



# DATA-DRIVEN DECISION MAKING IN EDUCATION

*An introduction to R on PISA datasets*



Agnes Salanki

June 2019



Hotels.com™

## Before we start

**1. Current presentation:**  
[http://bit.ly/wosr\\_slides](http://bit.ly/wosr_slides)

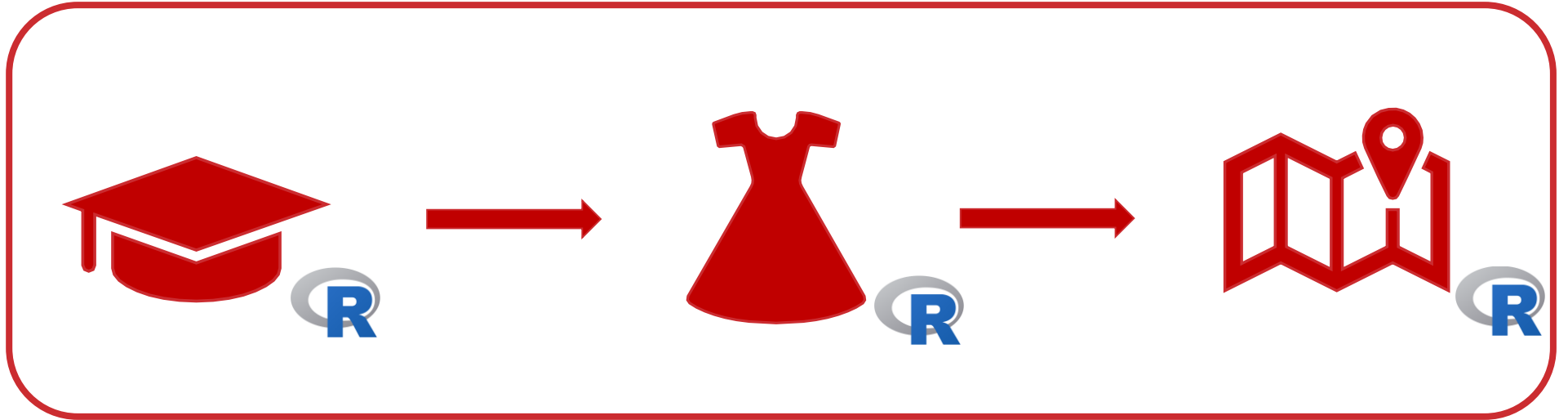
**2. Go to**

**<https://rstudio.cloud>**

**and create an account**



# Who am I?



Go to <https://rstudio.cloud> and create an account 

# What is R?

- Programming language/software environment
- Statistics and visualization
- Platform-independent



Go to <https://rstudio.cloud> and create an account



# Why am I talking about R here?

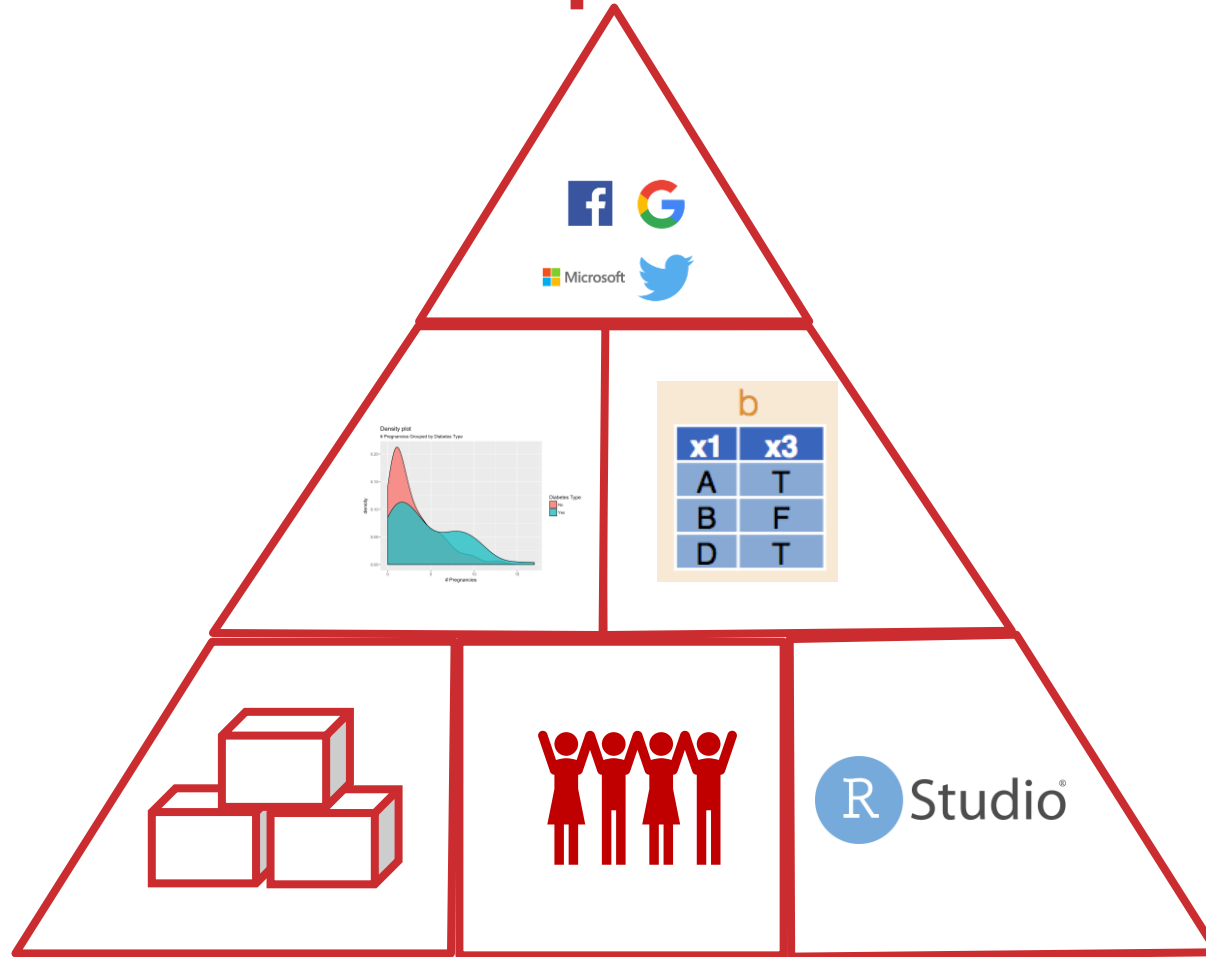
**I love**

- **the language**
- **the community**
- **sharing the word**



Go to <https://rstudio.cloud> and create an account

# Outline of the workshop



Go to <https://rstudio.cloud> and create an account



# Packages



# Why is it so easy to work with R?

## CRAN

Comprehensive R Archive Network

```
library(numbers)
primeFactors()
```

```
library(fun)
mine_sweeper()
```

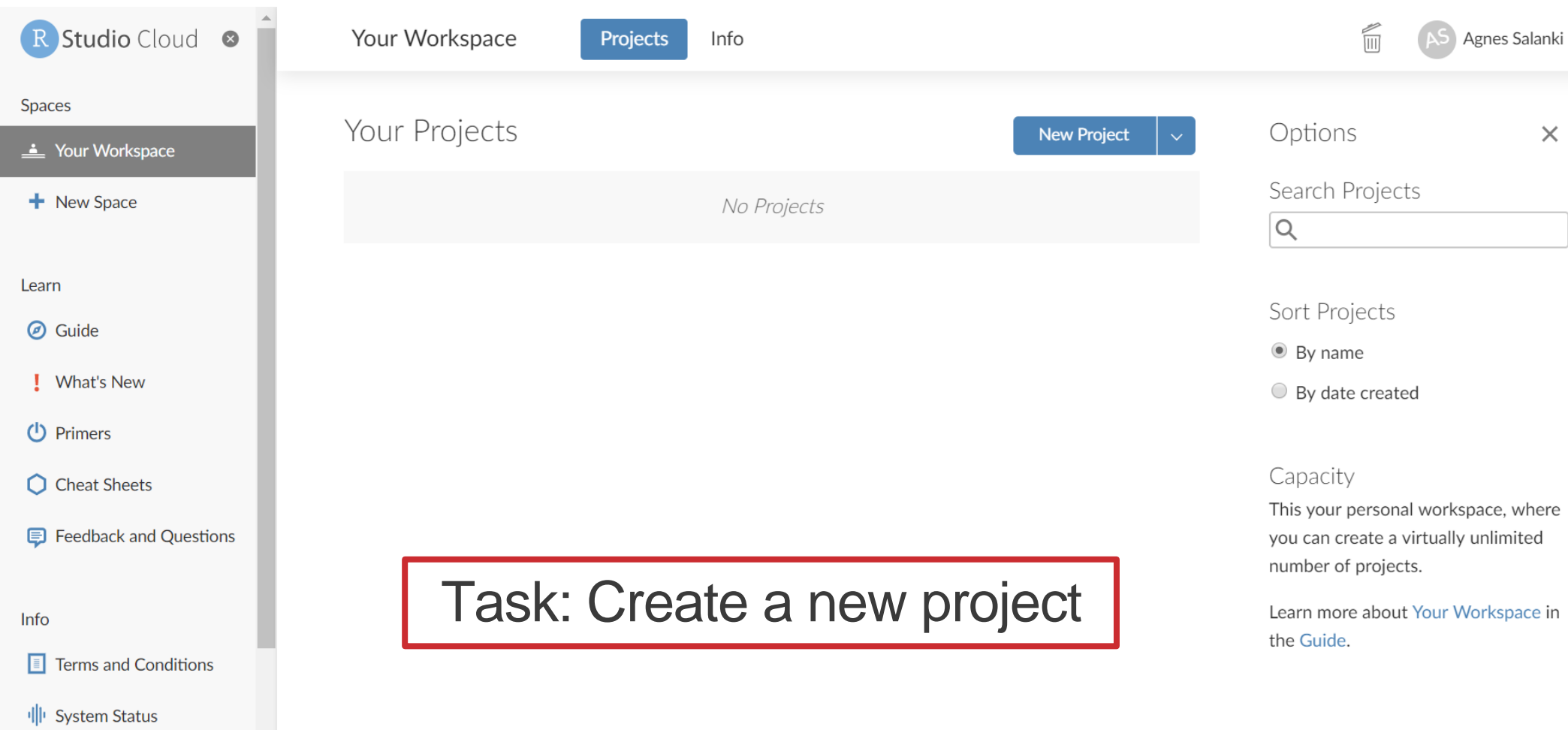
●	2	1	2	2	2	2	1	1	0
1	2	●	2	●	●	3	●	2	0
0	2	2	3	2	2	4	●	4	1
0	1	●	2	1	0	2	●	4	●
0	2	3	●	2	1	1	1	3	●
1	2	●	3	●	1	0	1	3	3
●	2	1	2	1	1	1	2	●	●
1	1	0	1	2	2	2	●	4	3
0	0	0	1	●	●	2	2	●	1
0	0	0	1	2	2	1	1	1	1



Go to <https://rstudio.cloud> and create an account



# Why is it so easy to work with R?



The screenshot displays the R Studio Cloud web interface. On the left is a sidebar with navigation links: Spaces (Your Workspace, New Space), Learn (Guide, What's New, Primers, Cheat Sheets, Feedback and Questions), and Info (Terms and Conditions, System Status). The main area has tabs for 'Your Workspace', 'Projects' (selected), and 'Info'. Below the 'Projects' tab, it says 'Your Projects' with a 'New Project' button and a dropdown arrow. A large grey box in the center contains the text 'No Projects'. On the right, there are 'Options' (with a close button), a 'Search Projects' search bar, 'Sort Projects' (with radio buttons for 'By name' and 'By date created'), and a 'Capacity' section explaining that it's a personal workspace for creating unlimited projects, with a link to the 'Guide'.

Task: Create a new project

# Why is it so easy to work with R?

## Task: Install the tidyverse package group

The screenshot shows the RStudio interface with the 'Install Packages' dialog box open. The dialog is titled 'Install Packages' and contains the following fields and options:

- Install from:** A dropdown menu set to 'Repository (CRAN, RSPM)' with a link to 'Configuring Repositories'.
- Packages (separate multiple with space or comma):** A text input field containing 'tidyverse'.
- Install to Library:** A dropdown menu set to '/home/rstudio-user/R/x86\_64-pc-linux-gnu-library/3.6 [Default]'.
- Install dependencies:** A checked checkbox.
- Buttons:** 'Install' and 'Cancel' buttons at the bottom.

The background shows the RStudio interface with the following panes:

- Console:** Displays R version 3.6.0, copyright information, and a prompt for the user to type commands like 'license()', 'demo()', 'help.start()', and 'q()'.
- Environment:** Shows the 'Global Environment' and a message 'Environment is empty'.
- Packages:** A table showing installed packages in the 'User Library'.

Name	Description	Version
base	The R Base Package	3.6.0
boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-22
class	Functions for Classification	7.3-15
cluster	"Finding Groups in Data":	2.0.8



# Community



# Where can I get help?

Task: load the tidyverse package group RUNNING  
*library(tidyverse) (Ctrl + Enter)*

## Meetups

e.g., LondonR,  
R-Ladies London

## Conferences

e.g., satRdays

## Help pages

?tidyverse  
??tidyverse

## Stackoverflow

+ Github  
+blogs

## Twitter

#rstats

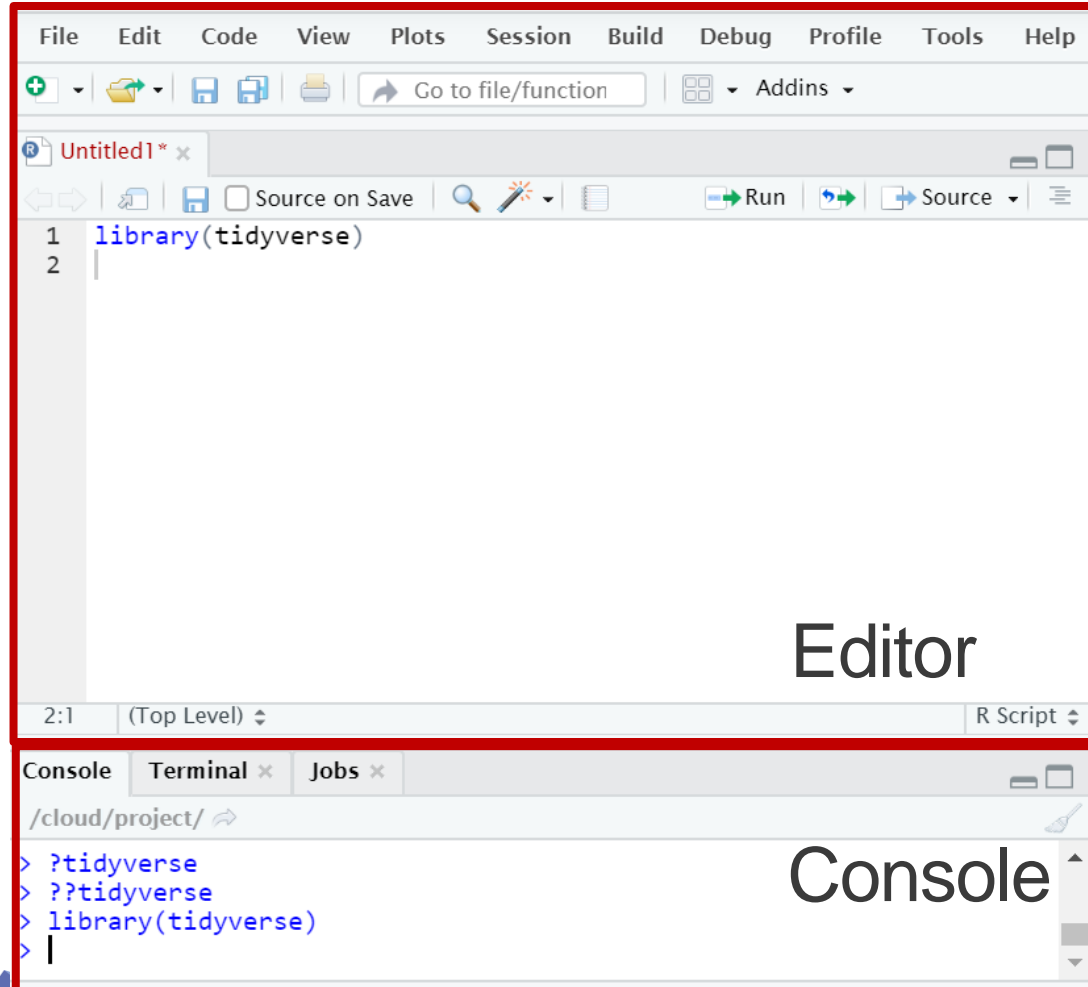




# RStudio

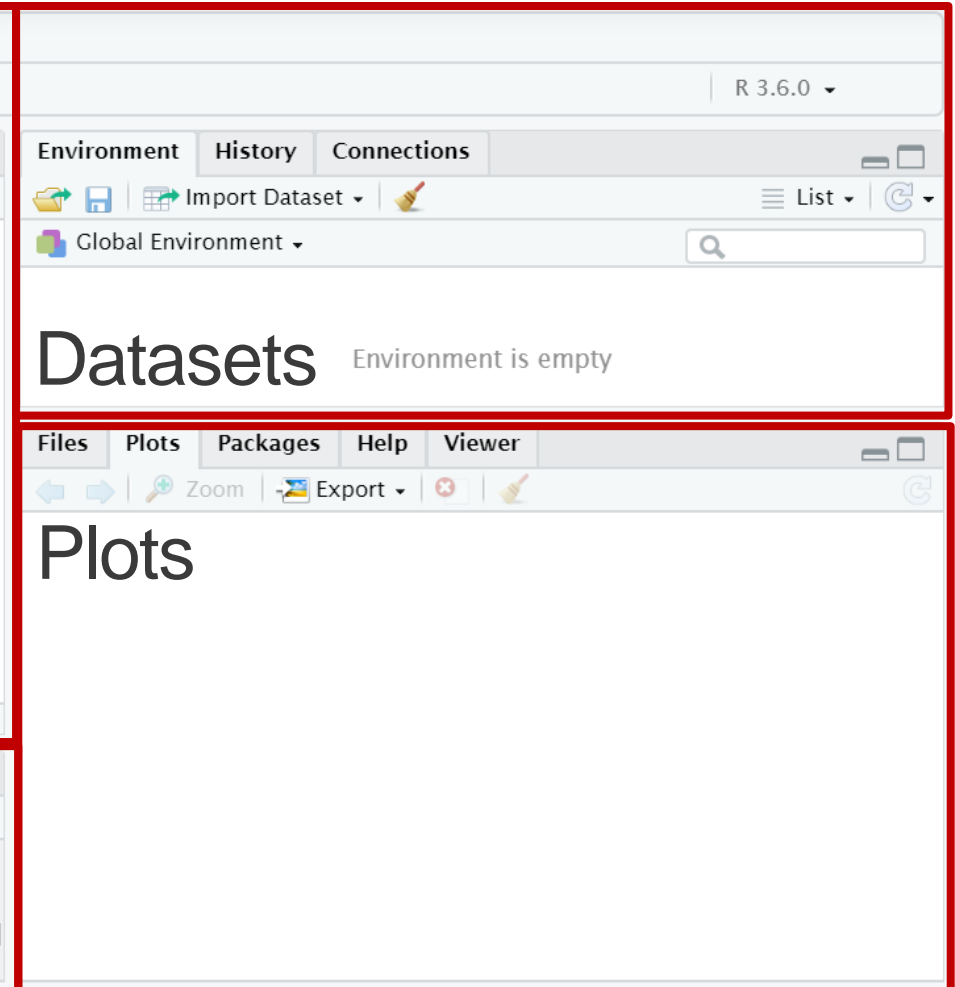


Hotels.com™ 13



The screenshot shows the RStudio Editor window. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu bar is a toolbar with icons for file operations and a search bar labeled 'Go to file/function'. The main editing area contains a file named 'Untitled1\*' with two lines of R code: `1 library(tidyverse)` and `2`. The status bar at the bottom indicates the cursor is at line 2, column 1, and the file is an R Script.

Editor



The screenshot shows the RStudio Environment and Plots panels. The top panel is the Environment pane, which shows the 'Global Environment' and a search bar. Below it is the Datasets pane, which displays 'Environment is empty'. The bottom panel is the Plots pane, which is currently empty. The status bar at the bottom indicates the cursor is at line 2, column 1, and the file is an R Script.

Environment

Datasets

Plots





# Data Transformation

dplyr



# Context: PISA tests

- Programme for International Student Assessment
- 15-year-old students
- Competence in three fields: math, science and reading
- Questionnaire about the students, their schools, parents, household, etc.
- Results and data are published in the following year



# Data description: a data frame

```
library(tidyverse)
```

```
pisa <- read_csv("http://bit.ly/wosr_pisa_data")
```

```
head(pisa)
```

```
## or
```

```
view(pisa)
```

```
## and
```

```
summary(pisa)
```



# Data transformation primitives

**Task: How many children got tested in each country?**

**SELECT** columns

**FROM** table

**WHERE** condition

**GROUP BY** columns

**ORDER BY** columns

```
pisa <- read_csv("http://bit.ly/wosr_pisa_data")
```

# Data transformation primitives

**Task: How many children got tested in each country?**

select

**SELECT** CNT, COUNT(1) as records

**FROM** pisa

%>%

summarise

filter

**WHERE** OECD = 'Yes'

**GROUP BY** CNT

group\_by

arrange

**ORDER BY** records

# Data transformation primitives

Statement: *“By age 15, students in the United Kingdom perform above the OECD average in science (509 score points) and reading (498 points) and around the OECD average in mathematics (492 points).”*

**Task: What is the median result for each field?**



# Data Visualization

ggplot2



# Approach of layers

Choose a data frame



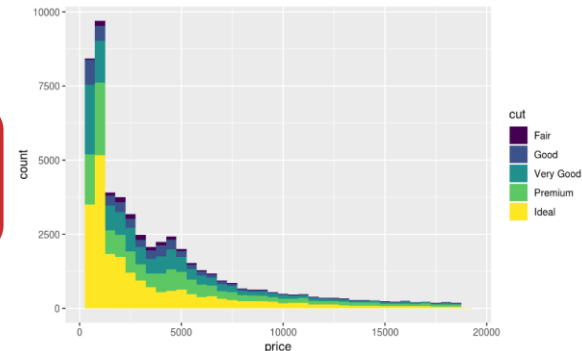
Choose a chart type



Choose the column to aesthetics mapping



Set the visual parameters



# Approach of layers

Choose a data frame

`ggplot()`

Choose a chart type

`geom_*()`

Choose the column to aesthetics mapping

`aes(x = ...,  
y = ...)`

Set the visual parameters

`theme(),  
scale_*()`

# Approach of layers

Statement: *“By age 15, students in the United Kingdom perform above the OECD average in science (509 score points) and reading (498 points) and around the OECD average in mathematics (492 points).”*

**Task: Plot the median science score for each OECD country**



# Approach of layers

Statement: *“By age 15, students in the United Kingdom perform above the OECD average in science (509 score points) and reading (498 points) and around the OECD average in mathematics (492 points).”*

**Task: Using boxplots, visualize science scores in each OECD country**

# Approach of layers

Statement: “*Students with an immigrant background (first or second generation) in the United Kingdom, as in many other OECD countries, do not perform as well in science as students without an immigrant background.*”

**Task: Using boxplots, visualize the distribution of science points in the UK for each immigrant status**

# Approach of layers

Statement: *“In the United Kingdom, boys and girls are equally likely to score at Level 5 or 6, the highest levels of proficiency, in science (12% of boys and 10% of girls) (Table I.2.6a).”*

**Task: Using boxplots, visualize the distribution of science points in the UK for each gender**

## Approach of layers

Statement: *“Even though gender differences in science performance tend to be small on average, in 33 countries and economies, the share of top performers in science is larger among boys than among girls. In the United Kingdom, as a whole, there is no significant difference in the share of top performers among boys and girls (Table I.2.6a), and this is also true in England, Northern Ireland, Scotland and Wales (Table B2.I.3).”*

**Task: Using boxplots, visualize the distribution of science points in the UK for each gender, separated for each region**

## Approach of layers

Statement: *“Even though gender differences in science performance tend to be small on average, in 33 countries and economies, the share of top performers in science is larger among boys than among girls. In the United Kingdom, as a whole, there is no significant difference in the share of top performers among boys and girls (Table I.2.6a), and this is also true in England, Northern Ireland, Scotland and Wales (Table B2.I.3).”*

**Task: Using boxplots, visualize the distribution of science points in OECD countries for each gender**

# Approach of layers

**Task: Modify the previous plot to have**

- *white background*
- *meaningful axis labels,*
- *bold fonts and*
- *different colors*



# Applications



# A few companies using R

