PERSONA CIÈNCIA EMPRESA
UNIVERSITAT RAMON LLULL

ASISTEMBE
Analytics, Simulations and Inquiry in STEM and Business Education
Research Group

Creating interactive visualizations with R, ggplot2 & Shiny

Vanessa Serrano, Francesc Martori, Jordi Cuadros

vanessa.serrano@iqs.edu, francesc.martori@iqs.edu, jordi.cuadros@iqs.edu

Learning Analytics Summer Institutes (LASI-Local)

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

LASI Spain 2016
Bilbao

Overview

- Software
- Why visualization?
- Grammar of graphics & ggplot2
- Recommended plots
- Shiny

Software

- **R**

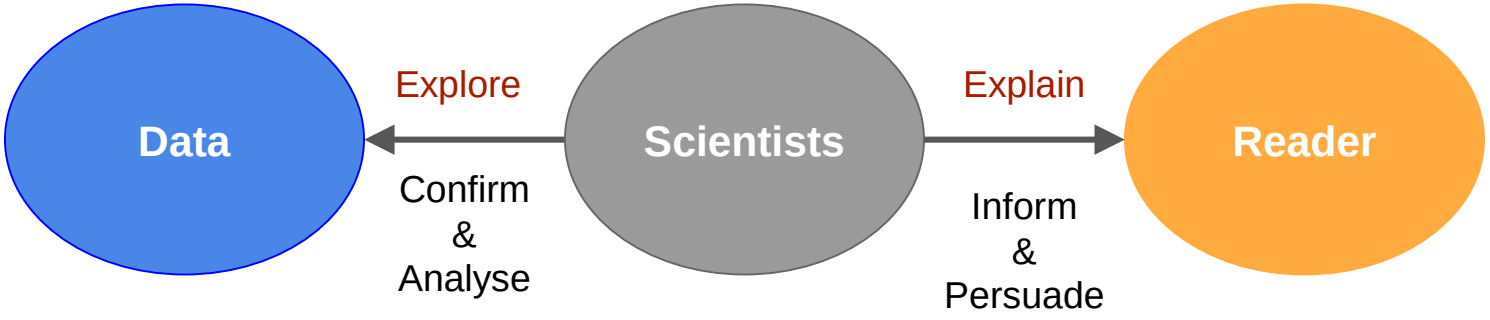
  http://cran.r-project.org/
- **RStudio**

  https://www.rstudio.com/ide/download/desktop
- **R packages: ggplot2, ggthemes, shiny**

Why visualization?

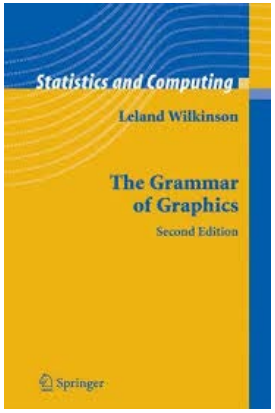
Essential component of the skills set of a data scientist

Graphical Data Analysis + Communication & Perception



Grammar of graphics

- **A tool that enables us to concisely describe the components of a graphic**
- **How to make good plots?**



Two main things:

1) Grammatical elements:

Graphics = distinct layers of grammatical elements

2) Aesthetic mappings:

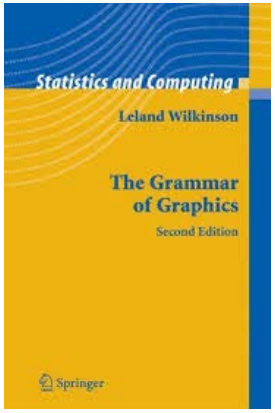
Meaningful plots through aesthetic mapping

How to map variables onto aesthetics

Aesthetics = scales = encoding elements

Grammar of graphics

- **All grammatical elements:**



Element	Description
Data	The dataset being plotted
Aesthetics	The scales onto which we map our the data
Geometries	The visual elements used for our data
Facets	Plotting small multiples
Statistics	Representations of our data to aid understanding
Coordinates	The space on which the data will be plotted
Themes	All non-data ink

Grammar of graphics & ggplot2

ggplot2

Open source implementation of the grammar of graphics for R

In ggplot2 a plot is an object that we can manipulate (not an image)

Example

```
g <- ggplot(tempdf,aes(y=PV1MATH,x=PV1READ))
g + geom_point()+theme_few()
```

Grammar of graphics & ggplot2

Aesthetic	Description
x	X axis position
y	Y axis position
colour	Colour of dots, outlines or other shapes
fill	Fill colour
size	Diameter of points, thickness of lines
alpha	Transparency
linetype	Line dash pattern
labels	Text on a plot or axes
shape	Shape

Continuous variables: x, y, size, alpha, colour, fill

Categorical variables: labels, fill, shape, alpha, linetype, size

Grammar of graphics & ggplot2

Common plot types	Geometries
Scatter plots	point
	jitter
	abline
Bar plots	histogram
	bar
	errorbar
Line plots	line

37 geometries in total

Recommended plots

Let's start with some data...

PISA 2012 - original dataset

- in OCDE
  - http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm
  - http://www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm (Codebooks)
- in R
  - https://github.com/pbiecek/PISA2012lite

Recommended plots

PISA 2012 - subset

- Data for Spanish students in Compulsory secondary education - 25152 obs.

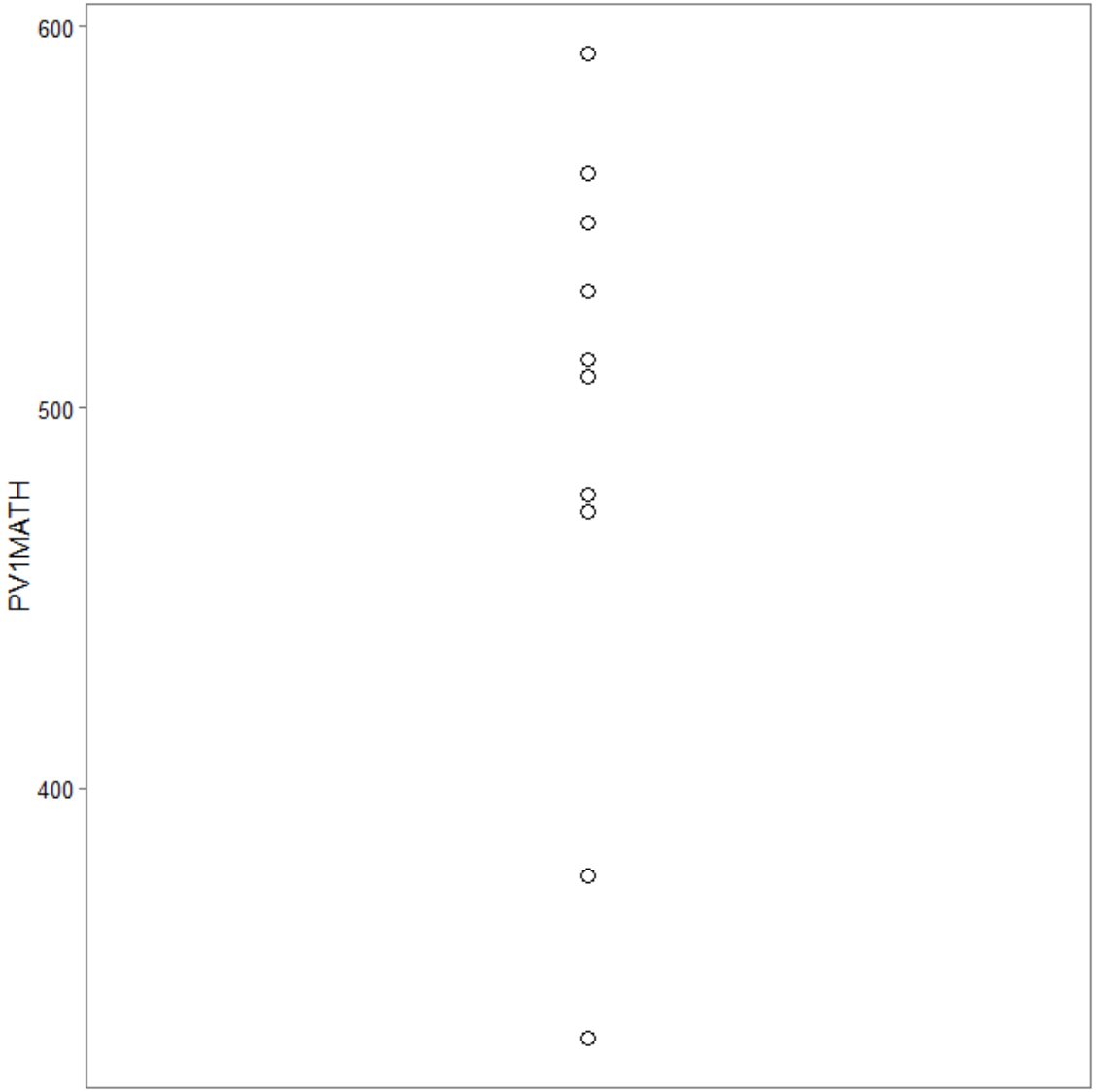
SUBNATIO, STRATUM, STIDSTD, BIRTHMONTH, GENDER, MOTHERQUAL, MOTHERJOB, FATHERQUAL, FATHERJOB, NUMBOOKS, TIMEINT, WEALTH, ATTSCHOOLOBJ, ATTSCHOOLACT, PV1MATH, PV1MACC, PV1MACQ, PV1MACS, PV1MACU, PV1MAPE, PV1MAPF, PV1MAPI, PV1READ, PV1SCIE

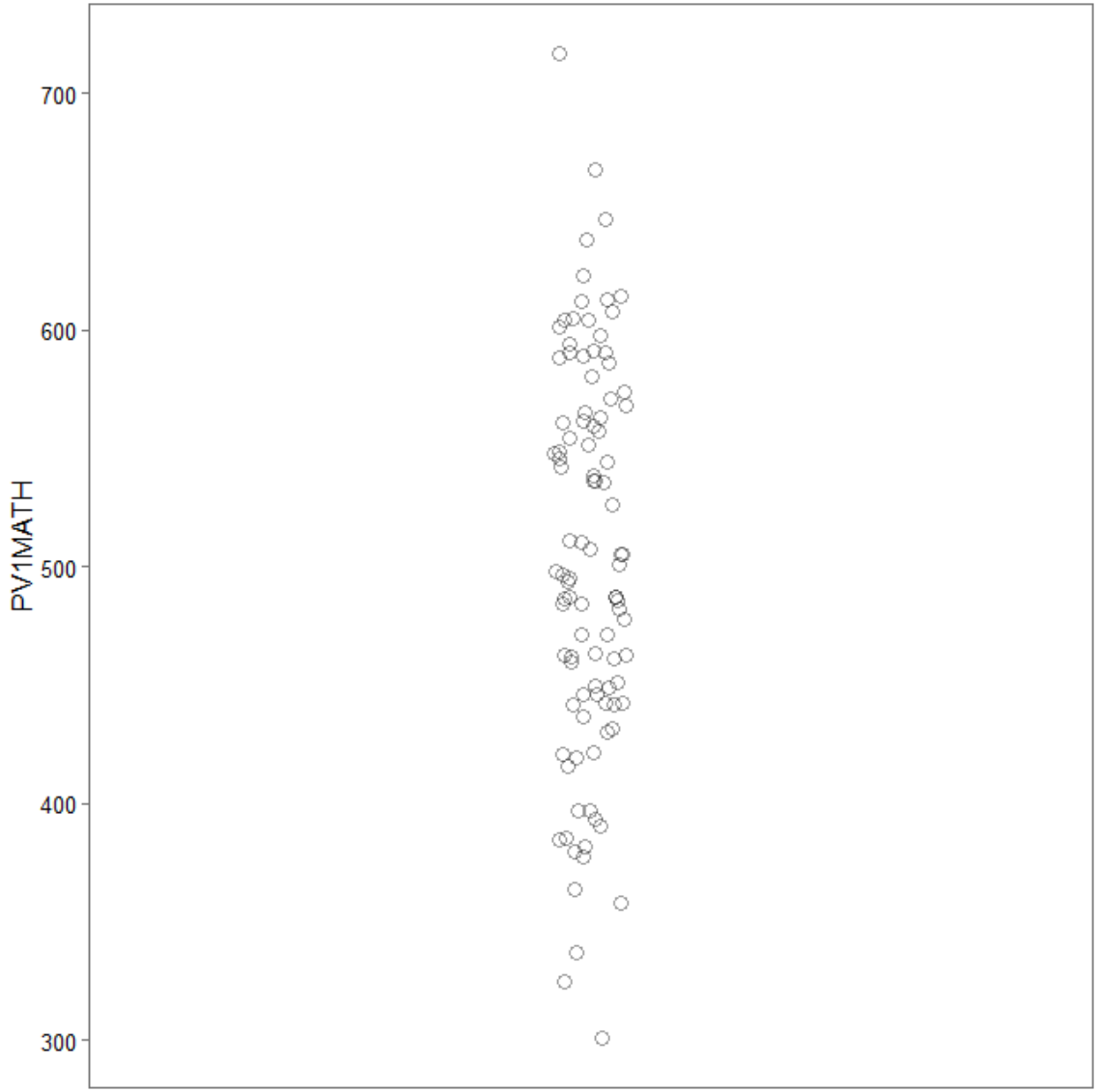
What’s the distribution of the results in maths?

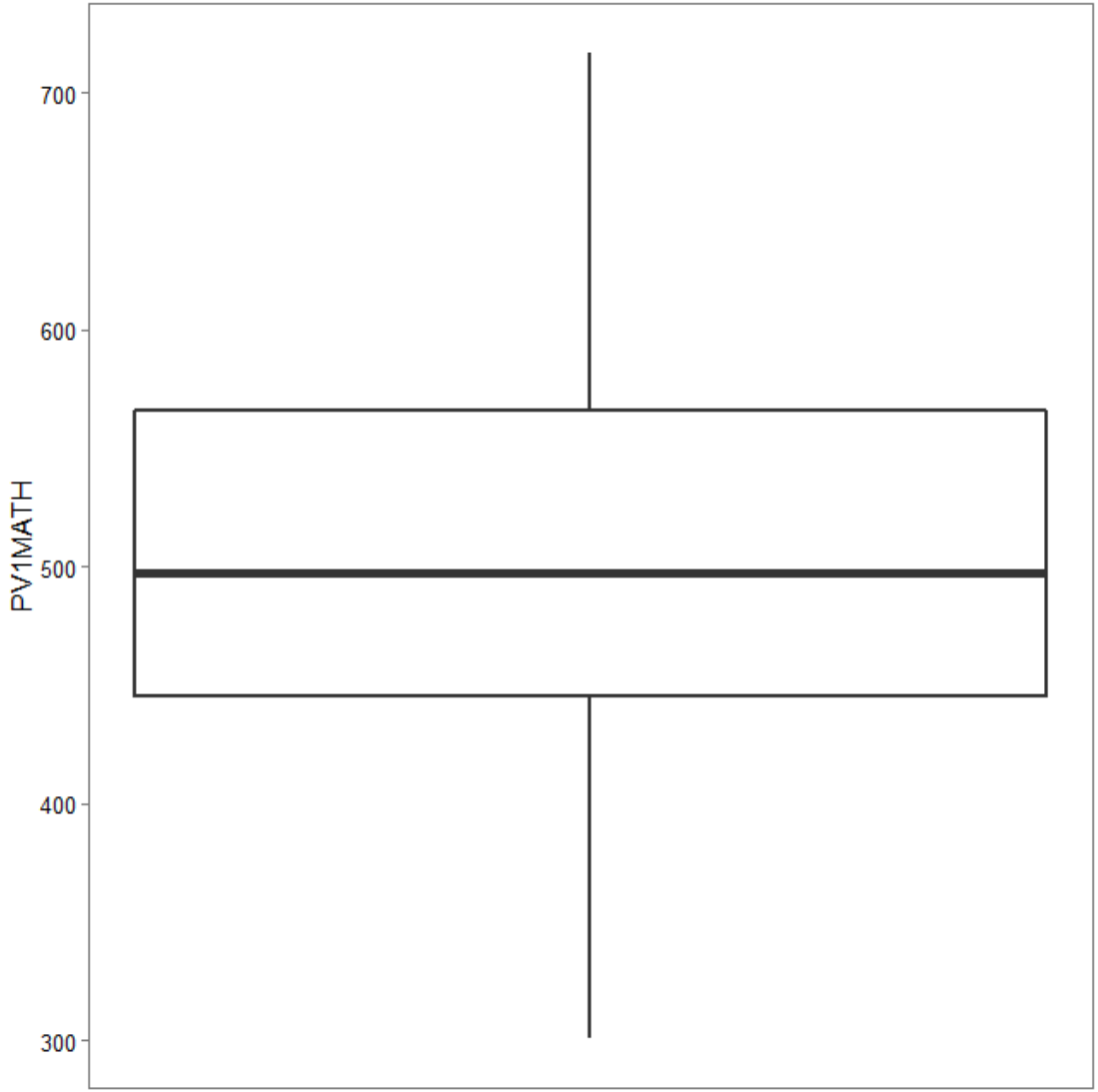
Recommended plots

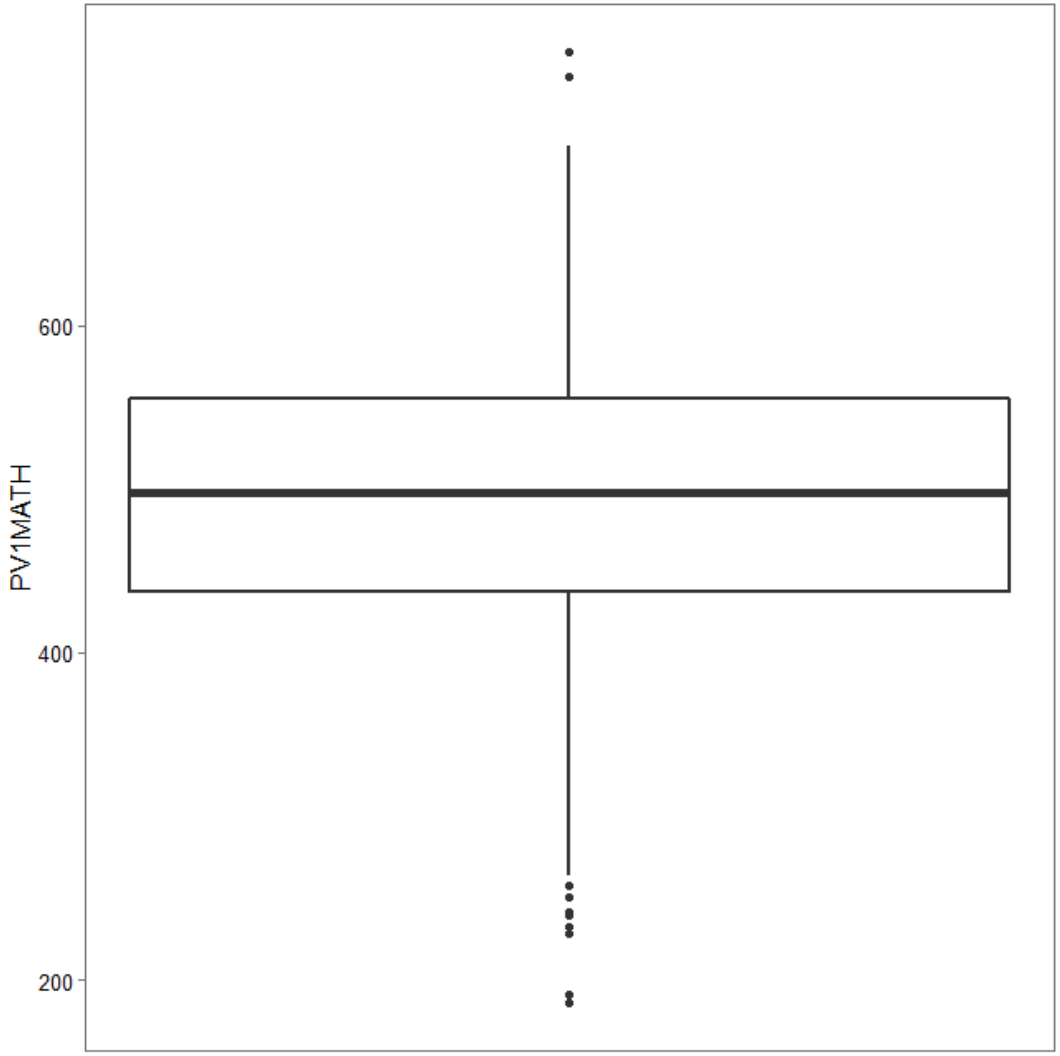
- **Visualize the distribution of values in a quantitative/continuous variable: PV1MATH**

- **"Dot-plot": useful for small data sizes**
  - Transparency, dodging, jittering
- **Boxplot: useful for middle data sizes**
  - Add dot-plot, remove outliers
- **Histogram: with density**


600
500
400
PV1MATH

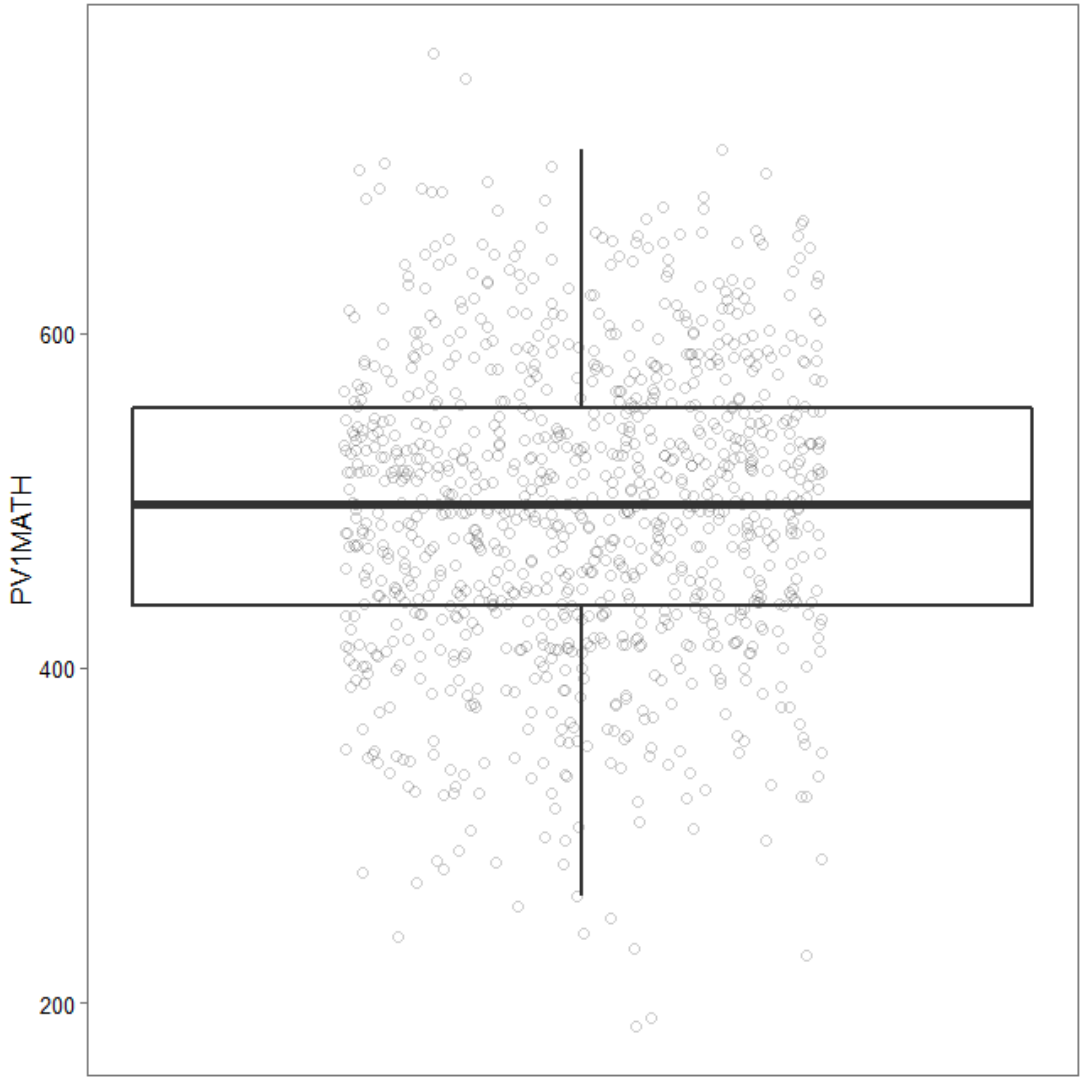

PV1MATH
700
600
500
400
300


700
600
500
400
300
PV1MATH


PV1MATH
600
400
200

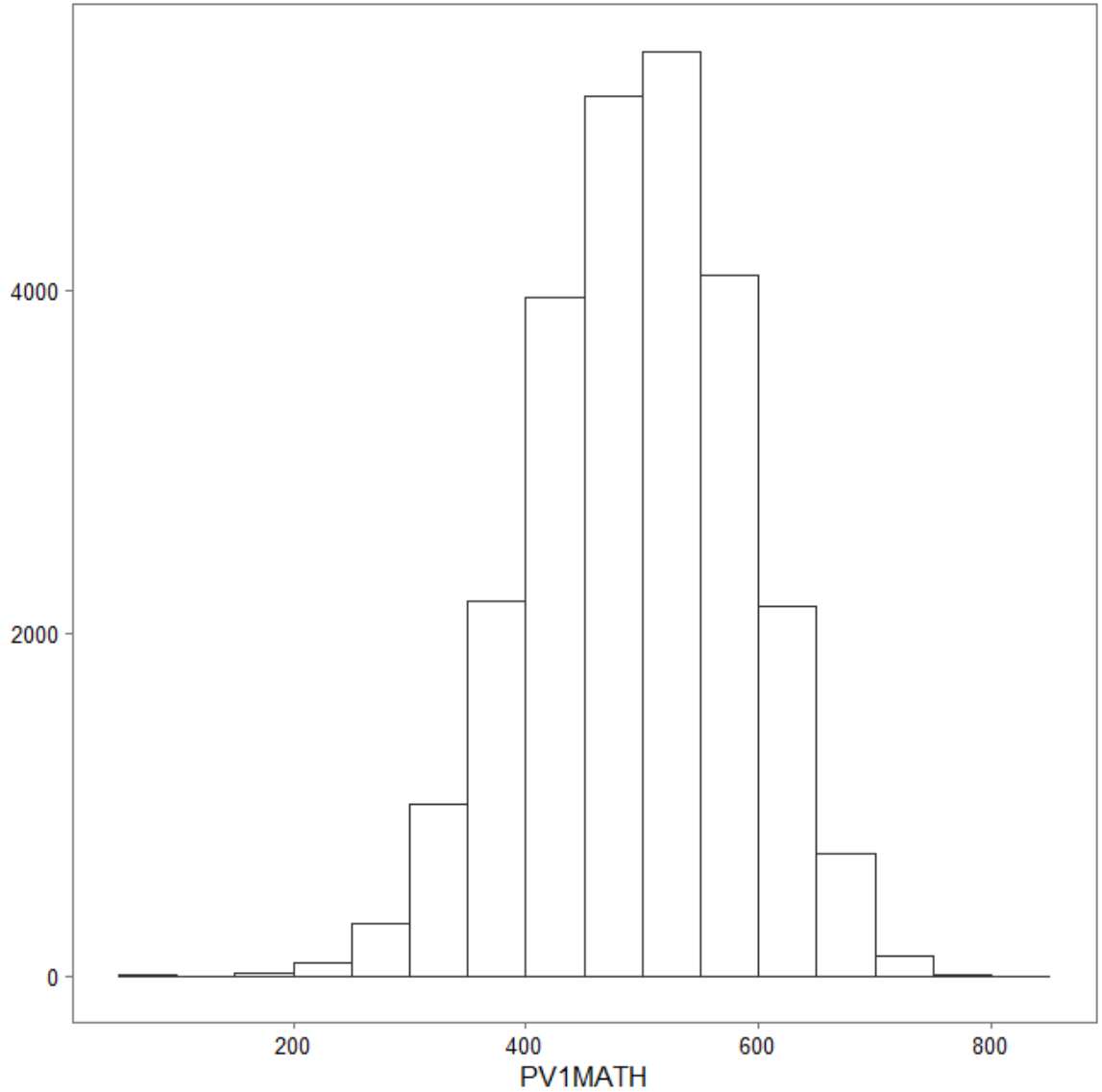
PERSONA CIÈNCIA EMPRESA
UNIVERSITAT RAMON LLULL

ASISTEMBE
Analytics, Simulations and Inquiry in STEM and Business Education
Research Group



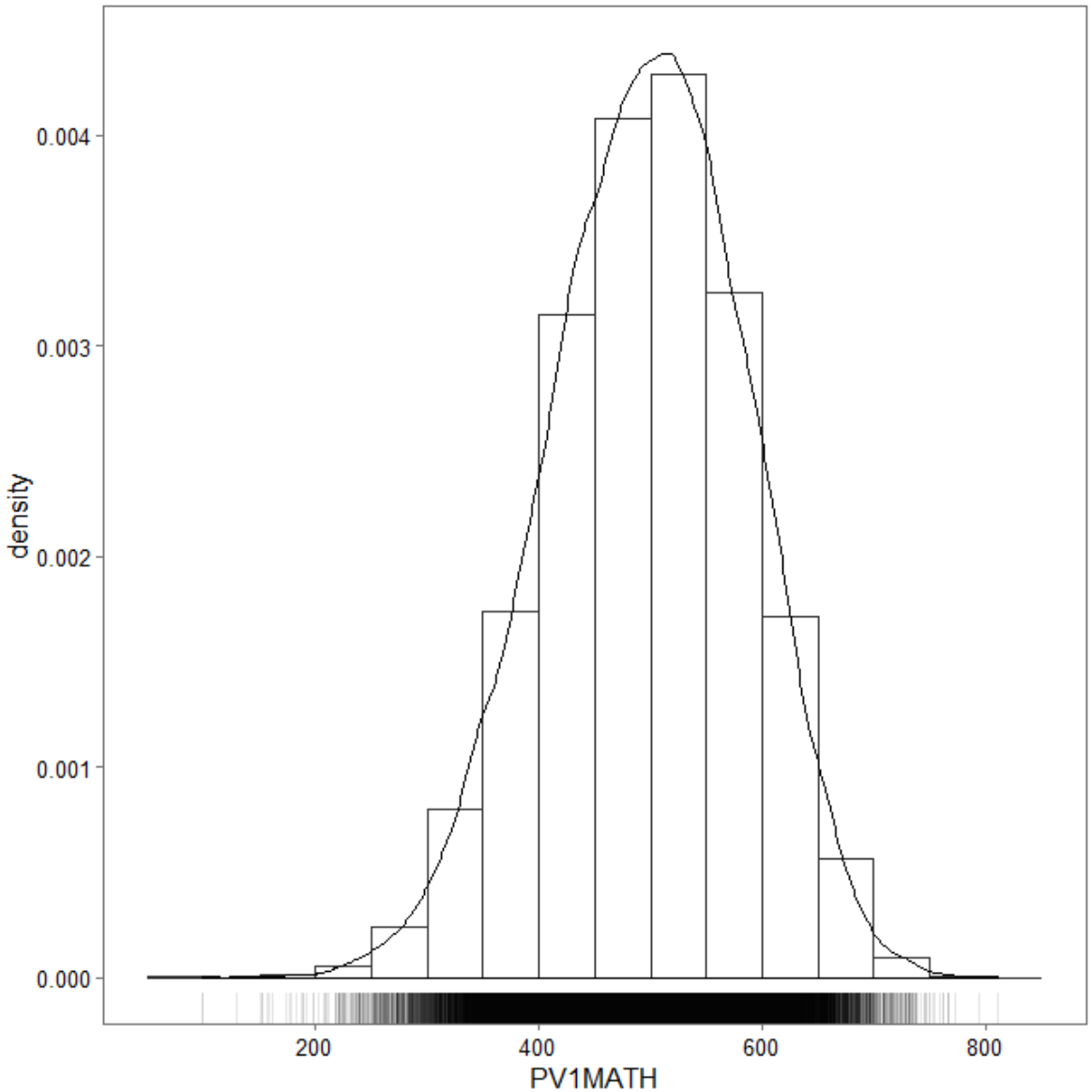
PERSONA CIÈNCIA EMPRESA
UNIVERSITAT RAMON LLULL

ASISTEMBE
Analytics, Simulations and Inquiry in STEM and Business Education
Research Group



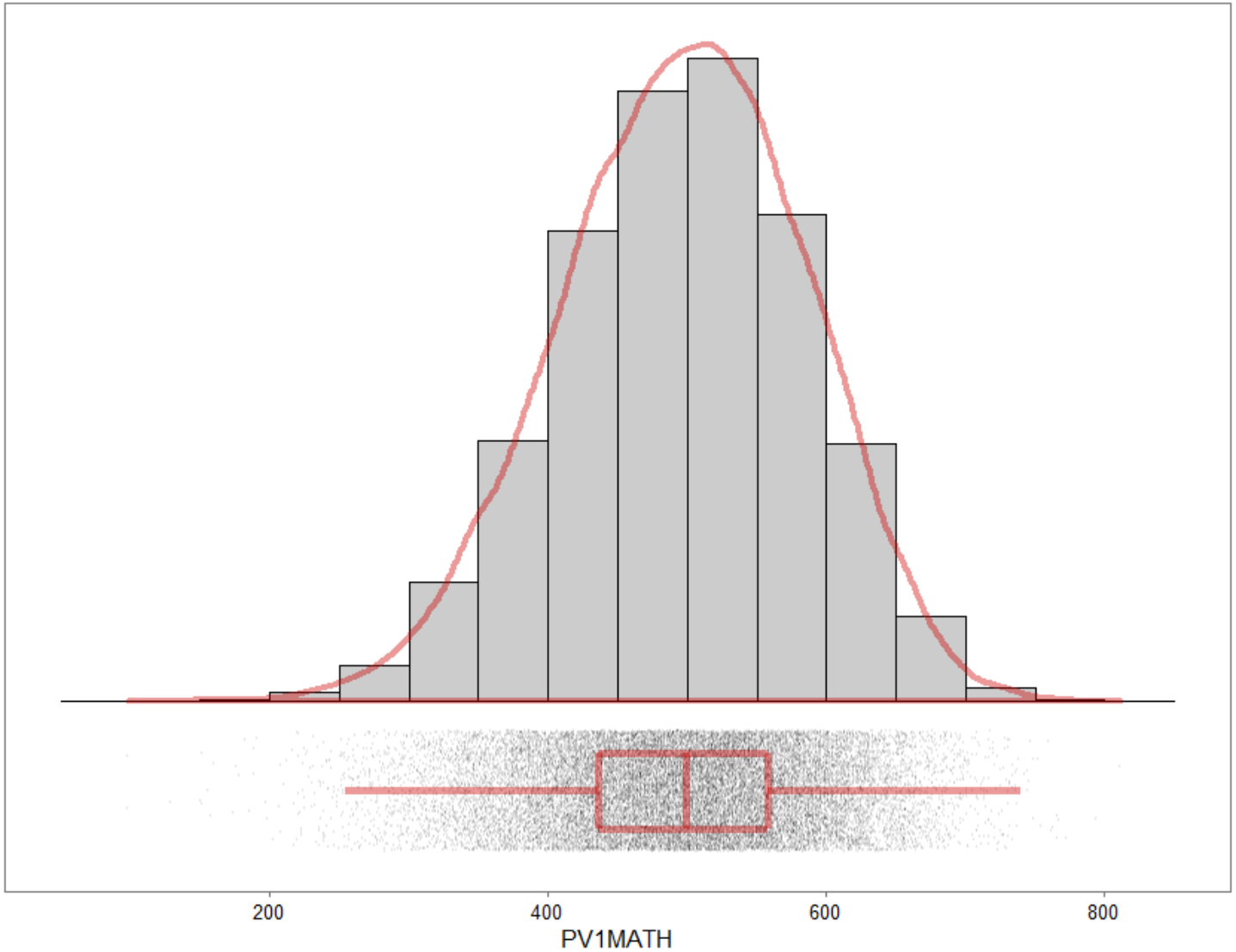
PERSONA CIÈNCIA EMPRESA
UNIVERSITAT RAMON LLULL

ASISTEMBE
Analytics, Simulations and Inquiry in STEM and Business Education
Research Group

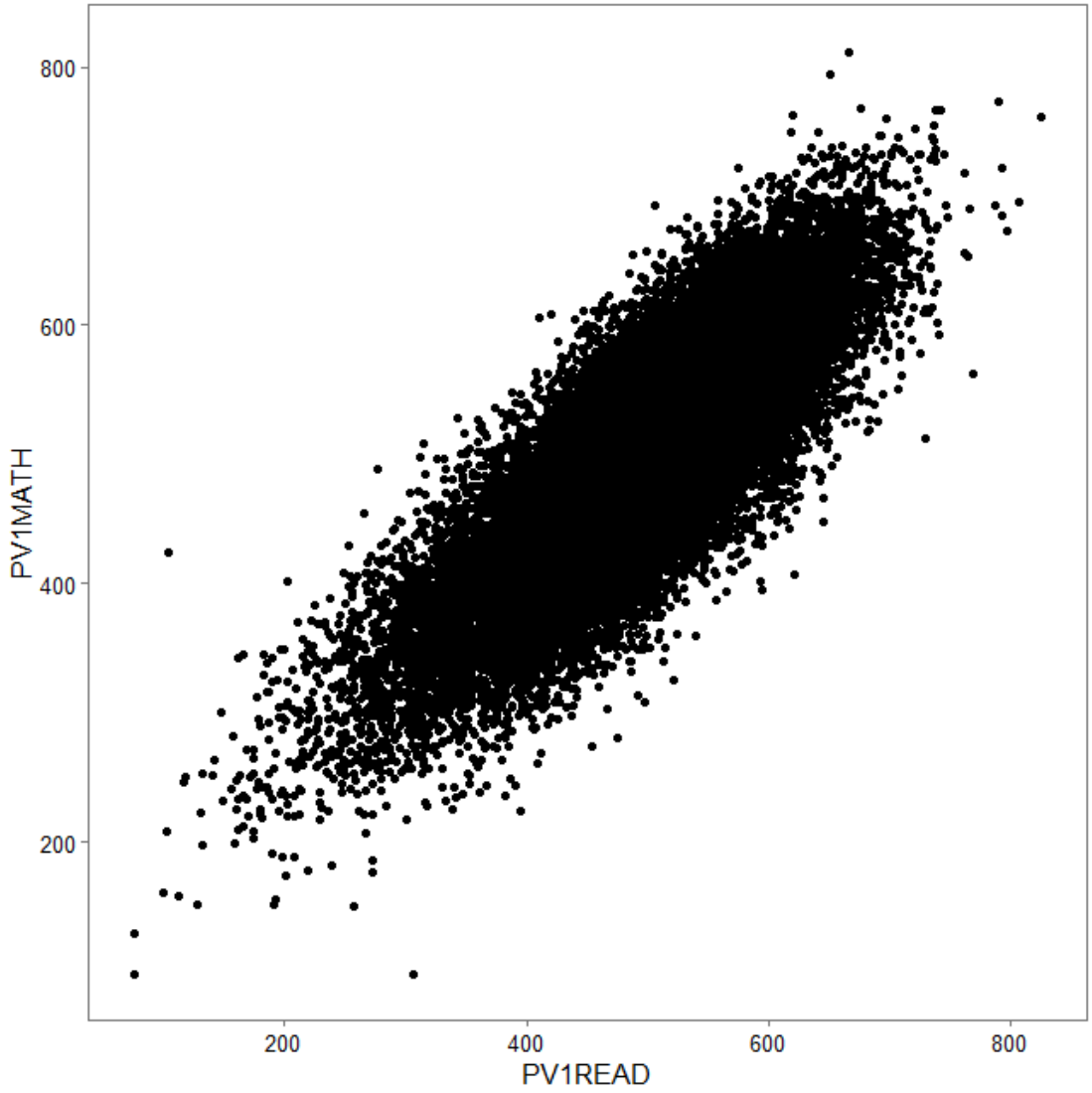


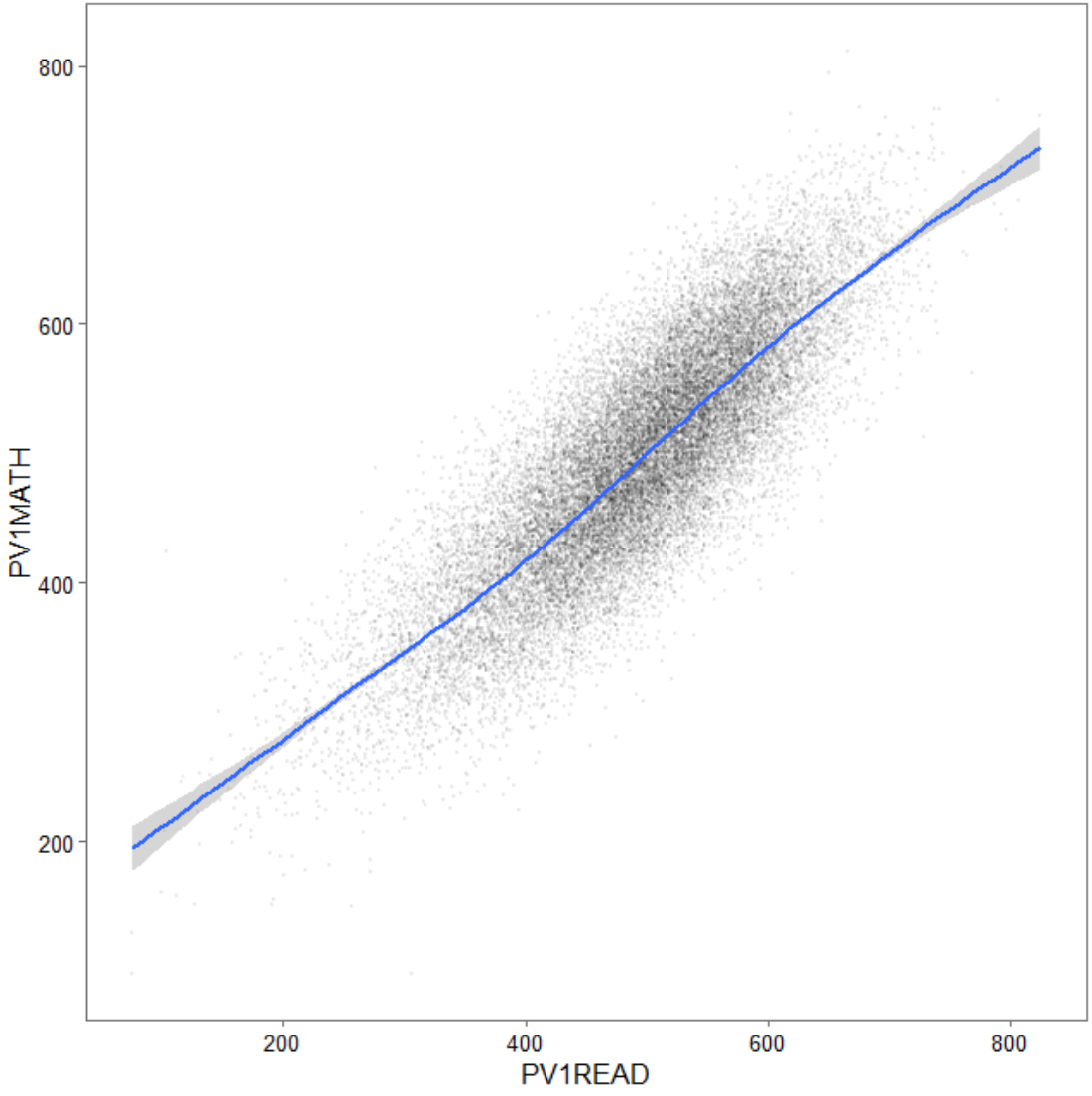
PERSONA CIÈNCIA EMPRESA
UNIVERSITAT RAMON LLULL

ASISTEMBE
Analytics, Simulations and Inquiry in STEM and Business Education
Research Group



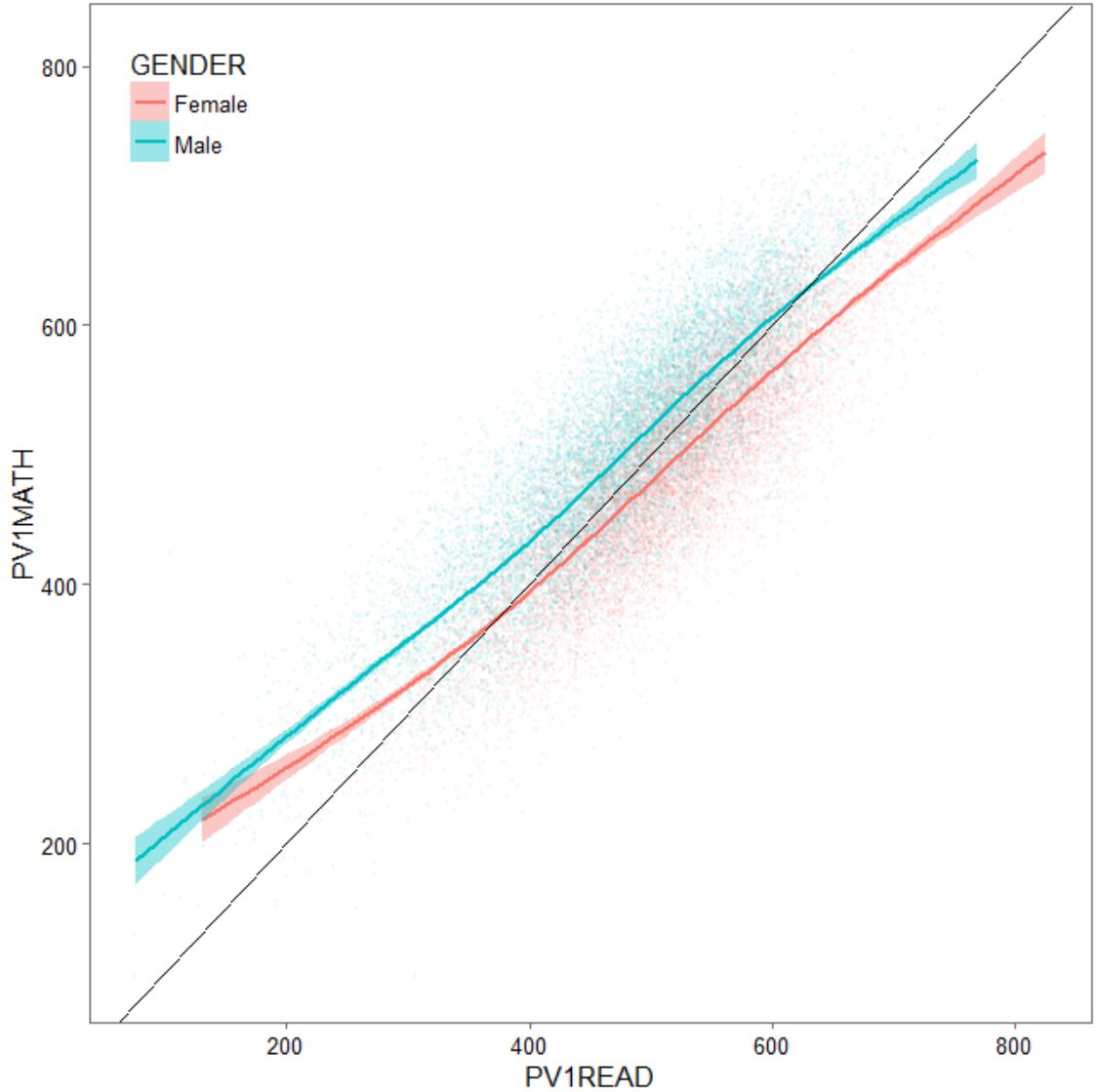
What’s the relation between the results in maths and the results in reading?


800
600
400
200
PV1MATH
200
400
600
800
PV1READ


800
600
400
200
PV1MATH
200
400
600
800
PV1READ

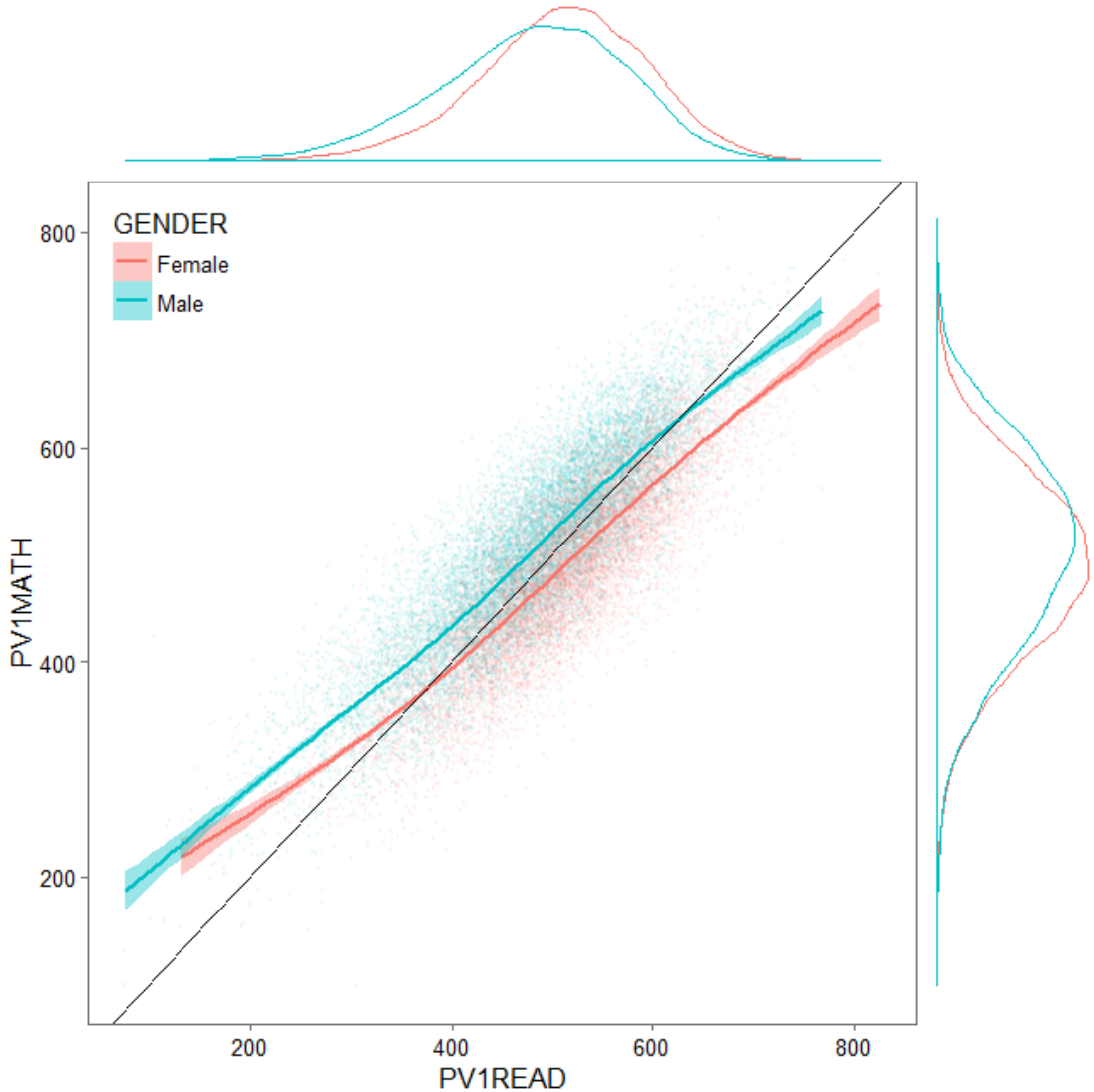
PERSONA CIÈNCIA EMPRESA
UNIVERSITAT RAMON LLULL

ASISTEMBE
Analytics, Simulations and Inquiry in STEM and Business Education
Research Group



PERSONA CIÈNCIA EMPRESA
UNIVERSITAT RAMON LLULL

ASISTEMBE
Analytics, Simulations and Inquiry in STEM and Business Education
Research Group



Are fathers and mothers equally educated?

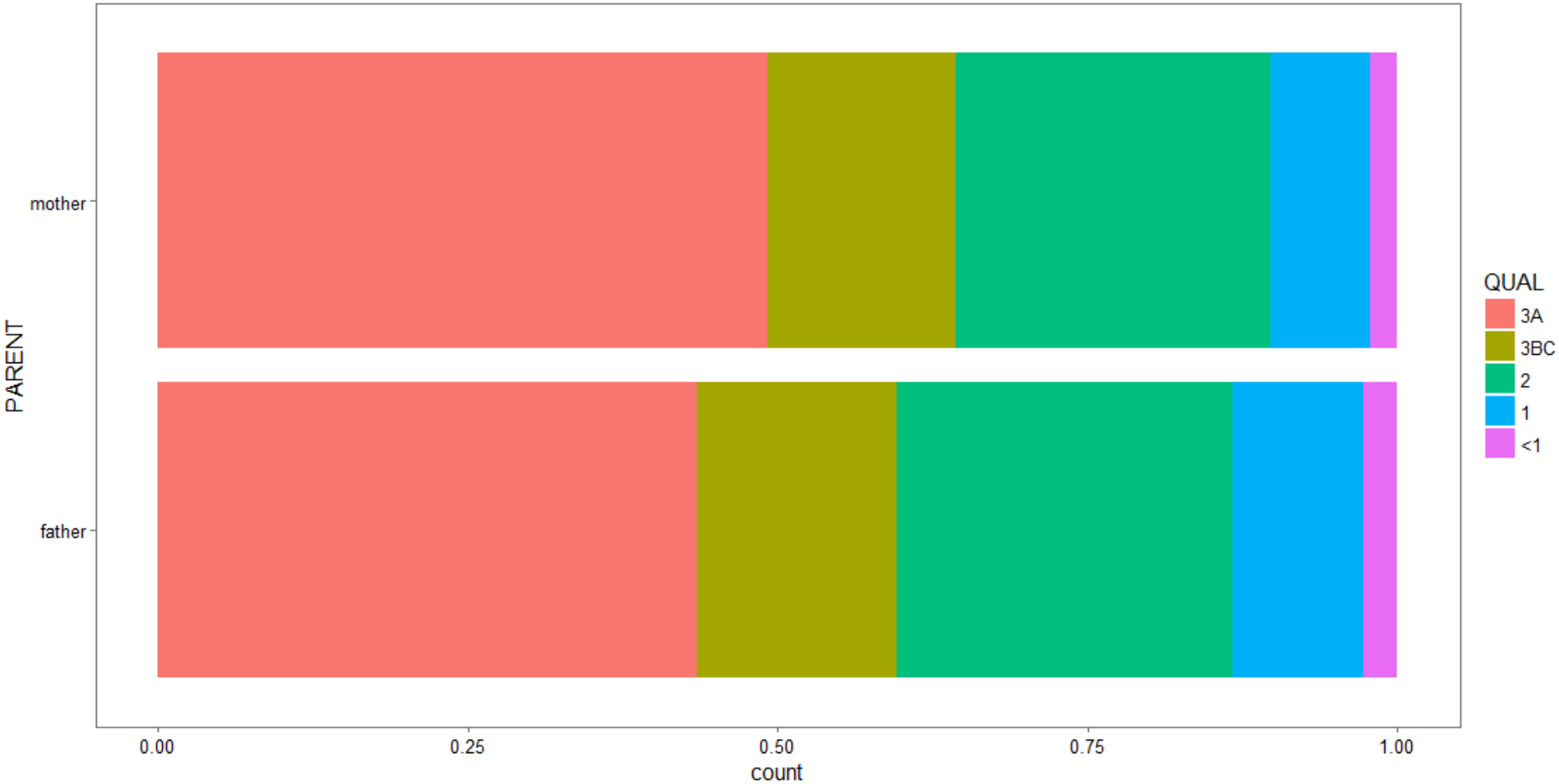
Recommended plots

- **Compare relative frequencies of two ordinal variables: MOTHERQUAL, FATHERQUAL**

- **Should we use a circular diagram (aka pie chart)?**
  - No...

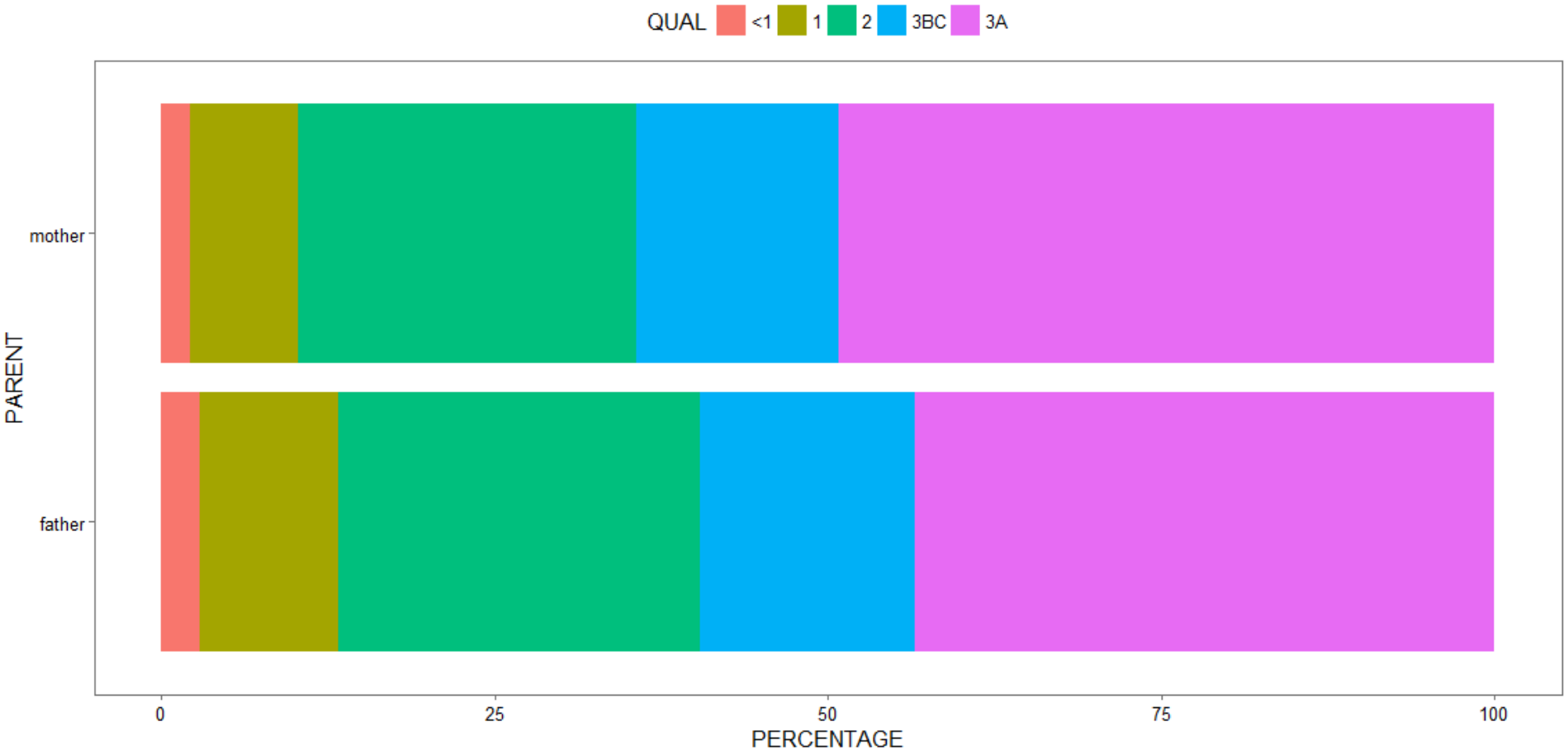
An alternative: stacked bar plot

http://www.r-chart.com/2010/07/pie-charts-in-ggplot2.html
http://www.storytellingwithdata.com/blog/2014/06/alternatives-to-pies


mother
father
PARENT
0.00
0.25
0.50
0.75
1.00
count
QUAL
3A
3BC
2
1
<1

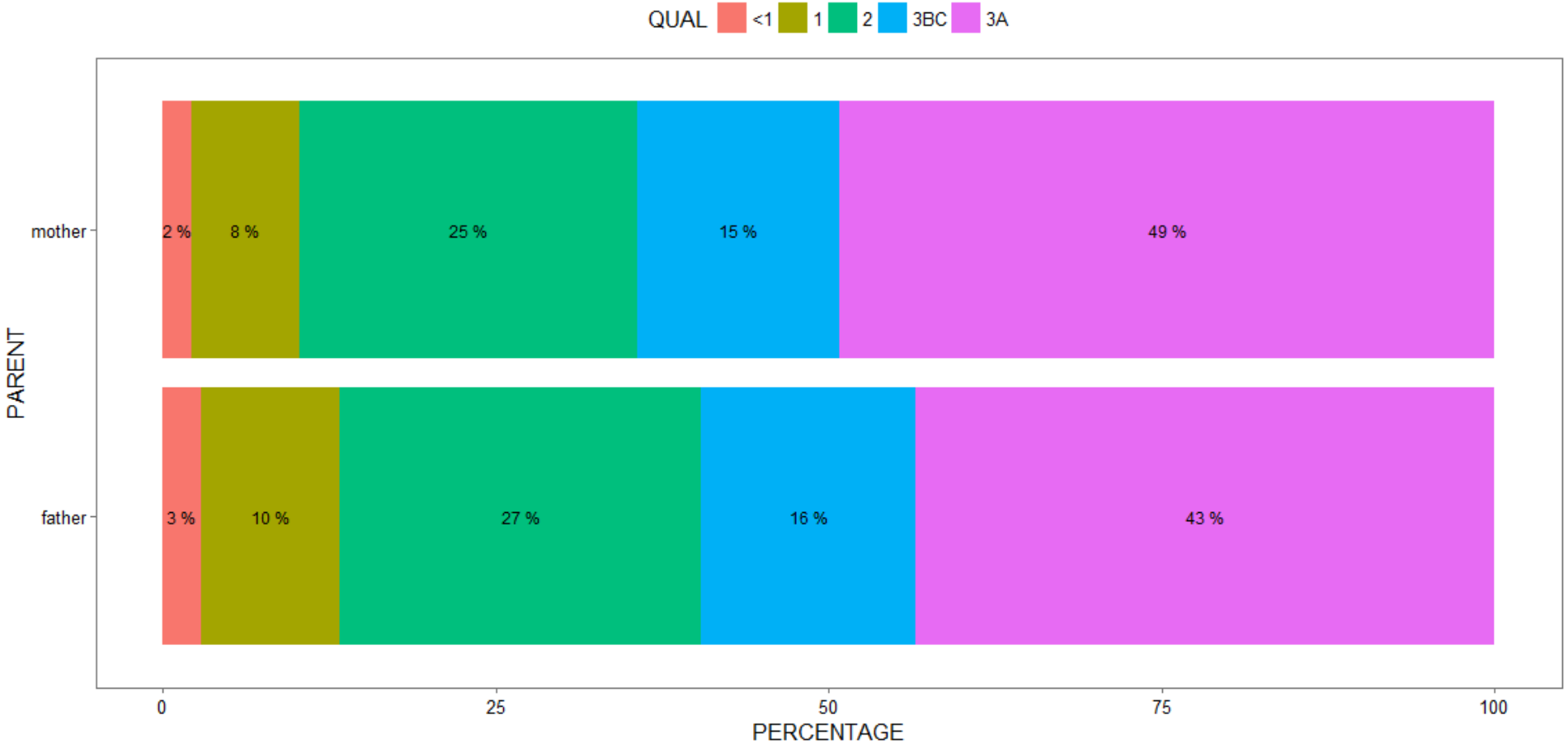
PERSONA CIÈNCIA EMPRESA
UNIVERSITAT RAMON LLULL

ASISTEMBE
Analytics, Simulations and Inquiry in STEM and Business Education
Research Group



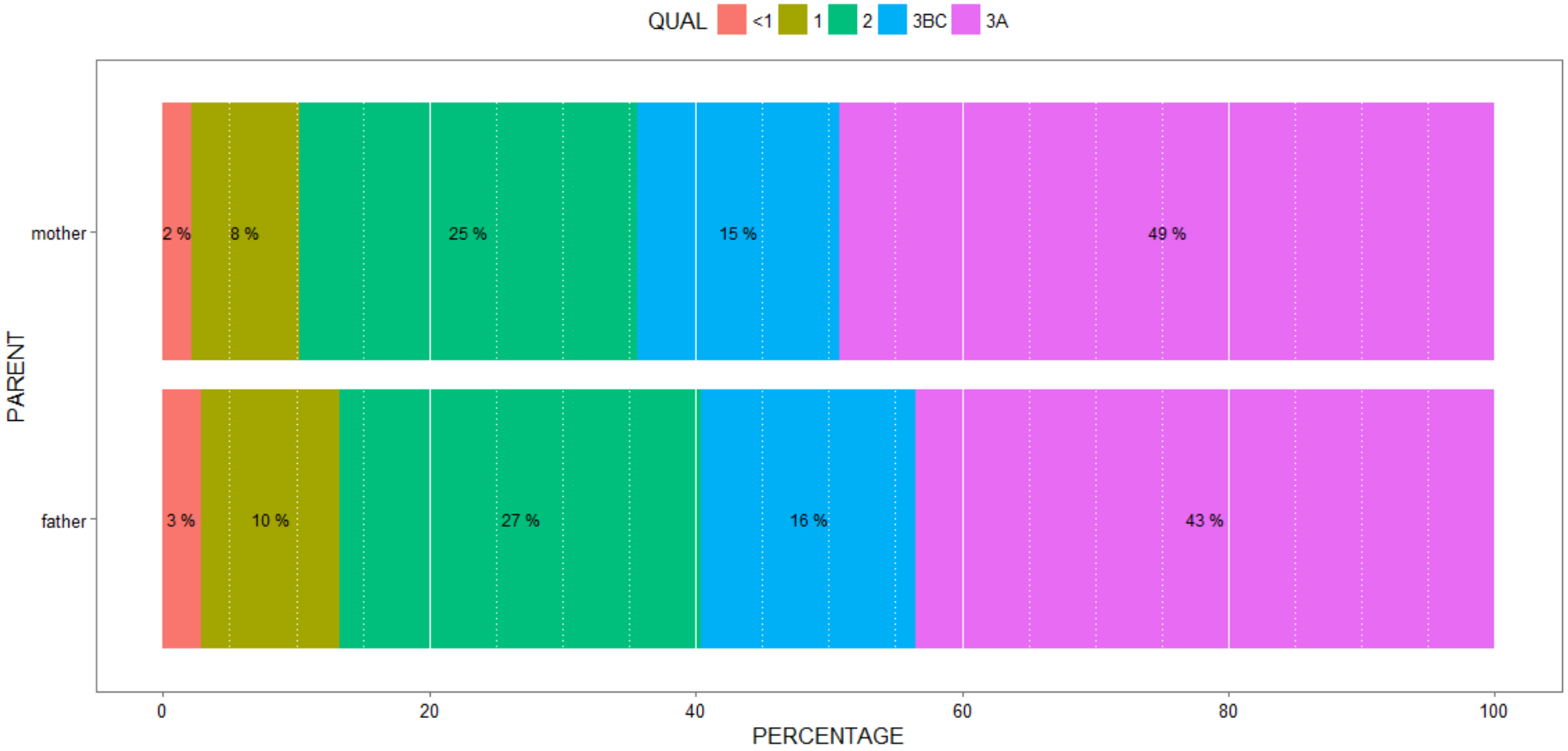
PERSONA CIÈNCIA EMPRESA
UNIVERSITAT RAMON LLULL

ASISTEMBE
Analytics, Simulations and Inquiry in STEM and Business Education
Research Group



PERSONA CIÈNCIA EMPRESA
UNIVERSITAT RAMON LLULL

ASISTEMBE
Analytics, Simulations and Inquiry in STEM and Business Education
Research Group



Shiny

Shiny is an open source R package that provides a web framework for building web applications using R. Shiny helps turn analyses into interactive web applications without requiring HTML, CSS, or JavaScript knowledge.

- **http://shiny.rstudio.com/**

Shiny

A Shiny app is a web page (ui.R) connected to a computer running a live R session (server.R)

It can be run locally from your computer or remotely from a server

- **http://www.rstudio.com/wp-content/uploads/2016/01/shiny-cheatsheet.pdf**

Two files and a Shiny App

- ui - nested R functions that assemble an HTML user interface for your app

- server - a function with instructions on how to build and rebuild the R objects displayed in the UI

- shinyApp - combines ui and server into a functioning app. Wrap with runApp() if calling from a sourced script or inside a function.

Two files and a Shiny App

ui.R is essentially an HTML document. Use Shiny's functions to assemble this HTML with R.

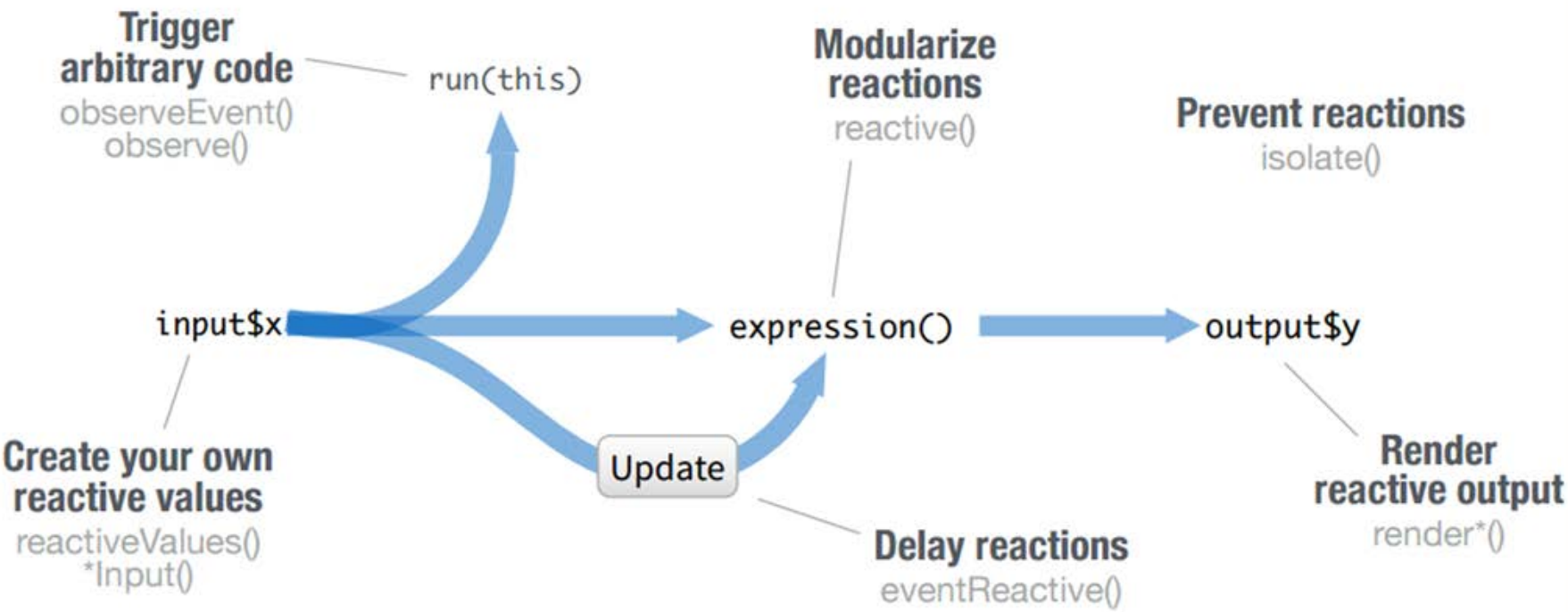
- Layouts to organize and combine multiple elements
- Inputs to collect values from the user
- Outputs to present results, plots …

server.R will render outputs with R using inputs and static information.

The key is automatic "reactive" binding between inputs and outputs

Reactivity

Reactive values work together with reactive functions. Call a reactive value from within the arguments of one of these functions to avoid the error `Operation not allowed without an active reactive context.`



Shiny resources

- http://rstudio.github.io/shiny/tutorial/
- https://plot.ly/r/shiny-gallery/
- https://plot.ly/r/shiny-tutorial/

PERSONA CIÈNCIA EMPRESA
UNIVERSITAT RAMON LLULL

ASISTEMBE

Analytics, Simulations and Inquiry
in STEM and Business Education
Research Group