LEAD-ME
Leading Platform for European Citizens, Industries, Academia and Policymakers in Media Accessibility
Funded by the Horizon 2020 Framework Programme of the European Union

Statistical Analysis of Eye Tracking Data

Krzysztof Krejtz

SWPS UNIVERSITY
UNIVERSITY OF WARSAW

LEAD-ME Summer Training School Warsaw 2021 (5-9 July 2021)
Eye tracking in media accessibility research - methods, technologies and data analyses

Funded by the Horizon 2020 Framework Programme of the European Union

Agenda

- R language & R-Studio IDE basics
- Descriptive statistics
- Moderation analysis in eye tracking studies
  - Mixed-design Analysis of Variance (ANOVA) with interaction effect
  - Pairwise comparisons
  - Graphical representations of interaction effects
- Multiple regression with interaction effect
  - Simple slopes analysis
  - Visualization of interaction effect in regression

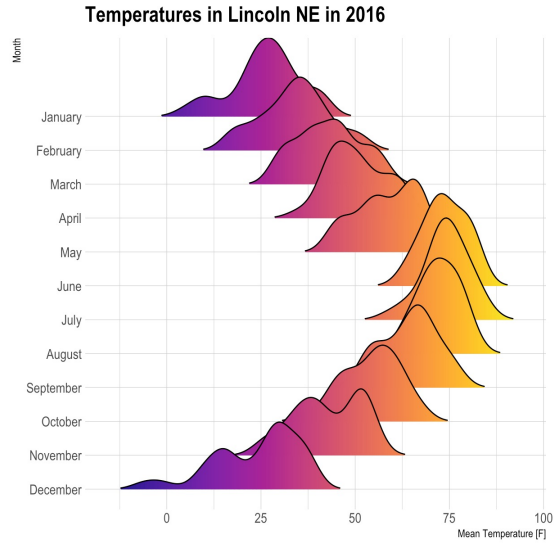
LEAD-ME

Funded by the Horizon 2020 Framework Programme of the European Union

SWPS UNIVERSITY

UNIVERSITY OF WARSAW

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY



R & R Studio Basics

LEAD-ME

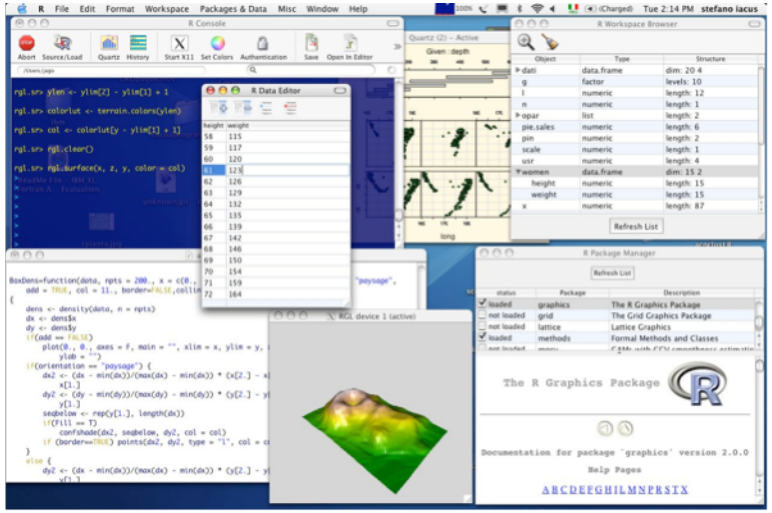
Funded by the Horizon 2020 Framework Programme of the European Union

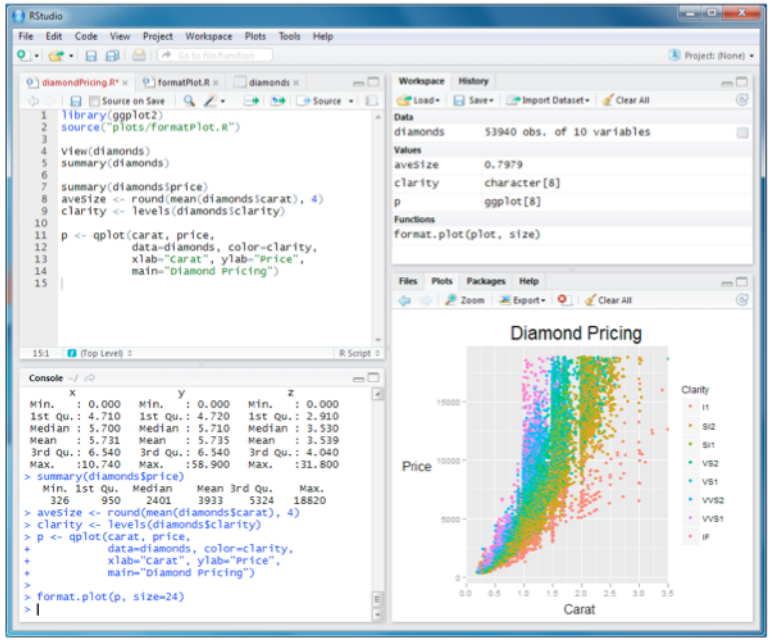
SWPS UNIVERSITY

UNIVERSITY OF WARSAW

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

R language and R Studio IDE



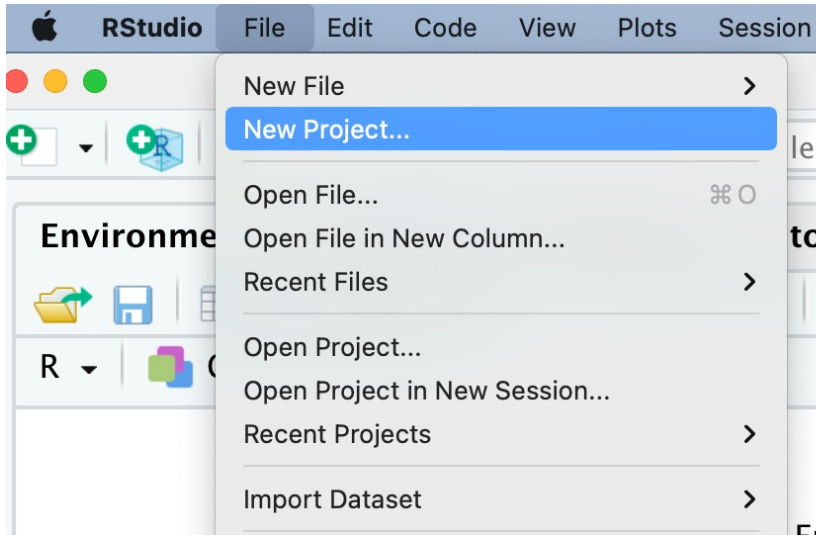


- R is a free software environment for statistical computing and graphics with large worldwide community.
  - http://www.r-project.org/
  - R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- RStudio is a free and open source Integrated Development Environment (IDE) for R.
  - http://www.rstudio.com/ide/

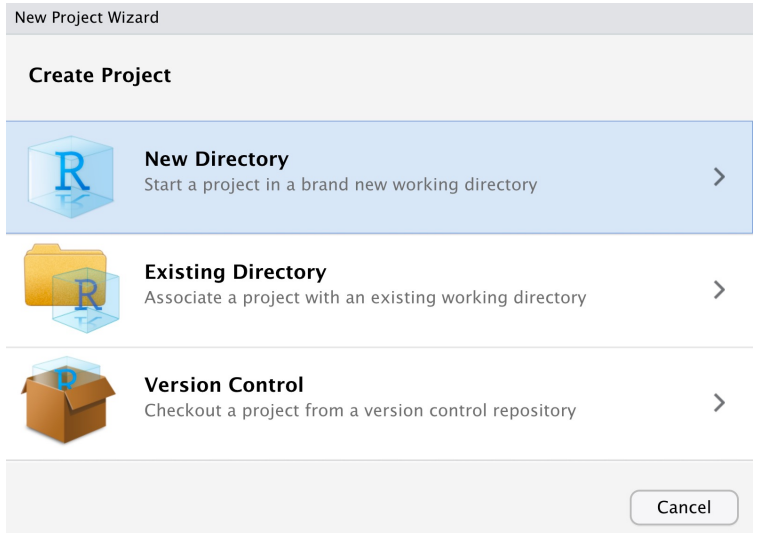
Funded by the Horizon 2020 Framework Programme of the European Union

Start a new project in 4 steps

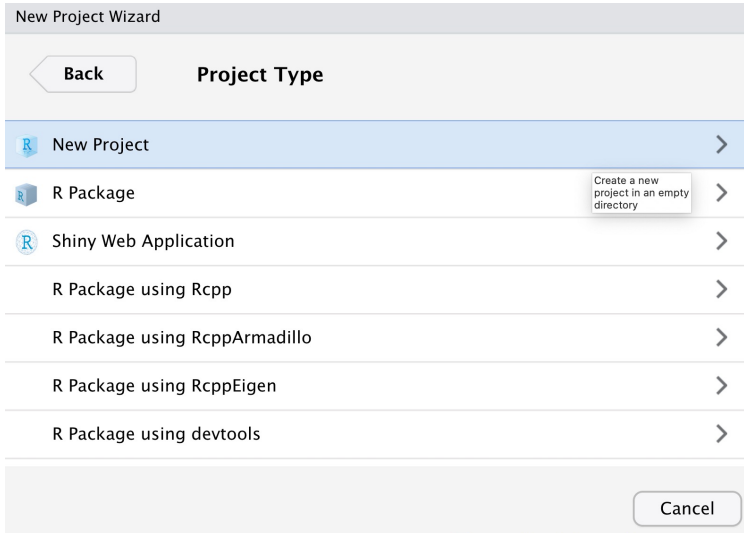
1



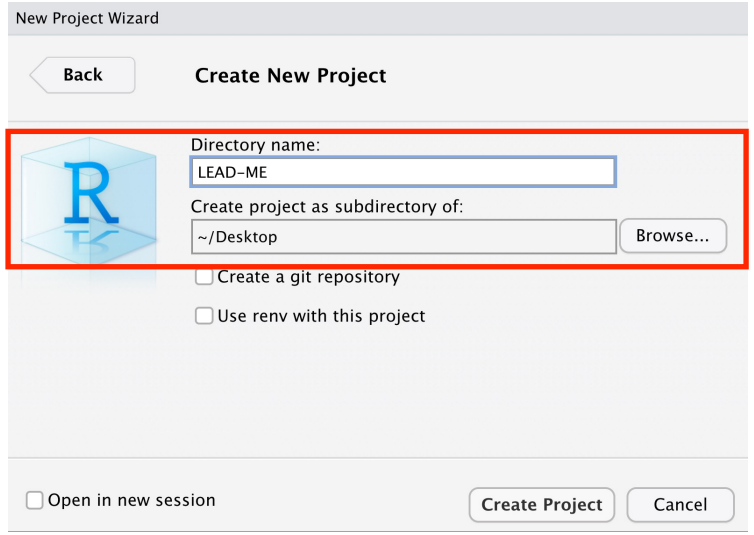
2



3



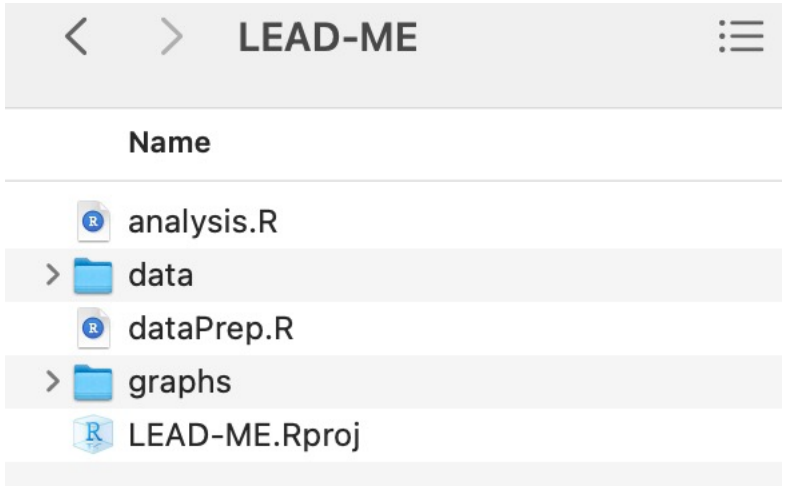
4



Funded by the Horizon 2020 Framework Programme of the European Union

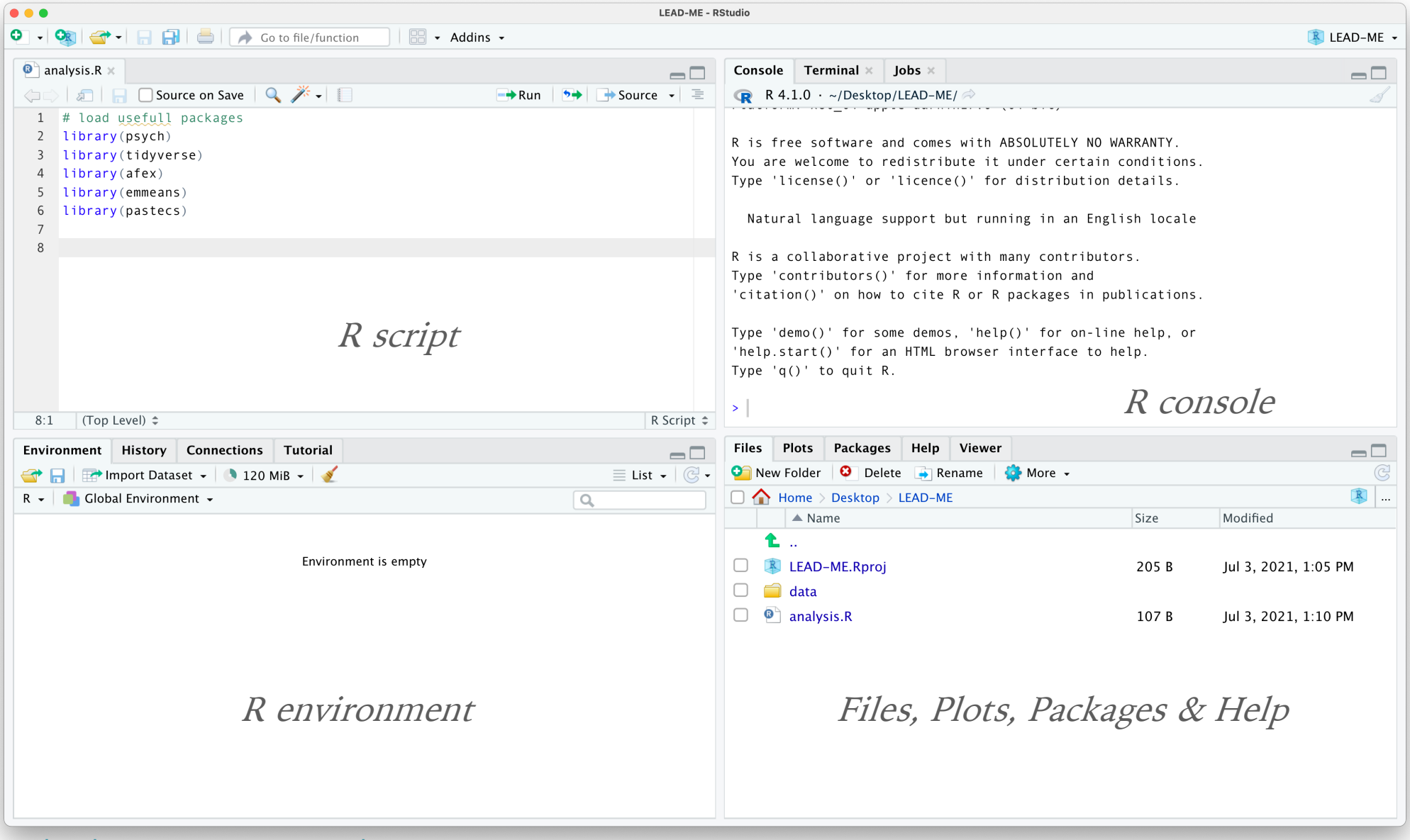
Outside RStudio

- Add data folder into your working folder
- Create 'graphs' folder in you working folder



Funded by the Horizon 2020 Framework Programme of the European Union

Complete RStudio IDE ready for statistical analyses



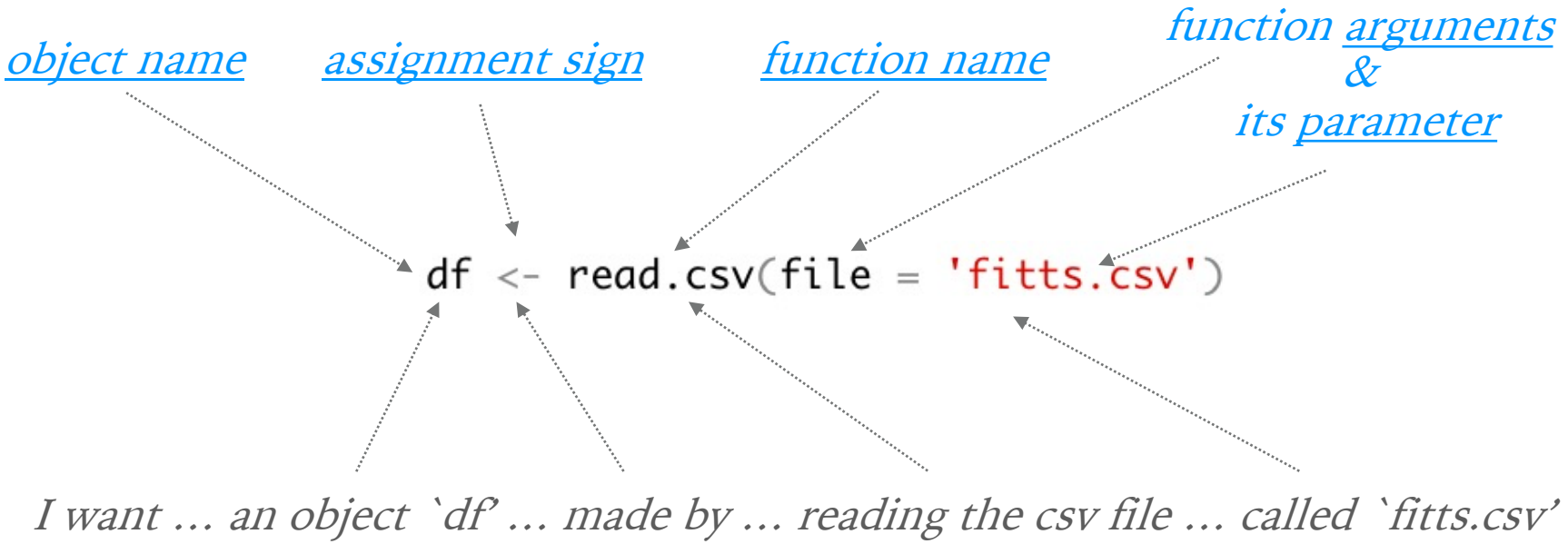
Funded by the Horizon 2020 Framework Programme of the European Union

Talking to R

```
lunch <- make.pizza(dough = "thin",
                    sauce = "tomato",
                    main.ingredient =  "mozzarella",
                    doubleCheese = FALSE)
```

Talking to R

Ordering pizza (*oops*… analyses) with R

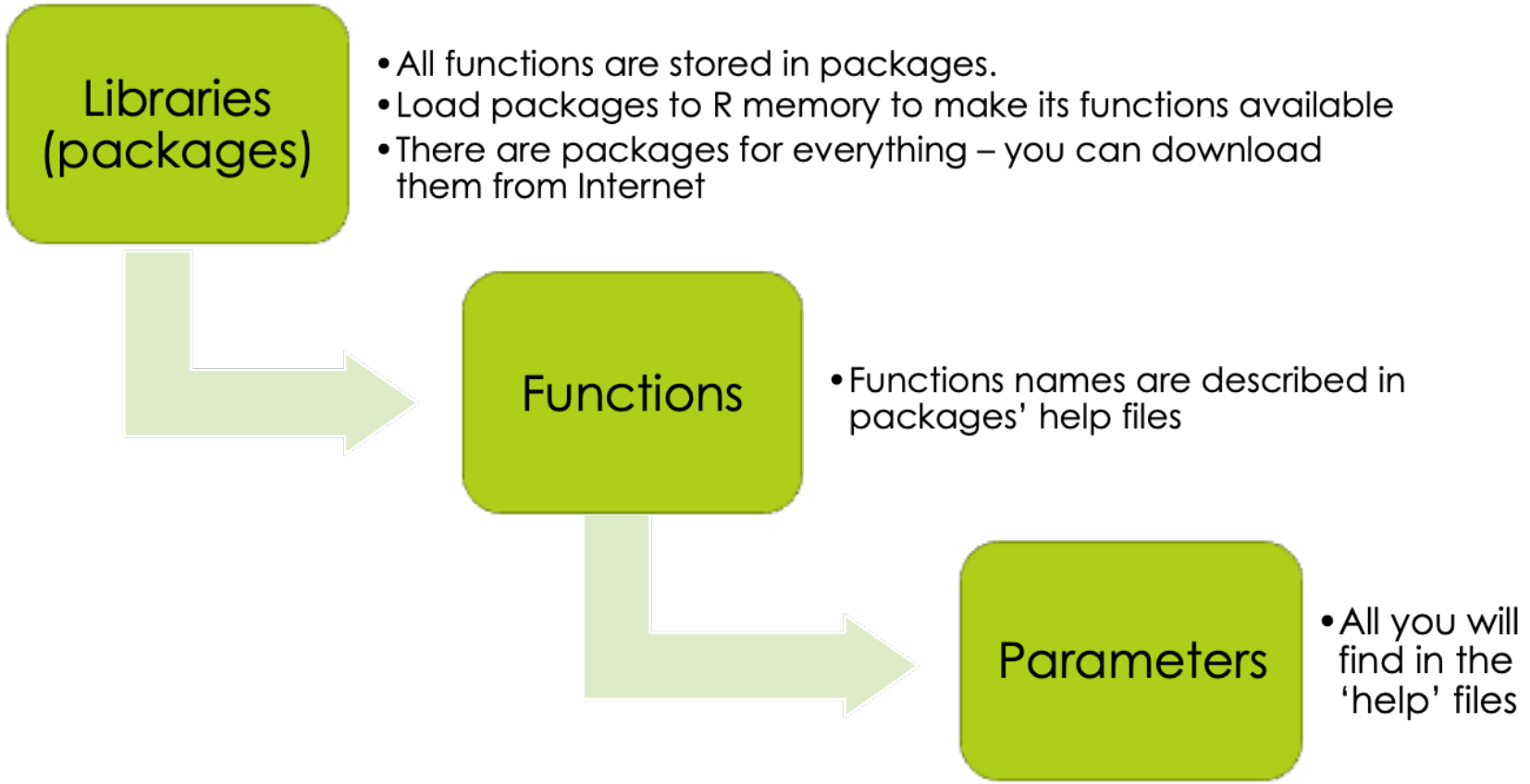


Funded by the Horizon 2020 Framework Programme of the European Union

libraries

functions

parameters



Funded by the Horizon 2020 Framework Programme of the European Union

libraries

functions

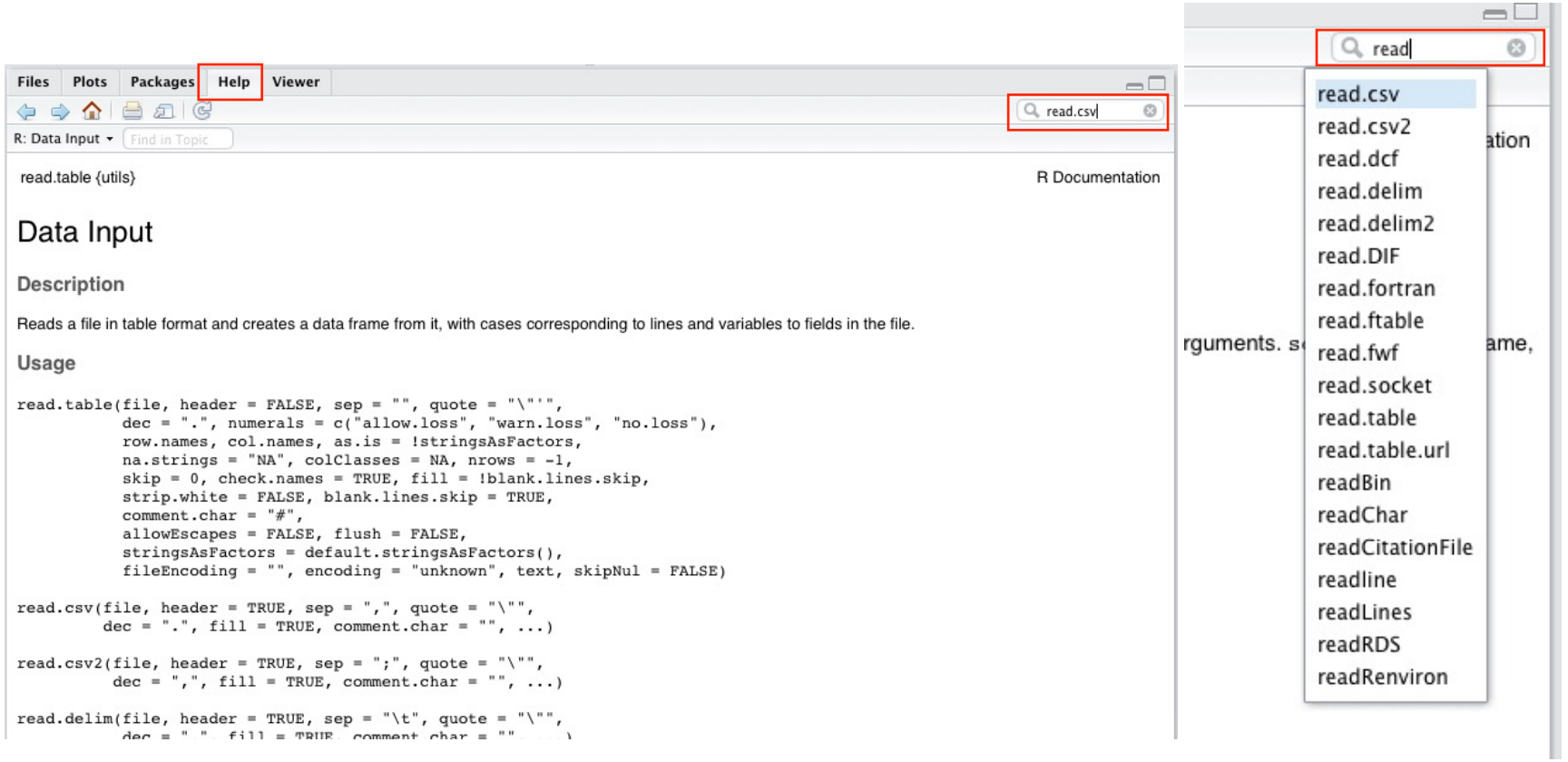
parameters

library(psych)

describe(x, na.rm = TRUE, interp=FALSE,skew = TRUE, ranges = TRUE, trim=.1, type=3, check=TRUE, fast=NULL, quant=NULL,IQR=FALSE)

Using help

Searching for functions' help: definition, arguments, examples of usage

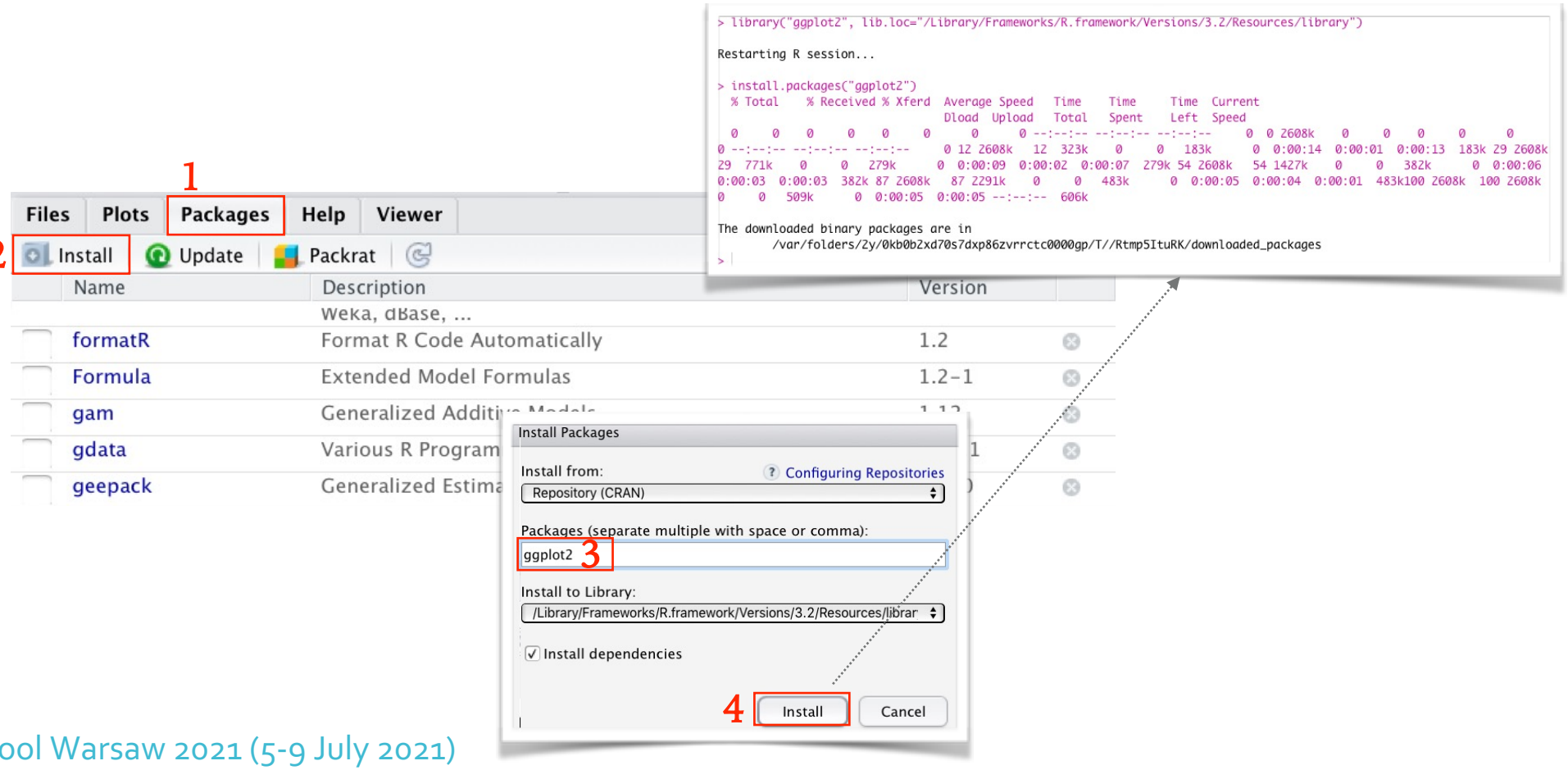


Funded by the Horizon 2020 Framework Programme of the European Union

Installing R libraries (packages)

```
install.packages(pkgs = "ggplot2", dependencies = TRUE)
```

if TRUE will install also all dependent packages



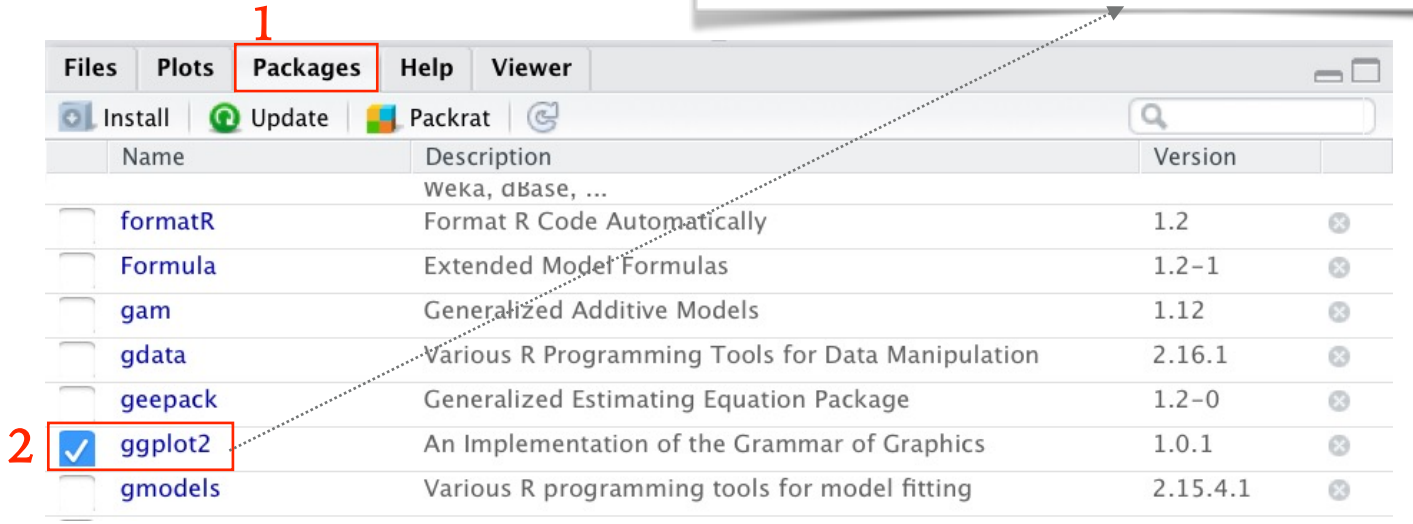
Loading installed libraries (packages)

Before using functions from the library we have to load them to the working memory.

```
library(ggplot2)
```

```
> library(ggplot2)
>
```

```
> library("ggplot2", lib.loc="/Library/Frameworks/R.framework/Versions/3.2/Resources/library")
```



Funded by the Horizon 2020 Framework Programme of the European Union

Loading large libraries

- Some libraries contain other useful packages e.g., tidyverse

```
> library(tidyverse)
── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──
✓ ggplot2 3.3.4     ✓ purrr   0.3.4
✓ tibble  3.1.2     ✓ dplyr   1.0.6
✓ tidyr   1.1.3     ✓ stringr 1.4.0
✓ readr   1.4.0     ✓ forcats 0.5.1
```

LEAD-ME

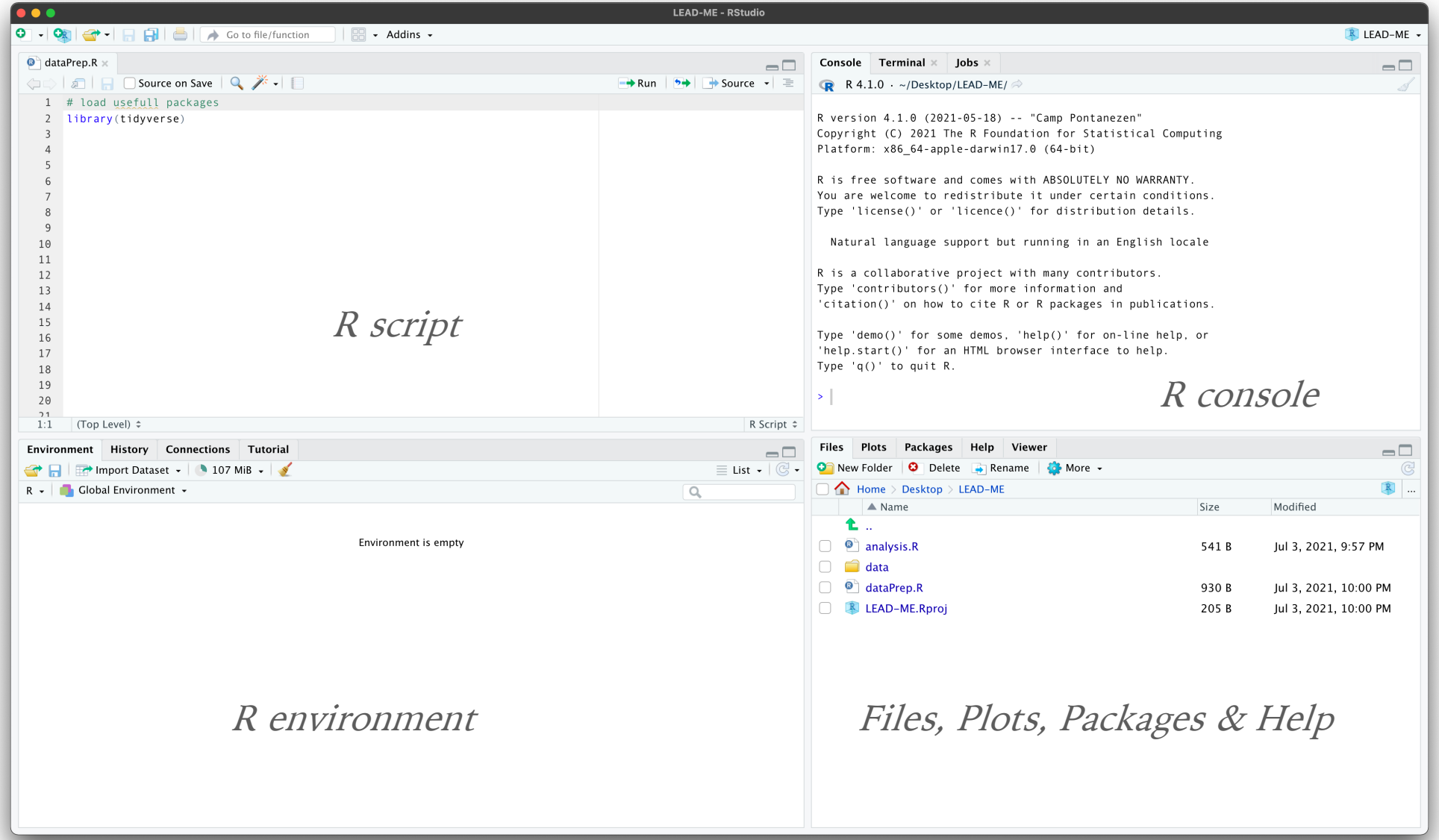
Funded by the Horizon 2020 Framework Programme of the European Union

SWPS UNIVERSITY

UNIVERSITY OF WARSAW

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

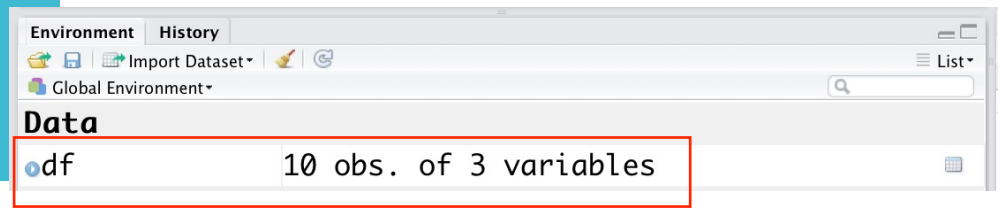
Complete RStudio IDE ready for statistical analyses



Data frames

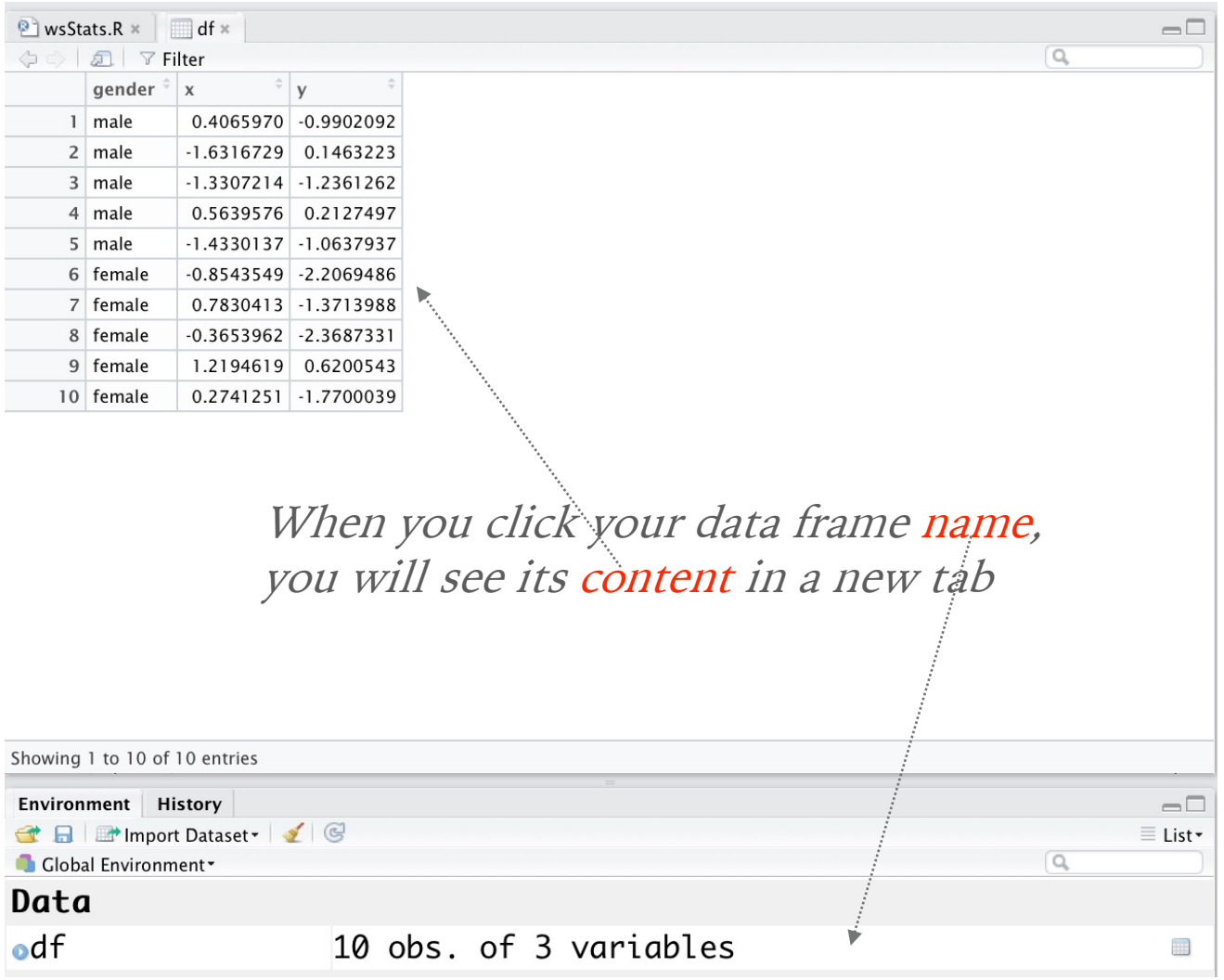
- Data frame is a quadratic table of data.
  - Different columns can be of different classes (numeric, character, factor, etc.).
  - All variables have to have the same length.
  - This is similar to SPSS datasets or other spreadsheets

```
20 # Data frames
21 x <- rnorm(10)                                  ▸ numeric vector of random 10 values
22 y <- rnorm(10)                                  ▸ numeric vector of random 10 values
23 gender <- c(rep("male", 5), rep("female", 5))   ▸ character vector of 5 „male" and 5 „female" values
24 gender <- factor(gender)                        ▸ factor of gender
25 df <- data.frame(gender, x, y)                  ▸ data frame with 3 variables: gender, x, and y
```



When done you'll see your data frame in Environment tab + information about its dimensions

Data frames



When you click your data frame name, you will see its content in a new tab

Funded by the Horizon 2020 Framework Programme of the European Union

Data frames & variables

A set of useful functions

(good to remember all of them)

```
27 class(df)
28 dim(df)
29 names(df)
30 str(df)
31 head(df)
```

information about the object class

dimensions of data frame or matrix

names of variables in the data frame

structure of data frame and all its variables

displays first 6 rows of data frame

```
> names(df)
[1] "gender" "x"      "y"
```

```
> head(df)
  gender          x          y
1   male  0.4065970 -0.9902092
2   male -1.6316729  0.1463223
3   male -1.3307214 -1.2361262
4   male  0.5639576  0.2127497
5   male -1.4330137 -1.0637937
6 female -0.8543549 -2.2069486
```

```
> str(df)
'data.frame':   10 obs. of  3 variables:
 $ gender: Factor w/ 2 levels "female","male": 2 2 2 2 2 1 1 1 1 1
 $ x     : num  0.407 -1.632 -1.331 0.564 -1.433 ...
 $ y     : num  -0.99 0.146 -1.236 0.213 -1.064 ...
```

Funded by the Horizon 2020 Framework Programme of the European Union

Data frames

Calling the variable

- Calling/using a specific variable from the data frame can be done in two ways:
  - referring to the name of the variable

    dataFrame$variableName

```
36 df$gender
```

```
> df$gender
 [1] male   male   male   male   male   female female female female female
Levels: female male
```

  - referring to the index (number) of the column which stores the variable in the data frame

    dataFrame[row, column]

```
38 df[ ,1]
```

```
> df[ ,1]
 [1] male   male   male   male   male   female female female female female
Levels: female male
```

Objects in R

- Objects are ***bricks***, which you can use for the analyses (data frames, variables, statistics results, functions, plots, etc.)
- Objects are stored temporarily in the R environment
- Objects can be loaded, eg. data frames or created by R

create the object

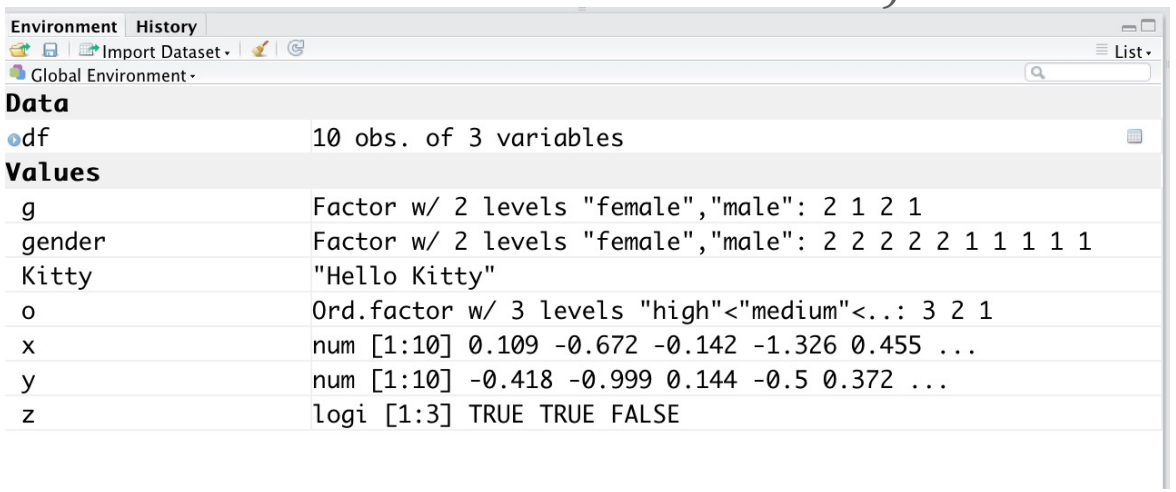
```
x <- rnorm(10)
```

object name — assignment sign — function defining object

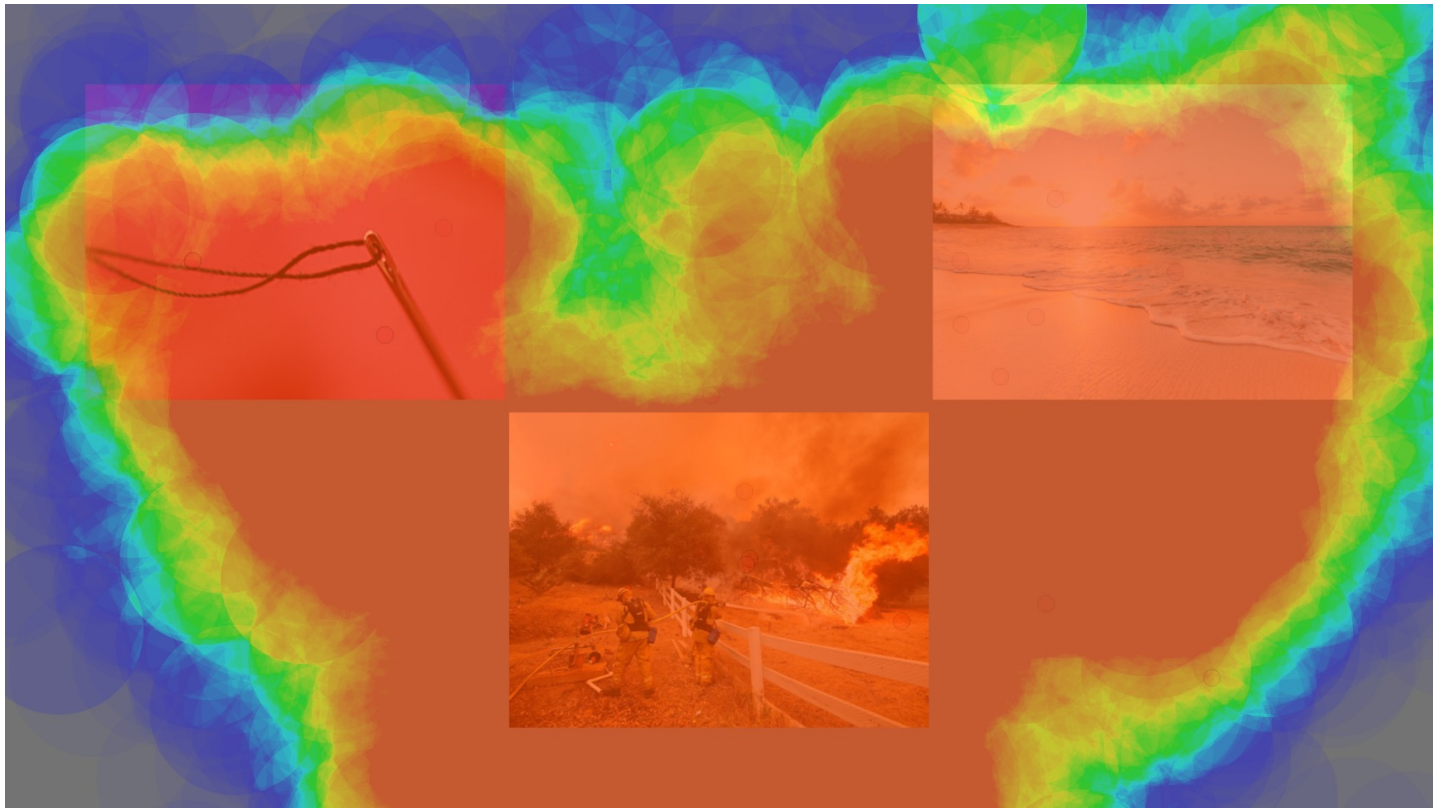
remove the object

```
rm(x)
```

the list of available objects

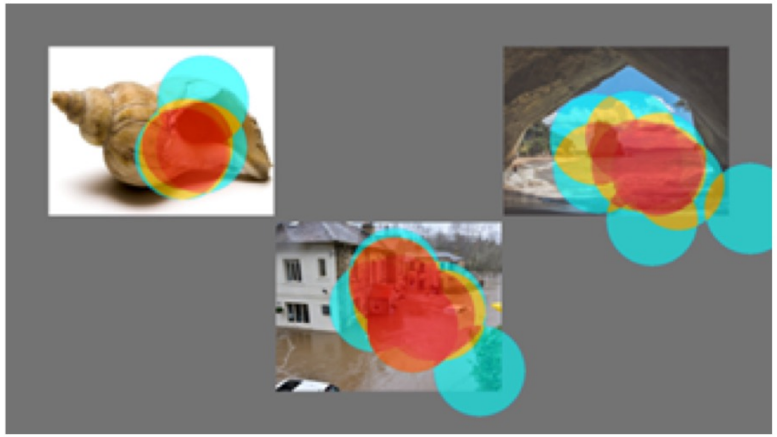


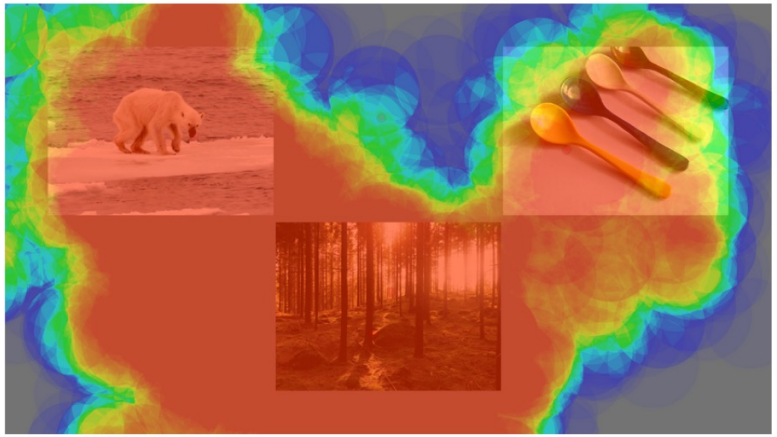
Funded by the Horizon 2020 Framework Programme of the European Union



Exemplary eye tracking project

Climate change overlooked. The role of attitudes and mood regulation in visual attention to global warming





Anna Mazurowska

SWPS University of Social Sciences and Humanities, Warsaw, Poland

Research Advisor:
dr. Krzysztof Krejtz

OBO
OŚRODEK BADAŃ OKULOGRAFICZNYCH

SWPS UNIVERSITY

Funded by the Horizon 2020 Framework Programme of the European Union

Hypothesis

Individuals with **environmental concerns** will be more likely to place visual attention on images depicting **negative consequences of climate change** than individuals without such concerns, where pro-environmental attitude is the mediator.

Funded by the Horizon 2020 Framework Programme of the European Union

Method

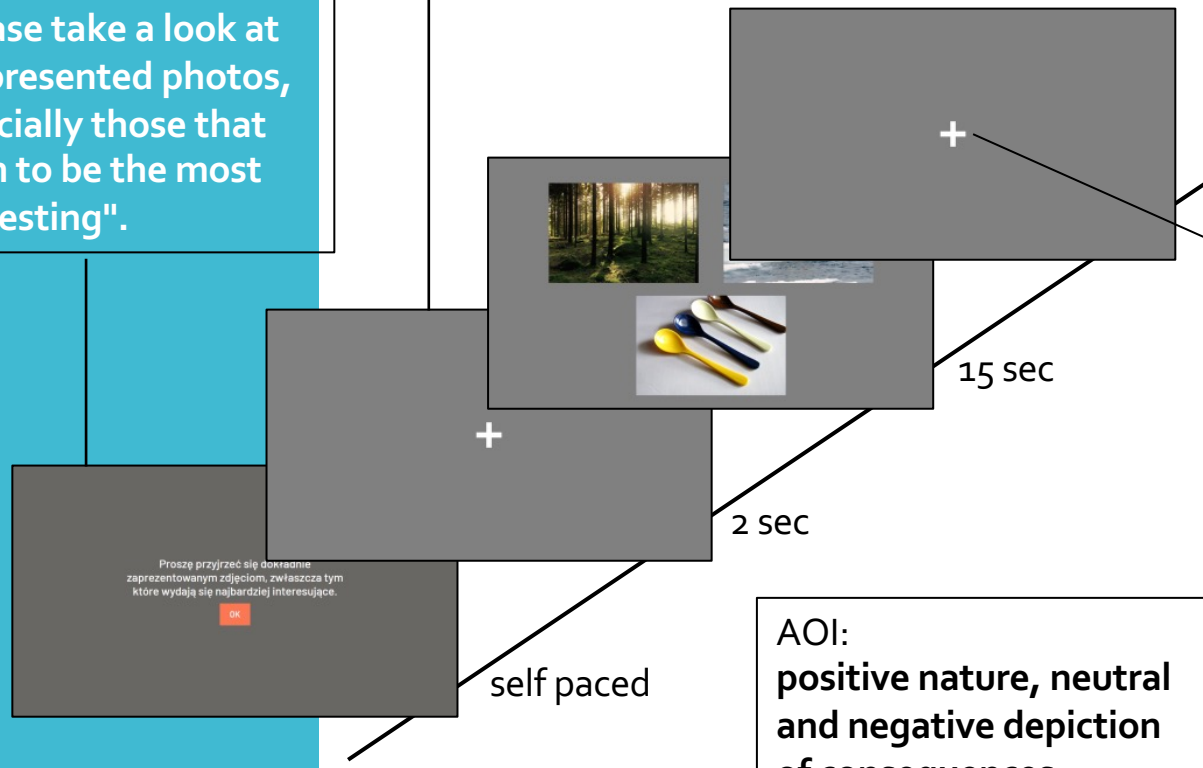
- Participants:
  - students of psychology at SWPS University (N = 48)
- Two stage online study:
  - Stage 1: Eye tracking experiment with RealEye
  - Stage 2: New Ecological Paradigm Scale (NEP) – a measure of pro-ecological attitudes (sensitivity to climate change)

Funded by the Horizon 2020 Framework Programme of the European Union

Experimantal procedure

Instruction:
"Please take a look at the presented photos, especially those that seem to be the most interesting".

Fixation point



Total 15 trials
in random order

15 sec

2 sec

AOI:
positive nature, neutral and negative depiction of consequences of climate change

Funded by the Horizon 2020 Framework Programme of the European Union

Questionnaire

Pro-environmental attitude measurement:

New Ecological Paradigm Scale (Dunlap et al. 2000), e.g.:

- "We are approaching the limit of the number of people the Earth can support".
- "Plants and animals have as much right as humans to exist".

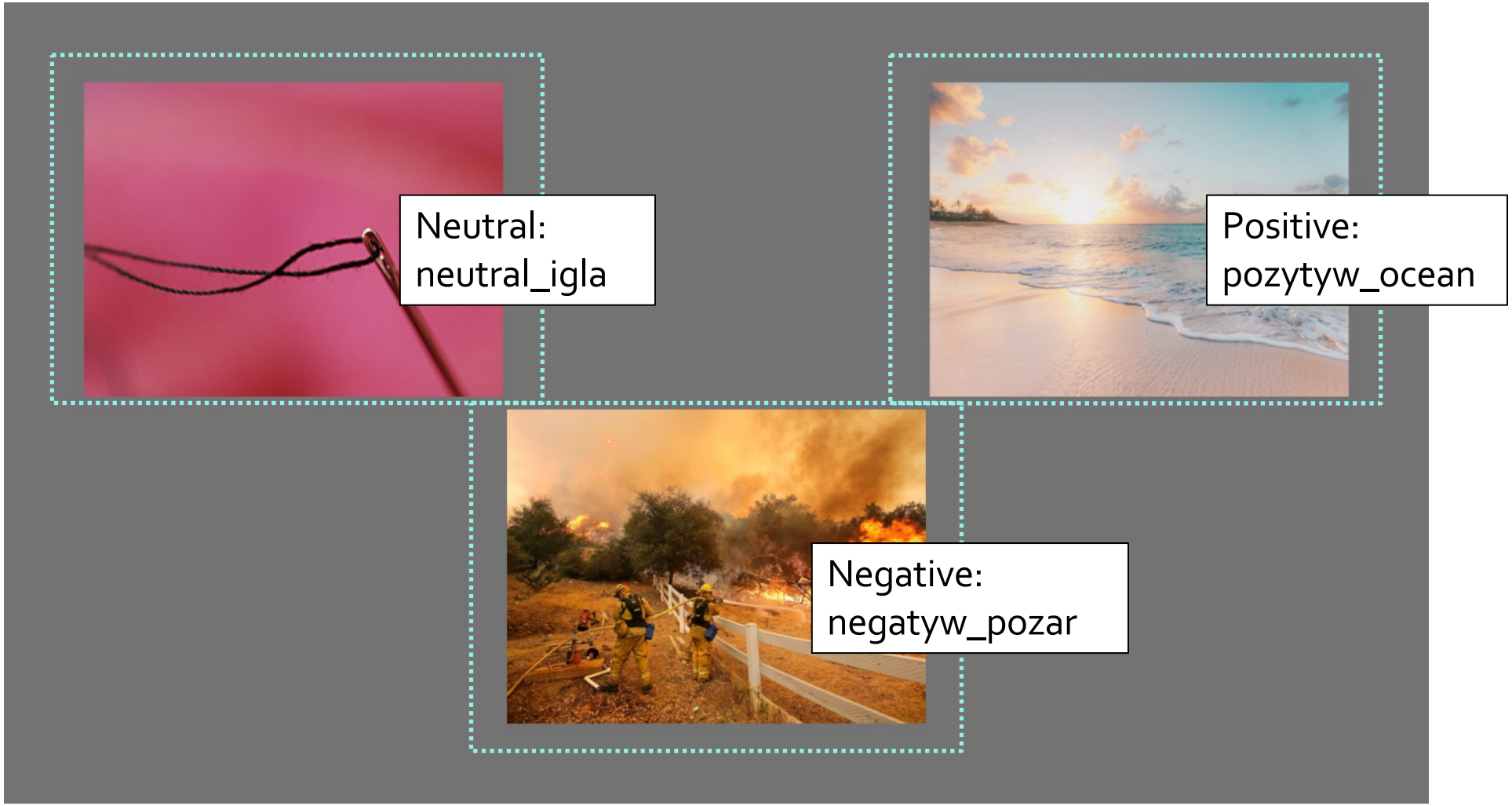
	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
We are approaching the limit of the number of people the Earth can support	○	○	○	○	◉
Plants and animals have as much right as humans to exist	○	○	○	◉	○

Funded by the Horizon 2020 Framework Programme of the European Union

Exemplary recording from RealEye

Funded by the Horizon 2020 Framework Programme of the European Union

Areas of interest definition



LEAD-ME

Funded by the Horizon 2020 Framework Programme of the European Union

SWPS UNIVERSITY

UNIVERSITY OF WARSAW

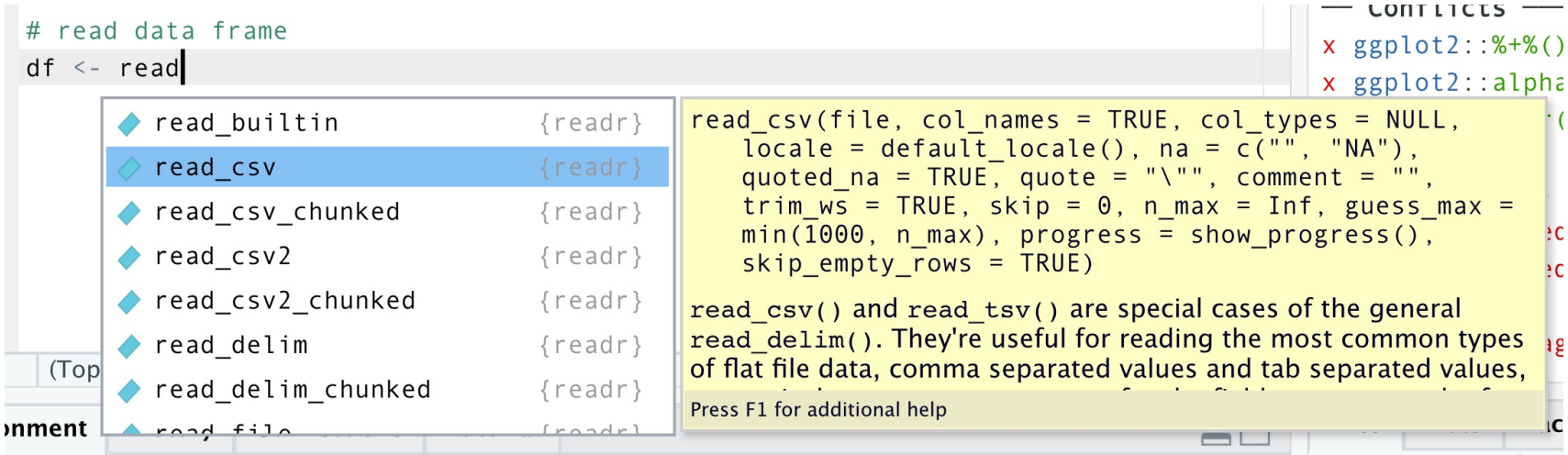
cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

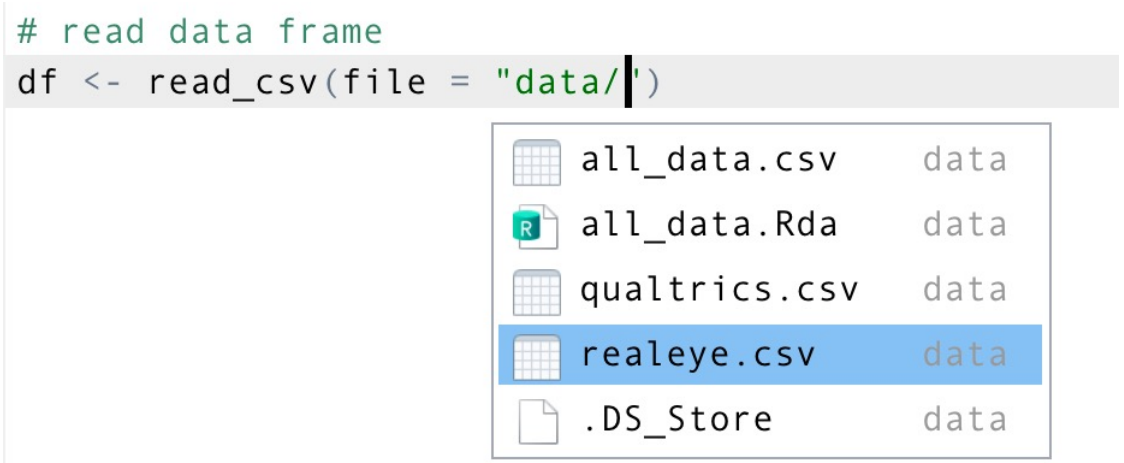
Preparing data for the analyses

read / filter / select / mutate / join / write

Funded by the Horizon 2020 Framework Programme of the European Union

Read data frame from file





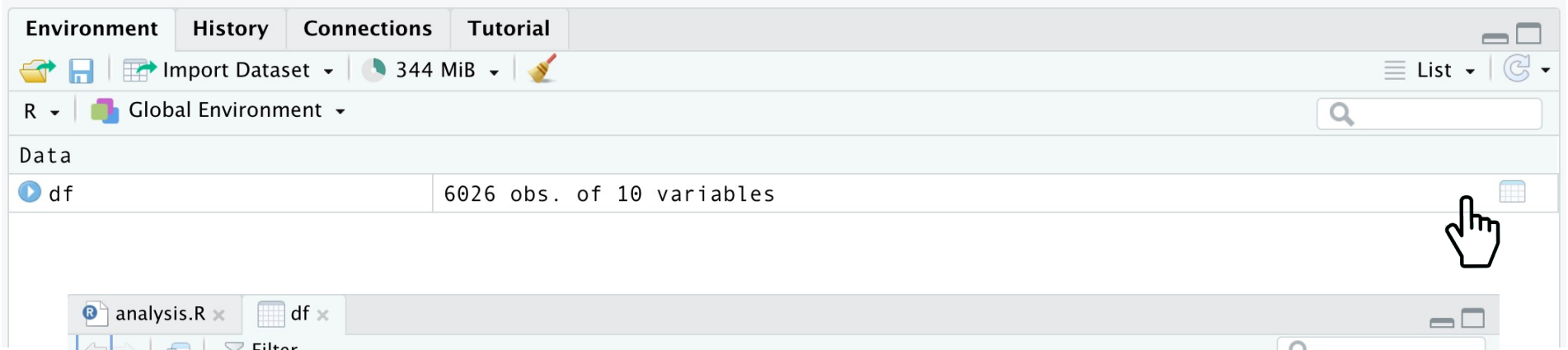
Funded by the Horizon 2020 Framework Programme of the European Union

Read data frame from file

```
> df <- read_csv(file = "data/realeye.csv")

── Column specification ────────────────────────────────────────────────────────
cols(
  ID = col_double(),
  aoi_id = col_character(),
  aoi_name = col_character(),
  aoi_size_percents = col_double(),
  aoi_fixation_total_count = col_double(),
  aoi_fixation_average_duration_ms = col_double(),
  aoi_fixation_ttff_ms = col_double(),
  aoi_fixation_average_total_time_spent_ms = col_double(),
  aoi_fixation_first_fixation_average_duration_ms = col_double(),
  notes = col_character()
)
```

Funded by the Horizon 2020 Framework Programme of the European Union

First look at the data



analysis.R | df

Filter

	ID	aoi_id	aoi_name	aoi_size_percents	aoi_fixation_total_count	aoi_fixation_averag
1	56827	7499dfa7-a9bd-4e17-a26a-9a62a9200d1f	negatyw_pozar	13.51776	7	
2	56827	c4d7557d-835a-496b-9757-be63a5a20754	neutral_igla	13.51776	7	
3	56827	7a6fb176-eed4-4ebc-871b-e1bdc483d039	pozytyw_ocean	13.48463	1	
4	63148	7499dfa7-a9bd-4e17-a26a-9a62a9200d1f	negatyw_pozar	13.51776	11	
5	63148	c4d7557d-835a-496b-9757-be63a5a20754	neutral_igla	13.51776	0	
6	63148	7a6fb176-eed4-4ebc-871b-e1bdc483d039	pozytyw_ocean	13.48463	2	
7	46018	7499dfa7-a9bd-4e17-a26a-9a62a9200d1f	negatyw_pozar	13.51776	4	
8	46018	c4d7557d-835a-496b-9757-be63a5a20754	neutral_igla	13.51776	0	
9	46018	7a6fb176-eed4-4ebc-871b-e1bdc483d039	pozytyw_ocean	13.48463	2	
10	45940	7499dfa7-a9bd-4e17-a26a-9a62a9200d1f	negatyw_pozar	13.51776	13	
11	45940	c4d7557d-835a-496b-9757-be63a5a20754	neutral_igla	13.51776	3	
12	45940	7a6fb176-eed4-4ebc-871b-e1bdc483d039	pozytyw_ocean	13.48463	12	
13	54370	7499dfa7-a9bd-4e17-a26a-9a62a9200d1f	negatyw_pozar	13.51776	10	
14	54370	c4d7557d-835a-496b-9757-be63a5a20754	neutral_igla	13.51776	2	

Showing 1 to 13 of 6,026 entries, 10 total columns

Funded by the Horizon 2020 Framework Programme of the European Union

Create AOI type independent variable

- There were three different types of AOI (positive, negative and neutral picture)
- Currently the data frame contains the following AOI names:

```
> unique(df$aoi_name)
 [1] "negatyw_pozar"       "neutral_igla"        "pozytyw_ocean"       "negatyw_lodowiec"
 [5] "neutral_kosci"       "pozytyw_gory"        "negatyw_niedzwiedz"  "neutral_lyzeczki"
 [9] "pozytyw_las"         "negatyw_podtopienia" "neutral_muszla"      "pozytyw_morze"
[13] "negatyw_susza"       "neutral_zegar"       "pozytyw_rzeka"
```

- We want to make AOI type variable which will be used as a factor in further analyses
- We will use the first element of unique AOI names as values of this new variable
- Technically we will split aoi_name variable into two columns by "_" , creating new "aoi_type" and "picture" variables

```
df <- df %>%
  separate(aoi_name, sep = "_", into = c("aoi_type", "picture"))
```

Funded by the Horizon 2020 Framework Programme of the European Union

Select & filter

- Sometimes you want to delete from your data frame some variables (columns)
- This time we want to get rid off three variables "notes", "aoi_id", and "picture"
- Also we want to remove four subjects who were not following the instruction

```
df <- df %>%
  select(-notes, -aoi_id, -picture) %>%
  filter(ID != "61567" & ID!="64828" & ID!="66001" & ID != "54127")
```

- The list of new df variables names shows that the operation was successful

```
> names(df)
[1] "ID"                                          "aoi_type"
[3] "aoi_size_percents"                           "aoi_fixation_total_count"
[5] "aoi_fixation_average_duration_ms"            "aoi_fixation_ttff_ms"
[7] "aoi_fixation_average_total_time_spent_ms"    "aoi_fixation_first_fixation_average_duration_ms"
```

Funded by the Horizon 2020 Framework Programme of the European Union

Missing values (NAs) a special case

- Missing values are annotated with NA
- NAs are not welcomed
- Always try to find out why you have NAs in your data set
- This time NAs mean that a participant did not fixate on AOI
  - It make sense to replace NAs with zeros

```
df <- df %>%
  replace_na(list(aoi_fixation_ttff_ms = 0, aoi_fixation_average_duration_ms = 0,
                  aoi_fixation_average_total_time_spent_ms = 0,
                  aoi_fixation_first_fixation_average_duration_ms = 0))
```

Funded by the Horizon 2020 Framework Programme of the European Union

Read new data file with questionnaire answers

```
# read data file with additional variables
dq <- read_csv(file = "data/qualtrics.csv")
```

```
── Column specification ──
cols(
  Q1 = col_double(),
  Q3 = col_double(),
  Q5 = col_double(),
  Q7 = col_double(),
  Q9 = col_double(),
  Q11 = col_double(),
  Q13 = col_double(),
  Q15 = col_double(),
  ID = col_double()
)
```

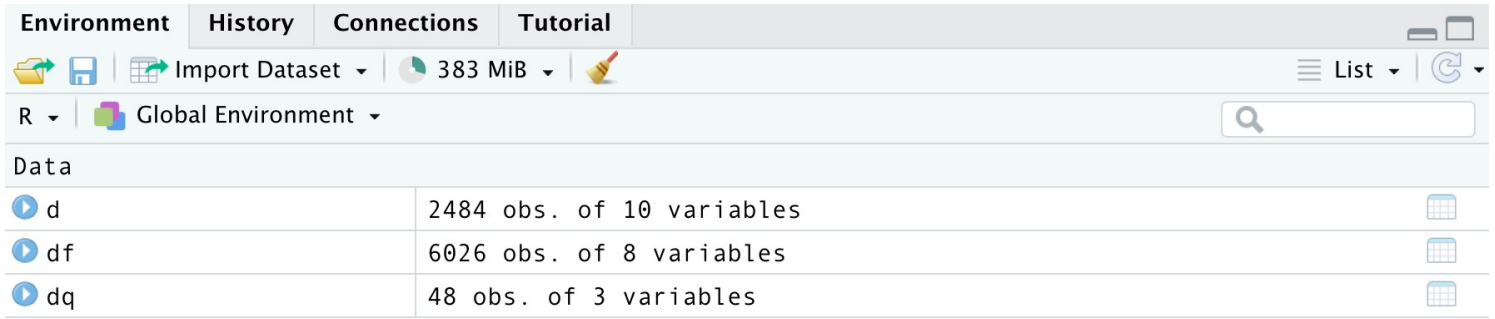
	Q1	Q3	Q5	Q7	Q9	Q11	Q13	Q15	ID
1	4	5	4	5	5	5	4	4	41428
2	4	5	2	5	4	5	3	2	64666
3	4	4	4	5	4	4	2	4	32185
4	2	2	4	2	5	4	3	3	65176
5	4	4	4	5	3	5	5	5	45652
6	1	5	5	4	5	4	2	5	60466
7	3	5	4	5	4	4	4	4	63148
8	3	4	4	4	4	4	5	4	45940
9	2	4	5	5	2	4	3	5	14539
10	3	4	4	2	4	3	4	3	29236
11	4	2	3	3	5	3	4	4	54619
12	4	4	3	4	2	5	5	4	46417
13	4	4	4	4	5	4	2	4	65905

Funded by the Horizon 2020 Framework Programme of the European Union

Joining two data frames by subject ID

- We need to join two data frames (with eye tracking data and questionnaire data)
- We will join to data frames by subject ID variable

```
# join two data frames by ID (subject ID variable)
d <- inner_join(df,dq, by = "ID")
```



- Note that new data frame "d" has lower number of rows.
- It is due to the fact that not all subjects who participated in eye tracking study completed the questionnaire.

Funded by the Horizon 2020 Framework Programme of the European Union

Calculate independent variables / factors

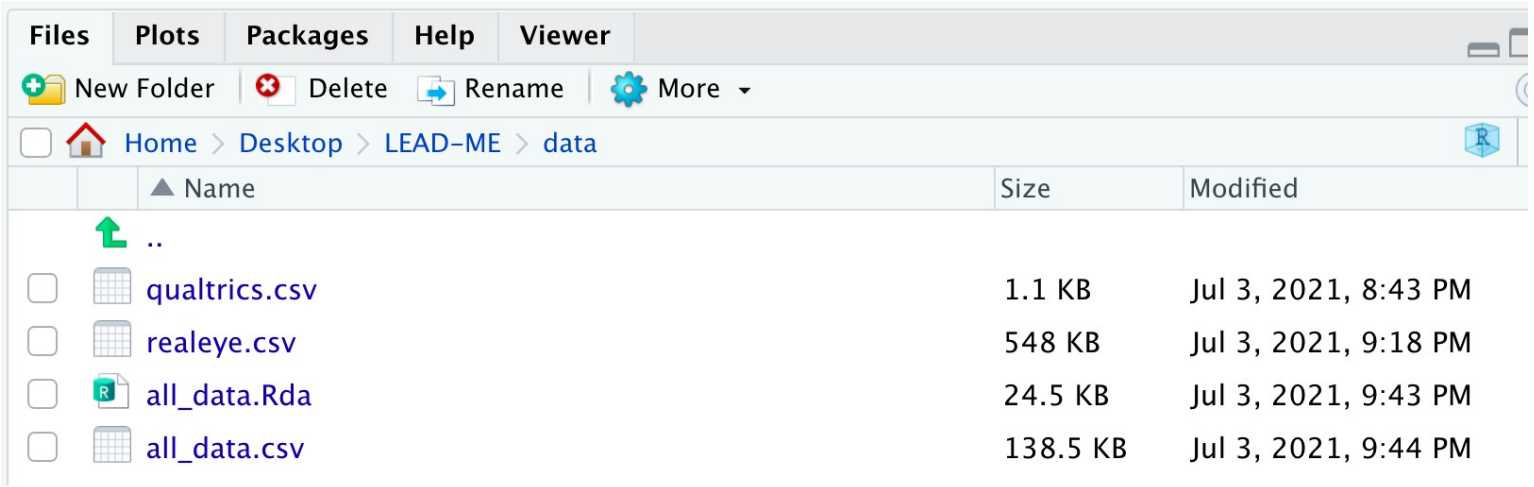
- Questionnaire data contains answers from the New Ecological Paradigm (NEP) questionnaire and **subject ID**
  - The answers were on 1-5 Likert-type scale (the higher value the higher sensitivity to climate change)
- We want to calculate one score of the sensitivity to climate change (NEP) which will be a mean of all given answers
- Next, we want to perform median split on NEP score (low vs high sensitivity to climate change) to use it as a factor in further analyses.
- We need also to set aoi_type as factor and make "neutral" value of it as a reference point (important for statistical analyses)
- Last, we do not need all raw answers to each NEP question (they all starts with "Q").

```
d <- d %>%
  mutate(NEP = rowSums(select(.,starts_with("Q")))) %>%
  mutate(NEPsplit = cut(NEP, 2, labels = c("low", "high"))) %>%
  mutate(aoi_type = factor(aoi_type)) %>%
  mutate(aoi_type = relevel(aoi_type, ref = "neutral")) %>%
  select(-starts_with("Q"))
```

Write data files

- Save the entire data frame into a new file
- We can do it in several formats. The most useful are:
  - .csv files (great for sharing even with those who do not use R)
  - Data format (.Rda) - great for further use within R and hard drive space saver

```
# write the entire data base into RData file
save(d, file = "data/all_data.Rda")
# or csv file
write_csv(d, file = "data/all_data.csv")
```



LEAD-ME

Funded by the Horizon 2020 Framework Programme of the European Union

SWPS UNIVERSITY

UNIVERSITY OF WARSAW

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

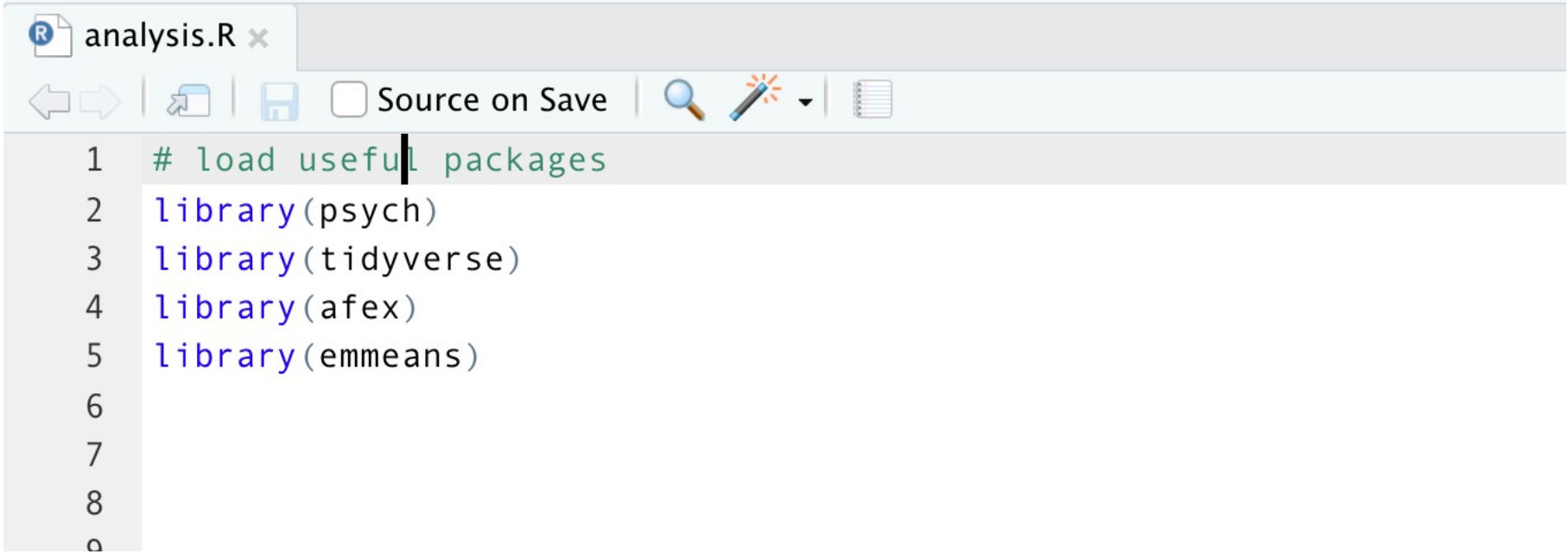
Descriptive statistics

With visualisations

Funded by the Horizon 2020 Framework Programme of the European Union

Create new R script "analysis.R"

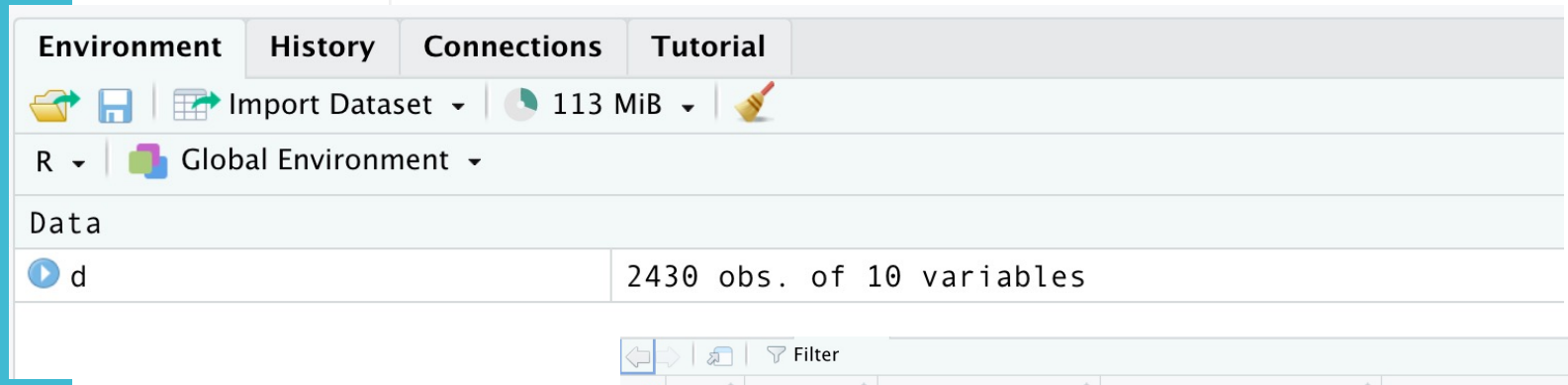
- It is good to have several R scripts for different purposes.
- We will create a new R script named "analysis.R"
- Start the script with useful libraries



Read data frame of RData format

- Reading RData format will automatically load data frame name

```
# read data frame from RData format
load("data/all_data.Rda")
```



Filter

	ID	aoi_type	aoi_size_percents	aoi_fixation_total_count	aoi_fixation_average_duration_ms	aoi_fixation_ttff_ms	a
1	63148	negatyw	13.51776	11	155	520	
2	63148	neutral	13.51776	0	NA	NA	
3	63148	pozytyw	13.48463	2	149	7384	
4	45940	negatyw	13.51776	13	192	1106	
5	45940	neutral	13.51776	3	150	747	
6	45940	pozytyw	13.48463	12	173	4859	
7	54370	negatyw	13.51776	10	145	767	
8	54370	neutral	13.51776	2	137	540	
9	54370	pozytyw	13.48463	9	151	1343	
10	61567	negatyw	13.51776	2	157	2392	
11	61567	neutral	13.51776	4	148	973	

Funded by the Horizon 2020 Framework Programme of the European Union

Describe data frame

- Summary is a generic function used to produce summaries of the results of various model fitting functions.
- The function invokes particular methods which depend on the class of the first argument.
- When applied to data frame it returns basic descriptive statistics for all numerical variables

```
> summary(d)
       ID          aoi_type         aoi_size_percents aoi_fixation_total_count aoi_fixation_average_duration_ms aoi_fixation_ttff_ms
 Min.   :10564   Length:2430        Min.   :11.97     Min.   : 0.000           Min.   :100.0                    Min.   :  500
 1st Qu.:41710   Class :character   1st Qu.:12.18     1st Qu.: 1.000           1st Qu.:144.0                    1st Qu.: 1044
 Median :55152   Mode  :character   Median :12.30     Median : 4.000           Median :162.0                    Median : 1918
 Mean   :51814                      Mean   :12.46     Mean   : 6.779           Mean   :169.1                    Mean   : 3119
 3rd Qu.:63148                      3rd Qu.:12.80     3rd Qu.:10.000           3rd Qu.:183.0                    3rd Qu.: 3998
 Max.   :67201                      Max.   :13.52     Max.   :44.000           Max.   :583.0                    Max.   :14882
                                                                               NA's   :486                      NA's   :486
 aoi_fixation_average_total_time_spent_ms aoi_fixation_first_fixation_average_duration_ms      NEP        NEPsplit
 Min.   :  100                           Min.   : 100.0                                    Min.   :3.125   low :1125
 1st Qu.:  450                           1st Qu.: 123.0                                    1st Qu.:3.750   high:1305
 Median : 1048                           Median : 143.0                                    Median :4.125
 Mean   : 1492                           Mean   : 163.9                                    Mean   :4.113
 3rd Qu.: 2020                           3rd Qu.: 179.0                                    3rd Qu.:4.500
 Max.   :13419                           Max.   :2051.0                                    Max.   :5.000
 NA's   :486                             NA's   :486
```

Can be hard to read however

Funded by the Horizon 2020 Framework Programme of the European Union

Descriptive statistics of selected variables

The **describe** function in the **psych** package is meant to produce the most frequently requested stats in psychometric and psychology studies, and to produce them in an easy to read data.frame.

```
> describe(d[,4:8])
                                              vars    n    mean      sd median trimmed     mad min   max range skew kurtosis    se
aoi_fixation_total_count                         1 2430    6.78    7.42    4.0    5.50    5.93   0    44    44 1.52     2.34  0.15
aoi_fixation_average_duration_ms                 2 1944  169.09   42.81  162.0  163.86   28.17 100   583   483 3.16    18.95  0.97
aoi_fixation_ttff_ms                             3 1944 3119.15 3063.71 1918.5 2495.53 1650.88 500 14882 14382 1.80     2.84 69.49
aoi_fixation_average_total_time_spent_ms         4 1944 1492.11 1536.47 1048.5 1230.19 1027.44 100 13419 13319 2.86    14.39 34.85
aoi_fixation_first_fixation_average_duration_ms  5 1944  163.92   79.41  143.0  151.01   35.58 100  2051  1951 8.71   169.99  1.80
```

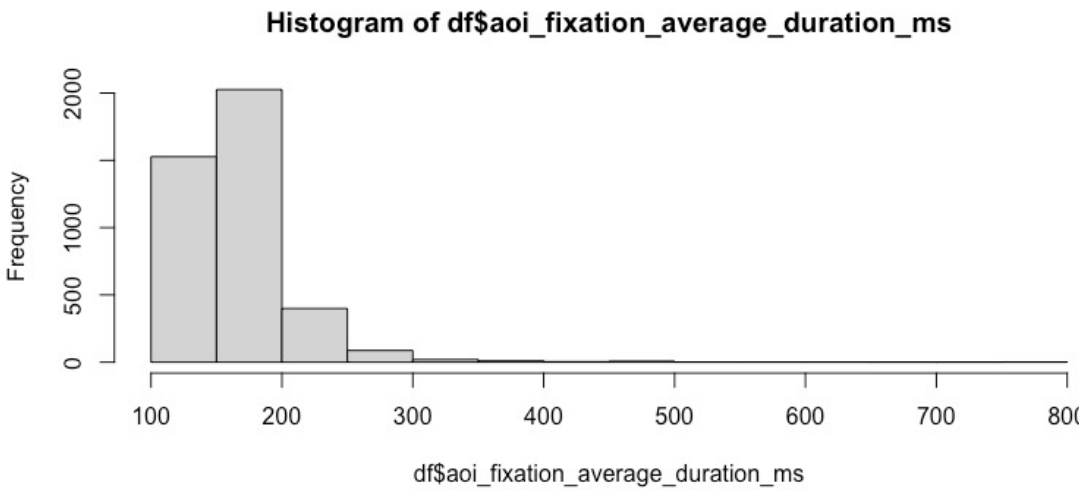
The **fast=TRUE** option will lead to a speed up of about 50% for larger data sets

It is also easy to read!

```
> describe(d[,4:8], fast = TRUE)
                                              vars    n    mean      sd min   max range    se
aoi_fixation_total_count                         1 2430    6.78    7.42   0    44    44  0.15
aoi_fixation_average_duration_ms                 2 1944  169.09   42.81 100   583   483  0.97
aoi_fixation_ttff_ms                             3 1944 3119.15 3063.71 500 14882 14382 69.49
aoi_fixation_average_total_time_spent_ms         4 1944 1492.11 1536.47 100 13419 13319 34.85
aoi_fixation_first_fixation_average_duration_ms  5 1944  163.92   79.41 100  2051  1951  1.80
```

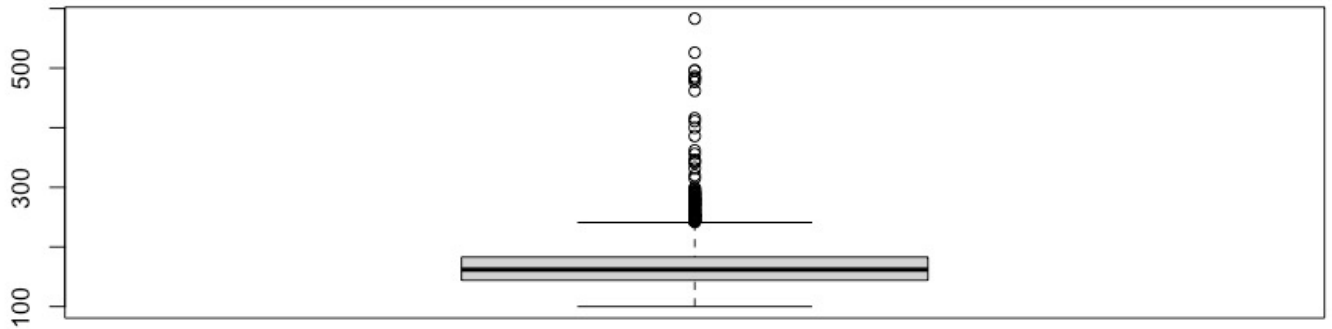
Funded by the Horizon 2020 Framework Programme of the European Union


```
# plot distribution of variable values with histogram
hist(d$aoi_fixation_average_duration_ms)
```



Useful to visually inspect normality of distribution

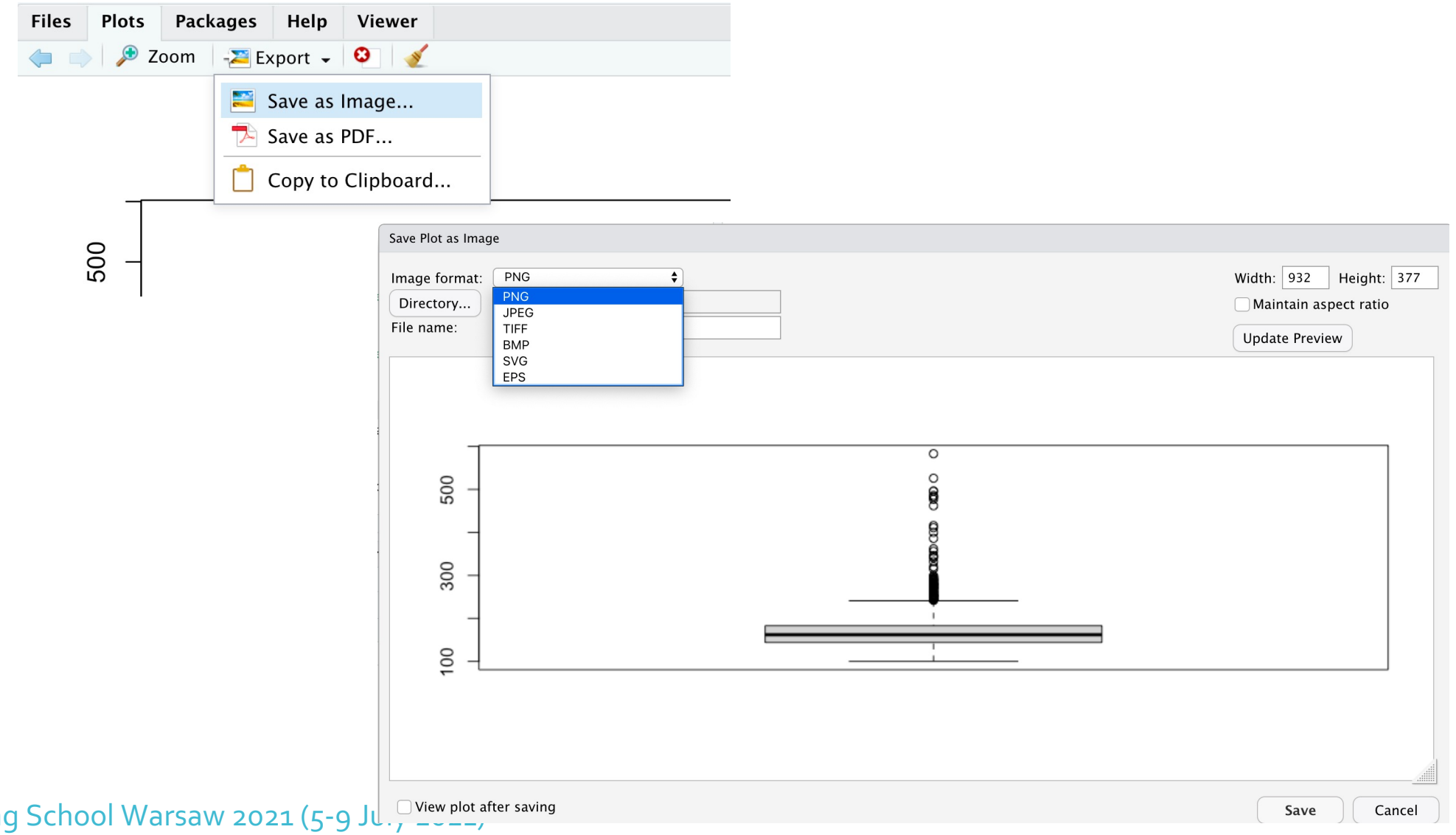
```
# plot distribution of variable values with boxplot
boxplot(d$aoi_fixation_average_duration_ms)
```



Useful to identify outlying values

Funded by the Horizon 2020 Framework Programme of the European Union

Saving graphs in RStudio graphical interface



LEAD-ME

Funded by the Horizon 2020 Framework Programme of the European Union

SWPS UNIVERSITY

UNIVERSITY OF WARSAW

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Moderation analysis with ANOVA

Mixed-design Analysis of Variance (ANOVA)

Funded by the Horizon 2020 Framework Programme of the European Union

Basic concept of ANOVA

- ANalysis Of VAriance generalizes t-tests to more than two groups.
- Tests variation among the means of several groups
  - Provides a statistical test of whether or not the means of several groups are all equal, or is there a difference between them?
- Null hypothesis
  - Several populations have the same means

Funded by the Horizon 2020 Framework Programme of the European Union

Variability of results and its sources

- **Between group variability**
  - Differences between group means
- **Within group variability / error variance**
  - Differences between each score in the sample and the sample mean

- **Between group variance** can come from:
  - The effect of IV☺️
    - Differences between group means – our expected effect
  - Individual differences ☹️
    - Everyone is different, different motivations, reactions to tasks, etc.
  - Measurement error ☹️
    - Nobody's perfect – our studies neither, non standardised instructions

Funded by the Horizon 2020 Framework Programme of the European Union

F - ratio

- Ratio of the between-groups variance (explained variability) to the within-groups variance (unexplained variability)

$$\text{statistic} = \frac{\text{between-groups variability}}{\text{error variability}}$$

- When the F is below or close to 1 …
- Reducing the error variability in the denominator of the equation will result in a larger computed statistical value, thereby making it easier to reject the Null hypothesis.

Funded by the Horizon 2020 Framework Programme of the European Union

Usefull libraries for ANOVA analysis

```
library(tidyverse)
library(afex)
library(emmeans)
```

```
> library(tidyverse)
── Attaching packages ─────────────────────
────────── tidyverse 1.3.0 ──
✓ ggplot2 3.3.3     ✓ purrr   0.3.4
✓ tibble  3.0.5     ✓ dplyr   1.0.3
✓ tidyr   1.1.2     ✓ stringr 1.4.0
✓ readr   1.4.0     ✓ forcats 0.5.0
```

Funded by the Horizon 2020 Framework Programme of the European Union

Load data

```
# read data frame from RData format
load("data/all_data.Rda")
```

```
> names(d)
 [1] "ID"
 [2] "aoi_type"
 [3] "aoi_size_percents"
 [4] "aoi_fixation_total_count"
 [5] "aoi_fixation_average_duration_ms"
 [6] "aoi_fixation_ttff_ms"
 [7] "aoi_fixation_average_total_time_spent_ms"
 [8] "aoi_fixation_first_fixation_average_duration_ms"
 [9] "NEP"
[10] "NEPsplit"
```

ID	aoi_type	aoi_size_percents	aoi_fixation_total_count	aoi_fixation_average_duration_ms	a
63148	negatyw	13.51776	11	155	
63148	neutral	13.51776	0	0	
63148	pozytyw	13.48463	2	149	
45940	negatyw	13.51776	13	192	
45940	neutral	13.51776	3	150	
45940	pozytyw	13.48463	12	173	
54370	negatyw	13.51776	10	145	
54370	neutral	13.51776	2	137	

Funded by the Horizon 2020 Framework Programme of the European Union

Simple function for ANOVA in R

- To run ANOVA in R you have plenty of options.
- We are going to use the **aov_ez()** function within the **afex package.**

```
library(afex)
```

afex package provides convenience functions for analyzing factorial experiments using ANOVA or mixed models. aov_ez(), aov_car(), and aov_4() allow specification of between, within, or mixed ANOVAs for data in long format.

- Generic form of **aov_ez()** function

```
aov_ez(id = "PARTICIPANTS ID VARIABLE",
       dv = "DEPENDENT VARIABLE",
       data = "DATA FRAME",
       between = "BETWEEN-SUBJECTS FACTOR",
       within = "WITHIN-SUBJECTS FACTOR",
       covariate = "BETWEEN-SUBJECTS COVARIATES",
       type = "ERROR TYPE - DEFAULT SET TO 3")
```

Funded by the Horizon 2020 Framework Programme of the European Union

Test the hypotheses about fixation duration

- We want to test two hypotheses:
  - In general, negative and positive pictures of environment will evoke longer fixation durations than neutral ones.
  - Negative and positive pictures of environment will evoke longer fixation durations than neutral ones BUT ONLY among people with high sensitivity to climate change.
    - *sensitivity to climate change will be a moderator of the effect of picture valence*

Funded by the Horizon 2020 Framework Programme of the European Union

Run two-way mixed-design ANOVA

RUN ANOVA TEST

```
# fixation count dependent on aoi valence and NEP score
fit <- aov_ez(data = d,
              id = "ID",
              dv = "aoi_fixation_average_duration_ms",
              within = "aoi_type",
              between = "NEPsplit")
```

Data

Subject ID variable

Dependent variable

Within-subject factor

Between-subject factor

PRINT ANOVA TEST RESULTS

```
> print(fit)
Anova Table (Type 3 tests)

Response: aoi_fixation_average_duration_ms
              Effect          df     MSE         F  ges p.value
1          NEPsplit        1, 42 1300.67      0.39 .005    .535
2          aoi_type 1.83, 76.89  546.00 10.54 *** .098   <.001
3 NEPsplit:aoi_type 1.83, 76.89  546.00    2.59 + .026    .086
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘+’ 0.1 ‘ ’ 1

Sphericity correction method: GG
```

Funded by the Horizon 2020 Framework Programme of the European Union

Estimated means and pairwise comparisons for a main effect of picture valence

```
em <- emmeans(fit, ~aoi_type)
```

```
> print(em)
 aoi_type emmean   SE  df lower.CL upper.CL
 neutral     127 4.19 102      118      135
 negatyw     141 4.19 102      133      150
 pozytyw     148 4.19 102      140      157

Results are averaged over the levels of: NEPsplit
Warning: EMMs are biased unless design is perfectly balanced
Confidence level used: 0.95
```

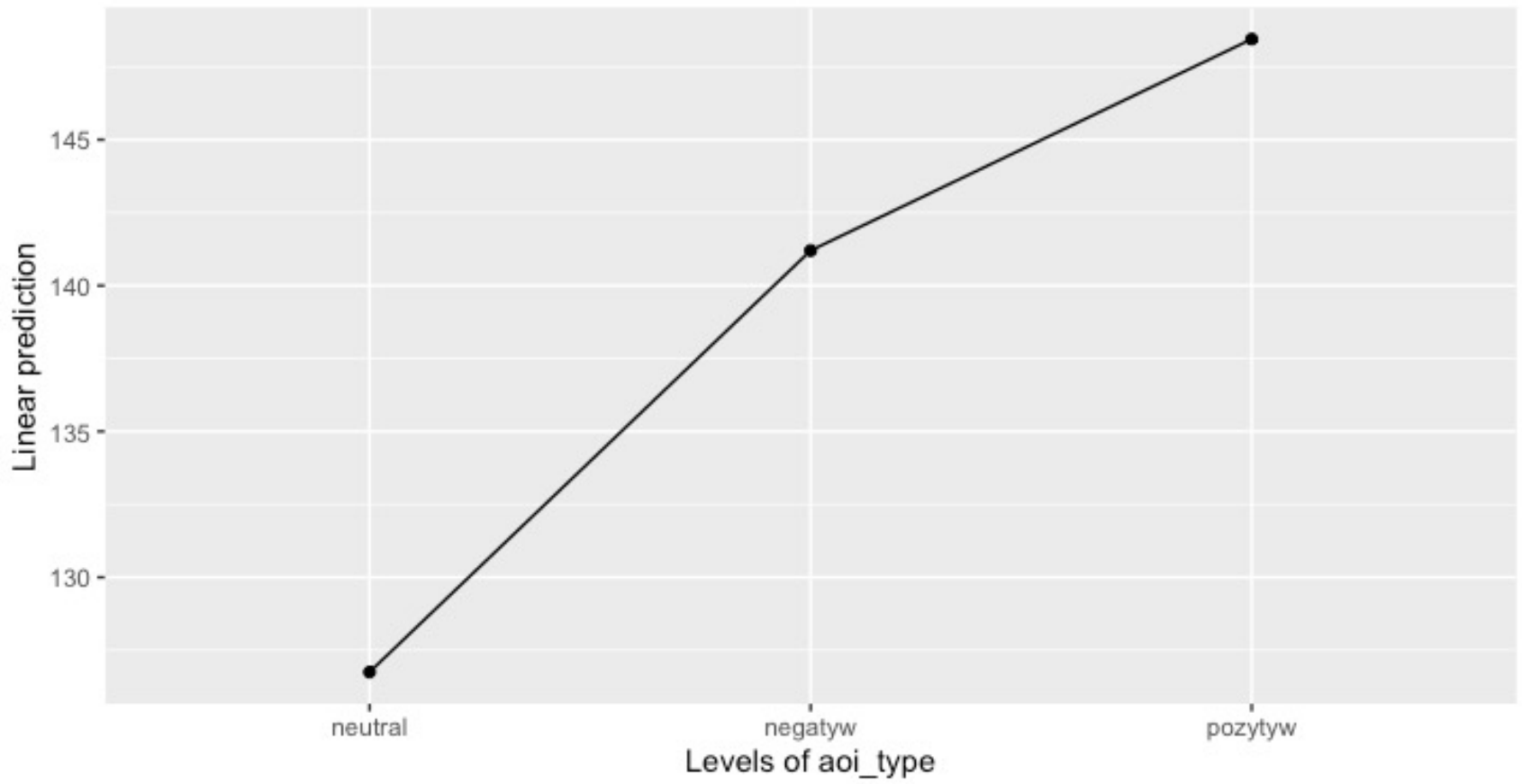
```
# post-hoc for main effect
pairs(em)
```

```
> pairs(em)
 contrast          estimate   SE df t.ratio p.value
 neutral - negatyw   -14.45 4.81 84 -3.003  0.0098
 neutral - pozytyw   -21.70 4.81 84 -4.510  0.0001
 negatyw - pozytyw    -7.25 4.81 84 -1.506  0.2931

Results are averaged over the levels of: NEPsplit
P value adjustment: tukey method for comparing a family of 3 estimates
```

Funded by the Horizon 2020 Framework Programme of the European Union

Analytical graph of the main effect

```
emmip(fit, ~ aoi_type)
```



Funded by the Horizon 2020 Framework Programme of the European Union

Estimated means for interaction effect

```
em <- emmeans(fit, ~aoi_type:NEPsplit)
```

```
> print(em)
 aoi_type NEPsplit emmean   SE    df lower.CL upper.CL
 neutral  low         134 6.12 107.1      122      147
 negatyw  low         138 6.12 107.1      126      150
 pozytyw  low         150 6.12 107.1      138      162
 neutral  high        119 5.76  95.7      108      131
 negatyw  high        144 5.76  95.7      133      156
 pozytyw  high        147 5.76  95.7      136      158
```

Funded by the Horizon 2020 Framework Programme of the European Union

Pairwise comparisons for the interaction effect

```
pairs(em, by = "NEPsplit")
```

```
> pairs(em, by = "NEPsplit")
NEPsplit = low:
 contrast          estimate   SE df t.ratio p.value
 neutral - negatyw    -3.54 7.25 84 -0.488  0.8774
 neutral - pozytyw   -15.48 7.25 84 -2.135  0.0890
 negatyw - pozytyw   -11.95 7.25 84 -1.647  0.2318

NEPsplit = high:
 contrast          estimate   SE df t.ratio p.value
 neutral - negatyw   -25.36 6.32 84 -4.011  0.0004
 neutral - pozytyw   -27.91 6.32 84 -4.414  0.0001
 negatyw - pozytyw    -2.55 6.32 84 -0.403  0.9144

P value adjustment: tukey method for comparing a family of 3 estimates
```

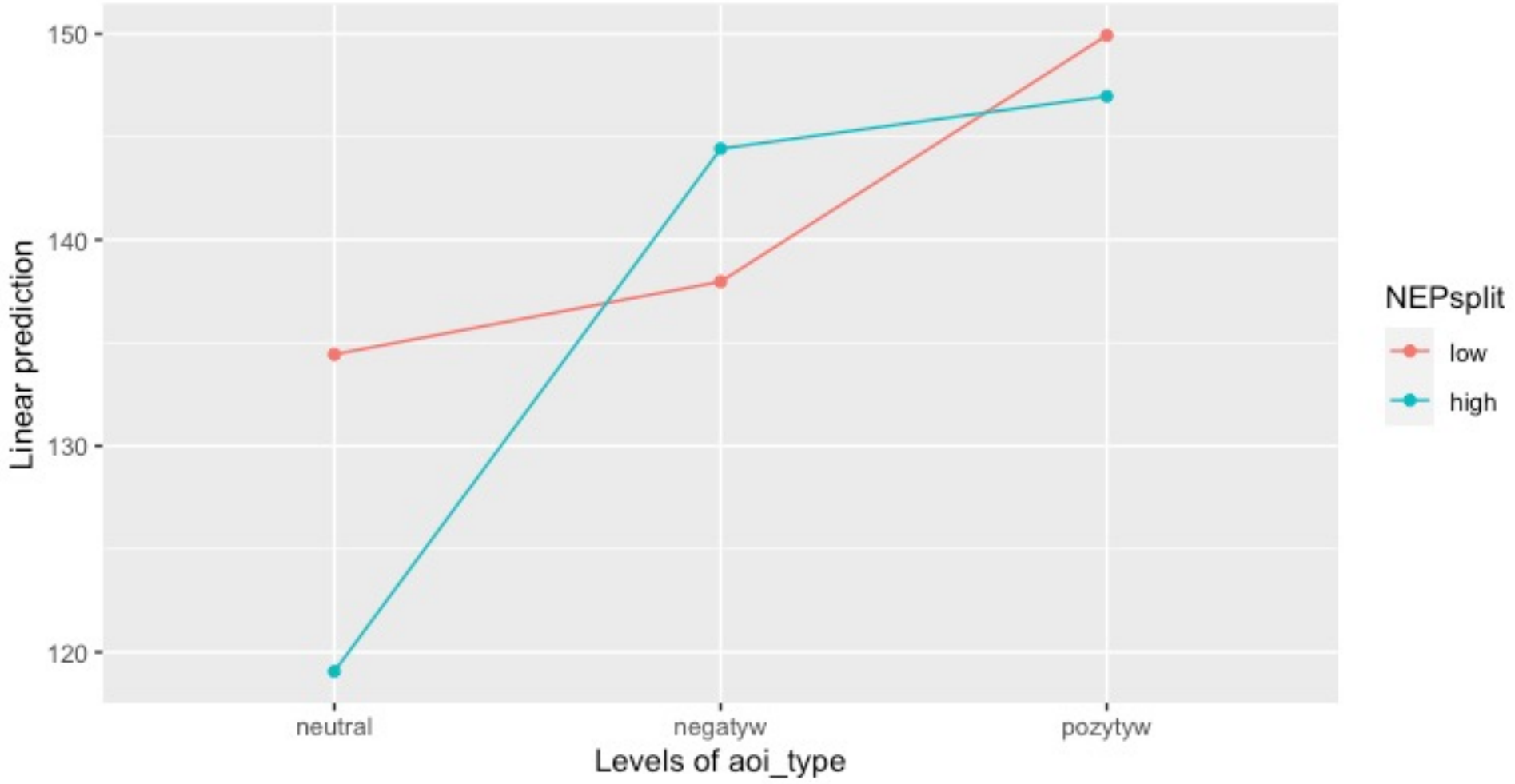
```
pairs(em, by = "aoi_type")
```

```
> pairs(em, by = "aoi_type")
aoi_type = neutral:
 contrast   estimate   SE  df t.ratio p.value
 low - high    15.38 8.43 101  1.825  0.0709

aoi_type = negatyw:
 contrast   estimate   SE  df t.ratio p.value
 low - high    -6.44 8.43 101 -0.765  0.4463

aoi_type = pozytyw:
 contrast   estimate   SE  df t.ratio p.value
 low - high     2.95 8.43 101  0.350  0.7267
```

Funded by the Horizon 2020 Framework Programme of the European Union

Analytical graph of the interaction effect

```
emmip(fit,  NEPsplit ~ aoi_type)
```



Funded by the Horizon 2020 Framework Programme of the European Union

Publication ready plot of interaction effect

```
tab <- summary(em)
```

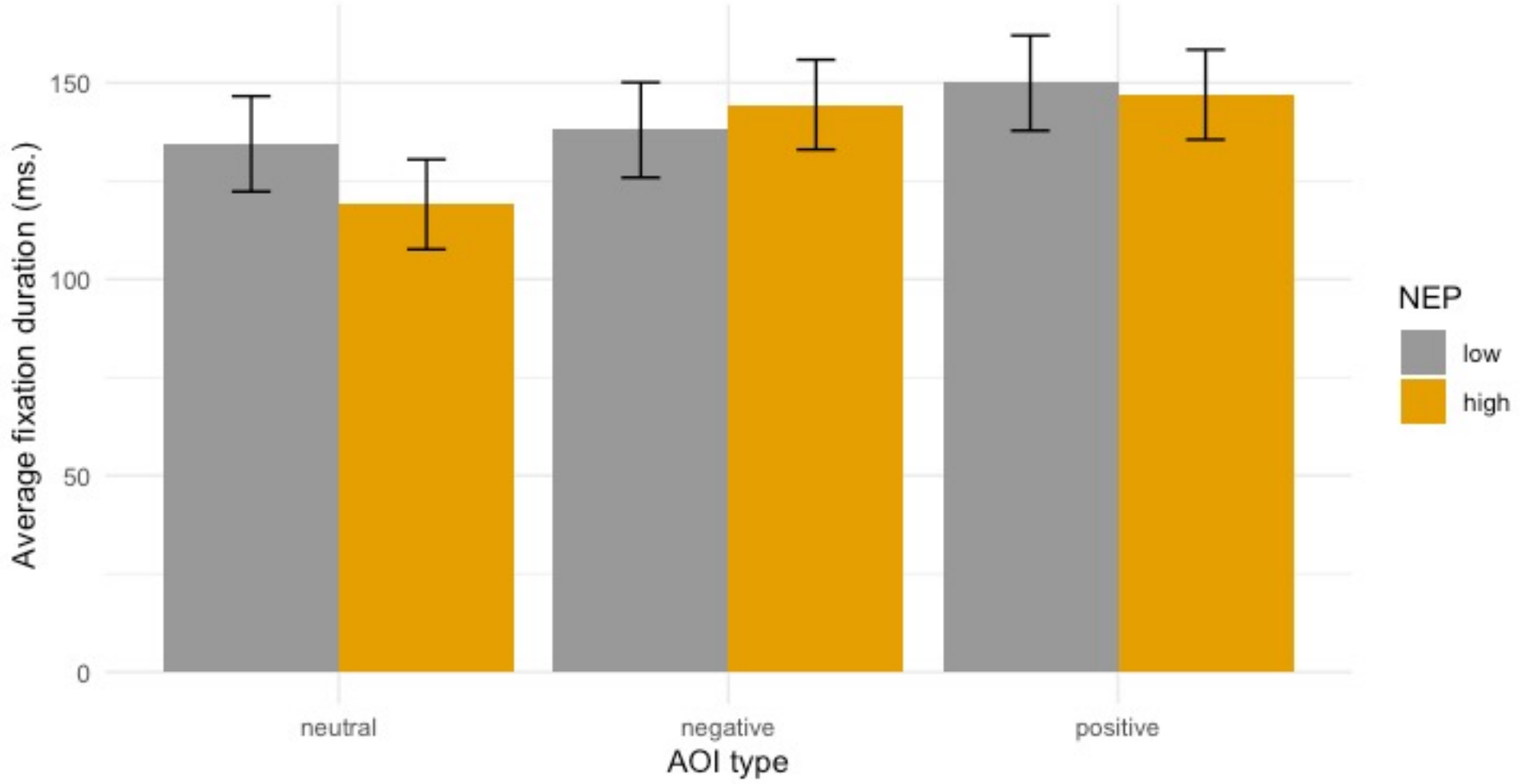
```
> print(tab)
 aoi_type NEPsplit emmean   SE    df lower.CL upper.CL
 neutral  low         134 6.12 107.1      122      147
 negatyw  low         138 6.12 107.1      126      150
 pozytyw  low         150 6.12 107.1      138      162
 neutral  high        119 5.76  95.7      108      131
 negatyw  high        144 5.76  95.7      133      156
 pozytyw  high        147 5.76  95.7      136      158
```

```
plo <- ggplot(tab, aes(x=aoi_type, y=emmean, fill=NEPsplit)) +
  geom_bar(position=position_dodge(), stat="identity") +
  geom_errorbar(aes(ymin=lower.CL, ymax=upper.CL),
                width=.2, position=position_dodge(.9)) +
  scale_y_continuous(name = "Average fixation duration (ms.)") +
  scale_x_discrete(name = "AOI type",
                   labels = c("neutral", "negative", "positive")) +
  scale_fill_manual(name = "NEP",
                    values=c("#999999", "#E69F00")) +
  theme_minimal()
```

Funded by the Horizon 2020 Framework Programme of the European Union

Publication ready plot of interaction effect

(bargraph with error bars representing confidence intervals)



Funded by the Horizon 2020 Framework Programme of the European Union

Save the graph to a file

Saving the graph to the file with a code.
You can choose a wide variety of formats (jpg, png, tiff, pdf, eps, etc.)

```
print(plo)
ggsave(plot = plo, filename = "graphs/AverageFixationDuration_NEP_AOI.jpg")
dev.off()
```

Funded by the Horizon 2020 Framework Programme of the European Union

Test hypotheses about fixation count

- We want to test two hypotheses:
  - In general, negative and positive pictures of environment will evoke more fixations than neutral pictures
  - Negative and positive pictures of environment will evoke more fixations than neutral pictures BUT ONLY for people with high sensitivity to climate change
    - *sensitivity to climate change will be a moderator for the effect of picture valence*

Funded by the Horizon 2020 Framework Programme of the European Union

Run two-way mixed-design ANOVA

RUN ANOVA TEST

```
fit <- aov_ez(data = d,
              id = "ID",
              dv = "aoi_fixation_total_count",
              within = "aoi_type",
              between = "NEPsplit")
```

Data

Subject ID variable

Dependent variable

Within-subject factor

Between-subject factor

PRINT ANOVA TEST RESULTS

```
> print(fit)
Anova Table (Type 3 tests)

Response: aoi_fixation_total_count
              Effect          df   MSE         F  ges p.value
1           NEPsplit       1, 42 12.08      0.57 .005    .455
2           aoi_type 1.76, 74.09 12.19 23.08 *** .260   <.001
3 NEPsplit:aoi_type 1.76, 74.09 12.19      0.67 .010    .497
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘+’ 0.1 ‘ ’ 1

Sphericity correction method: GG
```

Estimated means and pairwise comparisons for main effect

```
em <- emmeans(fit, ~aoi_type)
```

```
> print(em)
 aoi_type emmean    SE  df lower.CL upper.CL
 neutral    4.19 0.507 126     3.18     5.19
 negatyw    8.16 0.507 126     7.15     9.16
 pozytyw    8.50 0.507 126     7.50     9.50
```

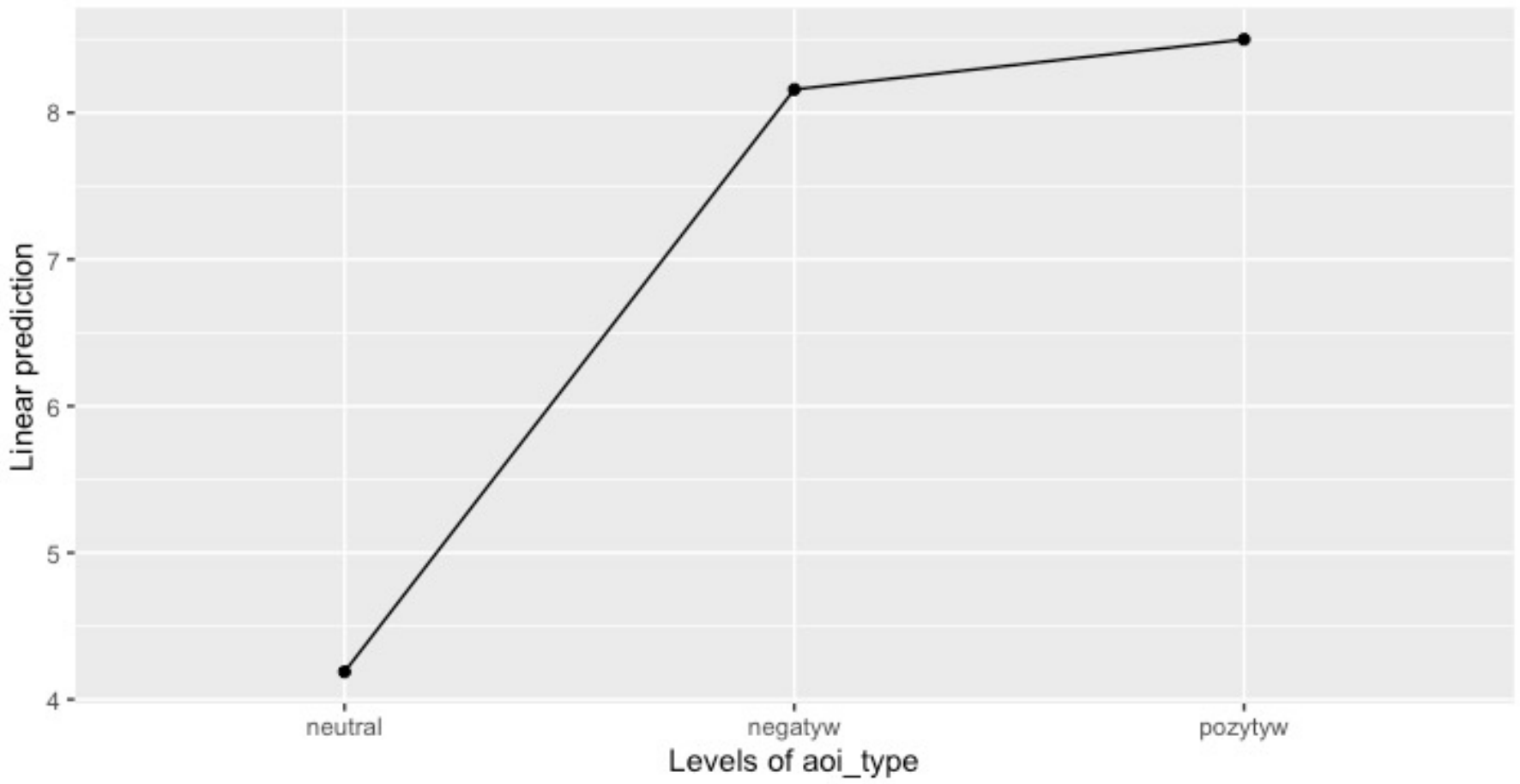
```
# post-hoc for main effect
pairs(em)
```

```
> pairs(em)
 contrast          estimate    SE df t.ratio p.value
 neutral - negatyw   -3.971 0.706 84 -5.626  <.0001
 neutral - pozytyw   -4.312 0.706 84 -6.110  <.0001
 negatyw - pozytyw   -0.341 0.706 84 -0.484  0.8792

Results are averaged over the levels of: NEPsplit
P value adjustment: tukey method for comparing a family of 3 estimates
```

Funded by the Horizon 2020 Framework Programme of the European Union

Analytical graph of the main effect

```
emmip(fit, ~ aoi_type)
```



Funded by the Horizon 2020 Framework Programme of the European Union

Publication ready plot of the main effect

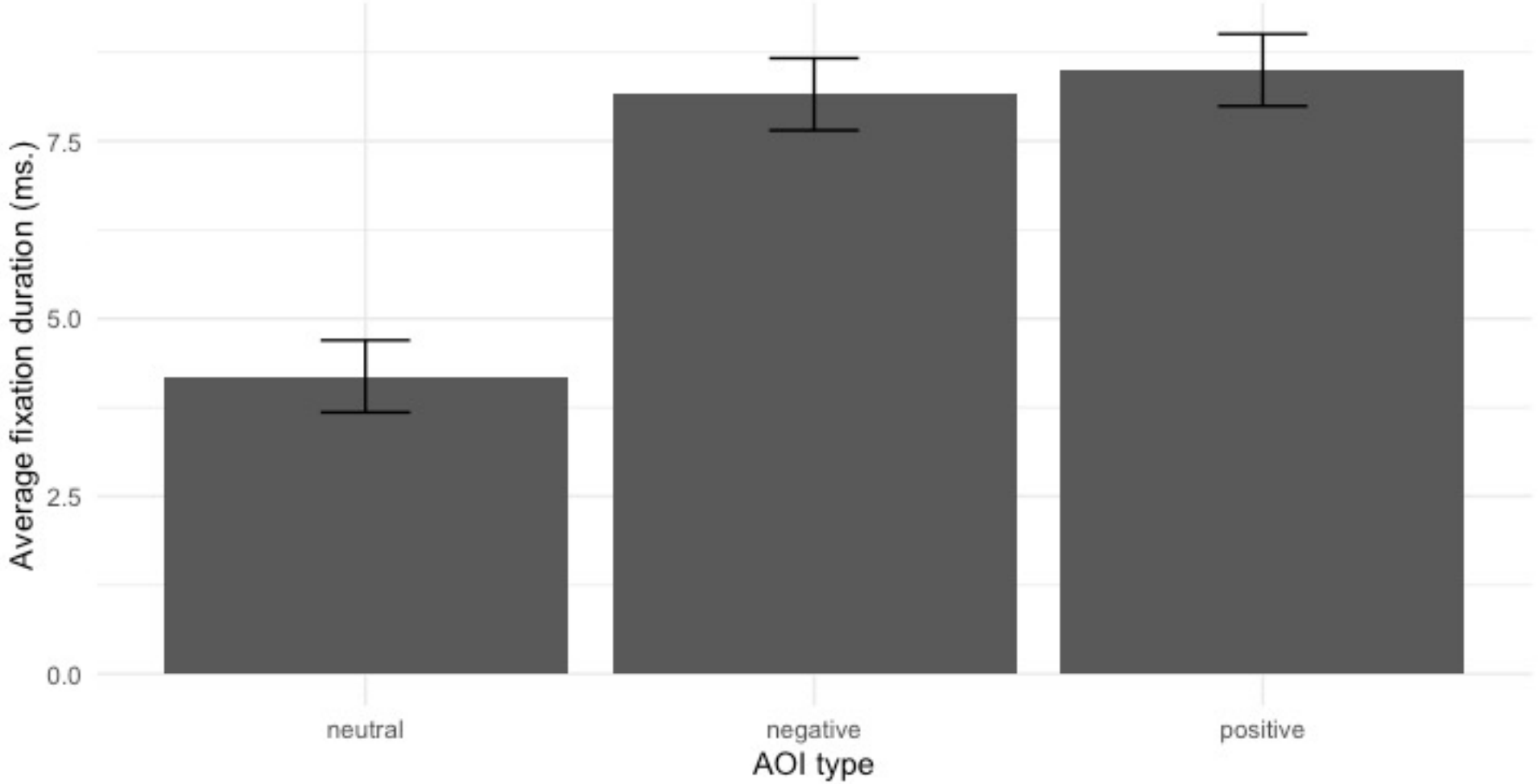
```
tab <- summary(em)
```

```
> print(tab)
 aoi_type emmean    SE  df lower.CL upper.CL
 neutral    4.19 0.507 126     3.18     5.19
 negatyw    8.16 0.507 126     7.15     9.16
 pozytyw    8.50 0.507 126     7.50     9.50
```

```
plo <- ggplot(tab, aes(x=aoi_type, y=emmean)) +
  geom_bar(position=position_dodge(), stat="identity") +
  geom_errorbar(aes(ymin=emmean-SE, ymax=emmean+SE),
                width=.2, position=position_dodge(.9)) +
  scale_y_continuous(name = "Average fixation duration (ms.)") +
  scale_x_discrete(name = "AOI type",
                   labels = c("neutral", "negative", "positive")) +
  theme_minimal()
```

Funded by the Horizon 2020 Framework Programme of the European Union

Publication ready plot of ithe main effect



Funded by the Horizon 2020 Framework Programme of the European Union

Save the graph to a file

Saving the graph to a file
You can choose a wide variety of formats (jpg, png, tiff, pdf, eps, etc.)

```
print(plo)
ggsave(plot = plo, filename = "graphs/fixationCount_AOI.jpg")
dev.off()
```

LEAD-ME

Funded by the Horizon 2020 Framework Programme of the European Union

SWPS UNIVERSITY

UNIVERSITY OF WARSAW

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Moderation analysis with linear regression

Mixed-design linear regression with continuous and nominal predictors and interaction term

Funded by the Horizon 2020 Framework Programme of the European Union

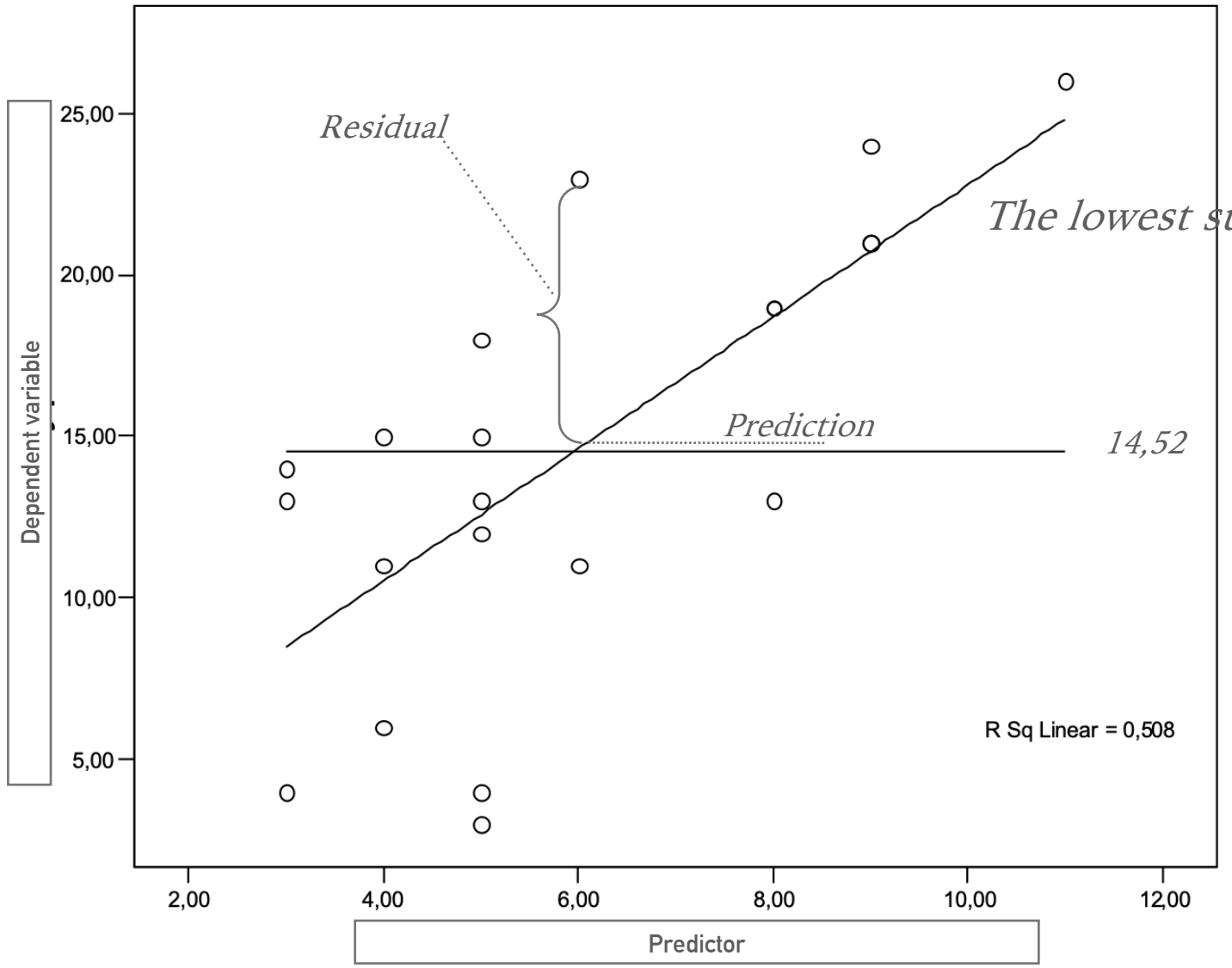
Linear regression

- In regression we are fitting a model to our data
  - not loosing variance if predictor is continous and does not need to be splitted into a categorical variable
  - predicting values of the dependent variable from one or more independent variable

$$\hat{Y} = a + bX + e$$

Funded by the Horizon 2020 Framework Programme of the European Union

Least squares Method



The lowest sum of squared differences

Funded by the Horizon 2020 Framework Programme of the European Union

Multiple regression without and with interaction

- Each predictor has its own regression coefficient
  - For every extra predictor you include, another coefficient need to be estimated
- We are looking for linear combination of predictors that correlate maximally with the outcome variable
  - multiple regression formula with two predictors (no interaction term)

$$\hat{Y} = a + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

  - multiple regression formula with two predictors and interaction term

$$\hat{Y} = a + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon$$

Funded by the Horizon 2020 Framework Programme of the European Union

Useful libraries

```
library(emmeans)
library(lme4)
```

lme4 provides functions for fitting and analyzing mixed models:
linear (lmer),
generalized linear (glmer), and
nonlinear (nlmer.)

Funded by the Horizon 2020 Framework Programme of the European Union

Hypotheses

- We want to test the following hypotheses:
  - The more sensitive to climate change people are the longer fixation duration on the environment pictures
  - Positive and negative pictures of environment will evoke longer fixation duration than neutral ones
  - Sensitivity to climate change will predict fixation duration differently while looking at pictures of different valence (interaction hypothesis).

Funded by the Horizon 2020 Framework Programme of the European Union

Multiple regression.

Model definition

Longer form of full model definition

```
fit <- lmer(aoi_fixation_average_duration_ms ~ NEP + aoi_type + NEP:aoi_type + (1 | ID),
            data = d)
```

Dependent variable

Predictor 1

Predictor 2

Interaction of predictors

Random intercept for subjects

Data frame

Shorter form of full model definition

```
fit <- lmer(aoi_fixation_average_duration_ms ~ NEP*aoi_type + (1 | ID),
            data = d)
```

Funded by the Horizon 2020 Framework Programme of the European Union

Model results.

Basic table

```
> anova(fit)
Type III Analysis of Variance Table with Satterthwaite's method
             Sum Sq Mean Sq NumDF DenDF F value  Pr(>F)
NEP            2269  2269.4     1    42  0.3989 0.53107
aoi_type      35625 17812.5     2  1932  3.1312 0.04389 *
NEP:aoi_type  51223 25611.7     2  1932  4.5022 0.01120 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Funded by the Horizon 2020 Framework Programme of the European Union

Model results.

Detailed table

```
> summary(fit)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: aoi_fixation_average_duration_ms ~ NEP + aoi_type + NEP:aoi_type +      (1 | ID)
   Data: d

REML criterion at convergence: 22769

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.2882 -0.1971  0.2103  0.5200  5.7394

Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept)  307.1   17.52
 Residual             5688.8   75.42
Number of obs: 1980, groups:  ID, 44

Fixed effects:
                    Estimate Std. Error        df t value Pr(>|t|)
(Intercept)          161.0240    31.1653  104.5102   5.167 1.15e-06 ***
NEP                   -1.0629     0.9374  104.5102  -1.134  0.25946
aoi_typenegatyw      -81.3558    32.7634 1932.0000  -2.483  0.01311 *
aoi_typepozytyw      -31.8670    32.7634 1932.0000  -0.973  0.33085
NEP:aoi_typenegatyw    2.9503     0.9855 1932.0000   2.994  0.00279 **
NEP:aoi_typepozytyw    1.6500     0.9855 1932.0000   1.674  0.09424 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) NEP    a_typn a_typp NEP:_typn
NEP         -0.992
ao_typngtyw -0.526  0.521
a_typpzytyw -0.526  0.521  0.500
NEP:_typngt  0.521 -0.526 -0.992 -0.496
NEP:_typpzy  0.521 -0.526 -0.496 -0.992  0.500
```

NOTE: In the report remember to provide all the values from the coefficients table: coefficient value, standard error, t-test value, degrees of freedom and p-value.

Funded by the Horizon 2020 Framework Programme of the European Union

Trends analysis for interaction

Code

```
em <- emtrends(fit, ~ NEP:aoi_type, var = "NEP")
```

Results of trends analysis

```
> print(em)
 NEP aoi_type NEP.trend    SE  df lower.CL upper.CL
  33 neutral     -1.063 0.937 105  -2.9218    0.796
  33 negatyw      1.887 0.937 105   0.0286    3.746
  33 pozytyw      0.587 0.937 105  -1.2718    2.446

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
```

Statistical comparison of slopes

```
> pairs(em)
 contrast                                                estimate    SE   df t.ratio p.value
 32.9772727272727 neutral - 32.9772727272727 negatyw    -2.95 0.986 1932 -2.994  0.0079
 32.9772727272727 neutral - 32.9772727272727 pozytyw    -1.65 0.986 1932 -1.674  0.2154
 32.9772727272727 negatyw - 32.9772727272727 pozytyw     1.30 0.986 1932  1.319  0.3845

Degrees-of-freedom method: kenward-roger
P value adjustment: tukey method for comparing a family of 3 estimates
```

Funded by the Horizon 2020 Framework Programme of the European Union

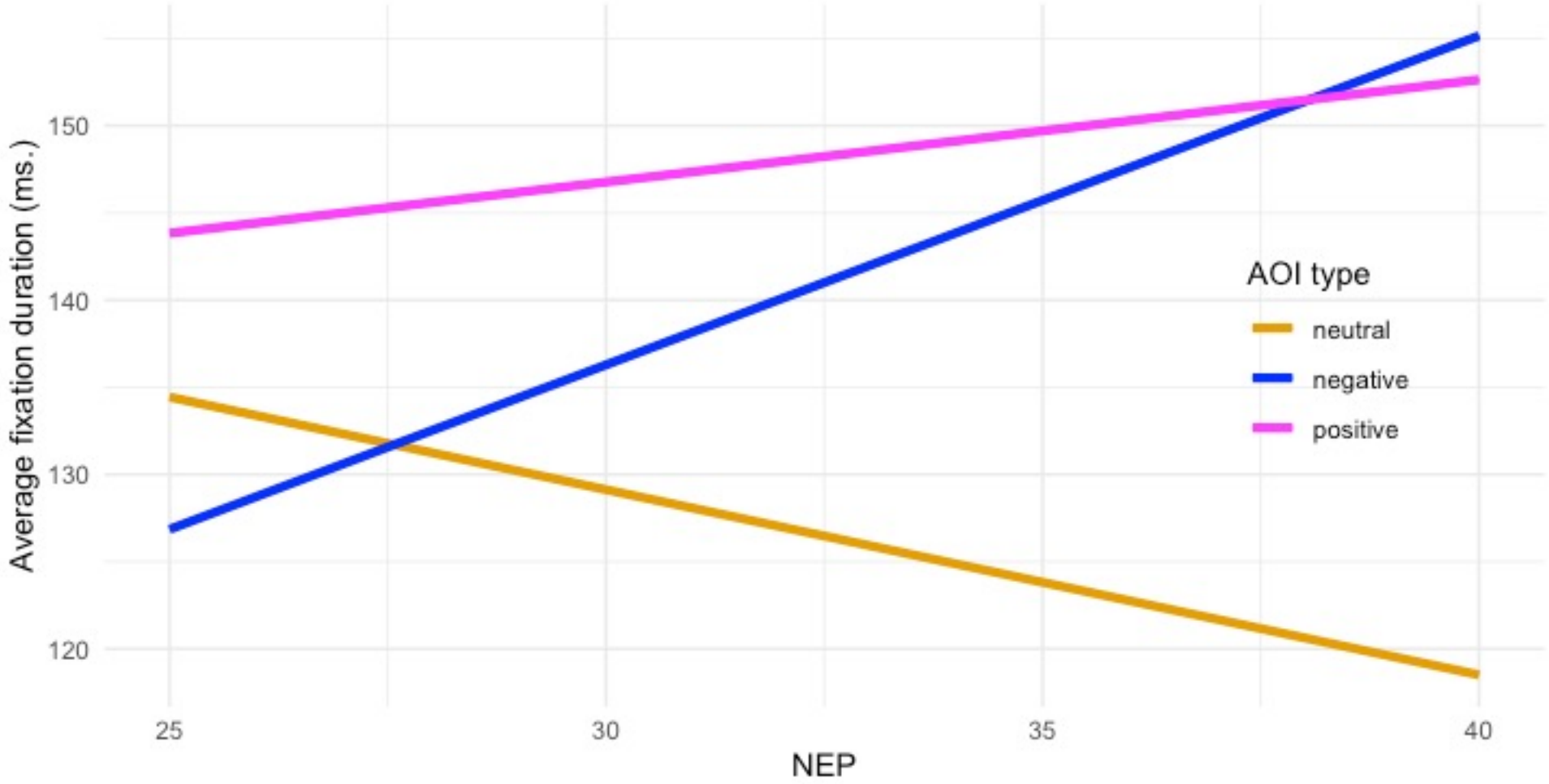
Interaction visualization.

Publication ready

```
plo <- emmip(fit, aoi_type ~ NEP, cov.reduce = range) +
        geom_line(size=1.5) +
        scale_y_continuous(name = "Average fixation duration (ms.)") +
        scale_x_continuous(name = "NEP") +
        scale_color_manual(name = "AOI type",
                           labels = c("neutral", "negative", "positive"),
                           values = c("#E69F00", "#0000FF", "#FF00FF")) +
        theme_minimal() +
        theme(legend.position = c(.85, .5))
```

Funded by the Horizon 2020 Framework Programme of the European Union

Save the graph




```
print(plo)
ggsave(plot = plo, filename = "graphs/fixationDuration_regression.jpg")
dev.off()
```

Funded by the Horizon 2020 Framework Programme of the European Union

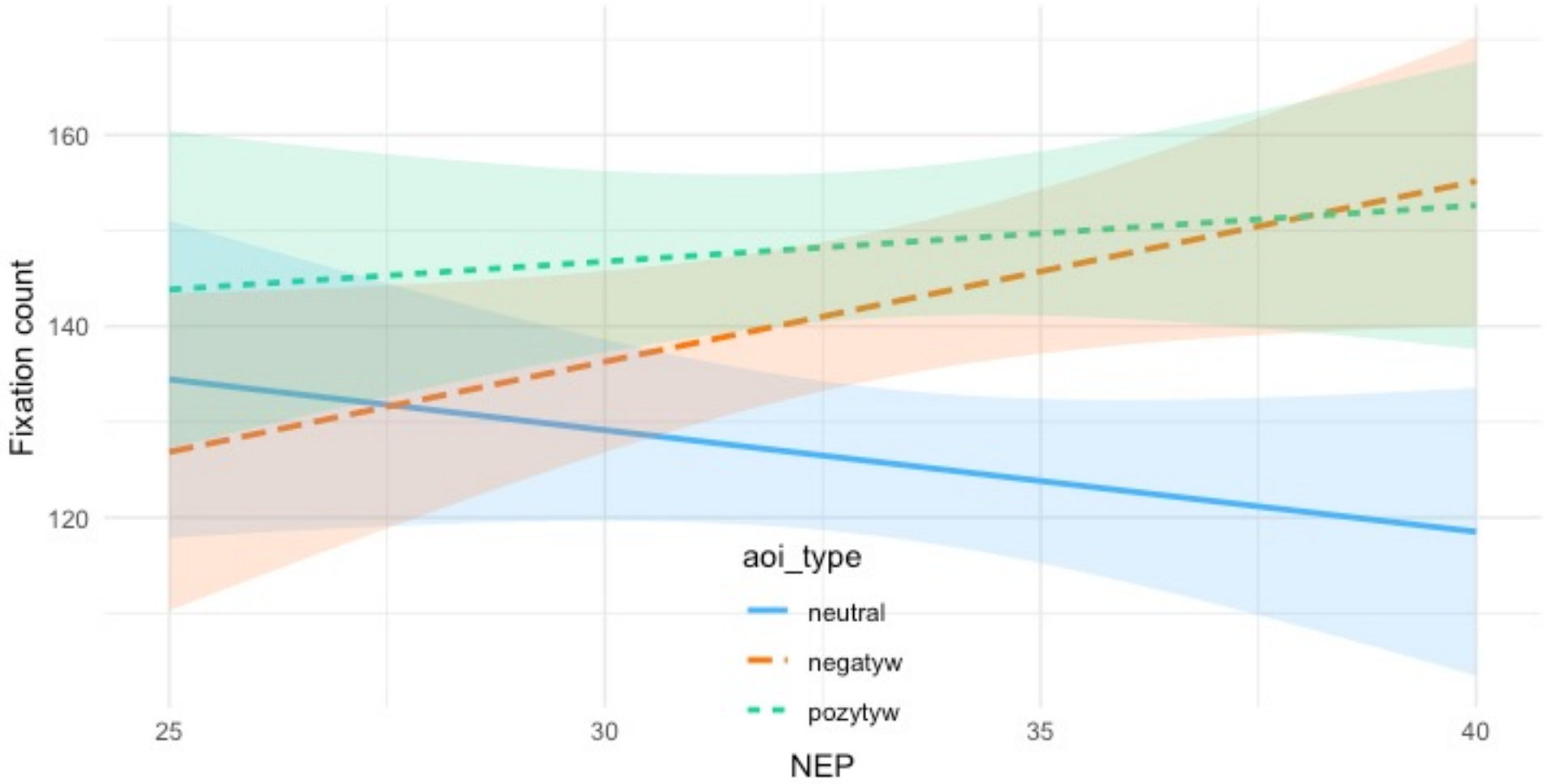
Alternative plot of the interaction

```
# alternatively
require(interactions)
interact_plot(fit, pred = NEP, modx = aoi_type, interval = TRUE) +
  scale_y_continuous(name = "Fixation count") +
  scale_x_continuous(name = "NEP") +
  theme_minimal() +
  theme(legend.position = c(.5, .1))
```

Funded by the Horizon 2020 Framework Programme of the European Union

Alternative plot of interaction

(with confidence intervals for slopes and different predefined line types)



Funded by the Horizon 2020 Framework Programme of the European Union

After this workshop … you are able to perform mixed-design ANOVA and multiple regression with interaction effects

- **Prepare the data frames for the analysis**
  - Read/load data into R
  - Calculate new variables
  - Prepare dichotomous factor with median-split
  - Select variables and filter observations
  - Merge two data frames
  - Write data frame to a file
- **Perform mixed-design ANOVA**
  - Read and interpret the results
  - Estimate means
  - Run pairwise comparisons (post hoc tests) for significant effects
  - Prepare publication ready bar graphs
- **Perform mixed-design linear multiple regression analysis**
  - Read and interpret the results
  - Perform trends (simple slopes) analysis
  - Compare statistically simple slopes
  - Prepare publication ready line graphs with confidence intervals

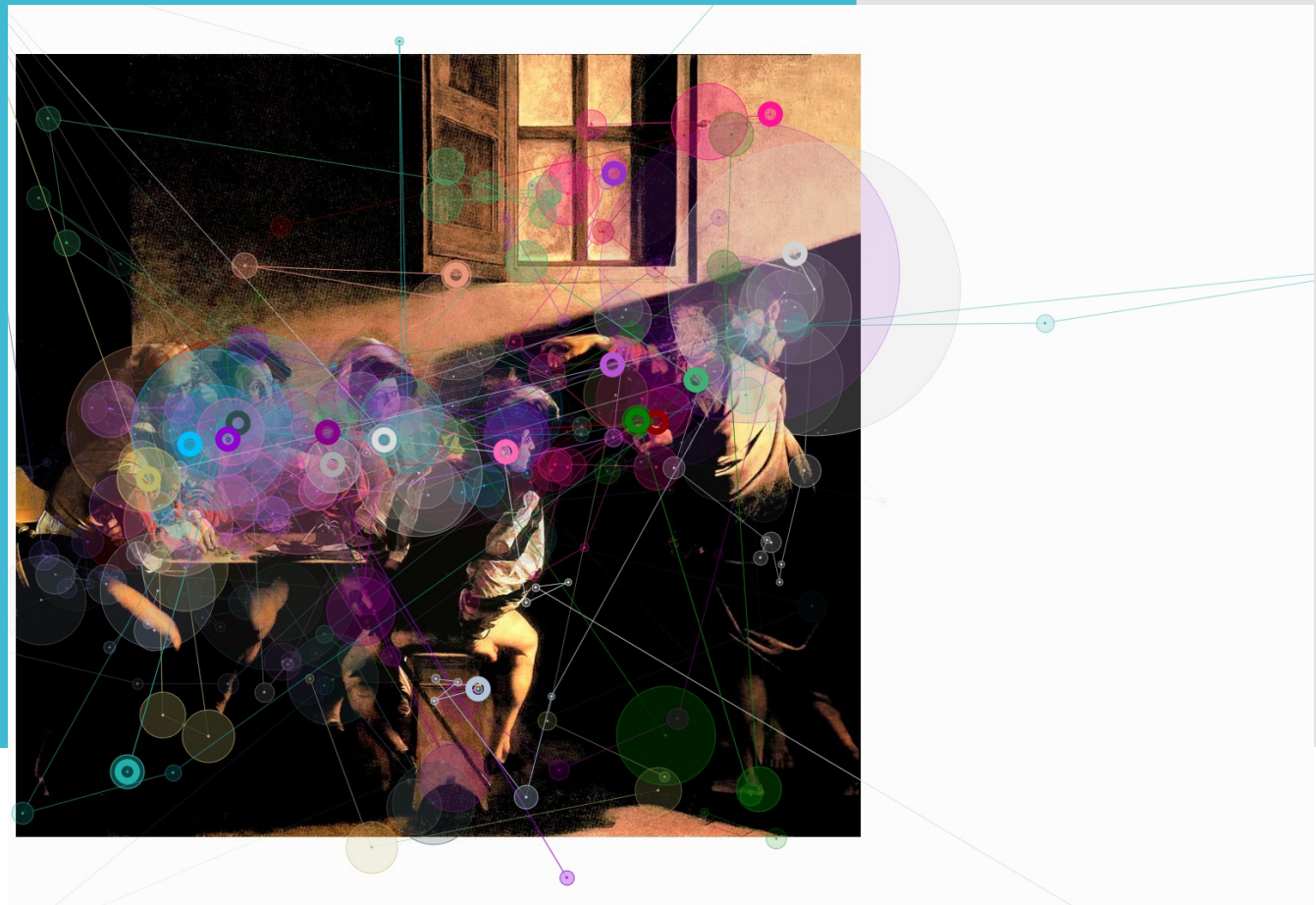
Funded by the Horizon 2020 Framework Programme of the European Union

SWPS UNIVERSITY

UNIVERSITY OF WARSAW

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

It's your turn to practice statistical analysis during hand-on sessions



Funded by the Horizon 2020 Framework Programme of the European Union

SWPS UNIVERSITY
UNIVERSITY OF WARSAW

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Thank you!

Any questions?