Advanced Quantitative Methods

Introduction

Instructor: Gregory Eady
Office: 18.2.10
Office hours: Fridays 13-15

Today

1. About me
2. About you
3. About the course and its goals
4. (Re-)Introduction to R

1. **Research**:
   - Methodological research
     - Ideological scaling
     - LLMs in survey research
     - Sensitive beliefs
   - Applied research
     - Effects of Russian disinformation on voting in the 2016 US election
     - Effects of January 6 insurrection on partisanship
     - Effects of approaches to political discourse on discourse quality
     - Electoral competitiveness and sharing of ideologically extreme content on social media
2. **Teaching:**
   - Advanced Quantitative Methods
   - Elective methods courses
   - SDS methods course
   - MA thesis supervision

Course website

gregoryeady.com/ResearchMethodsCourse/

If you need help

- E-mail
  - Try to use only for quick questions
  - Copy and paste code, don't screenshot
- In person
  - Schedule a meeting from the course website
    - Fridays, 13.00-15.00
  - If no time slots left, email me
  - You can drop by my office if my door is open

About you

- Go to the course website
- Weekly readings → 1. Introduction
- "Padlet for AQM"
- Make a post describing your background...

Make a post with the following

1. Title: Your full name
2. Research methods background
3. Statistical software background?
4. What would you like to use quantitative research methods for?

Aim of the course

1. Identify relevant designs & techniques to solve issues in the study of politics
2. Process data in structured and unstructured formats
3. Analyze those data
4. Reflect on the advantages and disadvantages of different designs and techniques

Assignment 1 (due Thursday, October 2)

- Download and load a dataset
- Clean data
- Create figures
- Run and interpret OLS regression models
- Already available on the course website

Assignment 2 (due Thursday, December 11)

Consists of three sub-assignments to replicate the main results from top-level political science articles:

- Explain the research question
- Explain the research design
- Replicate the main regression table(s) & figure(s)

Sub-assignments:

1. Instrumental variables (soft due date: Week 45)
2. Regression discontinuity design (soft due date: Week 47)
3. Difference-in-differences (soft due date: Week 50)

Class structure

1. First hour:
    - Lecture on a particular method and its applications
2. 10-15 minute break
3. Second hour:
    - Lab to implement the method in R

Lecture schedule (ELECTIVE)

Coding & Preliminaries	
Lecture 1	**Introduction**
Lecture 2	**R Introduction**
Lecture 3	**Visualization in R**
Lecture 4	**OLS refresher & OLS in R**
Lecture 4.5	**Assignment 1 workshop**

Experiments	
Lecture 5	**Credibility Revolution & LaTeX**
Lecture 6	**Experiments**
Lecture 7	**CACE & Randomization inference**

Lecture schedule (SPECIALIZATION)

Coding & Preliminaries	
Lecture 1	**Introduction**
Lecture 2	**Visualization in R**
Lecture 3	**OLS refresher & OLS in R**
Lecture 4	**Assignment 1 workshop**

Experiments	
Lecture 5	**Credibility Revolution & LaTeX**
Lecture 6	**Experiments**
Lecture 7	**CACE & Randomization inference**

About the course

Quasi-experiments	
Lecture 8	**Instrumental variables**
Lecture 9	**Panel data**
Lecture 10	**Regression discontinuity**
Lecture 11	**Difference-in-differences**
Lecture 12	**Event studies**
Lecture 13	**Staggered diff-in-diff**

Books & readings

- **Mastering 'Metrics: The Path from Cause to Effect**
  Princeton University Press, 2014.
  Joshua D. Angrist and Jörn-Steffen Pischke
- **Field Experiments: Design, Analysis, and Interpretation**
  W. W. Norton & Company, 2012.
  Alan S. Gerber and Donald P. Green.
- Articles posted to website

- **Causal Inference: The Mixtape** (optional)
  Yale University Press, 2021.
  Scott Cunningham.

Why does this course exist?

1. Data revolution
2. Computational revolution
3. Credibility revolution

"Advanced" Quantitative Methods

- The adjective "Advanced" in the course name is not about advanced statistical methods
- It is about cleverly thinking about research design
- Successful political science research is rarely about statistical sophistication
- Indeed, we almost solely rely on OLS

Statistical software

- SPSS (social sciences)
- Matlab (statistics, computer science, natural sciences)
- C++
- Stata (still dominate in economics)
- Python
- R

Why R?

1. Extremely flexible
2. Workflow is replicable
3. R versus Python
4. Free
5. Open-source and free libraries
6. Massive R community online
7. Cutting-edge methods are often implemented in R first
8. Data visualization

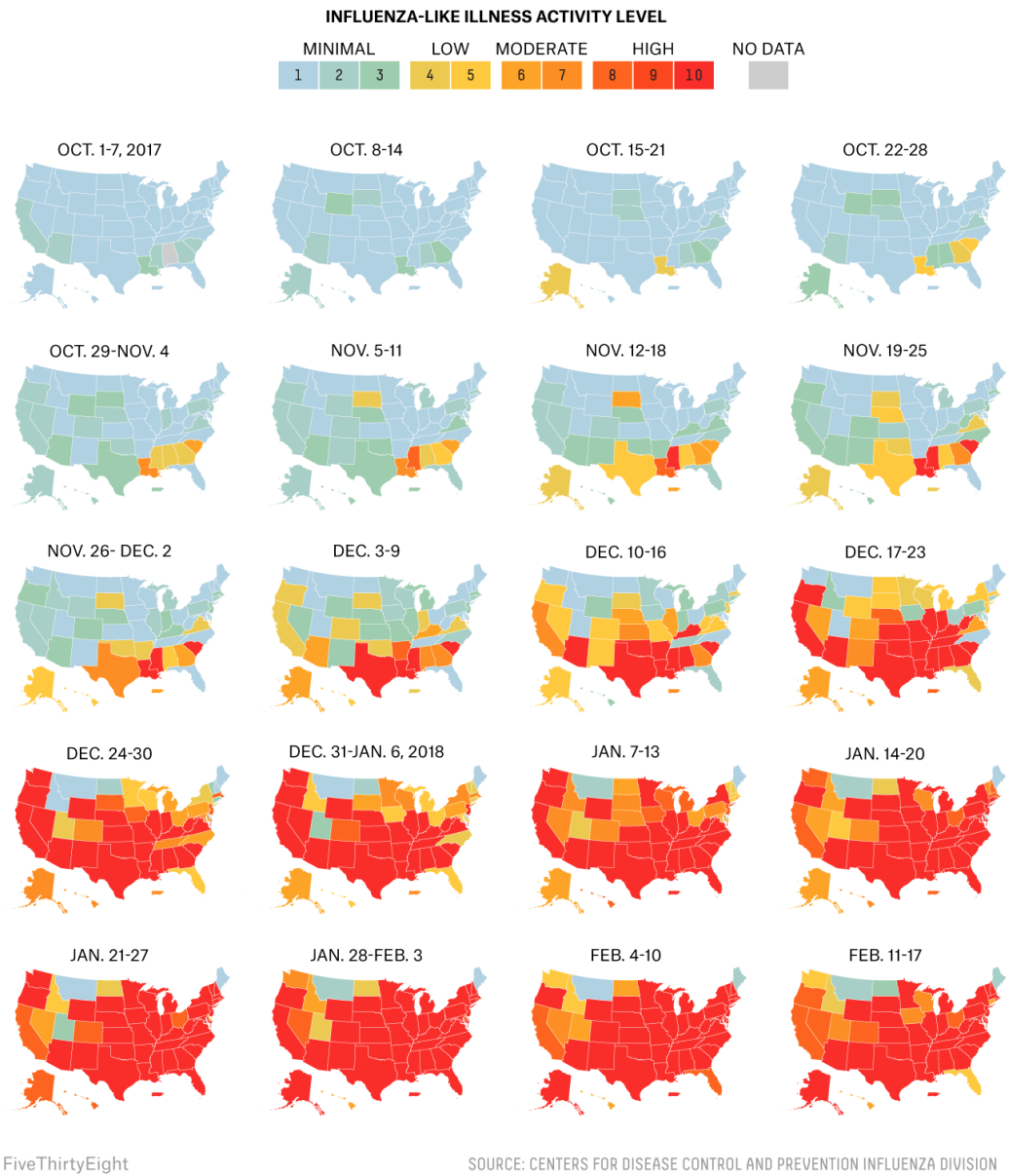
Hadley Wickham created 2 important R libraries:

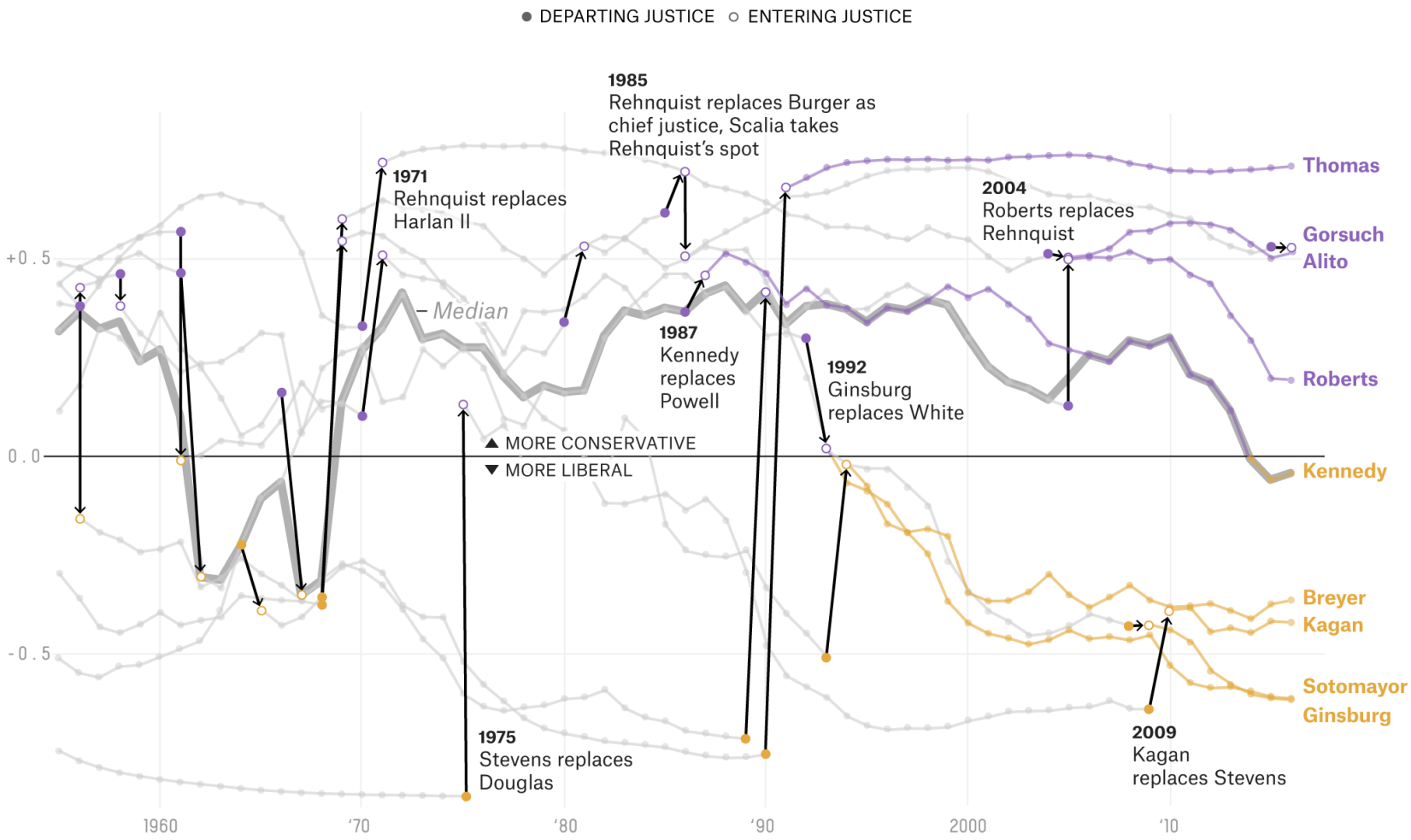
- tidyverse

```
Gender <- D %>%
      filter(respondent_type == "Citizen") %>%
      filter(!duplicated(id) & !is.na(resp_gender) & resp_gender != "Other") %>%
      group_by(country, resp_gender) %>%
      summarize(count_collected = n()) %>%
      ungroup() %>%
      group_by(country) %>%
      mutate(percentage = count_collected / sum(count_collected) * 100) %>%
      mutate(group = recode(resp_gender, "Male" = "Gender: Men",
                                         "Female" = "Gender: Women"))
```

- ggplot2

```
Figure_1B <- ggplot(G1B, aes(y = percent, x = respondent_type,
                             fill = category)) +
  my.theme(base_size = 7, borderless = 2) +
  labs(x = "% who want more\ngovernment action", y = "") +
  coord_cartesian(y = c(0, 85), expand = FALSE) +
  facet_wrap(~ group) +
  force_panelsizes(cols = c(0.5, 1)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.9), width = 0.7) +
  geom_text(aes(label = paste0(round(percent), "%")),
            position = position_dodge(width = 0.9),
            size = 2.5, vjust = -1) +
  scale_fill_manual(values = c("Women" = "#00BD8E", "Men" = "#5200FF", "Overall" = "grey20")) +
  theme(legend.position = "none",
        strip.text.x = element_text(angle = 0, hjust = 0, size = 7, face = "bold"))
```

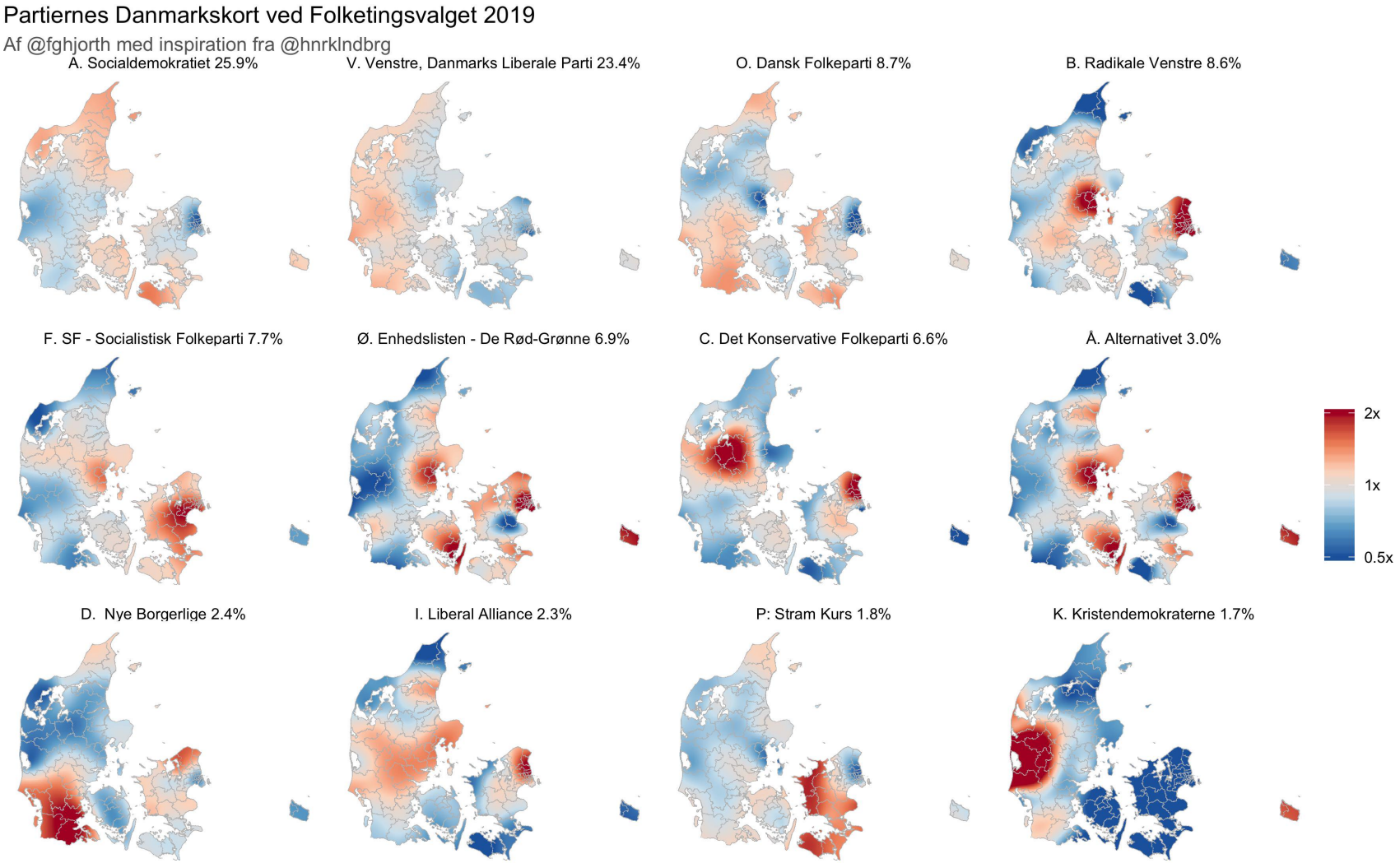

INFLUENZA-LIKE ILLNESS ACTIVITY LEVEL
MINIMAL
LOW
MODERATE
HIGH
NO DATA
1
2
3
4
5
6
7
8
9
10
OCT. 1-7, 2017
OCT. 8-14
OCT. 15-21
OCT. 22-28
OCT. 29-NOV. 4
NOV. 5-11
NOV. 12-18
NOV. 19-25
NOV. 26- DEC. 2
DEC. 3-9
DEC. 10-16
DEC. 17-23
DEC. 24-30
DEC. 31-JAN. 6, 2018
JAN. 7-13
JAN. 14-20
JAN. 21-27
JAN. 28-FEB. 3
FEB. 4-10
FEB. 11-17
FiveThirtyEight
SOURCE: CENTERS FOR DISEASE CONTROL AND PREVENTION INFLUENZA DIVISION

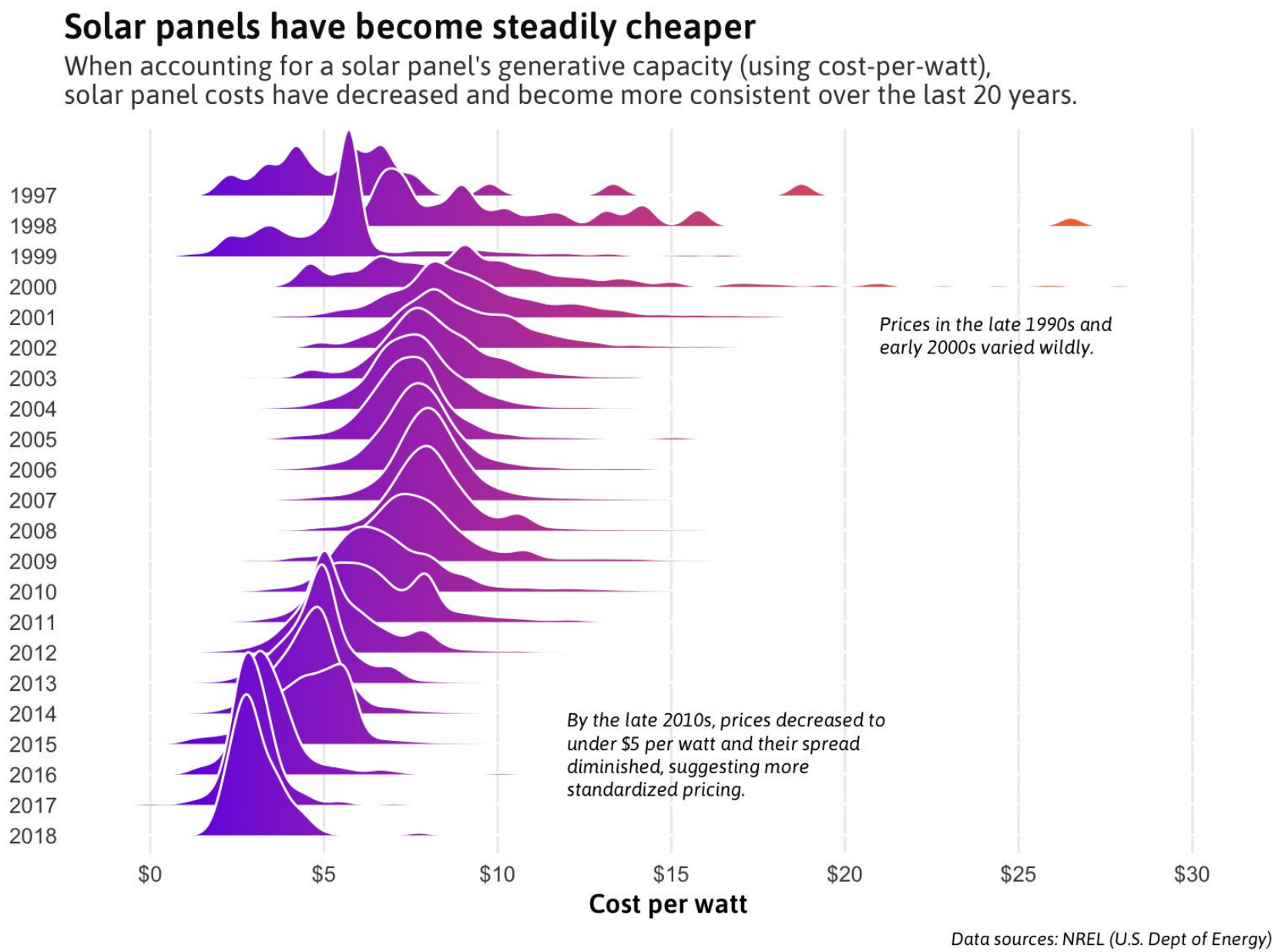


Years refer to Supreme Court terms, which run from October to September.

FiveThirtyEight

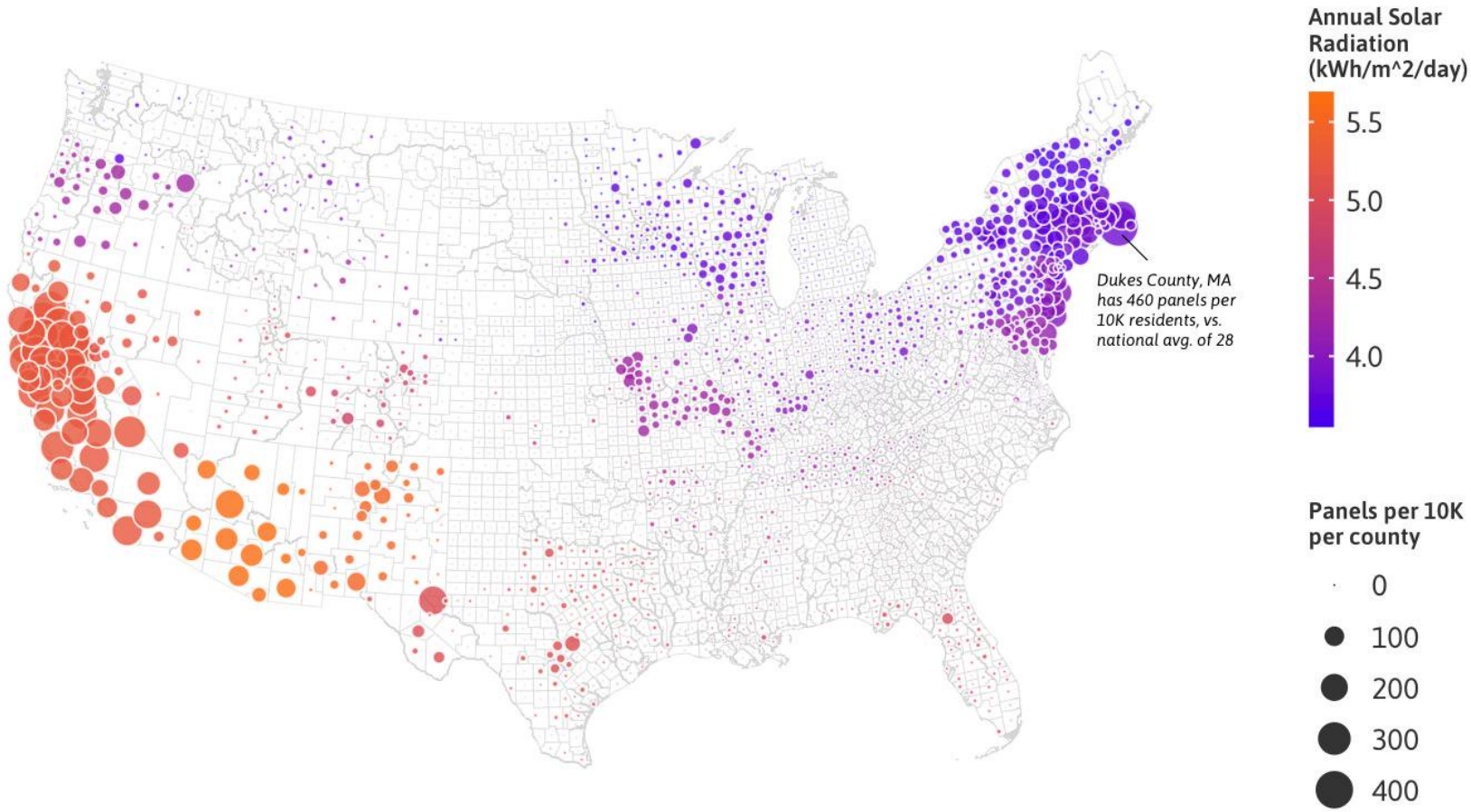
SOURCES: LEE EPSTEIN, CHAD WESTERLAND, THE JOURNAL OF LAW, ECONOMICS, AND ORGANIZATION


Partiernes Danmarkskort ved Folketingsvalget 2019
Af @fghjorth med inspiration fra @hnrklndbrg
A. Socialdemokratiet 25.9%
V. Venstre, Danmarks Liberale Parti 23.4%
O. Dansk Folkeparti 8.7%
B. Radikale Venstre 8.6%
F. SF - Socialistisk Folkeparti 7.7%
Ø. Enhedslisten - De Rød-Grønne 6.9%
C. Det Konservative Folkeparti 6.6%
Å. Alternativet 3.0%
D. Nye Borgerlige 2.4%
I. Liberal Alliance 2.3%
P: Stram Kurs 1.8%
K. Kristendemokraterne 1.7%
2x
1x
0.5x


Solar panels have become steadily cheaper
When accounting for a solar panel's generative capacity (using cost-per-watt),
solar panel costs have decreased and become more consistent over the last 20 years.
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Prices in the late 1990s and early 2000s varied wildly.
By the late 2010s, prices decreased to under $5 per watt and their spread diminished, suggesting more standardized pricing.
$0
$5
$10
$15
$20
$25
$30
Cost per watt
Data sources: NREL (U.S. Dept of Energy)

Large panels-per-capita in the Northeast, despite low annual solar radiation

While the Northeast region receives the least solar irradiance in the US, it also has counties with disproportionately high solar panel installation rates.



Data source: NREL (U.S. Dept of Energy)
Note: HI and AK are omitted due to lack of comparable solar resource data available.

Drawbacks to R

- No graphical interface out of the box
- Can be a bit intimidating at first
- Cryptic error messages
  - `object of type 'closure' is not subsettable`

You will code every week in this course

- If you have not coded before, it will be frustrating at first
- Everyone who finishes this course has learned how to code in R
- For those who already know R (perhaps many of you), you will probably need to learn more to complete the assignments
- We will use `tidyverse` throughout
  - You can use Base R functions as well, if you really want to

VICTORIA FINE • AUG 31, 2015 • 4:59 AM

I Was Intimidated by Coding Until I Learned This Secret Strategy: Googling

- Searching on Google and asking an LLM like ChatGPT is a skill
- You will develop this through experience
- It is extremely rare that one cannot find a solution by using Google or ChatGPT
- Code from ChatGPT that you copied and pasted into your assignments will be obvious
- Use it to help you debug, and to give suggestions, but you need to implement these things yourself

Latex

- Typesetting system for scientific writing
- As easy as or easier to learn than R
- Modern way to use LaTeX is through Overleaf
  - Overleaf is an online LaTeX interface for collaboration
- Free
- Does your bibliography for you
- Easy formatting of regression tables from R
- 'Easy' to do mathematical formulas
- Can work simultaneously on the same document (like Google Docs)

Let's learn some R

Install R:

`https://cran.r-project.org/mirrors.html`

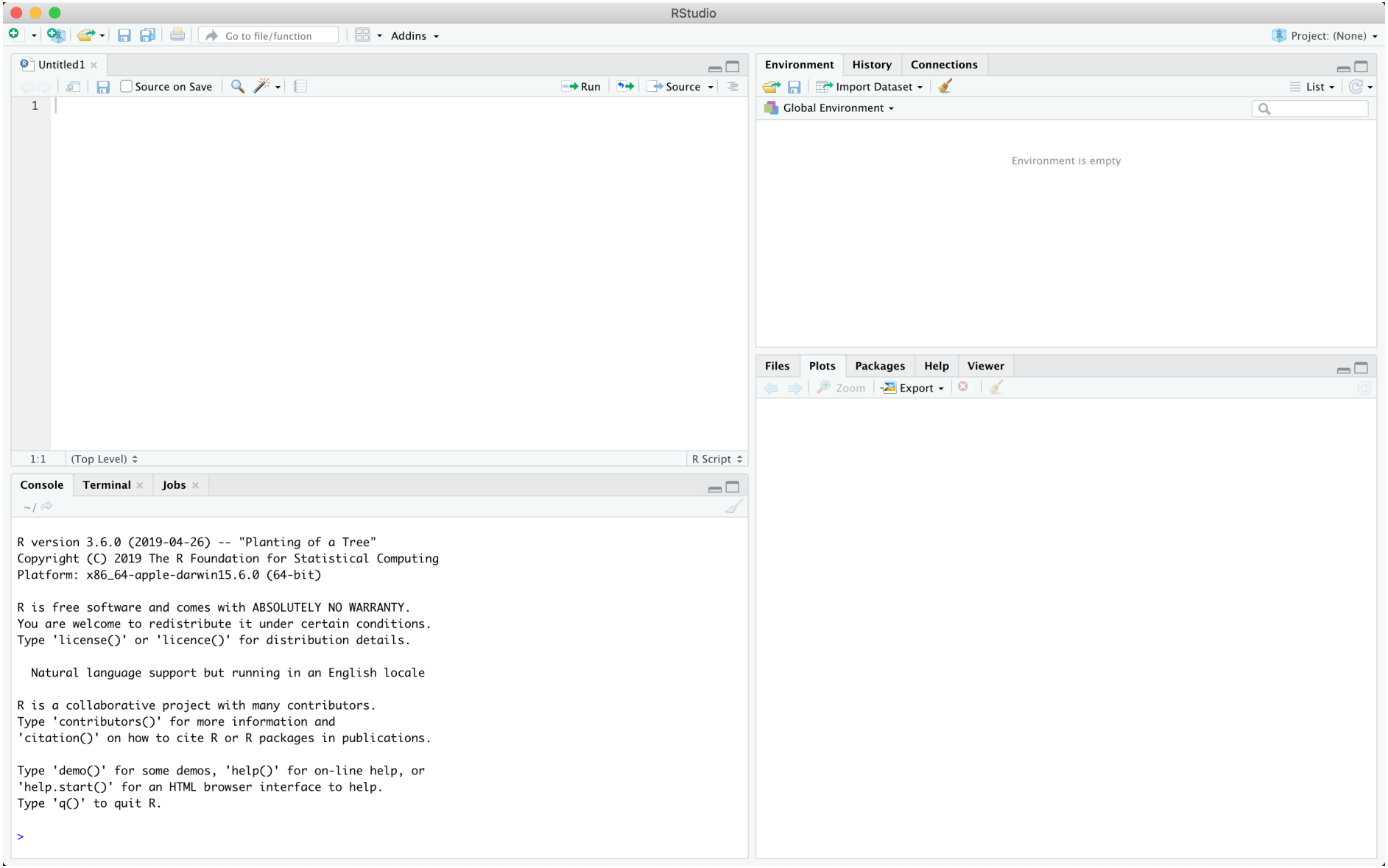
Install RStudio:

`https://posit.co/download/rstudio-desktop/`

These links are on the course website under the first "Weekly Readings"

What is RStudio?

- An Integrated Development Environment (IDE)


RStudio
Go to file/function
Addins
Project: (None)
Untitled1
Source on Save
Run
Source
1
1:1
(Top Level)
R Script
Console
Terminal
Jobs
~/
R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
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
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Zoom
Export

RStudio basics

- Comment and uncomment: `Ctrl + Shift + c`
- Run line of code: `Ctrl + Enter`
- Run code in console: `Enter`
- Clear console: `Ctrl + l`
- Indent: `Tab`
- Outdent: `Shift + Tab`

R basics

- Calculate: `(2 + 5) / 7`
- Create a variable:
  - `var_1 <- c(1, 4, 5, 7)`
- Create another variable:
  - `var_2 <- c(1, 15, 25, -1)`
- Elements of a vector: `var_1[1:3]`
- Arithmetic with variables:
  - `var_1 * var_2`
  - `var_1 * 2`
- Create a data.frame
  - `D <- data.frame(var_1, x = c(1, 2, 3, 4))`
  - `D <- data.frame(var_1, x = c(1, 2, 3))`

R basics

- Variables in data: `names(D)`
- Index variable within a data.frame:
  - `D$var1[2:4]`
- Create new variable within existing data.frame:
  - `D$new_var <- c(100, 200, 300, 400)`
- Basic sampling functions:
  - `sample(1:6, 10)`
  - `rnorm(10, mean = 5, sd = 0.25)`

R basics

- Install a package: `install.packages("tidyverse")`
- Load a package: `library(tidyverse)`
- Load data from a package: `data(iris)`
- What variables are in the data?: `names(iris)`
- Histogram of a variable:
  - `ggplot(iris, aes(x = Sepal.Length)) + geom_histogram()`

A "for" loop

```
x <- NA
for (i in 1:10) {
    print(i)
    x[i] <- i
}
```

Functions

```
get_mean <- function(x) {
  the_mean <- sum(x) / length(x)
  return(the_mean)
}
```

Loading data

- Download GM.csv from the course website
- Then load that dataset into R by doing the following:

```
# Find the directory and file name of the file
# Set your working _directory_
setwd("/Users/mkf215/Downloads/")

# Load the dataset GM into a data.frame object
# that we'll call "D"
D <- read_csv("GM.csv")
```

What's next

- Introduction to `tidyverse` by Hadley Wickham
- Data processing with `tidyverse`
- To do:
  - Complete the next DataCamp exercises before next class