

Creating Engaging Data Visualizations in R with ggplot2

Angela Li

2023-02-24

tigeRs: workshop 2!

- ▶ (re)-introductions
- ▶ Thanks to DDSS for sponsoring lunch!
- ▶ Sticky notes for communication during exercises
- ▶ Live coding with me

Motivation

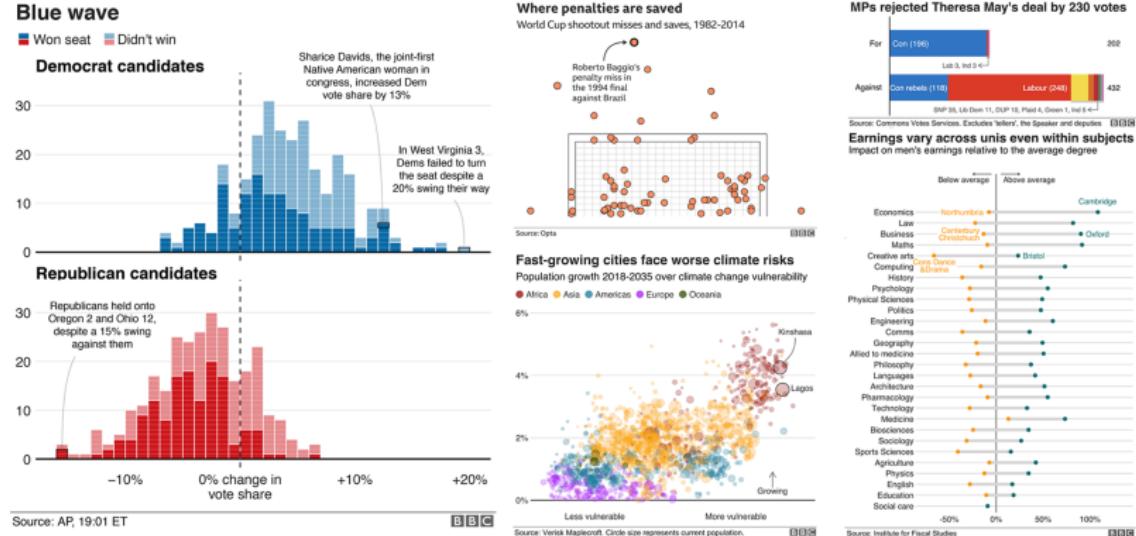
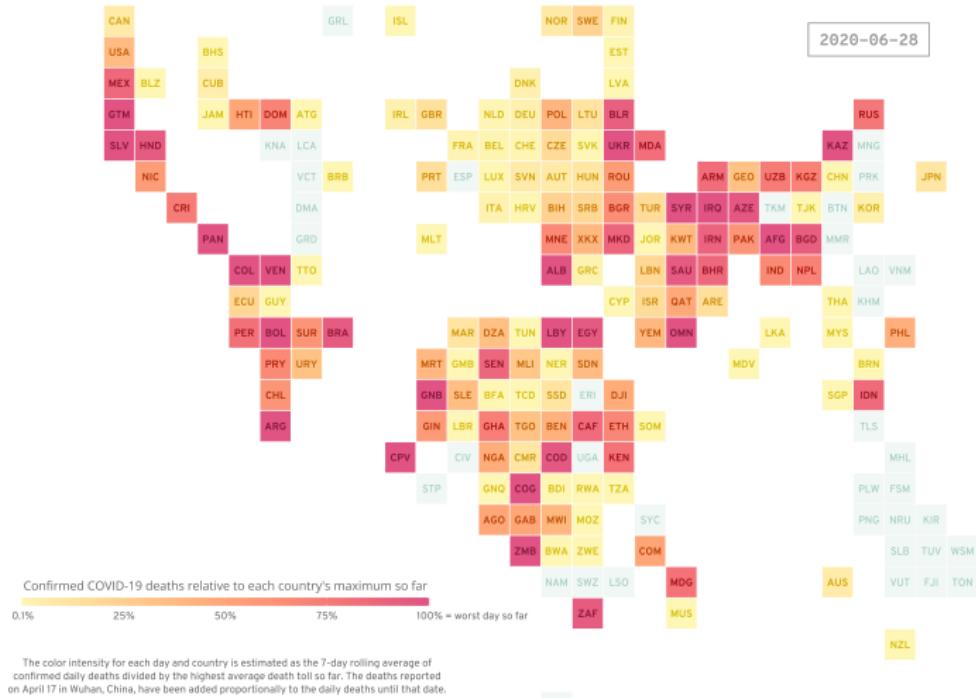


Figure 1: A selection of graphics by the BBC built with R and in-house package {bbplot}

Motivation

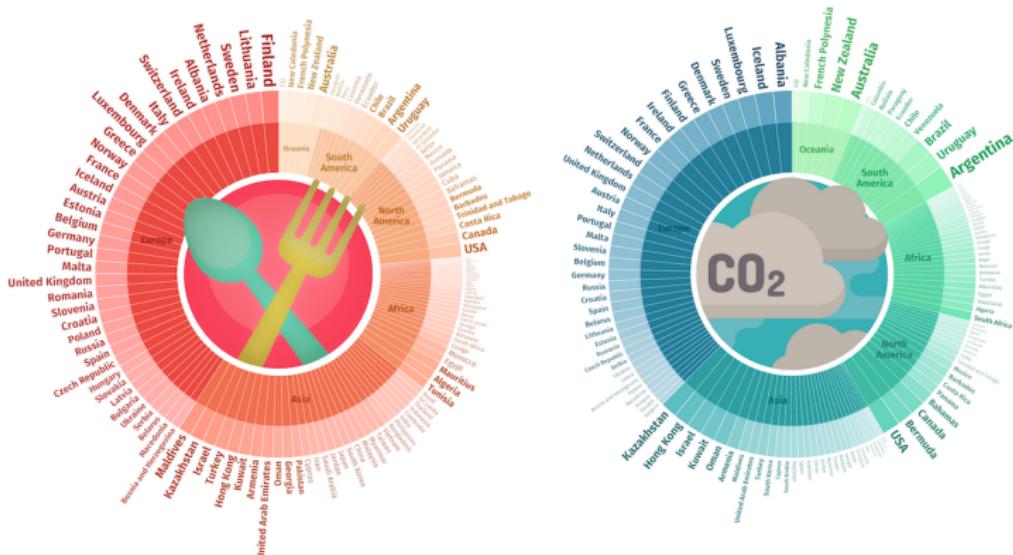
The Worst Days: Peaks in Confirmed Daily Deaths Due to COVID-19 So Far



Motivation

Food Carbon Footprint Index 2018

Global comparison of different diets in terms of **Average Consumption** (kg/person/year) of both animal and non-animal products as well as **Carbon Emissions** (kg CO₂/person/year) per continent and country. Font size and color intensity indicate each country's estimate with **countries printed in bold** belonging to the upper 50% of consumers and CO₂ emitters, respectively.

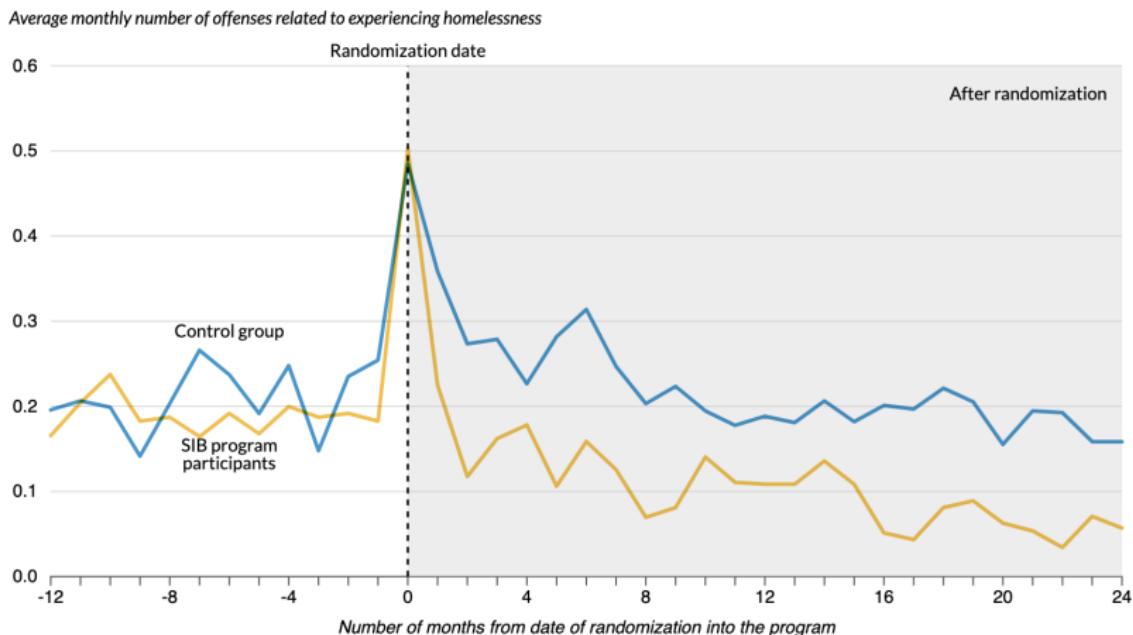


Visualization by Cédric Scherer • Data by Food and Agriculture Organization of the United Nations (FAO) via nu3 • Icons by Freepik

Figure 3: Cyclical charts in R

Motivation

People referred to SIB supportive housing were arrested for fewer offenses related to experiencing homelessness over time



Note: Because study participants were chosen based on having a police contact for an offense the day before randomization, everyone in the treatment and control groups had a police contact for an offense at month zero. The contact did not have to be for an offense associated with experiencing homelessness, which is why this number does not hit the 1 mark.

Goals

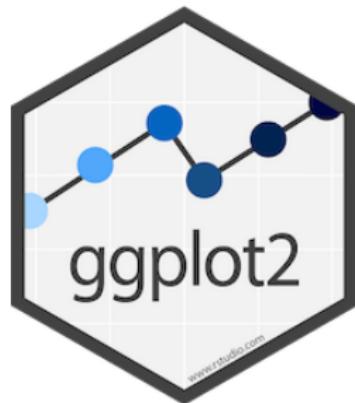
1. Get a taste of the logic behind the **grammar of graphics** (`ggplot2`)
2. Learn some good-enough figure design practices
3. Tell you about my favorite high-quality - and open-source! - learning resources

The Grammar of Graphics

The Grammar of Graphics

ggplot2 is a system for declaratively creating graphics, based on “The Grammar of Graphics” (Wilkinson, 2005).

You provide the data, tell ggplot2 how to **map variables to aesthetics**, what graphical features to use, and it takes care of the details.



The Grammar of Graphics

ggplot2 is based on the **grammar of graphics**, the idea that you can build every graph from the same components: a **data** set, a **coordinate system**, and geoms—visual marks that represent data points.

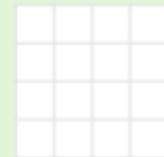


The Grammar of Graphics

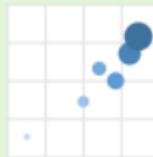
To display values, map variables in the data to visual properties of the geom (**aesthetics**) like **size**, **color**, and **x** and **y** locations.



data
geom
 $x = F$ • $y = A$
 $color = F$
 $size = A$



**coordinate
system**



plot

The Grammar of Graphics

Complete the template below to build a graph.

```
ggplot (data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>),  
    stat = <STAT>, position = <POSITION>) +  
  <COORDINATE_FUNCTION> +  
  <FACET_FUNCTION> +  
  <SCALE_FUNCTION> +  
  <THEME_FUNCTION>
```

required

Not required,
sensible
defaults
supplied

Advantages of {ggplot2}

- ▶ consistent underlying “grammar of graphics” (Wilkinson 2005)

Advantages of {ggplot2}

- ▶ consistent underlying “grammar of graphics” (Wilkinson 2005)
- ▶ very flexible, layered plot specification

Advantages of {ggplot2}

- ▶ consistent underlying “grammar of graphics” (Wilkinson 2005)
- ▶ very flexible, layered plot specification
- ▶ theme system for polishing plot appearance

Advantages of {ggplot2}

- ▶ consistent underlying “grammar of graphics” (Wilkinson 2005)
- ▶ very flexible, layered plot specification
- ▶ theme system for polishing plot appearance
- ▶ lots of additional functionality thanks to extensions

Advantages of {ggplot2}

- ▶ consistent underlying “grammar of graphics” (Wilkinson 2005)
- ▶ very flexible, layered plot specification
- ▶ theme system for polishing plot appearance
- ▶ lots of additional functionality thanks to extensions
- ▶ active and helpful community

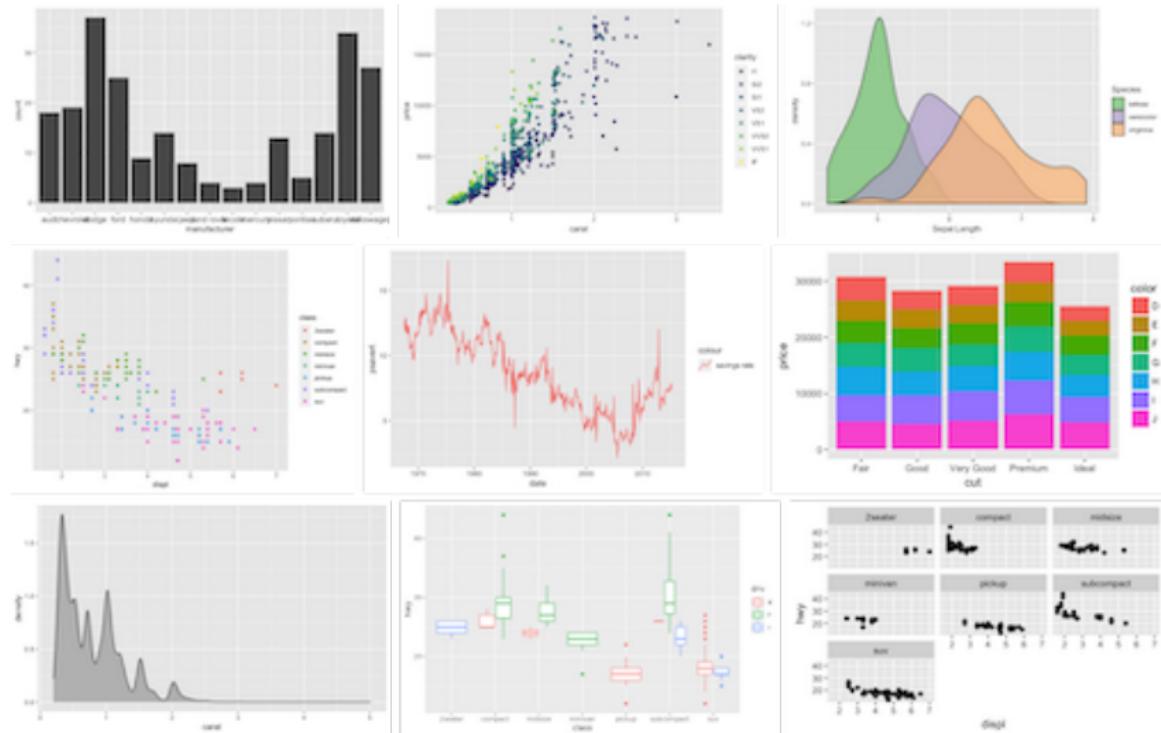


Figure 4: ggplot2 Basic ggplot2 examples featured on ggplot2.tidyverse.org

The ggplot2 Package

... is an **R package to visualize data** created by Hadley Wickham in 2005

```
# install.packages("ggplot2")
library(ggplot2)
```

... is part of the {tidyverse}

```
# install.packages("tidyverse")
library(tidyverse)
```

Let's jump in with examples!
`<ggplot-exercises.Rmd>`

Good Enough Figure Design

Why “Good Enough”?

- ▶ Depends on if visualization is exploratory or for presentation
- ▶ Not everyone is a data viz specialist
- ▶ But some practices are low-hanging fruit

ggplot2: VISUAL DATA EXPLORATION



ggplot2:

Build a data
MASTERPIECE



HORST '18

Angela's Top Figure Design Tips

1. Make your figure labels **larger**.

Angela's Top Figure Design Tips

1. Make your figure labels **larger**.
2. Include a descriptive title/subtitle with your main finding to **tell the story**. (Bonus: this sometimes allows you to remove obvious axis labels.)

Angela's Top Figure Design Tips

1. Make your figure labels **larger**.
2. Include a descriptive title/subtitle with your main finding to **tell the story**. (Bonus: this sometimes allows you to remove obvious axis labels.)
3. Reduce unnecessary “ink” to highlight the “data” in your graph (**data-ink ratio**)

Angela's Top Figure Design Tips

1. Make your figure labels **larger**.
2. Include a descriptive title/subtitle with your main finding to **tell the story**. (Bonus: this sometimes allows you to remove obvious axis labels.)
3. Reduce unnecessary “ink” to highlight the “data” in your graph (**data-ink ratio**)
4. **Draw out the graph** you want to make before jumping into R.

Angela's Top Figure Design Tips

1. Make your figure labels **larger**.
2. Include a descriptive title/subtitle with your main finding to **tell the story**. (Bonus: this sometimes allows you to remove obvious axis labels.)
3. Reduce unnecessary “ink” to highlight the “data” in your graph (**data-ink ratio**)
4. **Draw out the graph** you want to make before jumping into R.
5. When in doubt, **look at existing models** for best practices (I love Urban Institute’s style guide for policy communication, and PNAS academic journal articles tend to have nice visualizations in R).

Angela's Top Figure Design Tips

1. Make your figure labels **larger**.
2. Include a descriptive title/subtitle with your main finding to **tell the story**. (Bonus: this sometimes allows you to remove obvious axis labels.)
3. Reduce unnecessary “ink” to highlight the “data” in your graph (**data-ink ratio**)
4. **Draw out the graph** you want to make before jumping into R.
5. When in doubt, **look at existing models** for best practices (I love Urban Institute’s style guide for policy communication, and PNAS academic journal articles tend to have nice visualizations in R).
6. If you can’t make the figure work in ggplot2, you might need to **reshape the data** into long format

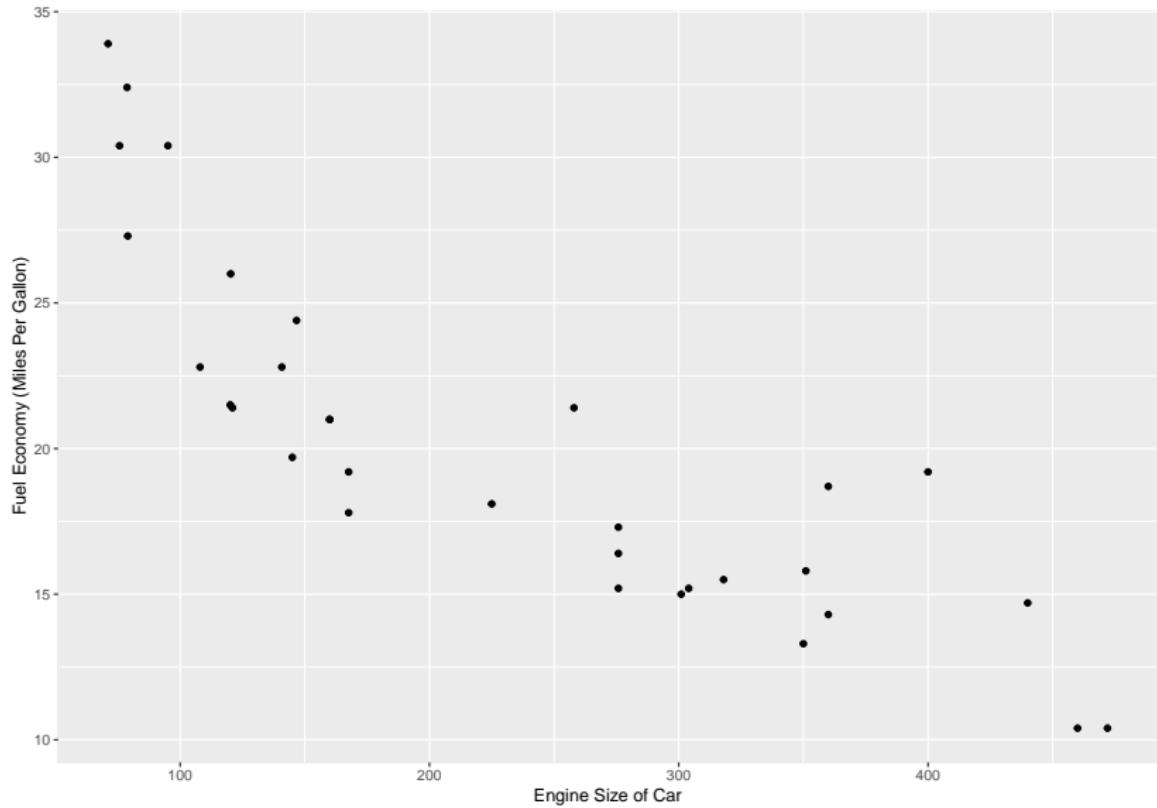
Angela's Top Figure Design Tips

1. Make your figure labels **larger**.
2. Include a descriptive title/subtitle with your main finding to **tell the story**. (Bonus: this sometimes allows you to remove obvious axis labels.)
3. Reduce unnecessary “ink” to highlight the “data” in your graph (**data-ink ratio**)
4. **Draw out the graph** you want to make before jumping into R.
5. When in doubt, **look at existing models** for best practices (I love Urban Institute’s style guide for policy communication, and PNAS academic journal articles tend to have nice visualizations in R).
6. If you can’t make the figure work in ggplot2, you might need to **reshape the data** into long format
7. There’s **probably an R package** that can make that {insert visualization here}.

OK Figure (“Ugly”)

```
ggplot(mtcars) +  
  geom_point(aes(x = disp, y = mpg)) +  
  labs(x = "Engine Size of Car",  
       y = "Fuel Economy (Miles Per Gallon)")
```

OK Figure (“Ugly”)



Better Figure

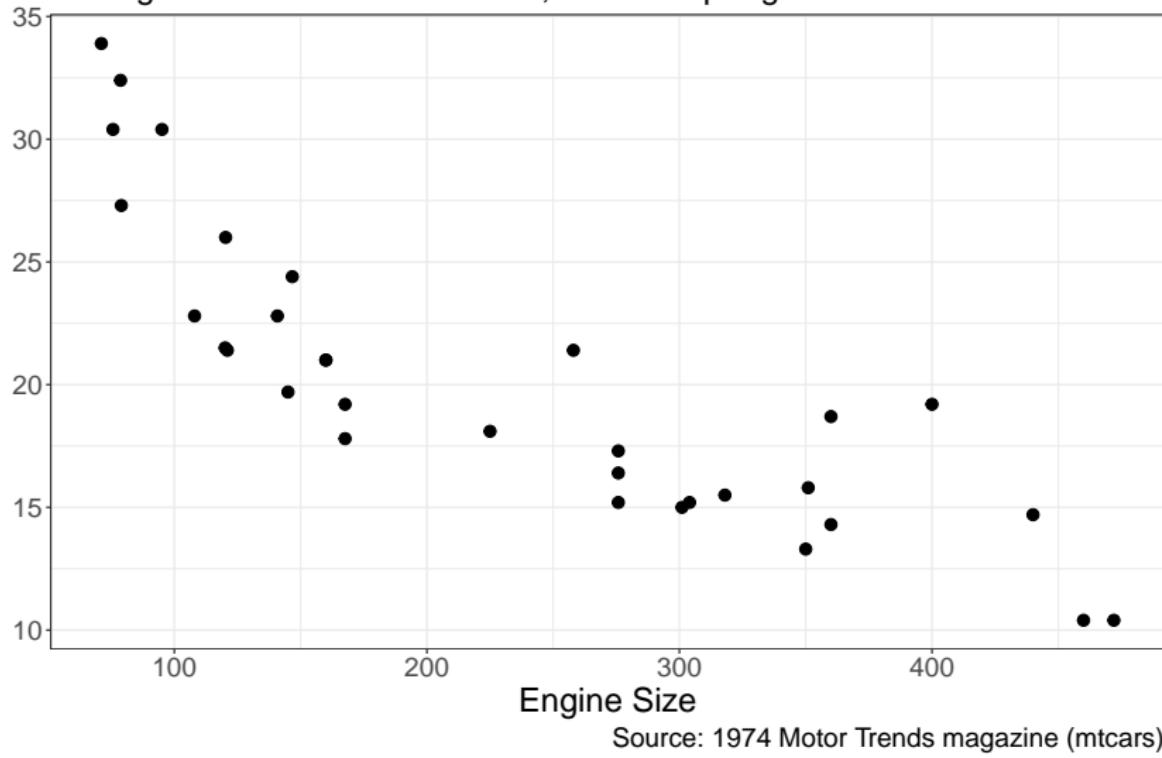
```
ggplot(mtcars) +  
  geom_point(aes(x = disp, y = mpg)) +  
  labs(x = "Engine Size",  
        y = "",  
        title = "Fuel Economy and Engine Size",  
        subtitle = "As engine size of a car increases, the fuel economy decreases.",  
        caption = "Source: 1974 Motor Trends magazine (mtcars)",  
        theme_bw() +  
        theme(text = element_text(size = 20)))
```

- ▶ See Wilke, Fundamentals of Data Visualization, Parts II and III for more tips.

Better Figure

Fuel Economy and Engine Size

As engine size of a car increases, the miles per gallon decrease



Better Figure Inspiration

ON THE WEB

- ▶ From Urban Institute Graphics Style Guide

Learning Resources

Open by Default

R is an **open-source** programming language. This means source code is public and can be remixed. Learning resources often are too. So...

- ▶ There's a very, very strong culture around **open-by-default** resources (with Creative Commons licenses, Github code posting, etc.)

Open by Default

R is an **open-source** programming language. This means source code is public and can be remixed. Learning resources often are too. So...

- ▶ There's a very, very strong culture around **open-by-default** resources (with Creative Commons licenses, Github code posting, etc.)
- ▶ The R community cares a lot about **helping folks learn** the language

Open by Default

R is an **open-source** programming language. This means source code is public and can be remixed. Learning resources often are too. So...

- ▶ There's a very, very strong culture around **open-by-default** resources (with Creative Commons licenses, Github code posting, etc.)
- ▶ The R community cares a lot about **helping folks learn** the language
- ▶ Many R users work in **collaborative** settings in academia, industry, and more.

My Favorite Open-Source Resources

Tutorials:

- ▶ **Posit Cloud Primers:** Interactive online tutorials to learn data viz in R, from the makers of RStudio
- ▶ **Data Visualization course by Andrew Heiss:** Designed as an entire course in R for policy school students

Cookbooks:

- ▶ **R Graphics Cookbook, 2nd edition:** Gives you recipes for creating various graphs in R - <https://r-graphics.org/>
- ▶ **R Graph Gallery:** Similarly, sorts R recipes by type of graph

My Favorite Open-Source Resources

Whole Open-Source Books:

- ▶ **ggplot2 book**: Nitty gritty on how ggplot2 all works, from the creator of ggplot2
- ▶ **Fundamentals of Data Visualization**: Less specific to ggplot2, more on best practices to learn data viz
- ▶ **Data Visualization: A Practical Introduction**: Online book by Soc Prof Kieran Healy walking you through exercises in R

Community:

- ▶ **Tidy Tuesday R Community**: A community of folks learning R!
- ▶ **#rstats Twitter**: Search this hashtag to find all the R folks online.
- ▶ **tigeRs**: Keep coming to our workshops!

Takeaway: The best R resources are free,
because the R community is awesome.

Thanks!

- ▶ Remixed slides and materials from:
 - ▶ Cedric Scherer's Graphic Design with ggplot2
 - ▶ Introduction to R for Geospatial Data
 - ▶ Emily Burchfield Courses
- ▶ This work is licensed under CC BY NC ND 4.0 as well, and on Github.
- ▶ Next tigeRs meeting: Lightning Talks after Spring Break!