

Development of an open application for teaching statistics: conception and results

Eduard Serrahima, Lluís Marco-Almagro, Xavier Tort-Martorell,
Pere Grima and Lourdes Rodero

UPC Universitat Politècnica de Catalunya. Barcelona Tech

Different commercial statistical software packages exist that are appropriate for teaching statistics. However, free alternatives are scarce, and those existing suffer from problems (such as difficult installations and bizarre interfaces) that make them not very usable. This paper presents the conception and implementation of a new application for statistical analysis with the following characteristics: free, simple to use, multi-platform, scalable and with a focus on statistical concepts and ideas.

The whole development cycle will be explained in the presentation: first list of user requirements, design decisions, implementation and testing. The application is developed in R language and uses the shiny package for the user interface. Although the application obviously allows the analysis of real data through graphs and statistical methods, its main objective is facilitating the acquisition of statistical concepts. To accomplish this, menus, configuration options and results are presented in a way that fosters reflection on basic statistical ideas. The application has a special emphasis on industrial statistics.

Initial feedback from users testing the application will be exposed, and ways to freely access and use the application will be presented.

Keywords: statistics teaching, data analysis, data visualization, free software, project-based learning

15th ENBIS Conference

**Development of an open application for
teaching statistics**

**Eduard Serrahima
Lluís Marco-Almagro
Xavier Tort-Martorell
Pere Grima
Lourdes Rodero**



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

7th September 2015



**Develop an application for statistical
data analysis that facilitates learning
statistical concepts to students in
secondary school (12-18 years old).**

Of course, other groups can benefit from the software!



Each year, secondary school students are invited to participate in a contest organized by Catalan universities. They have to submit a project (report + presentation) where data are analyzed using statistical tools.

Awards session: students and teachers with best projects are invited to attend



Projects are divided into categories (three age levels: 12-13, 14-15, 16-17)
According to a jury, the best project in each category wins a prize.

Nervous waiting for the result



Member of the jury (university teacher)
Secondary school students
Secondary school teacher
Faculty deans



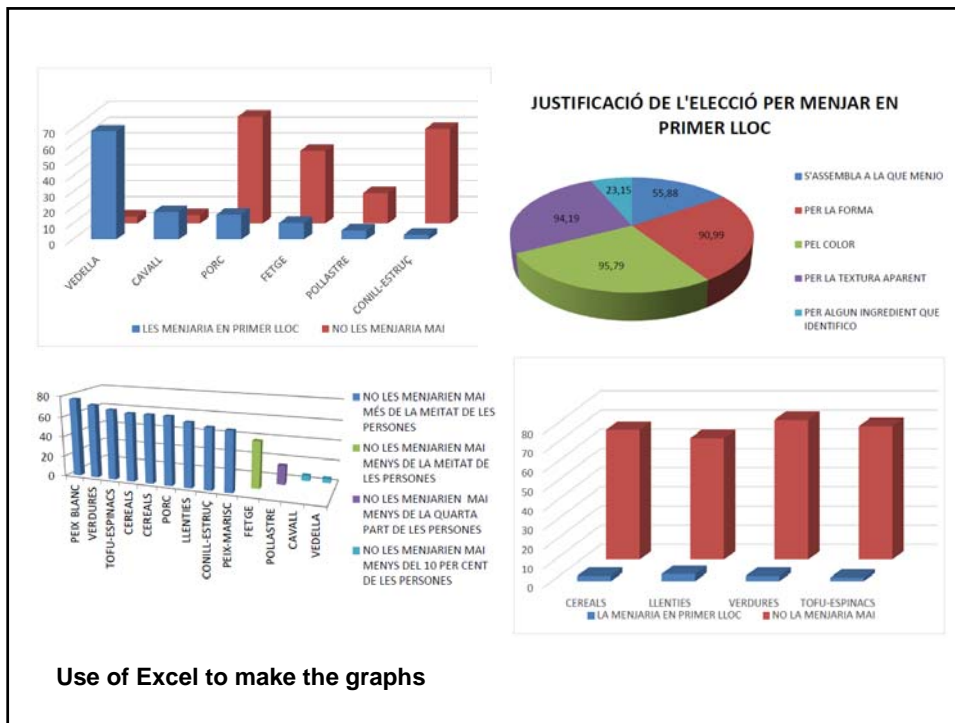
Examples of winning projects from 2015



13 years old
(2nd ESO)

**Online survey to
discover the favorite
hamburger**

**First prize in category
12-13 years old in 2015**



PROJECTE DE RECERCA
Perdent el To
2014-2015

15 years old (4th ESO)

Experimental design to discover which frequencies can be heard

First prize in category 14-15 years old in 2015

We have made this project to get more knowledge and information about how the ear works, how we perceive the sounds and the bad uses that we do of the new technological applications and musical gadgets which we use every day in our lives

**Manu Cazorla, Pau Viñas
Albert Olmedo, Pau Comas
4rt D
INS SANT QUIRZE
Tolosa | Judith Miró Calmes**

Use of tools freely available on the web
Excel to make the graphs

Determine Sample Size

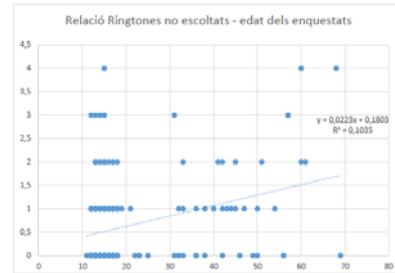
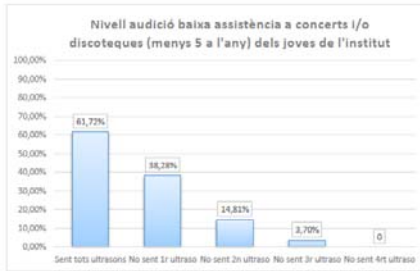
Confidence Level: 95% 99%

Confidence Interval:

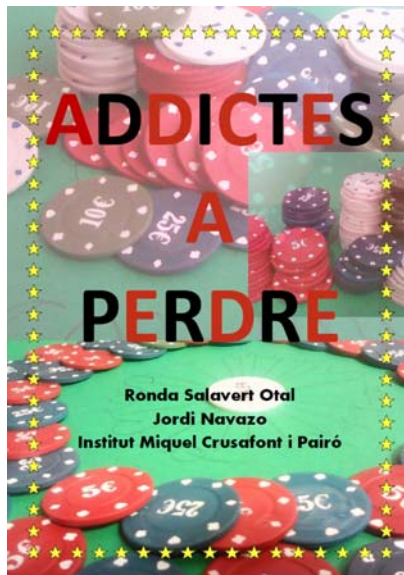
Population:

Calculate

Sample size needed:



“We obtained a Pearson coefficient of correlation of 0,1035. There is no relationship, because we need a coefficient of correlation greater than 0,5 (in absolute value) to assure correlation”



17 years old
(2nd batxillerat)

Simulation to show the casino and player benefits when using the roulette

First prize in category
16-17 years old in 2015

Simulation is done in R

```

RStudio [12 de 12] - [C:\Users\jorge\Documents\Programari\R\Ajuda\Fuencs\1.1\libre 1]
Archivo Editar Paquetes Ventanas Ayuda

#Primer de tot anomenem les variables que posteriorment utilitzarem.
t <- 0 #És pel comptador de partides.
d <- 0 #És per comptat les victòries.
p <- 0 #És per després comptat les derrotes.
n <- 1 #Aquí definim el nombre de partides que volem fer.
m <- 100 #Aquest és el diner amb què juguem.
k <- m #Faig una variable igual que el diner de l'aposta inicial.
y <- 1000 #És el diner amb el que nos retirarem de la partida quan guanyem.
m <- y - m #Aquesta resta nos mostra per després calcular els beneficis.
#És el que guanyem per partida (el que hem guanyat menys
#El diner amb què hem començat la partida).
#Comença el primer bucle on
#repetira les accions dependenc del nombre de partides que haguem determinat.
while (t < n) {
  #Inicialment d'apostat bucle i nos dirigirem a la funció dels beneficis.
  s <- 1 + 1 #F'''' És una mena de contador que compta les partides que hem fet.
  m <- m #És la reestabliment el valor de l'aposta per cada cop que comencem una nova partida.
  while(m > 0 && m < y) {
    if(s < (y - m)) {
      #Establim el valor de la aposta que farem, dependenc dels diners que
      #nos quedem al tenir menys diners dels que hauriem d'apostar, és el
      #d <- (y - m) / 2 #És que nos queda no apostar, i nos apostarem els diners que nos falten
      #per arribar als 1000 (que és el que hauriem d'apostar).
      #Nos apostem funció posem un nombre entre aleatori del 0 al 34,
      #com en les ruletes franceses.
      if(s > 18) {
        #Com que la nostra aposta està dels nombres entre 19 a 34, si el
        #nombre que surt és més gran que 18 guanyarem i nos donarem els diners
        m <- (s + m) * 2 #És que hauriem apostat, que se sumaria als diners que ja teníem.
        m <- (s - m) #És el cas que no sigui així i per tant nos quedarem amb el 0, és a dir, hem
        #perduts els diners que teníem i nos quedarem amb la nostra aposta, de manera que el que teníem
        #allí restarem el que hauriem apostat.
      }
      #Quan al que tenim és 0 (hauriem perdut), sumarem un punt al contador
      if(m == 0) {
        #És el cas que no sigui així i per tant nos quedarem amb el 0, és a dir, hem
        #perduts els diners que teníem i nos quedarem amb la nostra aposta, de manera que el que teníem
        #allí restarem el que hauriem apostat.
      }
      #Si arribem a 1000 nos retirarem (hauriem guanyat), en aquest cas,
      #anotarem un punt al contador de victòries que determinem.
    }
  }
  #Després de completar totes les partides que volem i de contar les que hem perdut i guanyat,
  #fem un exemple dels beneficis.
  d <- (d + m) - (y - m) #Aquesta resta nos mostra els 1000 (el que guanyem per cada victòria menys
  #els 1000 (el que posem per cada derrota.
  
```



Graphs are produced in Excel

General characteristics of the projects

- Basically, all projects collect and analyze data.
- Data is almost always analyzed graphically.
- Graphs are done with Excel (or students use the graphs provided by the website used to collect data, such as google forms).
- When some more sophisticated tool is used, it is something freely available on the web.
- The winning projects are in general really good.
- Some of the presented projects are rather poor, usually containing statistical mistakes.




- Why are these students only using Excel to produce graphs (often bad graphs)?
- Is it possible to use this project to really convey some important statistical concepts (such as the idea of variability)?
- Can we go a bit further (and even introduce some more complex statistical concepts)?



Use of software

Review of existing programs

Commercial software

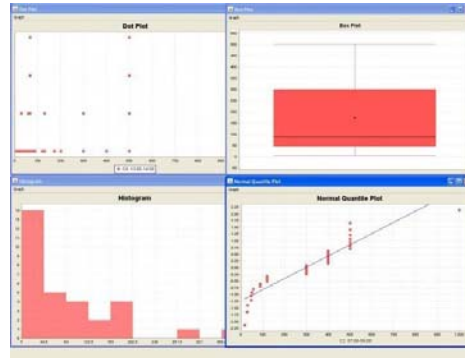
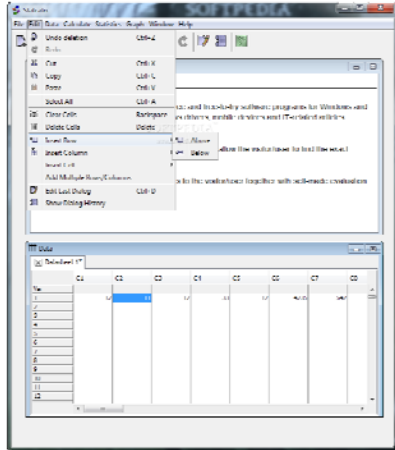
JMP	Minitab	SPSS
		

Free software

- Statcato
- Salstat
- PSPP
- Statistical Lab
- SOFA Statistics
- ...

Review of existing programs

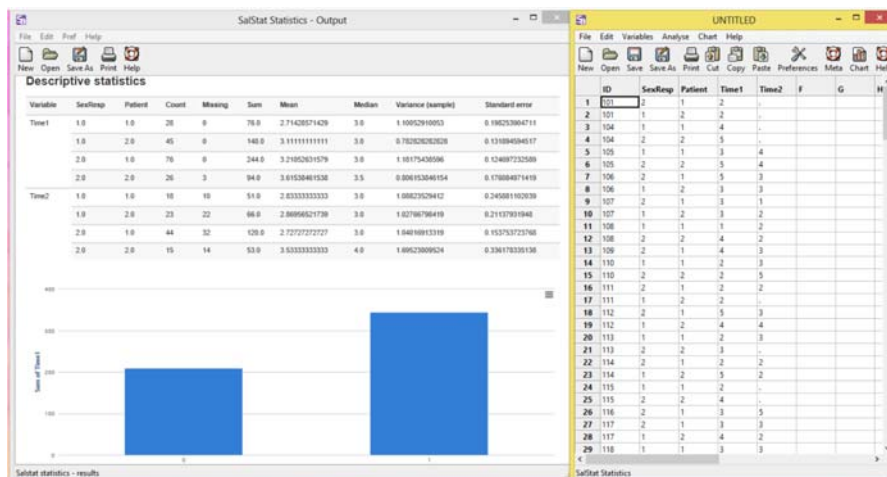
Statcato



Emulates Minitab in both user interface and presentation of results

Review of existing programs

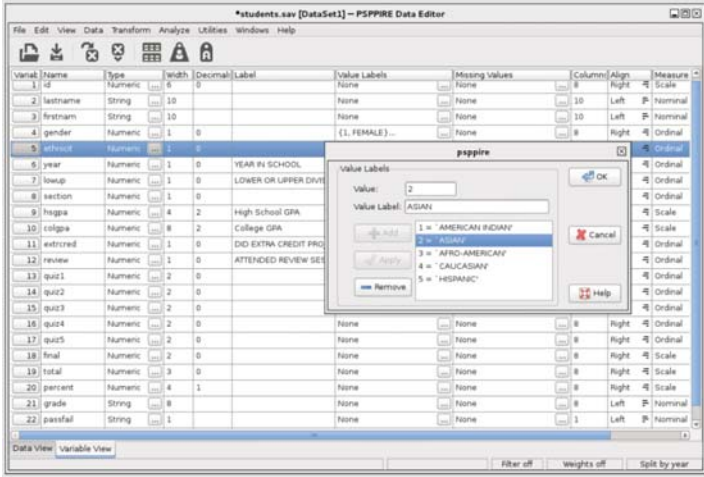
Salstat



Similar to Minitab

Review of existing programs

PSPP

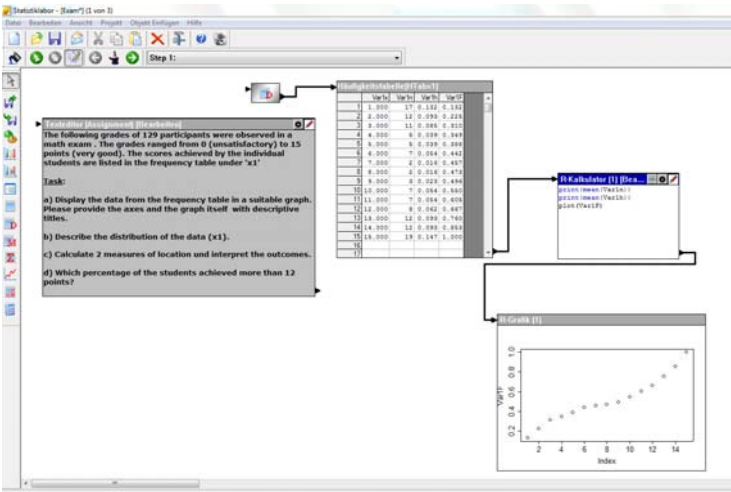


The screenshot shows the PSPP Data Editor window with a list of variables on the left. A dialog box titled 'ppspire' is open, allowing the user to define value labels for a variable. The dialog shows 'Value: 2' and 'Value Label: ASIAN'. Below, there are options to add, remove, or apply labels for other values (1, 3, 4, 5).

Tries to emulate SPSS

Review of existing programs

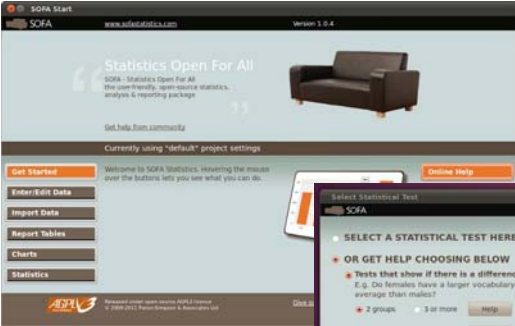
Statistical Lab




The screenshot shows the Statistical Lab interface. On the left, a task description is displayed: 'The following grades of 129 participants were observed in a math exam. The grades ranged from 0 (unsatisfactory) to 15 points (very good). The scores achieved by the individual students are listed in the frequency table under 'x1'. Task: a) Display the data from the frequency table in a suitable graph. Please provide the axes and the graph itself with descriptive titles. b) Describe the distribution of the data (x1). c) Calculate 2 measures of location and interpret the outcomes. d) Which percentage of the students achieved more than 12 points?'. In the center, a frequency table is shown with columns for 'x1', 'n(x1)', 'w(x1)', and 'W(x1)'. On the right, a graph titled 'Graph [1]' is displayed, showing a scatter plot of 'Index' (x-axis) versus 'VARI' (y-axis).

Uses R as calculation engine. You need some R knowledge

Review of existing programs




SOFA



Based on wizards

Main problems with existing free software



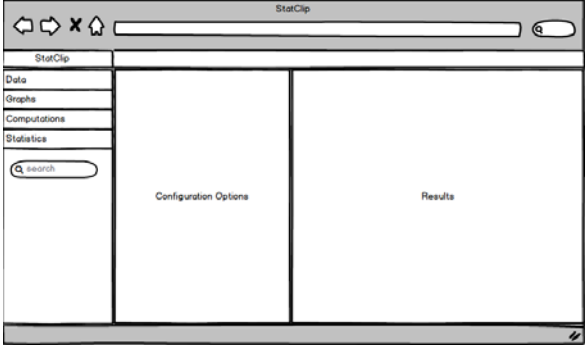
- **Often very difficult to install (sometimes not possible)**
- **Not available on all platforms**
- **Too complicated (many options, cluttered interfaces, ...)**
- **More designed to mimic commercial software than to explain statistical concepts**

Users and needs analysis

Users	Users		
	Students	Teachers	Developers
	Students	Teachers	Developers
Simplicity			
Accessibility			
Visual Impact			
Free			
Available material			
Code readability			
Easy to improve and expand			

Conceptual design

- App is held on a **website**
- **Window organization: three vertical regions**
 - Menu
 - Configuration Options
 - Results



Conceptual design

- Results change at every user input: **Reactivity**
- Regarding **data**, users can:
 - Upload data
 - Create a **simulated** data set
 - Use a **predefined set**
- **Menu Options**: four general groups
 - Data
 - Graphs
 - Computations
 - Statistics

Conceptual design

Data

Load data set

01	02	03	04
2344	31	34	31
123	31	456	56
3452	33	23	44
33	31	31	478
5644	4532	302	34

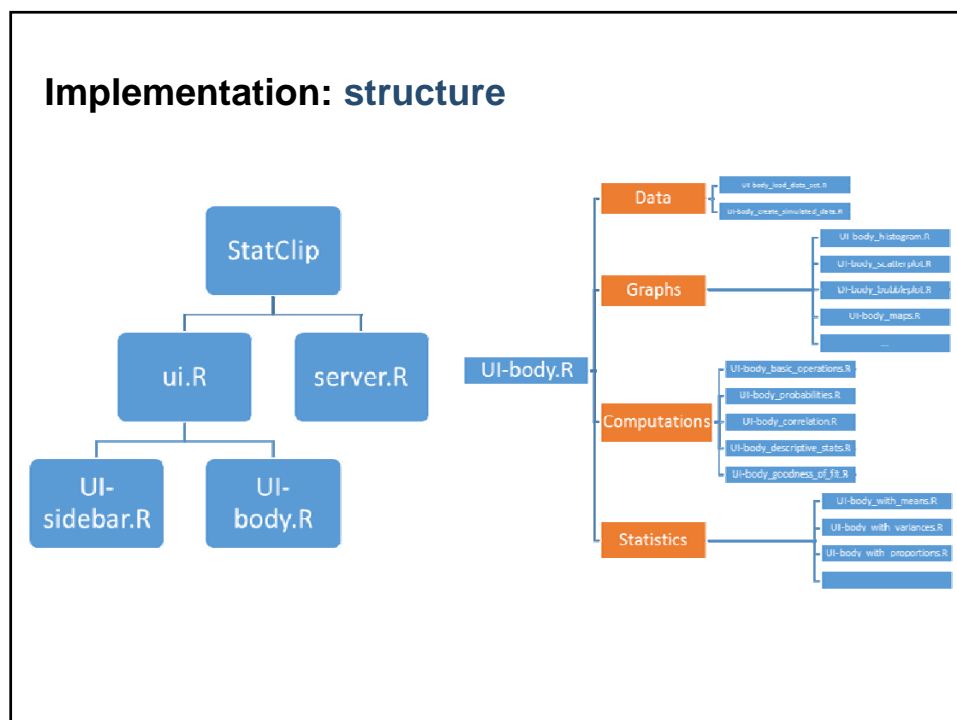
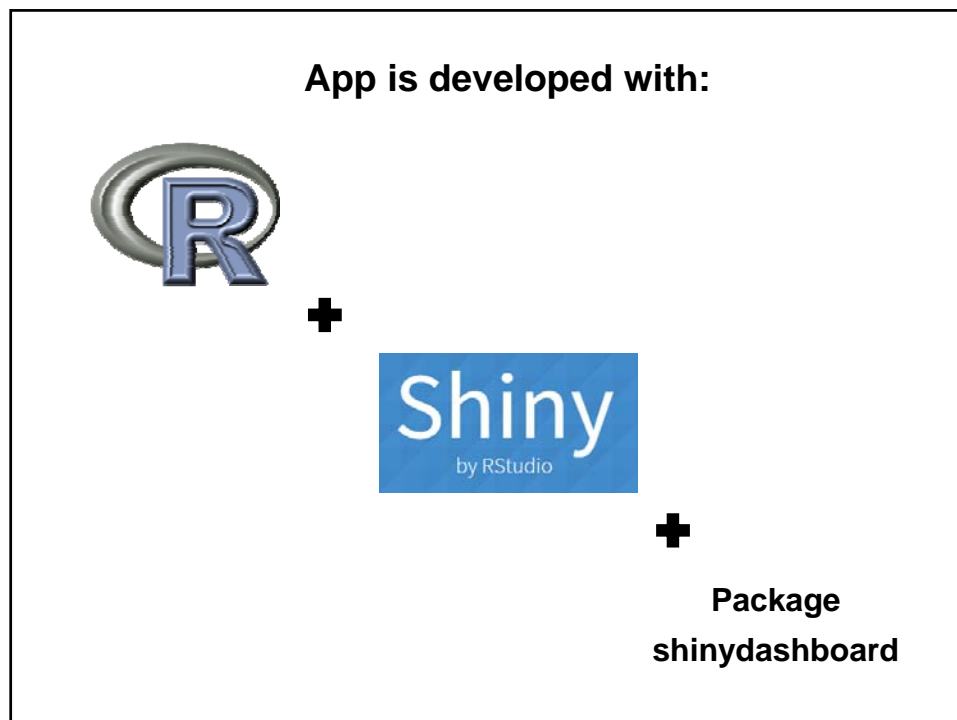
Create simulated data

01	02	03	04
2344	31	34	31
123	31	456	56
3452	33	23	44
33	31	31	478
5644	4532	302	34

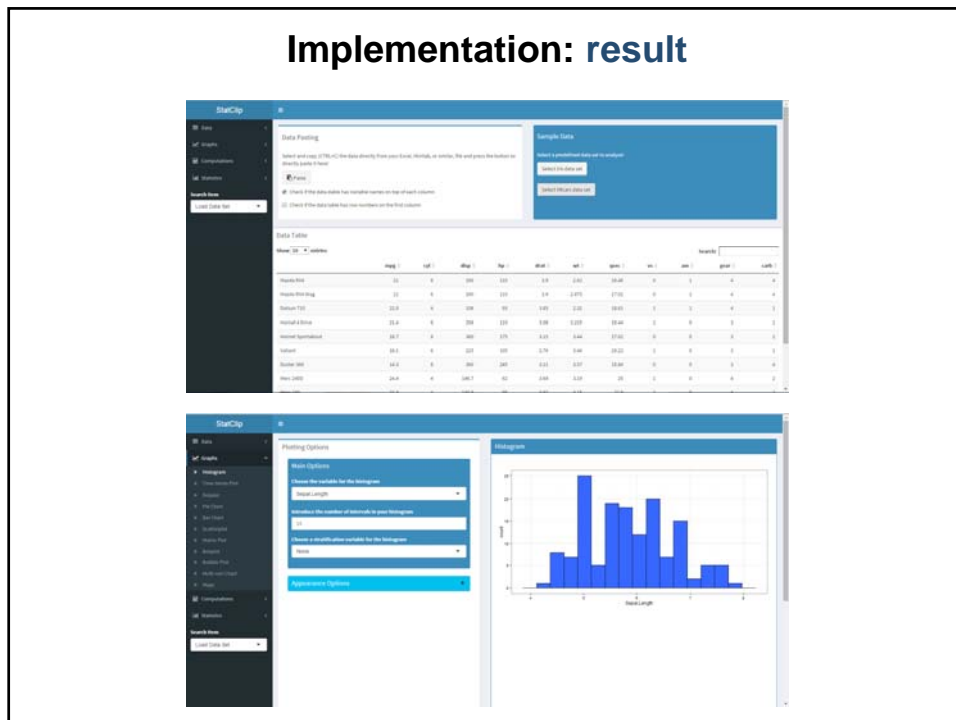
Graphs

Collapsed Options

Non collapsed Options



Implementation: result



Some examples

We'll call the app

StatClip

www.statclip.org

**We plan to have it ready during the
second semester of 2016**

Acknowledgments:



**Without the work of our student Eduard,
we wouldn't have been able to start this
project**



**for your
attention**

15th ENBIS Conference

**Development of an open application for
teaching statistics**

Eduard Serrahima

Lluís Marco-Almagro

Xavier Tort-Martorell

Pere Grima

Lourdes Rodero



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

7th September 2015