibles variaciones: “restauración”, “restaurado”, “restaurada”, “restaurados”, “restauradas”, “restaurar”, etc. Para conjugaciones especiales de verbos irregulares, como “demoler” y “demuelan”, tendremos que valorar ambas raíces.

Una vez que hemos obtenido los términos, se ha tratado de agruparlos lo máximo posible para reducir el número de motivos de salida. Son los siguientes:

<b>Términos positivos</b>	<b>Acción positiva</b>
Restauración	Restauración
Recuperación	
Consolidación	Consolidación
Rehabilitación	Rehabilitación
Conservación	Conservación
Intervención	Intervención
Reparación	
Reformas	
Obras	
Subsanación	
Intervención	
Corrección	
Limpieza	
<b>Términos negativos</b>	<b>Acción negativos</b>
Demolición	Demolición
Derrumbe	
Derribe	
Pérdida (de sus valores)	Alteración de sus valores
Alteración	
Atentado	

Tabla 10.32. Motivos de la retirada: términos y acciones

Haremos una distinción entre aquellos motivos que han supuesto la salida de una ficha para su inclusión en la Lista Verde y aquellos para su inclusión en la Lista Negra, aplicando un condicional previamente a la búsqueda de términos. De esta forma

evitamos detectar términos erróneos en caso de que apareciesen mencionados en la descripción del atributo sin ser la causa principal. Por ejemplo, considérese la frase “*La restauración no fue aprobada y se procedió a su demolición*”. En esta frase el motivo real de la retirada es una demolición, pero existe la posibilidad de que el sistema detecte antes el término “*restauración*”.

Los términos han sido dispuestos en dos pares de vectores según la Lista a la cual conducen. Dos vectores con las raíces de los términos y dos vectores con las acciones. Y dentro de cada uno, siguen el orden dispuesto en la *tabla 10.32*, de tal forma que se le da prioridad a aquellas actividades más frecuentes y significativas frente a otras de carácter secundario que puedan ser mencionadas en el texto. Así pues, si el edificio ha sido restaurado y limpiado, lo catalogamos como restaurado al considerarse que es más significativo.

Tras las primeras ejecuciones se ha comprobado que el sistema cataloga correctamente la gran mayoría de los motivos de salida, no obstante, permanecen una serie de ellos que resultan de interés. En su mayoría son descripciones poco concisas que nos indican una desaparición de las amenazas a las que se veía sometido el bien patrimonial. Así pues, se ha decidido implementar un segundo nivel de criba de datos.

Términos positivos	Acción positiva
Paralización [de obras]	Desaparición de amenazas
Anulación [de proyecto]	
Prohibición	
Protección	
Traslado	

Tabla 10.33. Anexo a Motivos de la retirada: términos y acciones

Todas aquellas fichas pertenecientes a la Lista Verde en las que no se haya detectado ninguno de los motivos “positivos”, se cumplimentarán con la etiqueta genérica “Otros”, y todas aquellas fichas pertenecientes a la Lista Negra en las que no se haya detectado ninguno de los motivos “negativos” listados, se cumplimentarán también con la etiqueta “Otros”. Finalmente, podrían existir fichas pertenecientes a la Lista Roja cuyo atributo **Motivo.de.la.retirada** no se encontrase vacío, pero en las que no se haya detectado ninguno de los motivos “positivos” ni “negativos” listados. En estos casos podemos deducir dos posibilidades: o que se trata de un fallo y por lo tanto debería estar vacío, o que es una ficha “reincidente” pero su motivo de retirada no es detectado por

nuestro sistema. Como para saber esto habría que revisar los casos manualmente, vamos a cumplimentarlos con un signo de interrogación, “?”.

Las fichas “reincidentes”, es decir, que estén en la Lista Roja pero tengan algo de información en **Motivo.de.la.retirada**, son almacenadas en un nuevo atributo que denominamos **Reincidencia**. El valor por defecto de este atributo será “No”, excepto en los casos de las fichas de la Lista Roja en que se detecte contenido en **Motivo.de.la.retirada** y este contenido no sea una interrogación. En caso de contener una interrogación, se almacenará también una interrogación en **Reincidencia**.

### 10.3.3.1. BATERÍA DE PRUEBAS

Las cadenas de entrada proporcionadas para los diferentes casos de prueba no se corresponden con ejemplos concretos del fichero de datos. Son casos creados para poner a prueba el sistema, en ocasiones más complejos de los que realmente aparecen en dicho fichero.

<b>Identificador de prueba</b>	<b>MOTIVO_P1</b>
<b>Entrada</b>	El edificio ha sido restaurado
<b>Tipo</b>	Lista Verde
<b>Resultado esperado</b>	<b>Motivo:</b> Restauración <b>Reincidencia:</b> No
<b>Resultado obtenido</b>	<b>Motivo:</b> Restauración <b>Reincidencia:</b> No

Tabla 10.34. Caso de prueba MOTIVO\_P1

<b>Identificador de prueba</b>	<b>MOTIVO_P2</b>
<b>Entrada</b>	El edificio ha sido recientemente demolido para levantar bloques de viviendas.
<b>Tipo</b>	Lista Negra
<b>Resultado esperado</b>	<b>Motivo:</b> Demolición <b>Reincidencia:</b> No
<b>Resultado obtenido</b>	<b>Motivo:</b> Demolición <b>Reincidencia:</b> No

Tabla 10.35. Caso de prueba MOTIVO\_P2

<b>Identificador de prueba</b>	<b>MOTIVO_P3</b>
<b>Entrada</b>	Se han llevado a cabo obras de consolidación.
<b>Tipo</b>	Lista Verde
<b>Resultado esperado</b>	<b>Motivo:</b> Consolidación <b>Reincidencia:</b> No
<b>Resultado obtenido</b>	<b>Motivo:</b> Consolidación <b>Reincidencia:</b> No

Tabla 10.36. Caso de prueba MOTIVO\_P3

<b>Identificador de prueba</b>	<b>MOTIVO_P4</b>
<b>Entrada</b>	Pérdida total de sus valores
<b>Tipo</b>	Lista Negra
<b>Resultado esperado</b>	<b>Motivo:</b> Alteración de sus valores <b>Reincidencia:</b> No
<b>Resultado obtenido</b>	<b>Motivo:</b> Alteración de sus valores <b>Reincidencia:</b> No

Tabla 10.37. Caso de prueba MOTIVO\_P4

<b>Identificador de prueba</b>	<b>MOTIVO_P5</b>
<b>Entrada</b>	El ayuntamiento paraliza los planes de demolición, desapareciendo todas las amenazas.
<b>Tipo</b>	Lista Verde
<b>Resultado esperado</b>	<b>Motivo:</b> Desaparición de amenazas <b>Reincidencia:</b> No
<b>Resultado obtenido</b>	<b>Motivo:</b> Desaparición de amenazas <b>Reincidencia:</b> No

Tabla 10.38. Caso de prueba MOTIVO\_P5

<b>Identificador de prueba</b>	<b>MOTIVO_P6</b>
<b>Entrada</b>	El ayuntamiento ha efectuado una inversión de dinero que ha facilitado su restauración
<b>Tipo</b>	Lista Verde
<b>Resultado esperado</b>	<b>Motivo:</b> Restauración <b>Reincidencia:</b> No
<b>Resultado obtenido</b>	<b>Motivo:</b> Restauración <b>Reincidencia:</b> No

Tabla 10.39. Caso de prueba MOTIVO\_P6



<b>Identificador de prueba</b>	<b>MOTIVO_P7</b>
<b>Entrada</b>	Se han efectuado labores de limpieza de los frescos, consolidación de los muros
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Motivo:</b> Consolidación <b>Reincidencia:</b> Si
<b>Resultado obtenido</b>	<b>Motivo:</b> Consolidación <b>Reincidencia:</b> Si

Tabla 10.40. Caso de prueba MOTIVO\_P7

<b>Identificador de prueba</b>	<b>MOTIVO_P8</b>
<b>Entrada</b>	La torre ha sido restaurada
<b>Tipo</b>	Lista Negra
<b>Resultado esperado</b>	<b>Motivo:</b> Otros <b>Reincidencia:</b> No
<b>Resultado obtenido</b>	<b>Motivo:</b> Otros <b>Reincidencia:</b> No

Tabla 10.41. Caso de prueba MOTIVO\_P8

<b>Identificador de prueba</b>	<b>MOTIVO_P9</b>
<b>Entrada</b>	Convento derrumbado
<b>Tipo</b>	Lista Verde
<b>Resultado esperado</b>	<b>Motivo:</b> Otros <b>Reincidencia:</b> No
<b>Resultado obtenido</b>	<b>Motivo:</b> Otros <b>Reincidencia:</b> No

Tabla 10.42. Caso de prueba MOTIVO\_P9

<b>Identificador de prueba</b>	<b>MOTIVO_P10</b>
<b>Entrada</b>	Hola mundo
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Motivo:</b> "?" <b>Reincidencia:</b> "?"
<b>Resultado obtenido</b>	<b>Motivo:</b> "?" <b>Reincidencia:</b> "?"

Tabla 10.43. Caso de prueba MOTIVO\_P10

### 10.3.4. ATRIBUTO “PROPIEDAD”

Pública
Privada
Junta de Andalucía
Ayuntamiento de Maella
Obispado de Santander
Ayuntamiento de Cuevas de Canart (en 2015 el Obispado de Teruel y Albarracín se lo vende al ayto.)
En la actualidad el edificio está dividido en viviendas pertenecientes a varios propietarios particulares

Figura 10.17. Muestras originales del atributo *Propiedad*

<b>Pública</b>
<b>Privada</b>
<b>Junta de Andalucía</b>
<b>Ayuntamiento de Maella</b>
<b>Obispado de Santander</b>
<b>Ayuntamiento de Cuevas de Canart (en 2015 el Obispado de Teruel y Albarracín se lo vende al ayto.)</b>
<b>En la actualidad el edificio está dividido en viviendas pertenecientes a varios propietarios <b>particulares</b></b>

Figura 10.18. Representación de los términos a detectar por el sistema

Pública
Privada
Gobierno Autonómico
Ayuntamiento
Obispado
Ayuntamiento
Particular

Figura 10.19. Resultado de la ejecución

El atributo **Propiedad** contiene información sobre el propietario de los bienes inmuebles almacenados en la Lista Roja. Su contenido, como se puede ver en la *figura 10.17*, se presenta en ocasiones de forma ordenada, en otros casos de forma más general (“*pública*” o “*privado*”) y otras veces aporta datos más precisos (“*Junta de Andalucía*”, “*Obispado de Santander*”). Pero no siempre es información ordenada o concisa, encontrándose frases con carácter descriptivo como las expuestas al final de la

figura anteriormente indicada. Así pues, nos interesa conocer si el bien es de propiedad pública o privada (o ambas) y, si tenemos acceso a información más específica, conocer qué clase de entidad es la propietaria.

Antes de entrar a extraer información, es importante tener en cuenta que un propietario puede ceder un bien a otra entidad o persona física. Así pues, vamos a eliminar “*cedido a / al / a / a la A*”. Con esto evitaremos que, sea quien sea “A”, pueda ser tomado como propietario por nuestro script. De igual forma, es posible que encontremos información dentro de este atributo que indique que un ex-propietario ha vendido el bien patrimonial a su nuevo propietario. Por tanto, vamos a eliminar también las construcciones “*vendido por*” o “*comprado a*”.

No obstante, existe un riesgo de cometer errores. Construcciones como “*cedido por*”, “*comprado por*” o “*vendido a*” deben mantenerse porque indican quién es el propietario real del bien. Sin embargo, puede darse el caso de que aparezcan así: “*cedido por A a B*”, “*comprado por A a B*” o “*vendido a A por B*”. En tal caso, el sistema detectaría tanto al propietario como al beneficiario de la cesión o el vendedor, quienes no pueden figurar como propietarios. Para solucionar esto, cada vez que las construcciones “*cedido por A*”, “*comprado por A*” o “*vendido a A*” aparezcan, serán extraídas con la función *strapply(data, pattern)* y sobrescribirán el contenido del atributo, borrando todo lo que quedase fuera de su expresión regular. De igual forma se hará con las construcciones “*cedido por el / la*”, “*comprado por el / la*” y “*vendido al / la*”.

Para analizar en profundidad el contenido del atributo se han generado nuevamente una serie de gráficas que muestran las palabras más frecuentes encontradas a lo largo de la variable y del emparejamiento de términos frecuentes, en ambos casos habiendo suprimido previamente las palabras vacías. Estos gráficos pueden observarse en las *figuras 10.20* y *10.21*. Podemos observar como existe una fuerte relación de emparejamiento entre las palabras “público” y “privado”. Por lo cual, durante la implementación de este script tendremos que valorar, no solo la aparición de múltiples propietarios, sino que estos sean de distinta titularidad.

Con los resultados obtenidos de las gráficas se ha creado una lista que almacena cada uno de estos términos. No obstante se han incluido otros términos valorando en todo momento que el fichero de datos pudiese ser ampliada y que apareciesen estas palabras. Se han valorado los principales Ministerios con bienes patrimoniales arquitectónicos, así como todas las posibles formas con las que se pueden denominar los gobiernos de cada autonomía. Cabe mencionar que también ha sido necesario tratar los acrónimos S.A. y S.L. antes de hacer la limpieza de palabras vacías, unificando su formato en letras mayúsculas para evitar que la preposición “a” sea confundida con la inicial de S.A.

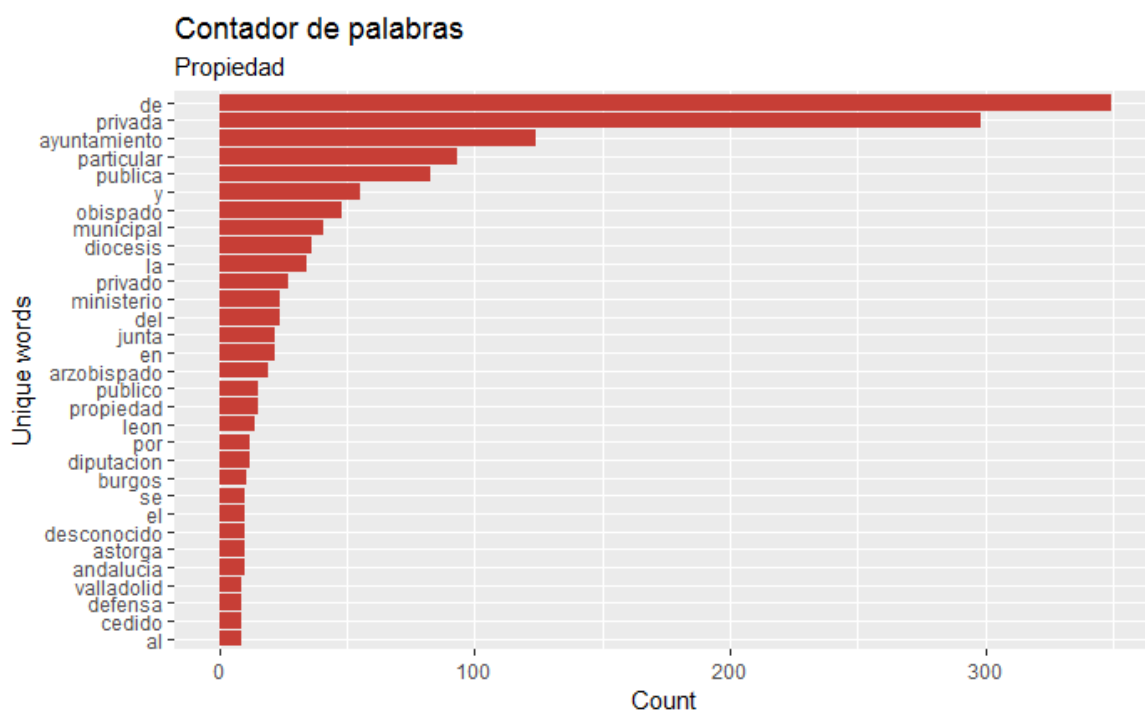


Figura 10.20. Términos más frecuentes hallados en *Propiedad*

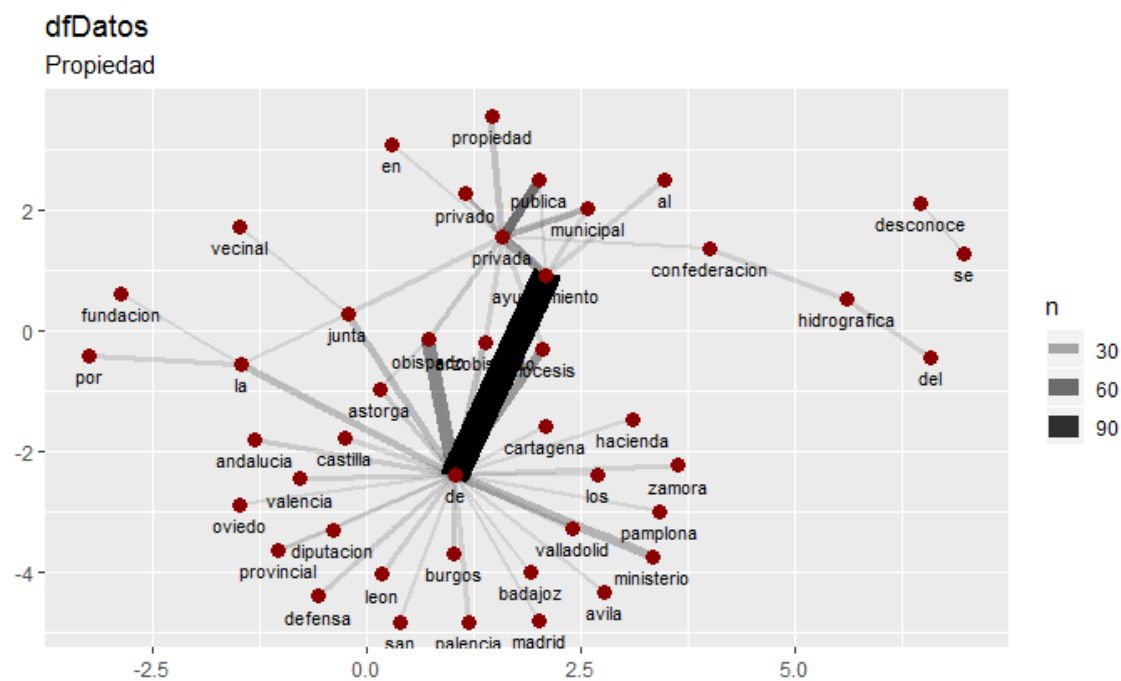


Figura 10.21. Emparejamientos más frecuentes de términos hallados en *Propiedad*

Propietario	Clasificación según su titularidad	
Particular	Privado	
Arzobispado, obispado, diócesis, orden		
Empresa, banco, S.A., S.L.		
Privado		
Municipal, ayuntamiento, ayto.	Público	
Ministerio de Hacienda, Ministerio de Defensa, Ministerio de Interior, Ministerio de Cultura, Ministerio de Educación, Ministerio de Fomento		
Gobierno Autonómico, Diputación, Junta de Andalucía, Junta de Extremadura, Junta de Castilla y León, Junta de Comunidades de Castilla-La Mancha, Junta/Xunta de Galicia, Gobierno Vasco, Gobierno del País Vasco, Gobierno Navarro/de Navarra, Govern/Gobierno Catalán, Gobierno Balear, Gobierno Cántabro, Gobierno de La Rioja, Diputación Aragonesa/de Aragón, Gobierno Aragonés, Gobierno Asturiano/del Principado de Asturias...		
Gobierno, Estado		
Banco de España		
Patrimonio Nacional		
Confederación Hidrográfica		
Público		
Fundación		Desconocido
Desconocido		

Tabla 10.44. Clasificación de propietarios

No es necesario establecer distinciones entre ayuntamientos; asumimos que en el caso de pertenecer un bien patrimonial a un ayuntamiento, será al de la localidad en la que se sitúa, información que queda recogida en el atributo **Localidad**. De igual forma lo hacemos con las juntas autonómicas o provinciales y con los obispados y arzobispados.

Con la información extraída de este atributo se han generado dos nuevas variables. La primera, que hemos denominado **Propietario**, va a contener la información más específica sobre la entidad. La segunda, **Propiedad**, nos indicará únicamente si es “Privado”, “Público” o “Desconocido”. Teniendo en cuenta esto se han valorado también los casos “Público y privado”, “Privado y otros” y “Público y otros”. Estas dos últimas aparecerán cuando, además de un propietario conocido, público o privado, exista otro propietario desconocido, lo cual nos impida determinar si es un bien compartido o no.

#### 10.3.4.1. BATERÍA DE PRUEBAS

Las cadenas de entrada proporcionadas para los diferentes casos de prueba no se corresponden con ejemplos concretos del fichero de datos. Son casos creados para poner a prueba el sistema, en ocasiones más complejos de los que realmente aparecen en dicho fichero.

<b>Identificador de prueba</b>	<b>PROPIEDAD_P1</b>
<b>Entrada</b>	Fue vendido por Jaime a Hermanos Ramírez S.L.
<b>Resultado esperado</b>	<b>Propietario:</b> Empresa <b>Propiedad:</b> Privada
<b>Resultado obtenido</b>	<b>Propietario:</b> Empresa <b>Propiedad:</b> Privada

Tabla 10.45. Caso de prueba PROPIEDAD\_P1

<b>Identificador de prueba</b>	<b>PROPIEDAD_P2</b>
<b>Entrada</b>	Cedido por el obispado al Ministerio de Fomento
<b>Resultado esperado</b>	<b>Propietario:</b> Obispado <b>Propiedad:</b> Privada
<b>Resultado obtenido</b>	<b>Propietario:</b> Obispado <b>Propiedad:</b> Privada

Tabla 10.46. Caso de prueba PROPIEDAD\_P2

<b>Identificador de prueba</b>	<b>PROPIEDAD_P3</b>
<b>Entrada</b>	Cedido al ayuntamiento
<b>Resultado esperado</b>	<b>Propietario:</b> Otros <b>Propiedad:</b> Desconocido
<b>Resultado obtenido</b>	<b>Propietario:</b> Otros <b>Propiedad:</b> Desconocido

Tabla 10.47. Caso de prueba PROPIEDAD\_P3

<b>Identificador de prueba</b>	<b>PROPIEDAD_P4</b>
<b>Entrada</b>	Propiedad privada
<b>Resultado esperado</b>	<b>Propietario:</b> Privada <b>Propiedad:</b> Privada
<b>Resultado obtenido</b>	<b>Propietario:</b> Privada <b>Propiedad:</b> Privada

Tabla 10.48. Caso de prueba PROPIEDAD\_P4

<b>Identificador de prueba</b>	<b>PROPIEDAD_P5</b>
<b>Entrada</b>	Xunta de Galicia y Arzobispado
<b>Resultado esperado</b>	<b>Propietario:</b> Gobierno Regional, Arzobispado <b>Propiedad:</b> Pública, Privada
<b>Resultado obtenido</b>	<b>Propietario:</b> Gobierno Regional, Arzobispado <b>Propiedad:</b> Pública, Privada

Tabla 10.49. Caso de prueba PROPIEDAD\_P5



<b>Identificador de prueba</b>	<b>PROPIEDAD_P6</b>
<b>Entrada</b>	Fundación Botín
<b>Resultado esperado</b>	<b>Propietario:</b> Fundación <b>Propiedad:</b> Desconocido
<b>Resultado obtenido</b>	<b>Propietario:</b> Fundación <b>Propiedad:</b> Desconocido

Tabla 10.50. Caso de prueba PROPIEDAD\_P6

<b>Identificador de prueba</b>	<b>PROPIEDAD_P7</b>
<b>Entrada</b>	Banco Santander
<b>Resultado esperado</b>	<b>Propietario:</b> Empresa <b>Propiedad:</b> Privada
<b>Resultado obtenido</b>	<b>Propietario:</b> Empresa <b>Propiedad:</b> Privada

Tabla 10.51. Caso de prueba PROPIEDAD\_P7

<b>Identificador de prueba</b>	<b>PROPIEDAD_P8</b>
<b>Entrada</b>	Banco de España
<b>Resultado esperado</b>	<b>Propietario:</b> Estado <b>Propiedad:</b> Pública
<b>Resultado obtenido</b>	<b>Propietario:</b> Estado <b>Propiedad:</b> Pública

Tabla 10.52. Caso de prueba PROPIEDAD\_P8

<b>Identificador de prueba</b>	<b>PROPIEDAD_P9</b>
<b>Entrada</b>	Diócesis de Madrid, propiedad privada
<b>Resultado esperado</b>	<b>Propietario:</b> Diócesis <b>Propiedad:</b> Privada
<b>Resultado obtenido</b>	<b>Propietario:</b> Diócesis <b>Propiedad:</b> Privada

Tabla 10.53. Caso de prueba PROPIEDAD\_P9

<b>Identificador de prueba</b>	<b>PROPIEDAD_P10</b>
<b>Entrada</b>	Propiedad de Patrimonio Nacional
<b>Resultado esperado</b>	<b>Propietario:</b> Patrimonio Nacional <b>Propiedad:</b> Pública
<b>Resultado obtenido</b>	<b>Propietario:</b> Patrimonio Nacional <b>Propiedad:</b> Pública

Tabla 10.54. Caso de prueba PROPIEDAD\_P10

<b>Identificador de prueba</b>	<b>PROPIEDAD_P11</b>
<b>Entrada</b>	Se desconoce quién es el propietario
<b>Resultado esperado</b>	<b>Propietario:</b> Desconocido <b>Propiedad:</b> Desconocido
<b>Resultado obtenido</b>	<b>Propietario:</b> Desconocido <b>Propiedad:</b> Desconocido

Tabla 10.55. Caso de prueba PROPIEDAD\_P11

<b>Identificador de prueba</b>	<b>PROPIEDAD_P12</b>
<b>Entrada</b>	Hola mundo
<b>Resultado esperado</b>	<b>Propietario:</b> Otros <b>Propiedad:</b> Desconocido
<b>Resultado obtenido</b>	<b>Propietario:</b> Otros <b>Propiedad:</b> Desconocido

Tabla 10.56. Caso de prueba PROPIEDAD\_P12

### 10.3.5. ATRIBUTO “GRADO DE PROTECCIÓN LEGAL”

Bien de Interés Cultural con categoría de Lugar Cultural (Sitio Histórico), 1 de abril de 2004.
Genérica Ley Patrimonio Histórico 16/1985
Sin protección específica
Declarado BIC como monumento histórico-artístico el 21 de marzo de 1970
La consejería esta en proceso de incoar expediente BIC desde 2007.

Figura 10.22. Muestras originales del atributo *Grado.de.protección.legal*

<b>Bien de Interés Cultural</b> con categoría de Lugar Cultural (Sitio Histórico), 1 de abril de 2004.
Genérica Ley Patrimonio Histórico <b>16/1985</b>
<b>Sin protección</b> específica
Declarado <b>BIC</b> como monumento histórico-artístico el 21 de marzo de 1970
La consejería esta en proceso de <b>incoar</b> expediente BIC desde 2007.

Figura 10.23. Representación de los términos a detectar por el sistema

BIC
16/1985
Sin protección
BIC
Expediente incoado

Figura 10.24. Resultado de la ejecución

El atributo **Grado.de.protección.legal**, como ya se ha descrito anteriormente, contiene información sobre el tipo de protección que tienen los bienes inmuebles almacenados en la Lista Roja. Su contenido, como se puede ver en la *figura 10.22*, se presenta nuevamente en forma de textos descriptivos de longitud variable. De este atributo nos interesa extraer, no solo el tipo de protección concedido a cada bien patrimonial, sino también el año de concesión, que en muchos casos está documentado y queda reflejado.

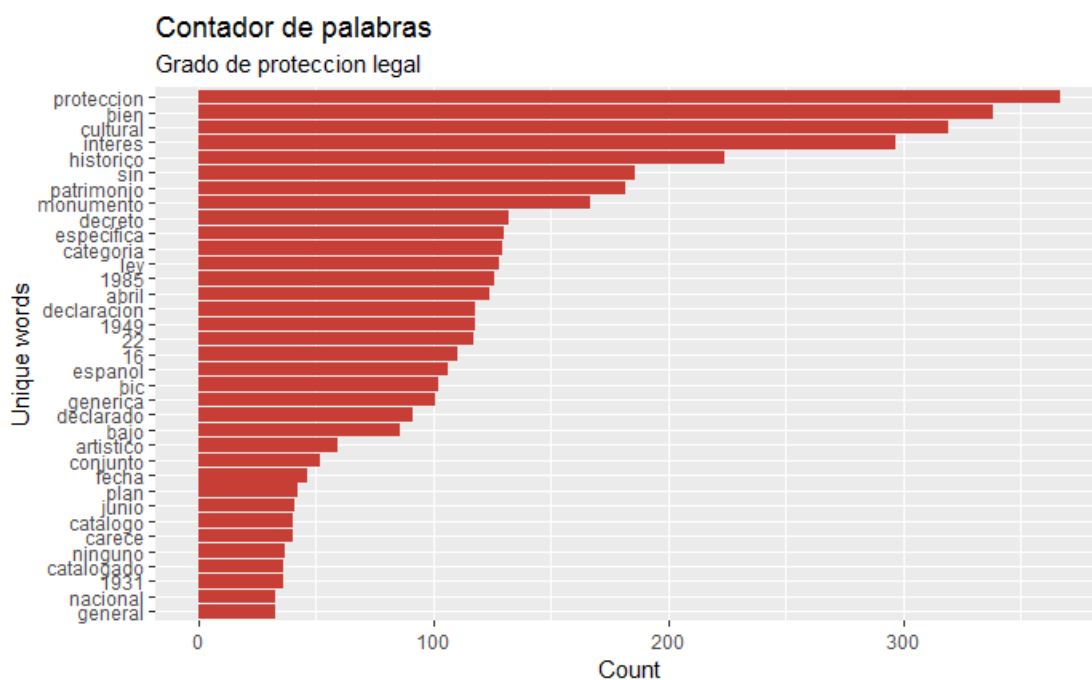


Figura 10.25. Términos más frecuentes hallados en *Grado.de.protección.legal*

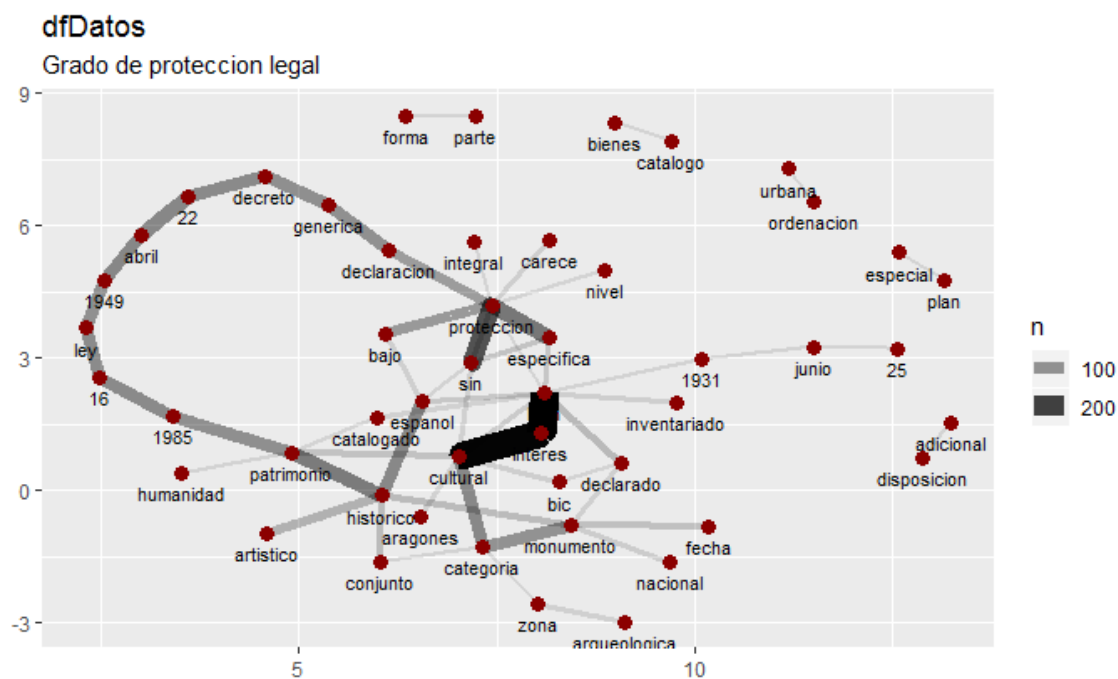


Figura 10.26. Emparejamientos más frecuentes de términos hallados en *Grado.de.protección.legal*

En este caso el gráfico de emparejamiento de palabras nos ofrece más información que el histograma de términos frecuentes. En el gráfico de emparejamiento podemos apreciar como destacan las construcciones “bienes interés cultural”, “sin protección específica”, “bajo protección específica”, “categoría monumento”, “ordenación urbana” o las diferentes leyes y decretos. El histograma por su parte confirma la relevancia de los diferentes términos que conforman estas categorías, pero no nos permite conocer los nombres reales de los grados de protección.

Tras analizar reiteradas veces el contenido se han valorado los siguientes tipos de protección:

### **Patrimonio Histórico Español**

Reúne todos los bienes inmuebles y muebles de interés histórico, artístico, arqueológico, paleontológico, etnográfico, científico o técnico. Se considera el grado mínimo de protección de cualquier bien. [26]

#### **BIC (Bien de Interés Cultural)**

Los bienes con categoría de BIC están incluidos en el Registro General de Bienes de Interés Cultural e incluye las siguientes categorías: Monumento histórico, Jardín histórico, Conjunto histórico, Sitio histórico, Zona arqueológica, Lugares de interés etnológico.

#### **Inventario General de Bienes Muebles**

Se trata de un registro administrativo adscrito al Ministerio de Cultura que recoge aquellos bienes patrimoniales que, pese a no haber sido catalogados como BIC, poseen una relevancia o valor singular.

#### **BRL (Bien de Relevancia Local)**

Se trata de una categoría establecida en algunas comunidades autónomas, tales como la Comunidad Valenciana, Cataluña o Aragón, en las cuales adquiere diferentes nombres, pero equivalentes entre sí. Recoge los bienes del Inventario General de Bienes Muebles de sus territorios, es decir, aquellos bienes de relevancia que no han sido catalogados como BIC.

#### **Ley 16/1985**

La Ley 16/1985, también conocida como Ley del Patrimonio Histórico Español, recogida en el BOE, es la que regula la declaración de Bienes de Interés Cultural pero entra también a valorar la confección de un Inventario General de Bienes Muebles del Patrimonio Histórico Español que, pese a no ser considerados BIC, tengan singular relevancia (artículo veintiséis) [27], por lo que comprende las categorías anteriormente descritas.

### Decreto del 22 de abril de 1949

Este decreto publicado en el BOE, y vigente durante los años del Franquismo, establecía que todos los castillos de España quedan bajo la protección del Estado con el objetivo de promover su conservación y la realización de un inventario. [28]

### PGOU (Plan General de Ordenación Urbana)

Planes que establecen las directrices de ordenación del territorio, el asentamiento urbano y las actividades que lo condicionan. Entran a valorar la protección patrimonial ya no solo de los monumentos o edificios singulares, sino también de su entorno. [29]

### PEPCH (Plan Especial de Protección del Casco Histórico)

Planes establecidos en el ámbito local para la protección de los conjuntos histórico artísticos que conforman el centro histórico de una urbe para lograr así la conservación de sus características arquitectónicas.

### Otros

Incluye inventarios de menor relevancia, catálogos o protecciones específicas que no nos permiten conocer con certeza si los bienes se encuentran en alguna de las categorías previamente descritas.

Se ha implementado nuevamente un script que permite buscar las diferentes categorías de protección dentro del atributo, así como fechas que puedan indicarnos el año en que entró en vigencia la protección. Esta información es almacenada en dos nuevas variables que denominamos **Grado.de.protección** y **Fecha.de.protección**.

#### 10.3.5.1. BATERÍA DE PRUEBAS

Las cadenas de entrada proporcionadas para los diferentes casos de prueba no se corresponden con ejemplos concretos del fichero de datos. Son casos creados para poner a prueba el sistema, en ocasiones más complejos de los que realmente aparecen en dicho fichero.

<b>Identificador de prueba</b>	<b>PROTECCIÓN_P1</b>
<b>Entrada</b>	Inventario de bienes de interés cultural desde 1935
<b>Resultado esperado</b>	<b>Grado.de.protección:</b> BIC <b>Fecha.de.protección:</b> 1935
<b>Resultado obtenido</b>	<b>Grado.de.protección:</b> BIC <b>Fecha.de.protección:</b> 1935

Tabla 10.57. Caso de prueba PROTECCIÓN\_P1

<b>Identificador de prueba</b>	<b>PROTECCIÓN_P2</b>
<b>Entrada</b>	El ayuntamiento ha incoado el expediente
<b>Resultado esperado</b>	<b>Grado.de.protección:</b> Expediente incoado
<b>Resultado obtenido</b>	<b>Grado.de.protección:</b> Expediente incoado

Tabla 10.58. Caso de prueba PROTECCIÓN\_P2

<b>Identificador de prueba</b>	<b>PROTECCIÓN_P3</b>
<b>Entrada</b>	Bajo la protección de Ley 16/1985
<b>Resultado esperado</b>	<b>Grado.de.protección:</b> 16/1985
<b>Resultado obtenido</b>	<b>Grado.de.protección:</b> 16/1985

Tabla 10.59. Caso de prueba PROTECCIÓN\_P3

<b>Identificador de prueba</b>	<b>PROTECCIÓN_P4</b>
<b>Entrada</b>	Sin protección específica
<b>Resultado esperado</b>	<b>Grado.de.protección:</b> Sin protección
<b>Resultado obtenido</b>	<b>Grado.de.protección:</b> Sin protección

Tabla 10.60. Caso de prueba PROTECCIÓN\_P4



<b>Identificador de prueba</b>	<b>PROTECCIÓN_P5</b>
<b>Entrada</b>	Decreto de protección especial
<b>Resultado esperado</b>	<b>Grado.de.protección:</b> Se desconoce
<b>Resultado obtenido</b>	<b>Grado.de.protección:</b> Se desconoce

Tabla 10.61. Caso de prueba PROTECCIÓN\_P5

<b>Identificador de prueba</b>	<b>PROTECCIÓN_P6</b>
<b>Entrada</b>	En 2004 se incluye en la lista de bienes de relevancia local
<b>Resultado esperado</b>	<b>Grado.de.protección:</b> Bien de Relevancia Local <b>Fecha.de.protección:</b> 2004
<b>Resultado obtenido</b>	<b>Grado.de.protección:</b> Bien de Relevancia Local <b>Fecha.de.protección:</b> 2004

Tabla 10.62. Caso de prueba PROTECCIÓN\_P6

### 10.3.6. ATRIBUTO “ESTADO DE CONSERVACIÓN”

Semiderruido después las lluvias y las crecidas del río del año 2013.
Techumbre hundida. Destejada y en progresiva ruina. Necesita urgente restauración.
Abandonado y en ruina progresiva con acumulación de vegetación.
En ruinas. Víctima de saqueos. Sólo se conservan las ruinas de la iglesia.
Pérdida del edificio por el abandono del mismo y la desprotección ante las inclemencias meteorológicas, la vegetación invasiva y los actos vandálicos.

Figura 10.27. Muestras originales del atributo *Estado.de.conservación*

Semiderruido después las lluvias y las crecidas del río del año 2013.
Techumbre <b>hundida</b> . Destejada y en <b>progresiva ruina</b> . Necesita urgente restauración.
<b>Abandonado</b> y en ruina <b>progresiva</b> con acumulación de <b>vegetación</b> .
En <b>ruinas</b> . Víctima de <b>saqueos</b> . Sólo se conservan las <b>ruinas</b> de la iglesia.
Pérdida del edificio por el <b>abandono</b> del mismo y la desprotección ante las inclemencias meteorológicas, la <b>vegetación</b> invasiva y los actos <b>vandálicos</b> .

Figura 10.28. Representación de los términos a detectar por el sistema

80
$80 + 90 = 170$
$60 + 90 + 20 = 170$
$90 + 40 = 130$
$60 + 20 + 60 = 140$

Figura 10.29. Representación de la cuantización de las variables

Amarillo
Naranja
Naranja
Naranja
Naranja

Figura 10.30. Resultado de la ejecución

El atributo **Estado.de.conservación** contiene un texto descriptivo que ofrece información relevante. El objetivo del tratamiento de este atributo ha sido cuantificar el grado de deterioro del bien. Esto ha supuesto otro de los grandes retos de este proyecto, ya que estamos hablando de convertir textos en variables numéricas. Y por supuesto, la objetividad de estos textos es relativa en cuanto a que han sido redactados por personas diferentes para las que cada adjetivo puede significar algo distinto.

Debido a la frecuente escasez de consistencia en el fichero de datos, múltiples fichas no contienen ningún tipo de información dentro de este atributo. A su vez, se ha comprobado que parte de la información del estado de conservación queda reflejada en la variable **Riesgo.por.el.que.entró**, que contiene información sobre el riesgo por el cual cada bien patrimonial ha sido incluido en la Lista Roja. Puesto que ambas referencian a conceptos muy relacionados entre sí, y para solventar esa ausencia de datos, se ha incluido el contenido del **Riesgo.por.el.que.entró** dentro de la variable **Estado.de.conservación** mediante la concatenación de sus contenidos. Posteriormente trataremos la variable **Riesgo.por.el.que.entró** con un objetivo distinto.

Para extraer la información que nos ayude a determinar el grado de deterioro, se ha analizado con un contador de palabras frecuentes, que muestra una lista ordenada de mayor a menor con las palabras según cuantas veces están repetidas. Entre las primeras encontramos términos como “ruina”, “ruinoso”, “abandono”, “abandonado” o “deterioro” entre otras.

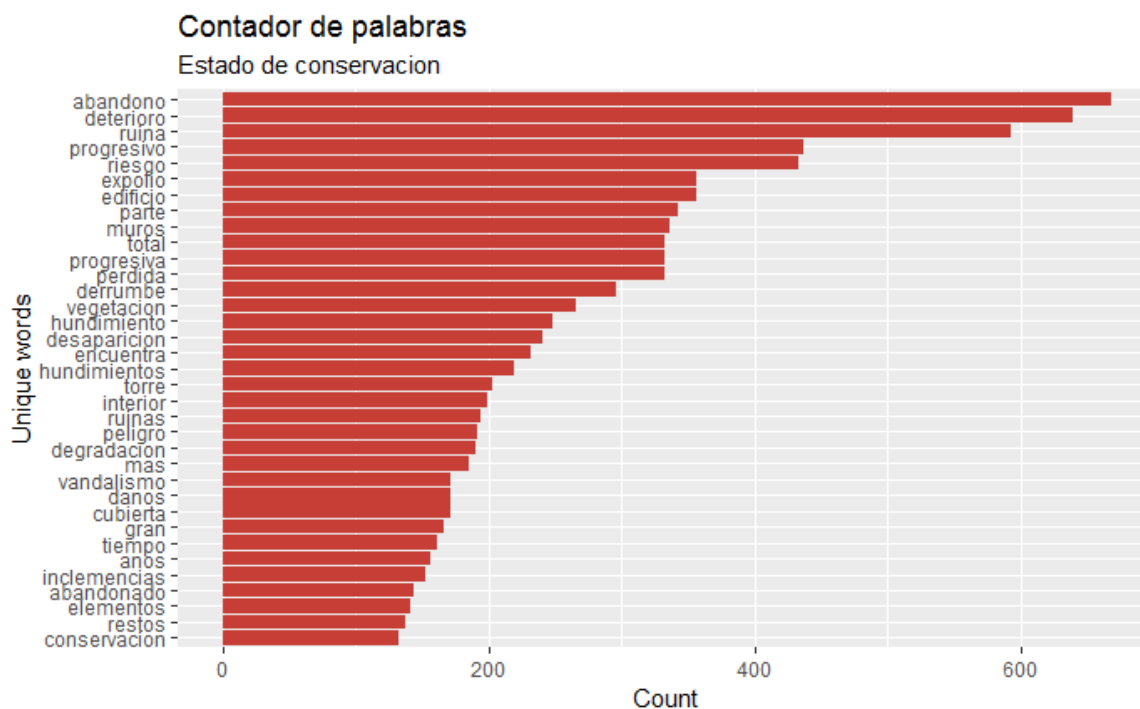


Figura 10.31. Términos más frecuentes hallados en *Estado.de.conservación*

## dfDatos

### Estado de conservación

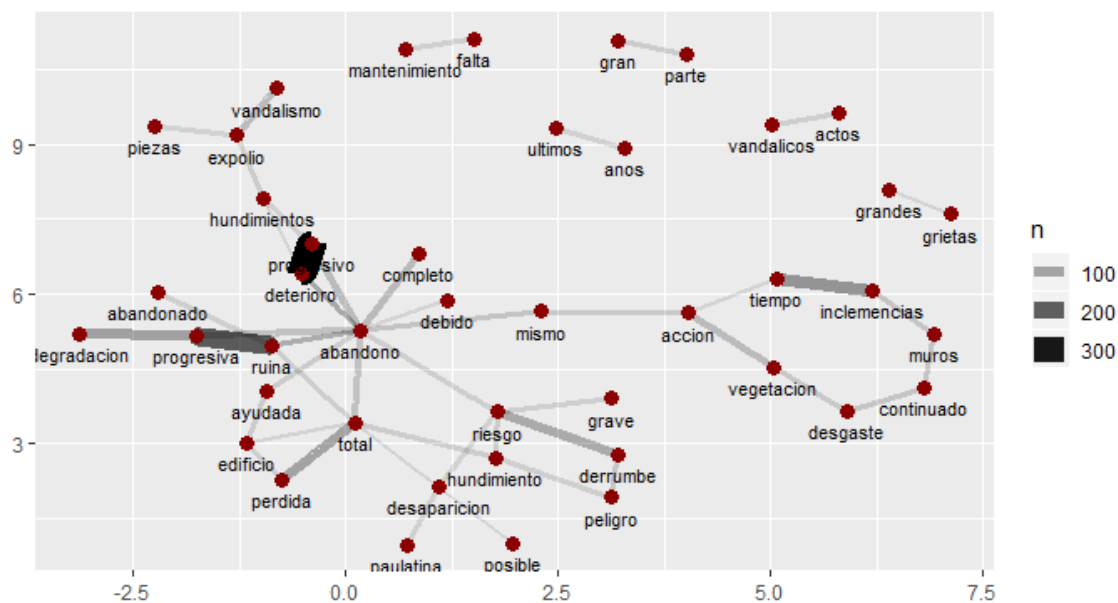


Figura 10.32. Emparejamientos más frecuentes de términos hallados en *Estado.de.conservación*

Basándome en el funcionamiento de los “análisis de sentimientos” en textos, se ha implementado un diccionario de términos asociados a un valor. De esta manera, se ha cuantificado la magnitud de muchos términos que aparecen o podrían aparecer. El valor asociado a estos términos es una aproximación según sus definiciones en la RAE.

Este diccionario puede verse en las *tablas 10.63 y 10.0.64*. Para simplificar la identificación de términos se ha recurrido a técnicas de lematización, almacenando las raíces de las palabras para poder trabajar indistintamente con cualquiera de sus terminaciones.

Cuando cada uno de los términos del diccionario era hallado en la variable **Estado.de.conservación** de una ficha, sumaba su puntuación a un contador que da como valor final su valoración negativa del estado. Es decir, a mayor puntuación, peor se encuentra. Sin embargo, esta idea tuvo que ser desechada puesto que las descripciones del estado de conservación varían mucho entre sí, de tal forma que una ficha descrita con gran detalle podía llegar a acumular más puntuación que otra más escueta cuyo estado pudiese ser realmente grave. Por ejemplo: “*el templo se encuentra en ruinas*” sumaría 90 puntos. Por su parte, “*el estado del templo es alarmante, se encuentra en ruinas y la piedra está erosionada*” aportarían respectivamente 70 + 90 + 30 puntos, acumulando un total de 190. Conociendo que dicho templo está en ruinas no nos resulta tan determinante que la piedra esté erosionada o que su estado sea lamentable, porque es algo intrínseco a su estado de ruina.

Categoría	Subcategoría	Subcategoría 2	Subcategoría 3	Subcategoría 4	Puntuación
Devastado					600 puntos
	Alarmante				70 puntos
	Derruido				180 puntos
		Bombardeado			170 puntos
			Ruina / ruinoso		80 puntos
				Deplorable	70 puntos
				Terrible / Horrible	60 puntos
				Lamentable / Pésimo	50 puntos
				Mal / malo	20 puntos
			Grieta		30 puntos
			Humedad		30 puntos
	Dañado				50 puntos
			Deteriorado		40 puntos
				Degradado	40 puntos
				Erosionado	30 puntos
	Desaparecido				70 puntos
			Perdido		60 puntos
				Alterado	40 puntos
	Hundido				80 puntos
			Desplomado		80 puntos
				Desprendido	60 puntos
	Quemado / Incendiado				80 puntos

Tabla 10.63. Lista de valoraciones ordenadas por categorías (parte 1)

Categoría	Puntuación
Abandonado	60 puntos
Expoliado / Saqueado	40 puntos
Vandalismo	60 puntos
Vegetación / Plantas	20 puntos

Tabla 10.64. Lista de valoraciones ordenadas por categorías (parte 2)

Para solucionar esto se ha trabajado con categorías en las que agrupamos los términos tal y como puede apreciarse en la *tabla 10.63*. De esta forma conseguimos que si se detecta un término que comprenda o implique otros términos menores, estos sean ignorados. En el último nivel de la jerarquía, las palabras aparecen listadas por orden de gravedad. Una vez aparezca uno, todos los que quedan por debajo de este serán ignorados si aparecen en el atributo. De esta manera evitamos que si alguien dice que algo está “horrible” valoremos también que está “mal”, algo que no tendría sentido. Los términos inferiores nunca deben superar, entre la suma de todos ellos, el valor del término que los engloba.

Sin embargo, basándonos en la información observada en la figura 10.32, se ha valorado también la aparición de adjetivos cuantificadores que acompañen a estos términos. No es lo mismo tener “deteriorado” que “muy deteriorado” o “progresivo deterioro”. Para ello hemos trabajado con bigramas, buscando la raíz de los términos de las tablas de valoraciones acompañados por cuantificadores. Los cuantificadores se han recogido en una nueva biblioteca, la cual puede visualizarse en la *tabla 10.65*. Y a cada uno de ellos se le ha proporcionado un valor menor de diez, que será el que sumen al valor del término al que acompañen.

Se ha valorado también el orden en la escritura. Así pues, “**ruina completa**” o “**completamente en ruinas**” representan lo mismo. Cabe añadir que previamente se ha realizado una eliminación de palabras vacías, tales como artículos o preposiciones, pero conservando ciertos adverbios de cantidad.

El inconveniente que puede suponer el uso de este método es que se empleen términos de gran magnitud para definir elementos que no representen la totalidad de un bien patrimonial, por ejemplo, “pinturas devastadas”. Este método de extracción de información ofrece otras posibilidades para ser tratado en mayor profundidad y con mayor precisión que podrían constituir una línea de investigación futura: el uso de estructuras más complejas, como son los trigramas.

Término	Puntuación
Devastador/devastadora	7 puntos
Absoluto/absolutamente	7 puntos
Total/totalmente	7 puntos
Terrible/terriblemente	6 puntos
Horrible/horriblemente	6 puntos
Completo/Completamente	5 puntos
Progresivo/Progresivamente	5 puntos
Gran	4 puntos
Muy/mucho	3 puntos

Tabla 10.65. Lista de cuantificadores

Finalmente, y con el objetivo de poder hacer análisis estadísticos más variados, se han obtenido diferentes atributos como resultado del tratamiento:

Un atributo de puntuación global, que es la que hemos obtenido anteriormente, llamada **Valoración.estado**. Si la valoración del estado de una ficha quedase con valor cero después de este proceso, se le dará el valor 10 por defecto.

Un atributo de puntuación asociada solo al deterioro estructural. Es decir, a aquellas valoraciones que no son ni el abandono, el expolio, la vegetación o el vandalismo. Esta variable se llama **Estado.deterioro**. Si la valoración del estado de una ficha quedase con valor cero después de este proceso, se le dará el valor 5 por defecto.

Cuatro atributos binarios para los restantes factores de conservación, “Abandono”, “Expolio”, “Vandalismo” y “Vegetación”. Estos almacenarán el valor “S” en caso de darse estos factores y permanecerán vacíos de no ser así, ya que no podemos determinar con certeza si realmente no sufren estos factores, únicamente sabemos que no han sido mencionados.

Se debe tener en cuenta que, aunque un bien patrimonial haya salido de la Lista Roja, conocer el estado de conservación que tenía en el momento de su inclusión puede resultar relevante, y esta información permanece almacenada después de su salida.

Llegados a este punto se han detectado algunos posibles errores en el tratamiento de este atributo. Supongamos que un atributo **Estado.de.conservación** contiene la cadena de caracteres “*Totalmente devastado y en progresivo abandono*”. Esto generaría los siguientes emparejamientos: “*totalmente devastado*”, “*devastado progresivo*” y “*progresivo abandono*”. Nosotros entendemos que la forma correcta de leer esto sería “*totalmente devastado*”, que se corresponde con 7 + 500 puntos; y “*progresivo abandono*”, que se corresponde con 60 + 5 puntos. En total, el estado de este bien alcanzaría 572 puntos.

Los bigramas se pueden ordenar de dos formas: que los más frecuentes aparezcan los primeros, o por orden alfabético de sus términos. Si el orden es alfabético, el sistema detectaría “*devastado progresivo*” antes que “*totalmente devastado*” y “*progresivo abandono*”, por lo que el resultado de la ejecución será “*Totalmente \_ y en \_ abandono*”. “*Totalmente*” sería ignorado y “*abandono*” contaría como si no estuviera acompañado de ningún cuantificador. Esto supondría una suma total de 565 puntos.

Ahora bien, supongamos que no ordenamos los bigramas de forma alfabética y lo hacemos según su frecuencia. Si tuviésemos más fichas que contuviesen “*devastado progresivo*” tendríamos el mismo problema. Para solucionar esto necesitamos un delimitador, es decir, un carácter no alfabético como un punto o una coma. Para ello, vamos a reemplazar previamente cualquier aparición de la conjunción “y” por puntos. La “y” desaparece una vez tratadas las palabras vacías unificando las frases que estuviera separando.

### 10.3.6.1. BATERÍA DE PRUEBAS

Las cadenas de entrada proporcionadas para los diferentes casos de prueba no se corresponden con ejemplos concretos del fichero de datos. Son casos creados para poner a prueba el sistema, en ocasiones más complejos de los que realmente aparecen en dicho fichero.

Identificador de prueba	ESTADO_P1
Entrada	Se encuentra deteriorado.
Tipo	Lista Roja
Resultado esperado	Valoración.estado: 40 Estado.deterioro: 40
Resultado obtenido	Valoración.estado: 40 Estado.deterioro: 40

Tabla 10.66. Caso de prueba ESTADO\_P1



<b>Identificador de prueba</b>	<b>ESTADO_P2</b>
<b>Entrada</b>	Alarmante y en ruinas.
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Valoración.estado: 150</b> <b>Estado.deterioro: 150</b>
<b>Resultado obtenido</b>	<b>Valoración.estado: 150</b> <b>Estado.deterioro: 150</b>

Tabla 10.67. Caso de prueba ESTADO\_P2

<b>Identificador de prueba</b>	<b>ESTADO_P3</b>
<b>Entrada</b>	Estado pésimo, completamente en ruinas.
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Valoración.estado: 85</b> <b>Estado.deterioro: 85</b>
<b>Resultado obtenido</b>	<b>Valoración.estado: 85</b> <b>Estado.deterioro: 85</b>

Tabla 10.68. Caso de prueba ESTADO\_P3

<b>Identificador de prueba</b>	<b>ESTADO_P4</b>
<b>Entrada</b>	Devastado, ruinoso y deplorable.
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Valoración.estado: 600</b> <b>Estado.deterioro: 600</b>
<b>Resultado obtenido</b>	<b>Valoración.estado: 600</b> <b>Estado.deterioro: 600</b>

Tabla 10.69. Caso de prueba ESTADO\_P4

<b>Identificador de prueba</b>	<b>ESTADO_P5</b>
<b>Entrada</b>	Quemado y expoliado
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Valoración.estado: 120</b> <b>Estado.deterioro: 80</b> <b>Estado.expolio: Sí</b>
<b>Resultado obtenido</b>	<b>Valoración.estado: 120</b> <b>Estado.deterioro: 80</b> <b>Estado.expolio: Sí</b>

Tabla 10.70. Caso de prueba ESTADO\_P5

<b>Identificador de prueba</b>	<b>ESTADO_P6</b>
<b>Entrada</b>	Se encuentra muy deteriorado.
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Valoración.estado: 43</b> <b>Estado.deterioro: 43</b>
<b>Resultado obtenido</b>	<b>Valoración.estado: 43</b> <b>Estado.deterioro: 43</b>

Tabla 10.71. Caso de prueba ESTADO\_P6

<b>Identificador de prueba</b>	<b>ESTADO_P7</b>
<b>Entrada</b>	En progresivo deterioro, muros erosionados
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Valoración.estado: 45</b> <b>Estado.deterioro: 45</b>
<b>Resultado obtenido</b>	<b>Valoración.estado: 45</b> <b>Estado.deterioro: 45</b>

Tabla 10.72. Caso de prueba ESTADO\_P7

<b>Identificador de prueba</b>	<b>ESTADO_P8</b>
<b>Entrada</b>	Abandonado, expoliado, vandalismo y vegetación
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Valoración.estado: 180</b> <b>Estado.deterioro: 5</b> <b>Estado.abandono: Sí</b> <b>Estado.expolio: Sí</b> <b>Estado.vandalismo: Sí</b> <b>Estado.vegetación: Sí</b>
<b>Resultado obtenido</b>	<b>Valoración.estado: 180</b> <b>Estado.deterioro: 5</b> <b>Estado.abandono: Sí</b> <b>Estado.expolio: Sí</b> <b>Estado.vandalismo: Sí</b> <b>Estado.vegetación: Sí</b>

Tabla 10.73. Caso de prueba ESTADO\_P8

<b>Identificador de prueba</b>	<b>ESTADO_P9</b>
<b>Entrada</b>	Devastado, abandonado, expoliado, vandalismo y vegetación
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Valoración.estado: 780</b> <b>Estado.deterioro: 600</b> <b>Estado.abandono: Sí</b> <b>Estado.expolio: Sí</b> <b>Estado.vandalismo: Sí</b> <b>Estado.vegetación: Sí</b>
<b>Resultado obtenido</b>	<b>Valoración.estado: 780</b> <b>Estado.deterioro: 600</b> <b>Estado.abandono: Sí</b> <b>Estado.expolio: Sí</b> <b>Estado.vandalismo: Sí</b> <b>Estado.vegetación: Sí</b>

Tabla 10.74. Caso de prueba ESTADO\_P9

<b>Identificador de prueba</b>	<b>ESTADO_P10</b>
<b>Entrada</b>	Totalmente devastado, en progresivo abandono. Muestras de gran vandalismo. Con muchas humedades. Mucha vegetación expandida por todo el edificio y ha sido totalmente expoliado
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Valoración.estado: 806</b> <b>Estado.deterioro: 607</b> <b>Estado.abandono: Sí</b> <b>Estado.expolio: Sí</b> <b>Estado.vandalismo: Sí</b> <b>Estado.vegetación: Sí</b>
<b>Resultado obtenido</b>	<b>Valoración.estado: 806</b> <b>Estado.deterioro: 607</b> <b>Estado.abandono: Sí</b> <b>Estado.expolio: Sí</b> <b>Estado.vandalismo: Sí</b> <b>Estado.vegetación: Sí</b>

Tabla 10.75. Caso de prueba ESTADO\_P10

<b>Identificador de prueba</b>	<b>ESTADO_P11</b>
<b>Entrada</b>	Totalmente devastado y en progresivo abandono. Muestras de gran vandalismo. Con muchas humedades. Mucha vegetación expandida por todo el edificio y ha sido totalmente expoliado.
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Valoración.estado:</b> 806 <b>Estado.deterioro:</b> 607 <b>Estado.abandono:</b> Sí <b>Estado.expolio:</b> Sí <b>Estado.vandalismo:</b> Sí <b>Estado.vegetación:</b> Sí
<b>Resultado obtenido</b>	<b>Valoración.estado:</b> 739 <b>Estado.deterioro:</b> 507 <b>Estado.abandono:</b> Sí <b>Estado.expolio:</b> Sí <b>Estado.vandalismo:</b> Sí <b>Estado.vegetación:</b> Sí

Tabla 10.76. Caso de prueba ESTADO\_P11

<b>Identificador de prueba</b>	<b>ESTADO_P12</b>
<b>Entrada</b>	Alarmante y en ruinas.
<b>Tipo</b>	Lista Negra
<b>Resultado esperado</b>	<b>Valoración.estado:</b> 150 <b>Estado.deterioro:</b> 150
<b>Resultado obtenido</b>	<b>Valoración.estado:</b> 150 <b>Estado.deterioro:</b> 150

Tabla 10.77. Caso de prueba ESTADO\_P12

<b>Identificador de prueba</b>	<b>ESTADO_P13</b>
<b>Entrada</b>	Derruido, ruinoso y con grietas.
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Valoración.estado: 180</b> <b>Estado.deterioro: 180</b>
<b>Resultado obtenido</b>	<b>Valoración.estado: 180</b> <b>Estado.deterioro: 180</b>

Tabla 10.78. Caso de prueba ESTADO\_P13

<b>Identificador de prueba</b>	<b>ESTADO_P14</b>
<b>Entrada</b>	Algo deteriorado
<b>Tipo</b>	Lista Verde
<b>Resultado esperado</b>	<b>Valoración.estado: 40</b> <b>Estado.deterioro: 40</b>
<b>Resultado obtenido</b>	<b>Valoración.estado: 40</b> <b>Estado.deterioro: 40</b>

Tabla 10.79. Caso de prueba ESTADO\_P14

<b>Identificador de prueba</b>	<b>ESTADO_P15</b>
<b>Entrada</b>	Hola mundo
<b>Tipo</b>	Lista Roja
<b>Resultado esperado</b>	<b>Valoración.estado: 10</b> <b>Estado.deterioro: 5</b>
<b>Resultado obtenido</b>	<b>Valoración.estado: 10</b> <b>Estado.deterioro: 5</b>

Tabla 10.80. Caso de prueba ESTADO\_P15

### 10.3.7. ATRIBUTO “RIESGO”

Riesgo de hundimiento total. Expolio y amenaza de derribo.
Progresivo deterioro por lamentable estado de abandono.
En buen estado aunque necesitaba obras de conservación y mantenimiento.
Las crecidas del rio causaron graves danos al desprender piedras de los tajamares. Riesgo de desplome.
Derrumbe de los muros de la nave; perdida total de las pinturas, que van desapareciendo de forma imparable por estar a la intemperie.

Figura 10.33. Muestras originales del atributo *Riesgo.por.el.que.entró*

Riesgo de <b>hundimiento</b> total. <b>Expolio</b> y amenaza de <b>derribo</b> .
Progresivo <b>deterioro</b> por lamentable estado de <b>abandono</b> .
En buen estado aunque necesitaba obras de conservación y mantenimiento.
Las crecidas del rio causaron graves <b>daños</b> al desprender piedras de los tajamares. Riesgo de <b>desplome</b> .
<b>Derrumbe</b> de los muros de la nave; <b>perdida</b> total de las <b>pinturas</b> , que van desapareciendo de forma imparable por estar a la intemperie.

Figura 10.34. Representación de los términos a detectar por el sistema

Riesgo de <b>hundimiento</b> total. <b>Expolio</b> y amenaza de <b>derribo</b> .
Progresivo <b>deterioro</b> por lamentable estado de <b>abandono</b> .
En buen estado aunque necesitaba obras de conservación y mantenimiento.
Las crecidas del rio causaron graves <b>daños</b> al desprender piedras de los tajamares. Riesgo de <b>desplome</b> .
<b>Derrumbe</b> de los muros de la nave; <b>perdida</b> total de las <b>pinturas</b> , que van desapareciendo de forma imparable por estar a la intemperie.

Figura 10.35. Resultado de la ejecución

El atributo **Riesgo.por.el.que.entró** refleja el motivo de la inclusión en la Lista Roja. Su tratamiento es similar al de **Propiedad** o **Grado.de.protección**, se pretende encontrar términos generalizados en los que puedan agruparse los múltiples riesgos. Es importante aclarar que no se va a valorar el riesgo “Desaparición” aplicado al conjunto total del bien patrimonial, ya que debe asumirse que un bien patrimonial ya está en riesgo de desaparición por el mero hecho de haber sido incluido en la Lista Roja.

Se han buscado nuevamente los términos más frecuentes dentro del atributo, tal y como se puede ver en las *figuras 10.36 y 10.37*, y se han almacenado sus raíces léxicas en una serie de listas según su categoría. Los diversos riesgos se han agrupado en cinco categorías, las cuales son “Demolición”, “Expolio”, “Daños estructurales”, “Ruina”, “Deterioro” y “Abandono”. Entendemos como “daños estructurales” aquellos que pongan en riesgo la estructura completa del edificio. “Deterioro” representará aquellos daños no estructurales que pueda sufrir el bien patrimonial. Aunque podríamos considerar una “ruina” como una consecuencia de los daños estructurales, vamos a recogerla por separado. Los términos clave identificados pueden visualizarse en la *Tabla 10.81*.

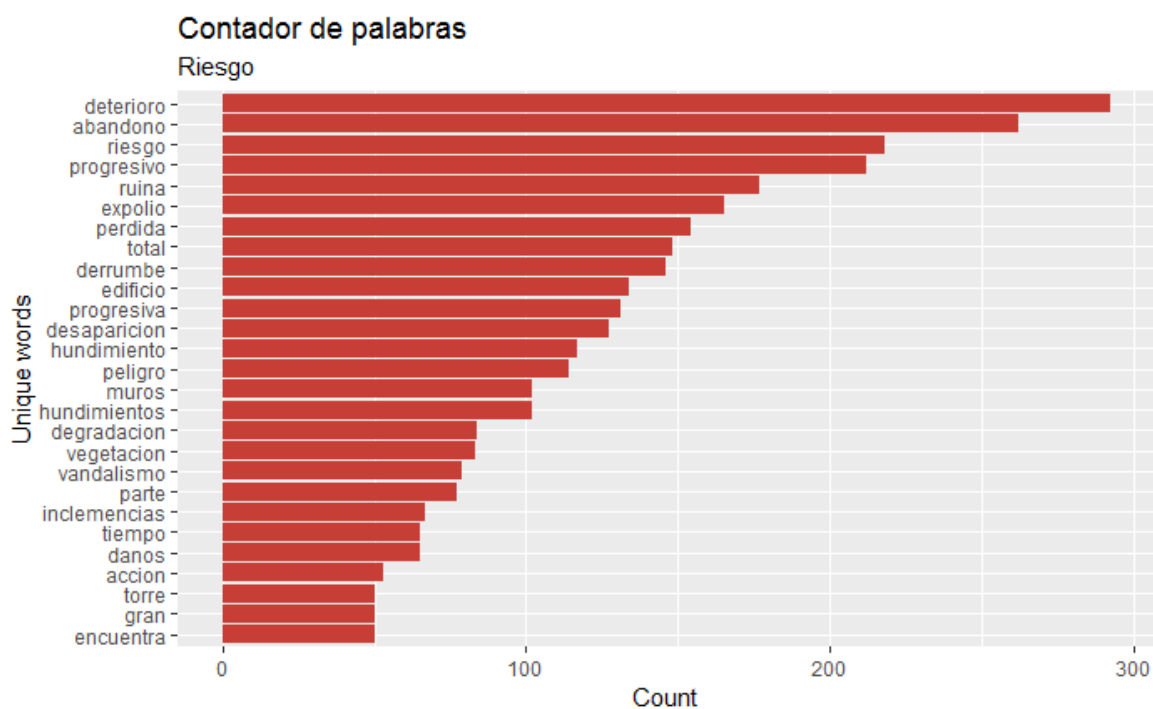


Figura 10.36. Emparejamientos más frecuentes de términos hallados en *Riesgo.por.el.que.entró*



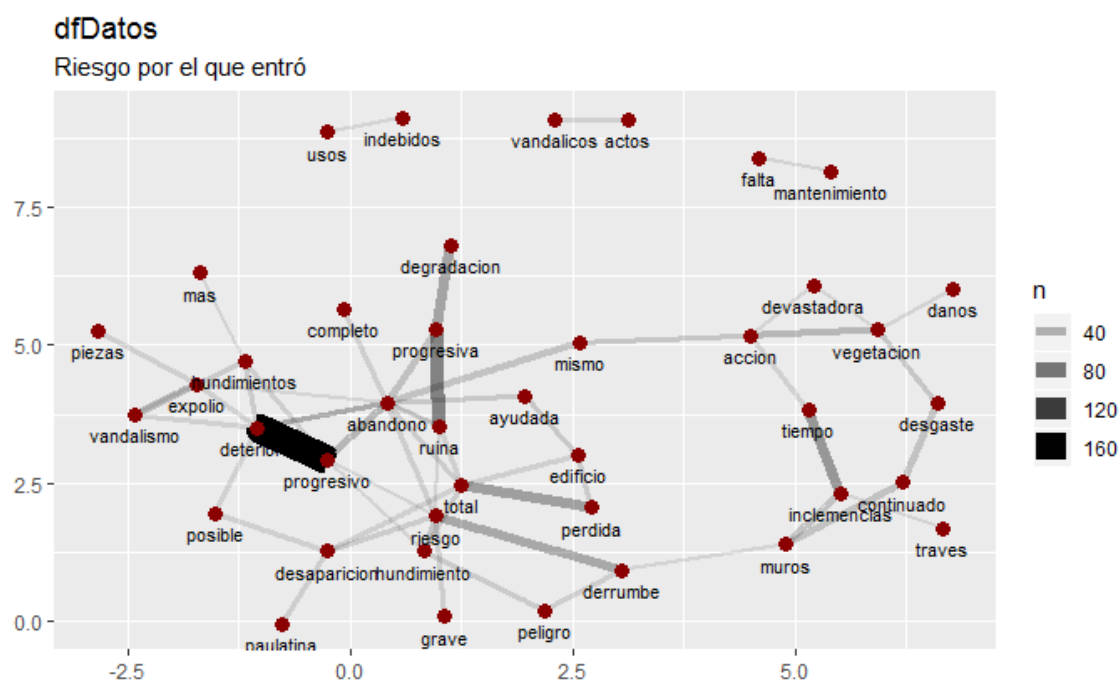


Figura 10.37. Emparejamientos más frecuentes de términos hallados en *Riesgo.por.el.que.entró*

Se asume, como es natural, que una ficha pueda verse sometida a varios riesgos. Por ello, el sistema busca coincidencias de las raíces de los términos dentro de cada texto, y almacena la categoría en que se engloban estos riesgos en un nuevo atributo llamado **Riesgo**. Una vez se obtiene la lista de riesgos de cada ficha, se separan en nuevos atributos binarios, uno por cada una de las categorías de riesgos valorados. En estos atributos se almacenará únicamente un “S” si la ficha cumple con el riesgo correspondiente a cada nuevo atributo.

Sin embargo, existen algunos términos que deben ser analizados en su contexto para determinar el riesgo al que hacen referencia. “Desprendimiento” es, por ejemplo, uno de ellos, ya que puede indicarnos dos posibilidades: el desprendimiento de un elemento decorativo, o el desprendimiento de partes esenciales de la construcción que pongan en riesgo su estabilidad. Para conocer mejor la vinculación entre las palabras se ha recurrido nuevamente al análisis de bigramas más frecuentes. Esta vez, en torno a cada uno de los términos que queremos analizar.

<b>Término</b>	<b>Categoría(s)</b>
Ruina	Ruina
Expolio	Expolio
Saqueo	Expolio
Abandono	Abandono
Deterioro	Deterioro
Vandalismo	Deterioro
Degradación	Deterioro
Erosión	Deterioro
Inclemencia meteorológica	Deterioro
Desplome	Riesgo estructural
Derrumbe	Riesgo estructural
Hundimiento	Riesgo estructural
Ceder	Riesgo estructural
Demolición	Demolición
Derribo	Demolición
Pérdida	Deterioro / Riesgo estructural
Desaparición	Deterioro / Riesgo estructural
Desprendimiento	Deterioro / Riesgo estructural
Caída	Deterioro / Riesgo estructural
Venirse abajo	Deterioro / Riesgo estructural
Derruir	Riesgo estructural / Demolición
Destrucción	Deterioro / Riesgo estructural / Demolición

Tabla 10.81. Términos clave del atributo *Riesgo.por.el.que.entró*

## Caída

Riesgo por el que entró

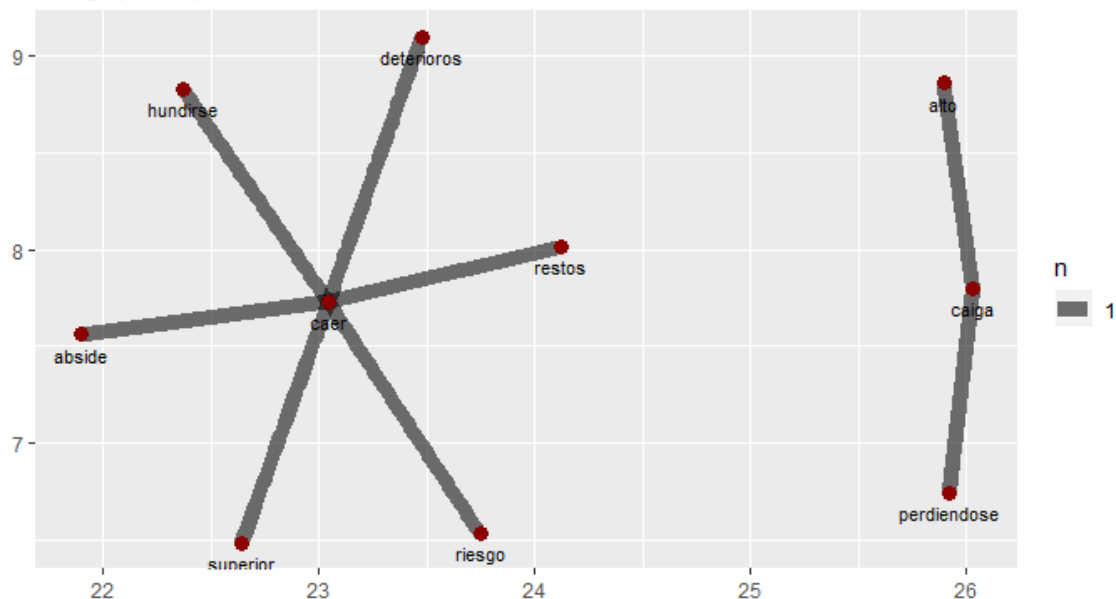


Figura 10.38. Emparejamientos más frecuentes de *caída* en *Riesgo.por.el.que.entró*

## Pérdida

Riesgo por el que entró

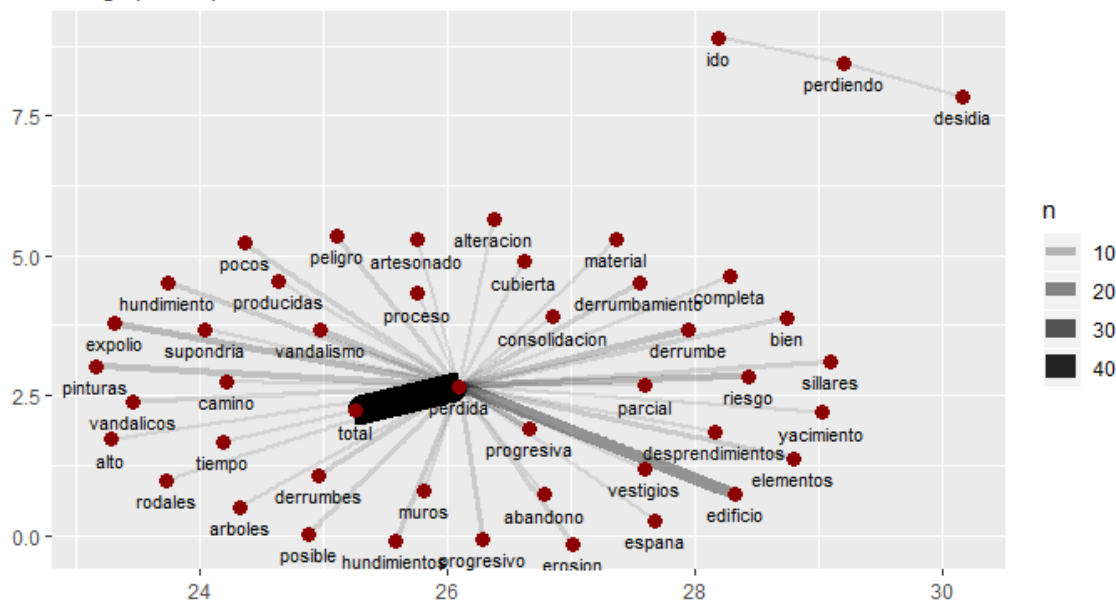


Figura 10.39. Emparejamientos más frecuentes de *pérdida* en *Riesgo.por.el.que.entró*

## Desaparición

Riesgo por el que entró

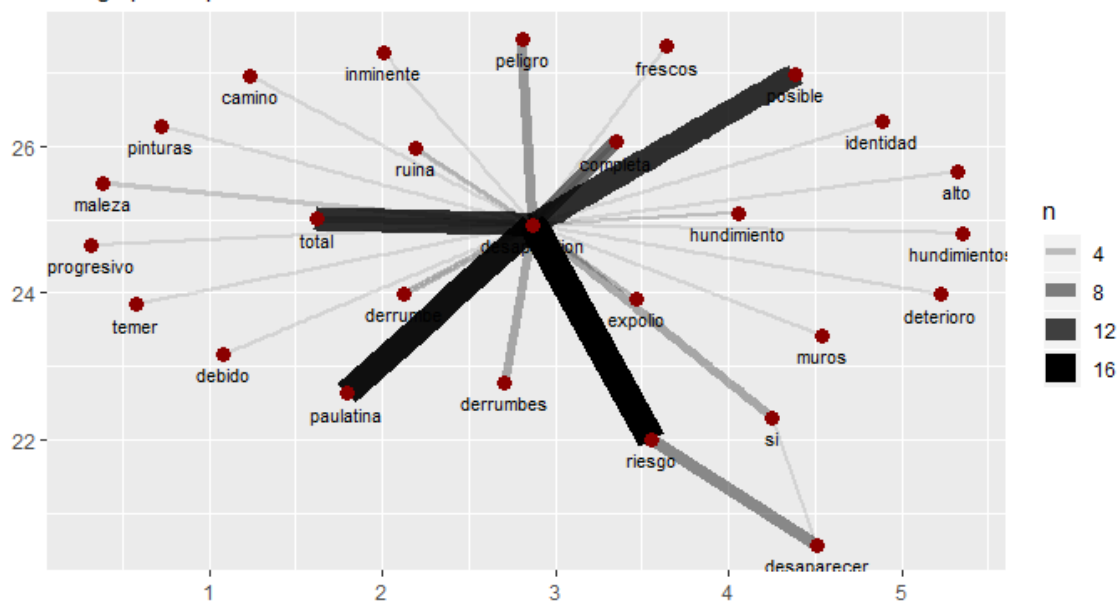


Figura 10.40. Emparejamientos más frecuentes de *desaparición* en *Riesgo por el que entró*

## Pérdida

Riesgo por el que entró

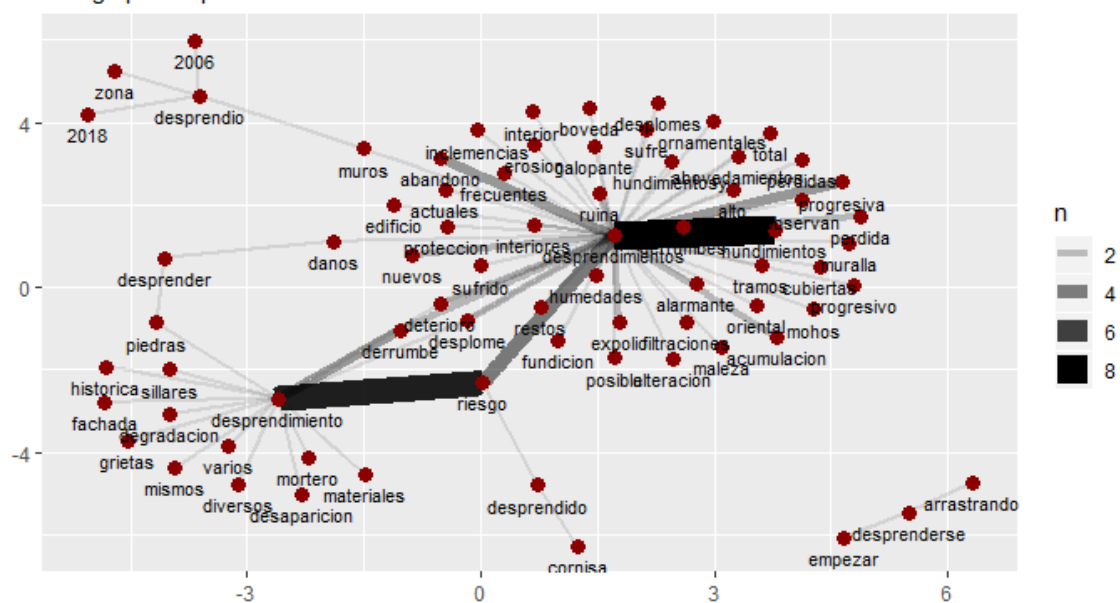


Figura 10.41. Emparejamientos más frecuentes de *desprendimiento* en *Riesgo por el que entró*

En base a los resultados obtenidos podemos hacer una diferenciación entre los elementos estructurales y los no estructurales. Se han listado los términos identificados, que quedan recogidos en las *tablas 10.82 y 10.83* según su clasificación. Podemos repetir los análisis de emparejamientos frecuentes reajustando el umbral a partir del cual se muestran los emparejamientos más frecuentes, e irán surgiendo nuevos términos.

Elementos estructurales
Muros
Paredes
Fachada
Pilares
Sillares
Artesonado
Bóveda
Cubierta
Tajamar

Tabla 10.82. Elementos estructurales

Elementos no estructurales
Pintura
Fresco
Relieve
Bajorrelieve
Cornisa
Sepulcro / sepultura
Elemento decorativo
Ornamento

Tabla 10.83. Elementos no estructurales

El sistema buscará construcciones en las cuales los términos referentes al riesgo vayan acompañados de los diversos elementos estructurales o no estructurales y los clasificará según sean detectados. Estas construcciones valorarán que el elemento estructural esté colocado antes o después del riesgo en la frase e incluso que pueda haber alguna palabra entre medias de ambos (ejemplo: “*desaparición total frescos*”). Por consiguiente, un riesgo que afecta a un elemento no estructural será catalogado como “*deterioro*”. Por su parte, si se trata de un elemento estructural será catalogado como “*daño estructural*”.

A continuación, se ha intentado tratar el término “derruir”, el cual, en función de la conjugación del verbo o del contexto de la frase, puede significar una demolición intencionada, o el efecto de ciertas inclemencias. No obstante, se ha comprobado que de entre todas las fichas de la Lista Roja, solo una contiene el término. Por ello nos resulta imposible recurrir al análisis de emparejamiento de términos. Sería necesario valorar todas las conjugaciones del verbo o estructuras semánticas en las que pudiese ir incluido para poder clasificarlas una a una. Se ha tratado, además, de acotar esta búsqueda incluyendo los términos “derruirla”, “derruirlo”, “derruyan” y “derruirán” como sinónimos de “*demolición*”, sin embargo, resulta imposible determinar si “derruir” o



### 10.3.7.1. BATERÍA DE PRUEBAS

Las cadenas de entrada proporcionadas para los diferentes casos de prueba no se corresponden con ejemplos concretos del fichero de datos. Son casos creados para poner a prueba el sistema, en ocasiones más complejos de los que realmente aparecen en dicho fichero.

<b>Identificador de prueba</b>	<b>RIESGO_P1</b>
<b>Entrada</b>	La torre amenaza con derrumbarse
<b>Resultado esperado</b>	<b>Riesgo.estructural: Sí</b>
<b>Resultado obtenido</b>	<b>Riesgo.estructural: Sí</b>

Tabla 10.84. Caso de prueba RIESGO\_P1

<b>Identificador de prueba</b>	<b>RIESGO_P2</b>
<b>Entrada</b>	Desprendimiento de las pinturas murales
<b>Resultado esperado</b>	<b>Riesgo.deterioro: Sí</b>
<b>Resultado obtenido</b>	<b>Riesgo.deterioro: Sí</b>

Tabla 10.85. Caso de prueba RIESGO\_P2

<b>Identificador de prueba</b>	<b>RIESGO_P3</b>
<b>Entrada</b>	El ayuntamiento planea derribar el edificio.
<b>Resultado esperado</b>	<b>Riesgo.demolición: Sí</b>
<b>Resultado obtenido</b>	<b>Riesgo.demolición: Sí</b>

Tabla 10.86. Caso de prueba RIESGO\_P3

<b>Identificador de prueba</b>	<b>RIESGO_P4</b>
<b>Entrada</b>	Riesgo de desprendimiento de los sillares. Pérdida total de los relieves, riesgo de expolio.
<b>Resultado esperado</b>	<b>Riesgo.expolio: Sí</b> <b>Riesgo.estructural: Sí</b> <b>Riesgo.deterioro: Sí</b>
<b>Resultado obtenido</b>	<b>Riesgo.expolio: Sí</b> <b>Riesgo.estructural: Sí</b> <b>Riesgo.deterioro: Sí</b>

Tabla 10.87. Caso de prueba RIESGO\_P4

<b>Identificador de prueba</b>	<b>RIESGO_P5</b>
<b>Entrada</b>	Destrucción de los pilares.
<b>Resultado esperado</b>	<b>Riesgo.estructural: Sí</b>
<b>Resultado obtenido</b>	<b>Riesgo.estructural: Sí</b>

Tabla 10.88. Caso de prueba RIESGO\_P5

<b>Identificador de prueba</b>	<b>RIESGO_P6</b>
<b>Entrada</b>	Destrucción de los sepulcros.
<b>Resultado esperado</b>	<b>Riesgo.deterioro: Sí</b>
<b>Resultado obtenido</b>	<b>Riesgo.deterioro: Sí</b>

Tabla 10.89. Caso de prueba RIESGO\_P6



<b>Identificador de prueba</b>	<b>RIESGO_P7</b>
<b>Entrada</b>	El edificio será destruido.
<b>Resultado esperado</b>	<b>Riesgo.demolición: Sí</b>
<b>Resultado obtenido</b>	El sistema no es capaz de detectar estos casos correctamente.

Tabla 10.90. Caso de prueba RIESGO\_P7

<b>Identificador de prueba</b>	<b>RIESGO_P8</b>
<b>Entrada</b>	Riesgo de ser derruido.
<b>Resultado esperado</b>	<b>Riesgo.demolición: Sí</b>
<b>Resultado obtenido</b>	<b>Riesgo.demolición: Sí</b>

Tabla 10.91. Caso de prueba RIESGO\_P8

<b>Identificador de prueba</b>	<b>RIESGO_P9</b>
<b>Entrada</b>	Van a derruirlo.
<b>Resultado esperado</b>	<b>Riesgo.demolición: Sí</b>
<b>Resultado obtenido</b>	<b>Riesgo.demolición: Sí</b>

Tabla 10.92. Caso de prueba RIESGO\_P9

<b>Identificador de prueba</b>	<b>RIESGO_P10</b>
<b>Entrada</b>	Otros.
<b>Resultado esperado</b>	No devuelve ningún riesgo.
<b>Resultado obtenido</b>	No devuelve ningún riesgo.

Tabla 10.93. Caso de prueba RIESGO\_P10

### 10.3.8. ATRIBUTO “DESCRIPCIÓN”

El atributo **Descripción** es un texto de longitud variable, aunque generalmente de extensión mucho mayor con respecto a los atributos anteriormente tratados, que carece de una estructura preestablecida. En muchos casos la información está extraída directamente de otras fuentes digitales y en otras redacta por los propios usuarios. Se han probado técnicas sencillas de *automatic summarization* (Resumen Automático de textos) con el objetivo de identificar información útil, pero sin mucho éxito debido a la enorme disparidad de la información contenida.

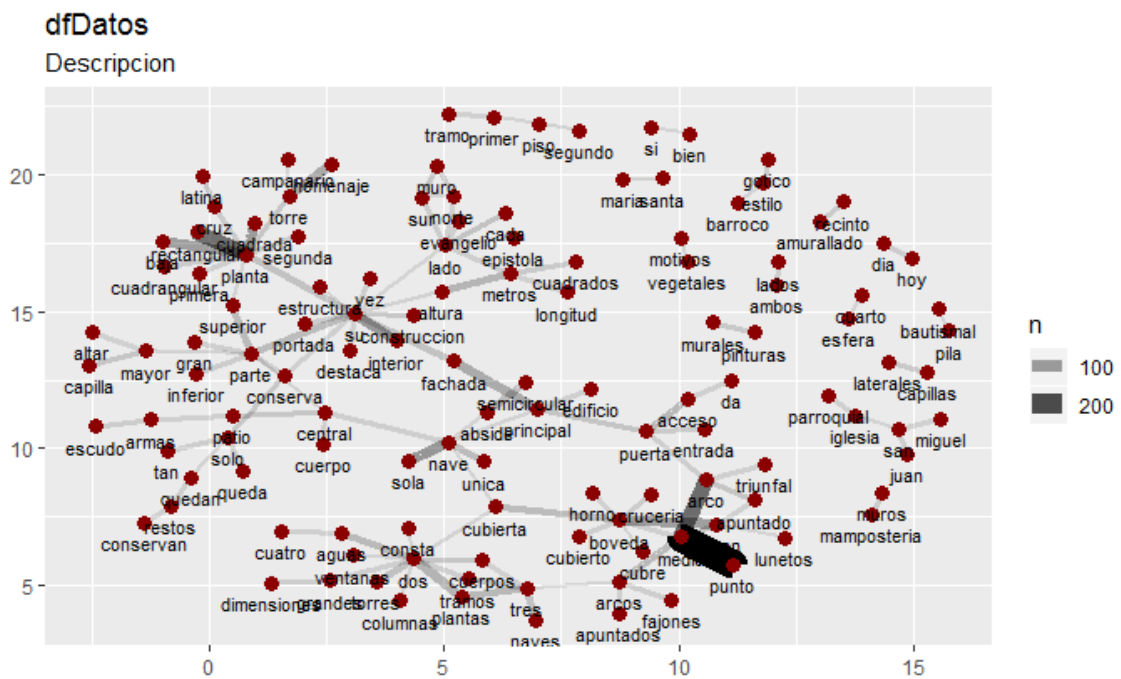


Figura 10.43. Emparejamientos más frecuentes hallados en *Descripción*

#### 10.3.8.1. MATERIALES

Tras realizar un análisis en busca de términos recurrentes llama la atención la frecuencia con la que aparecían términos como “*pedra*” o “*ladrillo*”. Considerando que los materiales de una construcción son claves en la perdurabilidad estructural del edificio, así como para el deterioro de otros elementos, como puede ser los decorativos, se ha hecho una recopilación de los materiales empleados con más frecuencia y se ha implementado un código que busque coincidencias de estos términos dentro de los textos.

Piedra caliza
Ladrillo
Granito
Cemento / Mortero
Hormigón
Madera
Adobe
Vidrio
Mármol
Metal
Hierro
Piedra
Cal

Tabla 10.94. Materiales más frecuentes empleados en la edificación

A su vez, se han creado nuevas variables, una por cada material de construcción. El código busca cada uno de los términos en las descripciones y conforme los va encontrando almacena un “S” en el atributo del material correspondiente de la ficha en cuestión.

### 10.3.8.2. ALTURAS

También se ha logrado extraer en una importante cantidad de fichas el número de pisos o alturas que tiene el bien patrimonial, con el objetivo de poder recopilar datos precisos. Esto ha podido hacerse mediante la búsqueda de términos como “alturas” o “pisos” acompañados de valores numéricos o la forma escrita de dichos valores numéricos. Se ha decidido no incluir los ordinales (primer, segundo, tercer, etc.) porque no nos da información del número total de alturas del edificio, sino a alguna planta determinada. Se desconoce la utilidad que pueda tener este dato para este estudio, pero puede tener cierto interés para la descripción básica del edificio.

### 10.3.8.3. SUPERFICIE

Se ha tratado de buscar información dentro de la variable **Descripción** que pudiese referenciar la superficie del bien arquitectónico en cuestión. Se desconoce la utilidad que pueda tener este dato para este estudio, pero se considera un dato de interés para documentar las características básicas del edificio. Se ha realizado una búsqueda de los términos “*superficie*”, “*área*”, “*metros cuadrados*” o “*m cuadrado*” acompañados de cifras numéricas, sin embargo, no se ha logrado encontrar nada más que unas pocas muestras, por lo que finalmente se han descartado. Considero que deberían buscarse construcciones más complejas tales como “*se extiende N kilómetros*” o “*N metros de altura*”, donde “*N*” sea un valor numérico. Para ello podría resultar de gran interés el uso de trigramas. Esto podrá conformar una nueva línea de investigación en el futuro.

### 10.3.8.4. BATERÍA DE PRUEBAS

Las cadenas de entrada proporcionadas para los diferentes casos de prueba no se corresponden con ejemplos concretos del fichero de datos. Son casos creados para poner a prueba el sistema, en ocasiones más complejos de los que realmente aparecen en dicho fichero.

Identificador de prueba	DESCRIPCIÓN_P1
Entrada	Se trata de una construcción de ladrillo, mortero y piedra, de 220 metros de largo y diecisiete arcos. Columnas de mármol.
Resultado esperado	Ladrillo: Sí Cemento: Sí Piedra: Sí Mármol: Sí
Resultado obtenido	Ladrillo: Sí Cemento: Sí Piedra: Sí Mármol: Sí

Tabla 10.95. Caso de prueba DESCRIPCIÓN\_P1

Identificador de prueba	DESCRIPCIÓN_P2
<b>Entrada</b>	Edificio exento de planta cuadrada y dos alturas con sillería en la fachada principal. Ordenada en torno a un patio central, estructurado en tres pisos, cada uno de los cuales está compuesto de columnas de madera.
<b>Resultado esperado</b>	<b>Madera:</b> Sí <b>Alturas:</b> 3
<b>Resultado obtenido</b>	<b>Madera:</b> Sí <b>Alturas:</b> 3

Tabla 10.96. Caso de prueba DESCRIPCIÓN\_P2

Identificador de prueba	DESCRIPCIÓN_P3
<b>Entrada</b>	El pantalán está formado de una estructura de hierro y hormigón que se adentra un kilómetro en el mar.
<b>Resultado esperado</b>	<b>Hormigón:</b> Sí <b>Hierro:</b> Sí
<b>Resultado obtenido</b>	<b>Hormigón:</b> Sí <b>Hierro:</b> Sí

Tabla 10.97. Caso de prueba DESCRIPCIÓN\_P3

Identificador de prueba	DESCRIPCIÓN_P4
<b>Entrada</b>	El suelo está conformado por grandes losas de granito. Los ventanales están cubiertos de elaboradas vidrieras.
<b>Resultado esperado</b>	<b>Granito:</b> Sí <b>Vidrio:</b> Sí
<b>Resultado obtenido</b>	<b>Granito:</b> Sí <b>Vidrio:</b> Sí

Tabla 10.98. Caso de prueba DESCRIPCIÓN\_P4



### 10.3.9.2. ACONTECIMIENTOS RELEVANTES

Finalizando el bloque de Búsqueda y Extracción de la Información se han creado una serie de variables relacionadas con sucesos o acontecimientos que pudiesen haber sucedido a los bienes patrimoniales y que pueden resultar de gran interés: desamortizaciones, guerra civil, francesada, incendios, terremotos e inundaciones.

Este código busca términos claves en el atributo **Historia** que referencien estos sucesos históricos, y que, de ser así, se almacene un “S” en el atributo correspondiente de la ficha. En la búsqueda se contemplan términos como “francesada”, “Napoleón”, “invasión francesa” o “invasión napoleónica” para el atributo **Francesada**. En el caso de **Incendios** se ha buscado “fuego”, “incendio” o “llamas”. Para **Terremotos** se han contemplado tanto “terremoto” como “seísmo”.

### 10.3.9.3. BATERÍA DE PRUEBAS

Las cadenas de entrada proporcionadas para los diferentes casos de prueba no se corresponden con ejemplos concretos del fichero de datos. Son casos creados para poner a prueba el sistema, en ocasiones más complejos de los que realmente aparecen en dicho fichero.

Identificador de prueba	HISTORIA_P1
Entrada	Edificado entre 1560 y 1585. Fue desamortizado durante la época de Mendizábal y sufrió un incendio en 1860.
Resultado esperado	<b>Desamortización:</b> Sí <b>Incendio:</b> Sí
Resultado obtenido	<b>Desamortización:</b> Sí <b>Incendio:</b> Sí

Tabla 10.99. Caso de prueba HISTORIA\_P1

<b>Identificador de prueba</b>	<b>HISTORIA_P2</b>
<b>Entrada</b>	Construido en el siglo XVII, fue usado como centro de operaciones durante la guerra civil española
<b>Resultado esperado</b>	<b>Desamortización:</b> Sí <b>Incendio:</b> Sí
<b>Resultado obtenido</b>	<b>Desamortización:</b> Sí <b>Incendio:</b> Sí

Tabla 10.100. Caso de prueba HISTORIA\_P2

<b>Identificador de prueba</b>	<b>HISTORIA_P3</b>
<b>Entrada</b>	Fue erigido en el siglo III d. C. y padeció los efectos de varios terremotos a lo largo de los últimos siglos.
<b>Resultado esperado</b>	<b>GuerraCivil:</b> Sí
<b>Resultado obtenido</b>	<b>GuerraCivil:</b> Sí

Tabla 10.101. Caso de prueba HISTORIA\_P3

<b>Identificador de prueba</b>	<b>HISTORIA_P4</b>
<b>Entrada</b>	El edificio fue saqueado durante la francesada por las tropas napoleónicas
<b>Resultado esperado</b>	<b>Francesada:</b> Sí
<b>Resultado obtenido</b>	<b>Francesada:</b> Sí

Tabla 10.102. Caso de prueba HISTORIA\_P4



## 10.4. SELECCIÓN DE OTROS ATRIBUTOS

Además de los atributos previamente tratados, se han incluido varios que se consideran de interés para el estudio y que no han necesitado intervención de ningún tipo. Estos son los expuestos a continuación.

### 10.4.1. ATRIBUTOS SOBRE LA LOCALIZACIÓN

Los atributos **Municipio**, **Provincia** y **Comunidad** son claves para la clasificación de las fichas y tendrán una gran utilidad en la realización de análisis geográficos. Nos permitirán también incluir datos externos segmentados por regiones geográficas.

Unos datos que no se han podido incluir y que habrían sido de enorme utilidad, son las coordenadas geográficas. Estas nos hubiesen permitido añadir datos más precisos, como los referentes al clima según su proximidad a las estaciones meteorológicas. Aunque existe un atributo llamado **Coordenadas**, no ha podido ser debidamente tratado: la ausencia de información en muchas de las fichas y la carencia de un formato unificado ha hecho imposible su correcta obtención. Esto constituiría una importante línea de ampliación en el futuro, y su aplicación podría ser de gran utilidad en la inclusión de datos externos, como veremos más adelante.

### 10.4.2. ATRIBUTO “FECHA DE INCLUSIÓN EN LA LISTA ROJA”

El atributo **Fecha.de.inclusión** contiene la fecha en la que el bien salió de la Lista Roja, ya fuese por pasar a la Lista Verde como por pasar a la Lista Negra. Esta fecha se encuentra inicialmente en formato *AAAA-DD-MM*. Al cargar el fichero de datos, RStudio las transforma automáticamente a formato *DD/MM/AAAA*. En caso de que una ficha no sea incluida tras la evaluación del comité científico, se almacena el valor “-000-00-00”.

### 10.4.3. ATRIBUTO “FECHA DE SALIDA DE LA LISTA ROJA”

El atributo **Fecha.de.salida** contiene la fecha en la que el bien entró en la Lista Roja. Esta fecha se encuentra inicialmente en formato *AAAA-DD-MM*. Al cargar el fichero de datos, RStudio las transforma automáticamente a formato *DD/MM/AAAA*. Las fichas que permanecen en la Lista Roja almacenan en este campo el valor “0000-00-00”.

## 10.5. REDUCCIÓN DE LA DIMENSIONALIDAD

El fichero de datos de la Lista Roja de Hispania Nostra consta en este momento de un total de 1195 filas (fichas) y 72 columnas (atributos). No todas las filas que contiene el *dataframe* son de interés, ya que están almacenadas todas las fichas sin distinguir las que han sido enviadas: muchas han sido rechazadas para agregarlas en la Lista Roja, otras son simples pruebas del formulario, carentes de información real. Para el correcto estudio ha sido necesario hacer una limpieza de estas variables.

Se han generado dos nuevos *dataframes*. En el primero de ellos se han suprimido todas aquellas fichas que no hayan sido incluidas públicamente en la Lista Roja, es decir, aquellas que han sido enviadas a través del formulario de usuario pero que no han sido aprobadas por el comité de evaluación científico de Hispania Nostra. Para ello se han guardado todas aquellas filas cuyo atributo **Date** sea distinto a "-0001-11-30", siendo este el valor por defecto con que se almacenan. Con ello hemos obtenido un *dataframe* que contenga únicamente las fichas válidas. Este *dataframe* se ha almacenado para posibles usos posteriores, así como para su cesión a Hispania Nostra.

El segundo *dataframe* que se ha generado es el que se ha usado para desarrollar el estudio, y es idéntico al anterior, salvo porque se han eliminado todas aquellas fichas que pertenezcan a Patrimonio Natural. Se ha exceptuado la subcategoría "*Jardines o huertas históricos*", ya que, como se ha explicado en un inicio, el tratamiento del Patrimonio Natural, puramente biológico, tiene grandes diferencias con el patrimonio arquitectónico, y debido a su magnitud no ha podido ser abarcado en este proyecto. Para ello, se han almacenado todas aquellas filas cuyo atributo **Subcategoría** sea distinto a "*Fauna y flora*" y "*Elementos o conjuntos de interés medioambiental o paisajístico*".

El *dataframe* final consta de 993 fichas, una cantidad muy limitada que no nos va a permitir aplicar ciertas técnicas de análisis como se había previsto inicialmente. Del total de fichas únicamente 171 están incluidas en la Lista Verde y 8 están en la Lista Negra.

## 11. INCLUSIÓN DE DATOS EXTERNOS

Una vez se ha terminado de tratar el contenido del fichero de datos proporcionada por Hispania Nostra podemos comenzar a incluir datos externos, procedentes de distintas fuentes, que puedan resultar de utilidad en el motivo del estudio. Como ya hemos visto, la adquisición de información conforma la Fase 2 del proyecto, ya que se trata de una fase iterativa a la que podemos volver de forma cíclica. La inclusión de nuevos datos puede realizarse a lo largo de cualquier proyecto de análisis sin que esto afecte al progreso alcanzado previamente.

### 11.1. DATOS CLIMATOLÓGICOS

Como ya se ha visto, al exponer la labor de entidades dedicadas a la conservación del patrimonio histórico, la monitorización, o por lo menos el control de las condiciones climáticas, es un factor de suma importancia en las labores de conservación. Esto es fácil de entender ya que es bien sabido que unas condiciones extremas pueden ser dañinas para ciertos materiales, así como los cambios bruscos de las mismas pueden afectar su perdurabilidad [30].

El acceso a los datos de la Agencia Estatal de Meteorología (AEMeT) está restringido al público desde el año 2015, en que se establecieron unos precios para tener acceso a ellos. Esto ha obligado a obtener la información de una fuente alternativa. Weather Spark es una web propiedad del grupo estadounidense Cedar Lake Ventures, Inc., que ofrece información sobre el clima en más de 150.000 localizaciones por todo el mundo. En dicha web es posible acceder a los valores promedio de casi cualquier ciudad a lo largo de los doce meses del año. Valores de la temperatura, nubes, probabilidad de precipitaciones, cantidad de agua recogida, horas de sol, humedad y velocidad y dirección del viento, entre otros.

Sin embargo, al no disponer de datos climatológicos concretos por cada municipio de España, así como al carecer de coordenadas definidas para todos los bienes patrimoniales que nos permitieran decretar la proximidad de estos a la estación meteorológica de la ciudad más cercana, se ha tenido que generalizar. Para ello se han tomado como referencia las capitales de provincia. Esto ha supuesto una considerable pérdida de precisión.

Los datos recopilados han sido la temperatura máxima promedio, la temperatura mínima promedio, la variación máxima de temperatura en un periodo de 24 horas, la humedad relativa máxima, la probabilidad de precipitación máxima, la cantidad máxima de agua recogida en lluvias y la velocidad máxima del viento.

## 11.2. VARIABLES SOCIOECONÓMICAS

El nivel socioeconómico del entorno podría llegar a ser un factor determinante. Tanto si se trata de una propiedad privada, como gestionada por la administración pública. Un indicador económico que probablemente aporte información es el Producto Interior Bruto per cápita provincial o Renta per cápita provincial, que mide la relación entre el nivel de renta de una región determinada por el número de habitantes [31]. Es posible acceder a estos datos a través del Instituto Nacional de Estadística. En el momento de la recopilación de datos, los más recientes que estaban disponibles eran los de 2017, por lo que son los que se han utilizado en todo momento.

En cuanto a los bienes pertenecientes a entidades públicas, como los de los gobiernos provinciales, hay otro factor que puede resultar de gran utilidad: la tasa de desempleo. Es interesante porque la mayoría de las intervenciones realizadas por las entidades públicas se suelen ver condicionadas por los presupuestos anuales de los que disponen (dato al que no se ha podido acceder). Sin embargo, es lógico pensar que la mayoría de los gobiernos provinciales van a priorizar el valor humano frente al valor patrimonial, y por consiguiente, una provincia con una tasa elevada de paro no va a tener su patrimonio como prioridad para invertir dinero. A estos datos se puede acceder a través del Instituto Nacional de Estadística [32]. En el momento de su recopilación, los más recientes que estaban disponibles eran los de 2019, pero se han tomado los del 2017 para que pertenezcan al mismo año que los datos de la Renta per cápita. Los datos se encuentran desglosados en los cuatro trimestres del año. Con ellos se ha calculado la media anual de cada provincia, siendo el dato final con el que se ha trabajado.

Esta información se ha almacenado en un archivo .csv y se ha incorporado al dataframe de la Lista Roja mediante una operación Natural Join por **Provincia**, que en R puede efectuarse mediante el método *merge(df1, df2, by = "column")*.

## 11.3. INDICADORES DE LA CALIDAD DEL AIRE

La contaminación supone también un factor de deterioro para el patrimonio histórico, y en general para cualquier material que se vea sometido a su efecto. La lluvia ácida se produce ante un descenso de los niveles de pH del agua, efecto generado por la acción de las partículas de ciertos contaminantes como son el dióxido de azufre (SO<sub>2</sub>) o el dióxido de nitrógeno (NO<sub>2</sub>). Estos, frecuentes en las zonas urbanas donde se producen altas emisiones, son los causantes de la formación de las “costras” negras que cubren fachadas, monumentos y esculturas, efecto de la corrosión de estos compuestos. Los principales causantes de esto son el hollín expulsado por los tubos de escape de los vehículos o la combustión de hidrocarburos [33].

Se han obtenido las estadísticas del año 2018 de los compuestos químicos SO<sub>2</sub>, NO<sub>2</sub> y NO<sub>x</sub> a través del INE [34]. Las mediciones de cada uno de los compuestos vienen almacenadas en un único fichero de datos, en el que se listan los diferentes centros de medición ordenados por comunidades autónomas y provincias. Por cada centro de medición están dispuestas las veinticuatro horas del día en columnas y los trescientos

sesenta y cuatro días del año en filas. En la *Tabla 11.1* puede apreciarse un esquema de la estructura del fichero de datos.

Comunidad	Provincia	Centro	Día del año	Hora 00	Hora 01	...
Andalucía	Córdoba	Centro 1	1-enero	[data]	[data]	...
			2-enero	[data]	[data]	...
			...	...	...	...
		Centro 2	1-enero	[data]	[data]	...
			...	...	...	...
			...	...	...	...
	Málaga	Centro N	1-enero	[data]	[data]	...
			...	...	...	...
		...	...	...	...	
		...	...	...	...	
Galicia	Lugo	Centro M	1-enero	[data]	[data]	...
...	...	...	...	...	...	...

Tabla 11.1. Estructura de los ficheros de mediciones de la calidad del aire

Para trabajar con estos datos se ha calculado la media diaria para cada fila de todas las mediciones obtenidas a cada hora del día. De igual forma, y para simplificar el problema, vamos a calcular la media anual de cada uno de los centros de medición. Podrían valorarse los máximos y los mínimos, pero eso lo vamos a dejar para otra ocasión. De manera idéntica a como se podrían haber planteado las variables climáticas, si dispusiéramos de coordenadas geográficas de los bienes, podríamos obtener los datos en función de su proximidad a los centros de medición. Pero como carecemos de esta información, se ha calculado nuevamente la media entre todos los centros de medición de cada provincia. Y así es como llegamos a un único dato por cada provincia, que podemos fusionar con el fichero de datos de la Lista Roja.

## 11.4. TASA SISMOGRÁFICA

Los niveles de peligrosidad sísmica son importantes y un evidente riesgo, ya que son causa de deterioro de las estructuras arquitectónicas. La información ha sido extraída de ZESIS, base de datos de información sismográfica del IGME [35]. Esta información viene dada en regiones sismográficas que no tienen correspondencia con las fronteras o límites administrativos municipales o provinciales. Por ello, se ha recurrido al uso de ArcGIS para transformar los datos. De esta forma, se ha calculado el centroide de cada una de las provincias españolas, abstrayendo cada una a un único punto, es decir, a una coordenada geográfica. Así se ha asociado a cada provincia el valor de peligrosidad sísmica correspondiente a la región sismográfica en la que se ubica cada centroide.

La decisión de obtener los datos en el ámbito provincial se debe a una cuestión puramente estadística. En España existen un total de 8131 municipios, de los cuales 6341 quedan comprendidos en regiones sismográficas [figura 11.1]. Es decir, aproximadamente el 78% de los municipios de la geografía española. Esto incluye también a las Islas Canarias. La posibilidad de fallo al realizar una operación Natural Join es muy elevada teniendo en cuenta la gran cantidad de municipios que pueden nombrarse de diferente forma según el idioma o dialecto regional.

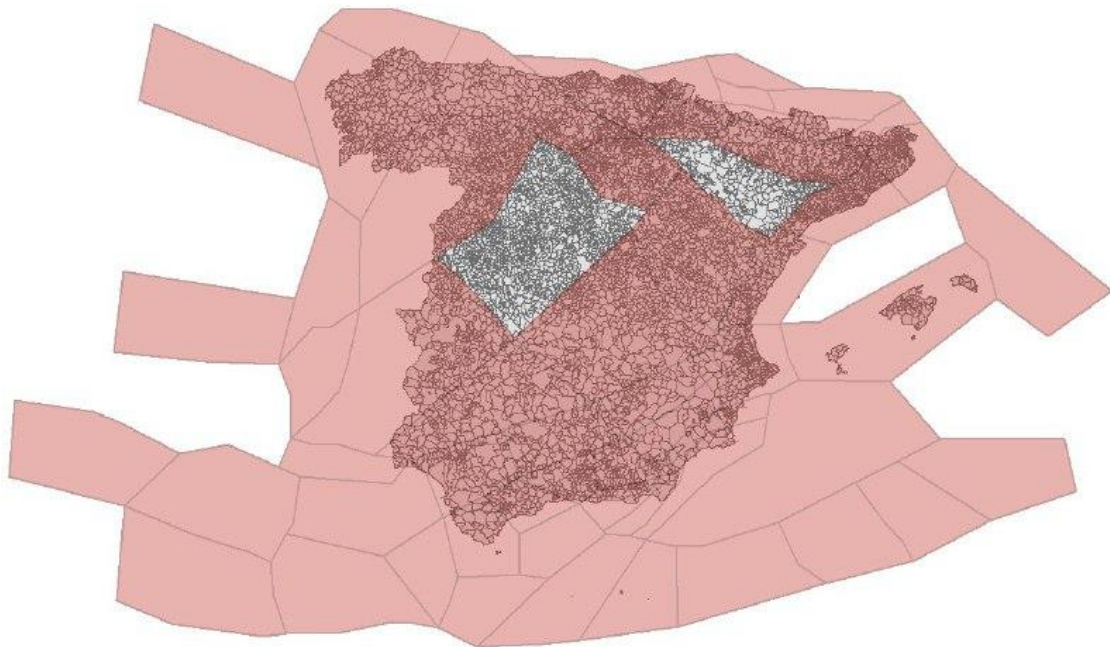


Figura 11.1. Regiones sismográficas de la Península Ibérica

## 11.5. PROXIMIDAD A LA COSTA

Además de la temperatura, la humedad, las precipitaciones u otros indicadores climatológicos, la proximidad al mar de un bien arquitectónico es determinante en su conservación. La salinidad en el ambiente de las zonas cercanas a la costa constituye un factor de corrosión en materiales metálicos [36] así como el arrastre de arena u otros sedimentos debido a la brisa marina contribuye a la erosión de los recubrimientos de los edificios [37]. Según algunos estudios, el efecto de la proximidad del mar en las estructuras se aprecia en un rango de aproximadamente 200 metros [38].

Disponer de las coordenadas geográficas de cada uno de los bienes patrimoniales sería de gran utilidad. Podría establecerse un perímetro de 200 metros en torno a la línea de costa y clasificar como “cercanas a la costa” todas las fichas cuyas coordenadas se encontrasen dentro de dicho perímetro. Por lo tanto, puesto que carecemos de las coordenadas, se ha tenido que establecer otro camino para obtener una aproximación. No podemos generalizar en el ámbito provincial e incluir todas las provincias costeras ya que la mayoría de ellas comparten zona de costa y zona de interior o montaña, por lo que sería muy impreciso. Por consiguiente, en esta situación se ha elegido trabajar con una precisión en el ámbito municipal. La cantidad de municipios costeros en España es de 474, lo que representa aproximadamente un 6% del total. Esto supone un riesgo muy inferior al que teníamos con la tasa sismográfica a la hora de unir los datos, y ha reforzado la decisión de trabajar con ello.

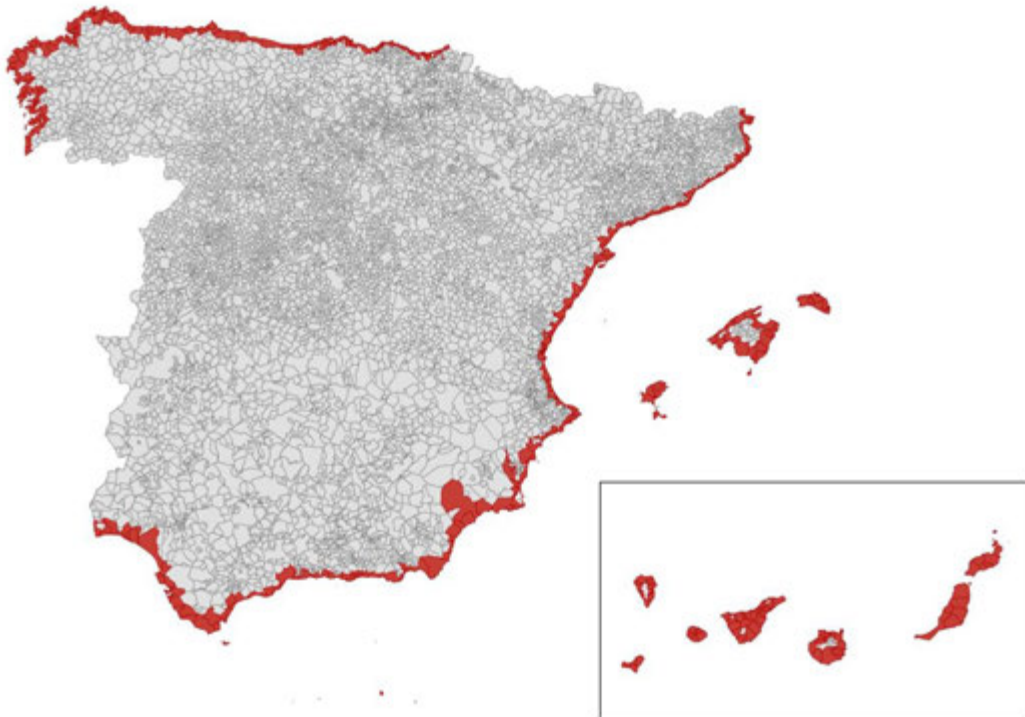


Figura 11.2. Mapa de municipios costeros

Se han extraído de la Base Topográfica Nacional del IGN las líneas de costa coincidentes con las fronteras española, y se han obtenido todos los municipios dispuestos a lo largo de ella [figura 11.2]. Estos datos se fusionarán con el fichero de datos de la Lista Roja de tal forma que se indique si un bien está o no cerca del mar con una precisión en el ámbito municipal. Esto quedará reflejado en una variable binaria llamada **Cercano.a.la.costa**.

## 11.6. NÚMERO DE SOCIOS DE HISPANIA NOSTRA

Una de las primeras hipótesis que se han establecido es que el número de fichas de cada provincia estuviese correlado con el número de socios de Hispania Nostra en dicha provincia. Es fácil deducir que un socio de Hispania Nostra estará más implicado en la conservación del patrimonio histórico, con el deseo aportar información de la Lista Roja para su crecimiento, así como es probable que esa persona conozca más los bienes patrimoniales de la región en la cual reside. Eso explicaría por qué hay provincias de las que se tienen gran cantidad de fichas, como puede ser el caso de Cáceres, y por qué provincias como Barcelona tienen tan pocas en proporción.

Así pues, se ha solicitado a Hispania Nostra una lista con el número de socios por provincia. Para evitar posibles violaciones de la Ley de Protección de Datos Personales 3/2018 no se han solicitado nombres ni apellidos, únicamente el número de socios.

## 11.7. EXTENSIÓN GEOGRÁFICA

Otro de los factores que pueden determinar el número de fichas por cada provincia, además de la hipótesis planteada en el anterior apartado, puede ser la superficie que tienen las distintas regiones. Se valorará tanto en el ámbito provincial como en el ámbito autonómico. Podemos deducir que una extensión de mayor superficie puede abarcar un mayor número de bienes patrimoniales frente a otras de menor extensión y que, por tanto, la Lista Roja recibirá fichas de forma proporcional a la cantidad de patrimonio que hay en cada región. No obstante, esto no tiene por qué cumplirse, ya que es sabido que Jaén es la provincia de España con mayor número de castillos (y también de Europa) [39], y que se trata de la decimocuarta provincia de mayor extensión, con un total de 14.496km. Por ello, trataremos de estudiar si realmente existe una relación entre ambas cosas.



## 12. ANÁLISIS DE LOS DATOS

Finalizado el tratamiento de los datos se requiere una serie de estudios preliminares sobre los contenidos del fichero de datos. Estos estudios se dividen en tres apartados: en el primero se visualizará la distribución geográfica de las fichas de los bienes patrimoniales del fichero de datos; en el segundo se desarrolla un breve estudio descriptivo de atributos numéricos, en concreto de los asociados al estado de conservación de los bienes patrimoniales; y la tercera parte consta de un estudio de los principales atributos categóricos.

En todo momento se ha tenido en cuenta que el fichero de datos de la Lista Roja es tan solo una muestra del total de datos existentes, que desconocemos y al que no tenemos acceso, y que abarcaría todo el patrimonio en riesgo de nuestro país. Esto va suponer un inevitable sesgo muestral, lo que puede conllevar que ninguno de los resultados de este estudio sea concluyente.

### 12.1. ANÁLISIS EXPLORATORIOS

#### 12.1.1. DISTRIBUCIÓN GEOGRÁFICA DE LOS BIENES PATRIMONIALES DE LA LISTA ROJA

La distribución geográfica de los bienes patrimoniales puede indicarnos muchas cosas al respecto de las condiciones de conservación de los mismos. Tal y como hemos visto durante el tratamiento de los datos y la inclusión de los datos externos. Tener una visión global de la situación de las fichas a nivel geográfico constituye algo esencial para conocer los datos con los que se trabaja. Para ello se han representado mediante diferentes gráficas el número de fichas incluidas en el fichero de la Lista Roja. En la *figura 12.1* puede observarse el número total de fichas por cada comunidad autónoma, desglosadas según su tipo (Roja, Verde o Negra). Incluye también las ciudades autónomas de Ceuta y Melilla.

Por otra parte, las *figuras 12.2, 12.3, 12.4 y 12.5* muestran el número de fichas distribuidas por comunidades autónomas en forma de mapa. Estas figuras se corresponden, respectivamente, al conjunto de todas las fichas, al subconjunto de fichas clasificadas como Lista Roja, al subconjunto de fichas clasificadas como Lista Verde y al conjunto de fichas clasificadas como Lista Negra. Esto ha sido posible gracias a la existencia de una biblioteca de R llamada *SP* que permite la visualización de datos asociados a regiones geográficas y representar mediante códigos de colores esos datos en polígonos prediseñados que se corresponden con la forma de las diferentes regiones geográficas.

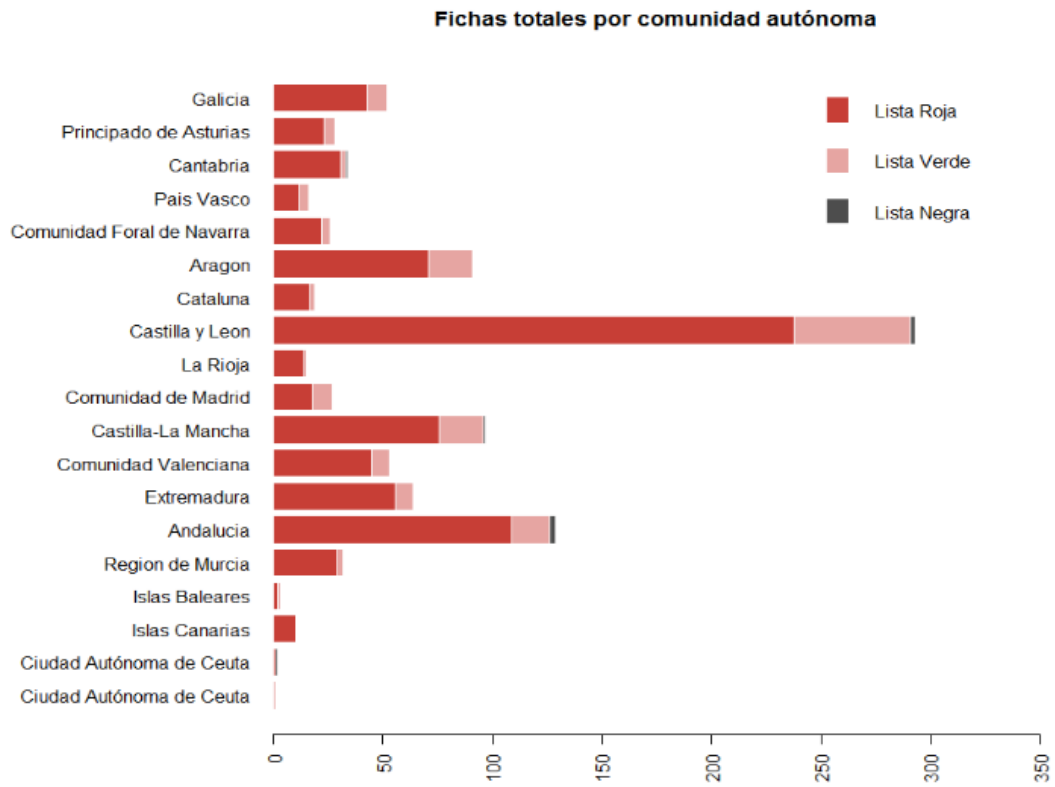


Figura 12.1. Fichas totales por comunidad autónoma

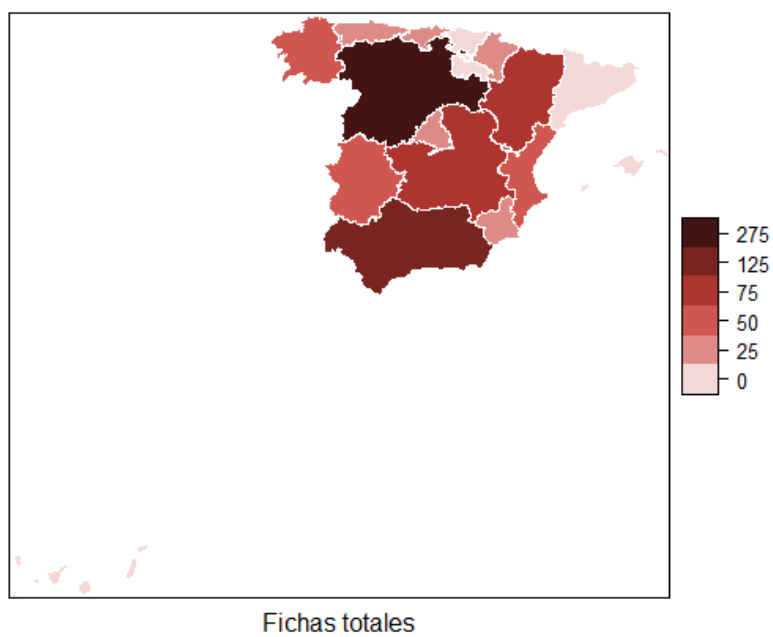
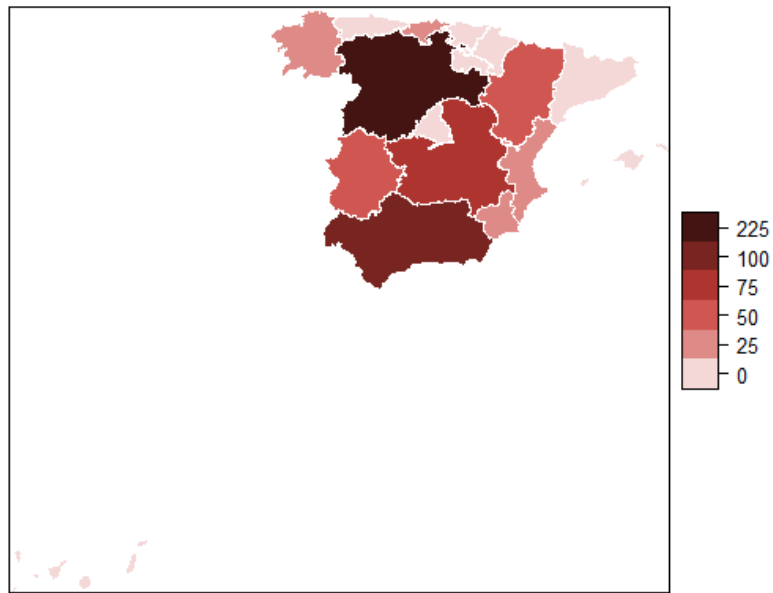
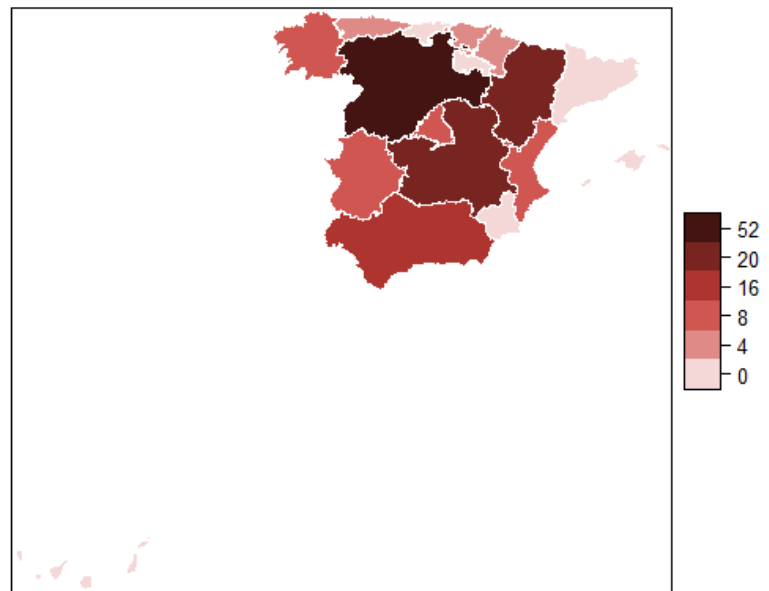


Figura 12.2. Número de fichas totales por comunidad autónoma



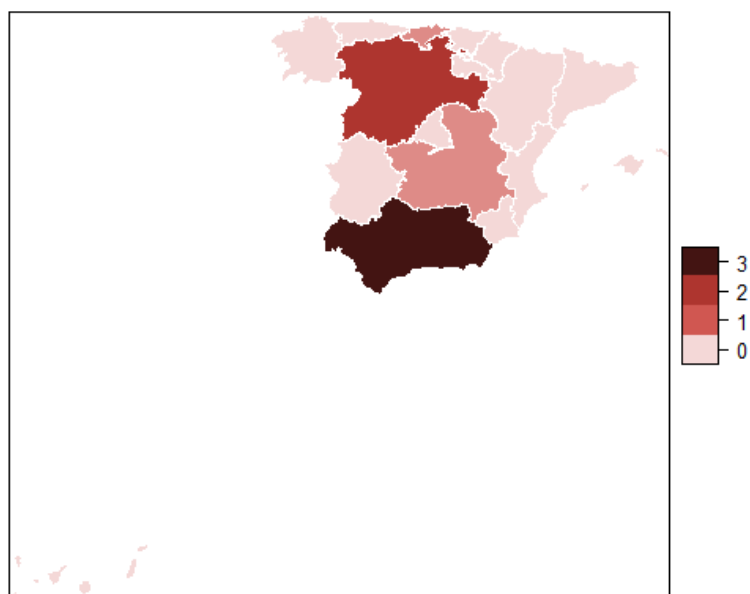
Fichas Lista Roja

Figura 12.3. Número de fichas de la Lista Roja por comunidad autónoma



Fichas Lista Verde

Figura 12.4. Número de fichas de la Lista Roja por comunidad autónoma



Fichas Lista Negra

Figura 12.5. Número de fichas de la Lista Negra por comunidad autónoma

Se observa claramente como Castilla y León es la comunidad autónoma que más destaca de toda la geografía española, poseyendo el mayor número de bienes patrimoniales de cuantos hay recogidos en el fichero de datos, con cerca de un tercio del total. Por ello es también la provincia con mayor número de fichas catalogadas como “Lista Roja”, así como también la que tiene mayor cantidad de bienes intervenidos y que han pasado a la Lista Verde. Estadísticamente, puesto que es la comunidad autónoma con mayor extensión del país, se podría asumir que es una de las comunidades con mayor cantidad de patrimonio histórico de España. Por otro lado, Andalucía le toma la delantera en cuanto a bienes en la Lista Negra, quedando Castilla y León en segundo lugar.

A continuación, se ha representado la información de manera idéntica, pero esta vez en el ámbito provincial. La *figura 12.6* representa el número total de fichas por provincias, desglosadas según su tipo (Roja, Verde o Negra). Las *figuras 12.7, 12.8, 12.9* y *12.10* muestran el número de fichas distribuidas por provincias, nuevamente en forma de mapa. Se corresponden, respectivamente, al conjunto de todas las fichas, al subconjunto de fichas clasificadas como Lista Roja, al subconjunto de fichas clasificadas como Lista Verde y al subconjunto de fichas clasificadas como Lista Negra.

### Fichas totales por comunidad autónoma

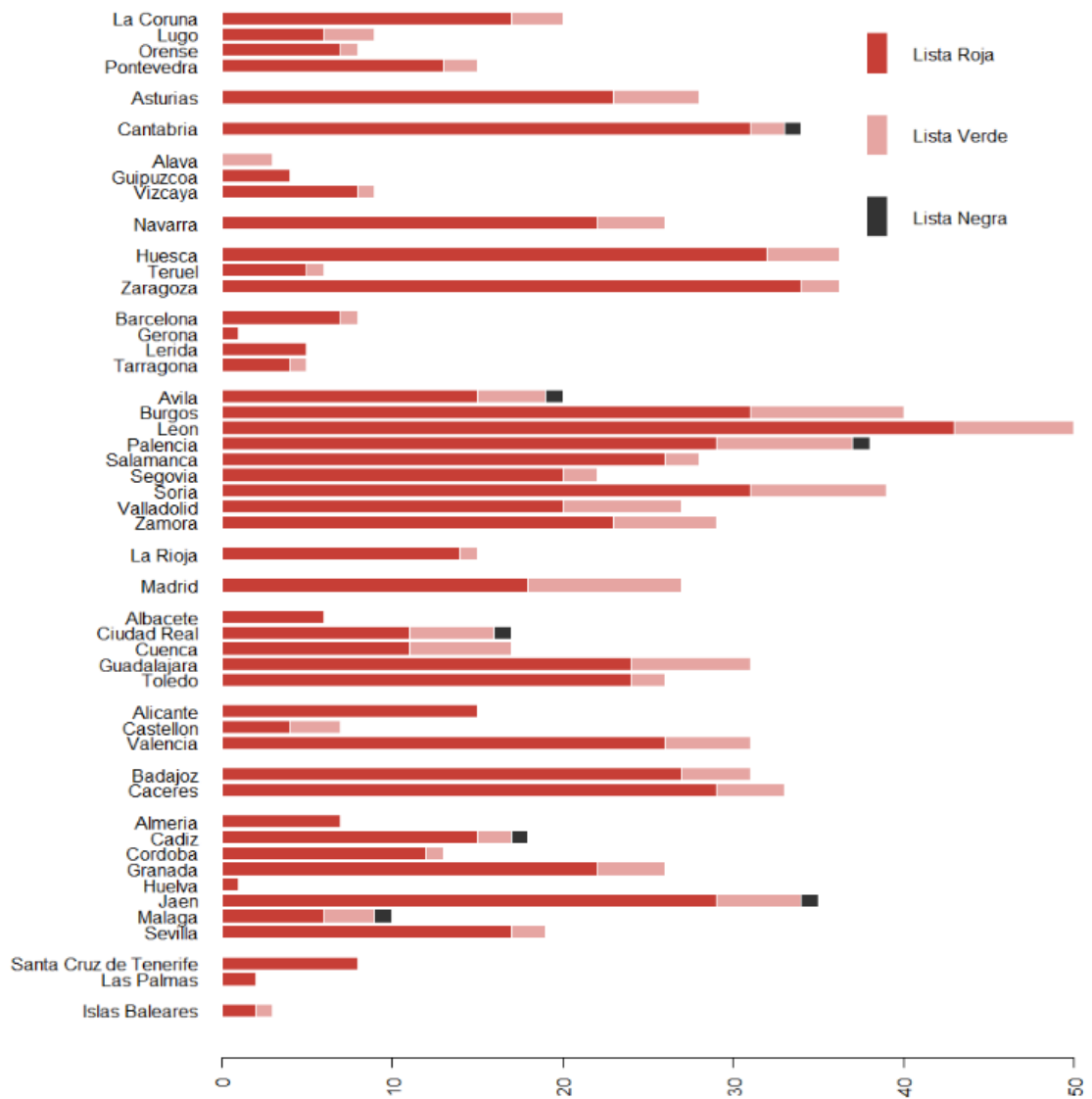
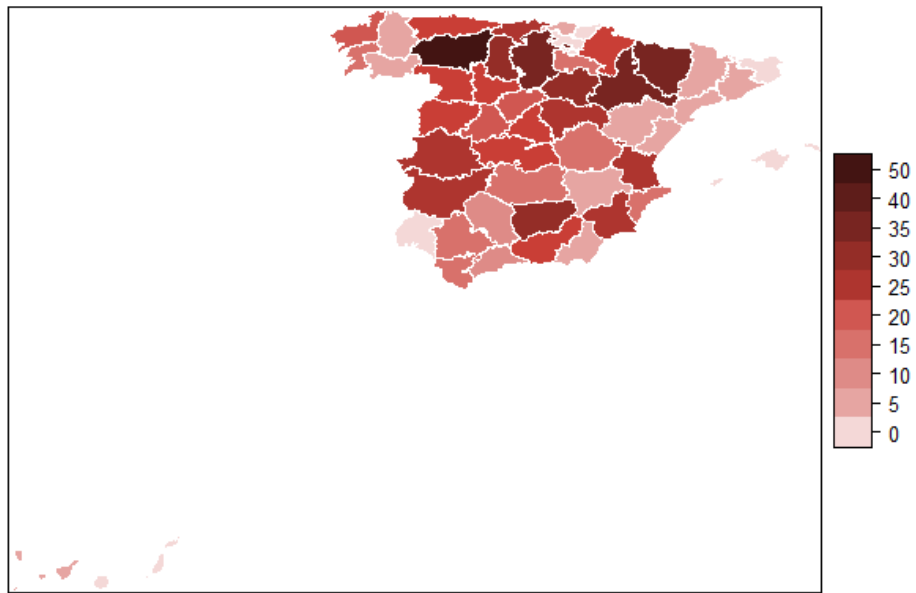
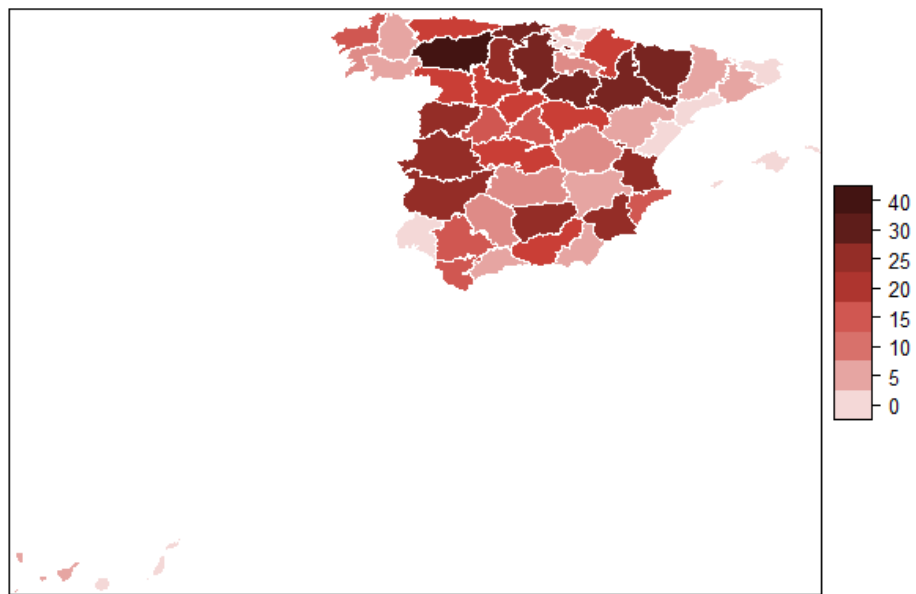


Figura 12.6. Fichas totales por provincia



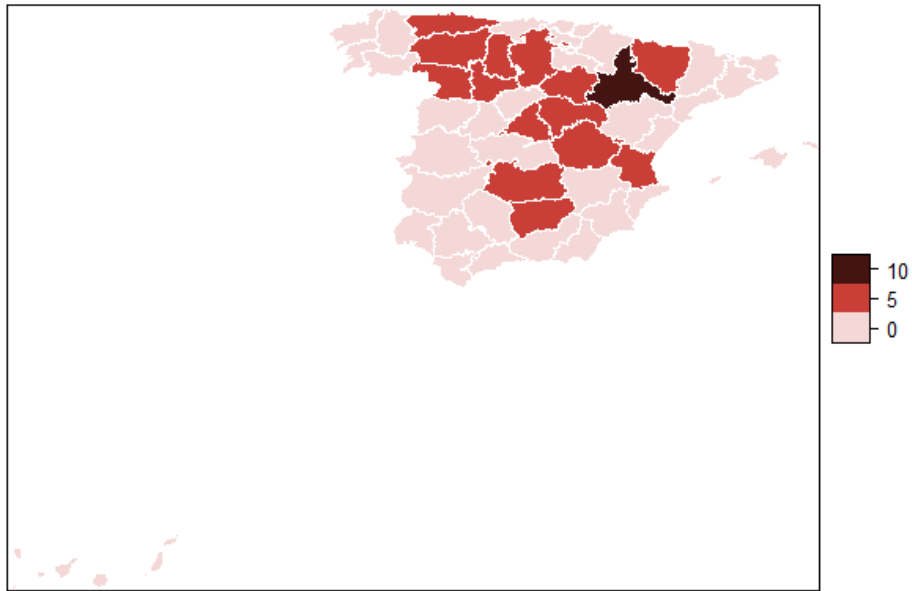
Fichas totales

Figura 12.7. Número de fichas totales por provincia



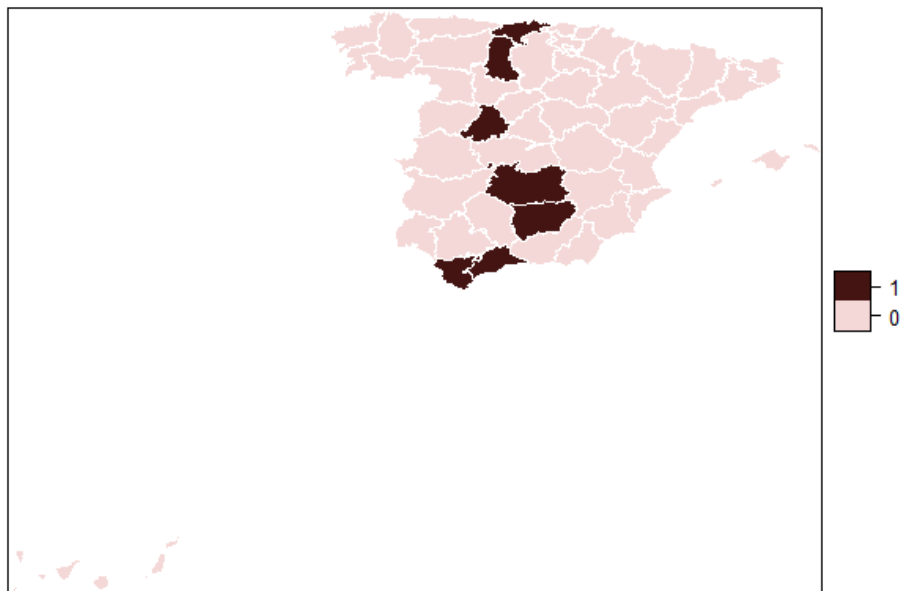
Fichas Lista Roja

Figura 12.8. Número de fichas de la Lista Roja por provincia



Fichas Lista Verde

Figura 12.9. Número de fichas de la Lista Verde por provincia



Fichas Lista Negra

Figura 12.10. Número de fichas de la Lista Negra por provincia

Como hemos visto en las gráficas en el ámbito autonómico, la comunidad autónoma con el mayor número de fichas es Castilla y León. En este caso, León es precisamente la provincia que más tiene a nivel nacional. Lo mismo sucede para las

fichas que se encuentran clasificadas como “Lista Roja”. En el caso de la Lista Verde, pese a que Castilla y León era la comunidad autónoma con mayor número de bienes salvados, en el ámbito provincial Zaragoza se pone a la cabeza. Para las fichas incluidas en la Lista Negra, las siete provincias que poseen bienes en dicha categoría tienen todas ellas la misma cantidad.

### 12.1.2. ESTADO DE CONSERVACIÓN DE LOS BIENES PATRIMONIALES

**Valoración.estado** y **Estado.deterioro** son los dos atributos numéricos que obtuvimos de analizar el estado de conservación de los bienes patrimoniales. Recordemos que **Valoración.estado** nos ofrece una aproximación del mal estado global de cada bien, mientras que **Estado.deterioro** solo contempla el estado estructural o “general” del edificio, sin entrar a valorar factores externos como el abandono, el expolio, el vandalismo o la vegetación. Estos factores externos quedan reflejados en sus respectivos atributos binarios. Se han obtenido las medidas descriptivas de ambos atributos para los bienes de la Lista Roja en tres ámbitos distintos: nacional, autonómico y provincial. A continuación, se recogen los resultados para cada uno de ellos.

#### 12.1.2.1. ANÁLISIS EN EL ÁMBITO NACIONAL

ESPAÑA		
Atributo	Valoración.estado	Estado.deterioro
Rango	10-520	5-355
Media	163.7983	110.6577
Desviación estándar	84.1861	63.553
Moda	80	80
Cuartiles	Q1 (25%)	100
	Q2 (50%)	155
	Q3 (75%)	220

Tabla 12.1. Medidas descriptivas del estado de bienes incluidos en la Lista Roja en el ámbito nacional



El segundo cuartil (Q2), que se corresponde con la mediana, indica que el 50% de datos se concentran entre los valores 5 y 160 para el atributo **Valoración.estado**, y entre los valores 5 y 85 para el atributo **Estado.deterioro**. Teniendo en cuenta que la “escala del mal estado”, según lo establecido por las ponderaciones de los términos durante su tratamiento, va desde 10 hasta 815 para **Valoración.estado**, y desde 5 hasta 607 para **Estado.deterioro**, podemos determinar que, en ambos casos, la mitad de los datos de los que disponemos tienen una valoración de estado muy inferior a la mitad de la escala establecida. Por ello, y valorando en todo momento que la cuantificación de una descripción carece de suficiente rigor y hace de estos datos poco fiables, podemos suponer que no es frecuente tener bienes patrimoniales en estado crítico, y que la mayoría se encuentran en un estado intermedio. Los bienes en estado más crítico rondan los 500 puntos en la “escala del mal estado” de conservación y son valores atípicos.

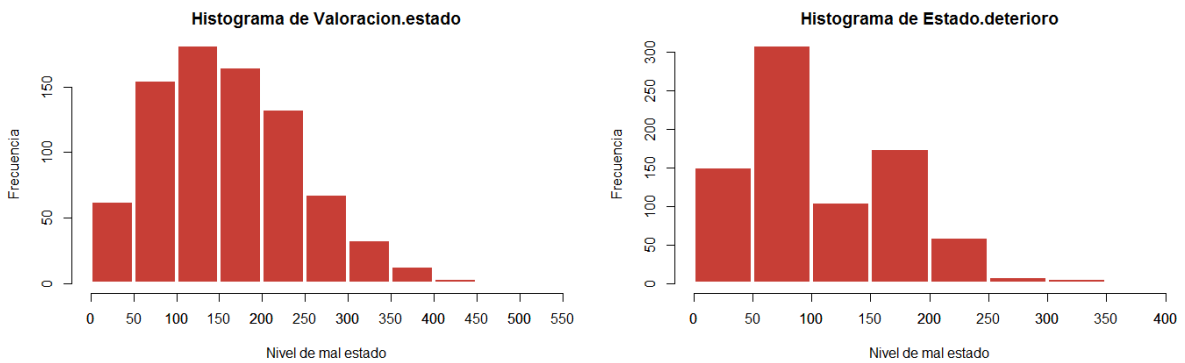


Figura 12.11. Histogramas del estado de conservación

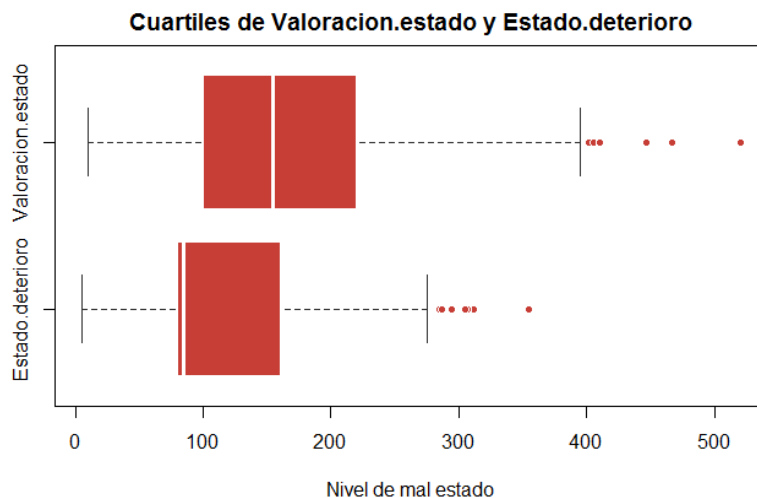


Figura 12.12. Diagramas de cajas del estado de conservación

Para poder entender toda esta información de una manera más visual se han representado los histogramas correspondientes a cada uno y sus gráficos de cajas [figuras 12.11 y 12.12]. Los histogramas nos muestran la frecuencia de los datos agrupados de cincuenta en cincuenta. Los diagramas de cajas nos indican la distribución de los datos en cuartiles, así como los valores atípicos.

**Valoración.estado** siempre toma valores superiores a **Estado.deterioro**, puesto que la primera comprende la información incluida en la segunda y además incluye otros factores que suman puntuación de mal estado. Según el gráfico de densidades representado en la figura 12.13, la información contenida en el atributo **Valoración.estado** se reparte a lo largo de los valores de la escala del estado de conservación, con curvas suaves, mostrando un máximo absoluto en torno al valor 150 y un máximo relativo en torno al valor 90. Por su parte, para **Estado.deterioro** los datos se encuentran más concentrados y tiene un comportamiento más abrupto, como puede apreciarse en su máximo absoluto situado en torno al valor 80 y su máximo relativo en torno a 160.

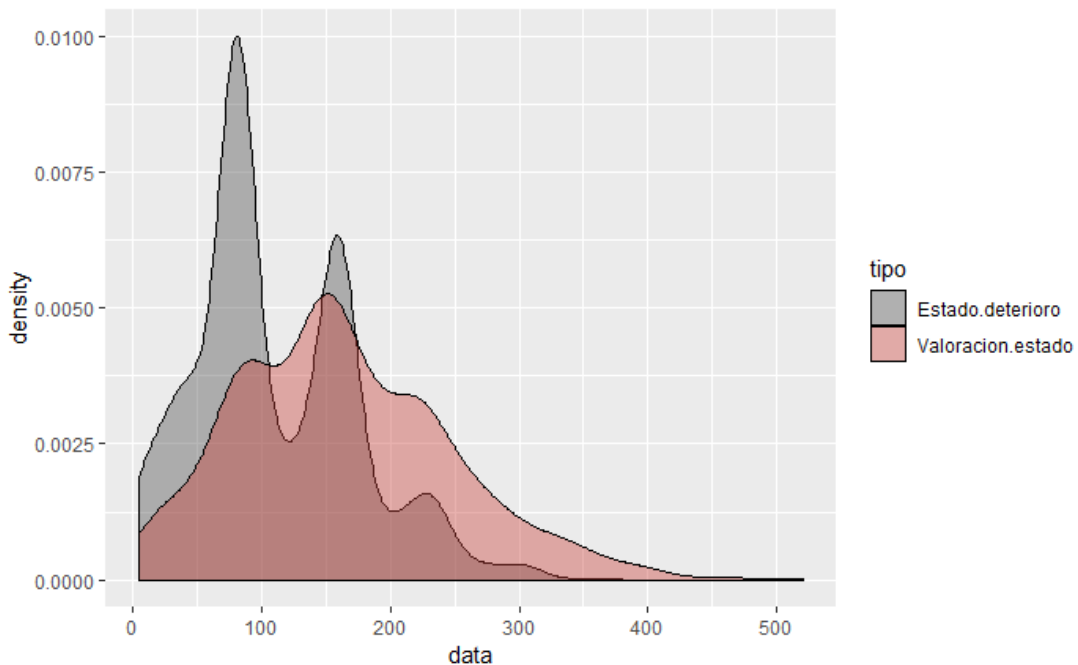


Figura 12.13. Diagramas de densidad del estado de conservación

Al sumarse en **Valoración.estado** los diversos valores de los factores según los atributos de cada ficha, comienza a haber una mayor distinción entre los dos atributos referentes al estado. Las curvas del diagrama de **Valoración.estado** se suavizan y los datos se “esparcen” a lo largo de la escala.

### 12.1.2.2. ANÁLISIS EN EL ÁMBITO AUTONÓMICO

De igual manera se ha realizado un análisis de las medidas descriptivas de ambos atributos pero segmentando la muestra por comunidades autónomas, lo que nos permite tener una visión más focalizada de cómo es la situación en cada una de ellas. En este apartado se expone la interpretación de los resultados. Las tablas con las medidas descriptivas de cada una de las provincias, así como sus histogramas, pueden encontrarse en el *Anexo A* que se encuentra al final del presente documento.

En base a los resultados obtenidos y tomando como referencia las medias de ambos atributos a nivel nacional, es decir, del total de datos incluidos en la categoría Lista Roja, se han identificado aquellas comunidades autónomas cuyos valores medios de estado de conservación se encuentran por encima de la media nacional. La lista es encabezada por las Islas Canarias, seguida por la Región de Murcia, Cataluña y Extremadura, y más moderadamente por la Comunidad Foral de Navarra, La Rioja y Castilla y León, todas ellas para ambos atributos referentes al estado de deterioro. Finalmente se puede apreciar que Aragón queda levemente por encima de la media pero únicamente en el atributo **Estado.deterioro**. Más adelante realizaremos este mismo análisis en el ámbito provincial y compararemos los resultados obtenidos.

Se puede observar una tendencia generalizada en la mayoría de comunidades autónomas por la cual, la distribución de los datos presenta una leve asimetría positiva. Esto nos indica que los datos se concentran más entre los valores bajos y medios de la denominada “escala del mal estado”, mientras que en los valores más altos los datos son más escasos y están más repartidos, algo que es también apreciable en los diagramas de cajas [*figuras 12.14 y 12.15*]. La excepción más destacable es la del Principado de Asturias, cuyo coeficiente de asimetría toma valores negativos, por lo que su distribución de los datos es opuesta a la del resto.

Los valores de los cuartiles nos indican que la mayoría de comunidades autónomas tienen la mayor parte de sus bienes patrimoniales con valores de entre 100 y 150 puntos en la “escala de mal estado”. Observando los rangos, se comprueba que existe un gran número de provincias cuyos valores iniciales de **Valoración.estado** y **Estado.deterioro** empiezan desde 10 y 5, siendo estos los valores por defecto introducidos para aquellas fichas, que tras el tratamiento del atributo **Estado**, permanecían con valor igual a cero. Esto fortalece la idea ya planteada de que el método empleado para la cuantificación del estado de conservación no sea el más indicado.

La desviación estándar es un indicador del nivel de dispersión de los datos con respecto a la media. El valor mínimo que toma es de 68.32 en Cataluña y el máximo es de hasta 130.81 en las Islas Baleares. Esto nos indica que, efectivamente, la dispersión de los datos en las muestras es muy elevada.

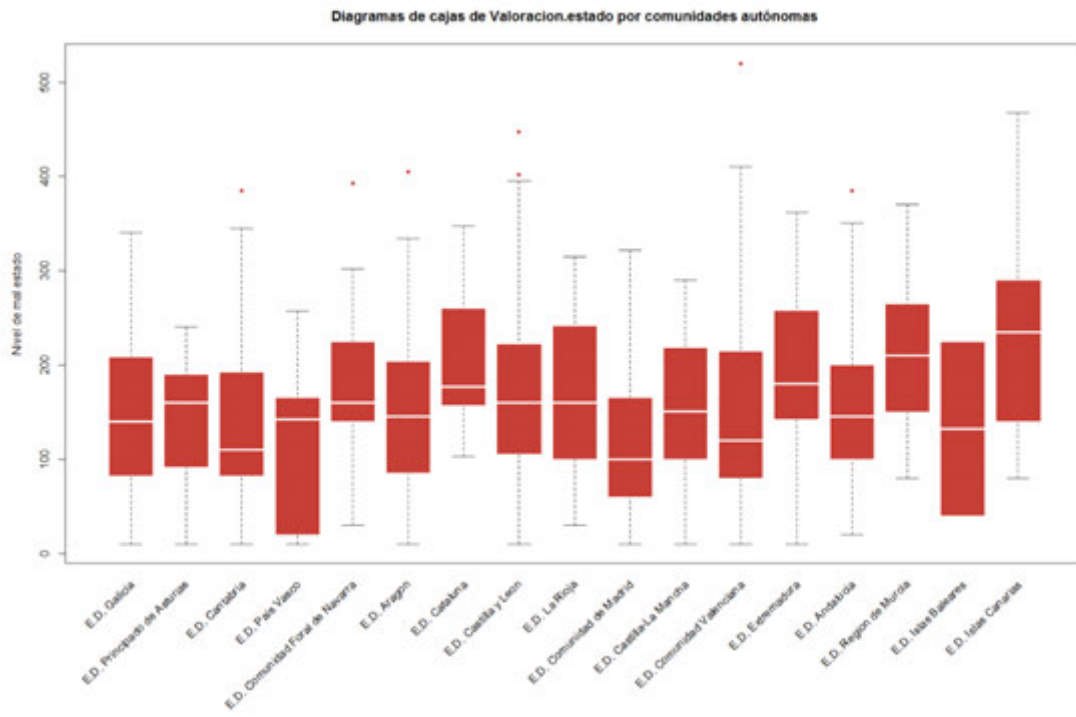


Figura 12.14. Diagramas de cajas de Valoración.estado por comunidades autónomas

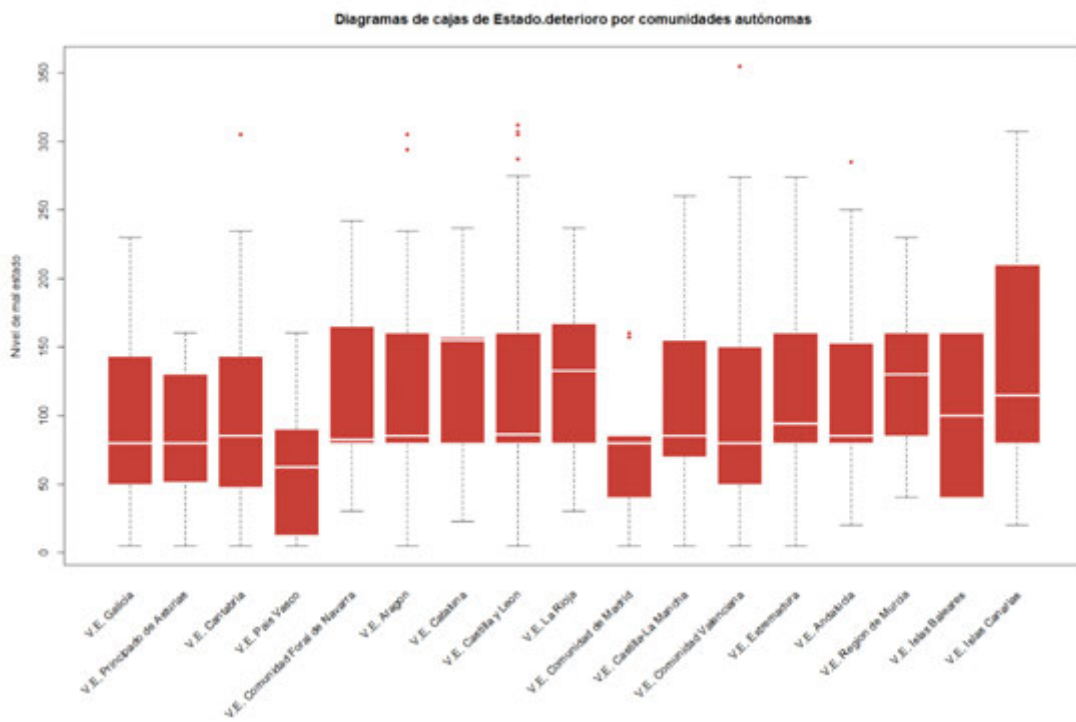


Figura 12.15. Diagramas de cajas de Estado.deterioro por comunidades autónomas

### 12.1.2.3. ANÁLISIS EN EL ÁMBITO PROVINCIAL

De igual forma que se han analizado previamente en el ámbito autonómico, se ha realizado un análisis de las medidas descriptivas de ambos atributos pero segmentando la muestra por provincias. En este apartado se expone la interpretación de los resultados obtenidos. Las tablas con las medidas descriptivas de cada una de las provincias, así como sus histogramas, pueden encontrarse en el *Anexo B* que se encuentra al final del presente documento.

Imitando el procedimiento seguido para el análisis autonómico previo, se han identificado aquellas provincias cuyos valores medios de estado de conservación se encuentran por encima de la media nacional. Un total de dieciséis provincias se encuentran por encima de la media de **Valoración.estado**, encabezadas por Santa Cruz de Tenerife seguida por Teruel, Lérida y Badajoz. Para **Estado.deterioro** el número de provincias con una media mayor a la nacional asciende a veinte, encabezadas por Teruel, y seguida muy de cerca por Lérida y Santa Cruz de Tenerife.

En comparación a los resultados obtenidos en el análisis autonómico previo, Teruel se reafirma como una de las provincias con datos más críticos, mientras que Islas Canarias y Cataluña se reafirman en el ámbito provincial como dos de las regiones geográficas con elementos en peor estado de conservación. No obstante, hay que tener en cuenta que las provincias de Santa Cruz de Tenerife y Lérida no representan ni el 1% de los datos totales, y que un dato mal tratado en estas provincias puede introducir un gran sesgo.

Aunque la mayor parte de las provincias sigue presentando una asimetría positiva tal y como ocurre en las comunidades autónomas, se aprecia una mayor cantidad de distribuciones de datos con asimetrías negativas, algo apreciable en los diagramas de cajas de las *figuras 12.16* y *12.17*. También los valores de los cuartiles nos muestran una mayor variación de valores con respecto al análisis autonómico, con una mayor concentración de los cuartiles segundo y tercero entre los valores 80 y 200. La amplitud de los rangos parece haber disminuido notablemente, lo que se refleja en una menor dispersión de los datos con respecto a la de las comunidades autónomas, algo observable en los valores que toma la desviación estándar.

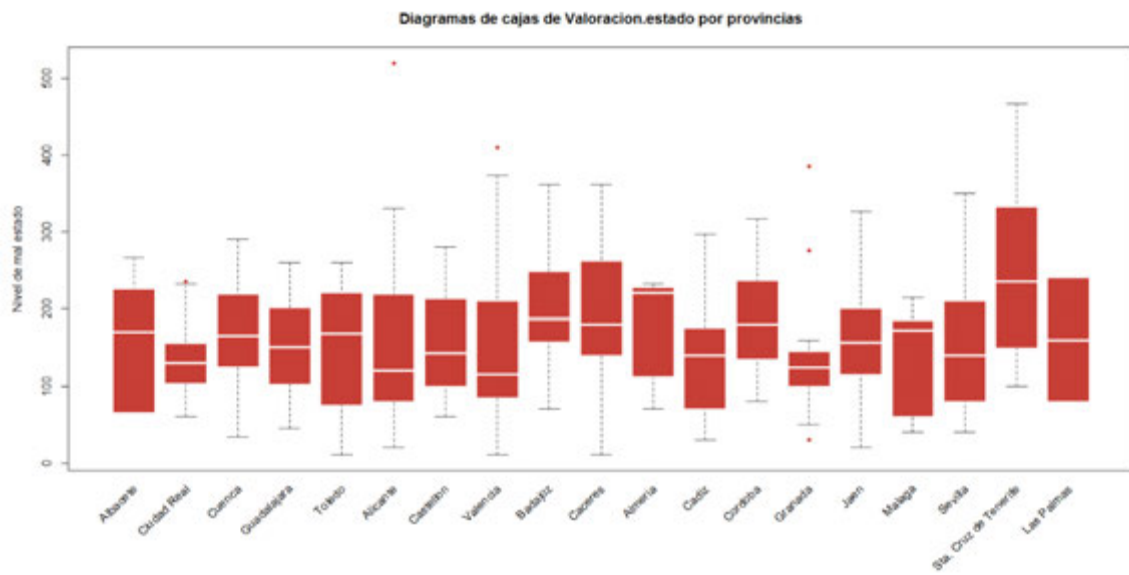
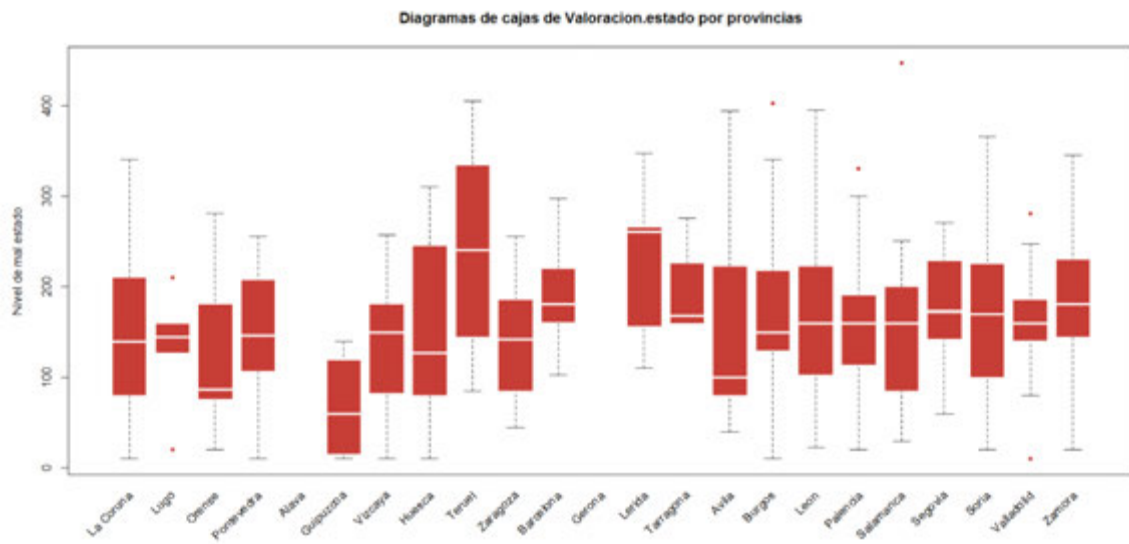


Figura 12.16. Diagramas de cajas de Valoración.estado por provincias

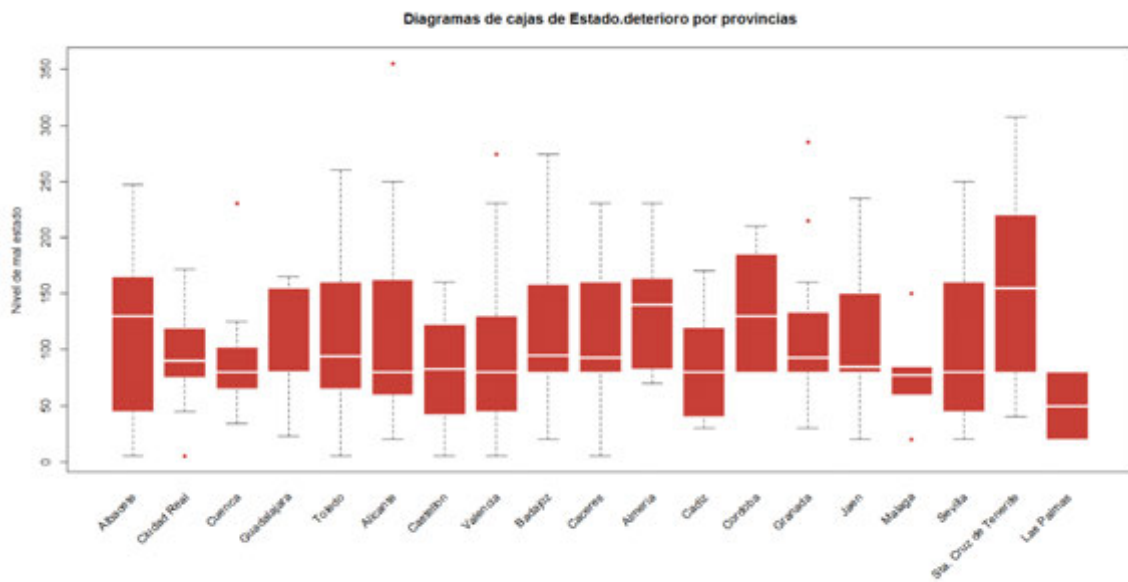
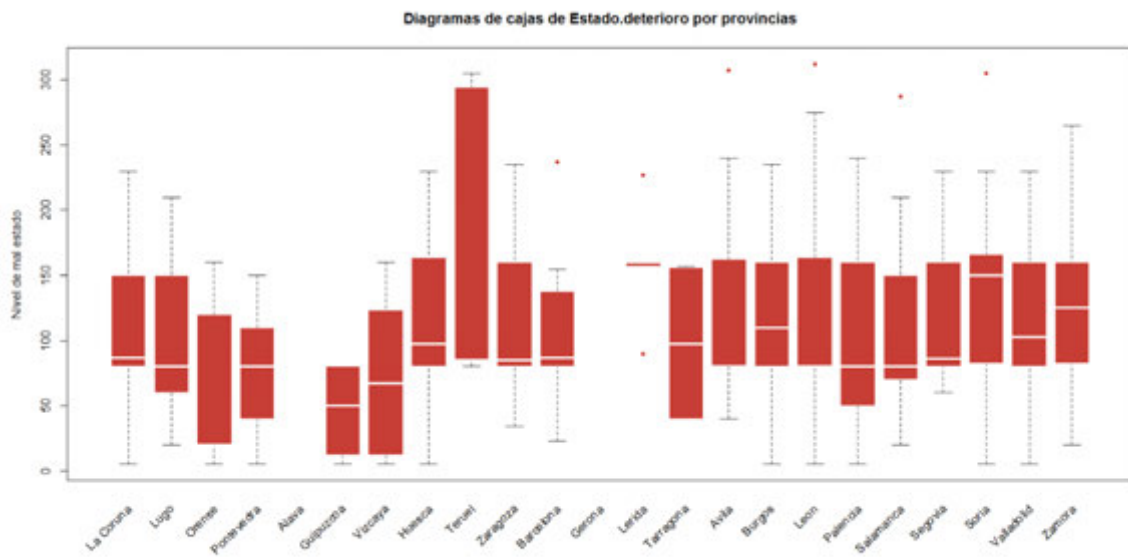


Figura 12.17. Diagramas de cajas de Estado.deterioro por provincias

### 12.1.3. ANÁLISIS EXPLORATORIO DE OTROS ATRIBUTOS NUMÉRICOS

Para finalizar con el análisis exploratorio de atributos numéricos se han visualizado las medidas más significativas de los atributos externos que han sido incluidos en el estudio, tales como los indicadores de la calidad del aire, las variables climatológicas o las variables socioeconómicas. Para tener una visión global se muestran únicamente el rango de valores de los atributos, su media y aquellas provincias en las que presentan sus valores máximo y mínimo.

TEMPERATURA MÁXIMA PROMEDIO (°C)	
<b>Provincia con valor máximo</b>	Córdoba Jaén Sevilla
<b>Provincia con valor mínimo</b>	Asturias Cantabria
<b>Rango</b>	23-36
<b>Media</b>	30.1149

Tabla 12.2. Medidas descriptivas de los valores de la temperatura máxima

TEMPERATURA MÍNIMA PROMEDIO (°C)	
<b>Provincia con valor máximo</b>	Las Palmas
<b>Provincia con valor mínimo</b>	Ávila Burgos Cuenca Salamanca Segovia Soria Teruel
<b>Rango</b>	- 1-16
<b>Media</b>	2.2097

Tabla 12.3. Medidas descriptivas de los valores de la temperatura mínima



VARIACIÓN MÁXIMA DE TEMPERATURA (°C)	
<b>Provincia con valor máximo</b>	Granada Jaén
<b>Provincia con valor mínimo</b>	Ceuta
<b>Rango</b>	5-18
<b>Media</b>	13.6875

Tabla 12.4. Medidas descriptivas de los valores de la variación de temperatura

HUMEDAD RELATIVA MÁXIMA (%)	
<b>Provincia con valor máximo</b>	Valencia
<b>Provincia con valor mínimo</b>	Ávila Guadalajara Jaén León Madrid Palencia Salamanca Segovia Soria Teruel Toledo Valladolid Zamora
<b>Rango</b>	0-78
<b>Media</b>	11.9576

Tabla 12.5. Medidas descriptivas de los valores de la humedad máxima

<b>VELOCIDAD MÁXIMA DEL VIENTO (%)</b>	
<b>Provincia con valor máximo</b>	S. C. de Tenerife
<b>Provincia con valor mínimo</b>	Lugo
<b>Rango</b>	8-24.1
<b>Media</b>	15.5446

Tabla 12.6. Medidas descriptivas de los valores de la velocidad máxima del viento

<b>PRECIPITACIÓN MÁXIMA (mm)</b>	
<b>Provincia con valor máximo</b>	Pontevedra
<b>Provincia con valor mínimo</b>	Las Palmas
<b>Rango</b>	15-151
<b>Media</b>	55.6119

Tabla 12.7. Medidas descriptivas de los valores de la precipitación máxima

<b>PROBABILIDAD DE PRECIPITACIÓN MÁX. (%)</b>	
<b>Provincia con valor máximo</b>	Lugo
<b>Provincia con valor mínimo</b>	Las Palmas
<b>Rango</b>	9-48
<b>Media</b>	25.1159

Tabla 12.8. Medidas descriptivas de los valores de la probabilidad de precipitación máxima

NOx ( $\mu\text{g}/\text{m}^3$ )	
<b>Provincia con valor máximo</b>	Barcelona
<b>Provincia con valor mínimo</b>	Zamora
<b>Rango</b>	1.3789-49.7822
<b>Media</b>	22.9449

Tabla 12.9. Medidas descriptivas de los valores de NOx

NO2 ( $\mu\text{g}/\text{m}^3$ )	
<b>Provincia con valor máximo</b>	Madrid
<b>Provincia con valor mínimo</b>	Zamora
<b>Rango</b>	3.5209-29.5093
<b>Media</b>	13.7137

Tabla 12.10. Medidas descriptivas de los valores de NO2

SO2 ( $\mu\text{g}/\text{m}^3$ )	
<b>Provincia con valor máximo</b>	Murcia
<b>Provincia con valor mínimo</b>	Cáceres
<b>Rango</b>	0.9160-8.2338
<b>Media</b>	3.6522

Tabla 12.11. Medidas descriptivas de los valores de SO2

PIB PER CÁPITA [2017] (€)	
Provincia con valor máximo	Murcia
Provincia con valor mínimo	Cádiz
Rango	17.231-36.921
Media	22.8530

Tabla 12.12. Medidas descriptivas de los valores del PIB per cápita

PARO [2017] (%)	
Provincia con valor máximo	Cádiz
Provincia con valor mínimo	Guipúzcoa
Rango	9.017-30.280
Media	16.7665

Tabla 12.13. Medidas descriptivas de los valores del paro

Finalmente se representa la distribución geográfica del número de socios con los que cuenta Hispania Nostra en cada provincia.

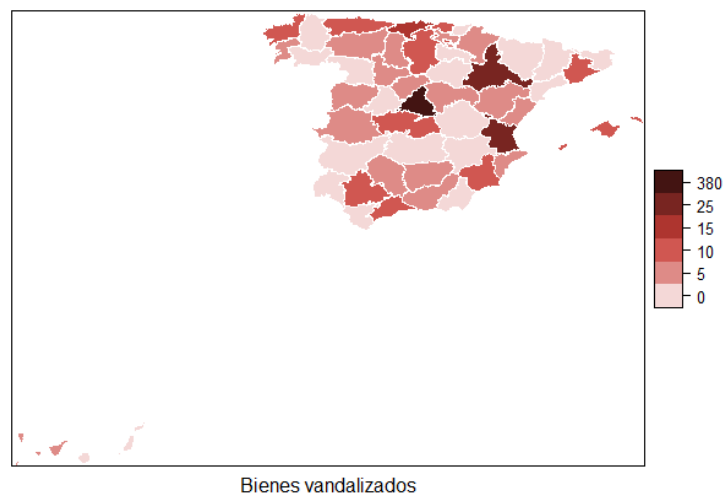


Figura 12.18. Distribución geográfica del número de socios

#### 12.1.4. ANÁLISIS EXPLORATORIO DE ATRIBUTOS BINARIOS

A continuación, se representan las distribuciones geográficas de los diferentes atributos binarios asociados al estado de conservación de los bienes patrimoniales. Estos son **Estado.abandono**, **Estado.expolio**, **Estado.vandalismo** y **Estado.vegetación**. Además, se muestran los atributos asociados a los diferentes riesgos que pueden presentar los bienes patrimoniales, tales como **Riesgo.de.abandono**, **Riesgo.expolio**, **Riesgo.deterioro**, **Riesgo.estructural**, **Riesgo.ruina** y **Riesgo.demolición**. En estos mapas quedan representados los valores positivos de los correspondientes atributos binarios.

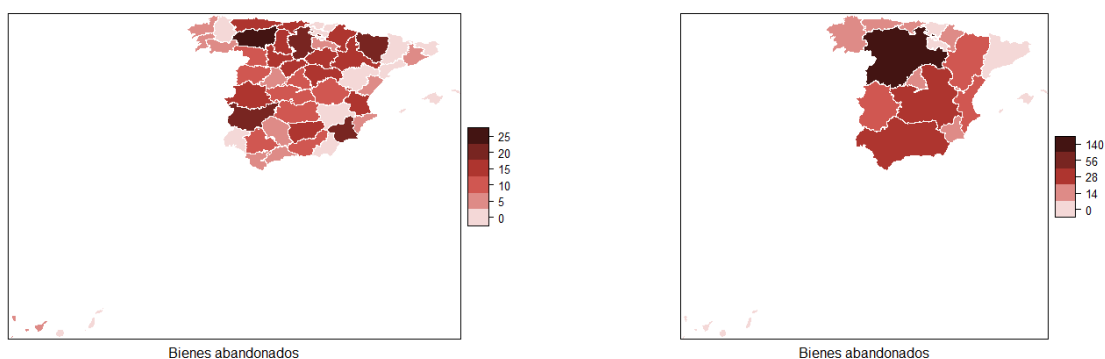


Figura 12.19. Distribuciones geográficas de bienes abandonados

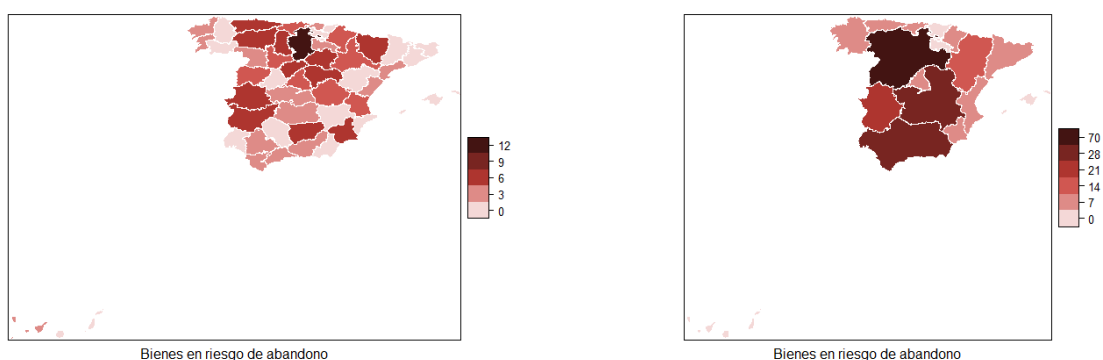


Figura 12.20. Distribuciones geográficas de bienes en riesgo de abandono

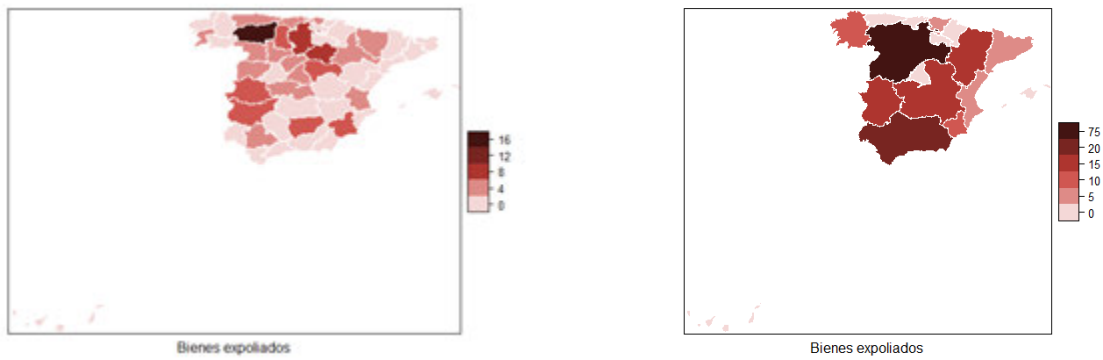


Figura 12.21. Distribuciones geográficas de bienes expoliados

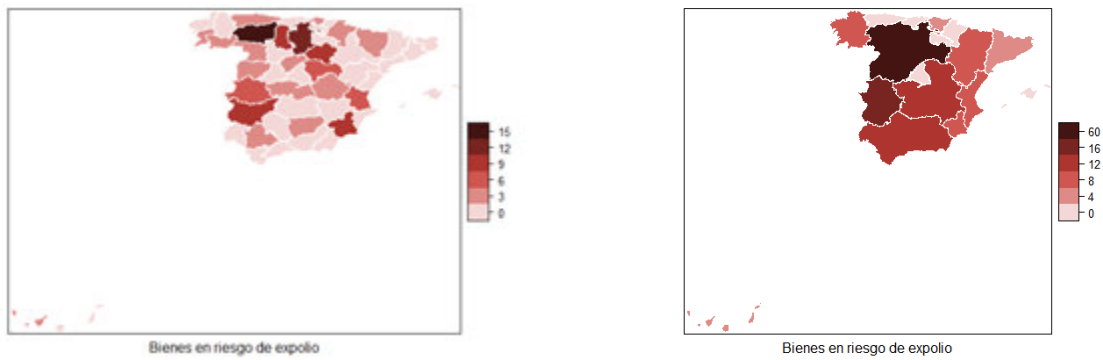


Figura 12.22. Distribuciones geográficas de bienes en riesgo de expolio

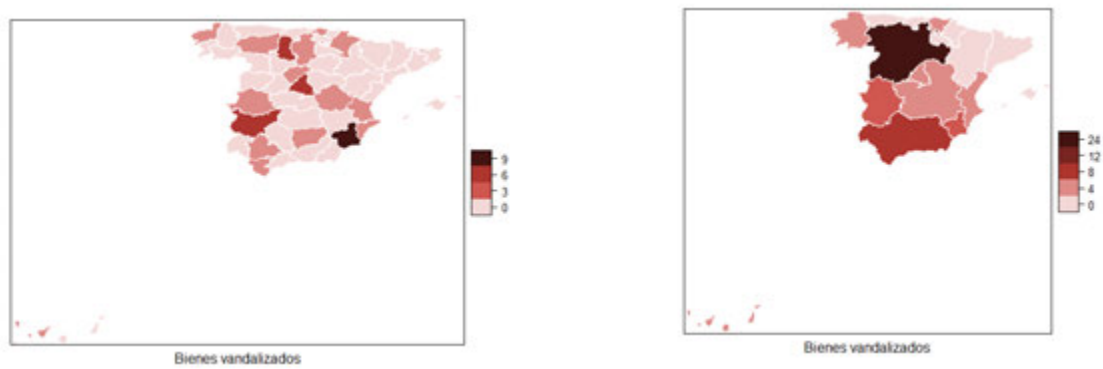


Figura 12.23. Distribuciones geográficas de bienes vandalizados

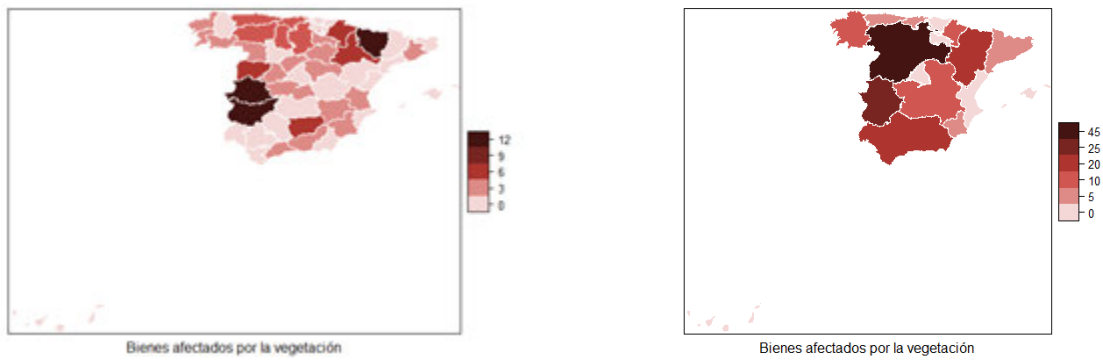


Figura 12.24. Distribuciones geográficas de bienes afectados por la vegetación

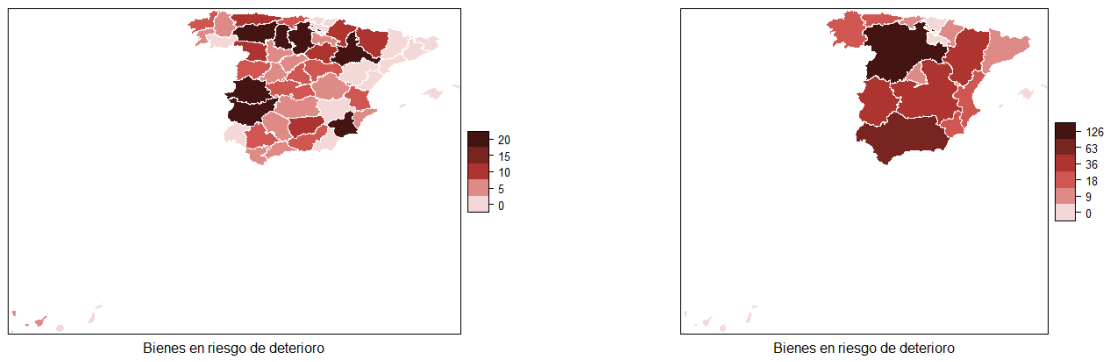


Figura 12.25. Distribuciones geográficas de bienes en riesgo de deterioro

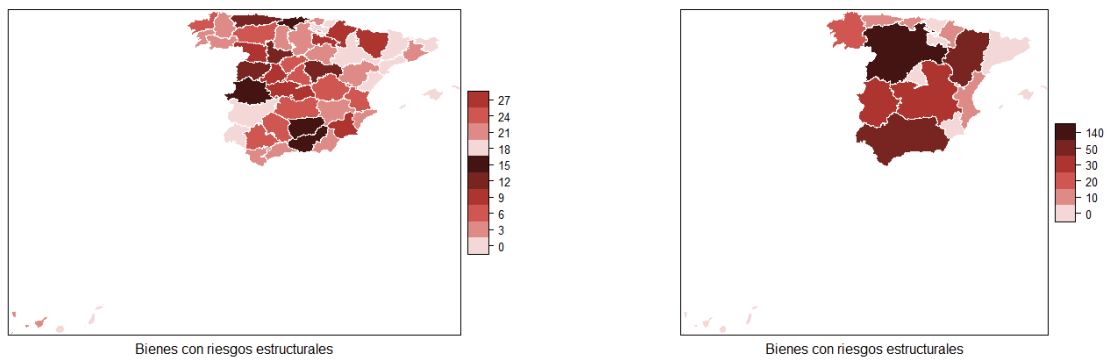


Figura 12.26. Distribuciones geográficas de bienes con riesgos estructurales

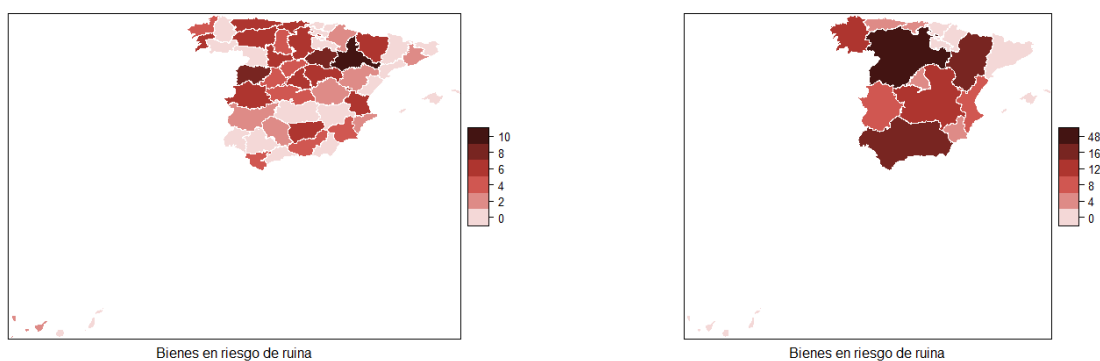


Figura 12.27. Distribuciones geográficas de bienes en riesgo de ruina

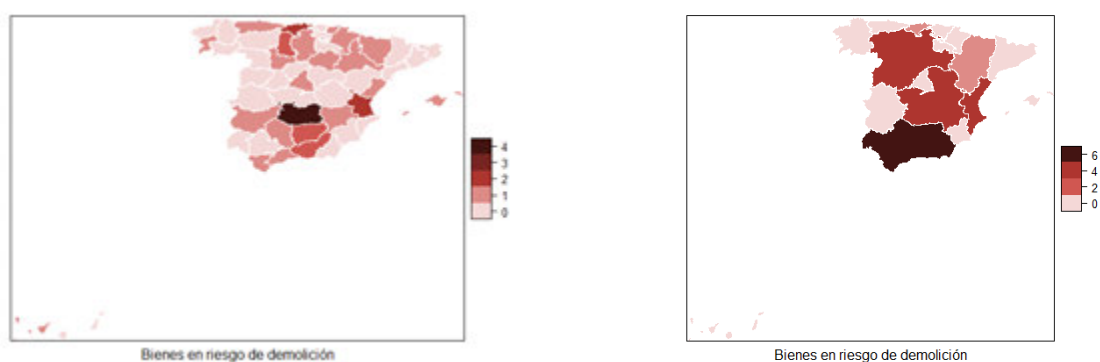


Figura 12.28. Distribuciones geográficas de bienes en riesgo de demolición

En estos mapas se aprecia una concentración significativa de los datos en la comunidad autónoma de Castilla y León, que como se ha visto anteriormente, es la que mayor número de fichas tiene, lo que induce a pensar que esta información podría encontrarse excesivamente condicionada por la muestra de datos de la que disponemos y que, por tanto, puede presentar un sesgo elevado. Algo similar pasa con Andalucía, la cual constituía otra de las comunidades autónomas con mayor concentración de datos. Más adelante se estudiarán las correlaciones entre los diversos atributos para comprobar si esto es así o no.

### 12.1.5. ANÁLISIS EXPLORATORIO DE ATRIBUTOS NOMINALES

A continuación, se expone el análisis de los principales atributos nominales tratados a lo largo del proyecto. Se han estudiado de manera superficial las diferentes categorías que pueden tomar los bienes patrimoniales en estos atributos, comprobando la cantidad de bienes asociados a cada una de ellas a nivel porcentual. No obstante, las afirmaciones realizadas no pueden considerarse categóricas, ya que solo tenemos conocimiento de los datos incluidos en la muestra del total que constituye la Lista Roja de Patrimonio.



Para el atributo **Grado.de.protección.legal** se ha hecho el correspondiente análisis. Los resultados nos indican que entre las categorías “Bien de Interés Cultural”, “Grado de protección desconocido”, “Sin protección específica” y “Decreto del 22 de abril de 1949” comprenden el 94% de fichas. El alto número de fichas cuyo grado de protección es desconocido resulta a priori alarmante. Esto ha motivado un análisis en mayor profundidad de la extracción de información del atributo para asegurar que no se esté perdiendo una importante cantidad de información útil debido a un tratamiento incorrecto. Una vez corregido el tratamiento, se ha repetido el análisis exploratorio.

GRADO DE PROTECCIÓN	Total	Roja	Verde	Negra
Sin protección específica	258	238	18	2
BIC	336	248	86	2
Decreto de 22 de abril de 1949	105	83	22	0
Se desconoce	140	123	16	1
BRL	17	16	1	0
P.G.O.U	15	12	3	0
Inventario	106	85	18	3

Tabla 12.14. Número de fichas según los tipos de protección más frecuentes

Se observa que, si bien los BIC son los bienes más frecuentes en la Lista Roja, el 25% de ellos han sido intervenidos, lo que los convierte en los bienes con mayor índice de recuperación. En oposición, los bienes sin protección constituyen el segundo conjunto con mayor presencia en la Lista Roja y su índice de recuperación es de un escaso 7%.

A continuación, analizaremos los atributos **Propiedad** y **Propietarios**. Es importante recordar que un bien patrimonial puede tener varios propietarios. Para esta ocasión no se ha entrado a valorar en profundidad el que cada propiedad pueda ser compartida entre varias entidades, por lo que los valores representados en la siguiente tabla no representan bienes exclusivos de cada propiedad, sino el total de ellos, incluyendo los que pudiesen ser compartidos.

<b>PROPIEDAD</b>	<b>Total</b>	<b>Rojo</b>	<b>Verde</b>	<b>Negra</b>
<b>Público</b>	302	239	60	3
Municipal	113	93	19	1
Diputación	7	6	1	0
Gobierno Autónomico	12	8	4	0
Ministerio	16	12	2	2
Estado	0	0	0	0
Confederación hidrográfica	4	3	1	0
<b>Privado</b>	543	473	65	5
Diócesis	30	19	11	0
Particular	86	77	8	1
Empresa	78	62	15	1
<b>Desconocido</b>	180	136	44	0
Fundación	5	2	3	0

Tabla 12.15. Número de fichas según los tipos de propietario más frecuentes

El 54% de la muestra está formado por bienes de propiedad privada frente a un 30% de bienes de propiedad pública. Un 11% del total pertenecen a entidades administrativas municipales. Las propiedades públicas son a su vez las que mayor índice de recuperación de bienes patrimoniales tienen, con aproximadamente un 20%, frente al 12% que presentan los bienes de propiedad privada.

En último lugar, se ha realizado el mismo análisis para el atributo **Motivo.de.la.retirada**. Las mediciones de las diversas categorías adoptadas por el atributo nos muestran que aproximadamente un 69% del total de bienes retirados ha sido gracias a la realización de labores de restauración, lo que constituye un 72% de los bienes que han pasado a la Lista Verde. Esto pone de manifiesto la importancia de la restauración de los bienes patrimoniales para su salida frente a otro tipo de intervenciones a priori más superficiales como la limpieza, la rehabilitación o la conservación.

MOTIVO DE LA RETIRADA	Total
Restauración	123
Rehabilitación	6
Limpieza	2
Conservación	1
Desaparición de amenazas	5
Consolidación	22
Demolición	5
Alteración de sus valores	3

Tabla 12.16. Número de fichas según los tipos de motivos de la retirada más frecuentes

Estos resultados podrían estar estrechamente vinculados a las distribuciones de los bienes patrimoniales en la escala de mal estado, anteriormente analizados. El estado de conservación de la mayoría de los bienes se concentraba en los valores intermedios de la “escala del mal estado”. Esto puede interpretarse como que la mayoría de los bienes se encuentran en un estado que requiere de intervenciones críticas, como una restauración, en lugar de limpiezas o rehabilitaciones que se corresponden con un estado de deterioro más leve.

## 12.2. CORRELACIONES LINEALES

El objetivo, llegados a este punto del estudio, es tratar de encontrar relaciones entre los diversos atributos que hemos ido recopilando y transformando, para aproximarnos a un modelo predictivo que nos permitiese conocer el estado o el riesgo en el que se encuentra un bien patrimonial en función de las condiciones, es decir, las variables, a las que se ve sometido. Para ello recurrimos al coeficiente de correlación de

Pearson, que indica si existe una relación de dependencia entre variables numéricas, la fuerza que tiene dicha relación lineal y su dirección, la cual puede ser positiva (ambas variables aumentan o disminuyen a la vez) o negativa (mientras una variable aumenta, la otra disminuye). Se asume para todos los próximos estudios de correlación un p-valor inferior a 0.05 para considerar significativo un resultado.

### 12.2.1. NÚMERO DE FICHAS

Para dar respuesta a algunas de las conjeturas formuladas a lo largo de los análisis exploratorios, se ha estudiado la correlación que pudiese existir entre el número de fichas y la cantidad de socios de Hispania Nostra o la extensión geográfica, tanto en el ámbito provincial como autonómico.

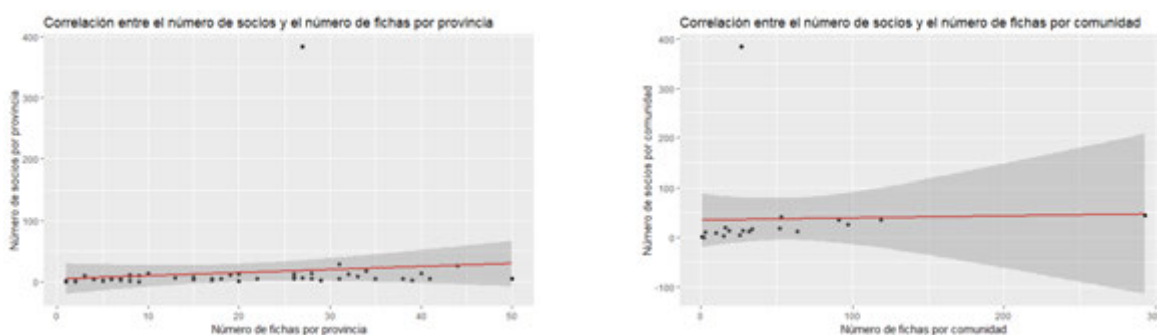


Figura 12.29. Diagramas de dispersión del número de socios y el número de fichas

En el ámbito provincial, los atributos referentes al número total de fichas y el número de socios de Hispania Nostra presentan una correlación de 0.12. El número de socios residentes en Madrid constituye un importante valor atípico, que sea seguramente debido a que la sede de Hispania Nostra esté asentada en la capital. Por su parte, la cantidad de fichas y la extensión de cada provincia presentan una correlación considerablemente mayor, de 0.55, aunque la dispersión de estas medidas es mucho mayor, tal y como puede apreciarse en la nube de puntos de la primera gráfica de la *figura 12.30*.

En el ámbito autonómico, la correlación entre el número de fichas incluidas en la Lista Roja y el número de socios de Hispania Nostra disminuye hasta un 0.03, mientras que la correlación entre el número de fichas y la extensión geográfica aumenta hasta un 0.87, presentando una dispersión menor que en el análisis en el ámbito provincial, como se aprecia en la segunda gráfica de la *figura 12.30*.

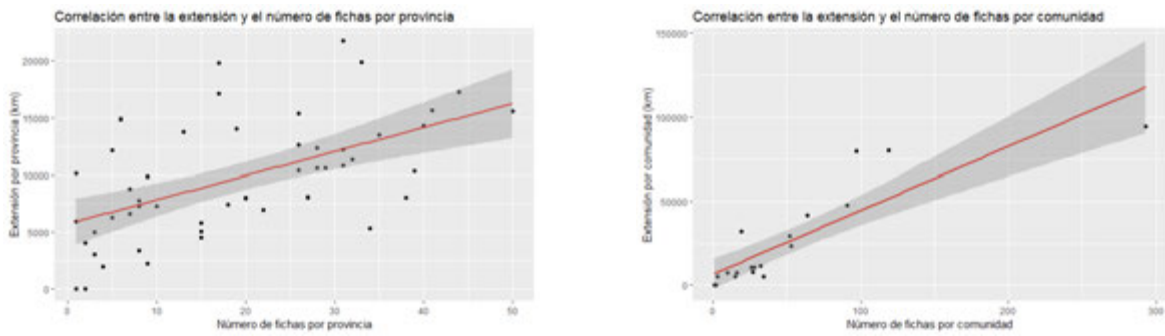


Figura 12.30. Diagramas de dispersión de la extensión geográfica y el número de fichas

Por otro lado, y con el objetivo de identificar si el número de socios depende de algún otro atributo, se ha comprobado la correlación entre la población de cada provincia y el número de socios. Esta toma un valor de 0.72. Sin embargo, podemos observar cómo los valores asociados a Madrid y Barcelona se presentan como valores atípicos al resto, completamente alejados de la nube de puntos. Por consiguiente, se ha repetido el análisis suprimiendo los valores de ambas provincias. El resultado presenta una correlación de 0.61. En el ámbito autonómico la correlación entre estos dos atributos disminuye hasta un 0.5.

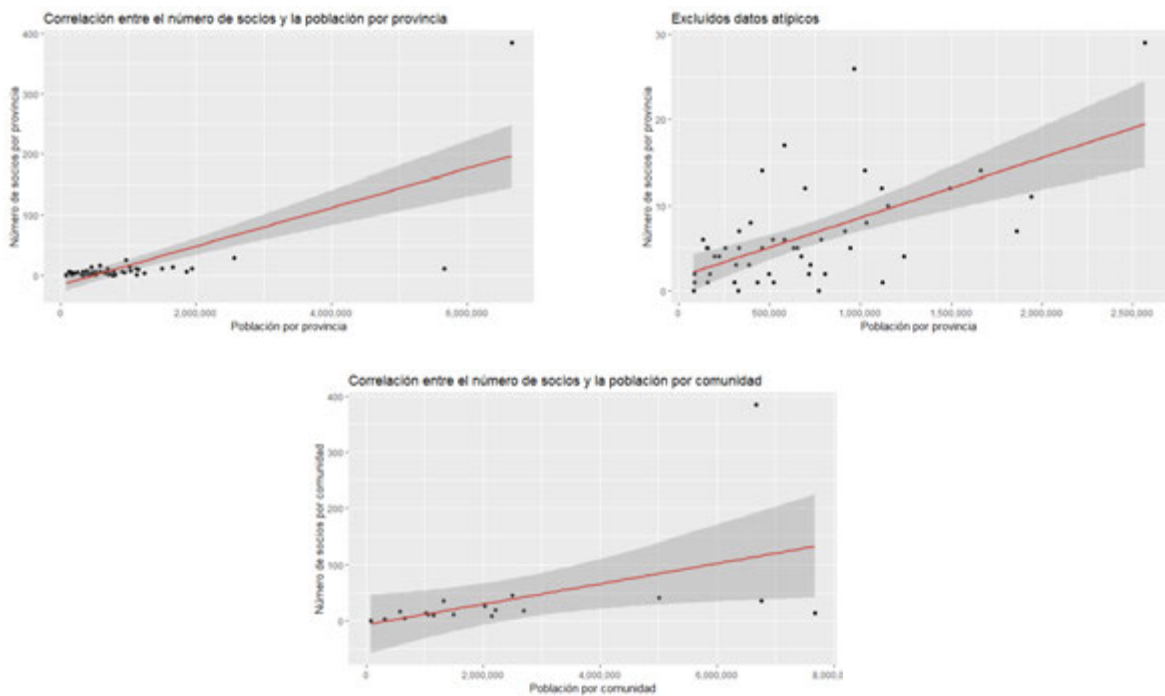


Figura 12.31. Diagramas de dispersión del número de socios y la población

Finalmente se ha comprobado si existe una relación entre el número de fichas por provincias y las variables socioeconómicas, climáticas y de la calidad del aire asociadas a cada una. Se observa que la mayoría de los resultados obtenidos muestran una correlación prácticamente nula. Los únicos resultados relevantes, recogidos en la matriz de correlaciones de la *figura 12.32*, nos indican que únicamente existe cierta correlación por la que, a mayor variación de temperatura, a menos temperatura mínima y a menos humedad, mayor es el número de fichas por provincia.

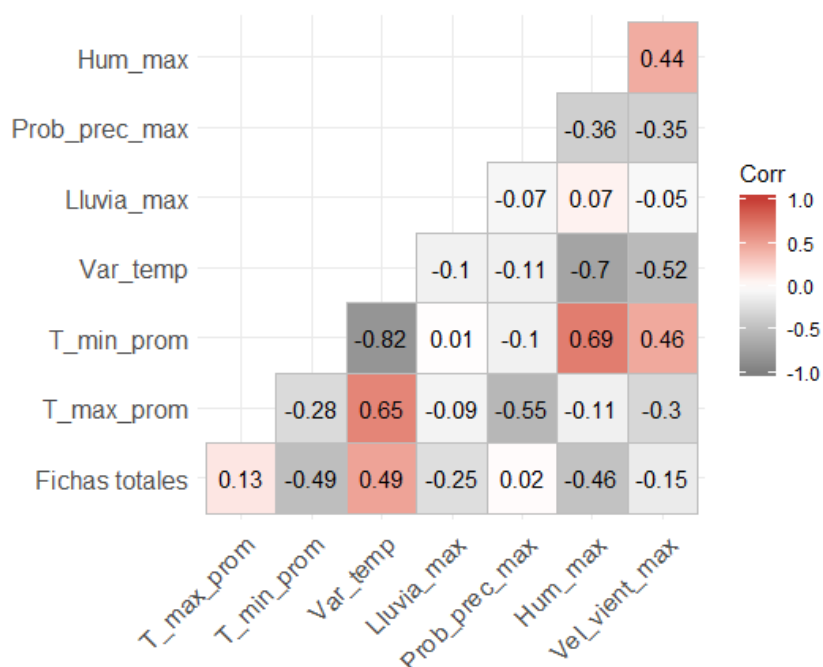


Figura 12.32. Matriz de correlaciones del número de fichas y las variables climáticas

Se ha tratado de realizar este mismo análisis con las fichas según su clasificación en Lista Roja, Verde o Negra, pero los resultados presentan un elevado error muestral, por lo que han sido suprimidos.

En resumen, y para concluir con el estudio de correlaciones lineales, podemos asumir que el número de fichas incluidas en la Lista Roja se encuentra más ligado a la extensión en superficie de la región geográfica en la que se encuentra, y en menor cantidad al número de socios con los que cuenta la asociación en dichas regiones. Por su parte, el número de socios se encuentra ligado a los datos de la población.

## 12.2.2. ESTADO DE CONSERVACIÓN

A continuación, se presentan los resultados del estudio de las correlaciones lineales entre el estado de conservación de un bien patrimonial y los diferentes atributos que representan factores de deterioro. Para realizar este análisis se han utilizado únicamente los datos catalogados como “Lista Roja”, es decir, aquellos que en el momento presente están sometidos a diversos riesgos. Así se evita introducir cierto sesgo resultante de la inclusión de bienes catalogados como “Lista Verde” o “Lista Negra”. Como hemos visto anteriormente, los diversos motivos de salida de la Lista Roja están más vinculados a intervenciones externas que al efecto de factores medioambientales.

Los primeros atributos estudiados han sido las mediciones de los contaminantes NOx, NO2 y SO2 presentes en el aire. Obtenemos por tanto la correlación que existe entre cada uno de ellos y las dos medidas del estado de conservación de los bienes, **Valoración.estado** y **Estado.deterioro**.

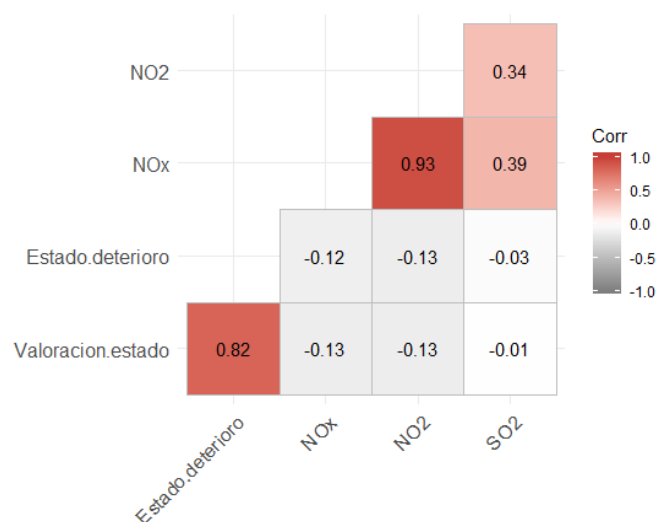


Figura 12.33. Matriz de correlaciones del estado de conservación y la calidad del aire

Se puede observar que las correlaciones entre estos atributos, representadas a la matriz correlaciones de la *figura 12.33*, son prácticamente nulas, lo que se puede visualizar en el elevado grado de dispersión de la nube de puntos resultante de cada análisis reunidas en la *figura 12.34*.

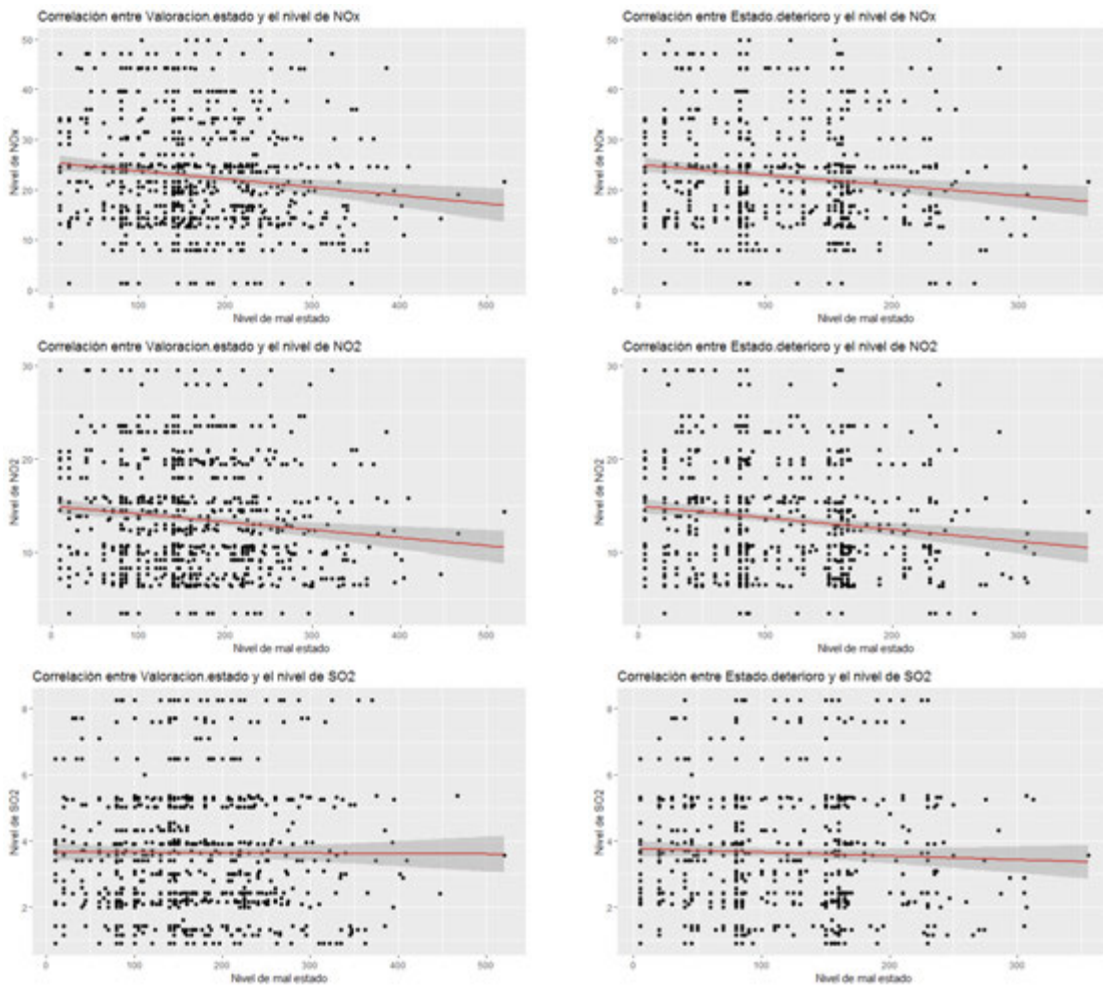


Figura 12.34. Diagramas de dispersión del estado de conservación y la calidad del aire

Puesto que estos resultados no son significativos se ha probado a calcular la correlación entre ambos atributos en el ámbito provincial y autonómico, tomando el valor medio para cada uno de estos atributos según se agrupan para cada división territorial.



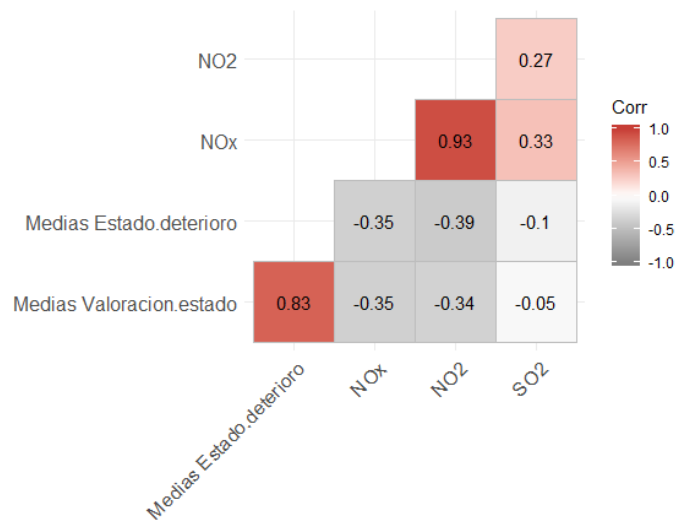


Figura 12.35. Matriz de correlaciones de la media del estado de conservación y la calidad del aire por provincias

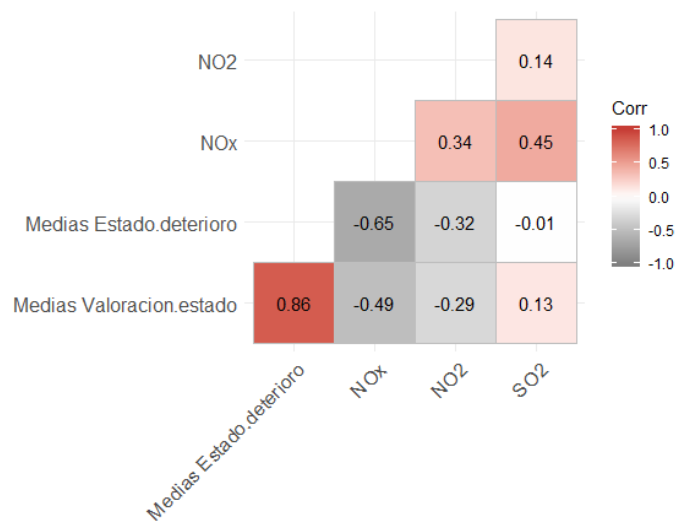


Figura 12.36. Matriz de correlaciones de la media del estado de conservación y la calidad del aire por comunidades autónomas

Aparentemente la correlación establece una asociación moderada entre los atributos, de tal forma que, a mayor concentración de contaminantes en el aire, menor deterioro presentan los bienes, contrariamente a lo que se había previsto y que se antoja más razonable. No obstante, observando los diagramas de dispersión, comprobamos que sigue siendo imposible aproximar estos datos a un modelo lineal. Se puede asumir que los contaminantes no son un factor relevante en el deterioro de los bienes patrimoniales, al menos no como para percibirlo en los datos de los que disponemos.

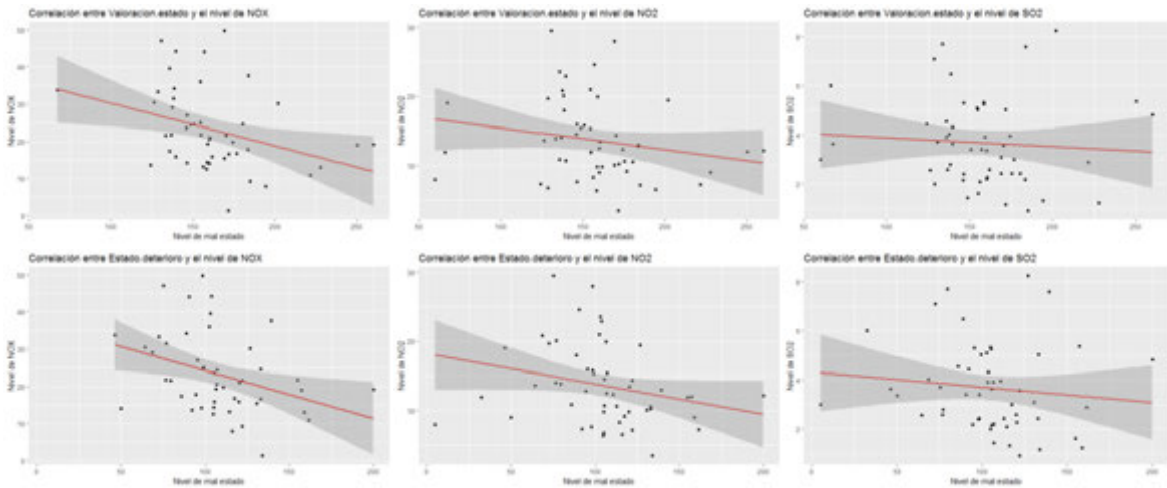


Figura 12.37. Diagramas de dispersión del estado de conservación y la calidad del aire por provincias

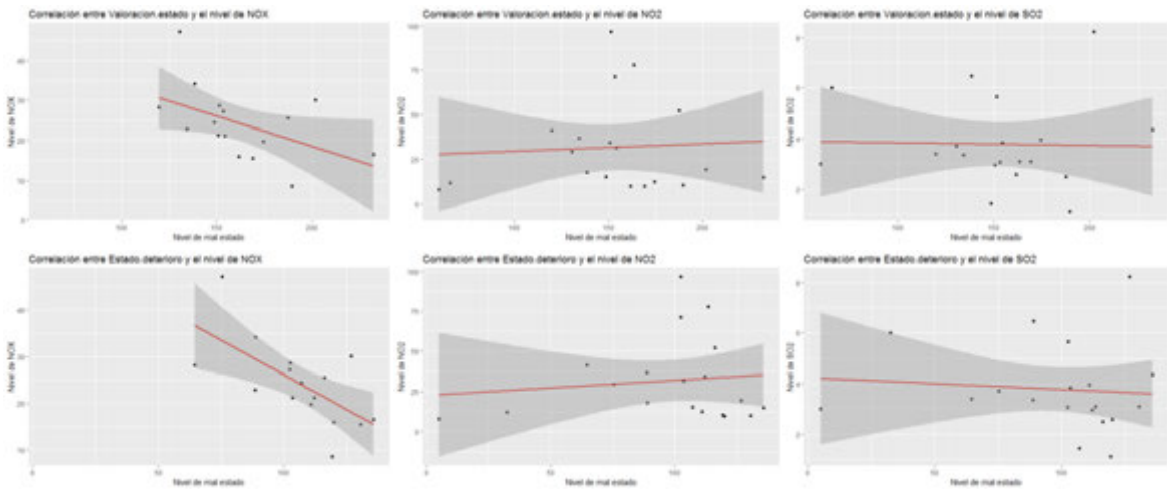


Figura 12.38. Diagramas de dispersión del estado de conservación y la calidad del aire por comunidades autónomas

Finalizado el estudio de los atributos relacionados con la calidad del aire, se pasa a analizar las variables climáticas: temperatura máxima y temperatura mínima, variación máxima de temperatura, probabilidad de precipitación máxima, cantidad de lluvia máxima registrada, humedad relativa máxima y velocidad máxima del viento. Sin embargo, el nivel de asociación entre estos atributos y el estado de conservación es prácticamente nulo, tal y como puede apreciarse en la *figura 12.39* y los correspondientes diagramas de dispersión representados en la *figura 12.40* y *figura 12.41*. Nuevamente se puede concluir que los datos son insuficientes o no son los correctos para realizar este estudio, algo que nuevamente podemos asociar a la falta de precisión de los datos climatológicos con respecto a las localizaciones exactas de los bienes patrimoniales, así como a las decisiones tomadas para la cuantificación del estado de deterioro de los susodichos.

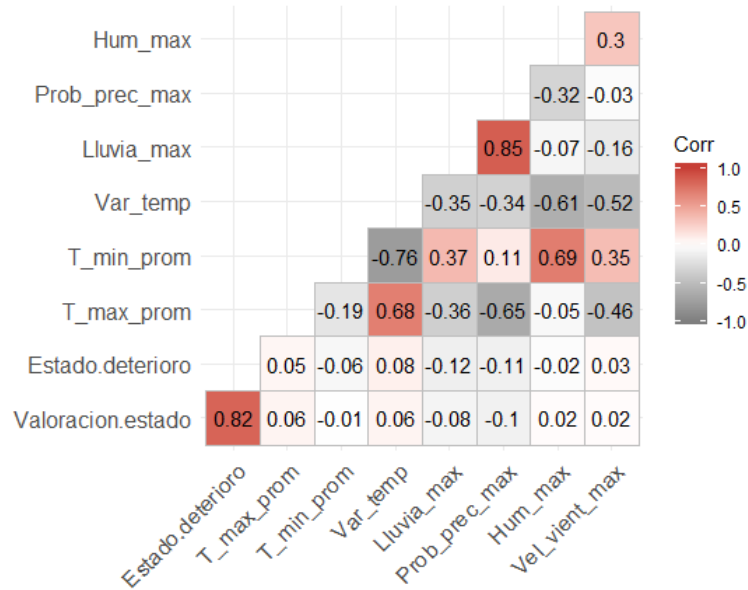


Figura 12.39. Matriz de correlaciones del estado de conservación y las variables climáticas

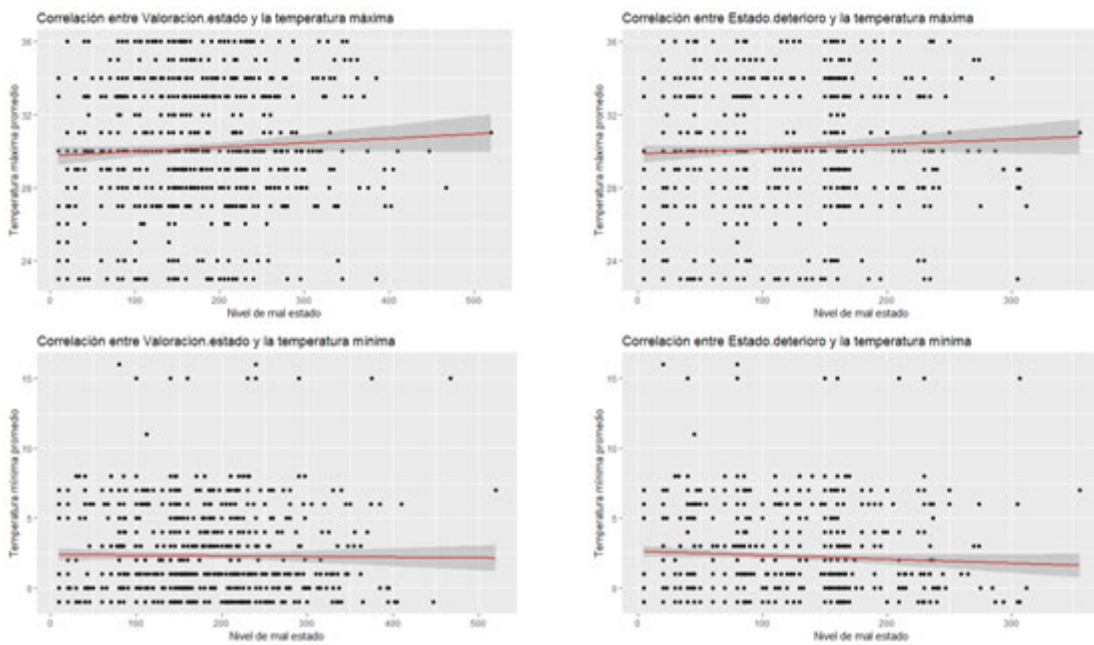


Figura 12.40. Diagramas de dispersión del estado de conservación y las variables climáticas (parte 1)

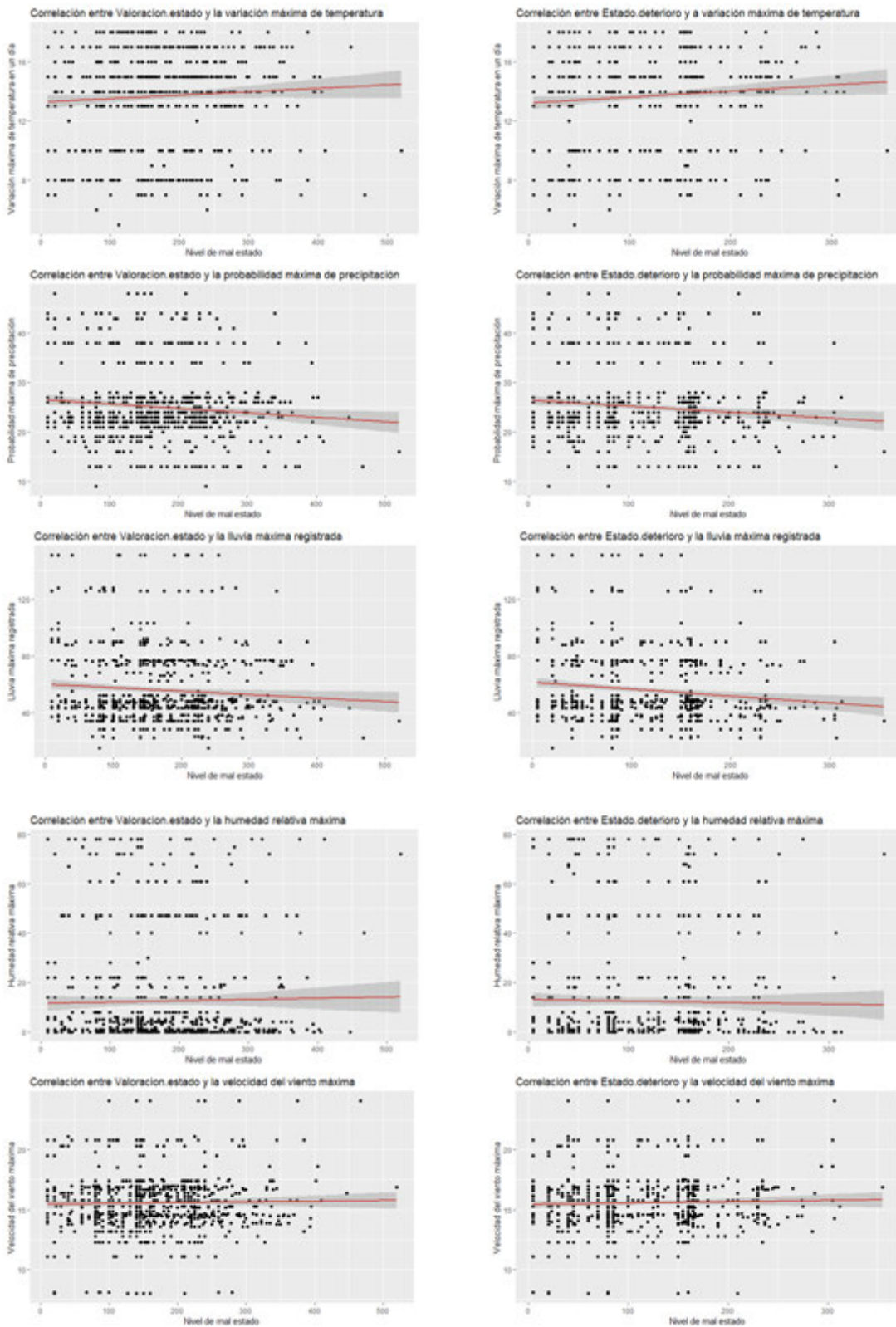


Figura 12.41. Diagramas de dispersión del estado de conservación y las variables climáticas (parte 2)

A la vista del poco éxito de los resultados obtenidos se ha vuelto a repetir el análisis tomando los valores medios por provincias y por comunidades autónomas, tal y como se hizo con los indicadores de la calidad del aire.

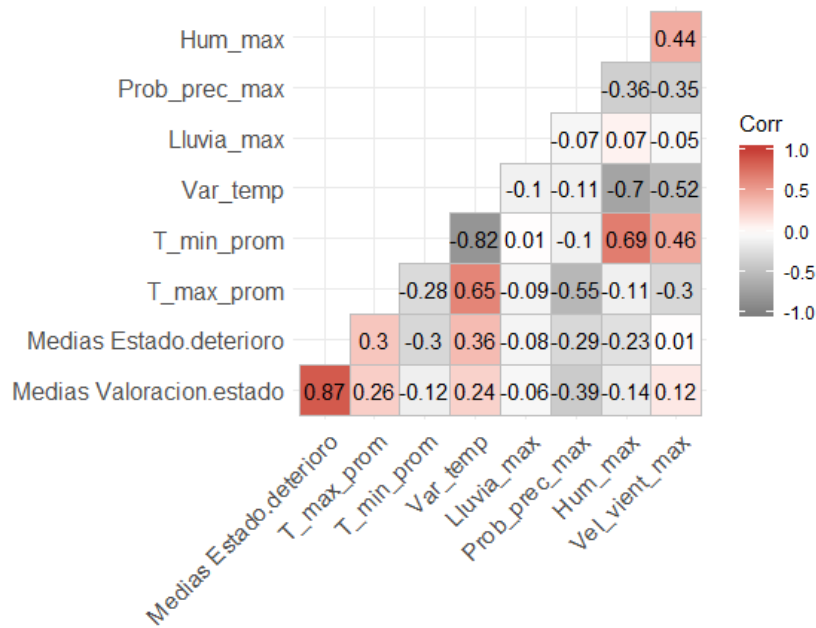


Figura 12.42. Matriz de correlaciones de la media del estado de conservación y las variables climáticas por provincias

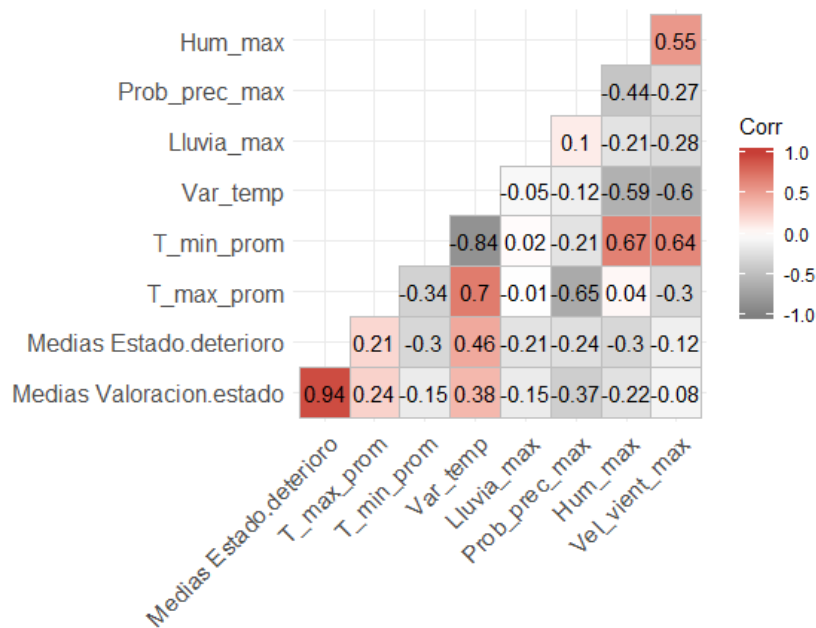


Figura 12.43. Matriz de correlaciones de la media del estado de conservación y las variables climáticas por comunidades autónomas

Los resultados obtenidos en el estudio en ambos ámbitos geográficos son algo más significativos. Destaca la asociación entre la temperatura máxima y su variación máxima, ambas muestran una correlación positiva con el estado de conservación, así como la probabilidad de precipitación y la humedad, que presentan correlación negativa.

Pasamos ahora a analizar la correlación entre el estado de conservación y los atributos relacionados con el nivel socioeconómico, es decir, el PIB per cápita y la tasa de paro. Se puede observar nuevamente que las correlaciones entre estos atributos son prácticamente nulas, lo que se puede visualizar en la matriz de la *figura 12.44* y en el elevado grado de dispersión de la nube de puntos resultante de cada análisis reunidas en la *figura 12.45* y *12.46*.



Figura 12.44. Matriz de correlaciones del estado de conservación y los indicadores socioeconómicos

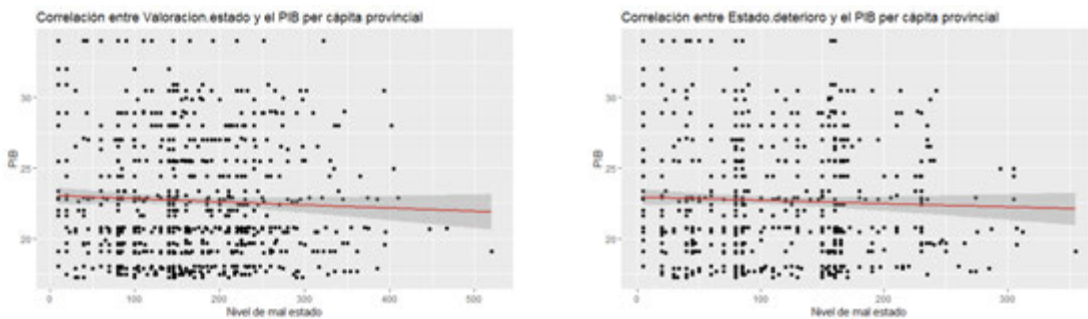


Figura 12.45. Diagramas de dispersión del estado de conservación con respecto al PIB per cápita

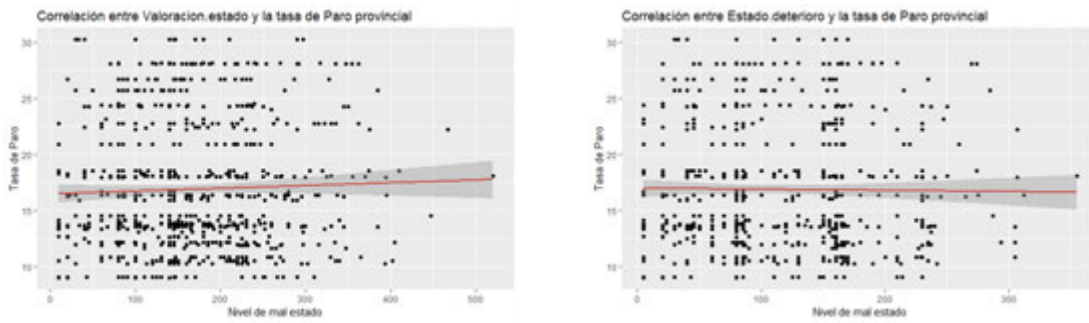


Figura 12.46. Diagramas de dispersión del estado de conservación con respecto a la tasa de paro

De manera idéntica a los análisis previos, se ha repetido el estudio usando los valores medios por provincias y por comunidades autónomas. Los resultados en el ámbito provincial nos indican una leve correlación negativa entre la tasa de paro y el estado de deterioro. Estos resultados se remarcan en el estudio en el ámbito autonómico, alcanzando unos valores de correlación de hasta -0.32.

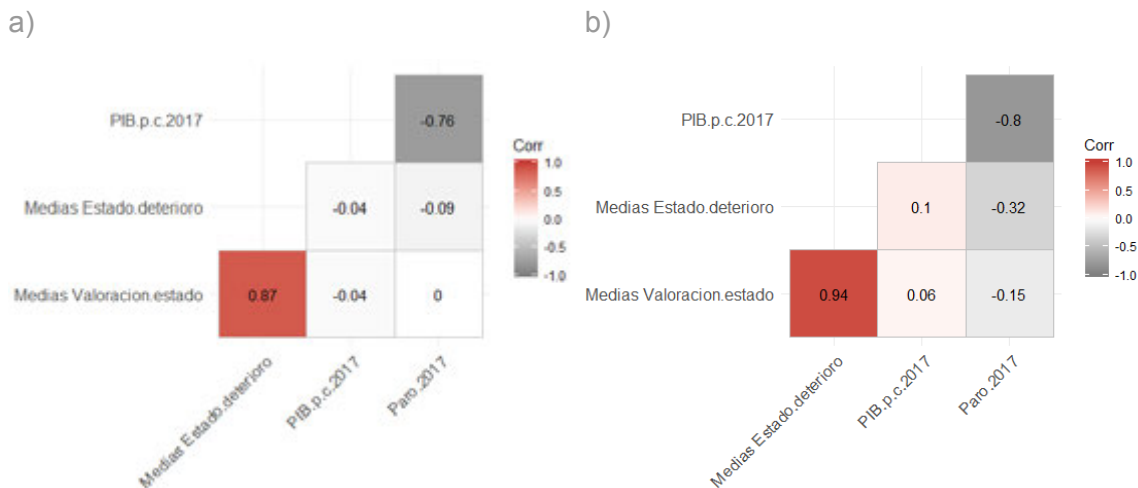


Figura 12.47. Matriz de correlaciones del estado de conservación y los indicadores socioeconómicos por provincias (a) y comunidades autónomas (b)

Finalmente, se ha estudiado si existe algún tipo de correlación entre el número de fichas por provincia que han pasado a la Lista Verde y la tasa de paro o en PIB per cápita de la provincia en la cual se encuentra ubicado. El resultado nos indica que existe una asociación entre ellas, de tal forma que podría confirmarse la hipótesis por la cual, a mayor cantidad de paro, menor número de bienes intervenidos para su recuperación (con una correlación de aproximadamente -0,31) y que a mayor PIB per cápita, mayor número de bienes intervenidos (con una correlación de aproximadamente 0.15).

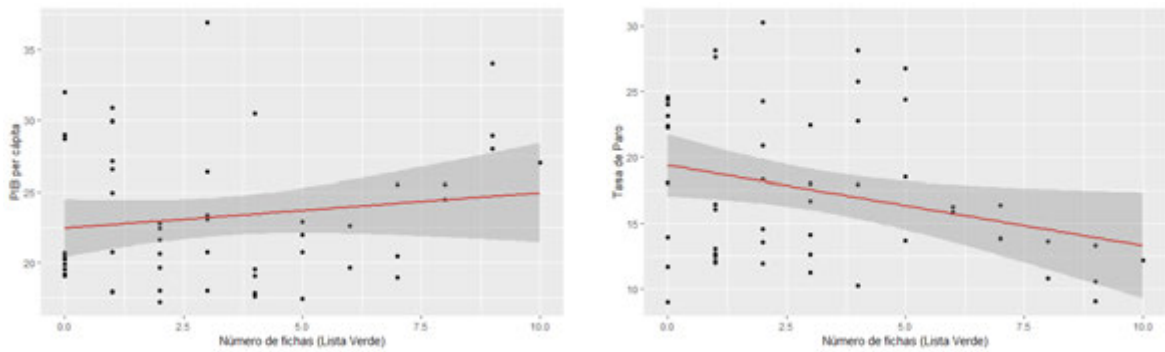


Figura 12.48. Diagramas de dispersión del número de fichas intervenidas con respecto al PIB per cápita y la tasa de paro

En último lugar, para finalizar el análisis de correlaciones lineales entre el estado de deterioro y otros atributos, se ha estudiado la que pueda tener con la época de edificación. Para ello ha sido necesario transformar el siglo de inicio, así como el siglo de finalización de la edificación en valores numéricos y dotarlos de signo en función del atributo **Era**, de tal forma que un siglo anterior a Cristo se referencie como un número negativo. Sin embargo, tal y como podemos observar en la matriz de correlaciones de la figura 12.49, estas son nulas.

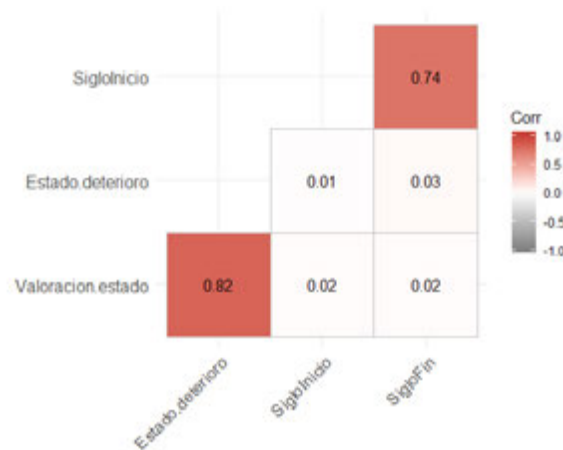


Figura 12.49. Matriz de correlaciones del estado de conservación y la época de edificación

Con esto se concluyen los análisis de correlaciones de los atributos numéricos incluidos para el estudio. Las limitaciones del fichero de datos del que se dispone para realizar el proyecto se han puesto de manifiesto, ya que los resultados obtenidos llegan incluso a contradecir los supuestos por los que se decidió incluirlos en el estudio. No podemos asumir la existencia de relaciones lineales entre la mayoría de los atributos.



### 12.3. CONTRASTE DE HIPÓTESIS

En vista de los resultados obtenidos en el estudio de los atributos numéricos mediante correlaciones lineales, se ha decidido realizar un estudio de atributos categóricos mediante el contraste de hipótesis. Para ello se ha recurrido a la prueba chi cuadrado de Pearson, que permite determinar la independencia entre dos variables. Se trata del método más extendido para el contraste de hipótesis dentro de una población estadística.

La formulación de hipótesis se realiza de la siguiente manera:

H0 (Hipótesis Nula) Los parámetros son independientes.

Se acepta H0 cuando  $X^2_{calculado} < X^2_{crítico}$ .

H1 (Hipótesis Alternativa) Los parámetros no son independientes.

Se acepta H1 cuando  $X^2_{calculado} > X^2_{crítico}$ .

Se va a recurrir también al análisis de los residuos de Pearson, los cuales recogen la desviación existente entre el valor observado y el valor esperado. Cuanto más cercana a cero es la desviación, significa que ambas distribuciones son más parecidas y por tanto, se confirma la Hipótesis Nula. Cuanto más lejos de cero se encuentre esta desviación, mayor es la probabilidad de que no sean independientes ambas variables y nos indica de qué forma se relacionan entre sí. Esto nos va a permitir determinar de qué forma se relacionan unas con otras en caso de no ser independientes.

Para realizar este estudio ha sido imprescindible la transformación de ciertos atributos numéricos en atributos nominales. Los más importantes son los que representan la valoración del estado de conservación de los bienes patrimoniales. Trataremos de agruparlos en cinco categorías según su estado. La primera y la última categoría serán la “Verde” y la “Negra” y dependen exclusivamente de que estén o no en dichas categorías de la Lista Roja. Las tres categorías intermedias consistirán en tres niveles según el grado de mal estado. Estos niveles son “Amarillo”, “Naranja” y “Rojo”. Los intervalos numéricos comprendidos en estos niveles han sido fijados tras el análisis de la distribución de los datos de cada uno de los dos atributos de estado. Estos nuevos atributos reciben los nombres de **Valoración.estado.nom** y **Estado.deterioro.nom** respectivamente. Los puntos de corte se han elegido con base en la distribución de los datos representado en la *figura 12.13* anteriormente vista. Esta figura mostraba máximos absolutos y relativos cercanos a los valores 90 y 160 para ambos atributos, y se ha decidido tomar estos puntos para dividir los datos en las tres categorías según su estado de conservación. Esto se ha hecho así debido a que, al ser una escala de medición de poca precisión, al basarnos en la distribución de los datos, permite que no haya grupos que se queden prácticamente vacíos debido a la elevada dispersión de los datos, o la existencia de *outliers*.

Otra posible alternativa de enfocar esto sería “fuzzificar” las variables, es decir, trabajar con variables borrosas. En las variables borrosas no existen estados absolutos de verdad, sino grados de verdad. Así pues, por poner un ejemplo, una ficha podría no ser “Roja”, sino “10% Amarilla”, “30% Naranja” y “60% Roja”.

De forma similar se han tratado los atributos referentes al PIB per cápita y el Paro. En este caso, los puntos de corte se han establecido dividiendo el rango en tres partes iguales ya que son datos obtenidos de una serie de mediciones estadísticas y económicas.

A continuación, se exponen en forma de tablas las pruebas chi cuadrado realizadas y los resultados obtenidos para cada una de ellas.

Identificador del test	TEST_01	TEST_02
<b>Variables</b>	Siglo.Inicio Valoración.estado	Siglo.Inicio Estado.deterioro
<b>X<sup>2</sup> calculado</b>	124.99	100.29
<b>Nivel de significancia (<math>\alpha</math>)</b>	0.05	0.05
<b>Grados de libertad</b>	88	88
<b>p-valor</b>	0.005	0.1746
<b>X<sup>2</sup> crítico</b>	110.898	110.898
<b>Hipótesis aceptada</b>	H1	H0

Tabla 12.17. Tests Chi Cuadrado de Pearson TEST\_01 y TEST\_02

Identificador del test	TEST_03	TEST_04
<b>Variables</b>	Siglo.Fin Valoración.estado	Siglo.Fin Estado.deterioro
<b>X<sup>2</sup> calculado</b>	126.94	94.7
<b>Nivel de significancia (<math>\alpha</math>)</b>	0.05	0.05
<b>Grados de libertad</b>	84	84
<b>p-valor</b>	0.0017	0.1994
<b>X<sup>2</sup> crítico</b>	106.3948	106.3948
<b>Hipótesis aceptada</b>	H1	H0

Tabla 12.18. Tests Chi Cuadrado de Pearson TEST\_03 y TEST\_04

Analizando los residuos para **Valoración.estado** se observa que los bienes terminados de edificar en el siglo II poseen una mayor probabilidad de tener estados catalogados como “Amarillo”, XIII y XXII “Verde”, y XV “Naranja”. No parece seguir una lógica determinada detrás de dichas probabilidades, esto puede ser debido a un sesgo elevado en la muestra de datos.

<b>Identificador del test</b>	TEST_05	TEST_06
<b>Variables</b>	Siglo.Inicio Riesgo.estructural	Siglo.Fin Riesgo.estructural
<b>X<sup>2</sup> calculado</b>	72.143	81.069
<b>Nivel de significancia (α)</b>	0.05	0.05
<b>Grados de libertad</b>	22	21
<b>p-valor</b>	0.0000	0.0000
<b>X<sup>2</sup> crítico</b>	33.9244	32.6706
<b>Hipótesis aceptada</b>	H1	H1

Tabla 12.19. Tests Chi Cuadrado de Pearson TEST\_05 y TEST\_06

Estudiando el residuo de Pearson se puede afirmar que los bienes de los siglos XIX y XX probablemente no presentan un riesgo estructural, mientras que los bienes del siglo XVII si tienen mayor probabilidad de presentarlo. Aunque no se puede determinar la probabilidad de los restantes siglos, se puede deducir que los bienes que no presentan este tipo de riesgo son los de más reciente edificación frente a los más antiguos.

<b>Identificador del test</b>	TEST_07	TEST_08
<b>Variables</b>	Siglo.Inicio Riesgo.deterioro	Siglo.Fin Riesgo.deterioro
<b>X<sup>2</sup> calculado</b>	31.193	29.723
<b>Nivel de significancia (α)</b>	0.05	0.05
<b>Grados de libertad</b>	22	21
<b>p-valor</b>	0.0922	0.0977
<b>X<sup>2</sup> crítico</b>	33.9244	32.6706
<b>Hipótesis aceptada</b>	H0	H0

Tabla 12.20. Tests Chi Cuadrado de Pearson TEST\_07 y TEST\_08

<b>Identificador del test</b>	TEST_09	TEST_10
<b>VARIABLES</b>	Siglo.Inicio Riesgo.de.ruina	Siglo.Fin Riesgo.de.ruina
<b>X<sup>2</sup> calculado</b>	27.186	26.602
<b>Nivel de significancia (<math>\alpha</math>)</b>	0.05	0.05
<b>Grados de libertad</b>	22	21
<b>p-valor</b>	0.2042	0.1844
<b>X<sup>2</sup> crítico</b>	33.9244	32.6706
<b>Hipótesis aceptada</b>	H0	H0

Tabla 12.21. Tests Chi Cuadrado de Pearson TEST\_09 y TEST\_10

<b>Identificador del test</b>	TEST_11	TEST_12
<b>VARIABLES</b>	Grado.de.protección Valoración.estado	Grado.de.protección Estado.deterioro
<b>X<sup>2</sup> calculado</b>	84.543	81.683
<b>Nivel de significancia (<math>\alpha</math>)</b>	0.05	0.05
<b>Grados de libertad</b>	44	44
<b>p-valor</b>	0.0002	0.0005
<b>X<sup>2</sup> crítico</b>	60.4809	60.4809
<b>Hipótesis aceptada</b>	H1	H1

Tabla 12.22. Tests Chi Cuadrado de Pearson TEST\_11 y TEST\_12

Estudiando el residuo de Pearson se puede afirmar que los BIC, con mayor probabilidad, se encuentran en la categoría “Verde” y no se encuentran en la categoría “Rojo”, así como los bienes sin protección específica no se encuentran en la categoría “Verde” con mayor probabilidad.

<b>Identificador del test</b>	TEST_13	TEST_14
<b>Variabes</b>	Propietario Valoración.estado	Propietario Estado.deterioro
<b>X<sup>2</sup> calculado</b>	438.46	445.51
<b>Nivel de significancia (<math>\alpha</math>)</b>	0.05	0.05
<b>Grados de libertad</b>	140	140
<b>p-valor</b>	0.0000	0.0000
<b>X<sup>2</sup> crítico</b>	168.613	168.613
<b>Hipótesis aceptada</b>	H1	H1

Tabla 12.23. Tests Chi Cuadrado de Pearson TEST\_13 y TEST\_14

<b>Identificador del test</b>	TEST_15	TEST_16
<b>Variabes</b>	Propiedad Valoración.estado	Propiedad Estado.deterioro
<b>X<sup>2</sup> calculado</b>	44.602	47.969
<b>Nivel de significancia (<math>\alpha</math>)</b>	0.05	0.05
<b>Grados de libertad</b>	20	20
<b>p-valor</b>	0.0012	0.0004
<b>X<sup>2</sup> crítico</b>	31.4104	31.4104
<b>Hipótesis aceptada</b>	H1	H1

Tabla 12.24. Tests Chi Cuadrado de Pearson TEST\_15 y TEST\_16

Estudiando el residuo de Pearson se comprueba que los bienes que son propiedad privada tienen una mayor probabilidad de no encontrarse en la categoría “Verde”, mientras que, por el contrario, los bienes de propiedad pública tienen mayor probabilidad de encontrarse en dicha categoría. No obstante, los bienes compartidos entre entidades privadas y fundaciones (desconociendo la clase titularidad de la fundación), sí presentan mayor probabilidad de incluirse en la categoría “Verde”.

<b>Identificador del test</b>	TEST_17	TEST_18
<b>Variabes</b>	Otras.intervenciones Valoración.estado	Otras.intervenciones Estado.deterioro
<b>X<sup>2</sup> calculado</b>	61.126	57.535
<b>Nivel de significancia (<math>\alpha</math>)</b>	0.05	0.05
<b>Grados de libertad</b>	4	4
<b>p-valor</b>	0.0000	0.0000
<b>X<sup>2</sup> crítico</b>	9.4877	9.4877
<b>Hipótesis aceptada</b>	H1	H1

Tabla 12.25. Tests Chi Cuadrado de Pearson TEST\_17 y TEST\_18

Estudiando el residuo de Pearson se confirma que los bienes intervenidos tienden con mayor probabilidad a estar en la categoría “Rojo”. Se debe recordar que este atributo recoge las intervenciones que no han supuesto la salida de un bien fuera de la Lista Roja.

<b>Identificador del test</b>	TEST_19	TEST_20
<b>Variabes</b>	Paro.nom Valoración.estado	Paro.nom Estado.deterioro
<b>X<sup>2</sup> calculado</b>	19.188	18.811
<b>Nivel de significancia (<math>\alpha</math>)</b>	0.05	0.05
<b>Grados de libertad</b>	12	12
<b>p-valor</b>	0.0841	0.0841
<b>X<sup>2</sup> crítico</b>	21.0261	21.0261
<b>Hipótesis aceptada</b>	H0	H0

Tabla 12.28. Tests Chi Cuadrado de Pearson TEST\_19 y TEST\_20

<b>Identificador del test</b>	TEST_21	TEST_22
<b>VARIABLES</b>	Riesgo.demolición Valoración.estado	Riesgo.demolición Estado.deterioro
<b>X<sup>2</sup> calculado</b>	101.04	84.874
<b>Nivel de significancia (α)</b>	0.05	0.05
<b>Grados de libertad</b>	4	4
<b>p-valor</b>	0.0000	0.0000
<b>X<sup>2</sup> crítico</b>	9.4877	9.4877
<b>Hipótesis aceptada</b>	H1	H1

Tabla 12.26. Tests Chi Cuadrado de Pearson TEST\_21 y TEST\_22

En este caso los residuos de Pearson solo nos confirman que la probabilidad de encontrarse en la categoría “Negro” es mayor en los bienes que tienen o han tenido riesgo de demolición.

<b>Identificador del test</b>	TEST_23	TEST_24
<b>VARIABLES</b>	PIB.p.c.nom Valoración.estado	PIB.p.c.nom Estado.deterioro
<b>X<sup>2</sup> calculado</b>	23.466	28.372
<b>Nivel de significancia (α)</b>	0.05	0.05
<b>Grados de libertad</b>	12	12
<b>p-valor</b>	0.02402	0.0049
<b>X<sup>2</sup> crítico</b>	21.0260	21.0260
<b>Hipótesis aceptada</b>	H1	H1

Tabla 12.27. Tests Chi Cuadrado de Pearson TEST\_21 y TEST\_22

El residuo de Pearson indica que los bienes situados en zonas con un PIB per cápita elevado tienen una mayor probabilidad de encontrarse en la categoría “Verde”.

<b>Identificador del test</b>	TEST_25	TEST_26
<b>Variabes</b>	Cercano.a.la.costa Valoración.estado	Cercano.a.la.costa Estado.deterioro
<b>X<sup>2</sup> calculado</b>	12.18	13.208
<b>Nivel de significancia (α)</b>	0.05	0.05
<b>Grados de libertad</b>	8	8
<b>p-valor</b>	0.1434	0.1049
<b>X<sup>2</sup> crítico</b>	15.5073	15.5073
<b>Hipótesis aceptada</b>	H0	H0

Tabla 12.29. Tests Chi Cuadrado de Pearson TEST\_25 y TEST\_26

<b>Identificador del test</b>	TEST_27	TEST_28
<b>Variabes</b>	PeligrosidadRelativa Valoración.estado	PeligrosidadRelativa Estado.deterioro
<b>X<sup>2</sup> calculado</b>	28.446	26.908
<b>Nivel de significancia (α)</b>	0.05	0.05
<b>Grados de libertad</b>	16	16
<b>p-valor</b>	0.0279	0.0425
<b>X<sup>2</sup> crítico</b>	26.2962	26.2962
<b>Hipótesis aceptada</b>	H1	H1

Tabla 12.30. Tests Chi Cuadrado de Pearson TEST\_27 y TEST\_28

El residuo de Pearson indica que los bienes situados en zonas de baja presencia sísmica tienen mayor probabilidad de no estar en la categoría “Rojo” y tienen mayor probabilidad de encontrarse en la categoría “Verde”.



<b>Identificador del test</b>	TEST_29	TEST_30
<b>Variabes</b>	Propietario Estado.abandono	Propiedad Estado.abandono
<b>X<sup>2</sup> calculado</b>	36.532	7.1161
<b>Nivel de significancia (<math>\alpha</math>)</b>	0.05	0.05
<b>Grados de libertad</b>	35	5
<b>p-valor</b>	0.3974	0.2121
<b>X<sup>2</sup> crítico</b>	49.8018	11.0705
<b>Hipótesis aceptada</b>	H0	H0

Tabla 12.31. Tests Chi Cuadrado de Pearson TEST\_29 y TEST\_30

<b>Identificador del test</b>	TEST_31	TEST_32
<b>Variabes</b>	Propietario Riesgo.de.abandono	Propiedad Riesgo.de.abandono
<b>X<sup>2</sup> calculado</b>	33.809	3.6059
<b>Nivel de significancia (<math>\alpha</math>)</b>	0.05	0.05
<b>Grados de libertad</b>	35	5
<b>p-valor</b>	0.5255	0.6074
<b>X<sup>2</sup> crítico</b>	49.8018	11.0705
<b>Hipótesis aceptada</b>	H0	H0

Tabla 12.32. Tests Chi Cuadrado de Pearson TEST\_31 y TEST\_32

<b>Identificador del test</b>	TEST_33	TEST_34
<b>Variabes</b>	Propietario Motivo.de.la.retirada	Propiedad Motivo.de.la.retirada
<b>X<sup>2</sup> calculado</b>	291.53	30.636
<b>Nivel de significancia (α)</b>	0.05	0.05
<b>Grados de libertad</b>	200	40
<b>p-valor</b>	0.0000	0.8567
<b>X<sup>2</sup> crítico</b>	233.9943	55.7585
<b>Hipótesis aceptada</b>	H1	H0

Tabla 12.33. Tests Chi Cuadrado de Pearson TEST\_33 y TEST\_34

El residuo de Pearson indica que un bien perteneciente a una diputación tiene mayor probabilidad de ser intervenido para su paso a la Lista Verde y a su vez que un bien perteneciente al Ministerio del Interior tiene más probabilidad de ser demolido. Estos resultados seguramente se deban a que existe una excesiva cantidad de categorías dentro del atributo **Propietario** y que cada una de estas tenga pocos datos. Por tanto no es posible generalizar con este atributo.

<b>Identificador del test</b>	TEST_35	TEST_36
<b>Variabes</b>	Propietario Grado.de.protección	Propiedad Grado.de.protección
<b>X<sup>2</sup> calculado</b>	404.89	98.752
<b>Nivel de significancia (α)</b>	0.05	0.05
<b>Grados de libertad</b>	385	55
<b>p-valor</b>	0.2331	0.0002
<b>X<sup>2</sup> crítico</b>	431.7513	73.3115
<b>Hipótesis aceptada</b>	H0	H1

Tabla 12.34. Tests Chi Cuadrado de Pearson TEST\_35 y TEST\_36

Analizando el residuo de Pearson se comprueba que un bien público tiene mayor probabilidad de estar catalogado como BIC, así como tiene una mayor probabilidad de no englobarse como bien “sin protección” o “protección desconocida”. Por otro lado, un bien privado presenta una mayor probabilidad de carecer de algún tipo de protección específica.

<b>Identificador del test</b>	TEST_37
<b>Variables</b>	Grado.de.protección Motivo.de.la.retirada
<b>X<sup>2</sup> calculado</b>	46.655
<b>Nivel de significancia (α)</b>	0.05
<b>Grados de libertad</b>	90
<b>p-valor</b>	1
<b>X<sup>2</sup> crítico</b>	113.1453
<b>Hipótesis aceptada</b>	H0

Tabla 12.35. Test Chi Cuadrado de Pearson TEST\_37

Finalmente se han llevado a cabo los test de hipótesis en busca de relación entre el estado de conservación y los atributos binarios referentes a acontecimientos importantes que haya podido experimentar un bien patrimonial a lo largo de su historia. Recordemos que estos atributos son **Francesada**, **GuerraCivil**, **Desamortizaciones**, **Terremotos**, **Inundaciones** e **Incendios**. Los resultados han mostrado todos p-valores superiores al 0.05, por lo que se acepta la Hipótesis Nula y, por tanto, las variables son independientes.

No conformes con estos resultados, y ante la posibilidad de que el atributo **Historia** contuviese información incorrectamente tratada, que asociaremos a la ausencia de información en el fichero de datos, se han llevado a cabo dos últimos test de hipótesis. Trataremos de relacionar las variables **Francesada** y **Desamortizaciones** con **Estado.expolio**. Puesto que la desamortización implica el expolio de bienes de la Iglesia, ambos atributos deberían mostrarse relacionados, de lo contrario, habrá que investigar si hay información errónea en el fichero de datos. Observando los resultados, podemos concluir que, efectivamente, existe una relación entre los bienes que han sido expoliados y los que han experimentado una desamortización.

<b>Identificador del test</b>	TEST_38	TEST_39
<b>Variabes</b>	Francesada Estado.expolio	Desamortizaciones Estado.expolio
<b>X<sup>2</sup> calculado</b>	0.4623	4.2523
<b>Nivel de significancia (<math>\alpha</math>)</b>	0.05	0.05
<b>Grados de libertad</b>	1	1
<b>p-valor</b>	0.4965	0.0392
<b>X<sup>2</sup> crítico</b>	3.8414	3.8414
<b>Hipótesis aceptada</b>	H0	H1

Tabla 12.36. Tests Chi Cuadrado de Pearson TEST\_38 y TEST\_39

## 13. ANÁLISIS DE LOS RESULTADOS

Aunque los resultados del análisis no han sido los deseados a causa de las diversas limitaciones identificadas en la muestra de datos, se ha logrado sacar algunas ideas en claro, las cuales se exponen a continuación.

Gracias al análisis de correlaciones se ha podido determinar que el número de fichas incluidas en el fichero de datos es dependiente de la extensión geográfica de la región en la que se hallan y, en menor medida, del número de socios que tiene Hispania Nostra en dicha región. Por su parte, el número de socios es dependiente de la población de la zona. Esto nos puede hacer suponer que existe mayor patrimonio en riesgo de desaparición del que no tenemos conocimiento puesto que no se encuentra en la Lista Roja en aquellas regiones donde menor es el número de socios de la entidad. Por otra parte, se ha comprobado que el paro se encuentra levemente relacionado con el estado de conservación, de tal forma que aquellas regiones con una mayor tasa de paro tienden a presentar bienes en peor estado de conservación. No obstante, el número de fichas no se ve relacionado con esto, lo que hace cuestionarse si se trata de una casualidad de la muestra.

Posteriormente las correlaciones nos han desvelado que los bienes situados en zonas con un mayor PIB per cápita tienden a encontrarse en la Lista Verde. Esto indica que efectivamente existe una relación entre los factores socioeconómicos de una región y lo descuidado o no que puedan tener su patrimonio. También se ha podido comprobar que el estado de conservación de un bien patrimonial es dependiente de la temperatura máxima, la variación máxima de temperatura, la probabilidad de precipitación y la humedad, de tal forma que un clima más seco, caluroso o con cambios bruscos de temperatura contribuye al deterioro de los bienes patrimoniales.

En cuanto a la titularidad, los bienes patrimoniales de propiedad pública tienen mayor probabilidad de encontrarse en la Lista Verde que los privados, y tienen también mayor probabilidad de poseer la etiqueta de BIC, lo que confirma que tienen mayor probabilidad de ser "Verde". Por su parte, los bienes de propiedad privada tienen mayor probabilidad de carecer de algún tipo de protección. Por tanto, los bienes que carecen de protección tienen mayor probabilidad de no encontrarse en la Lista Verde.

En resumidas cuentas, se puede afirmar que los resultados obtenidos están por debajo de las expectativas iniciales. Se ha podido comprobar que muchas variables que se sabe por medio de otros estudios que tienen un efecto crucial en el deterioro y la conservación del patrimonio inmueble, han dado resultados completamente opuestos en este estudio. No obstante, el análisis no representa el total de los objetivos del proyecto. Una parte de vital importancia sigue siendo el tratamiento de los datos. En el siguiente apartado se expone un análisis del funcionamiento del sistema de extracción y tratamiento de los datos.

## 14. ANÁLISIS DE LA SOLUCIÓN

El sistema de tratamiento de datos se puede asemejar a una criba: un utensilio que se usa para separar el grano de otros sólidos no deseados, como pueden ser la paja o la tierra. Si la rejilla no es lo suficientemente estrecha, el grano pasará a través de ella y se perderá junto con los sólidos más deseados.

En nuestro sistema podríamos comparar la rejilla con los diferentes scripts o “funciones” de tratamiento de cada atributo. A mayor número de scripts, más estrecha será la rejilla. Y, cuanto más estrecha es la rejilla que conforma esta criba, más seguridad hay de que no van a perderse los granos, pero, por otro lado, es más probable que también permanezcan aquellos sólidos no deseados de tamaño más grande que no han logrado atravesar la rejilla. Así pues, llega un punto en que hay ciertos datos cuya estructura es tan variante y compleja de tratar que resultaría más sencillo hacerlo a mano uno a uno en vez de implementar un código capaz de extraerlos de forma correcta sin llevarse con ello información no deseada. Basándose en esta idea, y con el objetivo de obtener los resultados más precisos dentro de lo posible, se ha estudiado cómo varía la cantidad de información útil extraída basándose en las funcionalidades implementadas por cada atributo. Debemos tener en cuenta que a la par que aumentan los scripts, aumentan las líneas de código y lo hacen también los tiempos de ejecución y el número de horas dedicadas a la implementación. Esto es algo que vamos a estudiar en los diferentes atributos tratados.

En este análisis se entienden por datos “tratados incorrectamente” aquellos datos que requieren de cierto tratamiento específico y que, hasta que el sistema no dispuso de una versión mejorada capaz de detectarlos, fueron indebidamente tratados. Por ejemplo, años no convertidos previamente en siglos, siglos dentro de intervalos, etc. Por ello, el número de datos incorrectamente tratados es siempre el valor mínimo, ya que podrían existir más que no hubiesen logrado detectarse.

### **ATRIBUTO “ÉPOCA”**

De las 1195 fichas totales se tiene conocimiento de que 13 de ellas están vacías, la mayoría debido a que pertenecen a elementos del tipo “Fauna y flora”, por lo que carecen de una fecha de origen. Por ello, el total de datos tratables es de 1182 datos.

#### **Caso 1:**

Scripts: - Detección de expresiones regulares de siglos.

Datos recuperados: 1046 ( $\approx$  88% de los datos totales)

Datos tratados incorrectamente: 303 ( $\approx$  28% de los datos recuperados)

Datos reales recuperados: 743 ( $\approx$  63% de los datos totales)

### **Caso 2:**

- Scripts:
- Conversión de años a siglos
  - Detección de expresiones regulares de siglos.

Datos recuperados: 1101 ( $\simeq$  93% de los datos totales)

Datos tratados incorrectamente: 314 ( $\simeq$  28% de los datos recuperados)

Datos reales recuperados: 787 ( $\simeq$  66% de los datos totales)

### **Caso 3:**

- Scripts:
- Tratamiento de intervalos
  - Conversión de años a siglos
  - Tratamiento de reformas
  - Detección de expresiones regulares de siglos.

Datos recuperados: 762 ( $\simeq$  64% de los datos totales)

Datos tratados incorrectamente: 2 ( $\simeq$  0,3% de los datos recuperados)

Datos reales recuperados: 760 ( $\simeq$  94% de los datos totales)

### **Caso 4:**

- Scripts:
- Tratamiento de intervalos
  - Conversión de años a siglos
  - Tratamiento de reformas
  - Detección de expresiones regulares de siglos.
  - Tratamiento de periodos históricos

Datos recuperados: 1096 ( $\simeq$  93% de los datos totales)

Datos tratados incorrectamente: 4 ( $\simeq$  0,3% de los datos recuperados)

Datos reales recuperados: 1092 ( $\simeq$  92% de los datos totales)

### **Caso 5:**

- Scripts:
- Tratamiento de intervalos
  - Conversión de años a siglos
  - Tratamiento de reformas
  - Detección de expresiones regulares de siglos.
  - Tratamiento de periodos históricos
  - Extracción de siglos de edificación del atributo Historia

Datos recuperados:  $\leq$  1162 ( $\simeq$  98% de los datos totales)

Podemos observar como un significativo aumento en la complejidad del algoritmo de tratamiento mediante el desarrollo de múltiples scripts se traduce en un incremento progresivamente menor del número de datos obtenidos. No obstante, también se traduce en una reducción más precisa de los datos, en ciertas ocasiones muy significativa, como en el Caso 3.

### **ATRIBUTO “MOTIVO.DE.LA.RETIRADA”**

De las 1195 fichas totales únicamente 179 han salido de la Lista Roja, ya sea para formar parte de la Lista Verde, la Lista Negra, así como las fichas reincidentes que han salido de la Lista Roja y posteriormente han vuelto a entrar en ella, por lo que son las únicas que contienen información en este atributo.

#### **Caso 1:**

Scripts: - Detección de términos.

Datos recuperados: 176 ( $\approx$  98% de los datos totales)

Datos tratados incorrectamente: 50 ( $\approx$  28% de los datos recuperados)

Datos reales recuperados: 126 ( $\approx$  70% de los datos totales)

#### **Caso 2:**

Scripts: - Inclusión de información procedente de “Intervención”.

- Detección de términos.

Datos recuperados: 178 ( $\approx$  99% de los datos totales)

### **ATRIBUTO “PROPIEDAD”**

De las 1195 fichas totales 151 se encuentran vacías, por lo que trabajaremos con un total de 1044 datos para realizar esta evaluación.

#### **Caso 1:**

Scripts: - Detección de términos.

Datos recuperados: 960 ( $\approx$  92% de los datos totales)

Datos tratados incorrectamente: 3 ( $\approx$  0.3% de los datos recuperados)

Datos reales recuperados: 957 ( $\approx$  92% de los datos totales)



### **Caso 2:**

Scripts:       - Tratamiento de expresiones “Vendido / Comprado / Cedido”.  
                  - Detección de términos.

Datos recuperados: 961 ( $\approx$  92% de los datos totales)

Datos tratados incorrectamente: 95 ( $\approx$  10% de los datos recuperados)

Datos reales recuperados: 866 ( $\approx$  83% de los datos totales)

### **Caso 3:**

Scripts:       - Tratamiento de expresiones “Vendido / Comprado / Cedido”.  
                  - Tratamiento de acrónimos.  
                  - Detección de términos.

Datos recuperados: 976 ( $\approx$  93% de los datos totales)

## **ATRIBUTO “GRADO.DE.PROTECCIÓN”**

De las 1195 fichas totales, 92 se encuentran vacías, por lo que trabajaremos con un total de 1103 datos para realizar esta evaluación.

### **Caso 1:**

Scripts:       - Detección de términos (versión 1).

Datos recuperados: 862 ( $\approx$  78% de los datos totales)

Datos tratados incorrectamente: 152 ( $\approx$  17% de los datos recuperados)

Datos reales recuperados: 710 ( $\approx$  64% de los datos totales)

### **Caso 2:**

Scripts:       - Tratamiento de datos vacíos.  
                  - Detección de términos (versión 2).

Datos recuperados: 1108 ( $\approx$  92% de los datos totales)

La principal ventaja que presenta esta solución es que permite la inclusión de nuevos métodos de búsqueda y limpieza de datos no contemplados previamente sin la necesidad de realizar cambios complejos en la estructura de los algoritmos diseñados para el tratamiento.

Por otro lado, presenta una serie de desventajas. La primera es que requiere de un aumento considerable de recursos (tiempo, esfuerzo, etc.) para recuperar cada vez una cantidad de información menor. Además, la inclusión de nuevos datos no garantiza que estos porcentajes se mantengan, ya que pueden aparecer nuevos términos en los atributos que no hayan aparecido previamente y que pueden no haber sido considerados.

Una solución alternativa para resolver este proceso de extracción de la información sería la implementación de una inteligencia artificial especializada en text summarization (resumen de textos) y que fuese capaz de extraer la información esencial de cada atributo. Existen ciertas bibliotecas especializadas para minería de datos que incluyen este tipo de inteligencia artificial, no obstante, no todas pueden analizar textos en castellano, y las más recomendadas por sus resultados en ocasiones han resultado ser de pago.

## 15. CONCLUSIONES

Los objetivos inicialmente fijados se han cumplidos de forma parcial. Por un lado, el tratamiento del fichero de datos y la transformación de su contenido para obtener la información esencial para su análisis ha sido bastante satisfactorio. Se han logrado corregir muchos aspectos mejorables del fichero de datos original y categorizar bastante información con considerable éxito. Sin embargo, la parte de cuantificación, pese a lograrse transformar descripciones tan dispares en valores numéricos siguiendo unos criterios preestablecidos, ha sido la menos satisfactoria del proceso en vista a los resultados obtenidos. A priori no existe una forma precisa y fiable de cuantizar el estado de un bien patrimonial, al menos partiendo de descripciones redactadas por los usuarios. Cada uno usa su propia terminología para expresarse y la percepción subjetiva suele quedar reflejada en la redacción del texto.

Por otro lado, los resultados nos muestran que no es posible realizar un análisis exhaustivo con los datos disponibles. Una parte considerable de los resultados obtenidos niega la existencia de relaciones entre el estado de conservación de los bienes patrimoniales y factores externos, cuando sabemos por otros estudios previos que sí son determinantes. Esto, tal y como se ha podido entrever durante las diferentes pruebas efectuadas, es posible que se deba esencialmente a tres razones. La primera es la citada en el párrafo anterior, que hace referencia al método de cuantificación del estado de deterioro, siendo este uno de los factores clave de un bien patrimonial. La segunda razón es la falta de precisión en la obtención de datos externos al fusionarlos con las fichas de los bienes patrimoniales. Tomar los datos climáticos o de la calidad del aire recogidos en una capital de provincia y generalizarlos para toda la extensión de dicha provincia, acarrea una alta imprecisión que conlleva la alteración de los resultados. La tercera razón es la escasez de datos y la heterogeneidad de la muestra. El fichero de datos es de 993 filas por 72 columnas. De esas fichas, en torno a un 17% forma parte de la Lista Verde y menos de un 1% está incluido en la Lista Negra. La muestra también presenta poca homogeneidad a nivel geográfico. Todo ello ha introducido un error de muestreo elevado

Es necesario hacer especial hincapié en la necesidad de digitalizar las bases de datos del patrimonio. No solo porque sea una de las principales formas de preservarlo y difundirlo, sino porque además permitiría realizar estudios con mayor profundidad. La labor de Hispania Nostra con la Lista Roja resulta admirable y necesaria, sin embargo, existe todavía mucho patrimonio olvidado que debe ser dado a conocer. Este estudio pone de manifiesto la importancia de establecer una sistematización en la digitalización del patrimonio; la implantación de un tesoro que permita controlar la ambigüedad de los términos empleados en estudios y descripciones; y la parametrización de factores debidamente consensuados que permitan determinar el estado de conservación de los bienes, similar (salvando las distancias) a las escalas del riesgo de extinción de especies animales o vegetales.

## 15.1. POSIBLES MEJORAS Y PLANES DE ACTUACIÓN FUTUROS

Como hemos ido viendo a lo largo del tratamiento de los datos, encontramos un considerable abanico de posibilidades. Por un lado, vamos a discernir estas posibles líneas en mejoras del sistema desarrollado durante este proyecto, y por otro lado posibles ampliaciones futuras que puedan suponer una mejora para los fines en este campo de investigación.

### Mejoras del sistema de tratamiento y análisis de datos

Inclusión de coordenadas geográficas en el análisis. Como ya hemos visto a lo largo del análisis, disponer de las coordenadas geográficas de todas las fichas de la Lista Roja ofrecería un conocimiento más preciso de muchos parámetros dependientes de la localización del bien patrimonial, tales como las condiciones climatológicas o la proximidad con respecto al mar.

El empleo de trigramas en lugar de bigramas, en las diferentes ocasiones en que se ha recurrido a ellos durante el tratamiento de los datos, permitiría conocer mejor el contexto en el cual aparecen los términos clave del tratamiento de cada atributo y, por consiguiente, la detección y extracción de información más precisa con estructuras lingüísticas más complejas.

Redes neuronales artificiales que permitan análisis de predicción y clasificación. Para esto es necesario disponer de una muestra de datos considerablemente mayor a la que se tiene en este momento. Para obtener resultados, el número de fichas de cada una de las tres clases debería ser más elevado. Puesto que la Lista Negra actualmente cuenta con menos de 10 fichas, el empleo de redes neuronales para la clasificación o la predicción es inviable.

### Otras posibles líneas de investigación futuras:

Análisis fotográfico para la valoración del estado de conservación. Esta idea, surgida por parte de miembros de la junta directiva de Hispania Nostra, propone el uso de fuentes documentales fotográficas para determinar el grado de deterioro sufrido en un bien patrimonial a lo largo de los años. Haciendo una comparación entre ambas imágenes mediante la implementación de un sistema de inteligencia artificial capaz de ello. Esto, a priori, implicaría una serie de requisitos técnicos muy exigentes, como la calidad de las imágenes a analizar.

Implementación de un sistema experto. Los sistemas expertos emplean un motor de inferencia estadística que les permite simular el razonamiento humano. Este sistema sería capaz, por ejemplo, de determinar qué nuevas fichas deben ser incluidas en la Lista Roja en función de los parámetros contenidos en la información aportada al enviarse la ficha.

Diseño de un modelo relacional para la transformación del fichero de datos de la Lista Roja de Patrimonio en una base de datos, lo que aportaría un almacenamiento más eficiente y un acceso más rápido a los datos.

Ampliación de las aplicaciones de este modelo de estudio al patrimonio mueble. Aunque este trabajo se ha centrado en el patrimonio inmueble, la mayoría de los parámetros que se están evaluando son aplicables al estudio del patrimonio mueble.

## 16. REFERENCIAS

- [1] UNESCO, «UNESCO World Heritage Centre», [En línea]. Available: <http://whc.unesco.org/es/list/?iso=es&search=&> . [Último acceso: 2020 marzo 4].
- [2] Ministerio de Cultura y Deporte, «Nota resumen. Empleo Cultural 2T2019-1T2020», Ministerio de Cultura y Deporte, 2020.
- [3] Ministerio de Cultura y Deporte, «Anuario de Estadísticas Culturales 2019», Ministerio de Cultura y Deporte, 2019.
- [4] J. Ruskin, *The Seven Lamps of Architecture*, Reino Unido: Smith, Elder & Co., 1984.
- [5] G. M. Piantadosi, «El 'big data' entra en el Reina Sofía», *El Independiente*, 26 octubre 17. [En línea]. Available: <https://www.elindependiente.com/tendencias/2017/10/25/big-data-arte-reina-sofia/>. [Último acceso: 1 marzo 2020].
- [6] J. G. Fernández, «Hispania Nostra», 20 junio 2017. [En línea]. Available: <https://www.hispanianostra.org/wp-content/uploads/2017/09/Estatutos-HN-2017.pdf>.
- [7] Hispania Nostra, «¿Qué es la Lista Roja del Patrimonio?», Hispania Nostra, [En línea]. Available: <https://listarojapatrimonio.org/quienes-somos/>. [Último acceso: 6 junio 2019].
- [8] GantaBI, «Qué es el Data Wrangling Parte I», 12 6 2019. [En línea]. Available: <https://www.gantabi.com/2019/06/12/que-es-el-data-wrangling-parte-i/>. [Último acceso: 8 enero 2020].
- [9] Sacyr, «'INROAD', el innovador sistema de mantenimiento predictivo para la gestión por indicadores en carreteras», Sacyr, 10 5 2019. [En línea]. Available: <https://canalinnovacion.sacyr.com/-/inroad-elinnovador-sistema-de-mantenimiento-predictivo-para-la-gestion-por-indicadores-encarreteras?redirect=%2Fcat%2F-%2Fcategories%2F25445976> . [Último acceso: 28 8 2019].
- [10] Fundación Santa María La Real, «MHS», [En línea]. Available: <http://www.mhsproject.com/es/inicio/>. [Último acceso: 2019 julio 12].
- [11] Comisión Europea, «¿Qué es el programa SUDOE?», [En línea]. Available: <https://interreg-sudoe.eu/inicio>. [Último acceso: 14 enero 2020].
- [12] I. Argerich Fernández, T. Díaz Fraile, L. García Hernández y A. Muñoz Cosme, «El inventario del Patrimonio Arquitectónico. 1978-1982», Instituto del Patrimonio Cultural de España.
- [13] Y. Dankov, *Digital Cultural Heritage and the Analytics as Innovative Tools for Processing Big Data*, Sofía: Universidad San Clemente de Ohrid de Sofía, 2017.
- [14] Padres de la Constitución, *Constitución Española*, Vols. %1 de %2Capítulo tercero. De las Comunidades Autónomas, Artículo 149, 1978.

- [15] Padres de la Constitución, *Constitución Española*, Vols. %1 de %2Capítulo tercero. De las Comunidades Autónomas, Artículo 148, 1978.
- [16] *Boletín Oficial del Estado*, Agencia Estatal Boletín Oficial del Estado, 1985.
- [17] *Boletín Oficial del Estado*, Agencia Estatal Boletín Oficial del Estado., 1985.
- [18] Patrimonio Nacional, «Conoce nuestra historia», [En línea]. Available: <https://www.patrimonionacional.es/sobre-patrimonio/conoce-nuestra-historia>. [Último acceso: 4 julio 2019].
- [19] J. M. Rodríguez Domingo, *La gestión del Patrimonio Eclesiástico en el ámbito local*, Sevilla: Universidad Internacional de Andalucía, 2009.
- [20] Secretaría de Estado de Presupuestos y Gasto, «Programa 337B. Conservación y restauración de bienes culturales», 2019. [En línea]. Available: [http://www.congreso.es/docu/pge2019/pge\\_2019-web/PGE-ROM/doc/1/3/23/3/2/15/N\\_19\\_A\\_R\\_31\\_124\\_1\\_2\\_3\\_1337B\\_C\\_1.PDF](http://www.congreso.es/docu/pge2019/pge_2019-web/PGE-ROM/doc/1/3/23/3/2/15/N_19_A_R_31_124_1_2_3_1337B_C_1.PDF).
- [21] Secretaría de Estado de Presupuestos y Gasto, «Programa 337C. Protección del Patrimonio Histórico», 2019. [En línea]. Available: [http://www.congreso.es/docu/pge2019/pge\\_2019-web/PGE-ROM/doc/1/3/23/3/2/16/N\\_19\\_A\\_R\\_31\\_124\\_1\\_2\\_3\\_1337C\\_C\\_1.PDF](http://www.congreso.es/docu/pge2019/pge_2019-web/PGE-ROM/doc/1/3/23/3/2/16/N_19_A_R_31_124_1_2_3_1337C_C_1.PDF).
- [22] Ministerio de Hacienda, «Ministerio de Hacienda. Financiación Autonómica», 4 julio 2019. [En línea]. Available: <https://www.hacienda.gob.es/es-ES/Areas%20Tematicas/Financiacion%20Autonomica/Paginas/Financiacion%20Autonomica.aspx>.
- [23] *Boletín Oficial del Estado*, Agencia Estatal Boletín Oficial del Estado, 2004.
- [24] Comisión Europea, «Protección del patrimonio cultural», [En línea]. Available: [https://ec.europa.eu/culture/policy/culture-policies/cultural-heritage\\_es](https://ec.europa.eu/culture/policy/culture-policies/cultural-heritage_es). [Último acceso: 4 julio 2019].
- [25] «Anexo: Períodos de la historia de España», [En línea]. Available: [https://es.wikipedia.org/wiki/Anexo:Per%C3%ADodos\\_de\\_la\\_historia\\_de\\_Espa%C3%B1a](https://es.wikipedia.org/wiki/Anexo:Per%C3%ADodos_de_la_historia_de_Espa%C3%B1a). [Último acceso: 13 octubre 2019].
- [26] Ministerio de Cultura y Deporte, «Régimen general de protección del Patrimonio Histórico», [En línea]. Available: <http://www.culturaydeporte.gob.es/cultura/patrimonio/bienes-culturales-protegidos/niveles-de-proteccion/regimen-general.html>. [Último acceso: 5 julio 2019].
- [27] *Boletín Oficial del Estado*, Agencia Estatal Boletín Oficial del Estado, 1985.
- [28] *Boletín Oficial del Estado*, Agencia Estatal Boletín Oficial del Estado, 1949.
- [29] *Boletín Oficial del Estado*, Agencia Estatal Boletín Oficial del Estado, 1978.
- [30] J. A. Herráez, G. Enríquez de Salamanca, M. J. Pastor Arenas y T. Gil Muñoz, «Manual de seguimiento y análisis de condiciones ambientales. Plan Nacional de Conservación Preventiva», Ministerio de Educación, Cultura y Deporte, 2014.

- [31] Instituto Nacional de Estadística, «P.I.B. a precios de mercado y valor añadido bruto a precios básicos por ramas de actividad: Precios corrientes por provincias y periodo», 2017. [En línea]. Available: <https://www.ine.es/jaxi/Datos.htm?path=/t35/p010/rev19/10/&file=02001.px#!tabs-tabla>.
- [32] Instituto Nacional de Estadística, «Tasas de actividad, paro y empleo por provincia y sexo», 2017. [En línea]. Available: <https://www.ine.es/jaxiT3/Datos.htm?t=3996#!tabs-tabla>.
- [33] B. Hermosín Campos, *Efectos de la contaminación atmosférica sobre el patrimonio histórico: Deposición de compuestos orgánicos y formación de costras negras sulfatadas*, Sevilla: Universidad de Sevilla. Facultad de Química., 1995.
- [34] Instituto Nacional de Estadística, «Estadística de Calidad del Aire», Ministerio para la Transición Ecológica y el Reto Demográfico, 2018. [En línea]. Available: <https://www.ine.es/dyngs/IOE/es/operacion.htm?id=1259931124792>. [Último acceso: 3 febrero 2020].
- [35] Instituto Geológico y Minero de España, «ZESIS - Zonación Sismogénica de Iberia», Ministerio de Ciencia, Innovación y Universidades, [En línea]. Available: <http://info.igme.es/zesis/>. [Último acceso: 6 febrero 2020].
- [36] Alsimet, «¿Cuáles son las causas de la corrosión en los materiales metálicos?», 27 noviembre 2019. [En línea]. [Último acceso: 7 marzo 2020].
- [37] E. Alario Catalá, «Con la Cámara a Cuestas... En la Playa. La importancia del mantenimiento de edificios», 29 diciembre 2014. [En línea]. Available: <https://enriquealario.com/con-la-camara-a-cuestas-en-la-playa/>. [Último acceso: 2 marzo 2020].
- [38] B. Chico, E. Otero, L. Mariaca y M. Morcillo, «La corrosión en atmósferas marinas. Efecto de la distancia a la costa», *Revista de Metalurgia*, vol. 34, nº Extra, pp. 71-74, mayo 1998.
- [39] N. S. Corbacho, «Viaje a través del territorio con más castillos de Europa», *El Confidencial*, 2 abril 2018.



## 17. ANEXO A: MEDIDAS DESCRIPTIVAS E HISTOGRAMAS DEL ESTADO DE CONSERVACIÓN A NIVEL AUTONÓMICO

GALICIA		
Atributo	Valoración.estado	Estado.deterioro
<b>Rango</b>	10-340	5-230
<b>Media</b>	140.0233	91.8837
<b>Desviación estándar</b>	81.0121	61.3387
<b>Moda</b>	10	80
<b>Cuartiles</b>	<b>Q1 (25%)</b>	82.5
	<b>Q2 (50%)</b>	20
	<b>Q3 (75%)</b>	208.5
		143

Tabla A.1. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Galicia.

PRINCIPADO DE ASTURIAS		
Atributo	Valoración.estado	Estado.deterioro
<b>Rango</b>	10-240	5-160
<b>Media</b>	139.3913	86.8261
<b>Desviación estándar</b>	71.6778	46.4481
<b>Moda</b>	160	80
<b>Cuartiles</b>	<b>Q1 (25%)</b>	91.5
	<b>Q2 (50%)</b>	160
	<b>Q3 (75%)</b>	190
		130

Tabla A.2. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en el Principado de Asturias.

CANTABRIA		
Atributo	Valoración.estado	Estado.deterioro
Rango	10-385	5-305
Media	142.6452	102.5806
Desviación estándar	96.6635	72.9889
Moda	110	80
Cuartiles	Q1 (25%)	82.5
	Q2 (50%)	110
	Q3 (75%)	192.5

Tabla A.3. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Cantabria.

PAÍS VASCO		
Atributo	Valoración.estado	Estado.deterioro
Rango	10-257	5-160
Media	113.5	63.0833
Desviación estándar	81.4008	55.9179
Moda	10	5
Cuartiles	Q1 (25%)	20
	Q2 (50%)	142.5
	Q3 (75%)	157.5

Tabla A.4. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Cantabria.

<b>COMUNIDAD FORAL DE NAVARRA</b>			
<b>Atributo</b>		<b>Valoración.estado</b>	<b>Estado.deterioro</b>
<b>Rango</b>		30-393	30-242
<b>Media</b>		179.7727	113.7727
<b>Desviación estándar</b>		81.9029	62.3560
<b>Moda</b>		160	80
<b>Cuartiles</b>	<b>Q1 (25%)</b>	140	80
	<b>Q2 (50%)</b>	160	82.5
	<b>Q3 (75%)</b>	223.75	165

Tabla A.5. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en la C.F. de Navarra.

<b>ARAGÓN</b>			
<b>Atributo</b>		<b>Valoración.estado</b>	<b>Estado.deterioro</b>
<b>Rango</b>		10-405	5-305
<b>Media</b>		154.0845	116.0986
<b>Desviación estándar</b>		81.2011	64.3004
<b>Moda</b>		80	80
<b>Cuartiles</b>	<b>Q1 (25%)</b>	85	80
	<b>Q2 (50%)</b>	145	85
	<b>Q3 (75%)</b>	203.5	160

Tabla A.6. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en País Vasco.

CATALUÑA			
Atributo		Valoración.estado	Estado.deterioro
Rango		103-375	23-237
Media		200.4706	124.8824
Desviación estándar		68.3237	62.1207
Moda		155	155
Cuartiles	Q1 (25%)	157	80
	Q2 (50%)	177	155
	Q3 (75%)	260	157

Tabla A.7. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Cataluña.

CASTILLA Y LEÓN			
Atributo		Valoración.estado	Estado.deterioro
Rango		10-447	5-312
Media		169.5252	118.4664
Desviación estándar		79.6879	63.9858
Moda		80	80
Cuartiles	Q1 (25%)	106.25	80
	Q2 (50%)	160	86
	Q3 (75%)	221.5	160

Tabla A.8. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Castilla y León.

LA RIOJA			
Atributo		Valoración.estado	Estado.deterioro
Rango		30-315	30-237
Media		170.7143	134.1429
Desviación estándar		86.8911	59.5907
Moda		100	80
Cuartiles	Q1 (25%)	102.5	85
	Q2 (50%)	160	132.5
	Q3 (75%)	236.5	165.25

Tabla A.9. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en La Rioja.

COMUNIDAD DE MADRID			
Atributo		Valoración.estado	Estado.deterioro
Rango		10-322	5-160
Media		121.8889	68.3333
Desviación estándar		82.9074	43.1127
Moda		80	80
Cuartiles	Q1 (25%)	65	40.75
	Q2 (50%)	100	80
	Q3 (75%)	160	85

Tabla A.10. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en la Comunidad de Madrid.

CASTILLA-LA MANCHA			
Atributo		Valoración.estado	Estado.deterioro
Rango		10-290	5-260
Media		153.7763	102.5
Desviación estándar		70.5441	59.4435
Moda		80	80
Cuartiles	Q1 (25%)	100	70
	Q2 (50%)	151	85
	Q3 (75%)	217.75	155

Tabla A.11. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Castilla-La Mancha.

COMUNIDAD VALENCIANA			
Atributo		Valoración.estado	Estado.deterioro
Rango		10-520	5-355
Media		158.1111	105.2444
Desviación estándar		110.0905	75.6574
Moda		80	80
Cuartiles	Q1 (25%)	80	50
	Q2 (50%)	120	80
	Q3 (75%)	215	150

Tabla A.12. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en la Comunidad Valenciana.

<b>EXTREMADURA</b>			
<b>Atributo</b>		<b>Valoración.estado</b>	<b>Estado.deterioro</b>
<b>Rango</b>		10-362	5-274
<b>Media</b>		198.2857	121.3214
<b>Desviación estándar</b>		87.7085	62.4823
<b>Moda</b>		145	80
<b>Cuartiles</b>	<b>Q1 (25%)</b>	143.75	80
	<b>Q2 (50%)</b>	180	94
	<b>Q3 (75%)</b>	256	160

Tabla A.13. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Extremadura.

<b>ANDALUCÍA</b>			
<b>Atributo</b>		<b>Valoración.estado</b>	<b>Estado.deterioro</b>
<b>Rango</b>		20-385	20-285
<b>Media</b>		154.4495	105.7523
<b>Desviación estándar</b>		76.8857	57.5311
<b>Moda</b>		140	80
<b>Cuartiles</b>	<b>Q1 (25%)</b>	100	80
	<b>Q2 (50%)</b>	145	85
	<b>Q3 (75%)</b>	200	153

Tabla A.14. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Andalucía.

REGIÓN DE MURCIA			
Atributo		Valoración.estado	Estado.deterioro
Rango		80-370	40-230
Media		206.6207	129.7241
Desviación estándar		76.5625	50.9978
Moda		265	80
Cuartiles	Q1 (25%)	150	85
	Q2 (50%)	210	130
	Q3 (75%)	265	160

Tabla A.15. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en la Región de Murcia.

ISLAS CANARIAS			
Atributo		Valoración.estado	Estado.deterioro
Rango		80-467	20-307
Media		232.2	135.7
Desviación estándar		121.7217	135.7
Moda		240	80
Cuartiles	Q1 (25%)	145	80
	Q2 (50%)	235	115
	Q3 (75%)	277.5	197.5

Tabla A.16. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en las Islas Canarias.



ISLAS BALEARES			
Atributo		Valoración.estado	Estado.deterioro
Rango		40-225	40-160
Media		132.5	100
Desviación estándar		130.8148	84.8528
Moda		40	40
Cuartiles	Q1 (25%)	86.25	70
	Q2 (50%)	132.5	100
	Q3 (75%)	178.75	130

Tabla A.17. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en las Islas Baleares.

CIUDAD AUTÓNOMA DE CEUTA		
Atributo	Valoración.estado	Estado.deterioro
Valor	112	45

Tabla A.18. Valores del estado del único bien incluido en la Lista Roja en Ceuta.

*Nota: No existe tabla de la Ciudad Autónoma de Melilla porque carece de fichas en la Lista Roja actualmente*

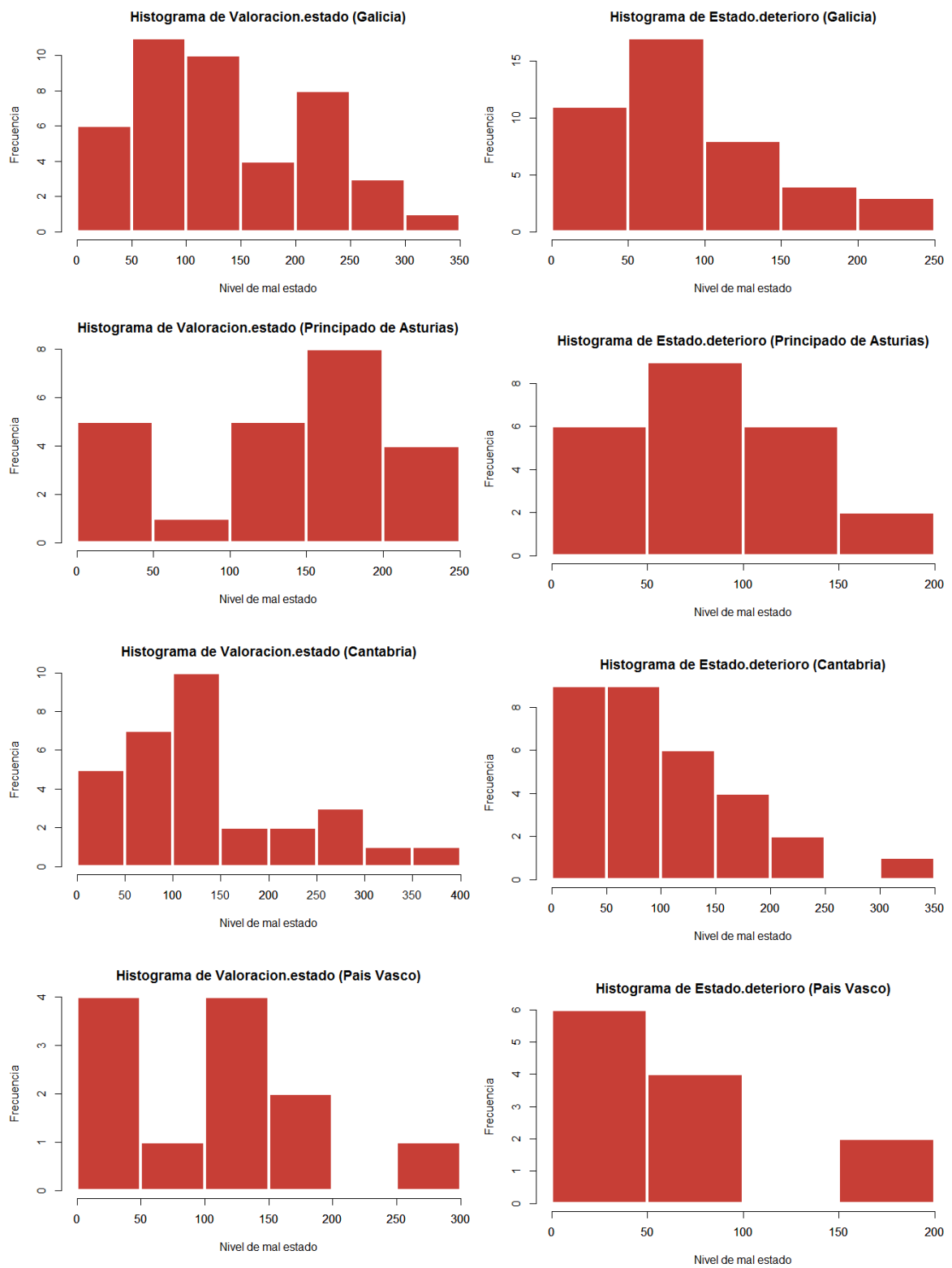


Figura A.1. Histogramas del estado por comunidades autónomas (parte 1)

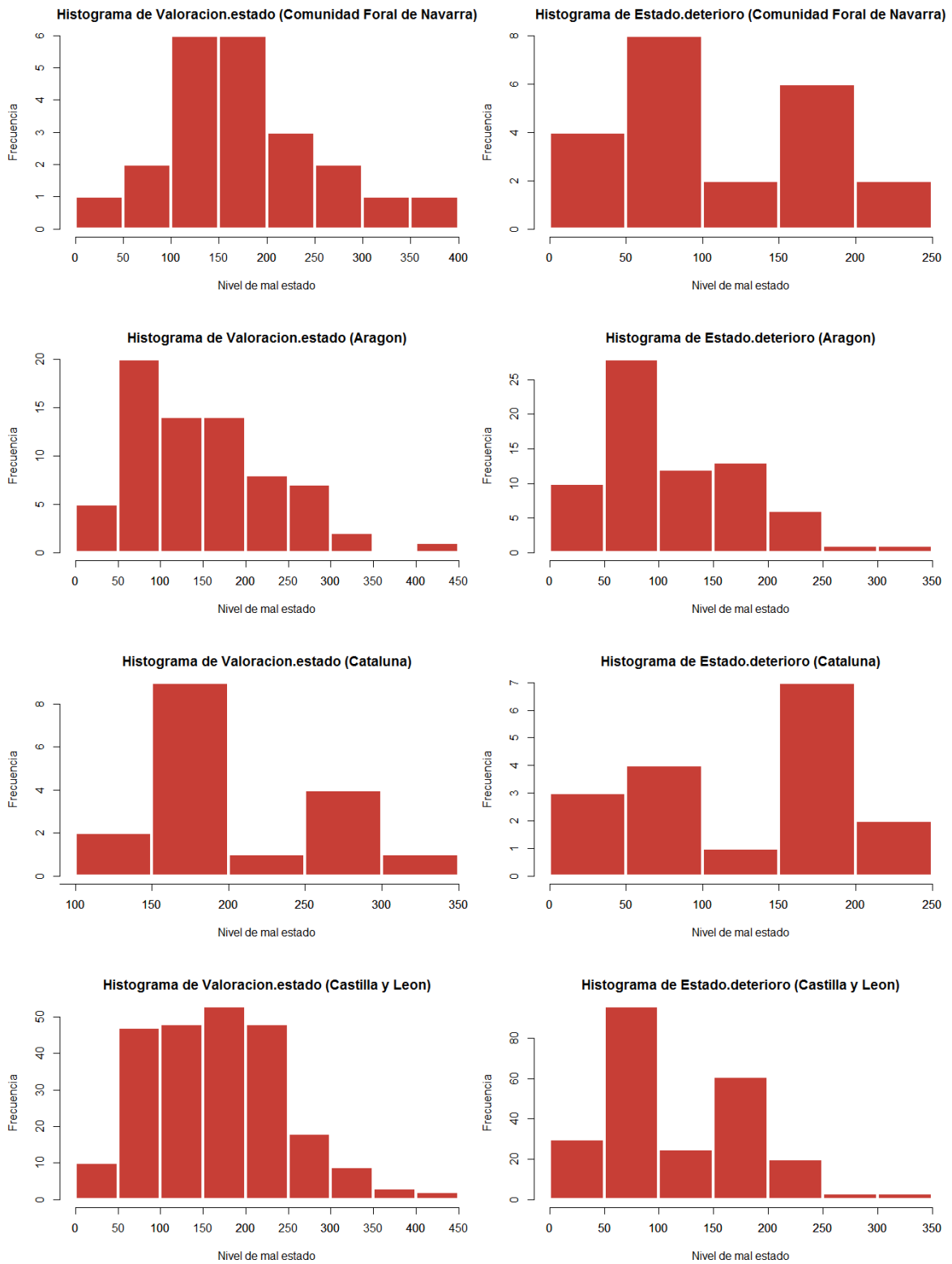


Figura A.2. Histogramas del estado por comunidades autónomas (parte 2)

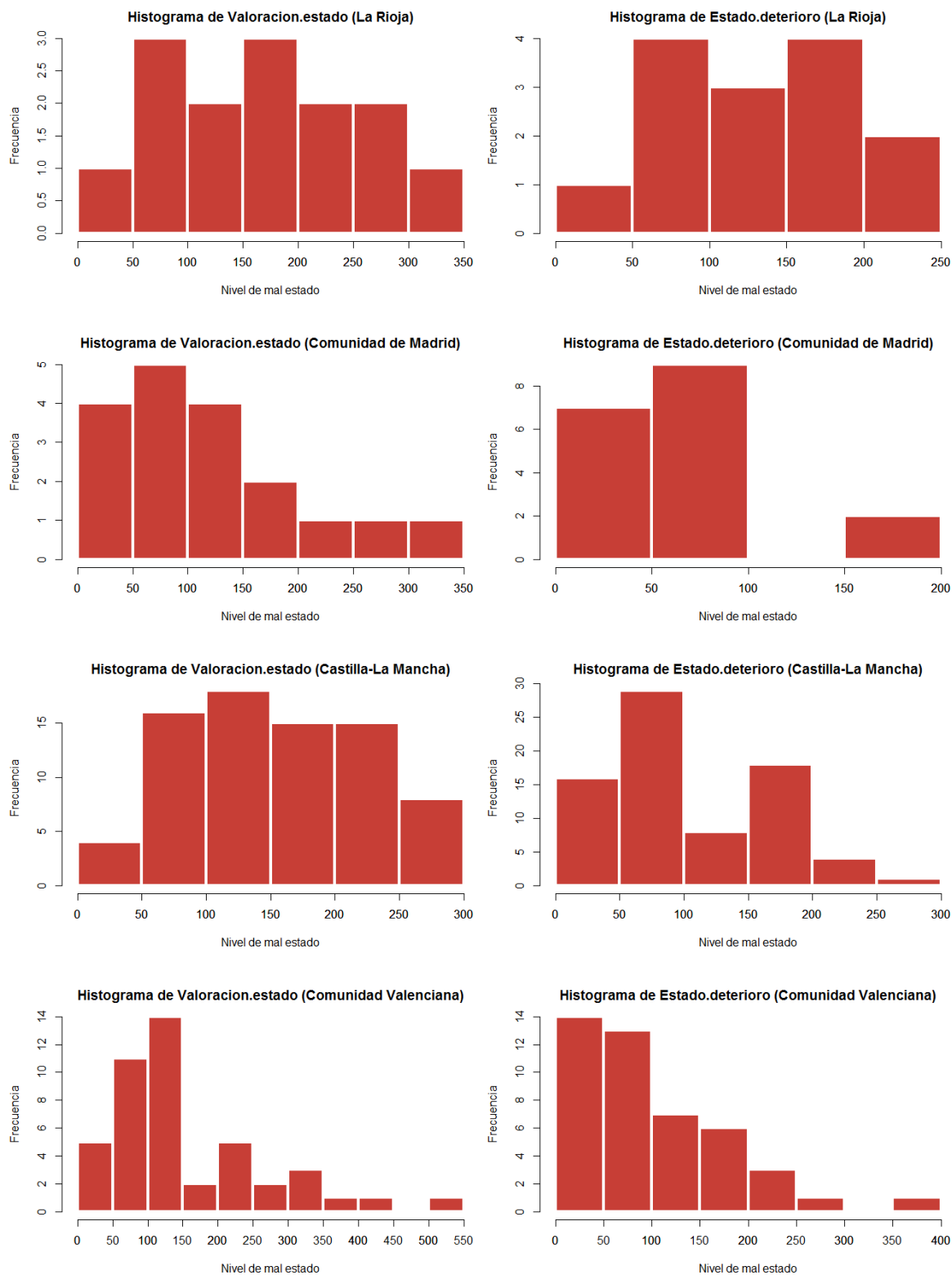


Figura A.3. Histogramas del estado por comunidades autónomas (parte 3)

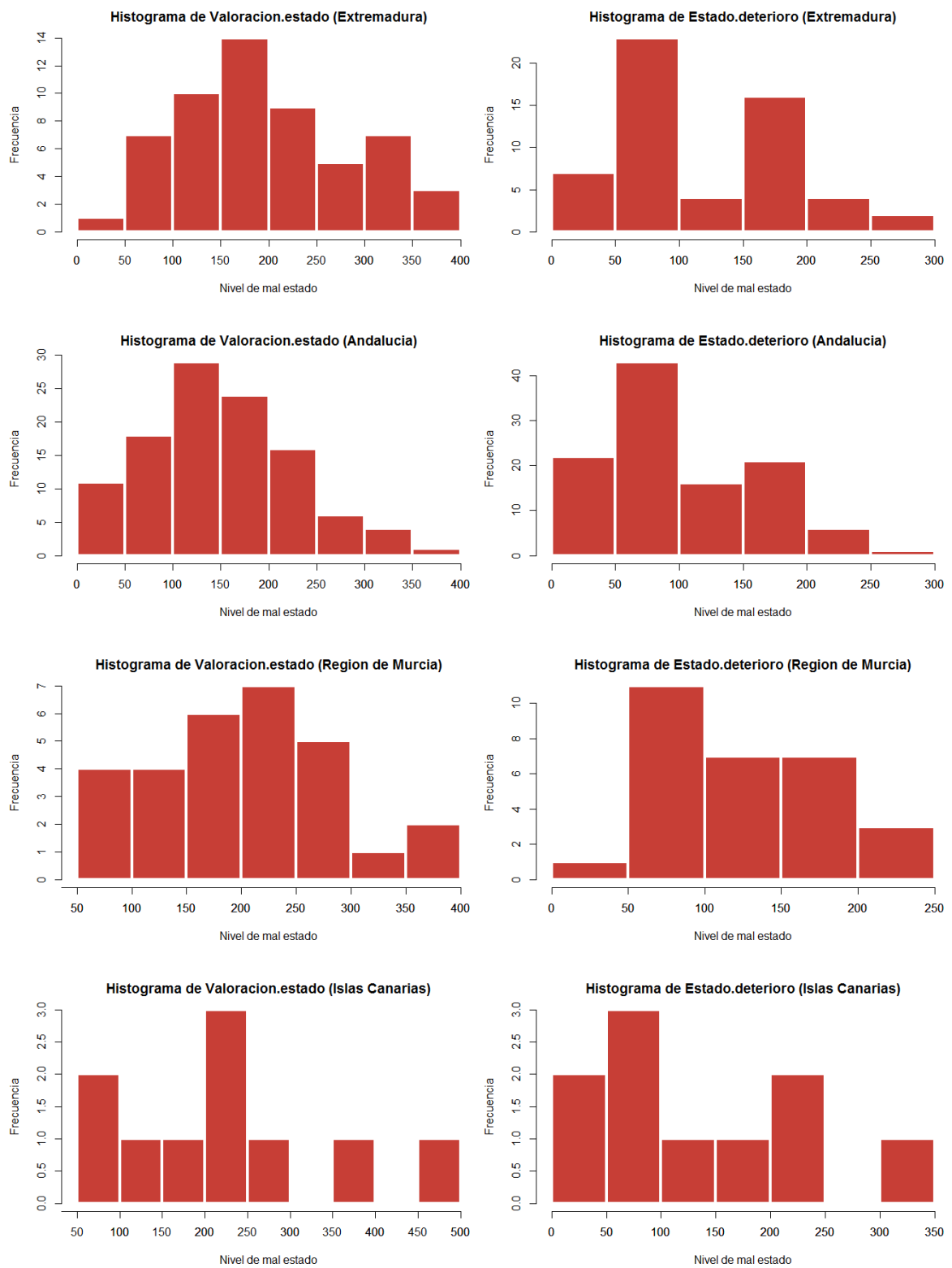


Figura A.4. Histogramas del estado por comunidades autónomas (parte 4)

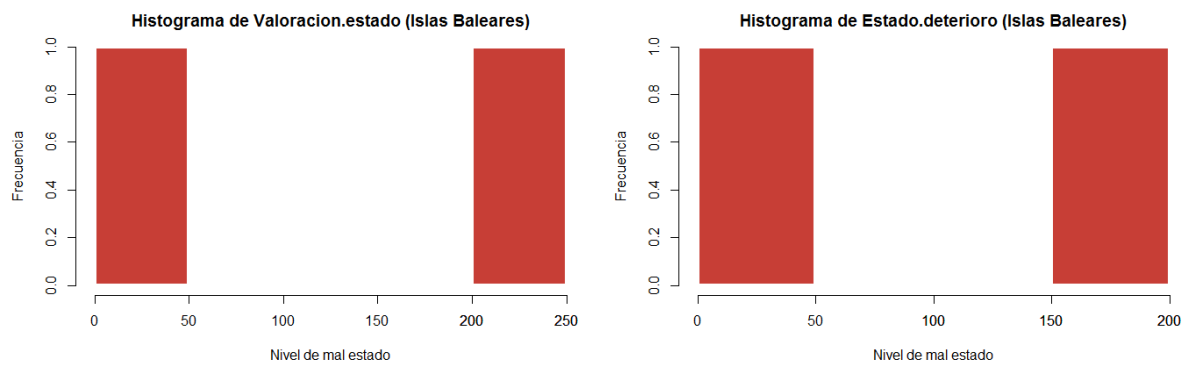


Figura A.5. Histogramas del estado por comunidades autónomas (parte 5)

*Nota: No existen histogramas de la Ciudades Autónomas de Ceuta y Melilla porque la primera consta de un único dato y la segunda carece de fichas en la Lista Roja actualmente.*

## 18. ANEXO B: MEDIDAS DESCRIPTIVAS E HISTOGRAMAS DEL ESTADO DE CONSERVACIÓN A NIVEL PROVINCIAL

LA CORUÑA			
Atributo		Valoración.estado	Estado.deterioro
Rango		10-340	5-230
Media		147.1765	111.2941
Desviación estándar		85.5264	63.7247
Moda		80	80
Cuartiles	Q1 (25%)	80	80
	Q2 (50%)	140	87
	Q3 (75%)	210	150

Tabla B.1. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en La Coruña

LUGO			
Atributo		Valoración.estado	Estado.deterioro
Rango		20-210	20-210
Media		134.5	100
Desviación estándar		62.90	68.4105
Moda		20	80
Cuartiles	Q1 (25%)	130.25	65
	Q2 (50%)	145	80
	Q3 (75%)	157.5	132.5

Tabla B.2. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Lugo

ORENSE			
Atributo		Valoración.estado	Estado.deterioro
Rango		20-280	5-160
Media		128.4286	66.4286
Desviación estándar		100.1879	68.2345
Moda		20	20
Cuartiles	Q1 (25%)	76	20
	Q2 (50%)	87	20
	Q3 (75%)	180	120

Tabla B.3. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Orense

PONTEVEDRA			
Atributo		Valoración.estado	Estado.deterioro
Rango		10-255	5-150
Media		139.4615	76.4615
Desviación estándar		79.8943	47.6840
Moda		147	5
Cuartiles	Q1 (25%)	107	40
	Q2 (50%)	147	80
	Q3 (75%)	207	110

Tabla B.4. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Pontevedra



GUIPÚZCOA		
Atributo	Valoración.estado	Estado.deterioro
Rango	10-140	5-80
Media	67.5	46.25
Desviación estándar	62.9152	39.4493
Moda	10	80
Cuartiles	Q1 (25%)	17.5
	Q2 (50%)	60
	Q3 (75%)	110

Tabla B.5. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Guipuzcoa

VIZCAYA		
Atributo	Valoración.estado	Estado.deterioro
Rango	10-257	5-160
Media	136.5	71.5
Desviación estándar	83.08	63.2749
Moda	150	5
Cuartiles	Q1 (25%)	113.75
	Q2 (50%)	150
	Q3 (75%)	180

Tabla B.6. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Vizcaya

*Nota: No existe tabla de Álava porque carece de fichas en la Lista Roja actualmente.*

<b>HUESCA</b>			
<b>Atributo</b>		<b>Valoración.estado</b>	<b>Estado.deterioro</b>
<b>Rango</b>		10-310	5-230
<b>Media</b>		151.3438	111.3438
<b>Desviación estándar</b>		88.3372	65.9123
<b>Moda</b>		80	80
<b>Cuartiles</b>	<b>Q1 (25%)</b>	80	80
	<b>Q2 (50%)</b>	127.5	97.5
	<b>Q3 (75%)</b>	242.5	161.75

Tabla B.7. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Huesca

<b>TERUEL</b>			
<b>Atributo</b>		<b>Valoración.estado</b>	<b>Estado.deterioro</b>
<b>Rango</b>		85-405	80-305
<b>Media</b>		241.8	169.8
<b>Desviación estándar</b>		131.4294	118.4808
<b>Moda</b>		85	85
<b>Cuartiles</b>	<b>Q1 (25%)</b>	145	85
	<b>Q2 (50%)</b>	240	85
	<b>Q3 (75%)</b>	334	294

Tabla B.8. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Teruel

ZARAGOZA			
Atributo		Valoración.estado	Estado.deterioro
Rango		45-255	34-235
Media		143.7647	112.6765
Desviación estándar		57.2660	50.1662
Moda		140	80
Cuartiles	Q1 (25%)	85	80
	Q2 (50%)	142.5	85
	Q3 (75%)	185	157.5

Tabla B.9. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Zaragoza

BARCELONA			
Atributo		Valoración.estado	Estado.deterioro
Rango		103-297	23-237
Media		62.5186	68.4731
Desviación estándar		191.7143	111.7143
Moda		103	80
Cuartiles	Q1 (25%)	161	80
	Q2 (50%)	180	87
	Q3 (75%)	220	137.5

Tabla B.10. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Barcelona

GERONA		
Atributo	Valoración.estado	Estado.deterioro
Valor	155	155

Tabla B.11. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Gerona

LÉRIDA		
Atributo	Valoración.estado	Estado.deterioro
Rango	110-347	90-227
Media	227.8	158.8
Desviación estándar	94.2321	48.4530
Moda	110	160
Cuartiles	Q1 (25%)	157
	Q2 (50%)	260
	Q3 (75%)	265

Tabla B.12. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Lérida

TARRAGONA		
Atributo	Valoración.estado	Estado.deterioro
Rango	160-275	40-157
Media	193	98
Desviación estándar	55.2509	66.9776
Moda	160	40
Cuartiles	Q1 (25%)	160
	Q2 (50%)	168.5
	Q3 (75%)	201.15

Tabla B.13. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Tarragona

ÁVILA			
Atributo		Valoración.estado	Estado.deterioro
Rango		40-394	40-307
Media		151.6	127.1333
Desviación estándar		101.6463	78.8106
Moda		80	80
Cuartiles	Q1 (25%)	80	80
	Q2 (50%)	100	80
	Q3 (75%)	222.5	162.5

Tabla B.14. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Ávila

BURGOS			
Atributo		Valoración.estado	Estado.deterioro
Rango		10-402	5-235
Media		178.2258	119.4839
Desviación estándar		84.1786	54.6686
Moda		80	80
Cuartiles	Q1 (25%)	130	80
	Q2 (50%)	150	110
	Q3 (75%)	217.5	160

Tabla B.15. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Burgos

LEÓN			
Atributo		Valoración.estado	Estado.deterioro
Rango		23-395	5-312
Media		171	114.6977
Desviación estándar		89.1462	70.2297
Moda		80	80
Cuartiles	Q1 (25%)	102.5	80
	Q2 (50%)	160	80
	Q3 (75%)	222.5	163.5

Tabla B.16. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en León

PALENCIA			
Atributo		Valoración.estado	Estado.deterioro
Rango		20-330	5-240
Media		167.069	108.2069
Desviación estándar		69.6045	68.5567
Moda		100	80
Cuartiles	Q1 (25%)	114	50
	Q2 (50%)	160	80
	Q3 (75%)	190	160

Tabla B.17. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Ávila

SALAMANCA			
Atributo		Valoración.estado	Estado.deterioro
Rango		30-447	20-287
Media		155.6154	103.4231
Desviación estándar		86.3602	60.7152
Moda		60	80
Cuartiles	Q1 (25%)	88.75	72.5
	Q2 (50%)	160	80
	Q3 (75%)	200	145

Tabla B.18. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Salamanca

SEGOVIA			
Atributo		Valoración.estado	Estado.deterioro
Rango		60-270	60-230
Media		174.3	118.55
Desviación estándar		65.9754	47.2646
Moda		145	80
Cuartiles	Q1 (25%)	143.75	80
	Q2 (50%)	172.5	86
	Q3 (75%)	227.75	160

Tabla B.19. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Segovia

SORIA			
Atributo		Valoración.estado	Estado.deterioro
Rango		20-365	5-305
Media		176.3548	132.9677
Desviación estándar		82.2239	70.7877
Moda		85	165
Cuartiles	Q1 (25%)	100	82.5
	Q2 (50%)	170	150
	Q3 (75%)	225	166

Tabla B.20. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Soria

VALLADOLID			
Atributo		Valoración.estado	Estado.deterioro
Rango		10-280	5-230
Media		161	114.8
Desviación estándar		56.7923	54.6603
Moda		185	80
Cuartiles	Q1 (25%)	140.75	80
	Q2 (50%)	160	102.5
	Q3 (75%)	185	230

Tabla B.21. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Valladolid



ZAMORA			
Atributo		Valoración.estado	Estado.deterioro
Rango		20-345	20-265
Media		179.6087	132
Desviación estándar		75.9520	64.5079
Moda		145	80
Cuartiles	Q1 (25%)	145	82.5
	Q2 (50%)	180	125
	Q3 (75%)	229.5	160

Tabla B.22. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Zamora

ALBACETE			
Atributo		Valoración.estado	Estado.deterioro
Rango		65-267	5-247
Media		160.3333	120.3333
Desviación estándar		82.6115	88.3214
Moda		65	5
Cuartiles	Q1 (25%)	88.78	58.75
	Q2 (50%)	170	130
	Q3 (75%)	213.75	163.75

Tabla B.23. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Albacete

CIUDAD REAL			
Atributo	Valoración.estado	Estado.deterioro	
Rango	60-235	5-175	
Media	138.1818	95	
Desviación estándar	55.6953	48.3384	
Moda	60	80	
Cuartiles	Q1 (25%)	104	75
	Q2 (50%)	130	90
	Q3 (75%)	156	119

Tabla B.24. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Ciudad Real

CUENCA			
Atributo	Valoración.estado	Estado.deterioro	
Rango	34-290	34-230	
Media	171.9091	91.7273	
Desviación estándar	79.5644	54.1093	
Moda	185	80	
Cuartiles	Q1 (25%)	125	65
	Q2 (50%)	165	80
	Q3 (75%)	218.5	102.5

Tabla B.25. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Cuenca

GUADALAJARA			
Atributo		Valoración.estado	Estado.deterioro
Rango		45-260	23-165
Media		153.2917	99.5
Desviación estándar		62.4712	45.5154
Moda		80	80
Cuartiles	Q1 (25%)	103.75	80
	Q2 (50%)	151	80
	Q3 (75%)	194	155

Tabla B.26. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Guadalajara

TOLEDO			
Atributo		Valoración.estado	Estado.deterioro
Rango		10-260	5-260
Media		151.4583	109.4167
Desviación estándar		80.3097	72.3974
Moda		10	5
Cuartiles	Q1 (25%)	77.5	67.5
	Q2 (50%)	168.5	94.5
	Q3 (75%)	220	160

Tabla B.27. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Toledo

ALICANTE			
Atributo		Valoración.estado	Estado.deterioro
Rango		20-520	20-355
Media		170.0667	122.1333
Desviación estándar		128.5171	92.5989
Moda		80	20
Cuartiles	Q1 (25%)	80	60
	Q2 (50%)	120	80
	Q3 (75%)	219	162.5

Tabla B.28. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Alicante

CASTELLÓN			
Atributo		Valoración.estado	Estado.deterioro
Rango		60-280	5-160
Media		156.25	82.5
Desviación estándar		91.23	63.3114
Moda		60	5
Cuartiles	Q1 (25%)	120	61.25
	Q2 (50%)	142.5	82.5
	Q3 (75%)	178.75	103.75

Tabla B.29. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Castellón

VALENCIA			
Atributo		Valoración.estado	Estado.deterioro
Rango		10-410	5-274
Media		151.5	99
Desviación estándar		104.6481	67.0606
Moda		115	80
Cuartiles	Q1 (25%)	85	46.25
	Q2 (50%)	115	80
	Q3 (75%)	204.25	129.25

Tabla B.30. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Valencia

CÁCERES			
Atributo		Valoración.estado	Estado.deterioro
Rango		10-362	5-230
Media		189.3793	124.2759
Desviación estándar		93.2027	62.2993
Moda		145	80
Cuartiles	Q1 (25%)	140	80
	Q2 (50%)	180	93
	Q3 (75%)	262	160

Tabla B.31. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Cáceres

<b>BADAJOS</b>			
<b>Atributo</b>		<b>Valoración.estado</b>	<b>Estado.deterioro</b>
<b>Rango</b>		70-362	20-274
<b>Media</b>		207.8519	118.1481
<b>Desviación estándar</b>		82.0655	63.7071
<b>Moda</b>		242	80
<b>Cuartiles</b>	<b>Q1 (25%)</b>	157.5	80
	<b>Q2 (50%)</b>	187	95
	<b>Q3 (75%)</b>	248	158.5

Tabla B.32. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Badajoz

<b>ALMERÍA</b>			
<b>Atributo</b>		<b>Valoración.estado</b>	<b>Estado.deterioro</b>
<b>Rango</b>		70-232	70-230
<b>Media</b>		171.7143	133.1429
<b>Desviación estándar</b>		71.9646	58.3450
<b>Moda</b>		70	70
<b>Cuartiles</b>	<b>Q1 (25%)</b>	112.5	82.5
	<b>Q2 (50%)</b>	220	140
	<b>Q3 (75%)</b>	227.5	163.5

Tabla B.33. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Almería

<b>CÁDIZ</b>			
<b>Atributo</b>		<b>Valoración.estado</b>	<b>Estado.deterioro</b>
<b>Rango</b>		30-297	30-170
<b>Media</b>		139.4667	83.5333
<b>Desviación estándar</b>		85.3010	49.7291
<b>Moda</b>		140	80
<b>Cuartiles</b>	<b>Q1 (25%)</b>	70	40
	<b>Q2 (50%)</b>	140	80
	<b>Q3 (75%)</b>	175	120

Tabla B.34. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Cádiz

<b>CÓRDOBA</b>			
<b>Atributo</b>		<b>Valoración.estado</b>	<b>Estado.deterioro</b>
<b>Rango</b>		80-317	80-210
<b>Media</b>		184.9167	136.8333
<b>Desviación estándar</b>		68.9420	50.2301
<b>Moda</b>		140	80
<b>Cuartiles</b>	<b>Q1 (25%)</b>	137.5	80
	<b>Q2 (50%)</b>	180	130
	<b>Q3 (75%)</b>	234	182.5

Tabla B.35. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Córdoba

GRANADA		
Atributo	Valoración.estado	Estado.deterioro
Rango	30-385	30-285
Media	133.4545	105.8182
Desviación estándar	74.0269	60.8446
Moda	100	80
Cuartiles	Q1 (25%)	100
	Q2 (50%)	124.5
	Q3 (75%)	145

Tabla B.36. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Granada

HUELVA		
Atributo	Valoración.estado	Estado.deterioro
Valor	260	200

Tabla B.37. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Huelva

JAÉN		
Atributo	Valoración.estado	Estado.deterioro
Rango	20-327	20-235
Media	159.0345	100.4483
Desviación estándar	66.5837	48.4529
Moda	180	80
Cuartiles	Q1 (25%)	115
	Q2 (50%)	157
	Q3 (75%)	200

Tabla B.38. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Jaén



<b>MÁLAGA</b>			
<b>Atributo</b>		<b>Valoración.estado</b>	<b>Estado.deterioro</b>
<b>Rango</b>		40-215	20-150
<b>Media</b>		140.8333	78.3333
<b>Desviación estándar</b>		72.3475	42.5049
<b>Moda</b>		40	85
<b>Cuartiles</b>	<b>Q1 (25%)</b>	87.5	62.5
	<b>Q2 (50%)</b>	172.5	77.5
	<b>Q3 (75%)</b>	182.5	85

Tabla B.39. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Málaga

<b>SEVILLA</b>			
<b>Atributo</b>		<b>Valoración.estado</b>	<b>Estado.deterioro</b>
<b>Rango</b>		40-350	20-250
<b>Media</b>		157	105.2353
<b>Desviación estándar</b>		96.3587	72.3218
<b>Moda</b>		40	80
<b>Cuartiles</b>	<b>Q1 (25%)</b>	80	45
	<b>Q2 (50%)</b>	140	80
	<b>Q3 (75%)</b>	210	160

Tabla B.40. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Sevilla

SANTA CRUZ DE TENERIFE			
Atributo		Valoración.estado	Estado.deterioro
Rango		100-467	40-307
Media		250.25	157.125
Desviación estándar		123.9317	89.6365
Moda		100	80
Cuartiles	Q1 (25%)	155	80
	Q2 (50%)	235	155
	Q3 (75%)	311.25	215

Tabla B.41. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Sta. Cruz de Tenerife

LAS PALMAS			
Atributo		Valoración.estado	Estado.deterioro
Rango		80-240	20-80
Media		160	50
Desviación estándar		113.1371	42.4264
Moda		80	20
Cuartiles	Q1 (25%)	120	35
	Q2 (50%)	160	50
	Q3 (75%)	200	65

Tabla B.42. Medidas descriptivas del estado de los bienes incluidos en la Lista Roja en Las Palmas

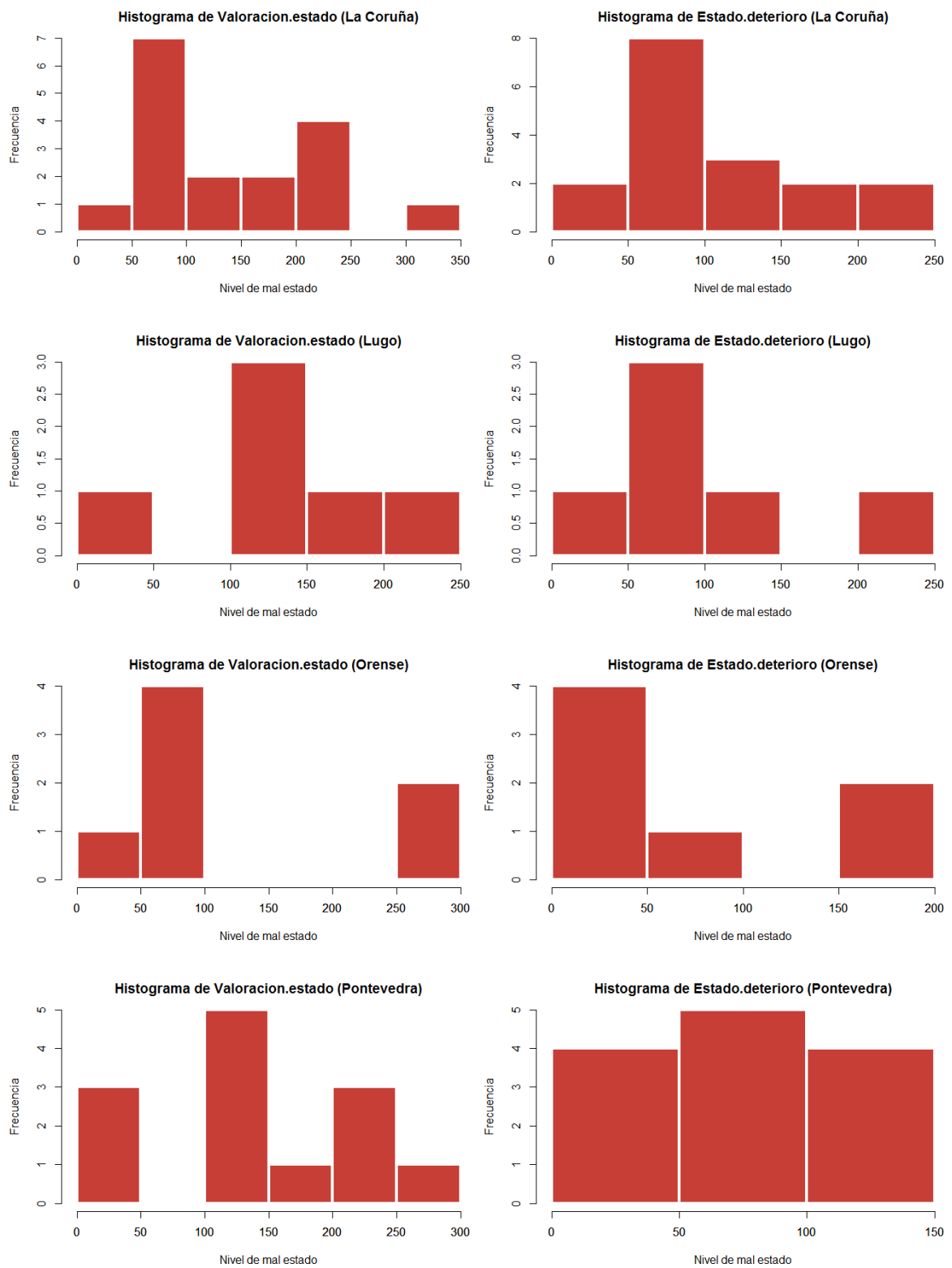


Figura B.1. Histogramas del estado por provincias en Galicia.

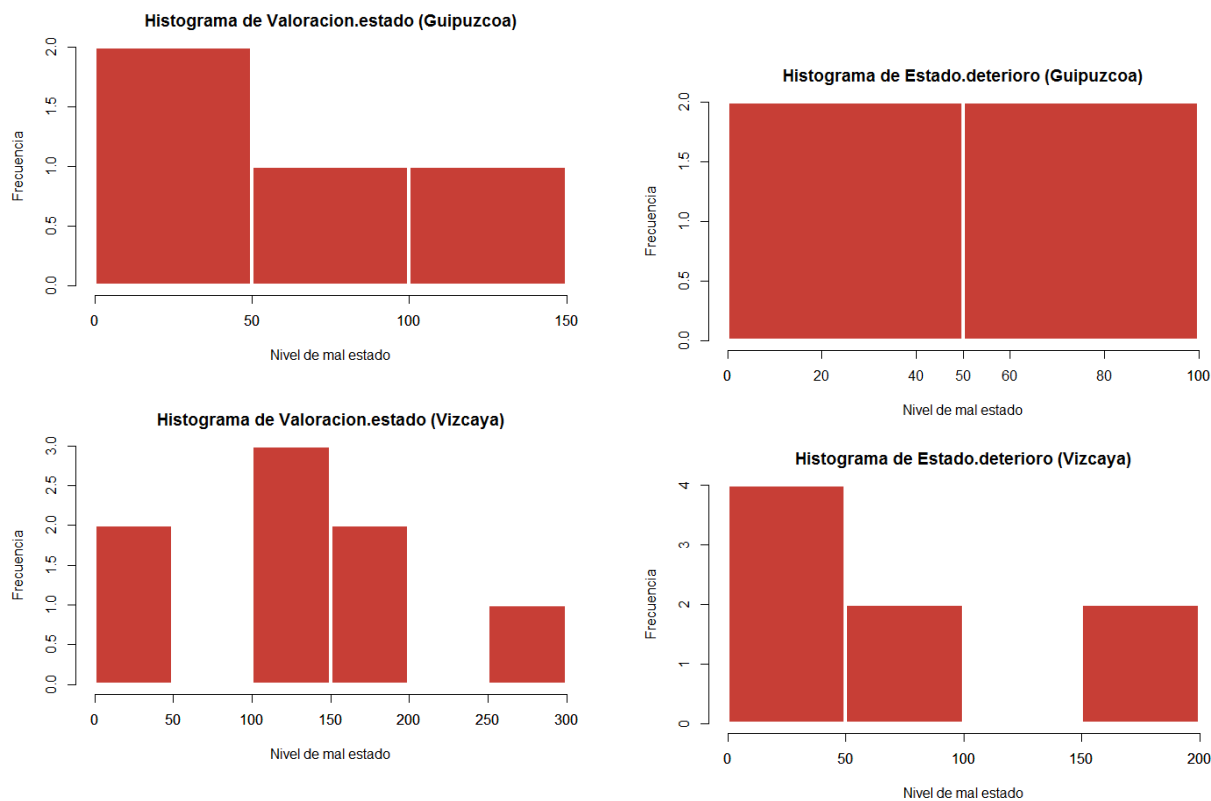


Figura B.2. Histogramas del estado por provincias en País Vasco.

*Nota: No existe histograma de Álava porque carece de fichas en la Lista Roja actualmente.*

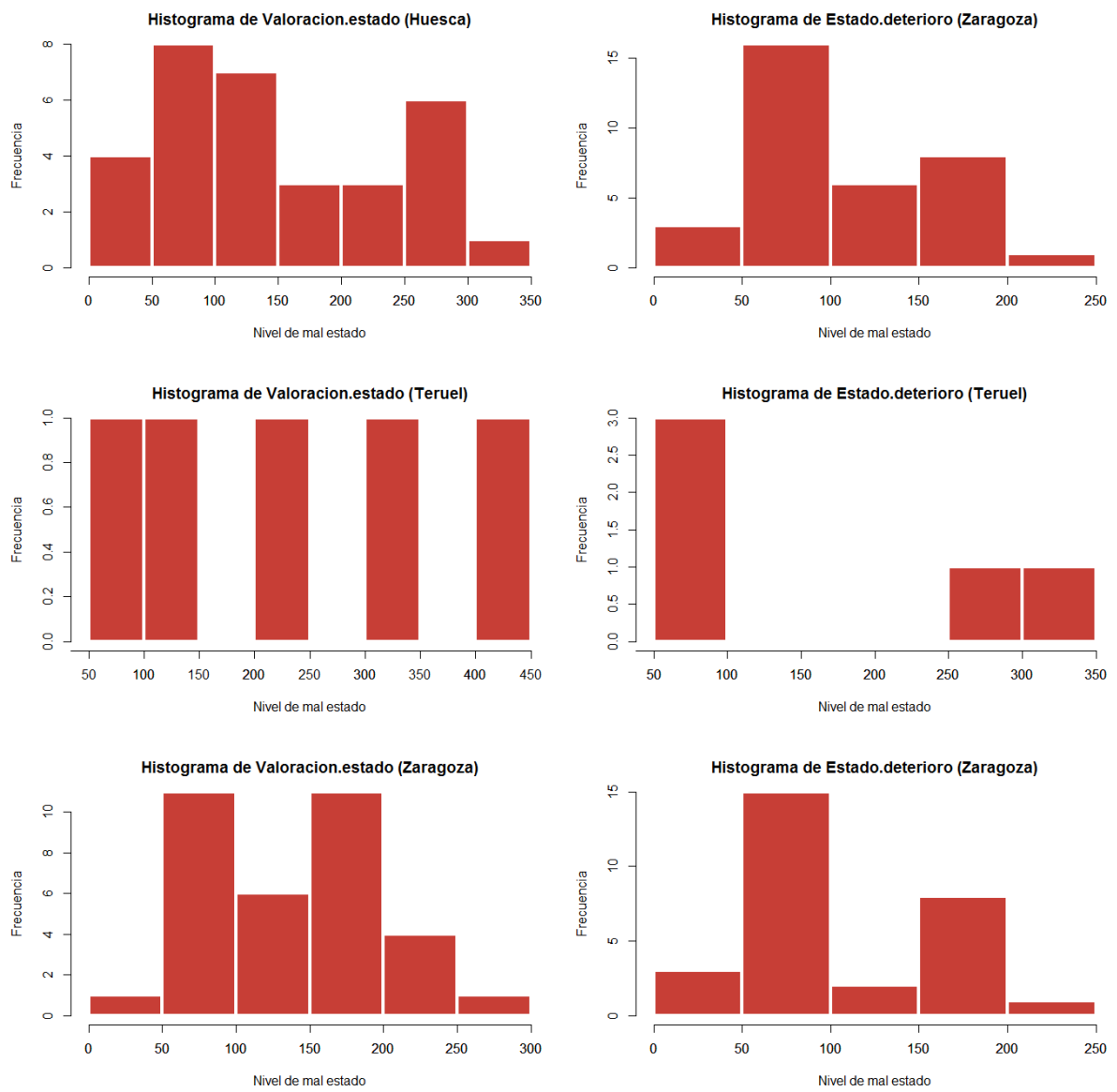


Figura B.3. Histogramas del estado por provincias en Aragón.

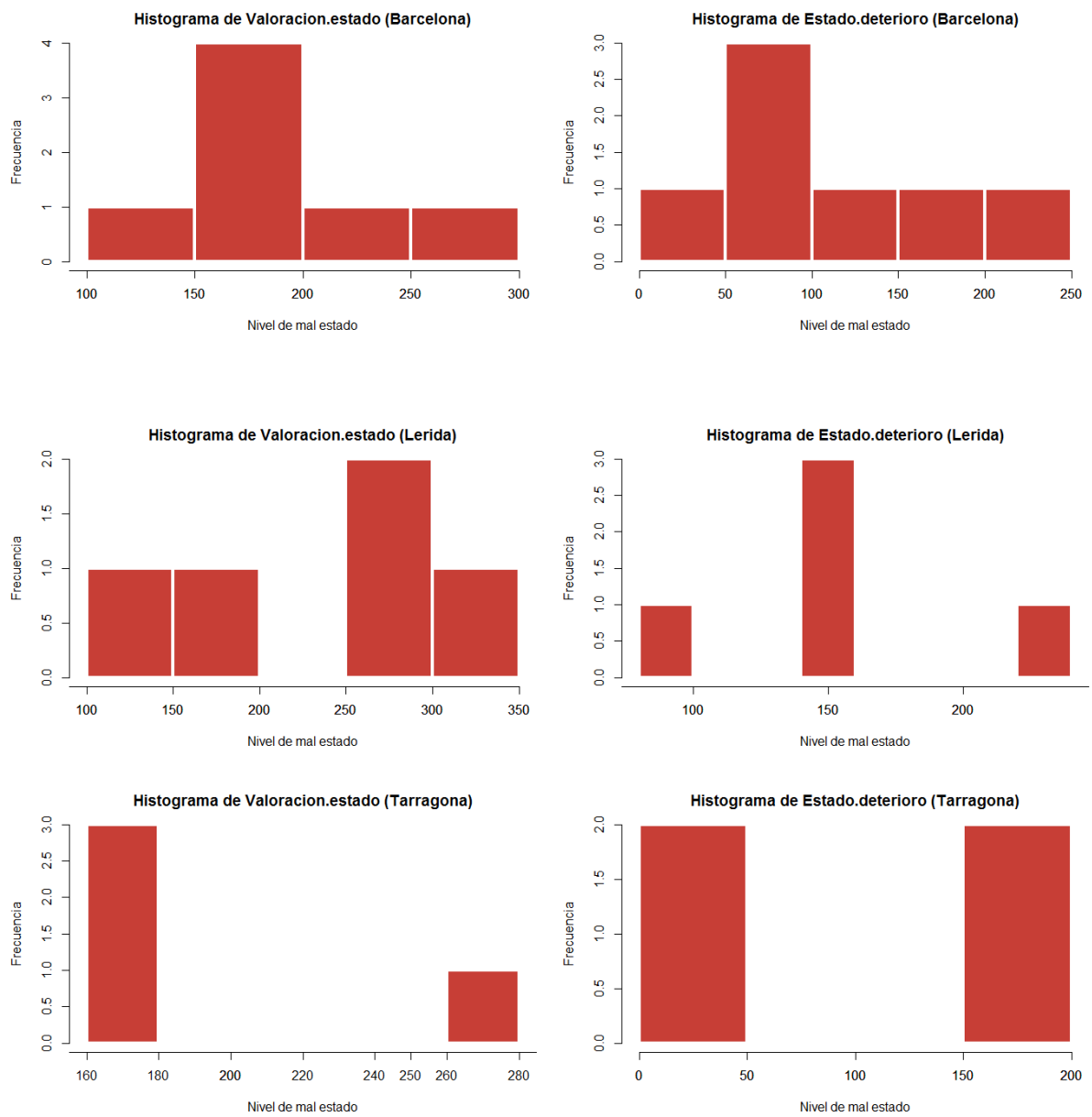


Figura B.4. Histogramas del estado por provincias en Cataluña.

*Nota: No se ha representado histograma de Gerona porque contiene datos con valores idénticos, por lo que consta de una única columna.*

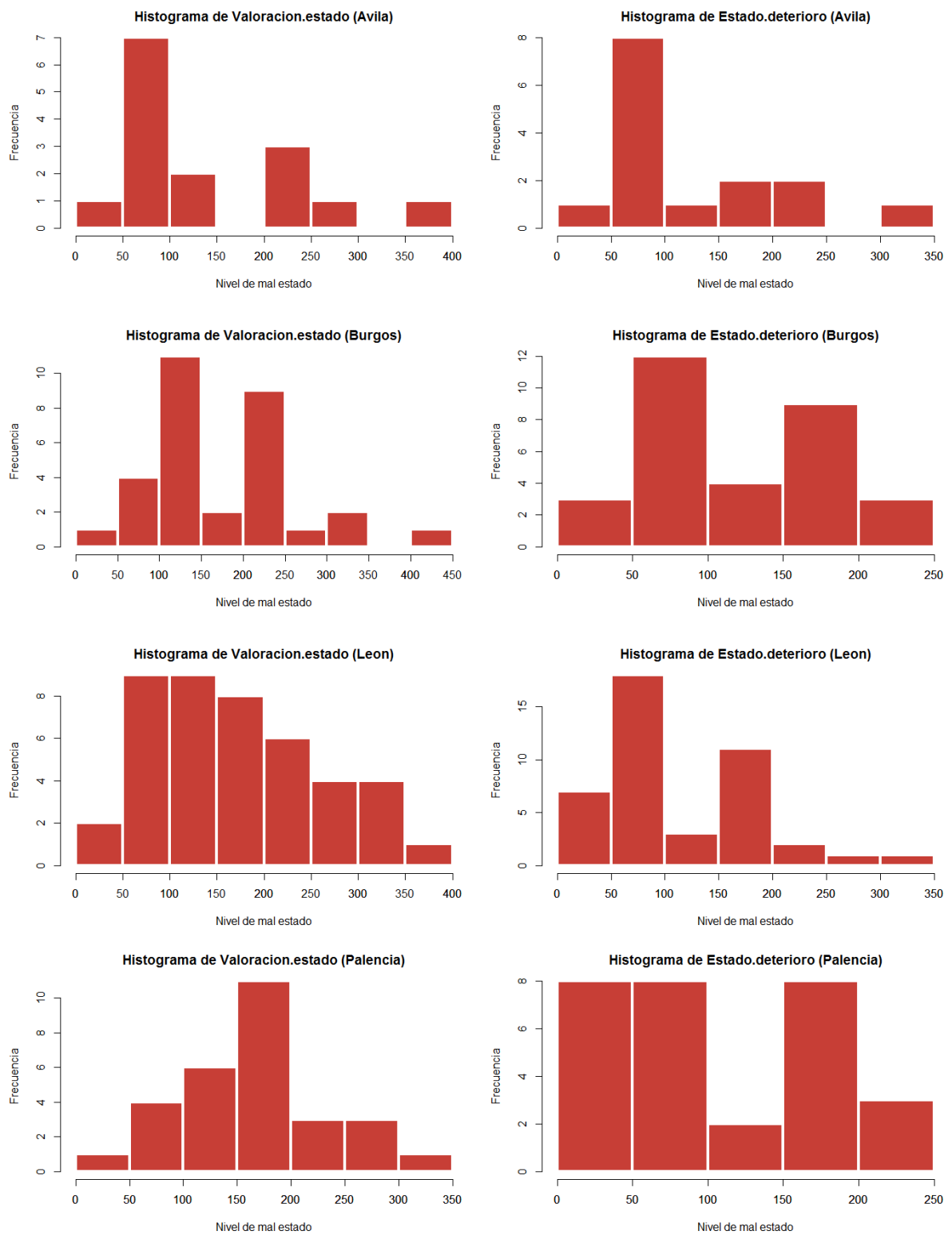


Figura B.5. Histogramas del estado por provincias en Castilla y León (parte 1).

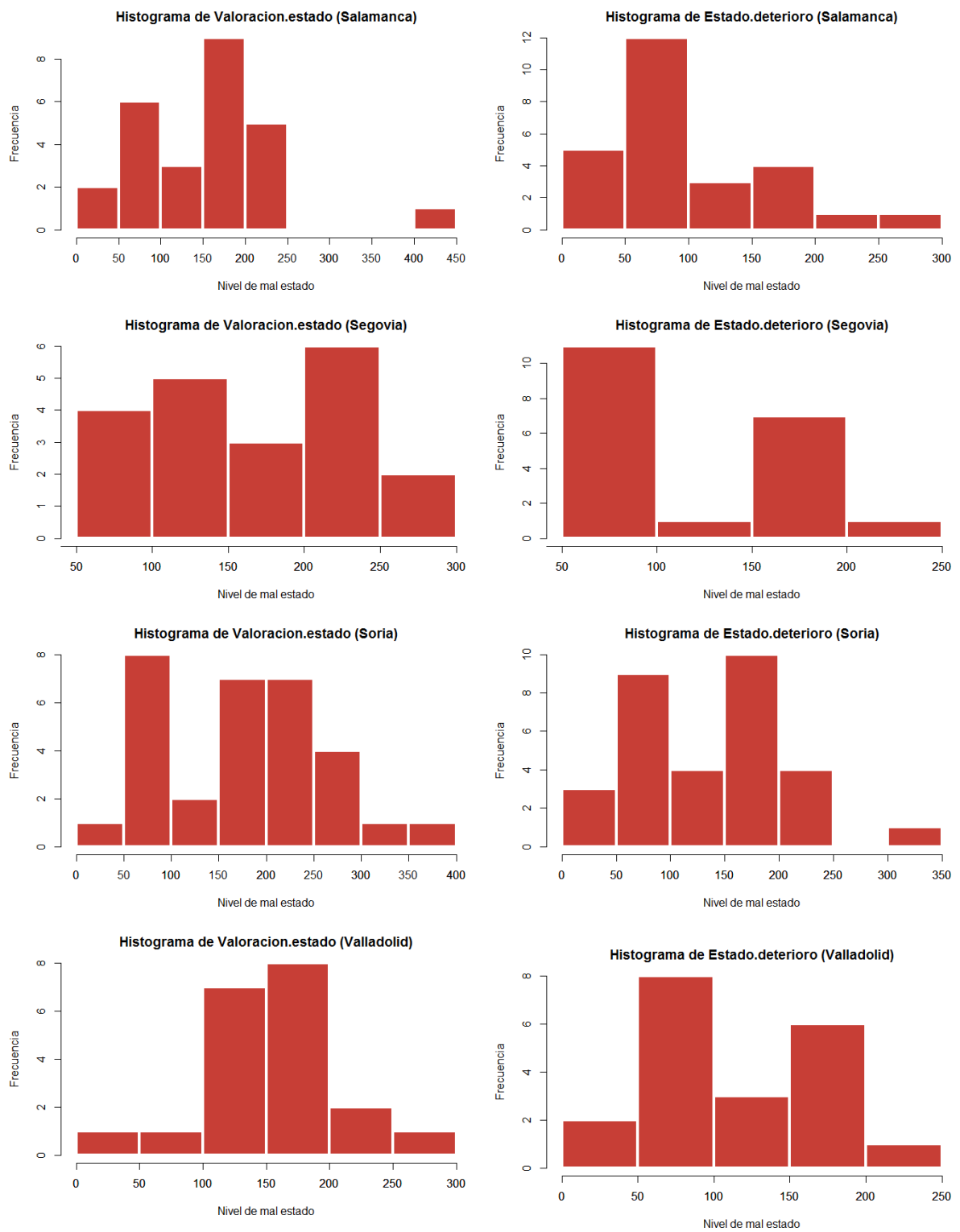


Figura B.6. Histogramas del estado por provincias en Castilla y León (parte 2).



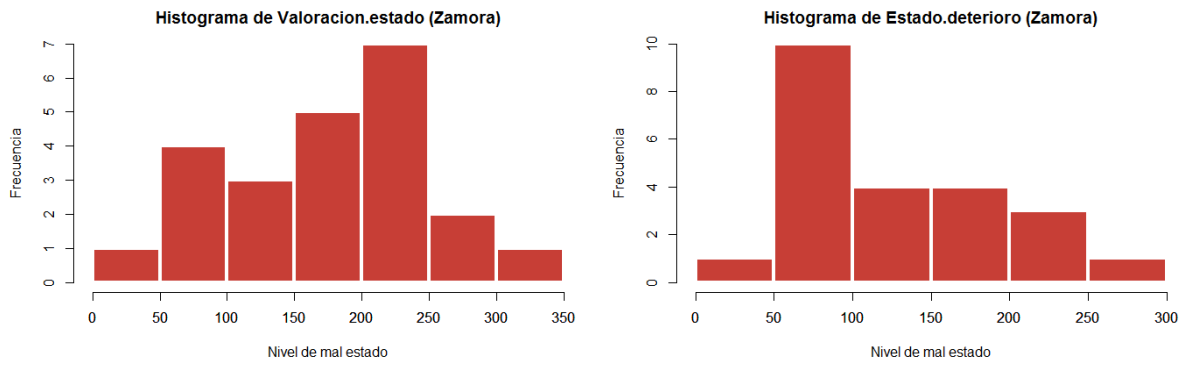


Figura B.7. Histogramas del estado por provincias en Castilla y León (parte 3).

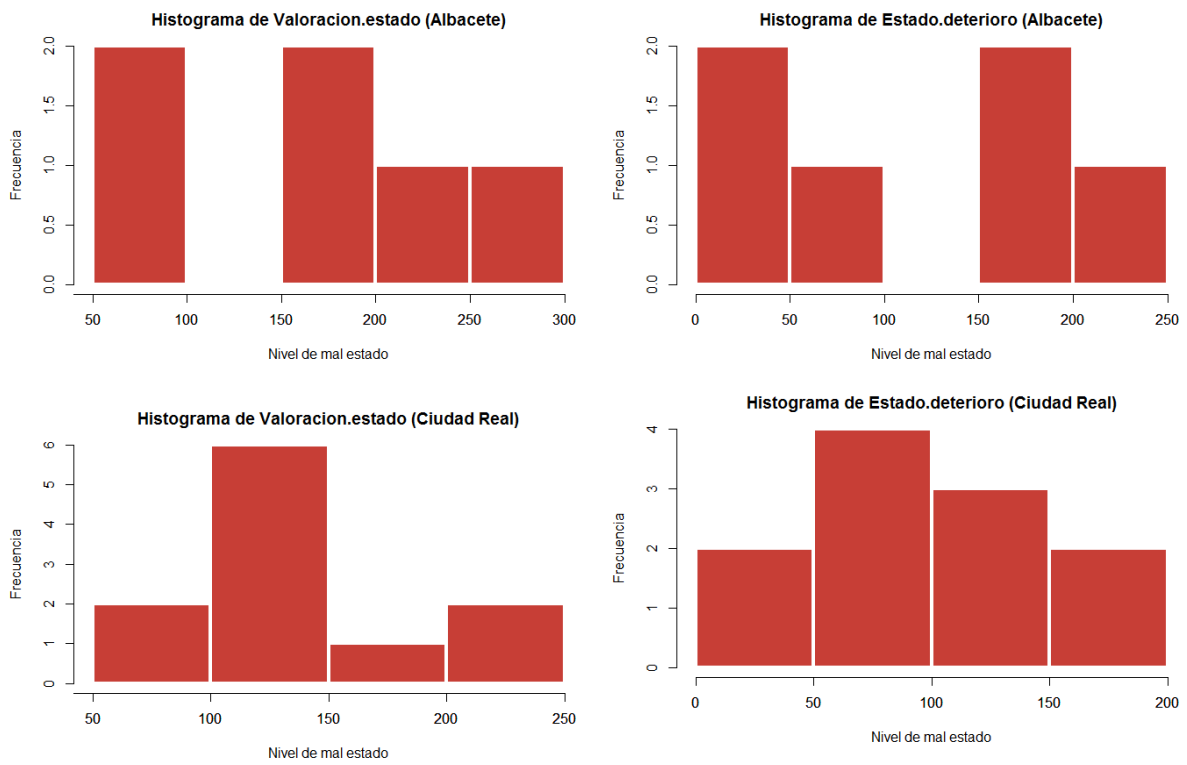


Figura B.8. Histogramas del estado por provincias en Castilla-La Mancha (Parte 1).

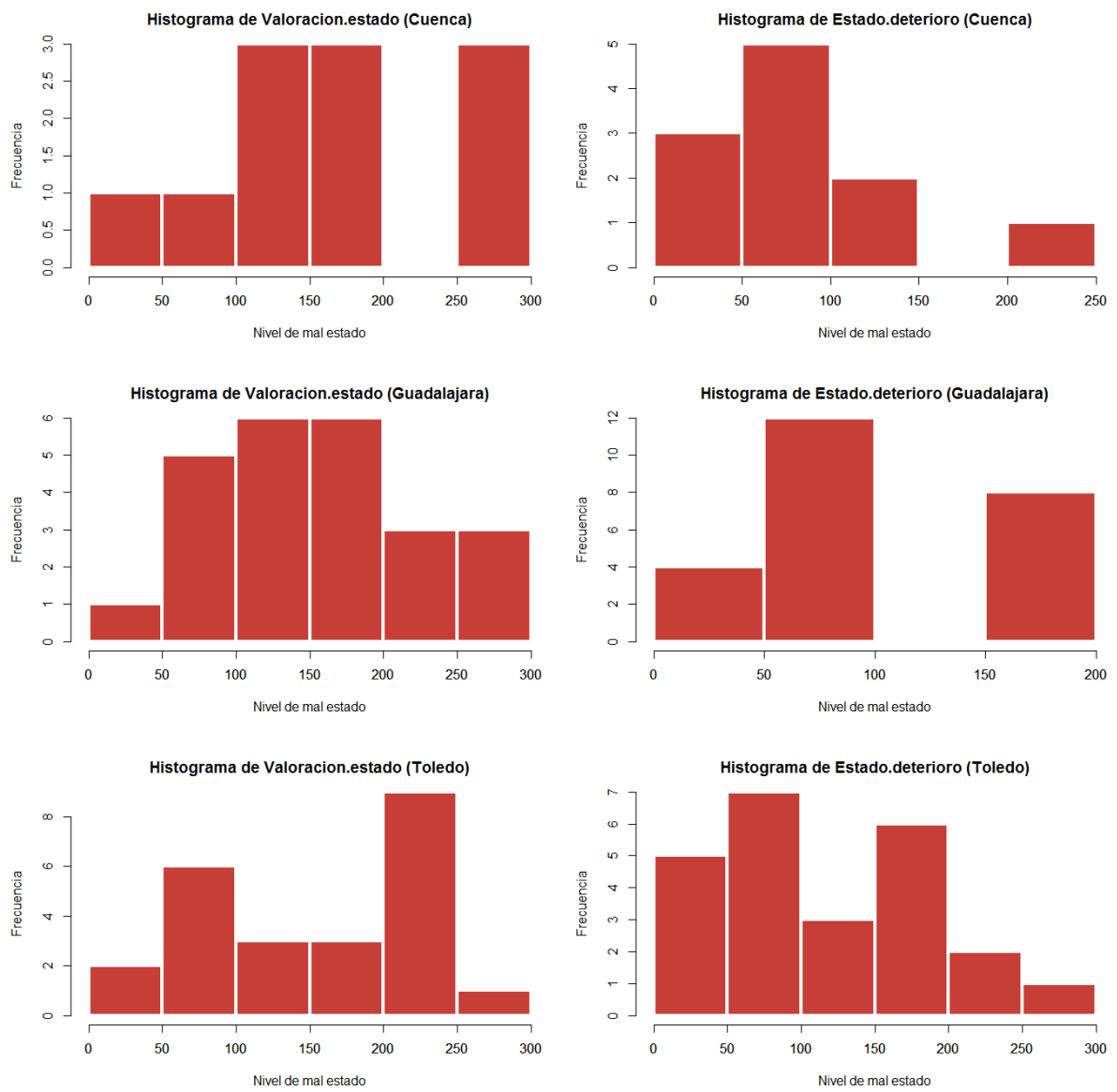


Figura B.9. Histogramas del estado por provincias en Castilla-La Mancha (Parte 2).

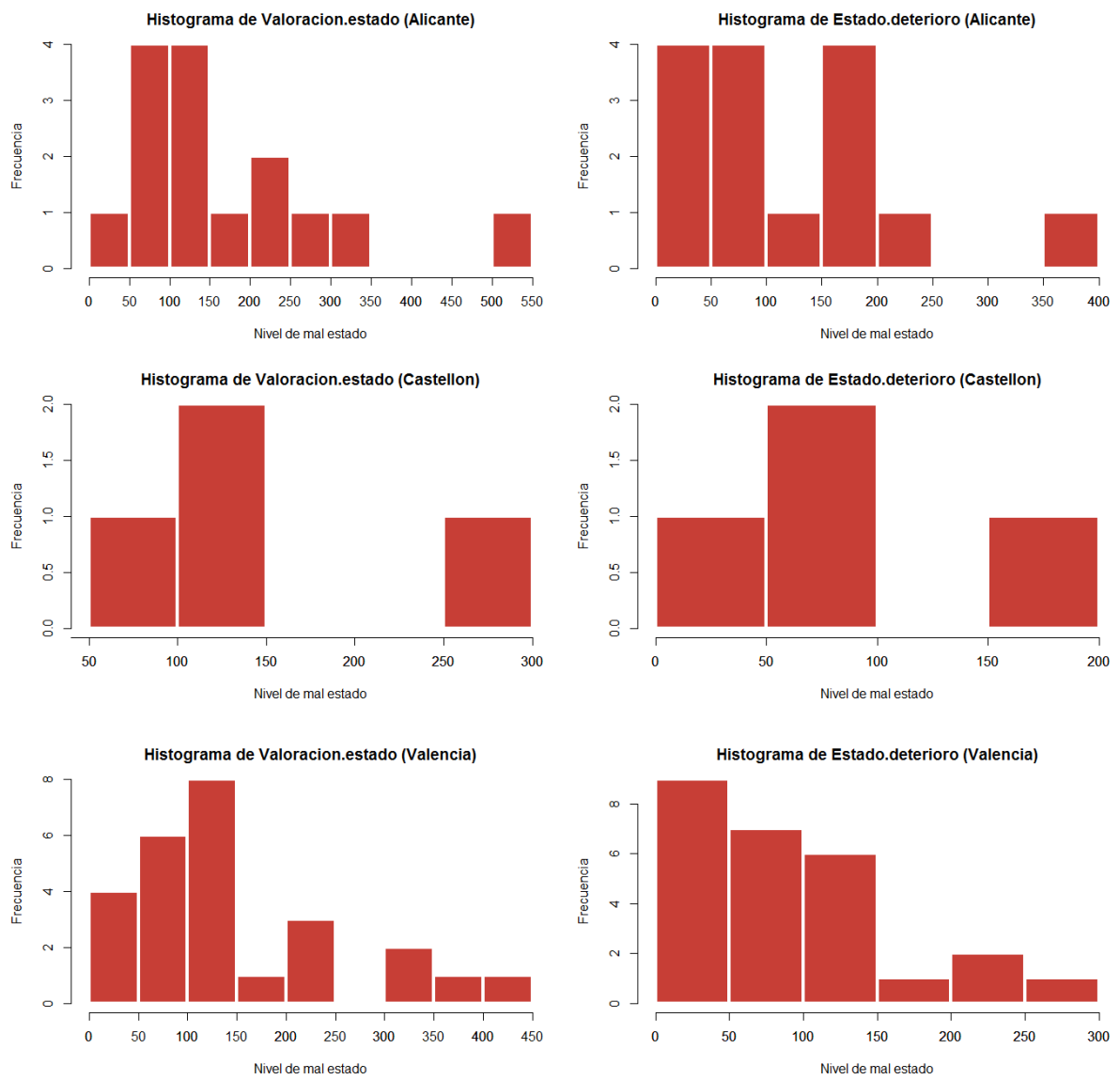


Figura B.10. Histogramas del estado por provincias en Valencia.

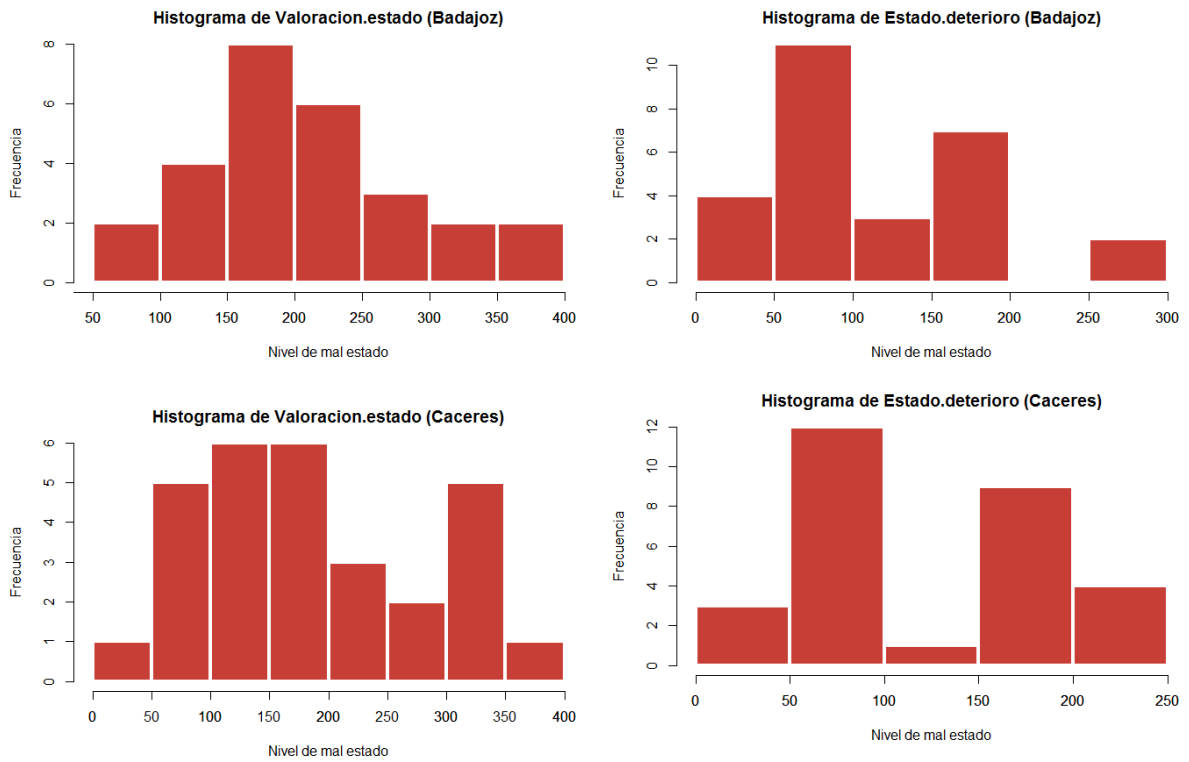


Figura B.11. Histogramas del estado por provincias en Extremadura.

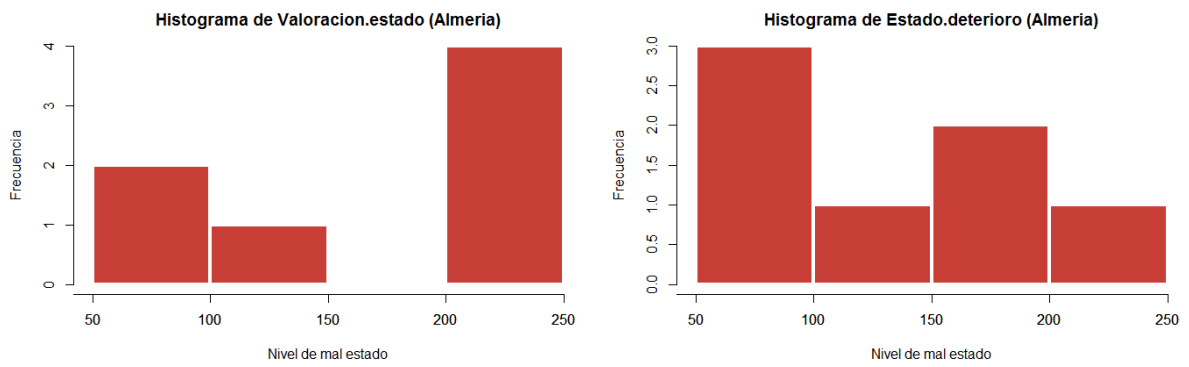


Figura B.12. Histogramas del estado por provincias en Andalucía (parte 1).

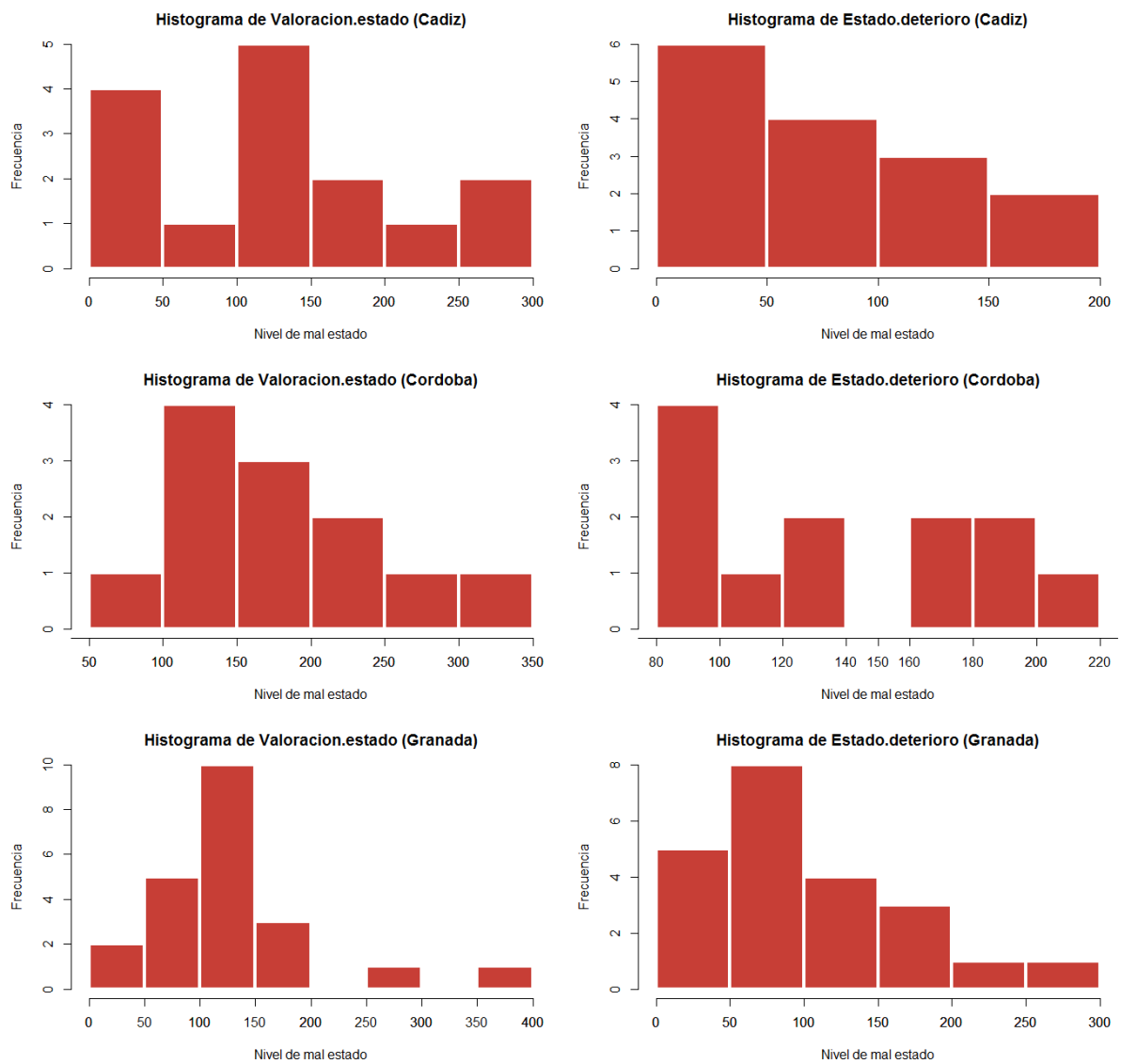


Figura B.13. Histogramas del estado por provincias en Andalucía (parte 2).

*Nota: No se ha representado histograma de Huelva porque contiene datos idénticos entre sí.*

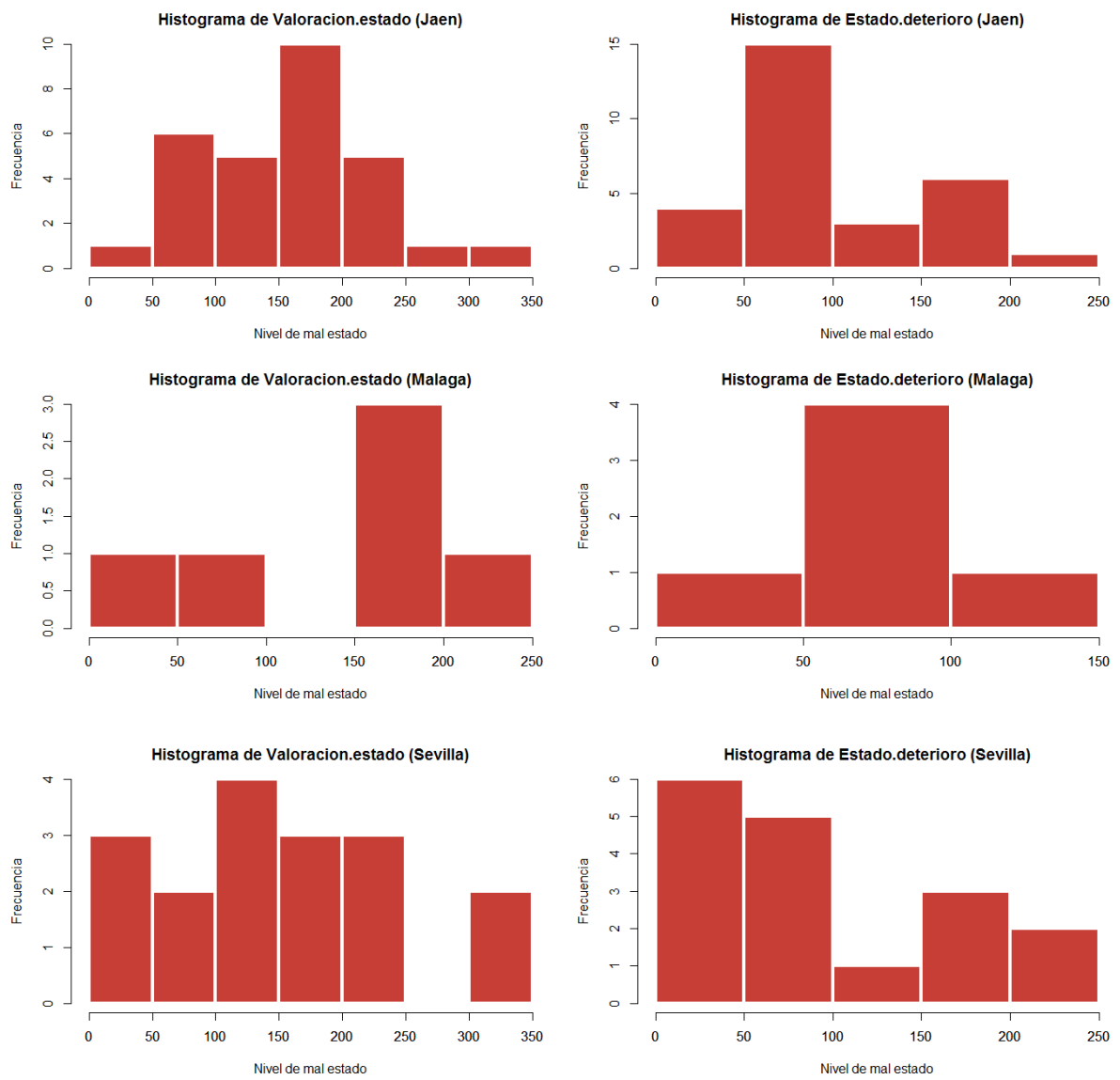


Figura B.14. Histogramas del estado por provincias en Andalucía (parte 3).

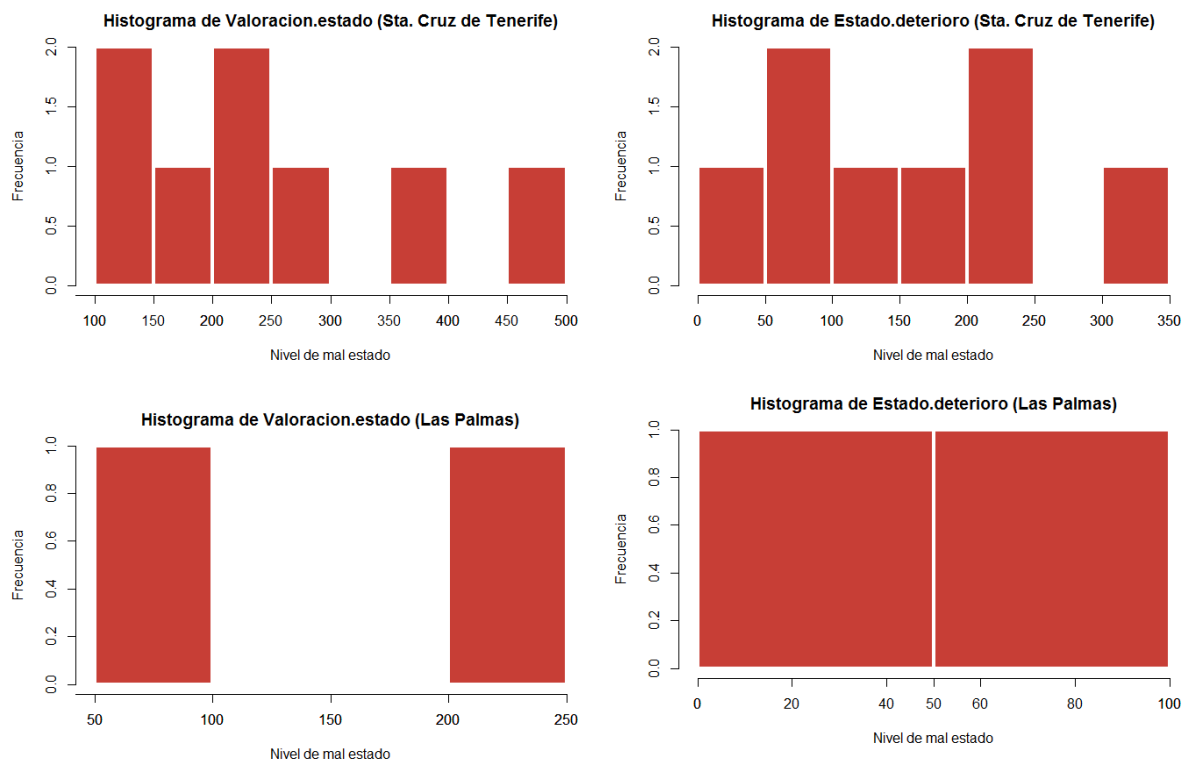


Figura B.15. Histogramas del estado por provincias en Islas Canarias.

*Nota: Para provincias de Comunidades Autónomas provinciales, consultar el Anexo A.*

## 19. ANEXO C: PROPUESTAS PARA LA MEJORA DE LA LISTA ROJA DE PATRIMONIO

En febrero de 2020 tuvo lugar una nueva reunión con los responsables de Hispania Nostra para dar parte del progreso del estudio.

El objetivo de este TFG, más allá del interés académico, es también que la entidad cesora de los datos se vea beneficiada en todo momento del análisis. Por ello se ha realizado un informe que recopila recomendaciones para su mejora tras los primeros análisis de la Lista Roja de Patrimonio. Los puntos contemplados fueron los siguientes:

### 19.1. ELIMINACIÓN DE LA REDUNDANCIA Y AGRUPACIÓN DE LOS ATRIBUTOS

Existe una considerable cantidad de atributos redundantes que contienen u ofrecen la misma información. Muchos de ellos es posible que se deban a las actualizaciones y modificaciones que ha sufrido el fichero de datos, así como el formulario de inclusión de bienes a lo largo de los años que lleva en funcionamiento.

Se recomienda unificar todas aquellas variables redundantes y buscar siempre la división de atributos según la información que ofrecen. Se recomienda evitar unir diferentes aspectos en uno solo, como se ha hecho en **Localización** (que contiene los datos de la provincia y la comunidad autónoma) e **Historia/Descripción**.

### 19.2. UNIFORMIDAD DEL FICHERO DE DATOS

**Época.** Debido a la ambigüedad del nombre no existe un formato ni tipo de dato consensuado. Se recomienda establecer un consenso y un formato específico que se indique con más claridad en el formulario. Lo ideal sería almacenar el siglo de edificación, y esto podría hacerse, por ejemplo, mediante un campo a cumplimentar con limitación de caracteres.

**Coordenadas.** No existe un formato consensuado. Se recomienda la implementación de una plantilla estructurada dentro del formulario, similares a los que ofrecen las entidades comerciales para los datos de las tarjetas de crédito, en base al formato deseado. La implementación de un sistema que permitiese el uso de un mapa (similar al que dispone la Sede Electrónica del Catastro) podría ser de enorme utilidad.

**Propiedad.** Se recomienda utilizar un “option button” con las opciones “Privado”, “Público” y “Público y Privado” y añadir un campo para explayarse. De esta forma, el usuario puede redactar los detalles que estime oportunos, pero siempre tendremos la clasificación del bien según su titularidad.



### 19.3. ERRORES EN EL FORMULARIO DE ENVÍO DE FICHAS

Se han detectado una serie de errores en el formulario de envío de fichas de la Lista Roja cuya corrección debería ser considerada en pos de una mejor experiencia de usuario y una gestión de los datos más precisa. Para hacer esta demostración más práctica y visual se buscó información sobre un bien patrimonial que pudiese ser incluido en la Lista Roja, se elaboró su ficha y se cumplimentaron los campos sobre dicho bien, mostrando los diferentes aspectos a mejorar. Esta ficha puede verse en el *Anexo D*.

#### Provincia

El campo “Provincia” consiste en un desplegable que contiene entre sus opciones todas las provincias españolas, así como una lista de las comunidades autónomas, tal y como se muestra en la *figura C.1*. Esto puede generar el envío de incongruencias como las que se muestran en la *figura C.2*, ya que el formulario las permite como opción válida. Para evitar esto, el desplegable de “Provincia” debería contener las opciones limitadas en función de la selección realizada previamente en “Comunidad Autónoma”. Para trabajar con fichas cuya localización sea toda la Península Ibérica, podría introducirse “España” o “Toda la geografía española” dentro de “Comunidad Autónoma”, y que esta deshabilite el desplegable de “Provincia”.



Figura C.1. Desplegable de Provincia en el formulario.



Figura C.2. Posibles combinaciones erróneas permitidas en el formulario.

**Tipología.** El campo “Tipología” consiste en un desplegable [figura C.3] que contiene entre sus opciones las distintas clases agrupadas de patrimonio, así como su subcategoría dentro de cada uno. Esto permite al usuario ser más concreto, pero también puede aportar menos información de la que conoce, pudiendo introducir simplemente como dato el tipo de patrimonio. Una manera de solventar esto podría ser implementar una opción similar a la descrita en “Provincia”, habilitando un segundo desplegable. cuyo nombre podría ser “Subtipo”, que contenga las diferentes subcategorías en función de la clase de patrimonio marcada anteriormente.

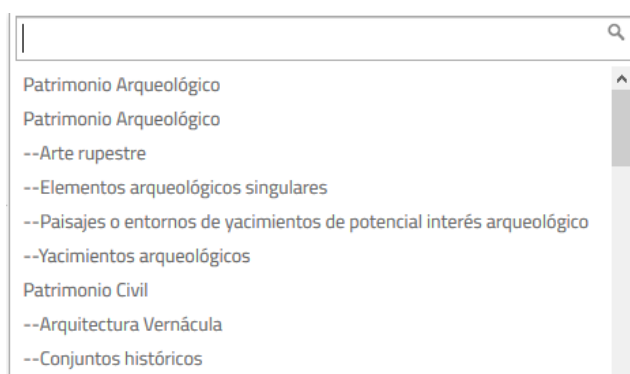


Figura C.3. Desplegable de Tipología en el formulario.

## 19.4. SISTEMATIZACIÓN DEL ESTADO DE CONSERVACIÓN

Se recomienda prescindir de descripciones no estructuradas para definir el estado de conservación de un bien patrimonial. Lo ideal sería determinar una serie de factores fijos que sean claves inherentes al deterioro y asociar a cada uno de ellos un valor de una escala predefinida en base a un estudio realizado por el comité científico. De esta forma se conseguiría que todos los bienes puedan ser catalogado en base a los mismos criterios y por tanto a una misma escala de medición.

## 20. ANEXO D: FICHA EMPLEADA PARA ANALIZAR EL FORMULARIO DE ENVÍO DE BIENES DE LA LISTA ROJA

<b>Nombre</b>	Pantalán de la Compañía Minera de Sierra Menera
<b>Comunidad Autónoma</b>	Comunidad Valenciana
<b>Provincia</b>	Valencia
<b>Localidad</b>	Puerto de Sagunto (Sagunto)
<b>Coordenadas</b>	39°39'07.6"N 0°12'20.9"W
<b>Tipología</b>	Patrimonio Industrial
<b>Época</b>	Siglo XX
<b>Propietario</b>	APV
<b>Descripción breve</b>	Muelle de uso minero.
<b>Descripción</b>	Estructura de metal y hormigón que se adentra 1,1km sobre el mar.
<b>Historia</b>	Construido en 1975, este pantalán fue proyectado por la Compañía Minera de Sierra Menera para la carga de mineral de hierro, procedente de las minas de Ojos Negros (Teruel), en barcos. Tras la disolución de la CMSM en 1987 pasó a ser propiedad de APV.
<b>Estado de conservación</b>	Mal estado de conservación. Se ib a poner en marcha un proceso de licitación de obras con carácter de urgencia para la restauración y consolidación de la estructura.
<b>Grado de protección</b>	<i>[Este campo lo cumplimenta el comité de evaluación de la Lista Roja]</i>
<b>Razones de inclusión</b>	Se ha derrumbado el primer tramo del pantalán, de aproximadamente 30 metros de longitud, debido a los fuertes temporales y el efecto de las olas.

Tabla D.1. Ficha de prueba (parte 1)

<b>Fuentes</b>	<p>“Pantalán del puerto de Sagunto”, Sagunto Turismo.  <a href="https://saguntoturismoblog.wordpress.com/2014/12/07/pantalan-de-puerto-de-sagunto">https://saguntoturismoblog.wordpress.com/2014/12/07/pantalan-de-puerto-de-sagunto</a></p> <p>“El alcalde de Sagunto sobre el pantalán: “Ya habíamos trasladado a APV que no estaba en buenas condiciones””, 20 Minutos.  <a href="http://20minutos.es/noticia/4079444/0/el-alcalde-de-sagunto-sobre-el-pantalan-ya-habiamos-trasladado-a-apv-que-no-estaba-en-buenas-condiciones">20minutos.es/noticia/4079444/0/el-alcalde-de-sagunto-sobre-el-pantalan-ya-habiamos-trasladado-a-apv-que-no-estaba-en-buenas-condiciones</a></p>
<b>Entidades defensa</b>	Iniciativa Porteña
<b>Contacto entidad</b>	<a href="http://www.elpuertoexiste.es">www.elpuertoexiste.es</a>

Tabla D.2. Ficha de prueba (parte 2)