PUBPOL 750 DATA ANALYSIS FOR PUBLIC POLICY I: DESCRIPTIVE STATISTICS

Master of Public Policy in Digital Society

Instructor: Justin Savoie

Week 1 - Introduction

Course general objectives

- R programming and statistics
- “Foundations of foundations”
- Preparing you for a variety of roles (data scientists, project manager, project analyst, etc.)
- Examining data, knowing exactly what is involved in a data analysis

Specific objectives

- Become familiar with R
- Know how to load data, clean data, wrangle data, reformat data, plot data
- Prepare you for Data Analysis II and III
- Prepare you for doing quantitative research

Format

- ~ 45 minutes presentation
- 5 minutes break
- Exercises

Assessment

- Homework
- 2 projects
- Attendance and participation

- PASS / FAIL / PASS with distinction

- Late policy: flexibility - treat it professionally

Office hours + contacts

- Wednesday 3-5pm

- I'm flexible, you can email me. I'll always answer in 48h (probably less). Feel free to send me a reminder if I missed your email.

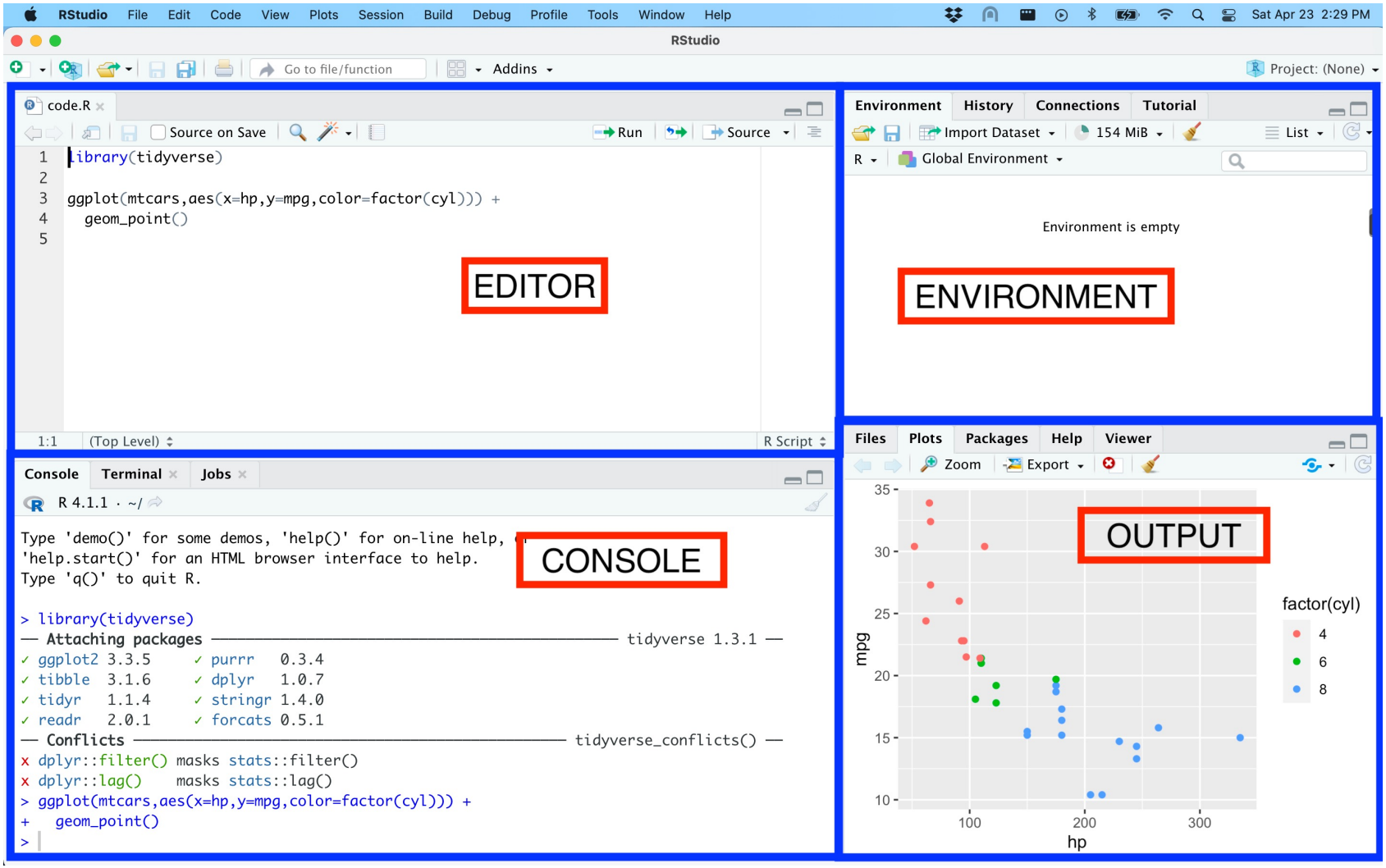
Material

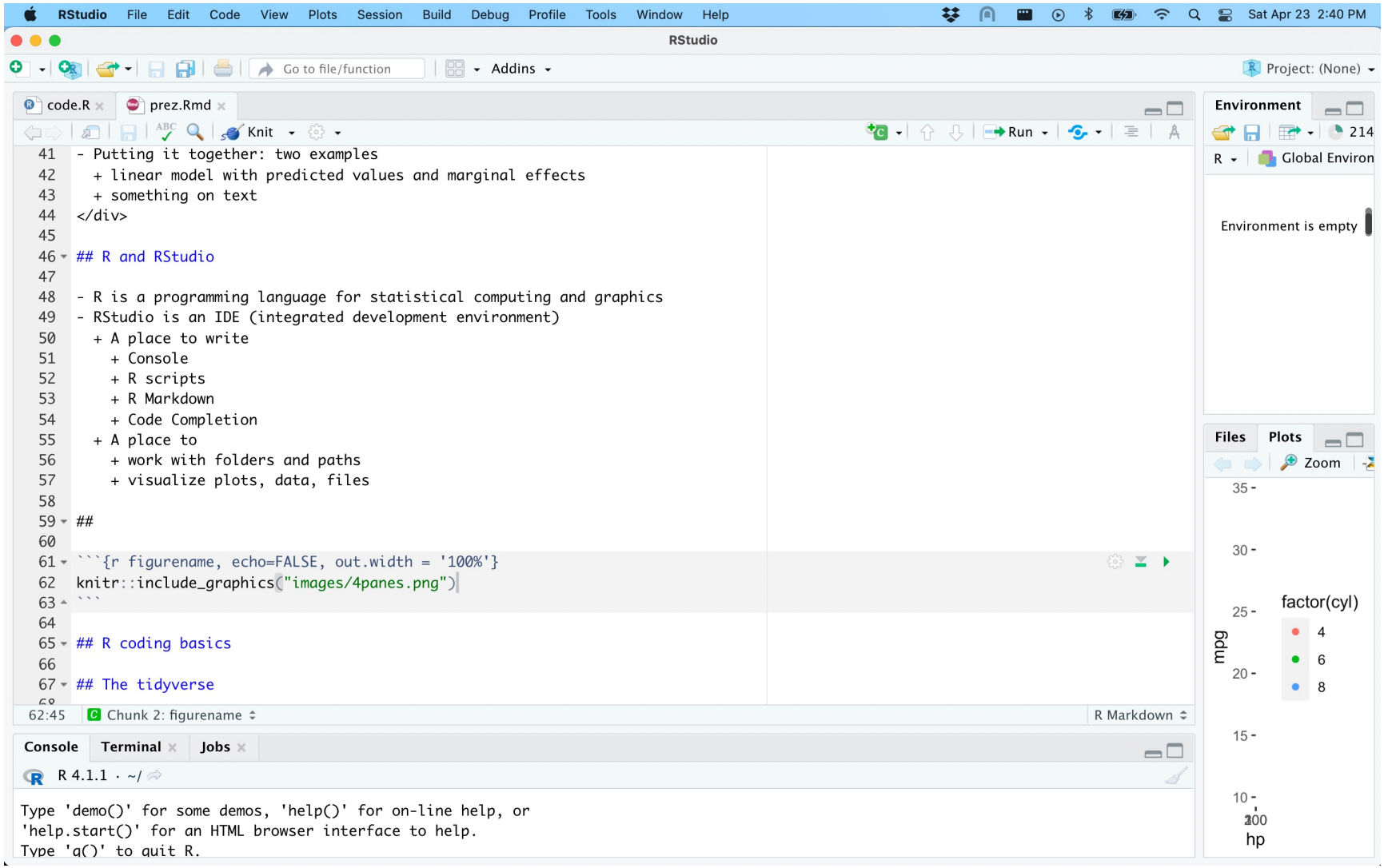
- R for Data Science (R4DS) by Hadley Wickham and Garrett Gromelund Available: https://r4ds.had.co.nz/

- Modern Dive: Statistical Inference via Data Science (Modern) by Chester Ismay and Albert Y. Kim Available: https://moderndive.com/

R and Rstudio

- R is a programming language for statistical computing and graphics
- RStudio is an IDE (integrated development environment)


RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Sat Apr 23 2:29 PM
RStudio
Go to file/function
Addins
Project: (None)
code.R
Source on Save
Run
Source
1 library(tidyverse)
2
3 ggplot(mtcars,aes(x=hp,y=mpg,color=factor(cyl))) +
4 geom_point()
5
EDITOR
1:1
(Top Level)
R Script
Environment
History
Connections
Tutorial
Import Dataset
154 MiB
List
R
Global Environment
Environment is empty
ENVIRONMENT
Console
Terminal
Jobs
R 4.1.1 · ~/
Type 'demo()' for some demos, 'help()' for on-line help,
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
CONSOLE
> library(tidyverse)
── Attaching packages ──── tidyverse 1.3.1 ──
✓ ggplot2 3.3.5 ✓ purrr 0.3.4
✓ tibble 3.1.6 ✓ dplyr 1.0.7
✓ tidyr 1.1.4 ✓ stringr 1.4.0
✓ readr 2.0.1 ✓ forcats 0.5.1
── Conflicts ──── tidyverse_conflicts() ──
x dplyr::filter() masks stats::filter()
x dplyr::lag() masks stats::lag()
> ggplot(mtcars,aes(x=hp,y=mpg,color=factor(cyl))) +
+ geom_point()
>
Files
Plots
Packages
Help
Viewer
Zoom
Export
OUTPUT
mpg
hp
factor(cyl)
4
6
8


RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Sat Apr 23 2:40 PM
RStudio
Go to file/function
Addins
Project: (None)
code.R
prez.Rmd
Knit
Run
41 - Putting it together: two examples
42 + linear model with predicted values and marginal effects
43 + something on text
44
45
46 ## R and RStudio
47
48 - R is a programming language for statistical computing and graphics
49 - RStudio is an IDE (integrated development environment)
50 + A place to write
51 + Console
52 + R scripts
53 + R Markdown
54 + Code Completion
55 + A place to
56 + work with folders and paths
57 + visualize plots, data, files
58
59 ##
60
61 ```{r figurename, echo=FALSE, out.width = '100%'}
62 knitr::include_graphics("images/4panes.png")
63 ```
64
65 ## R coding basics
66
67 ## The tidyverse
62:45
Chunk 2: figurename
R Markdown
Console
Terminal
Jobs
R 4.1.1 · ~/
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Environment
R
Global Environ
Environment is empty
Files
Plots
Zoom
35
30
25
20
15
10
mpg
100
200
hp
factor(cyl)
4
6
8

About me

- Year 5 Political Science PhD at UofT
- Working on text analysis
- 4 years as data scientist and data science lead at Vox Pop Labs working on projects like Vote Compass
- In academia, published on election forecasting, survey analysis
- Work on survey data

About you