```
library(dplyr)

rladies_global %>%
  filter(city == 'Amsterdam')
```

Wifi: Uva Open Wifi

Please make sure you downloaded

1) **R & RStudio**
2) **Workshop materials:**
   https://github.com/rladiesamsterdam/workshop_may_2018

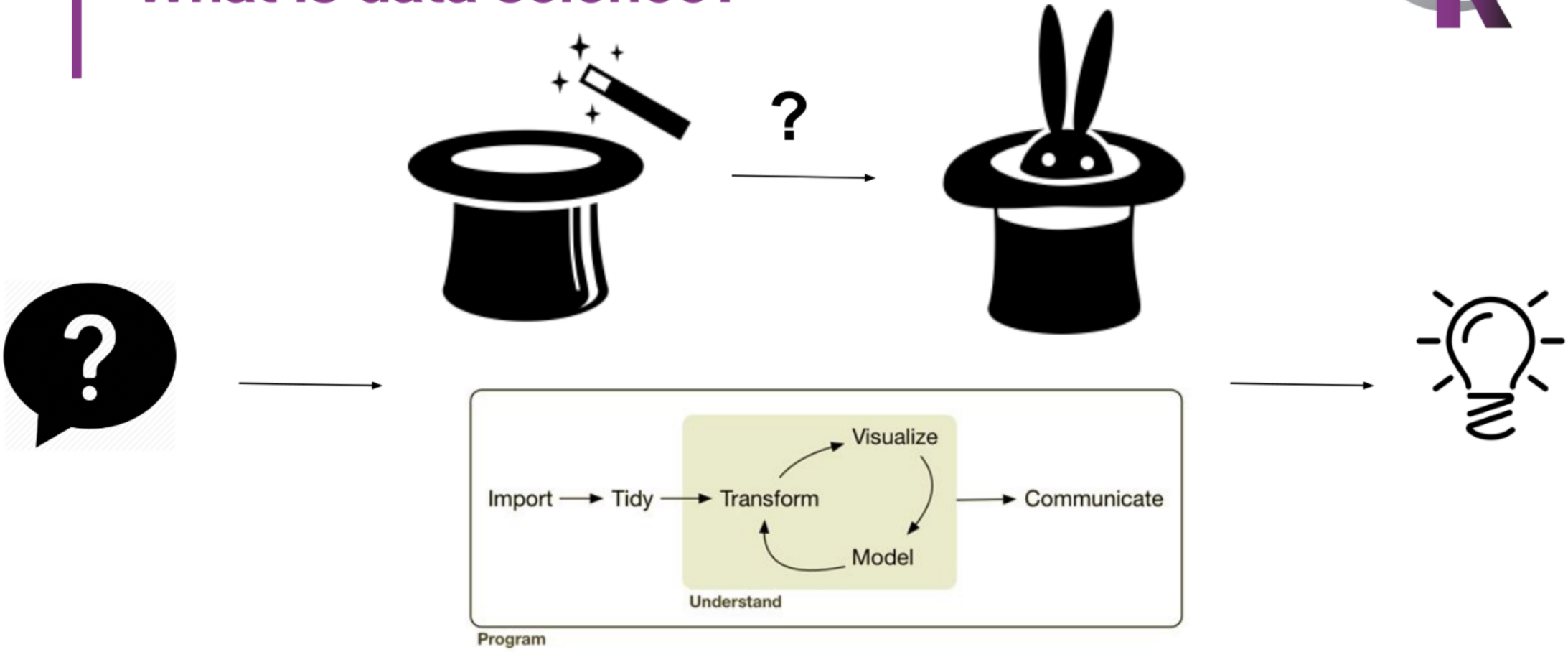
Introduction to R (for Data Science)

May 24th, 2018

Introduction to R (for Data Science)

@RLadiesAMS
#RLadies #rstats

What is data science?



Source: Hadley Wickham's Introduction to R for Data Science

What is R?

- **Statistical computing & graphics programming language and environment**
- Leading tool for machine learning, statistics & data analysis
- **Free open source** alternative to Matlab, SAS, Excel & SPSS
- Maintained and developed by a community of developers

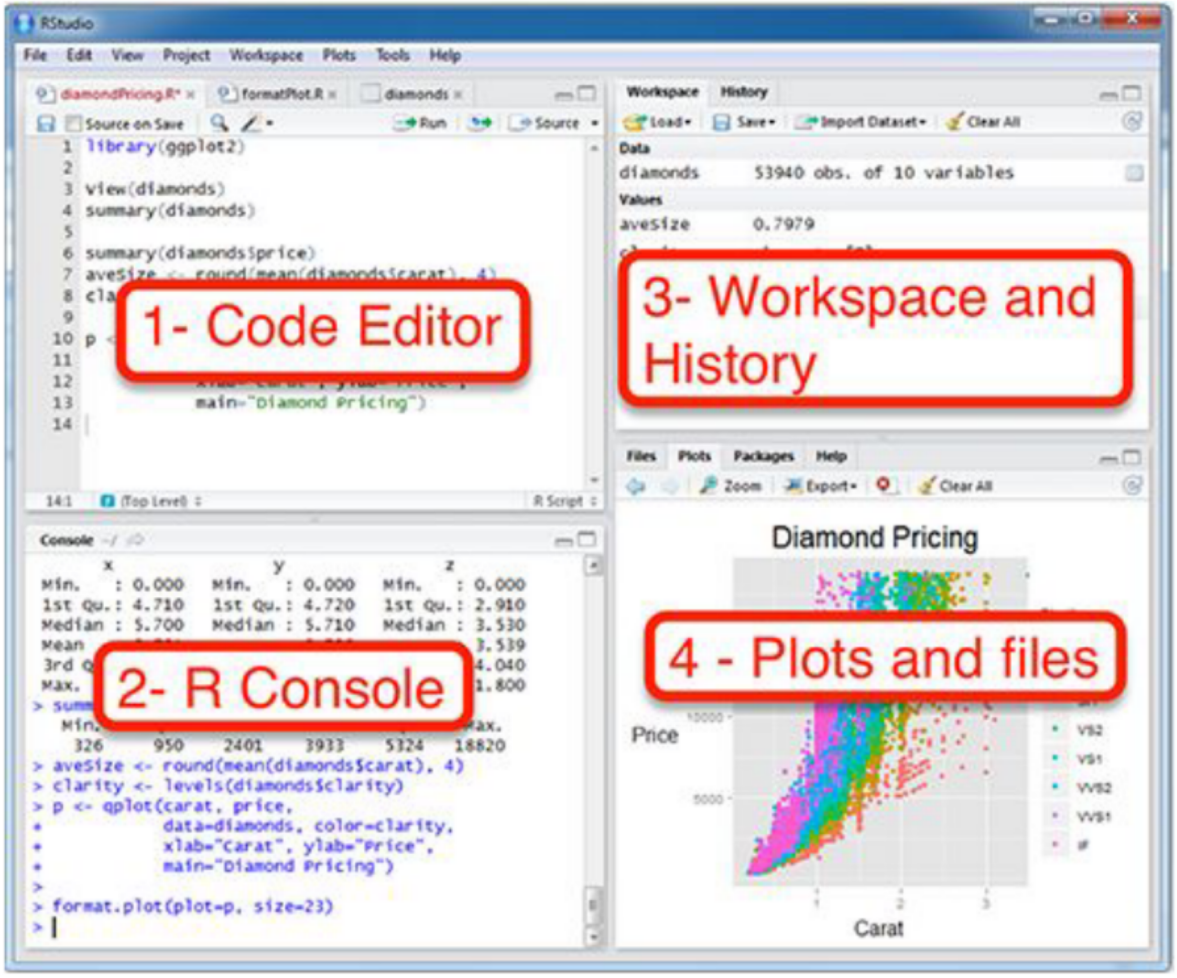
Why learn R?

- **Rapid growing community**
- **Most comprehensive statistical package**- newest ideas come up in R first
- More than 10K topic specific packages (from education, finance to time series)
- **Best tools for visualization, reporting and interactivity**
- capable of executing code written in other languages such as C++ or Java, so resources coded in those languages can be made available.
- can easily be ported between Unix, Windows or Mac environments.
- **Free & open source** (anyone can examine source code and see what's going on)
- **Collaborate with others & track computations**
- **Great community support**- support via the canonical R mailing lists, R- users groups and, more importantly, with StackOverflow, RStudio Community & Twitter

Intro to R using R studio

Use R Studio to open the file ‘tutorial1_RLadiesMay2018.R’ from rladiesamsterdam/workshop_may_2018/tutorial/

R studio


RStudio
File Edit View Project Workspace Plots Tools Help
1- Code Editor
3- Workspace and History
2- R Console
4 - Plots and files
Diamond Pricing
Price
Carat

R studio

Some useful tips to start

# shortcut for running a line of code :	ctrl + enter
# shortcut for running the entire script:	ctrl + shift + enter
# shortcut for '<-' :	alt + -(minus sign)

(for mac, use cmnd instead of ctrl)

?help function

google is your friend!

more Rstudio shortcuts: https://support.rstudio.com/hc/en-us/articles/200711853-Keyboard-Shortcuts

R as a calculator

Arithmetic operators

```
# addition:            +
# subtraction:         -
# multiplication:      *
# division:            /
# exponentiation:      ^ or **
# modulus:             %%
```

Logical operators

```
# less than:                  <
# less than or equal to:      <=
# greater than:               >
# greater than or equal to:   >=
# exactly equal to:           ==
# not equal to:               !=
# OR:                         |
# AND:                        &
```


Source: https://www.statmethods.net/management/operators.html

Variables

assign values to variables using either <- or =

variable names are case sensitive! *My_Variable != my_variable*

valid variable names include only letters, numbers, _ and .
valid variable names start with letters or .

be careful not to use names of existing functions in R to avoid confusion

be careful overwriting information you want to keep

Types of data

# numeric & integer:	quantitative information E.g. *5.78 ; 6*
# character:	strings with any type of data E.g. *“Hello, world!”* ; *‘five’* ; *“5”*
# logical:	either True or False E.g. *T; TRUE; FALSE*

you can use the function *class()* to find out what type of data you are working with.

Vectors

A vector is a sequence of data elements of the same basic type
any_vector = c()

indexing
any_vector[index]

note: many programming languages start counting at 0, but R starts counting at 1

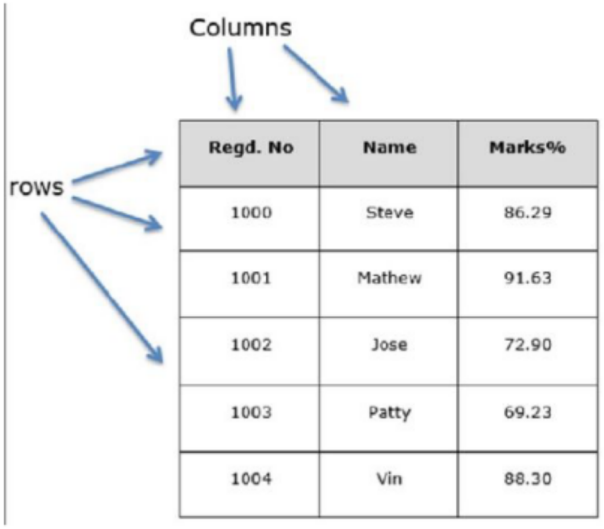
Data Frames

create dataframes: *my_df = data.frame()*

Access column/ row names: *rownames() ; colnames()*
Access number of columns/ rows: *nrow(); ncol()*

Index values in data frames: *my_df[row,column]*
rows and columns can be referred to either by name or number

View(my_df)



Regd. No	Name	Marks%
1000	Steve	86.29
1001	Mathew	91.63
1002	Jose	72.90
1003	Patty	69.23
1004	Vin	88.30

Plotting

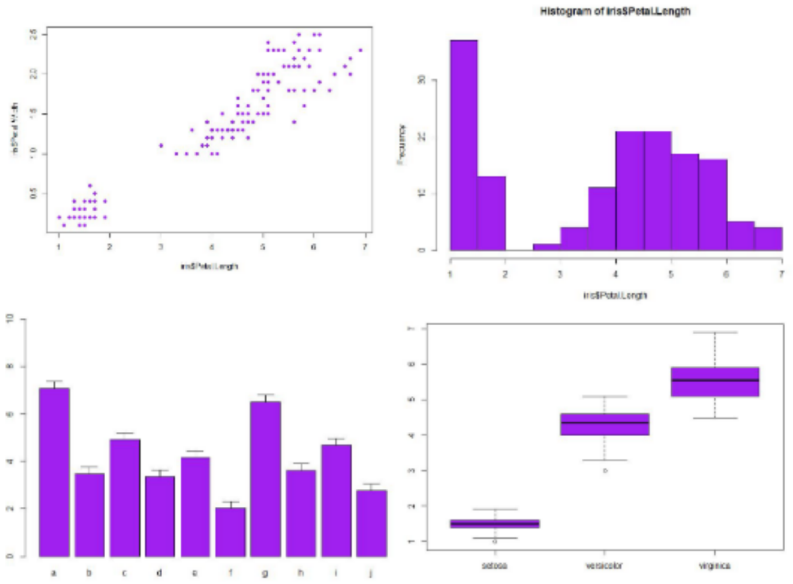
For quick plotting, you can use R's base functions, e.g.

```
# plot(numeric_x, numeric_y)

# barplot(numeric_x)

# hist(factor_x)

# boxplot(numeric_x ~ factor_y)
```



For more advanced plotting, you could use the ggplot2 library

Saving things

# to save a data frame as csv:	write.csv() write.table()
# save your script as a .R file	
# to save a workspace:	see lightning talk ;)

Titanic assignment

Open the file ‘tutorial2.r’ from rladiesamsterdam/workshop_may_2018/assignment

Form groups of 2-3 to work on through the script together.

Additional resources

 # Packages:
You can install a lot of additional packages/libraries with a lot of functions that are written and shared by people all over the world

 # RStudio cheatsheet:
https://www.rstudio.com/wp-content/uploads/2016/01/rstudio-IDE-cheatsheet.pdf

Google everything!

Post on **Stackoverflow** (always include a reproducible example of your problem)

 # RLadies Slack channel

MOOCs

Additional Resources 2

Interactive

Datacamp: https://www.datacamp.com
Coursera:https://www.coursera.org/specializations/jhu-data-science
Udacity:https://eu.udacity.com/course/data-analysis-with-r--ud651
Lynda:https://www.lynda.com/R-training-tutorials/1570-0.html

Books

Hadley Wickhams- Data Science in R:http://r4ds.had.co.nz/
Norman Matloff- The art of programming in R:
http://diytranscriptomics.com/Reading/files/The%20Art%20of%20R%20Programming.pdf

Video Lectures:

RStudio video lectures & webinars:https://www.rstudio.com/resources/webinars/
LearnR: https://www.youtube.com/user/TheLearnR/featured
GoogleDevelopers:
https://www.youtube.com/playlist?list=PLOU2XLYxmsIK9qQfztXeybpHvru-TrqAP
MarinStatsLectures:
https://www.youtube.com/user/marinstatlectures

Blogs for latest news/ tutorials:

https://www.r-bloggers.com/
http://flowingdata.com/
https://www.rweekly.org/
https://www.r4stats.com/
https://www.r-exercises.com/
https://www.blog.rstudio.com/