Social Media and Political Participation

Lab 1

pablo.barbera@nyu.edu

January 6, 2014

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

Hola!

TA: **Pablo Barberá**

- PhD candidate, Dept. of Politics, New York University
- Graduate student associate of the SMaPP Lab
- I study the effects of social media on political polarization
- Contact:

email pablo.barbera@nyu.edu
chat pba220@nyu.edu
twitter @p_barbera
web www.pablobarbera.com
office Room 230, 19w4th Building.

About the lab sessions

- Six lab sessions, 2pm-4pm
  1. Jan 6. First steps in R
  2. Jan 7. Introduction to statistical analysis using R
  3. Jan 8. Introduction to Twitter and the Twitter API
  4. Jan 10. Introduction to Facebook and Facebook API
  5. Jan 16. Analyzing Facebook and Twitter data I
  6. Jan 21. Analyzing Facebook and Twitter data II
- Office hours: 4pm-6pm, Monday to Friday, on Google Hangout or my office (room 230)

Class logistics

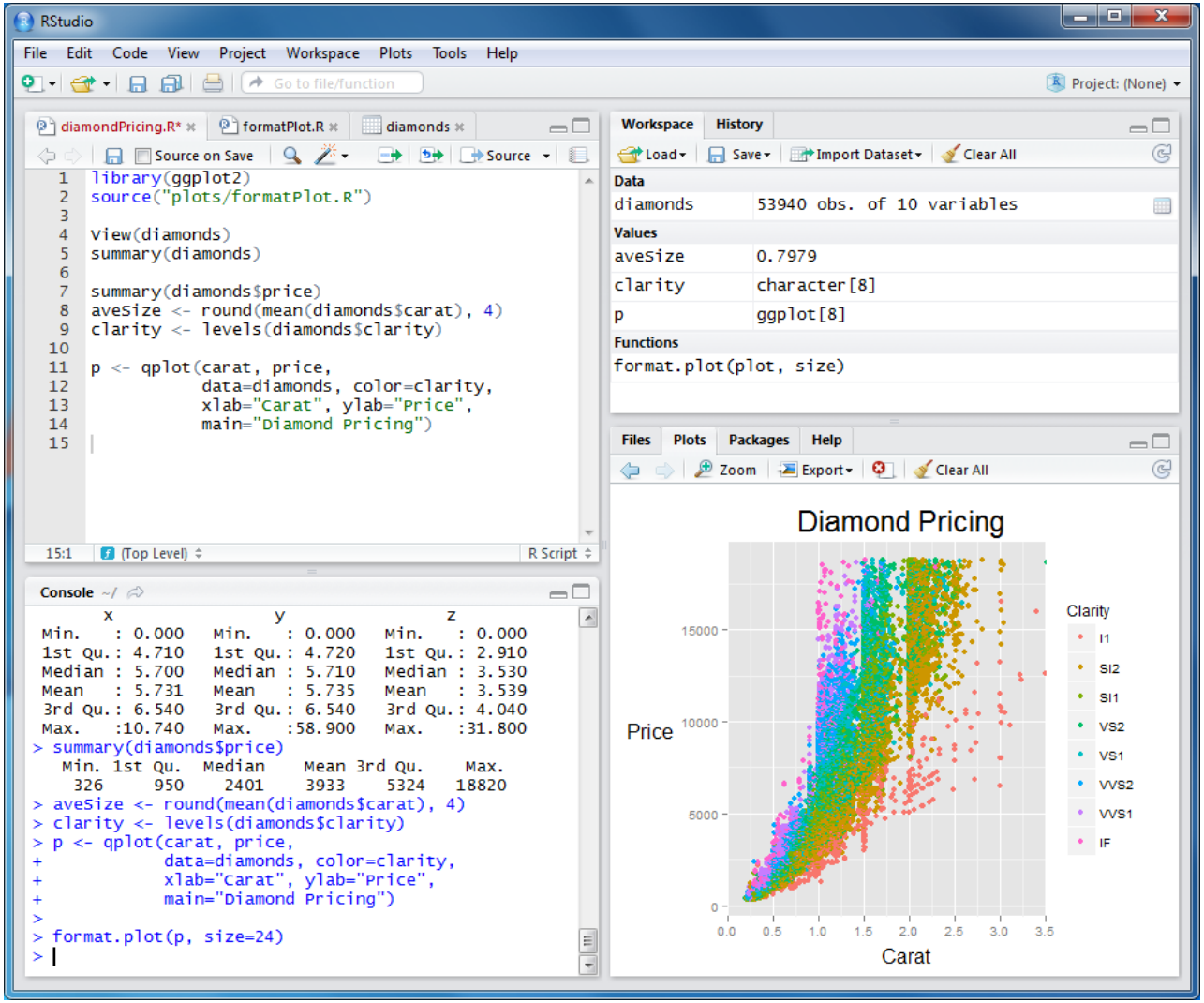
- Each lab session will have three parts:
  1. Introduction to a topic
  2. Interactive R session
  3. In-class exercise and quiz
- You're welcome to bring your own laptop.
- All materials (slides, code, quiz) are available on NYU Classes, in "Resources" > "Labs" folder
- Ask questions!

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
File Edit Code View Project Workspace Plots Tools Help
diamondPricing.R* formatPlot.R diamonds
Source on Save Source
1 library(ggplot2)
2 source("plots/formatPlot.R")
4 View(diamonds)
5 summary(diamonds)
7 summary(diamonds$price)
8 aveSize <- round(mean(diamonds$carat), 4)
9 clarity <- levels(diamonds$clarity)
11 p <- qplot(carat, price,
12 data=diamonds, color=clarity,
13 xlab="Carat", ylab="Price",
14 main="Diamond Pricing")
15:1 (Top Level) R Script
Console ~/
x y z
Min. : 0.000 Min. : 0.000 Min. : 0.000
1st Qu.: 4.710 1st Qu.: 4.720 1st Qu.: 2.910
Median : 5.700 Median : 5.710 Median : 3.530
Mean : 5.731 Mean : 5.735 Mean : 3.539
3rd Qu.: 6.540 3rd Qu.: 6.540 3rd Qu.: 4.040
Max. :10.740 Max. :58.900 Max. :31.800
> summary(diamonds$price)
Min. 1st Qu. Median Mean 3rd Qu. Max.
326 950 2401 3933 5324 18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+ data=diamonds, color=clarity,
+ xlab="Carat", ylab="Price",
+ main="Diamond Pricing")
> format.plot(p, size=24)
Workspace History
Load Save Import Dataset Clear All
Data
diamonds 53940 obs. of 10 variables
Values
aveSize 0.7979
clarity character[8]
p ggplot[8]
Functions
format.plot(plot, size)
Files Plots Packages Help
Zoom Export Clear All
Diamond Pricing
Price
Carat
Clarity
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF

RStudio

Installation

- Should be installed in all lab computers
- To install it on your laptops, go to rproject.com and look for "download R". Then, go to rstudio.com to download and install Rstudio.

Four panels (describe what they are even if they're on next slide)

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
load(Rfacebook)
# an example function
searchFacebook(string="obama", token=my_token)
```

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*, Springr. Available online through NYU library. We will cover 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_obama_data.csv`
2. Runs different commands that help you answer the following questions:
   1. How many status updates did Barack Obama posted on his page in 2013?
   2. How many comments did each post receive, on average?
   3. What is the maximum number of comments that a post received?
   4. What was the content of the **last** status update of 2013?