

Social Media and Political Participation

Lab 1¹

drewd@nyu.edu

January 5, 2015

¹slides modified from Pablo Barberá

Information to access the lab computers:

`username` your netID (e.g. ddd281)

`password` P@ssw0rd (note the zero instead of an “o”)

After logging in, press Ctrl+Alt+Esc to change your password.

Today

- Introductions and lab logistics
- First steps in R:
 - What is R?
 - Downloading and installing Rstudio
 - Using R as a calculator
 - Working with R scripts
- Reading and analyzing data using R
- In-class exercise: your first R script

Introductions and class logistics

Hi!

TA: **Drew Dimmery**

- PhD candidate, Dept. of Politics, New York University
- Graduate student associate of the SMaPP Lab
- I like experiments! I use them to help understand why people take costly political action.
- Contact:

email drewd@nyu.edu

chat [ddd281@nyu.edu](https://chat.ddd281@nyu.edu)

twitter [@DrewDim](https://twitter.com/DrewDim)

web drewdimmery.com

office Room 317, 19w4th Building.

About the lab sessions

- Six lab sessions, 2pm-4pm
 - 1 Jan 5. First steps in R
 - 2 Jan 6. Introduction to statistical analysis using R
 - 3 Jan 8. Introduction to Twitter and the Twitter API
 - 4 Jan 12. Introduction to Facebook and Facebook API
 - 5 Jan 15. Analyzing Facebook and Twitter data I
 - 6 Jan 19. Analyzing Facebook and Twitter data II
- Office hours: 4pm-6pm, Monday to Friday, on Google Hangout or my office (room 317)

Class logistics

- Each lab session will have three parts:
 - 1 Introduction to a topic
 - 2 Interactive R session
 - 3 In-class exercise and quiz
- You're welcome to bring your own laptop. I can help you installing RStudio. But recommended to use lab computers.
- All materials (slides, code, quiz) are available on NYU Classes, in "Resources" > "Labs" folder
- Ask questions!

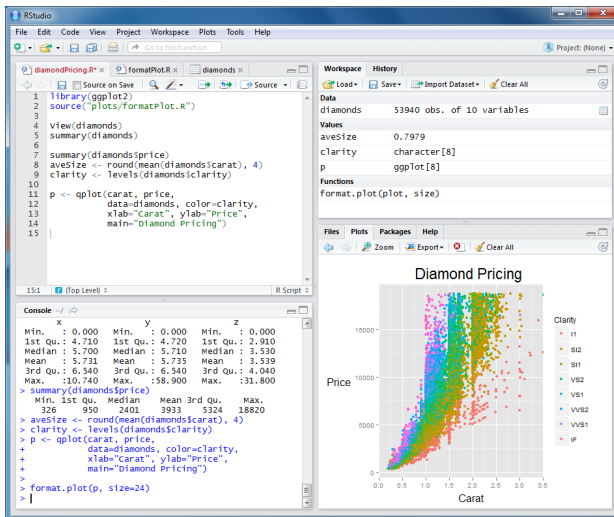
First steps in R

What is R?

- R is a free programming language for statistical analysis
- Used by most statisticians and social scientists interested in data analysis, it's becoming the standard in Data Science
- Open-source: highly customizable and easily extensible through “packages”.
- Powerful tool to generate elegant and effective plots.
- Command-line interface: steeper learning curve, but allows easy replication of analysis through “scripts”.
- Excellent documentation and online help resources.
- RStudio: software that provides a more friendly interface to R.



RStudio



RStudio

Installation

- Should be installed in all lab computers
- To install it on your laptops, read the document “Installing_RStudio.pdf” in the NYU Classes Resources folder.

Four panels

- 1 Console: where you type R commands interactively and see the output and error messages
- 2 Workspace: shows all objects (data) in memory
- 3 Viewer: shows plots produced by R, helps you find files, displays help menus, etc.
- 4 Script: where you write R code

Working with R scripts

- An R script is a text document that contains a list of commands that you wish to execute
- Why should you use scripts instead of typing commands on the console?
 - ① Replicability (audit trail)
 - ② Creates a library of code examples
 - ③ Easier to correct errors: fix it and then run script again
- R scripts basics:
 - ① Each line contains a different command
 - ② Add “comments” using # sign at beginning of line
 - ③ In RStudio, select block of code you want to execute and then click on “Run” or press `Ctrl+Enter`

Installing and loading packages

- A “package” is a collection of functions that expands the basic functionality of R.
- For example, Rfacebook is a package that allows R to capture Facebook data.
- You need to install them (once) and then load them every time you want to use them:

```
# how to install the Rfacebook package
install.packages("Rfacebook")
# how to load a package
library(Rfacebook)
# an example function
getPage("barackobama", token=my_token, n=100)
```

Getting help

Where to look for help:

- 1 R manuals: clicking on the “Help” tab on the Viewer panel (bottom-right)
- 2 Documentation for each function in R

```
# these two commands do the same
?mean
help(mean)
```
- 3 Online:
 - [Stack Overflow](#)
 - [CRAN](#): repository of R packages and documentation
 - Google your problem mentioning “rstats”

First steps in R

- Go to NYU Classes, Resources > Labs > Lab 1, and download the following R scripts to your desktop:
 - lab1_first_steps.R
 - lab1_data_analysis.R
- Now click on the Start button and type “RStudio”. Click on the first result.
- Using the bottom right panel, navigate to the folder where you saved the R scripts and click on them
- We will start interacting with R using the first script, lab1_first_steps.R

R resources

- Textbook for first part of lab sessions:
 - “A Beginner’s Guide to R”, by Alain F. Zuur *et al*, Springer. Available online through NYU library. We will cover the equivalent of chapters 1–3, 5–6 in class.
- How to prepare elegant plots with R.
 - “The R Graphics Cookbook”, by Winston Chang. O’Reilly.
- Additional resources:
 - Google R tutorials on Youtube.
 - Drew Dimmery’s “Thinking in R”
 - Book: “The Art of R Programming”, by Matloff. Available online through NYU library.

Reading and analyzing data

Reading and analyzing data

The R script `lab1_data_analysis.R` shows how to:

- Work with vectors: access specific elements, subset by certain conditions, compute length...
- Use statistical functions: compute the mean, minimum, maximum...
- Do basic calculations with vectors: products, divisions...
- Create and subset data frames
- Import data from a spreadsheet

In-class exercise

In-class exercise: your first R script

Create your own R script (with comments) that:

- ① Opens the dataset `lab1_nyu_data.csv`
- ② Runs different commands that help you answer the following questions:
 - ① How many status updates has NYU Abu Dhabi posted on its page?
 - ② What is the **average** number of likes AND comments that its posts receive?
 - ③ What is the **maximum** number of likes AND comments that its posts receive?
 - ④ What was the content of the **last** status update of 2014?