Week 7 – Still further into Data Viz

Joshua Rosenberg and Alex Lishinski

March 4, 2021

Welcome!

Welcome to *week 7*!

LET ME SHOW YOU JUST HOW FAR DOWN THE RABBIT HOLE GOES

quickmeme.com

Record the meeting

Breakout rooms!

Starting with whomever likes spicy food the most . . .

One question:

- What is one interesting or exciting thing you did or found out about when using R this past week?

One reflection/discussion:

- What is one take–away (large or small) that you can use in your own visualizations from the two readings from last week?

- 1: https://clauswilke.com/dataviz/histograms–density–plots.html
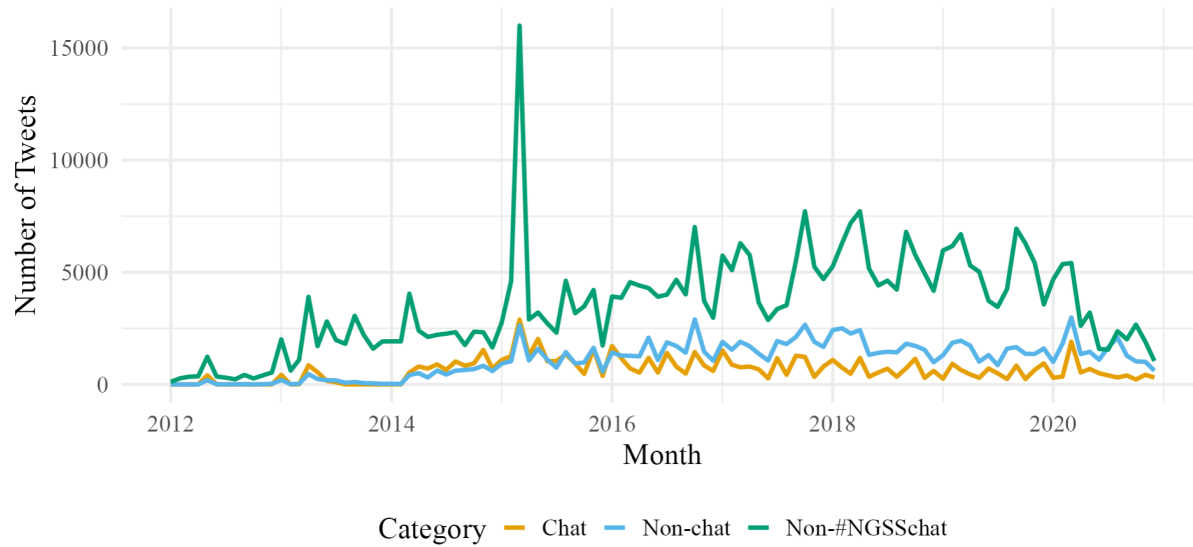- 2: https://clauswilke.com/dataviz/visualizing–proportions.html

Review of last week's class

- Review of the grammar of graphics

- Understanding visualizations by layers

- Understanding mapping of data to geoms

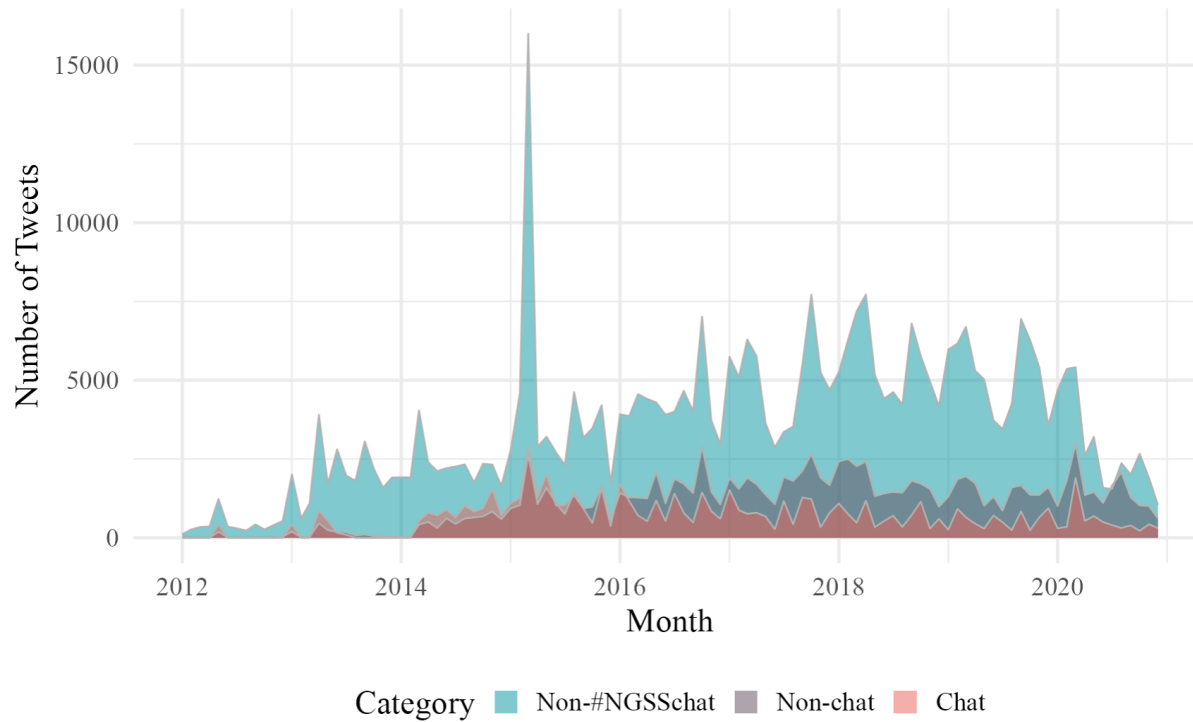- Homework: counting, grouping and summarizing, recoding using a factor, and telling a story with data
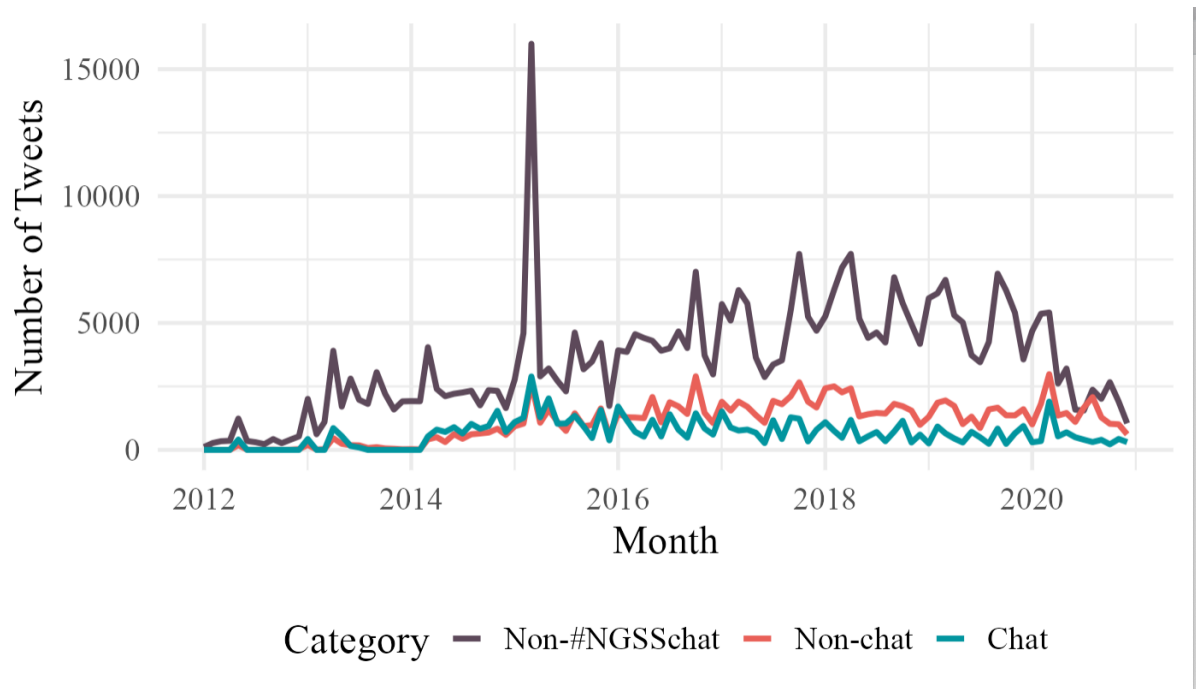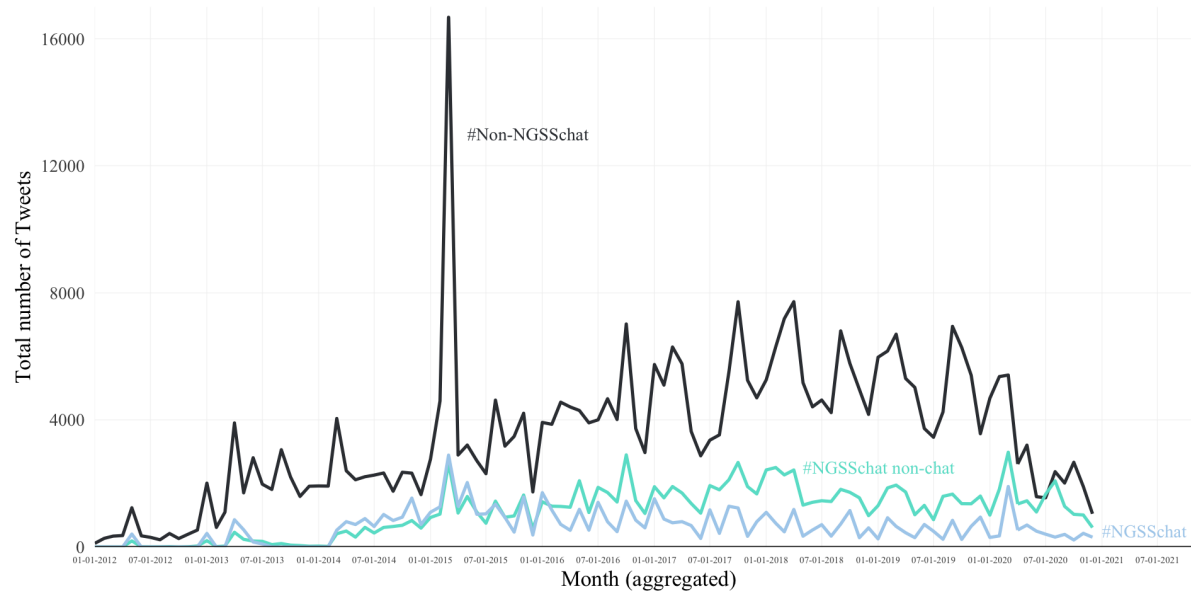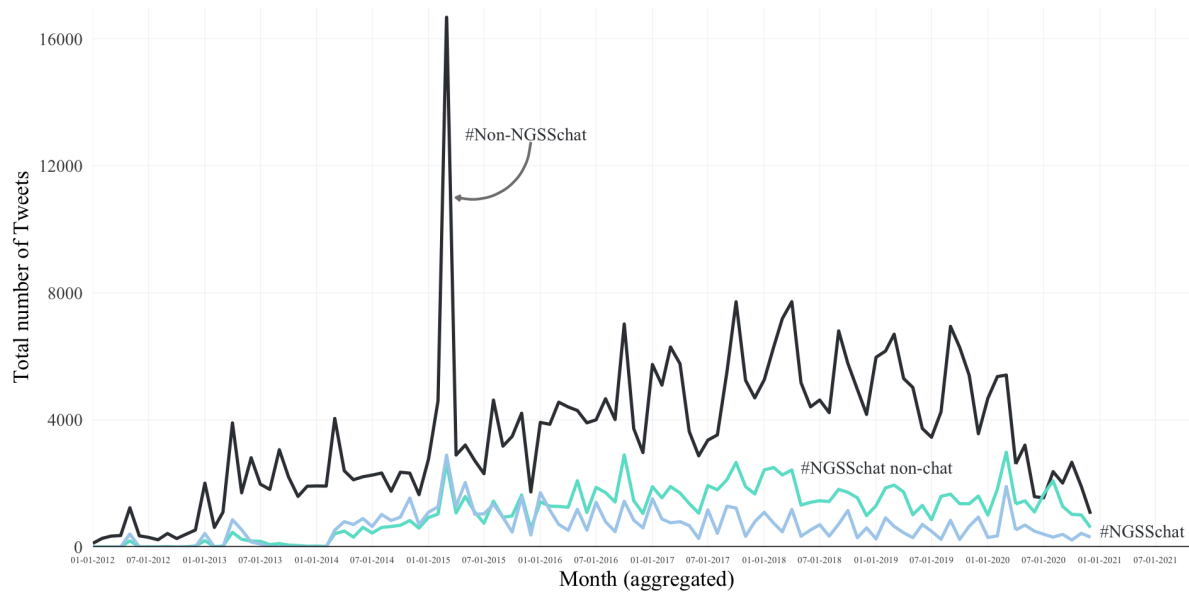
But first: an example
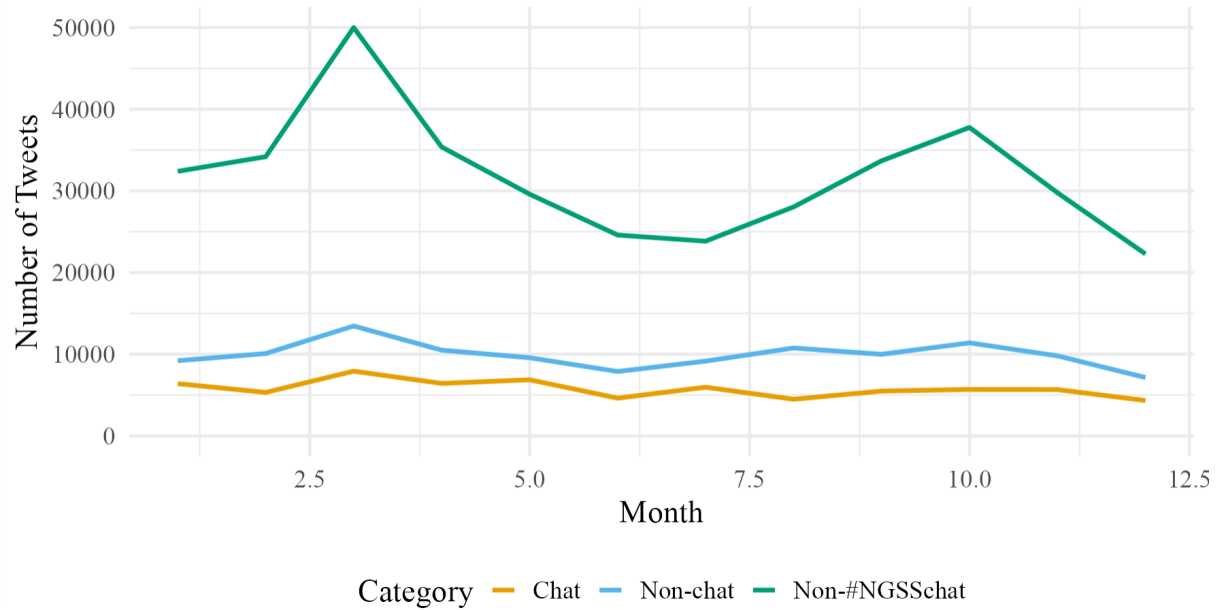
But first: an example

But first: an example

But first: an example

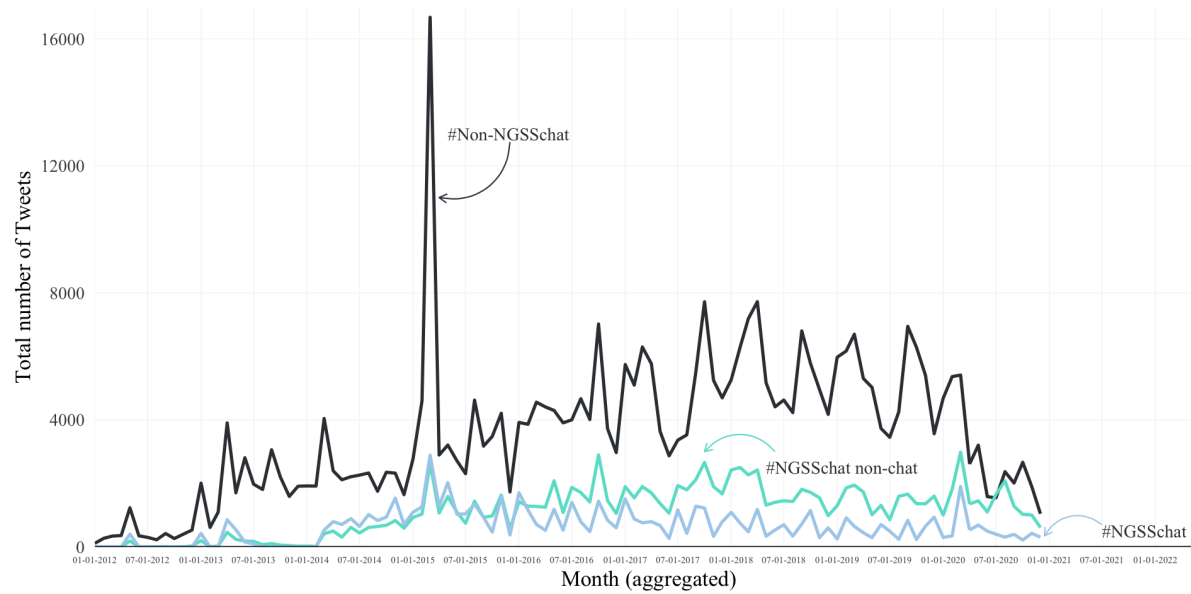# But first: an example

But first: an example

A line chart titled with axes "Total number of Tweets" (y-axis) and "Month (aggregated)" (x-axis). The y-axis ranges from 0 to 16000. The x-axis ranges from 01-01-2012 to 07-01-2021. Three lines are labeled: #Non-NGSSchat, #NGSSchat non-chat, and #NGSSchat.
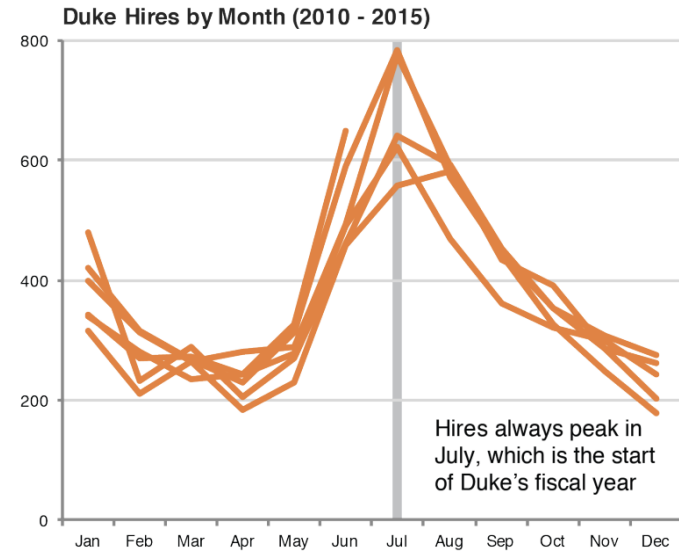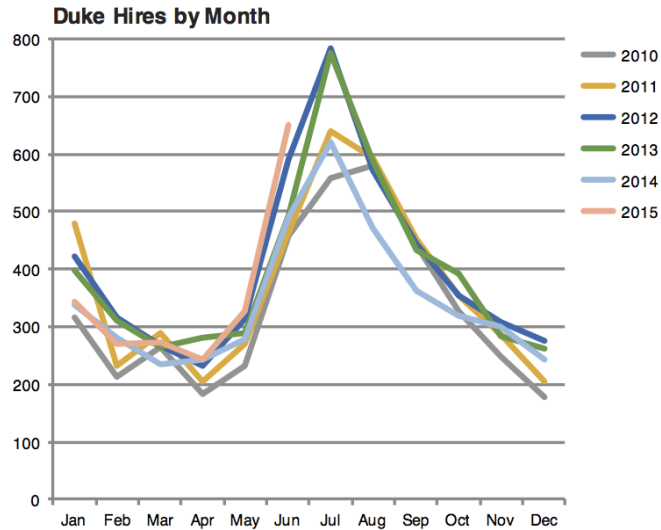
But first: an example

Two overarching goals of learning data viz in R

- Conceptual framework of visualization

- Grammar of graphics and different mappings of data onto visual elements

- Details of implementation

- How to build and refine plots layer by layer
- Eventually: Interactive data viz with ggviz and shiny

Homework highlights

Duke Hires by Month

2010, 2011, 2012, 2013, 2014, 2015

Duke Hires by Month (2010 - 2015)

Hires always peak in July, which is the start of Duke's fiscal year

Homework highlights

This year's Big Ten is one of the highest rated ever

Simple Rating System score by season for the six* major men's college basketball conferences, since 1949-50
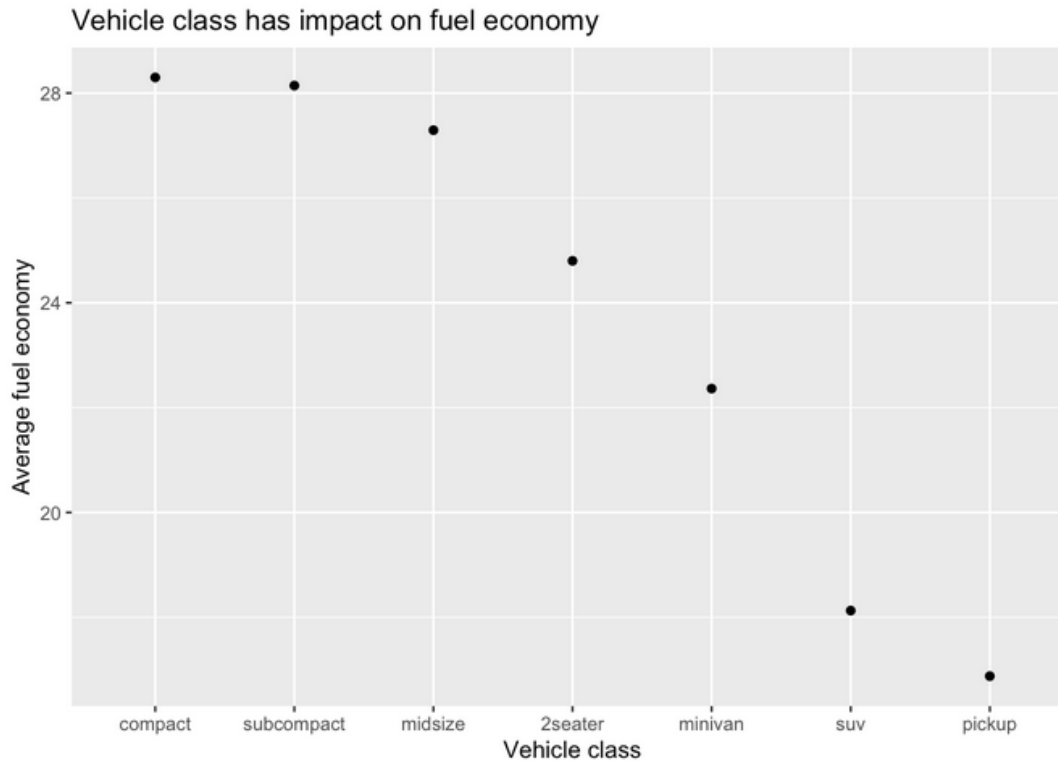
Big Ten

16.86

16.52
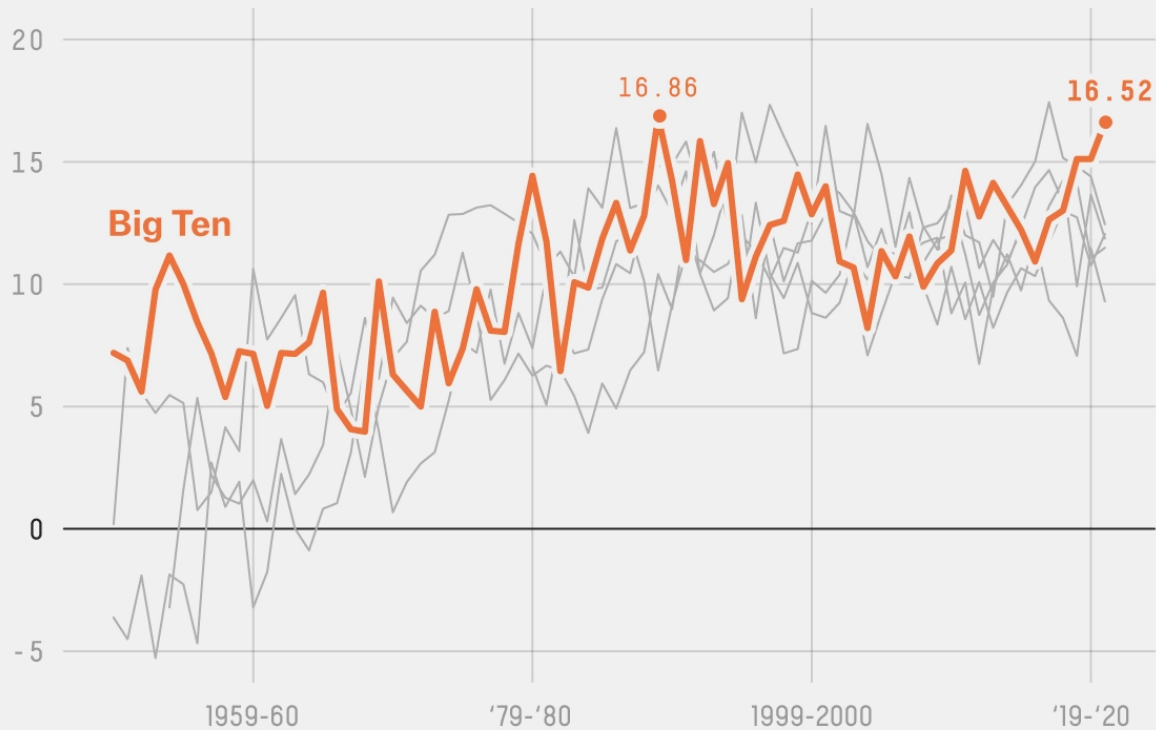
*The ACC, Big 12, Big East, Big Ten, Pac-12 and SEC.

FiveThirtyEight

SOURCE: SPORTS-REFERENCE.COM

Homework highlights

```
sum_df4 <- mpg %>% group_by(class) %>%
   summarize(MPG=mean(hwy))
   ggplot(sum_df4, aes(x = reorder(class, -MPG), y = MPG)) + geom_point() +
   labs(
     x="Vehicle class",
     y= "Average fuel economy",
     title = "Vehicle class has impact on fuel economy")
```


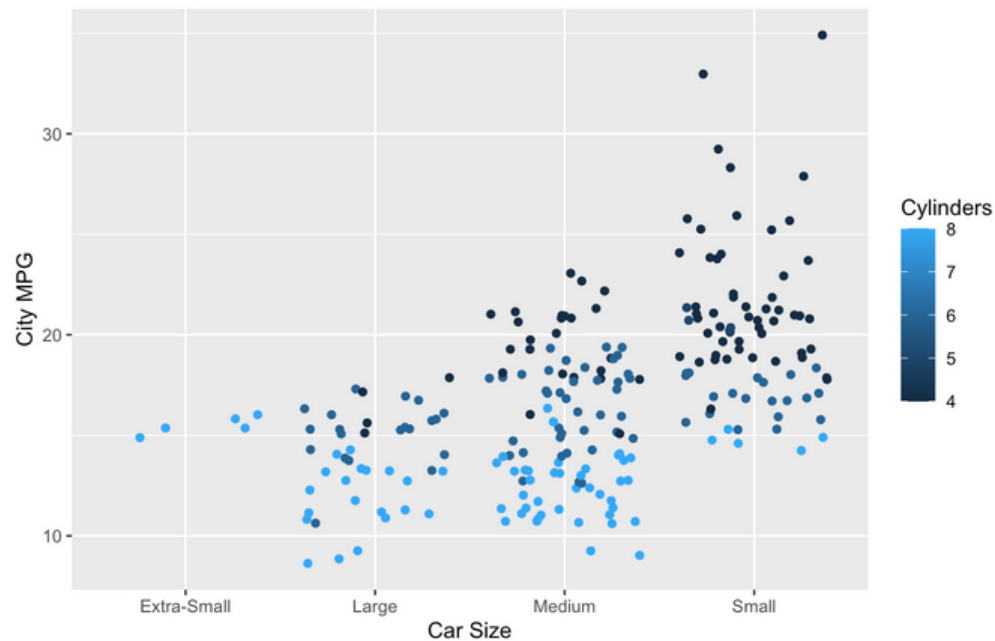Vehicle class has impact on fuel economy

Homework highlights

```
ggplot(mpg_cars, aes(class_group, cty)) + geom_jitter(aes(color = cyl)) + labs(
  title = "Size of car reveals little connection to fuel efficiency",
  subtitle = "Much overlap in size means cylinders are more impactful",
  x = "Car Size",
  y = "City MPG",
  color = "Cylinders")
```

Homework highlights

```
library(ggthemes)
ggplot(summ_mpg_2, aes(manufacturer,cfe, color="Combined"))+
  geom_point()+
  geom_point(aes(manufacturer,mean_hwy,color="Highway"))+
  geom_point(aes(manufacturer,mean_cty,color="City"))+
  labs(title = "Mean City, Highway, and Combined Fuel Efficiency by Manufacturer",x="Manufacturer", y="Miles per
Gallon")+
  labs(color='Fuel Economy') +
  theme_stata()+
  theme(axis.text.x = element_text(angle = 90))
```


Mean City, Highway, and Combined Fuel Efficiency by Manufacturer
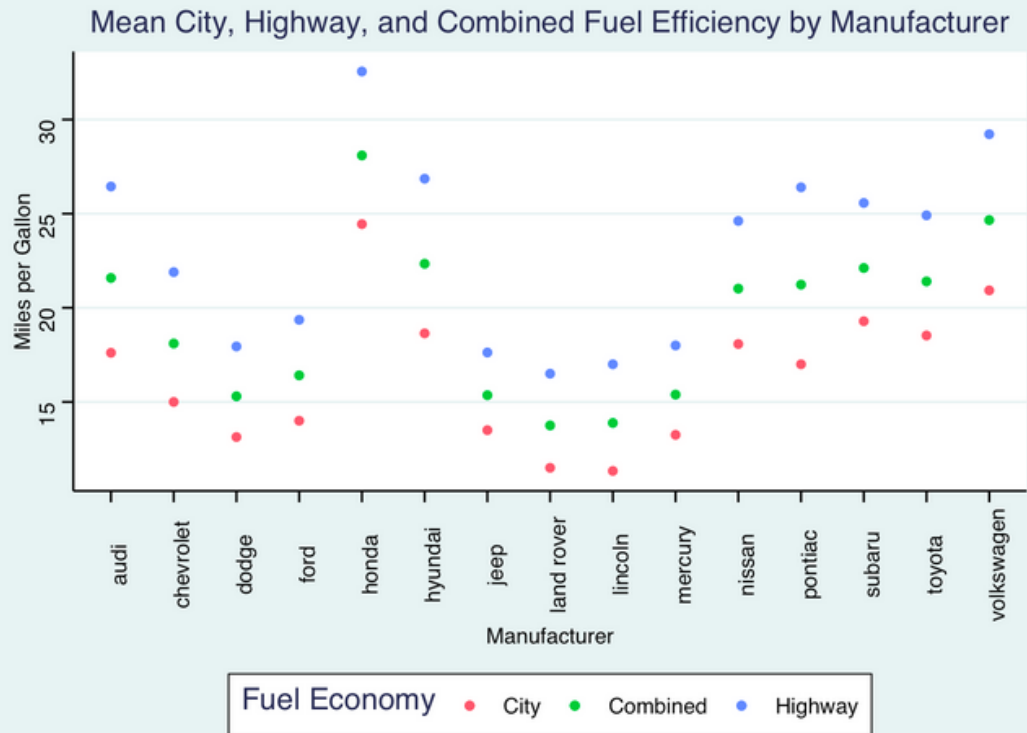
This week's topics

Overview

A. Using color

B. Grouping and stacking bar charts

C. Faceting plots

A. Using color

One high–level distinction to consider:

- Assigning a color to a geom
- Mapping a variable (with `aes()`) a color to a geom

A. Using color

Assigning a color to a geom

```
colors()
```
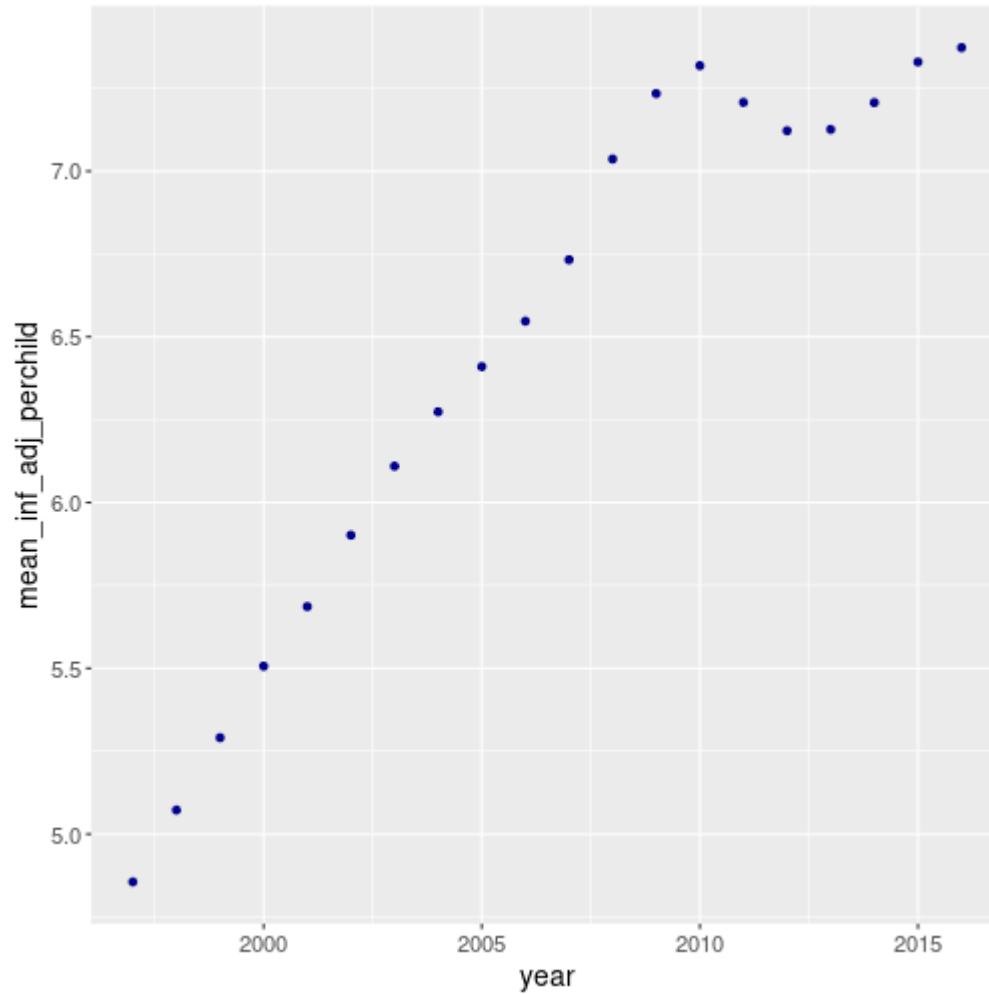
```
tidykids <- read_csv(here("data", "tidykids.csv"))

tidykids %>%
  filter(variable == "PK12ed") %>%
  group_by(year) %>%
  summarize(mean_inf_adj_perchild = mean(inf_adj_perchild)) %>% # b/c of multiple states
  ggplot(aes(x = year, y = mean_inf_adj_perchild)) +
  geom_point(color = "darkblue") +
  theme(text = element_text(size = 14))
```

A. Using color
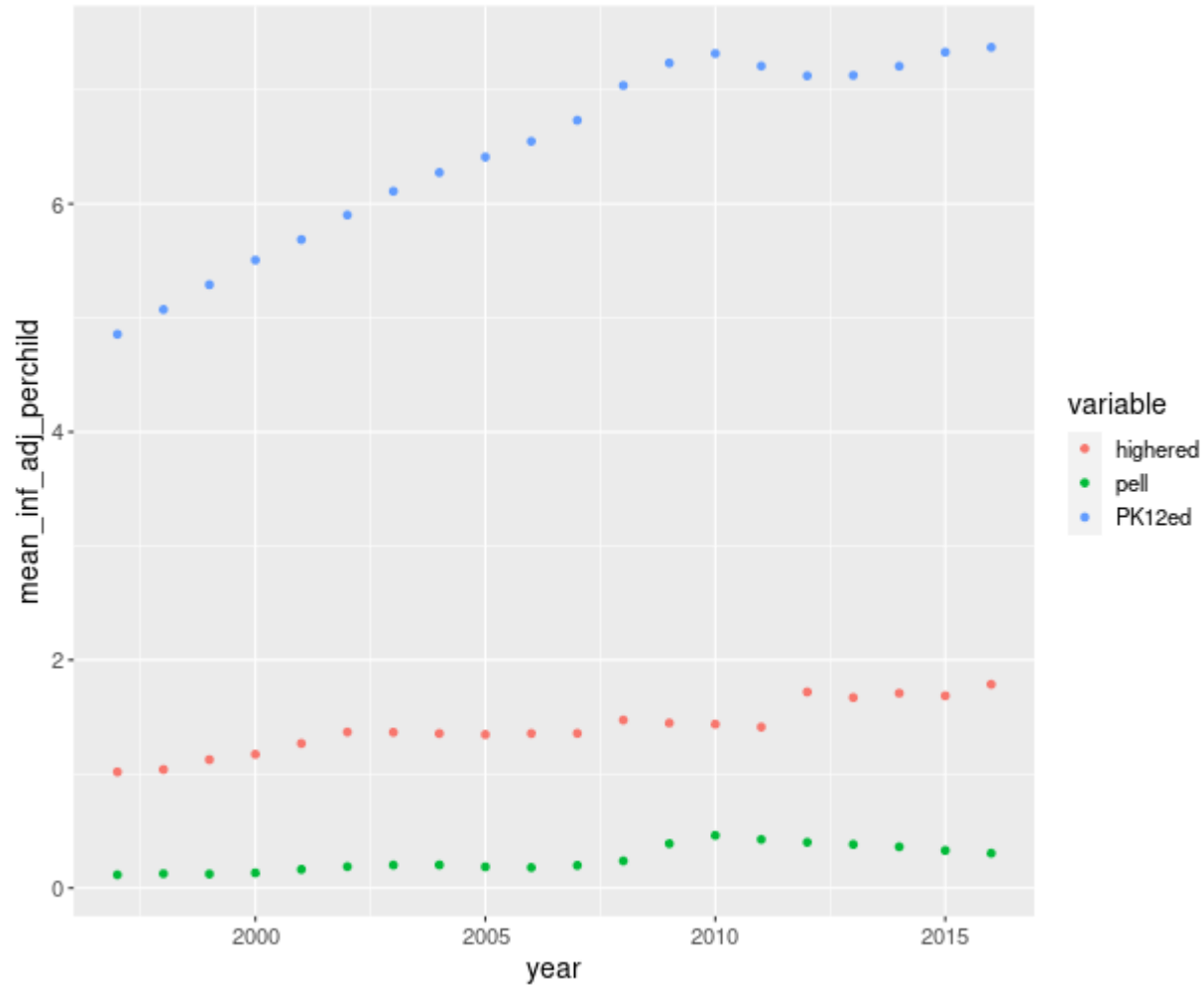
Assigning a color to a geom

A. Using color

Mapping a color to a geom

```r
tidykids %>%
  filter(variable %in% c("PK12ed", "highered", "pell")) %>%
  group_by(variable, year) %>%
  summarize(mean_inf_adj_perchild = mean(inf_adj_perchild)) %>% # b/c of multiple states
  ggplot(aes(x = year, y = mean_inf_adj_perchild, color = variable)) +
  geom_point() +
  theme(text = element_text(size = 14))
```

A. Using color

Mapping a color to a geom

A. Using color

Use a scale function to modify a scale

- scale_color_*
  - scale_color_discrete
  - scale_color_continous
  - scale_color_brewer
  - scale_color_manual()
- scale_fill_*
  - scale_fill_discrete
  - scale_fill_continous
  - scale_fill_brewer
  - scale_color_manual()
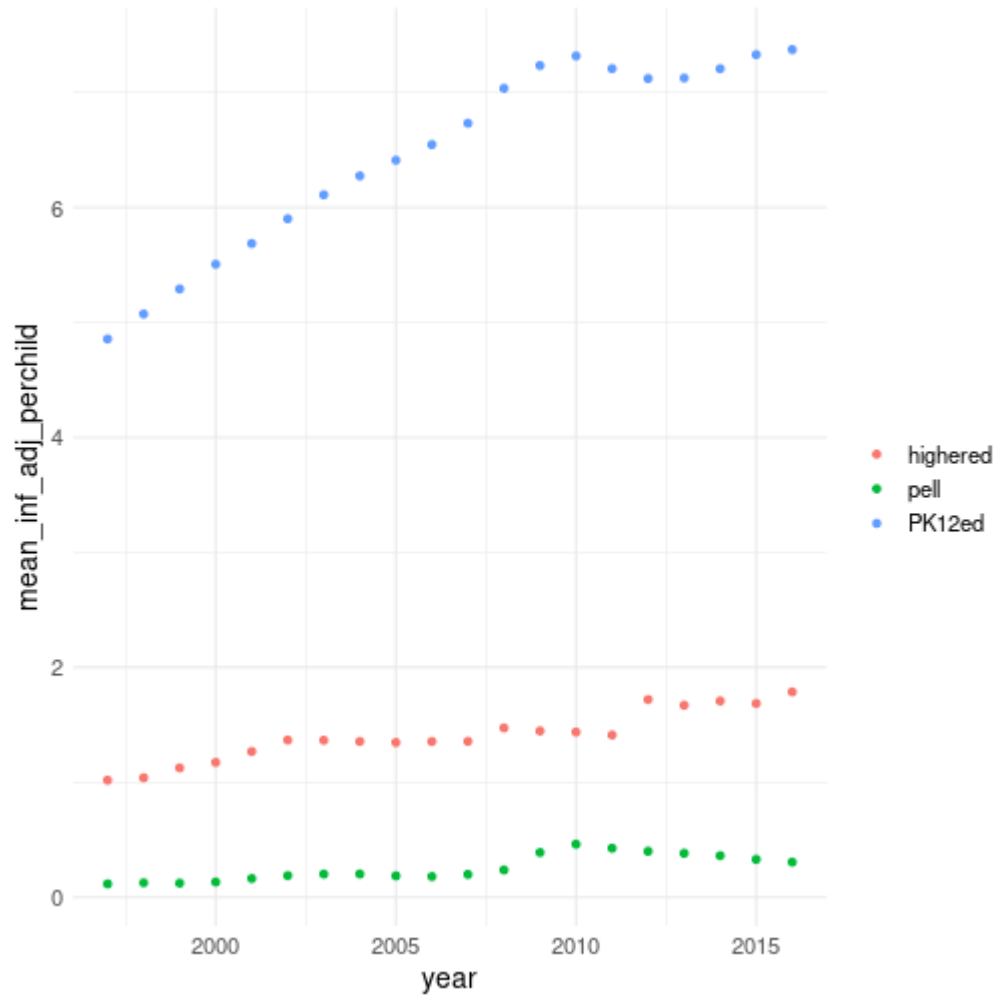
A. Using color

Removing the title from a scale

```
p <- tidykids %>%
  filter(variable %in% c("PK12ed", "highered", "pell")) %>%
  group_by(variable, year) %>%
  summarize(mean_inf_adj_perchild = mean(inf_adj_perchild)) %>% # b/c of multiple states
  ggplot(aes(x = year, y = mean_inf_adj_perchild, color = variable)) +
  geom_point() +
  theme(text = element_text(size = 14))

p +
  scale_color_discrete("")
```

See the result on the next slide.
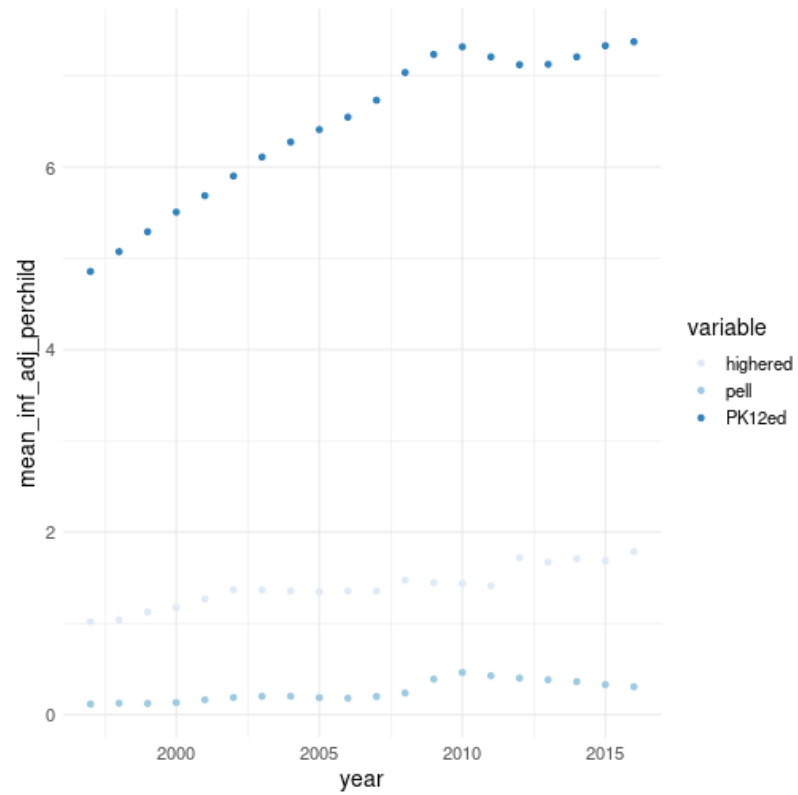
A. Using color

Removing the title from a scale

A. Using color

Changing the colors of a scale using a "color brewer" scale:

```
p +
  scale_color_brewer()
```

A. Using color

Changing the colors of a scale using a "color brewer" scale:

The "color brewer" functions support three kinds of scales:

- sequential (`type = "seq"`) for ordered data
- divergent (`type = "div"`) for data with a natural midpoint and two extremes
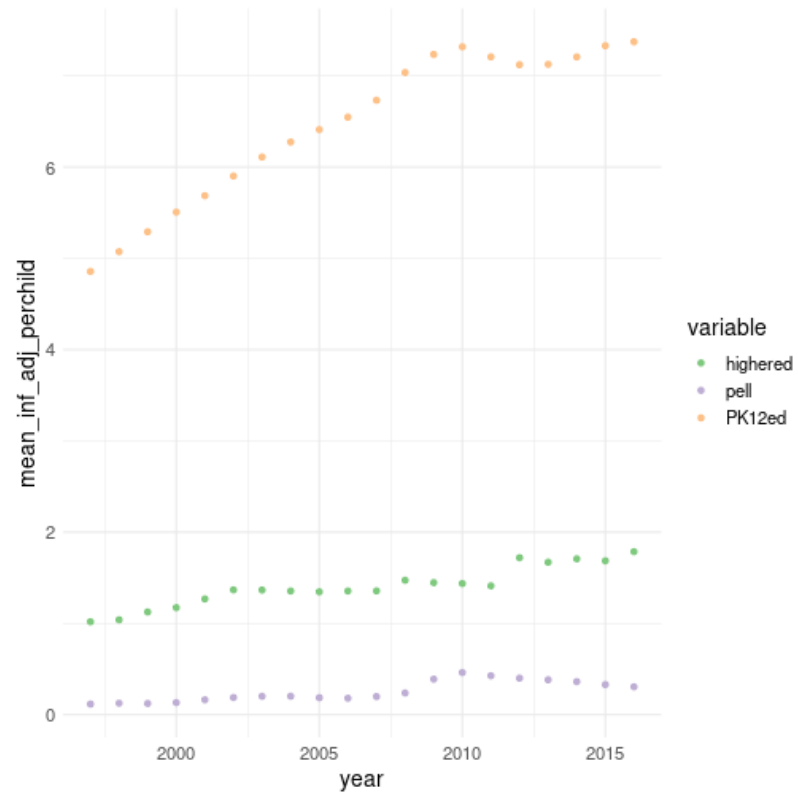- qualitative (`type = "qual"`) for categorical/qualitative data

https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3 https://www.r–graph–gallery.com/38–rcolorbrewers–palettes.html

What kind of palette would be best for our last plot?

A. Using color

Changing the colors of a scale using a "color brewer" scale:
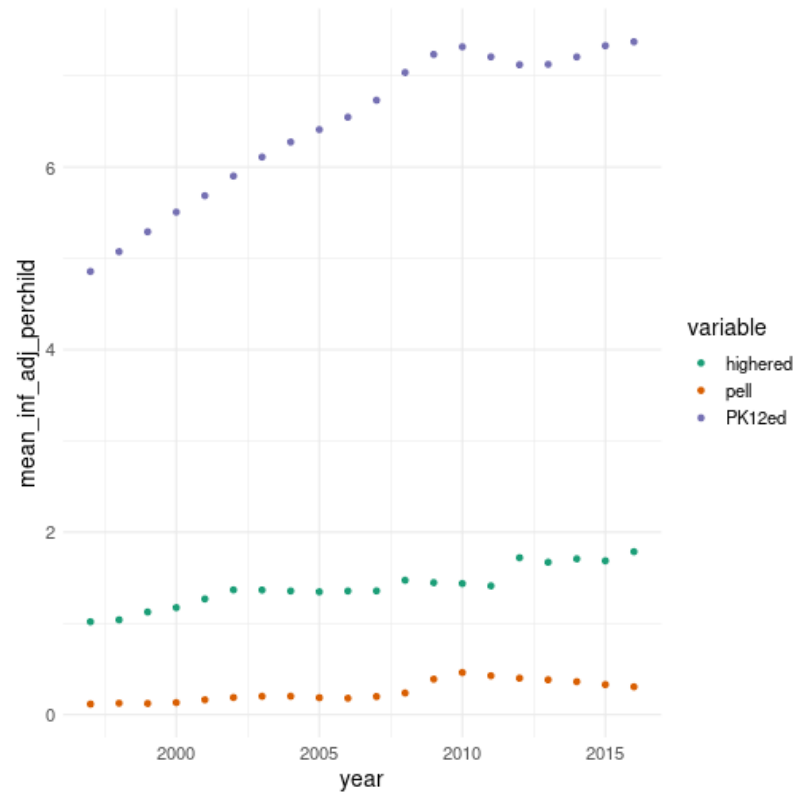
```
p +
  scale_color_brewer(type = "qual")
```

A. Using color

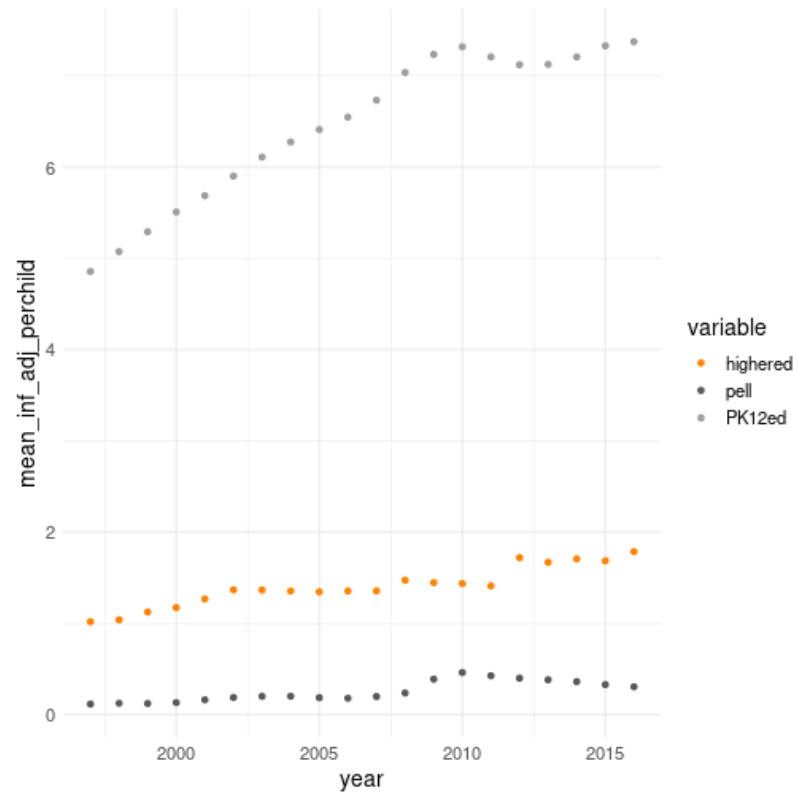**Changing the colors of a scale using a "color brewer" scale:**

```
p +
  scale_color_brewer(type = "qual", palette = 2)
```

A. Using color

https://brand.utk.edu/standards/colors/

```
p +
  scale_color_manual(values = c("#FF8200", "#58595B", "8D2048"))
```

B. Grouping and stacking bar charts

Removing the title from a scale

```
p <- tidykids %>%
  filter(variable %in% c("PK12ed", "highered", "pell")) %>%
  group_by(variable, year) %>%
  summarize(mean_inf_adj_perchild = mean(inf_adj_perchild)) %>% # b/c of multiple states
  ggplot(aes(x = year, y = mean_inf_adj_perchild, color = variable)) +
  geom_col() +
  theme_minimal() +
  theme(text = element_text(size = 14))

p +
  scale_color_discrete("")
```
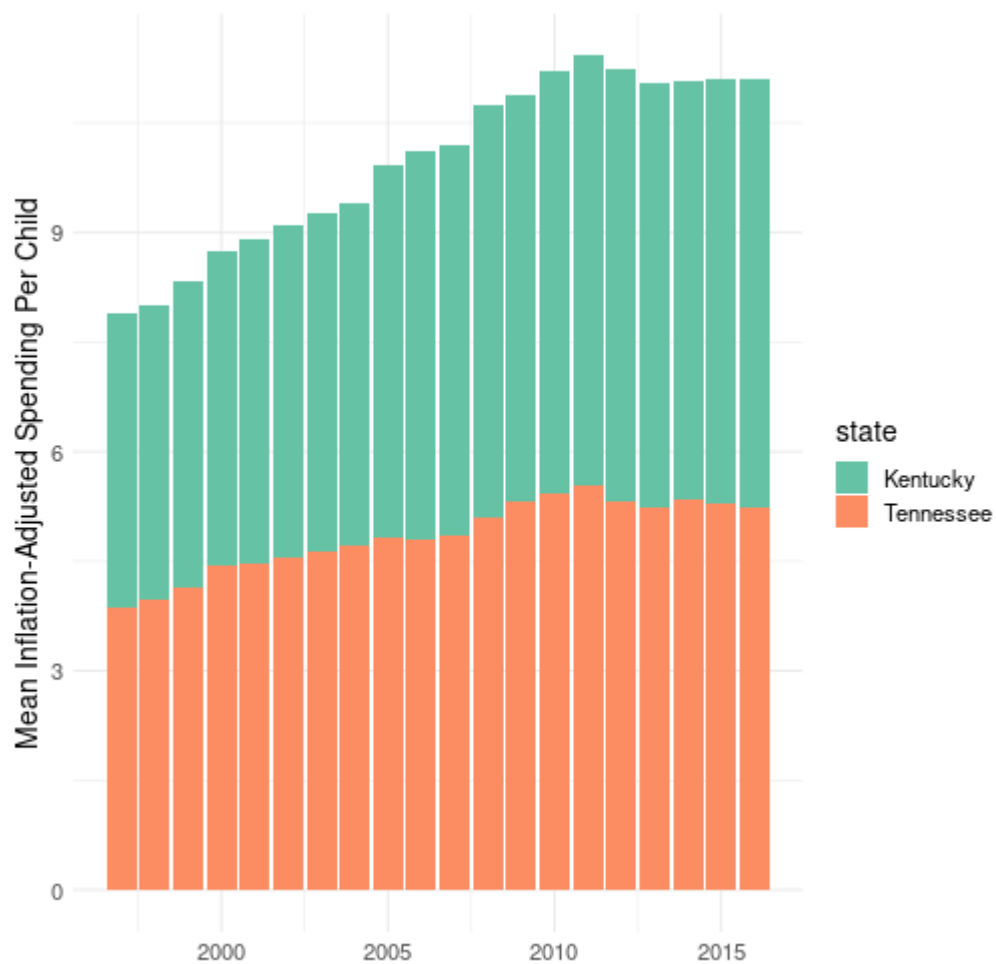
See the result on the next slide.

B. Grouping and stacking bar charts

What might we change?

```
tidykids %>%
  filter(variable == "PK12ed",
         state %in% c("Tennessee", "Kentucky")) %>%
  ggplot(aes(x = year, y = inf_adj_perchild, fill = state)) +
  geom_col() +
  theme_minimal() +
  theme(text = element_text(size = 14)) +
  scale_fill_brewer(type = "qual", palette = 7) +
  xlab("") +
  ylab("Mean Inflation-Adjusted Spending Per Child")
```

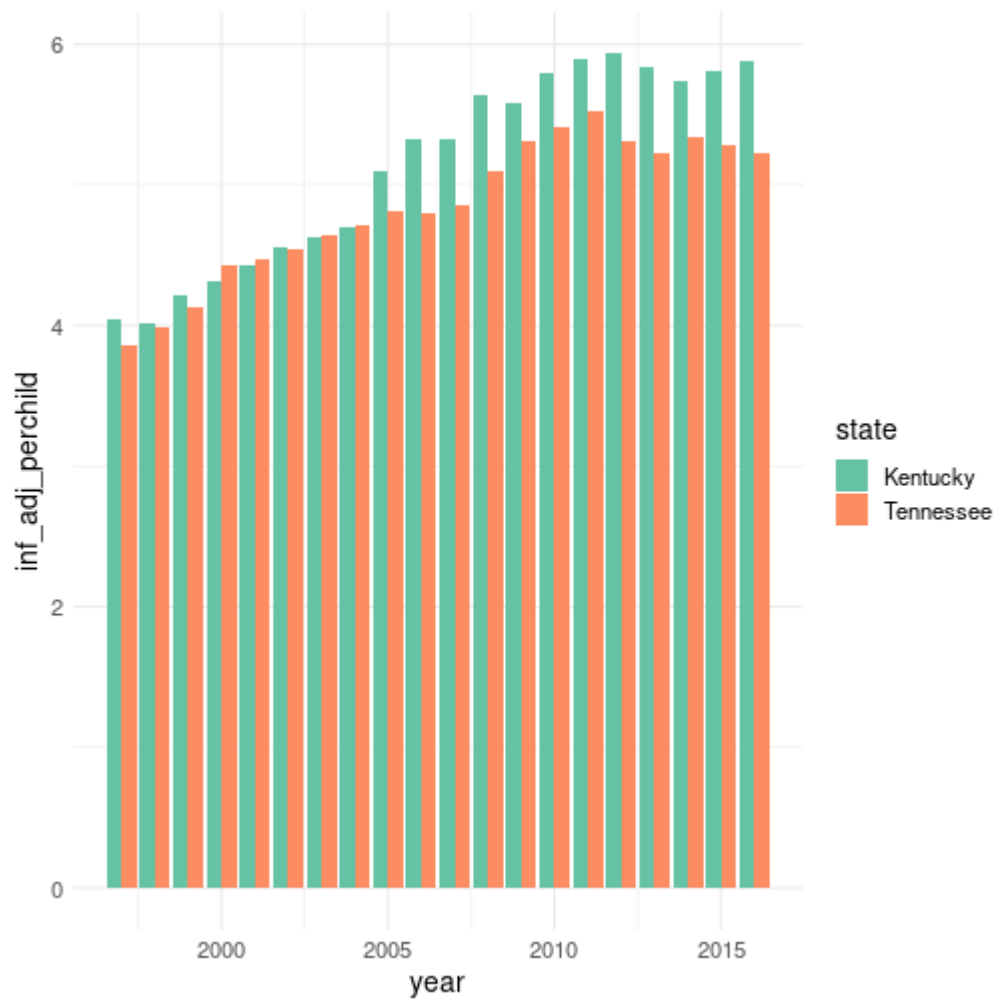# B. Grouping and stacking bar charts

B. Grouping and stacking bar charts

Using the `position = "dodge"` argument

```r
tidykids %>%
  filter(variable == "PK12ed",
         state %in% c("Tennessee", "Kentucky")) %>%
  ggplot(aes(x = year, y = inf_adj_perchild, fill = state)) +
  geom_col(position = "dodge") +
  theme_minimal() +
  theme(text = element_text(size = 14)) +
  scale_fill_brewer(type = "qual", palette = 7)
```

# B. Grouping and stacking bar charts

B. Grouping and stacking bar charts

A use for stacking

```
tidykids %>%
  filter(variable %in% c("PK12ed", "highered"),
         state %in% c("Tennessee")) %>%
  group_by(year) %>%
  mutate(prop_of_edu_spending = inf_adj_perchild / sum(inf_adj_perchild)) %>%
  ggplot(aes(x = year, y = prop_of_edu_spending, fill = variable)) +
  geom_col(position = "stack") +
  theme_minimal() +
  theme(text = element_text(size = 14)) +
  scale_fill_brewer(type = "qual", palette = 7) +
  xlab("") +
  ylab("Mean Inflation-Adjusted Spending Per Child")
```

B. Grouping and stacking bar charts
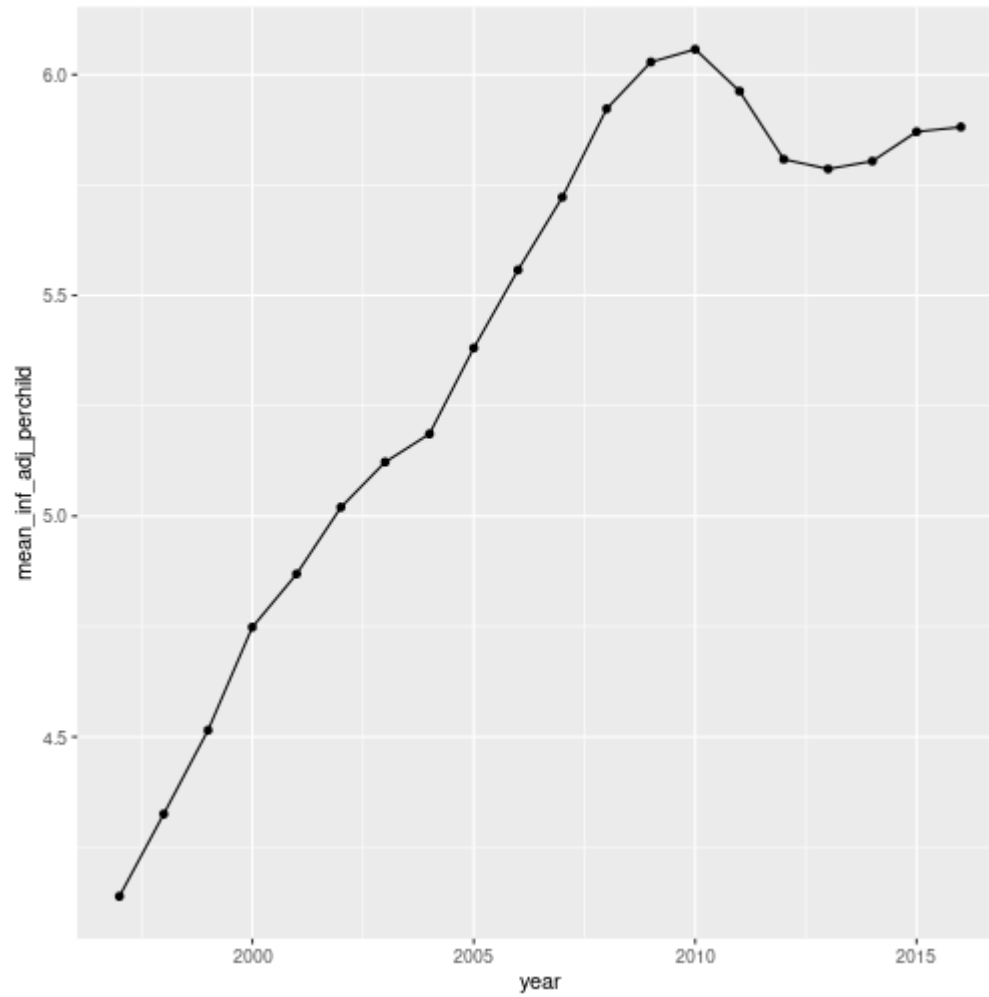
A use for stacking
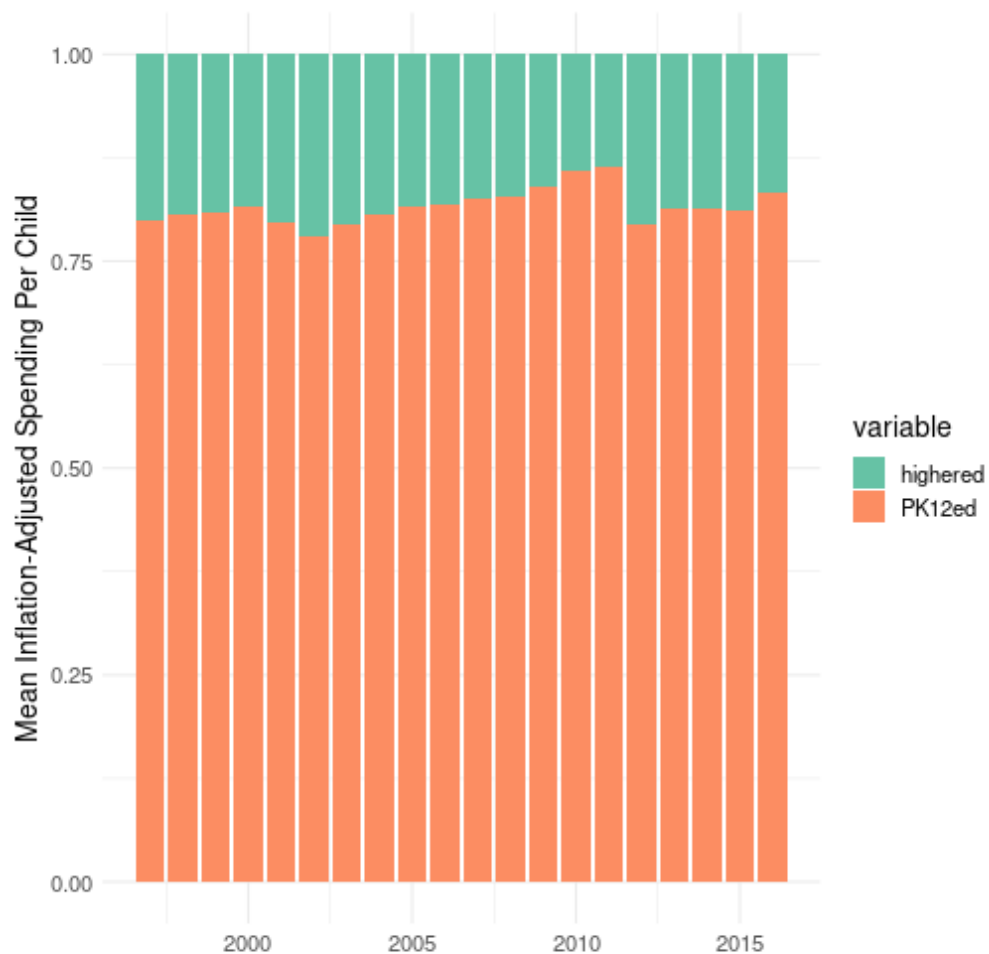
C. Faceting plots

Faceting plots with `facet_wrap()`

States combined

```
tidykids %>%
  filter(variable == "PK12ed",
         state %in% c("Tennessee", "Kentucky", "North Carolina", "Virginia", "Georgia", "Alabama",
  group_by(year) %>%
  summarize(mean_inf_adj_perchild = mean(inf_adj_perchild)) %>%
  ggplot(aes(x = year, y = mean_inf_adj_perchild)) +
  geom_point() +
  geom_line()
```
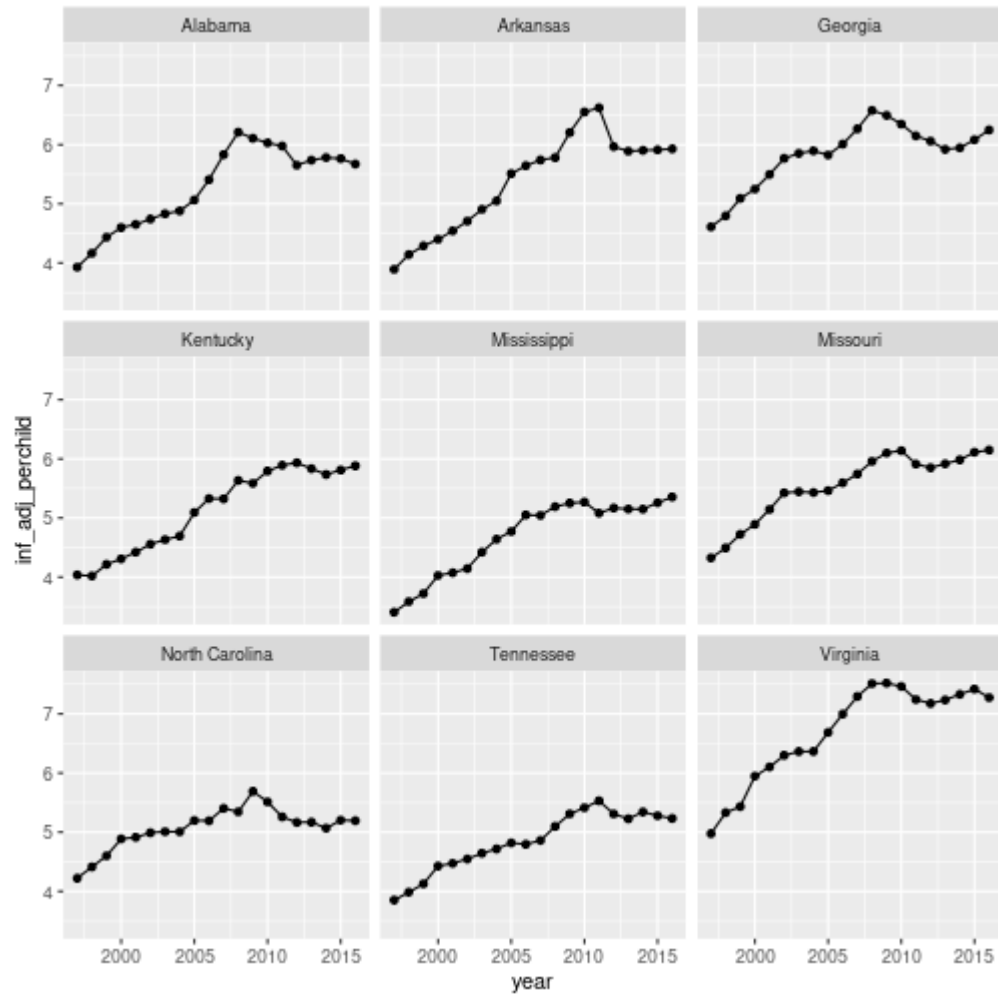
C. Faceting plots

C. Faceting plots

States faceted

```
tidykids %>%
  filter(variable == "PK12ed",
         state %in% c("Tennessee", "Kentucky", "North Carolina", "Virginia", "Georgia", "Alabama",
  ggplot(aes(x = year, y = inf_adj_perchild)) +
  geom_point() +
  geom_line() +
  facet_wrap(~state)
```
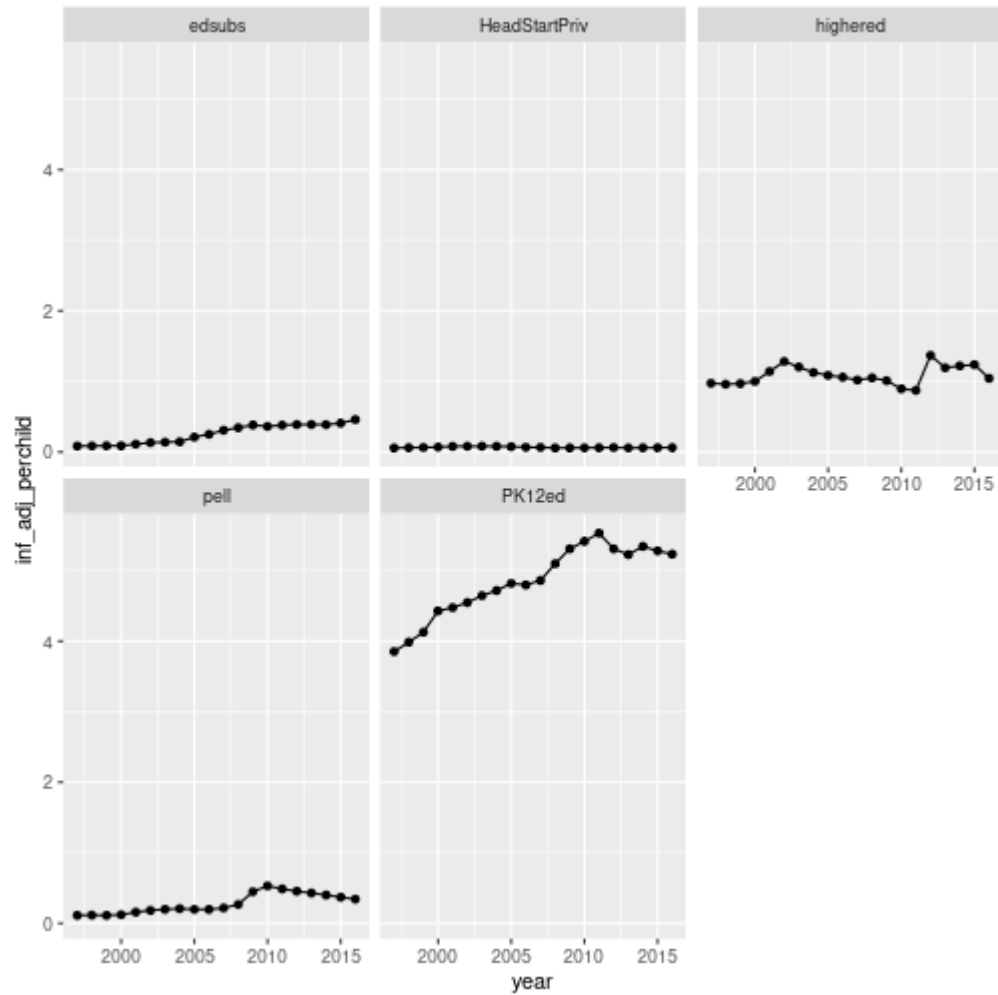
C. Faceting plots

C. Faceting plots

Variables of spending faceted

```
tidykids %>%
  filter(state == "Tennessee",
         variable %in% c("HeadStartPriv", "highered", "PK12ed", "pell", "edserv", "edsubs")) %>%
  group_by(year) %>%
  ggplot(aes(x = year, y = inf_adj_perchild)) +
  geom_point() +
  geom_line() +
  facet_wrap(~variable)
```

C. Faceting plots

Live coding

Let's head over to the following file for a demonstration:

`week-7-demo.R`
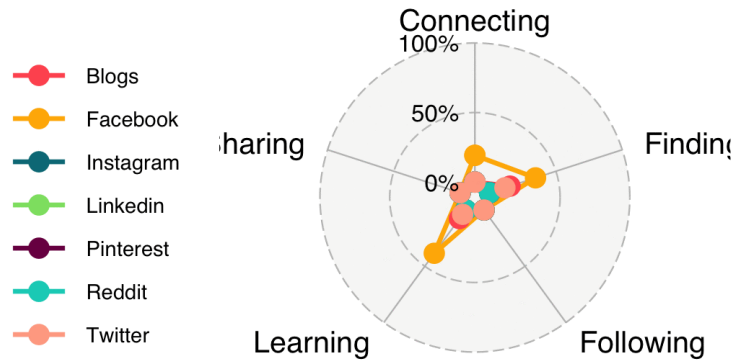
Why `.R` instead of `.Rmd`?

Data ethics statement

- Next, you will proviide feedback on your base group member's data ethics statement
- To do so, provide at least three stars (elements you liked) and three wishes (elements for which you suggest improvements)
- Do so by providing comments on your base group members' Google Doc
- When complete, message the person to whom you provided feedback in your base group's Slack channel

Where to next with respect to data viz?

- Interactive visualizations?

  - gganimate: https://gganimate.com/articles/gganimate.html
  - magick: https://cran.r-project.org/web/packages/magick/vignettes/intro.html
  - Shiny: https://shiny.rstudio.com

- Books:

  - https://socviz.co/
  - https://clauswilke.com/dataviz

- Resources

  - http://rweekly.org/
  - https://github.com/rfordatascience/tidytuesday; #tidytuesday on Twitter

Educator
1

This week

- Homework 7: Available tomorrow by noon; **Due by Thursday, 3/11**

  - Theming with color
  - Stacking and dodging
  - Faceting

- Readings

  - https://r4ds.had.co.nz/graphics-for-communication.html
  - https://clauswilke.com/dataviz/figure-titles-captions.html
  - https://clauswilke.com/dataviz/color-pitfalls.html

- Data ethics feedback; **Due by Thursday, 3/11**

- Final project

Random

- Webinar on open science: https://cos-io.zoom.us/webinar/register/1216134199427/WN_OpHYNc0PQhSkOVPSB3Qaag
- xaringan: https://bookdown.org/yihui/rmarkdown/xaringan.html

Wrapping up

In your base group's Slack channel:

- What is one thing you learned today?
- What is something you want to learn more about?
- Share your feelings in GIF form!

We really appreciate being able to see these reactions and get this feedback!