

GRD 610A

Data Visualization II

January 20, 2021 | Introduction





About the Course





Land Acknowledgement

“The College for Creative Studies...[sits] on the traditional, contemporary, and ancestral homelands of the Anishinaabe - Council of Three Fires: the Ojibwe/Chippewa, Odawa/Ottawa, and Potawatomi/Bodéwadmí along with their neighbors the Seneca, Delaware, Fox, Shawnee, Loups, Miami and Wyandot who maintained, and continue to preserve lifeways along Detroit’s river banks and throughout the Great Lakes region. Through signing the Treaty of Detroit in 1807, Anishinaabek tribes ceded the land now occupied by the city we stand on. We recognize Michigan is home to 12 federally recognized tribes who continue to steward this land, in remembrance of their ancestors and thinking of future generations.”



Who is your instructor?



Jenn Schilling (she/her/hers)

Senior Data Analyst, University Analytics & Institutional Research, University of Arizona

Experience as a middle school teacher, data visualization consultant, operations research engineer in the supply chain, statistician in advertising and market research

Originally from Maryland



Quick Round of Introductions

- Preferred Name and Pronouns
- Why are you pursuing the Systems Design Thinking MFA?
- What has been your favorite part of the Systems Design Thinking MFA program?
- What was your favorite part of winter break?



Class Survey

Please complete the following short survey so I can get to know you a bit better and understand your expectations of the course:

<https://forms.gle/EaVWhy71jVrXuybq8>



Course Learning Outcomes

- Apply the basics of programming and data manipulation in R
- Communicate analyses with charts and plots using R visualization tools
- Design and evaluate information visualizations and other forms of visual presentation
- Apply a workflow for moving from a dataset to a fully designed static visualization



Course Expectations

- Be prepared
- Be kind
- Be honest
- Be creative



Syllabus & Semester Schedule





Grading

Understanding of class material and preparation for discussion	30%
Breadth and depth of lab exploration and homework	30%
Execution and Presentation of midterm and final projects	30%
Classroom attendance and participation	10%

- Competency
- Clarity
- Creativity

BREAK - 15 Minutes

Complete the class survey during the break.
Link on Blackboard and in Zoom Chat.



Data Visualization





Current Data Visualization Perspective

Visit this Padlet and post at least one comment in each column:

https://padlet.com/jschilling_ccs/mmzv9zq2ic8e6iqs



Why visualize data?

- Understand data
- Show patterns
- Engage viewer
- Convince viewer
- Tell a story



13 Datasets: Data

x <dbl>	y <dbl>
55.3846	97.1795
51.5385	96.0256
46.1538	94.4872
42.8205	91.4103
40.7692	88.3333
38.7179	84.8718
35.6410	79.8718
33.0769	77.5641
28.9744	74.4872
26.1538	71.4103

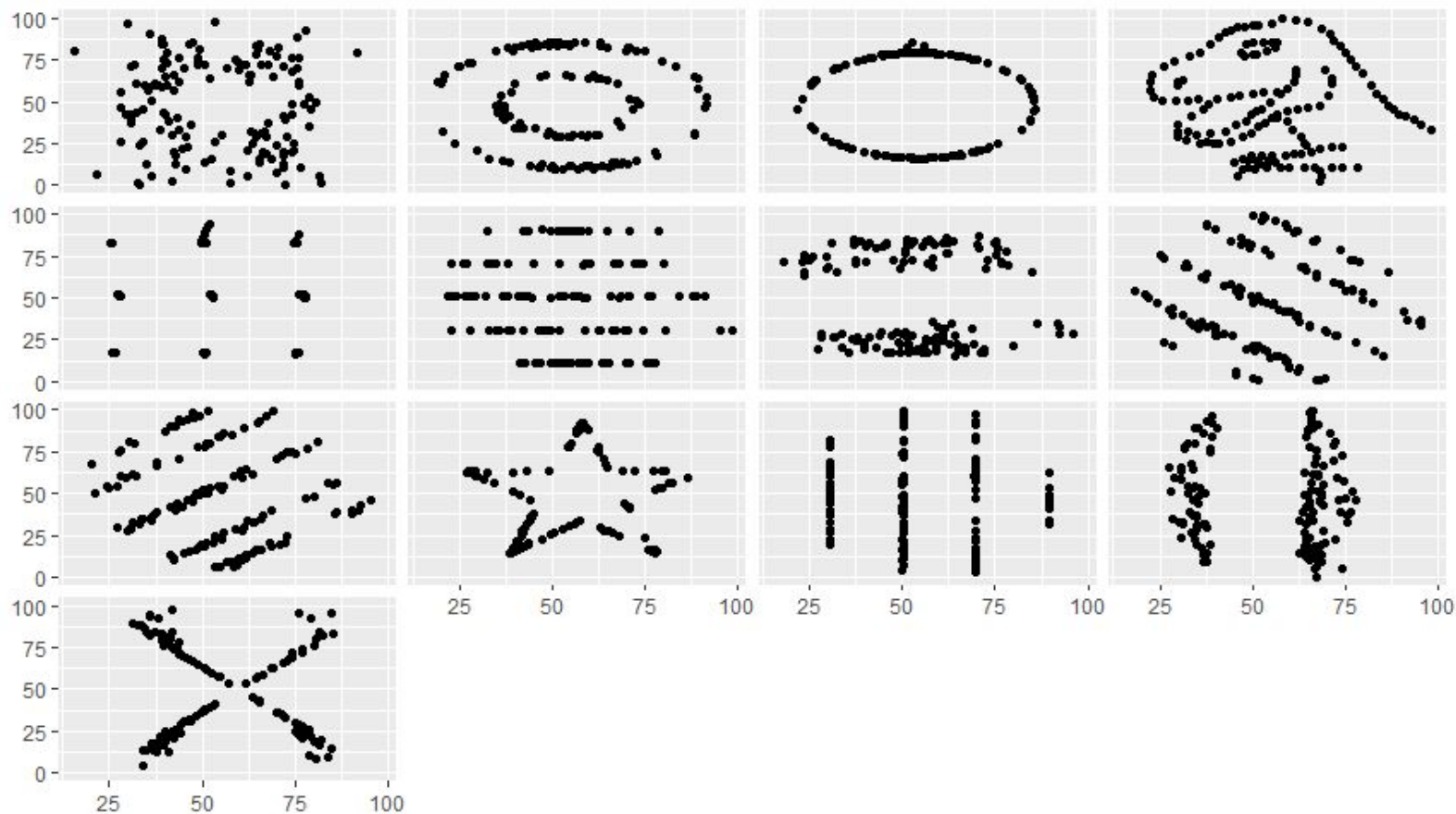


13 Datasets: Summary Statistics

Dataset	Mean of X	Mean of Y	Standard Dev. of X	Standard Dev. of Y	Correlation
1	54.266	47.835	16.770	26.940	-0.064
2	54.269	47.831	16.769	26.936	-0.069
3	54.267	47.838	16.760	26.930	-0.068
4	54.263	47.832	16.765	26.935	-0.064
5	54.260	47.840	16.768	26.930	-0.060
6	54.261	47.830	16.766	26.940	-0.062
7	54.269	47.835	16.767	26.940	-0.069
8	54.268	47.836	16.767	26.936	-0.069
9	54.266	47.831	16.769	26.939	-0.069
10	54.267	47.840	16.769	26.930	-0.063
11	54.270	47.837	16.770	26.938	-0.069
12	54.267	47.832	16.770	26.938	-0.067
13	54.260	47.840	16.770	26.930	-0.066

The Datasaurus Dozen

Never trust summary statistics alone; always visualize your data - Alberto Cairo





**What makes an effective
data visualization?**





“The greatest value of a picture
is when it forces us to notice
what we never expected to see.”

John W. Tukey, *Exploratory Data Analysis*



Some Thoughts from Tufte

“Confusion and clutter are failures of design, not attributes of information. And so the point is to **find design strategies that reveal detail and complexity** - rather than to fault the data for an excess of complication. Or, worse, to fault viewers for a lack of understanding.” - Edward Tufte

“Graphical excellence is the **well-designed presentation of interesting data** - a matter of substance, of statistics, and of design...[It] consists of **complex ideas communicated with clarity, precision, and efficiency**. ...[It] is that which gives to the viewer the **greatest number of ideas** in the **shortest time** with the **least ink** in the **smallest space**...[It] is nearly always multivariate...And graphical excellence requires **telling the truth about the data**.” - Edward Tufte, *The Visual Display of Quantitative Information*, p. 51



Some Thoughts from Cairo

“A good visualization is:

1. reliable information,
2. visually encoded so relevant patterns become noticeable,
3. organized in a way that enables at least some exploration, when it's appropriate,
4. and presented in an attractive manner, but always remembering that honesty, clarity, and depth come first.”

Alberto Cairo, *The Truthful Art*, p. 12



Some Thoughts from Cairo

The Five Qualities of Great Visualizations:

1. Truthful
2. Functional
3. Beautiful
4. Insightful
5. Enlightening

Alberto Cairo, *The Truthful Art*, p. 45



Some Thoughts from Healy

“The graphs you make are **meant to be looked at by someone**. The effectiveness of any particular graph is **not just a matter of how it looks** in the abstract but **also a question of who is looking at it, and why**.” - Kieran Healy, Data Visualization: A Practical Introduction, p. 1

“Good visualization methods offer **extremely valuable tools** that we should use in the process of **exploring, understanding, and explaining data**. But they are not a magical means of seeing the world as it really is. **They will not stop you from trying to fool** other people if that is what you want to do, and they may not stop you from fooling yourself either.” - Kieran Healy, Data Visualization: A Practical Introduction, p. 1



Evaluating a Data Visualization

[The Junk Charts Trifecta Checkup:](#)

- What is the question?
- What does the data say?
- What does the chart say?



What's Wrong?

From Facebook Q3 2016 Report

Worldwide



US & Canada





What's Wrong?

From the 2012 World Happiness Report

Figure 11: GNH index by gender

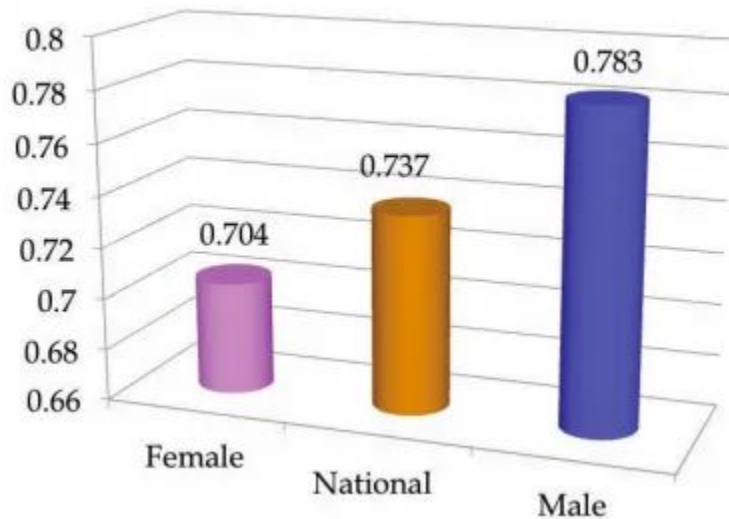
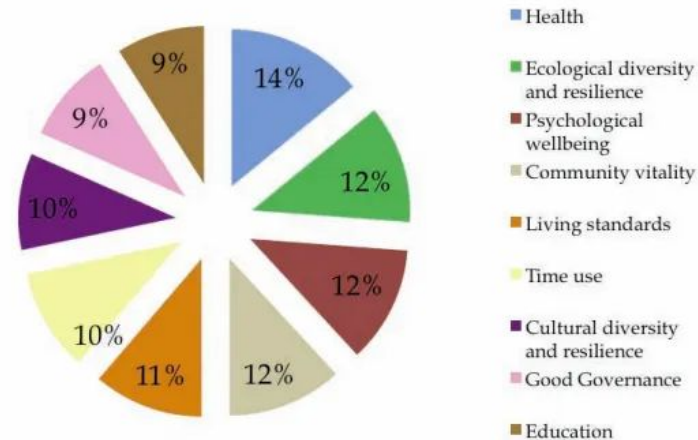
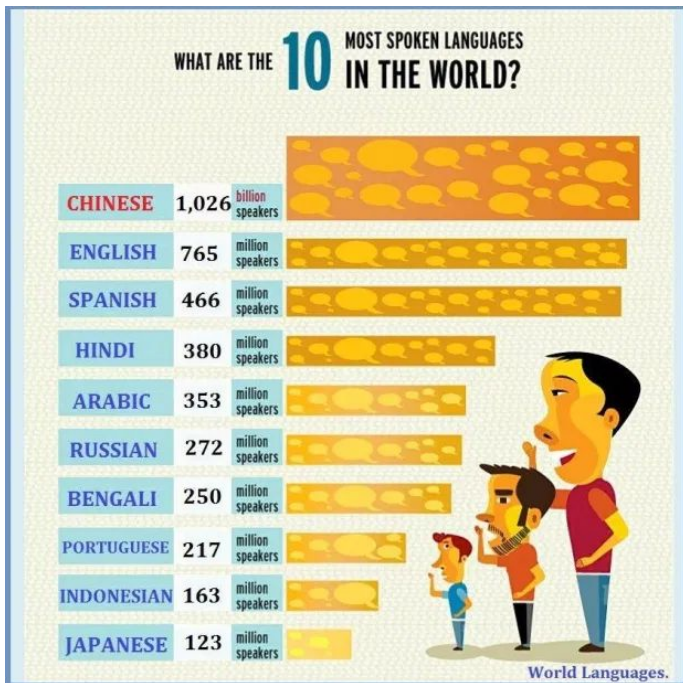


Figure 4: In which domains do happy people enjoy sufficiency?



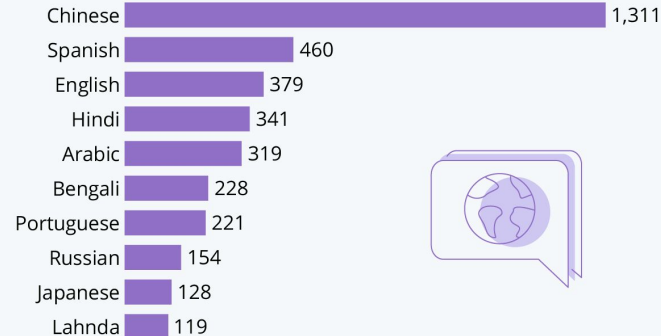
What's Wrong?

Source unknown



The World's Most Spoken Languages

Estimated number of first-language speakers worldwide in 2019 (millions)*



* Each language also includes associated member languages and varieties

Source: Ethnologue



statista



R



R



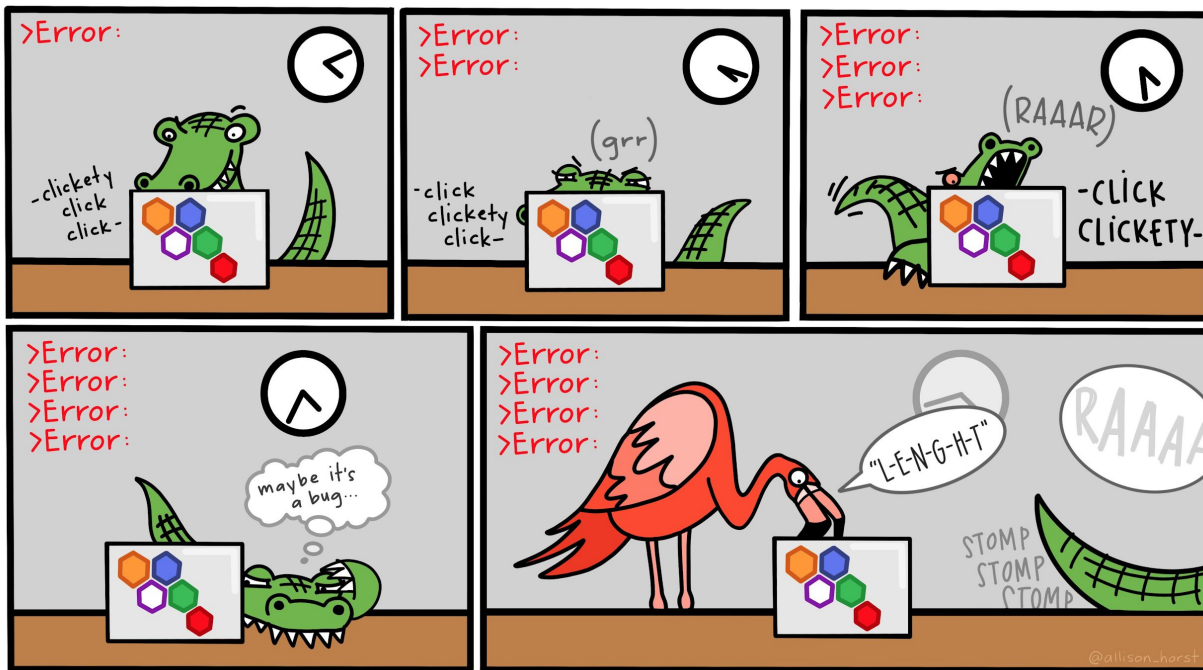
This course is based around the programming language **R**, which we will be using through **RStudio**, an integrated development environment (IDE).


If this is your first time programming, that's okay! This course is intended to build your R skills over the course of the semester.

Remember that learning R is like learning a new language, and it can be challenging.

If you're struggling with an error - take a break, email me, ask a classmate, search online (include keywords like "ggplot" when searching Google and use [Stack Overflow](#) and the [RStudio Community](#) as additional resources).

Even Experienced Programmers Struggle...



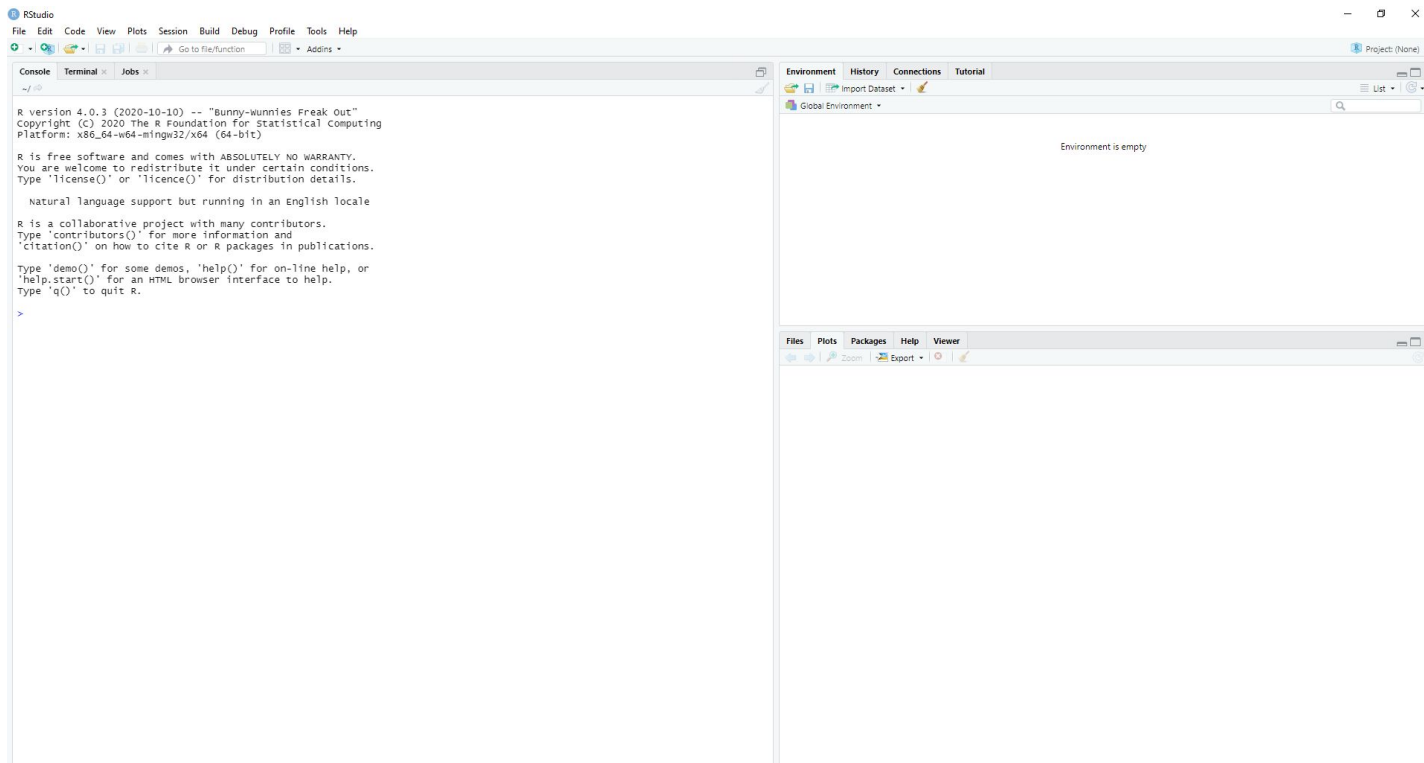


“It’s easy when you start out programming to get really frustrated and think, “Oh it’s me, I’m really stupid,” or, “I’m not made out to program.” But, **that is absolutely not the case**. Everyone gets frustrated. I still get frustrated occasionally when writing R code. It’s just a natural part of programming. So, it happens to everyone and gets less and less over time. **Don’t blame yourself. Just take a break, do something fun, and then come back and try again later.**”

Hadley Wickham, Chief Scientist at RStudio



Installing R & RStudio





Resources: Installing R & RStudio

See p. xviii of the Preface of *Data Visualization: A Practical Introduction* by Kieran Healy

See Blackboard -> Week 1 - Introductions -> How to Install R and RStudio PDF



Step 1: Install R



To install R, visit the Comprehensive R Network (CRAN) (<https://cran.r-project.org/>), and follow the links to download R for your operating system.

For Windows, click "base" or "install R for the first time".

For Mac OS, download the latest .pkg file; you will also need to download and install XQuartz.

Run the downloaded executable and follow the prompts to install R.

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.



Step 2: Install RStudio



To install RStudio, visit the downloads section of the RStudio website (<https://rstudio.com/products/rstudio/download/#download>), and click on the download button.

The website should automatically detect your operating system and show the correct version on the button.

If that is not the case, then scroll down to the table on the page and find the correct version for your operating system. Run the downloaded executable and follow the prompts.

RStudio Desktop 1.3.1093 - Release Notes

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:



Requires Windows 10/8/7 (64-bit)



All Installers

Linux users may need to [import RStudio's public code-signing key](#) prior to installation, depending on the operating system's security policy.

RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an [older version of RStudio](#).

OS	Download	Size	SHA-256
Windows 10/8/7	RStudio-1.3.1093.exe	171.62 MB	62b9e60a
macOS 10.13+	RStudio-1.3.1093.dmg	148.66 MB	bdc4d3a4
Linux 64-bit	RStudio-1.3.1093-macos.dmg	171.62 MB	77960a08

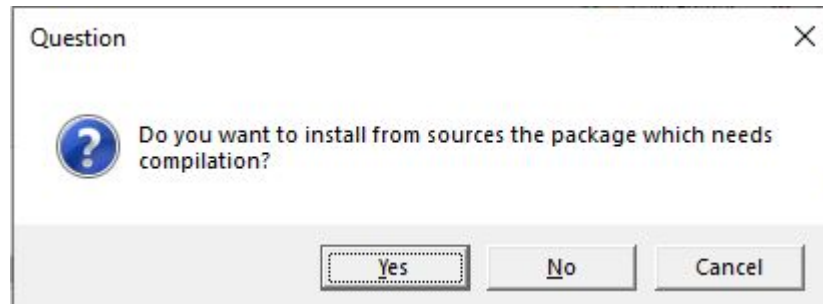


Step 3: Install Packages

R packages are collections of functions and data sets that have been developed by members of the R community. They are useful for improving and expanding the functionality of R. We will use several packages in this course related to data visualization.

Open RStudio, and copy the code on the next slide into the Console to install the packages from *Data Visualization: A Practical Introduction* by Kieran Healy.

Press Enter to run the lines of code that you copied. If you are prompted with the question "Do you want to install from sources the package which needs compilation?", select "No".





Step 3: Install Packages

```
my_packages <- c("tidyverse", "broom", "coefplot", "cowplot",  
  "gapminder", "GGally", "ggrepel", "ggribbles",  
  "gridExtra", "here", "interplot", "margins",  
  "maps", "mapproj", "mapdata", "MASS", "quantreg",  
  "rlang", "scales", "survey", "srvyr", "viridis",  
  "viridisLite", "devtools")  
  
install.packages(my_packages, repos = "http://cran.rstudio.com")  
  
devtools::install_github("kjhealy/socviz")
```



Next Class





Visualization of the Week

- An example of an effective or ineffective visualization
- Related to your interests, passions, or career interests
- Related to the current or prior week's course content
- How might you make the visualization in R?



Visualization of the Week

1 person / week

12 weeks / 4 people = 3 times a semester per person

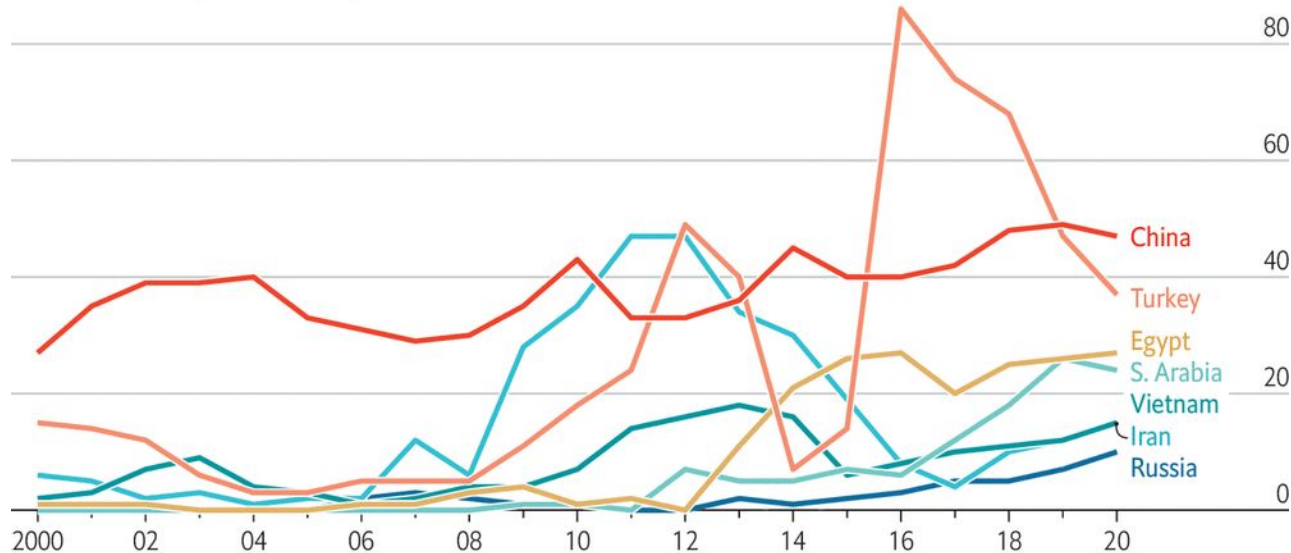
Skip: Week 1 (Today), Week 7 (Midterm Due), Week 15 (Final Class)

TODO: Create schedule for visualizations of the week

Visualization of the Week | Example

Stop the press

Total number of journalists imprisoned, 2000-20, selected countries



Source: Committee to Protect Journalists

The Economist



Tasks to Complete

- Reading (see Syllabus)
- Install R and RStudio (see Blackboard)
- Prepare for your Visualization of the Week