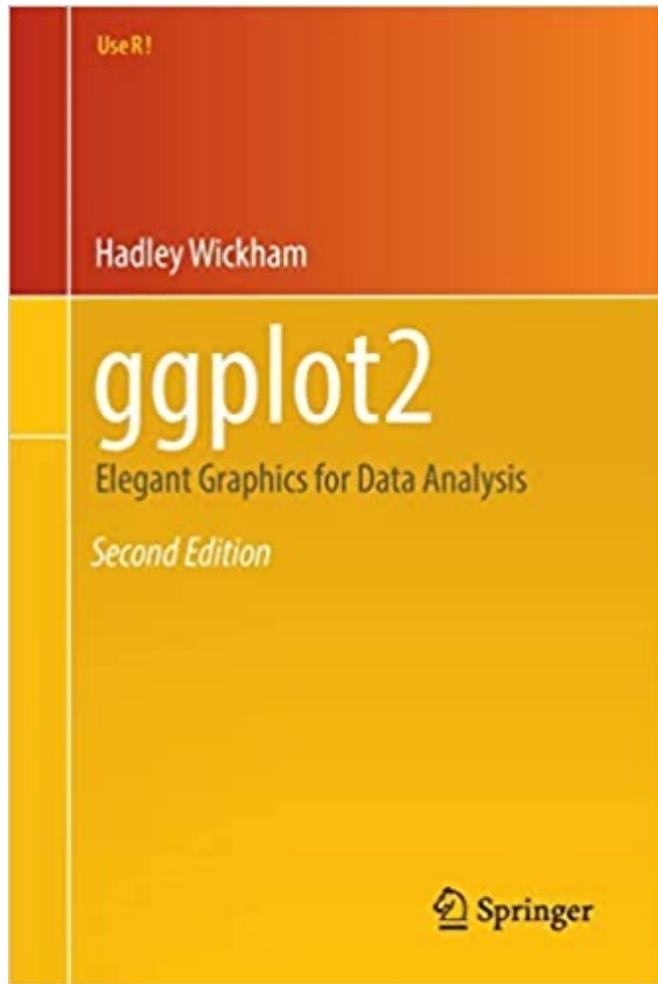
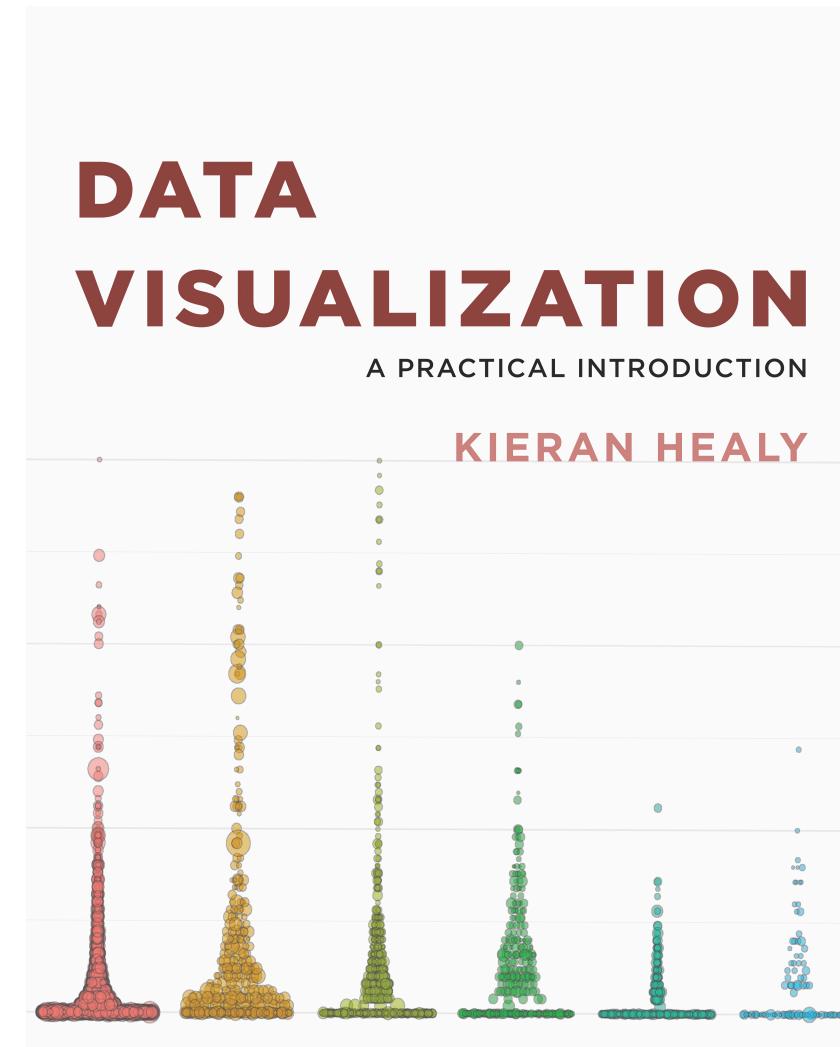


# Data visualization

February 23, 2022



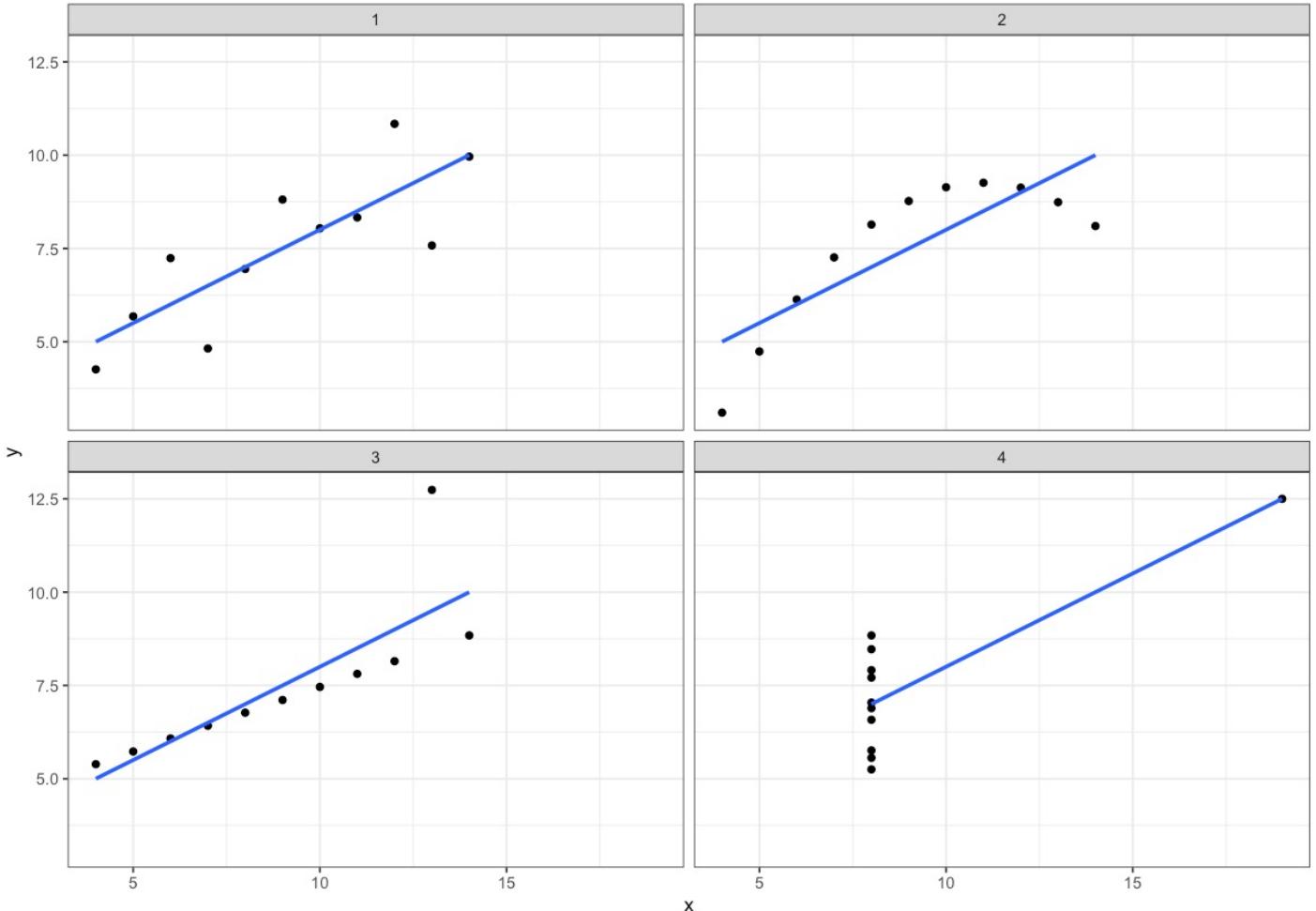
<https://ggplot2-book.org>

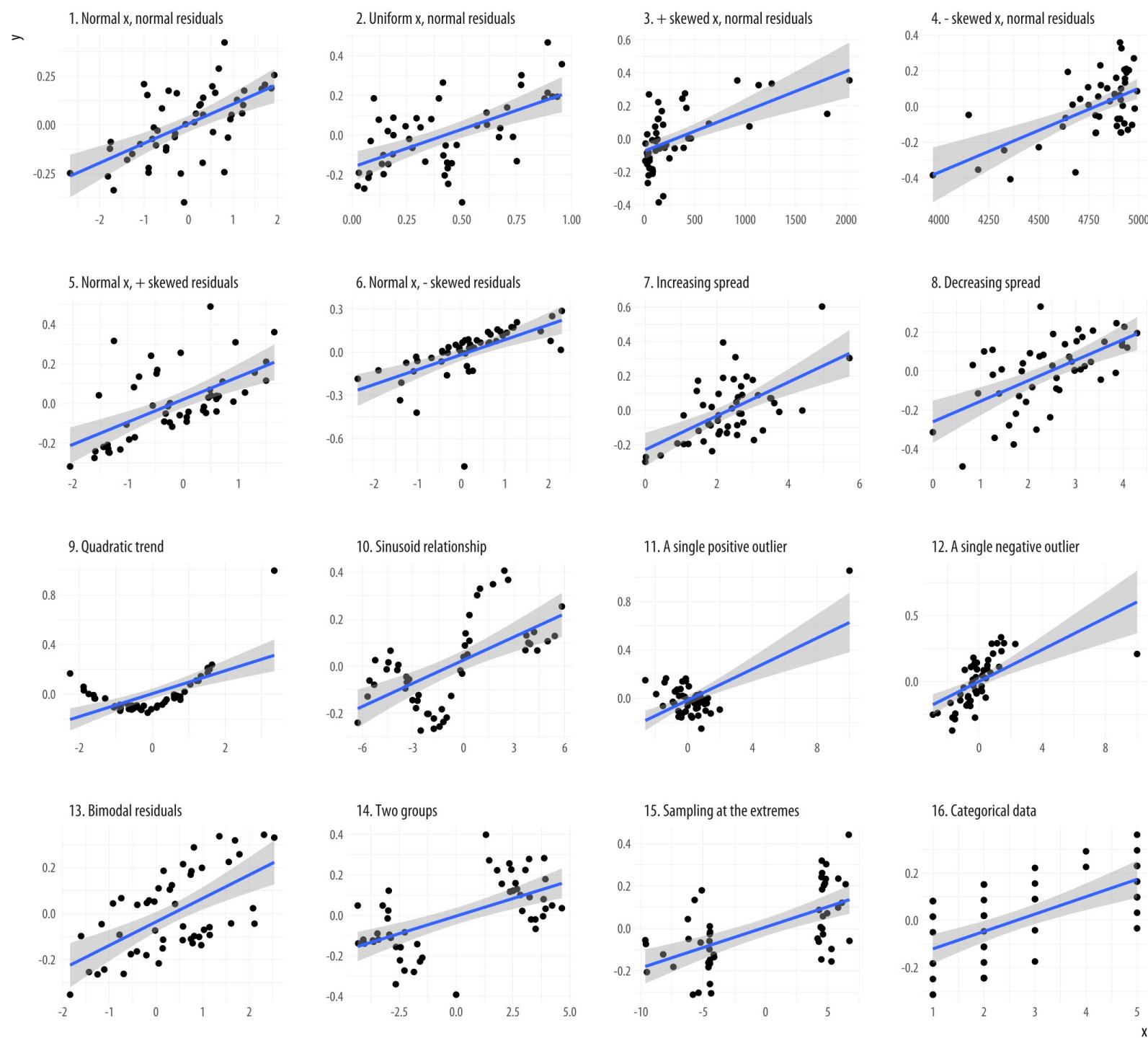


<https://socviz.co/lookatdata.html>

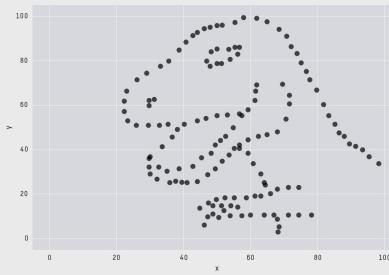
- graph data
- "Anscombe's quartet"

```
> ans_data %>% map_dfr(~ tidy(lm(y~x, data = .)), .id = "model")
# A tibble: 8 x 6
  model term      estimate std.error statistic p.value
  <chr> <chr>      <dbl>     <dbl>      <dbl>    <dbl>
1 1   (Intercept)  3.00      1.12      2.67  0.0257
2 1   x            0.500     0.118      4.24  0.00217
3 2   (Intercept)  3.00      1.13      2.67  0.0258
4 2   x            0.5       0.118      4.24  0.00218
5 3   (Intercept)  3.00      1.12      2.67  0.0256
6 3   x            0.500     0.118      4.24  0.00218
7 4   (Intercept)  3.00      1.12      2.67  0.0256
8 4   x            0.500     0.118      4.24  0.00216
```

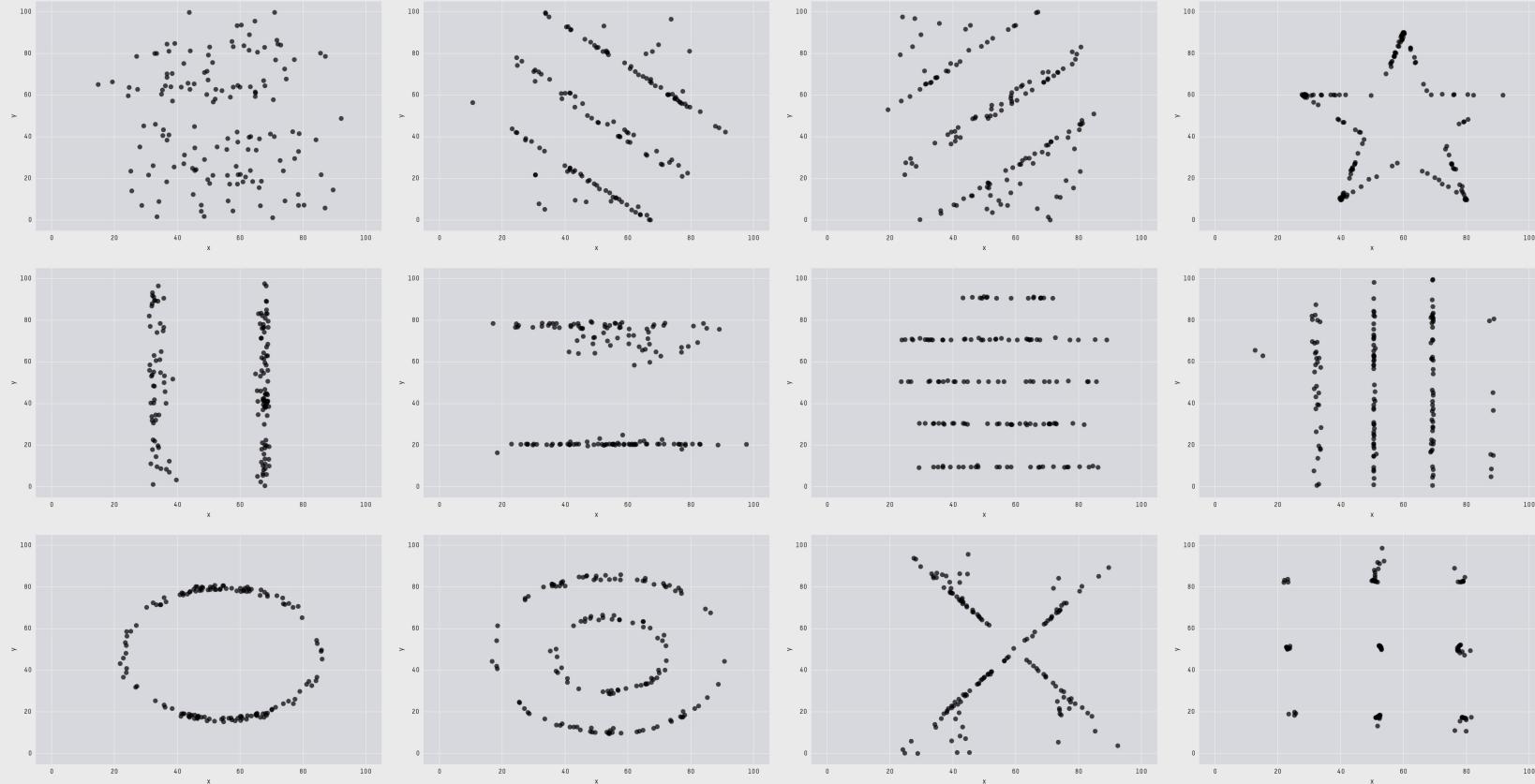




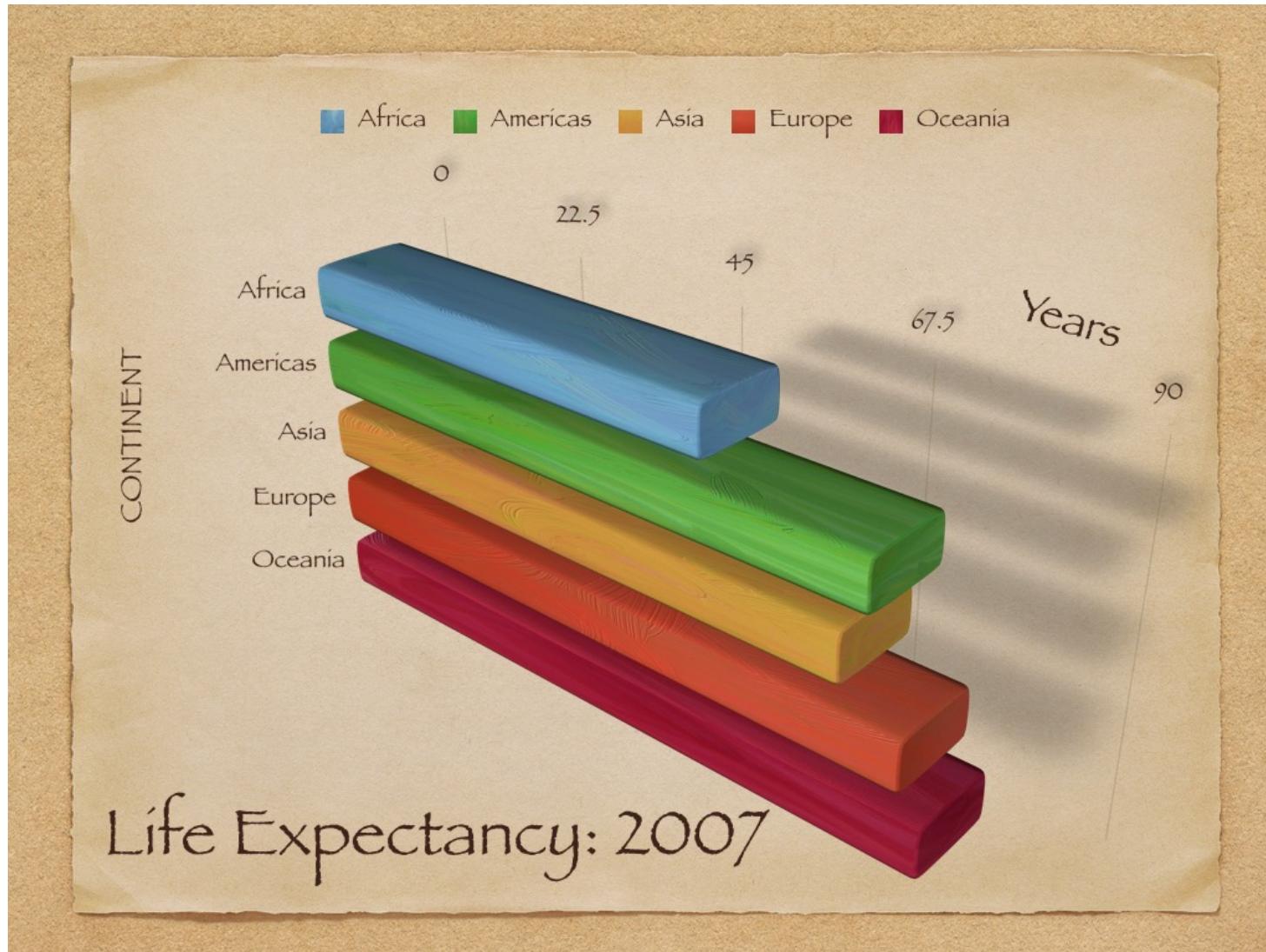
## The "Datasaurus Dozen"



X Mean: 54.26  
Y Mean: 47.83  
X SD : 16.76  
Y SD : 26.93  
Corr. : -0.06

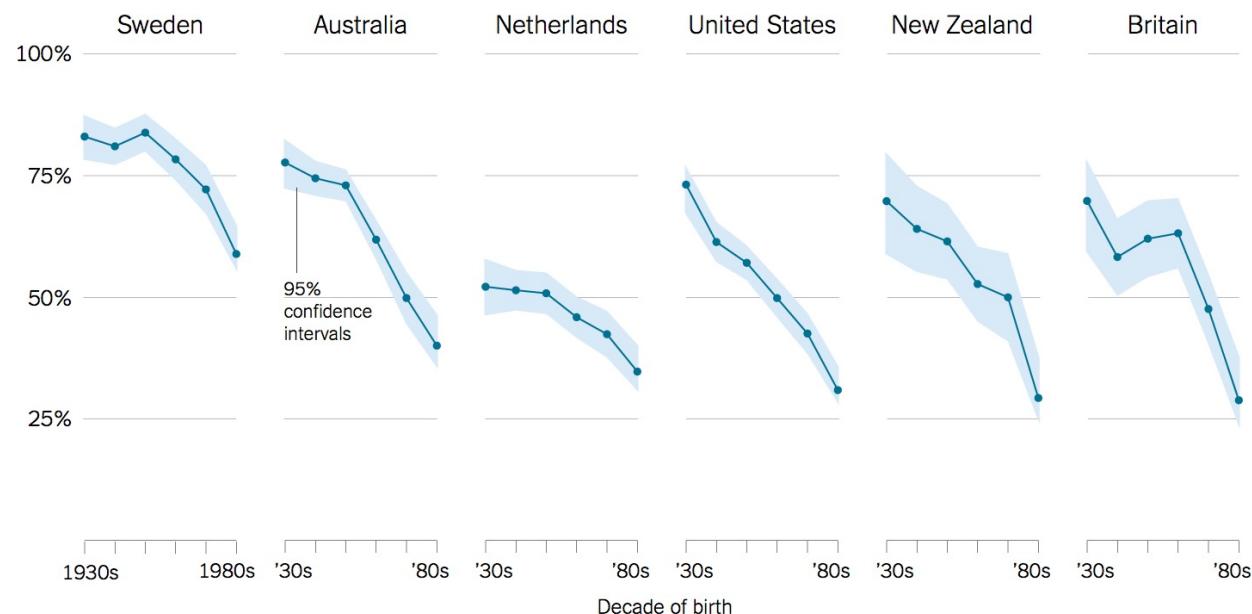


# Chart junk



# Misleading labels

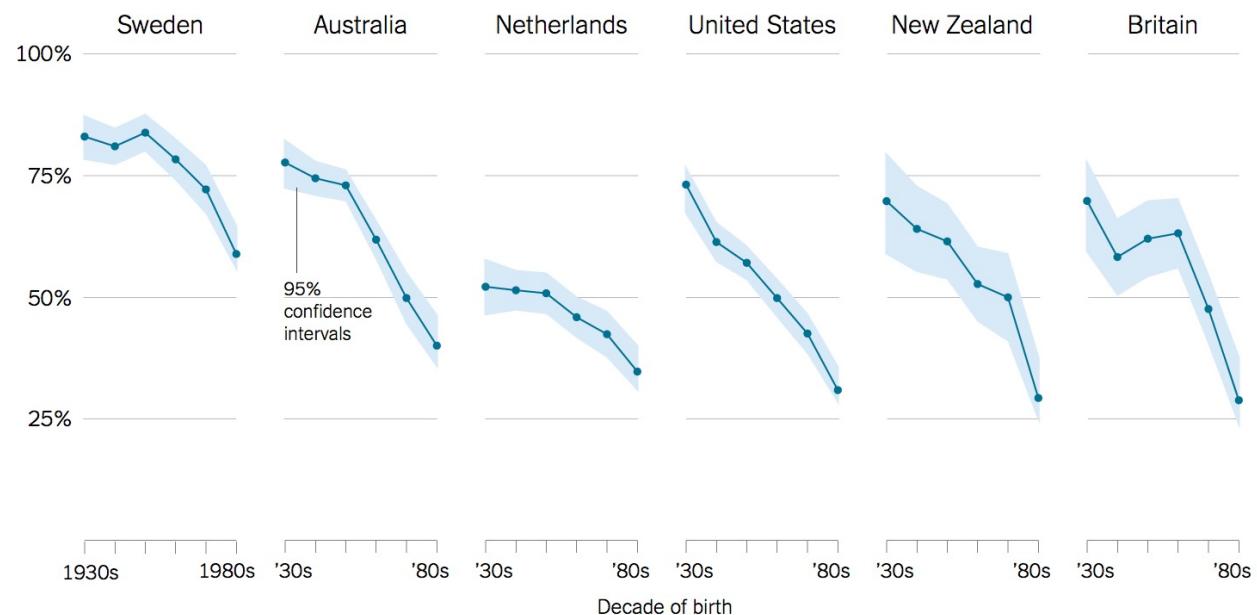
Percentage of people who say it is “essential” to live in a democracy



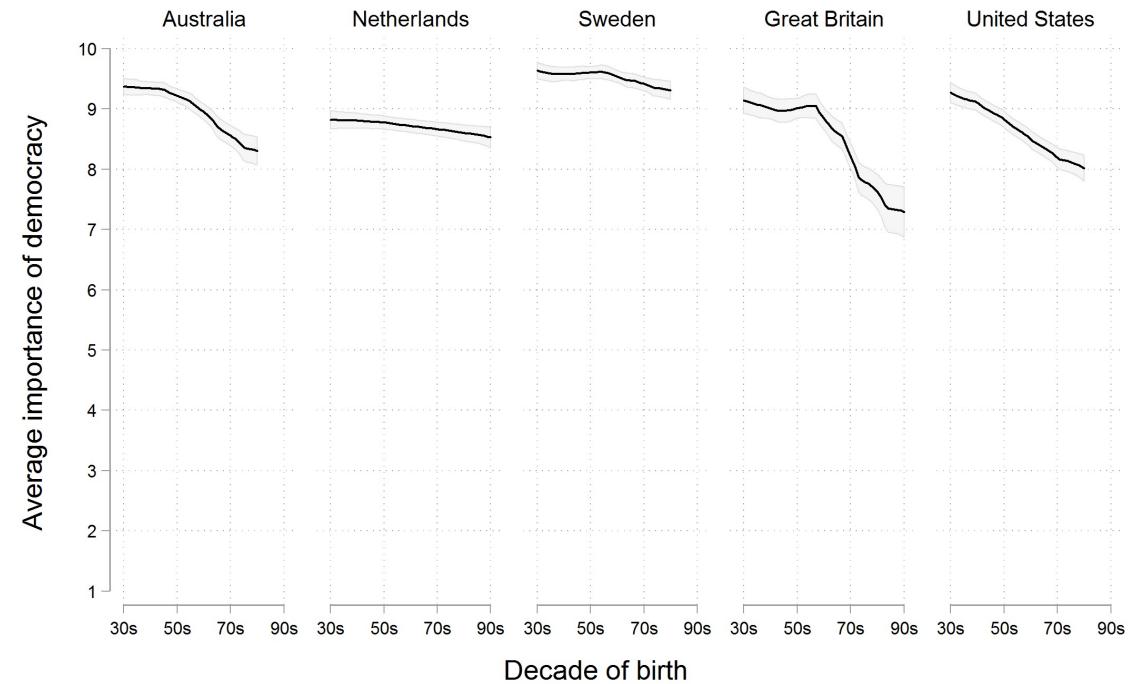
Source: Yascha Mounk and Roberto Stefan Foa, “The Signs of Democratic Deconsolidation,” Journal of Democracy | By The New York Times

# Misleading labels

Percentage of people who say it is “essential” to live in a democracy



Source: Yascha Mounk and Roberto Stefan Foa, “The Signs of Democratic Deconsolidation,” Journal of Democracy | By The New York Times

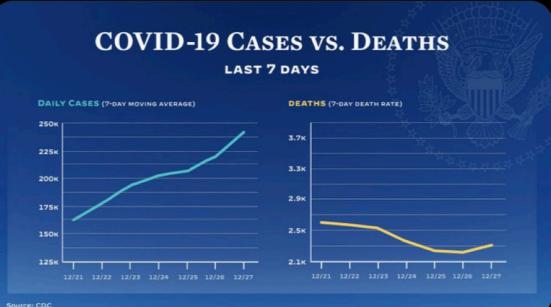


Graph by Erik Voeten, based on WVS 5

# Misleading labels

 **The White House**  @WhiteHouse · 16h ...  
Omicron cases are on the rise, but it's clear that vaccines and boosters are making a difference. Vaccines and boosters help prevent severe illness and death — if you haven't already, go get your vaccine and booster.

**COVID-19 CASES VS. DEATHS**  
**LAST 7 DAYS**



DAILY CASES (7-DAY MOVING AVERAGE)

DEATHS (7-DAY DEATH RATE)

Source: CDC

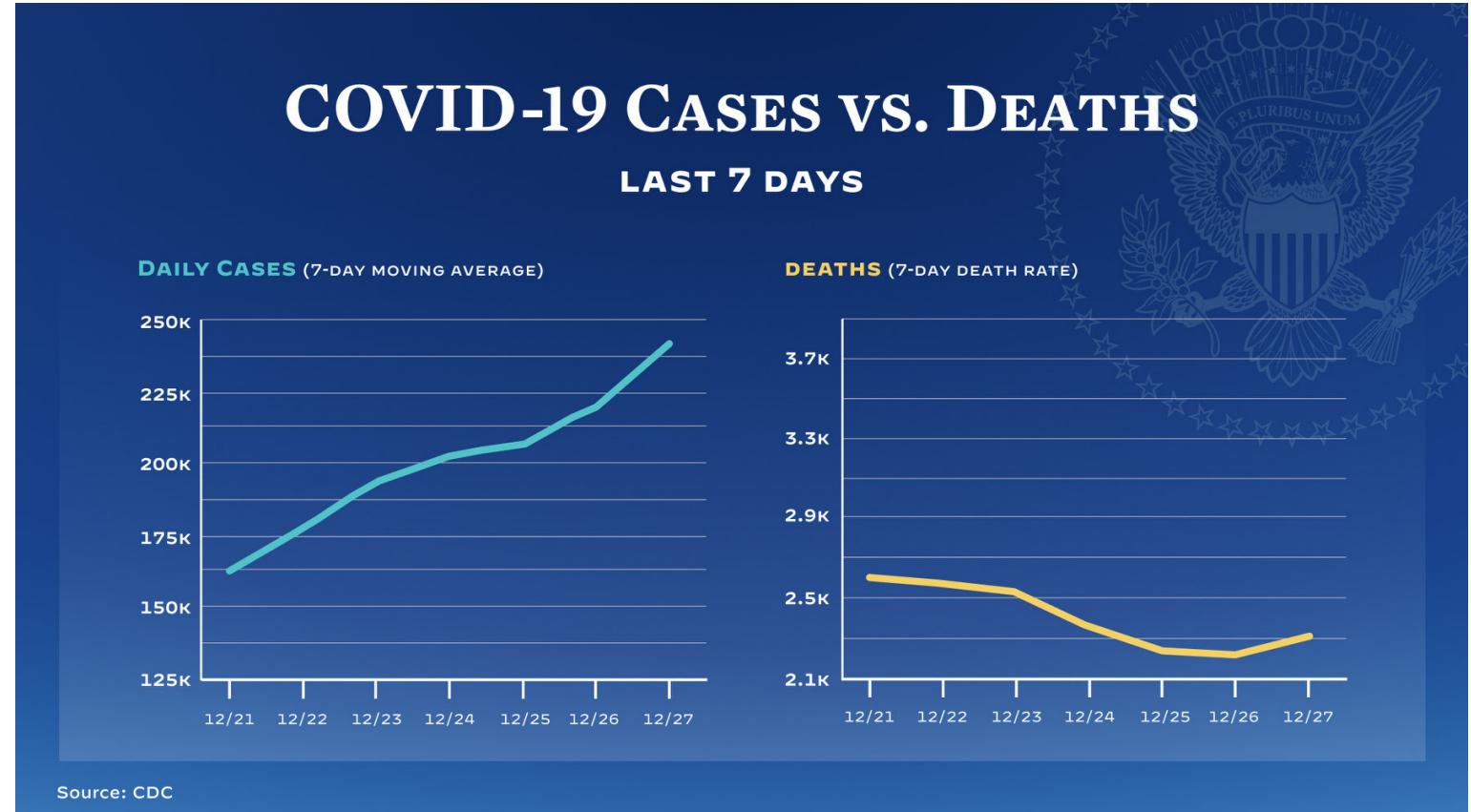
1,040 1,054 4,048

Joey Burnett  
@cjoseph\_burnett

Replying to @WhiteHouse

Another graph for your stats class,  
@markgbaxter?

12:30 PM · 12/30/21 · Twitter for iPhone

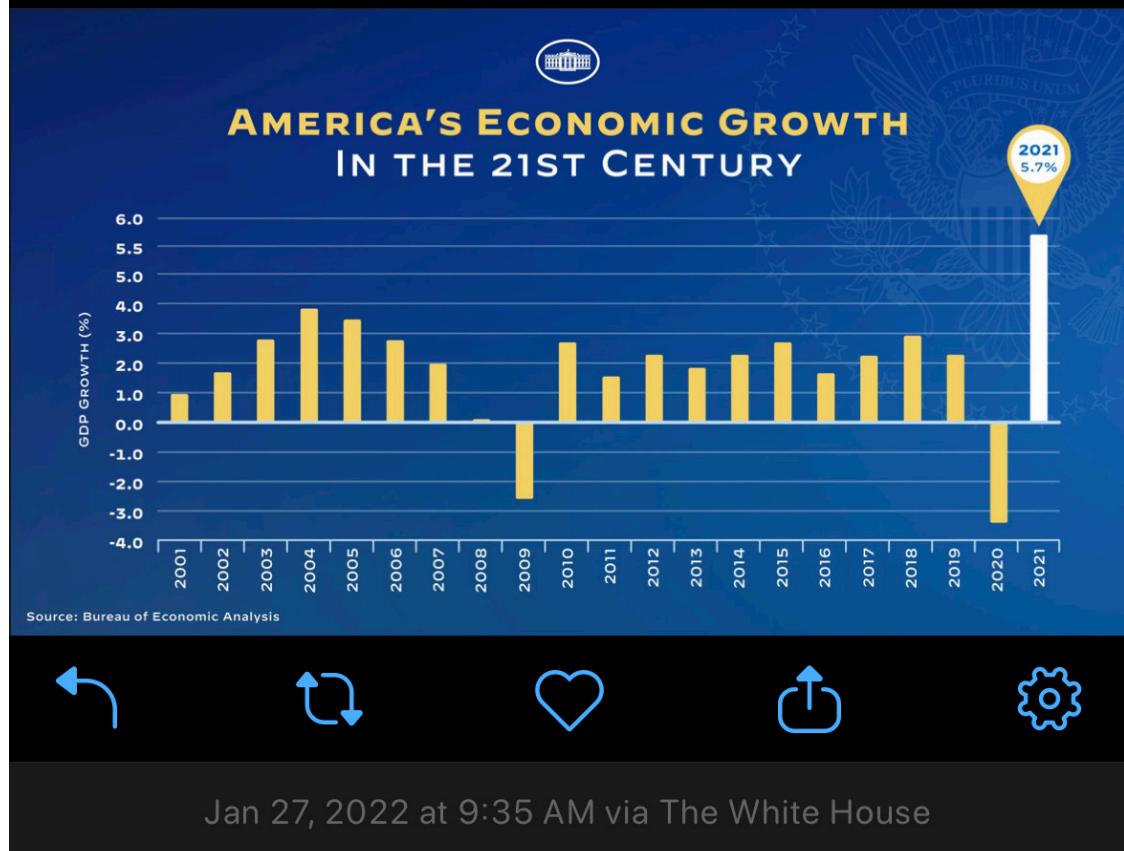




# The White House

@WhiteHouse

We just learned that President Biden's first year in office was the strongest year for economic growth since 1984.



**The White House** @WhiteHouse  
We just learned that President Biden's first year in office was the strongest year for economic growth since 1984.

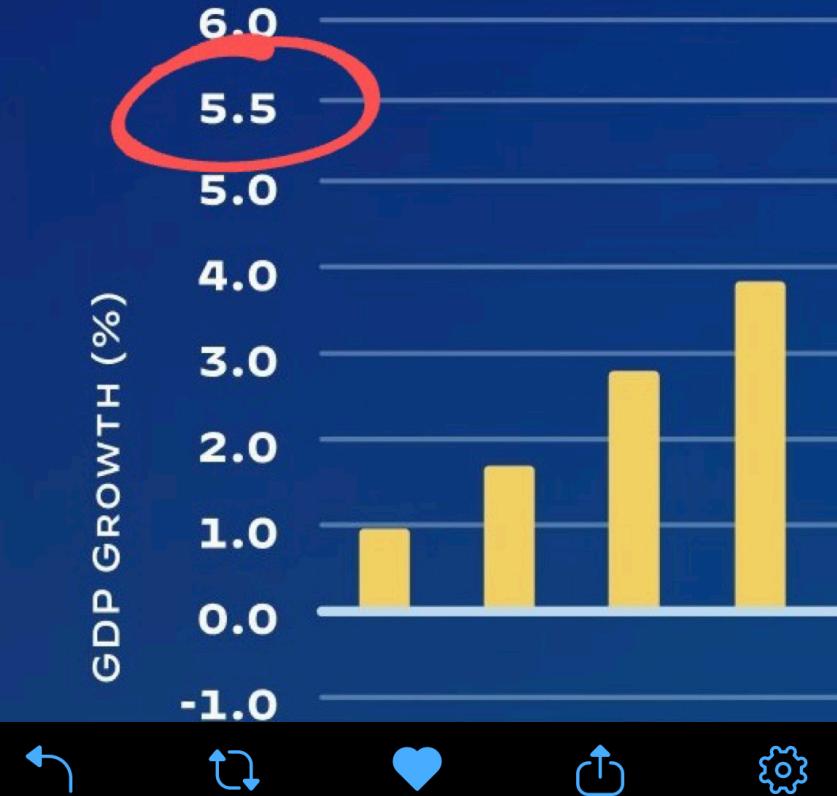


6h

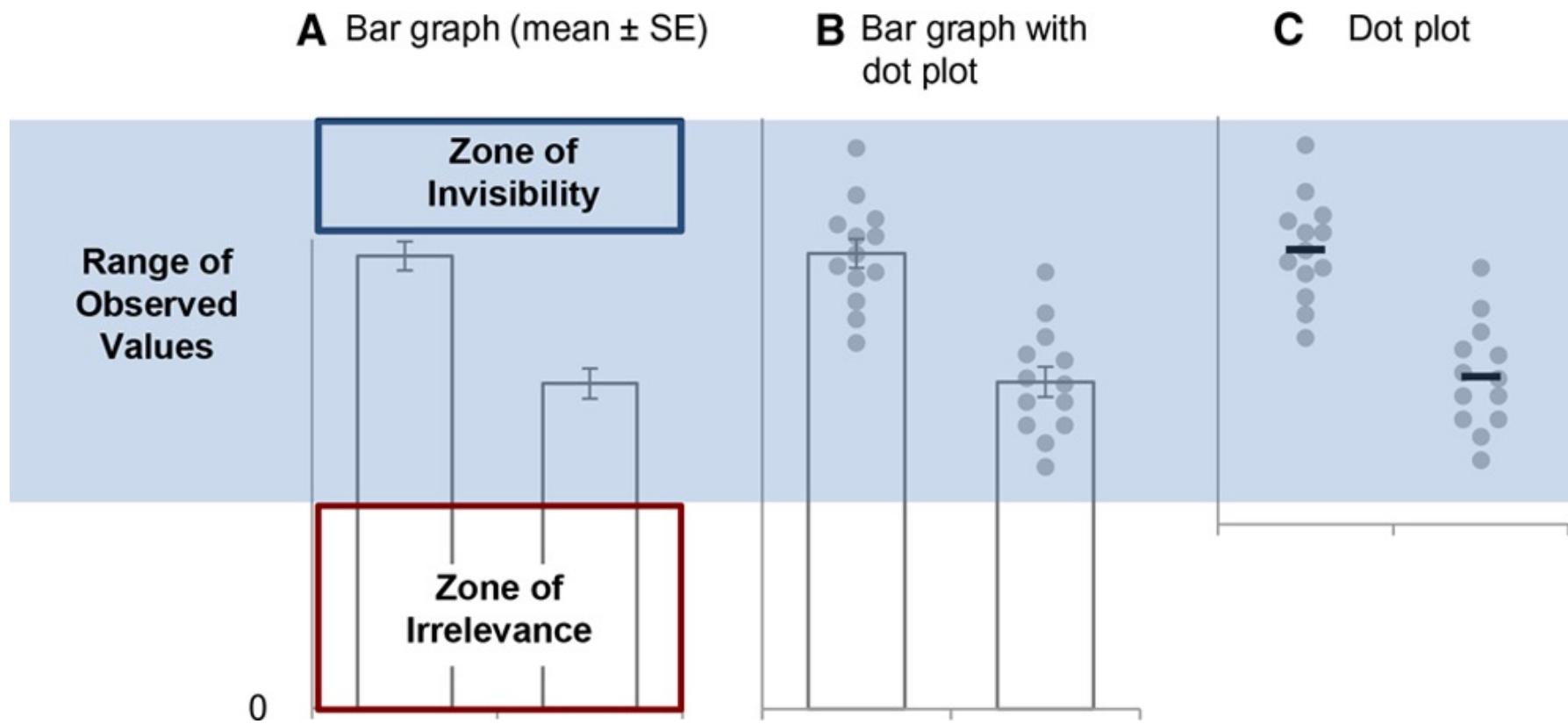


**Chad Scherrer**  
@ChadScherrer

@WhiteHouse C'mon man



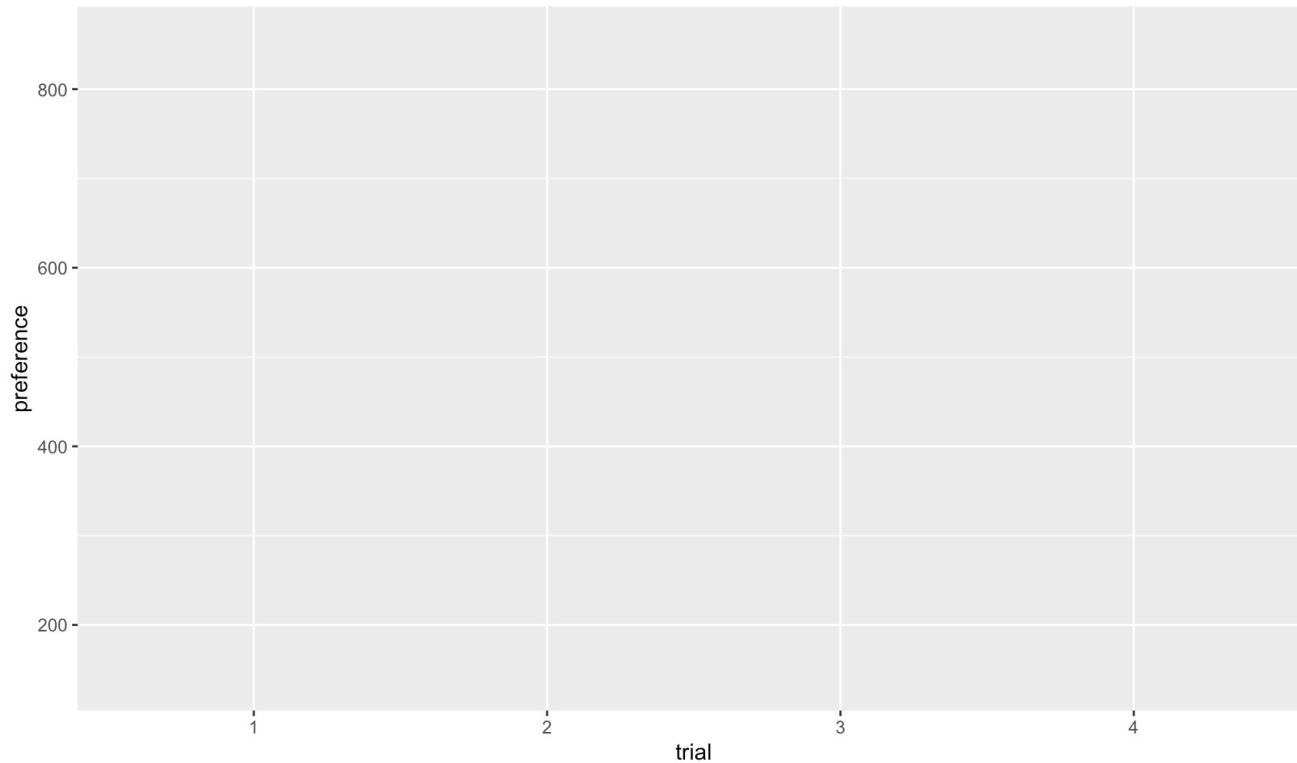
# Optimizing data display



# ggplot2 fundamentals

- `ggplot(data = your_data, aes(x = ..., y = ..., color = ...)) +  
geom_point()`
- ggplot followed by + not %>%!
- ggplot specifies blank canvas
- **aesthetics** specify mapping of data frame to elements of canvas
- `geom_line()`, `geom_point()`, ... placement of graphical elements (**geoms**) on top of ggplot "canvas"
- each geom can have data, aes arguments; if not specified, inherits from ggplot statement

```
> hw4_data %>% glimpse
Rows: 80
Columns: 4
$ subject_ID <fct> 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 6,
$ group      <fct> control, control, control, control, control, control,
$ trial       <fct> 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1,
$ preference <dbl> 323, 474, 440, 345, 282, 379, 412, 425, 250, 281, 303, 213, 353
```

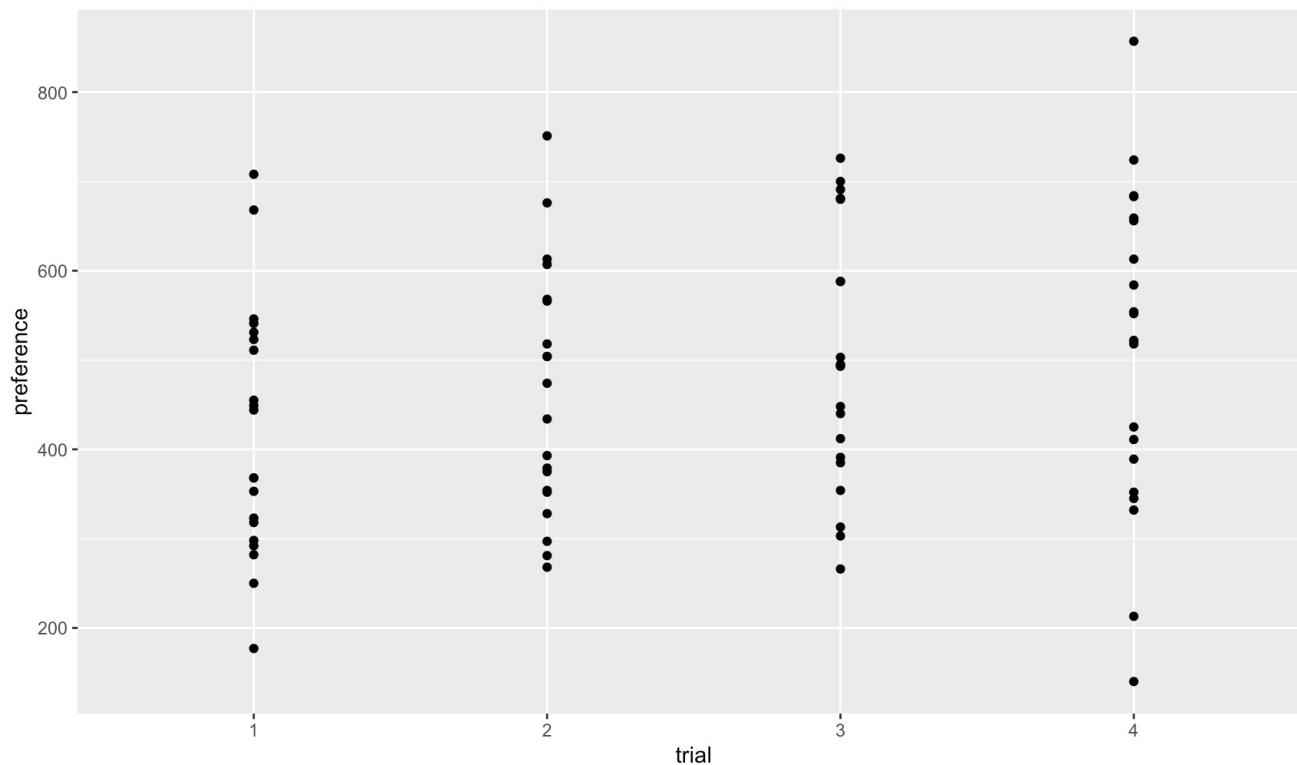


```
hw4_data %>% ggplot(aes(x=trial,y=preference))
```

sets axes, scales, labels  
maps "trial" to x and "preference" to y  
trial is a factor – discrete categories on x axis

must add geoms to plot data

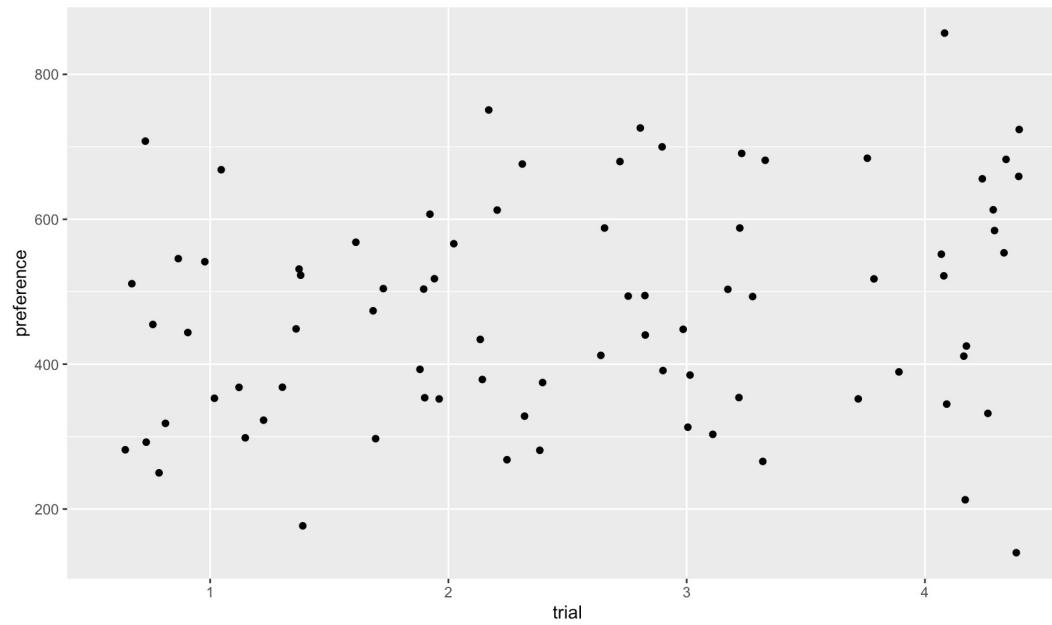
```
> hw4_data %>% glimpse  
Rows: 80  
Columns: 4  
$ subject_ID <fct> 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 6,  
$ group      <fct> control, control, control, control, control, control,  
$ trial       <fct> 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1,  
$ preference <dbl> 323, 474, 440, 345, 282, 379, 412, 425, 250, 281, 303, 213, 353
```



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_point()
```

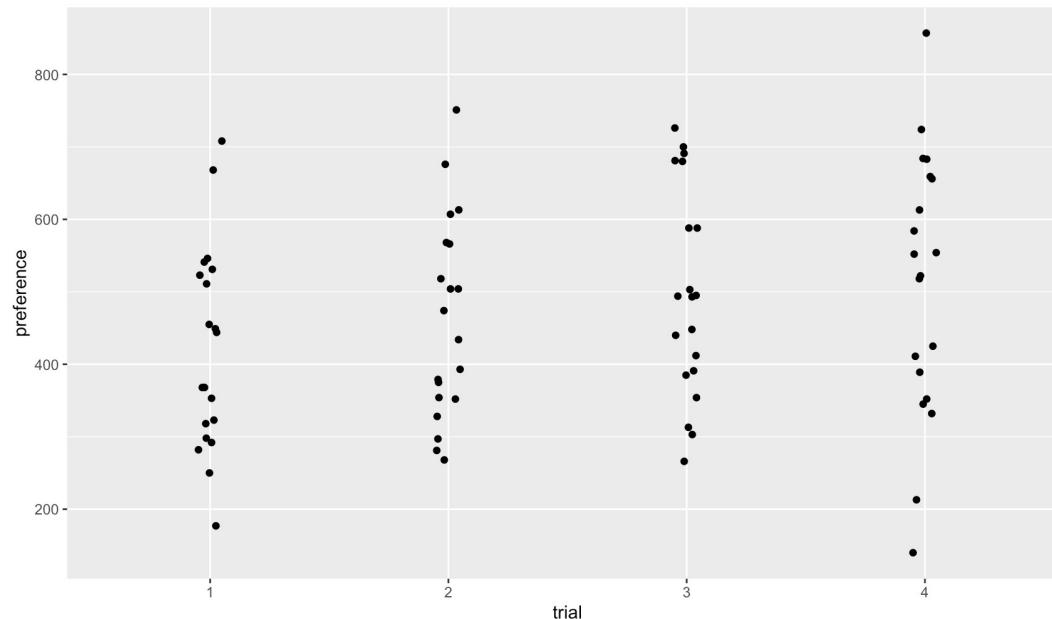
add a "geom" – specifies data display

interprets mapping of data based on aesthetics, in this case looking for (x, y) points



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_jitter()
```

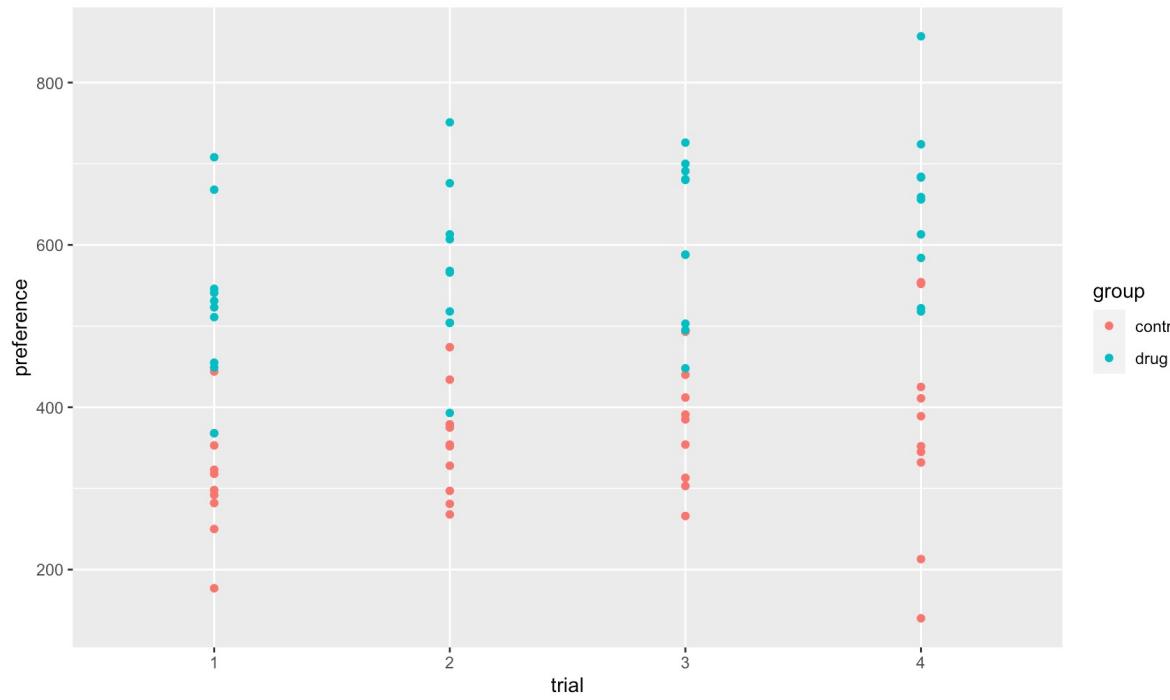
useful with overlapping points  
but they're too jittered!



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_jitter(height = 0.05, width = 0.05)
```

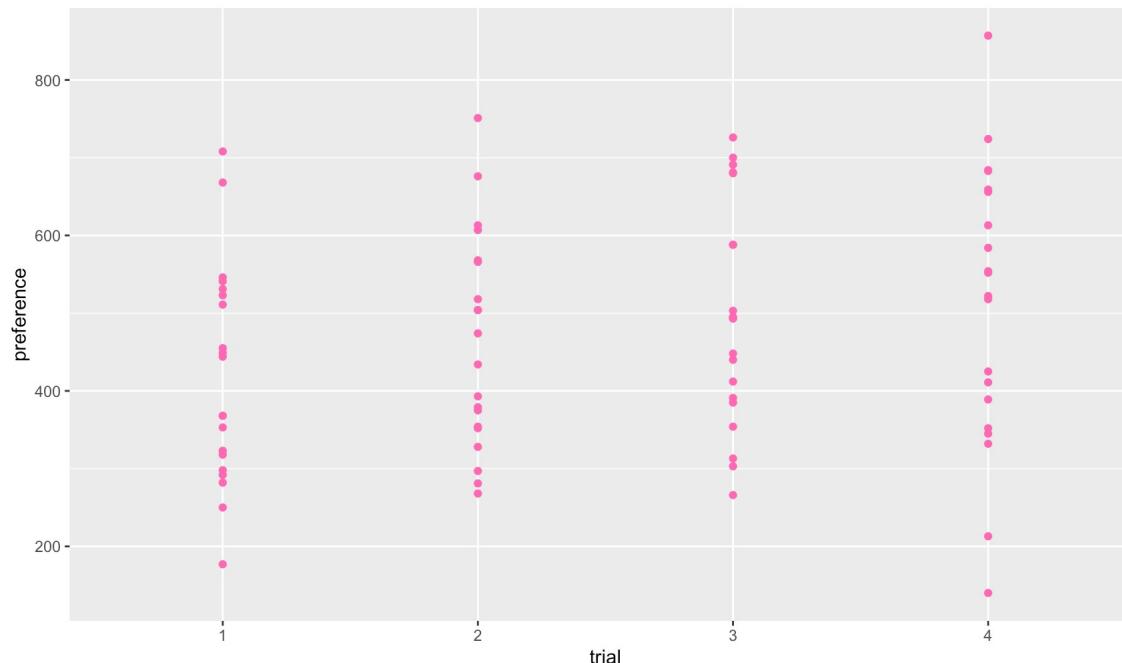
can specify options for geoms  
in this case amount of vertical and horizontal jitter

nicer visual effect



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_point(aes(color = group))
```

as before

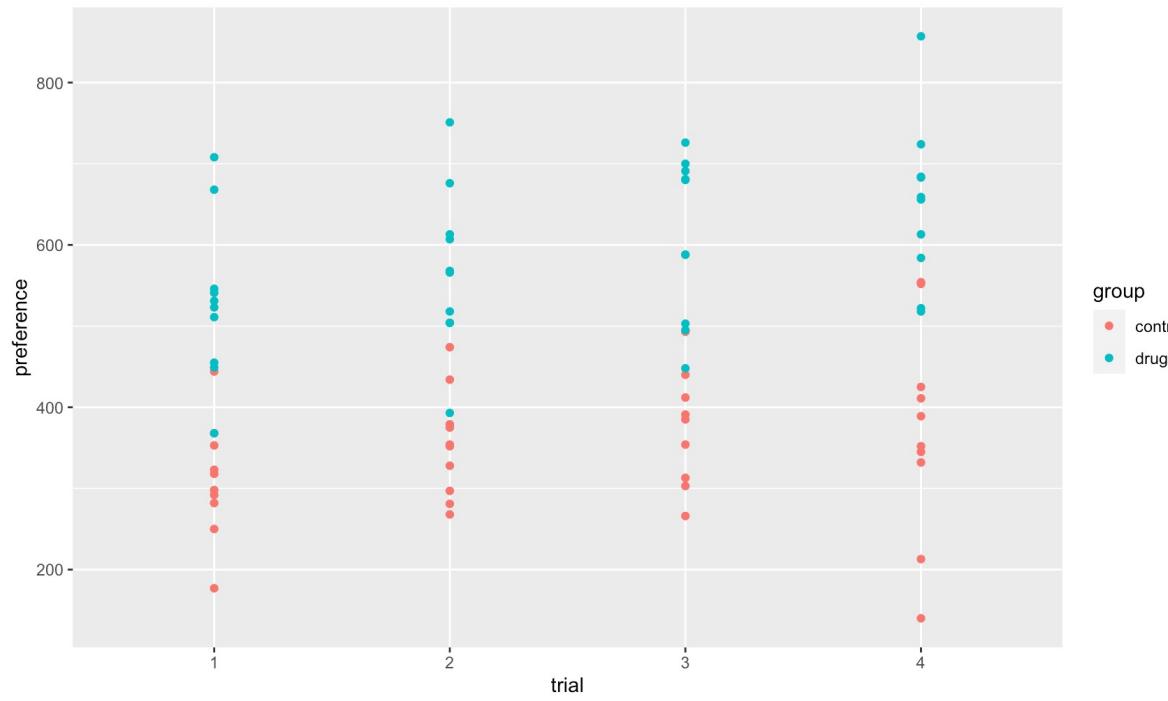


```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_point(color = "hotpink")
```

"color" can be set outside the aesthetic

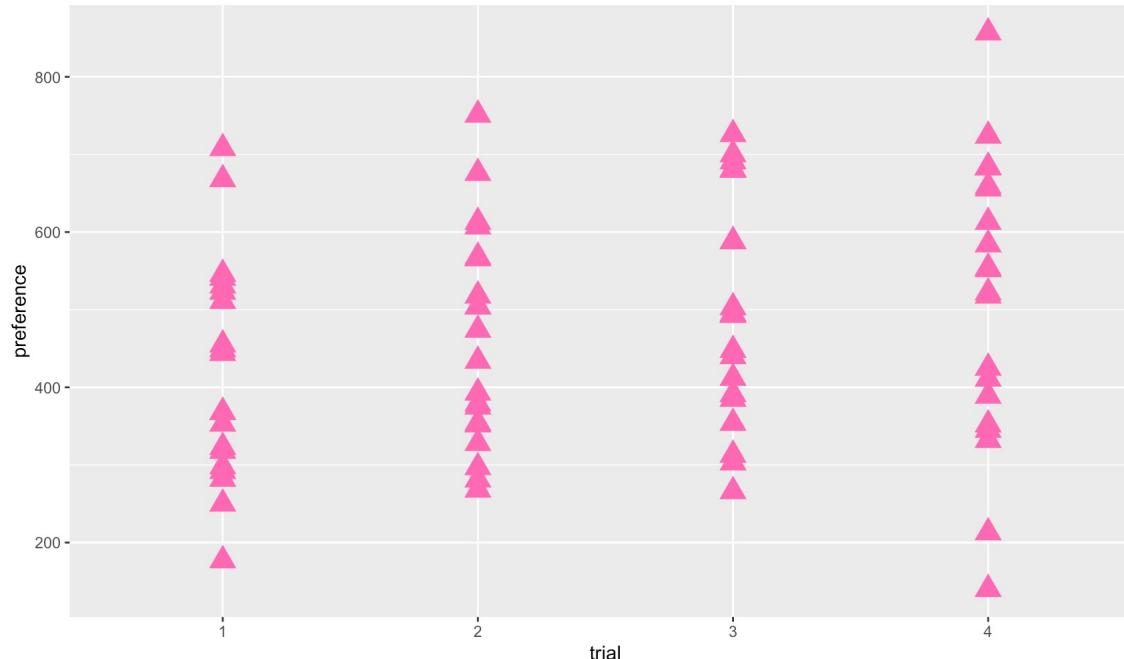
value of color is not mapped to an element of the data

if color = group not inside aes() you get an error



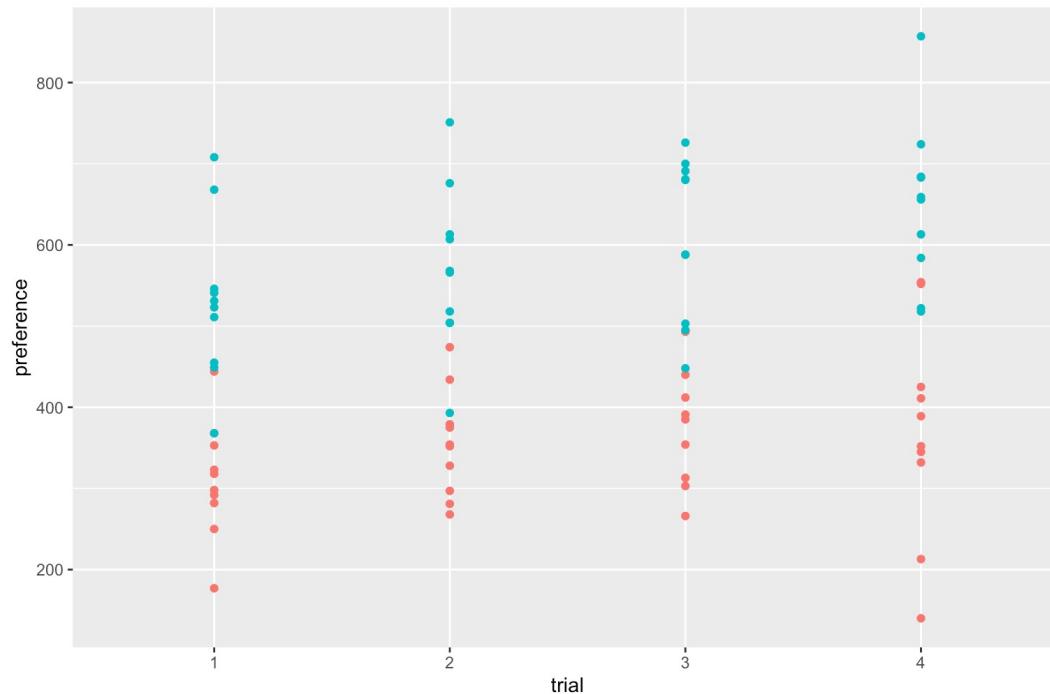
```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_point(aes(color = group))
```

as before



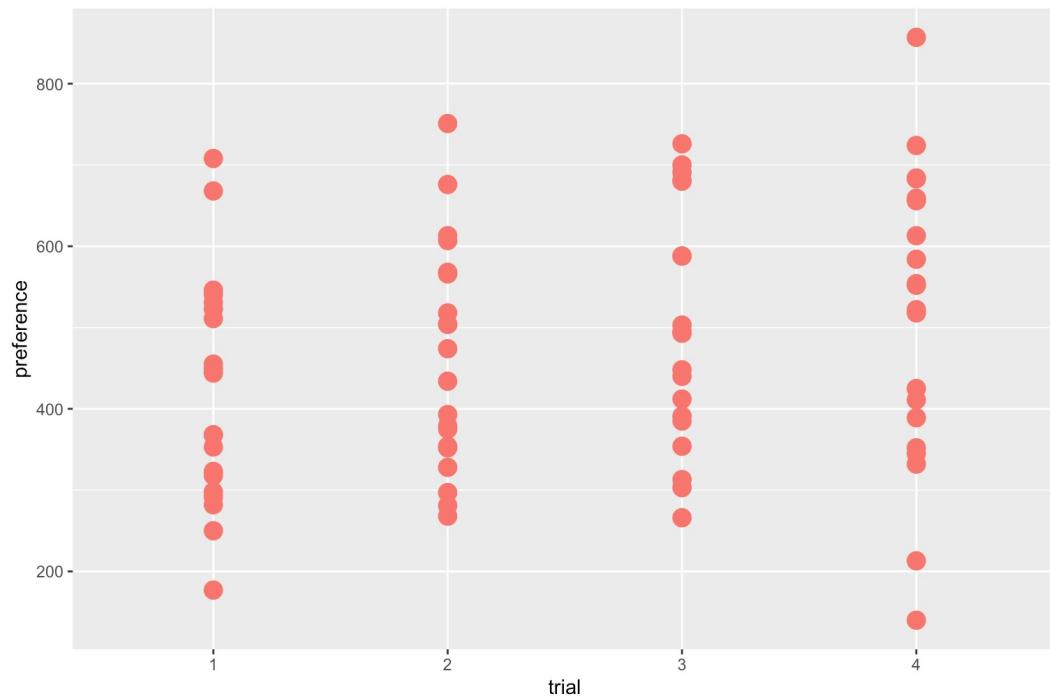
```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_point(color = "hotpink", size = 5, shape = "triangle")
```

can modify other elements as well



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_point(aes(color = group))
```

as before



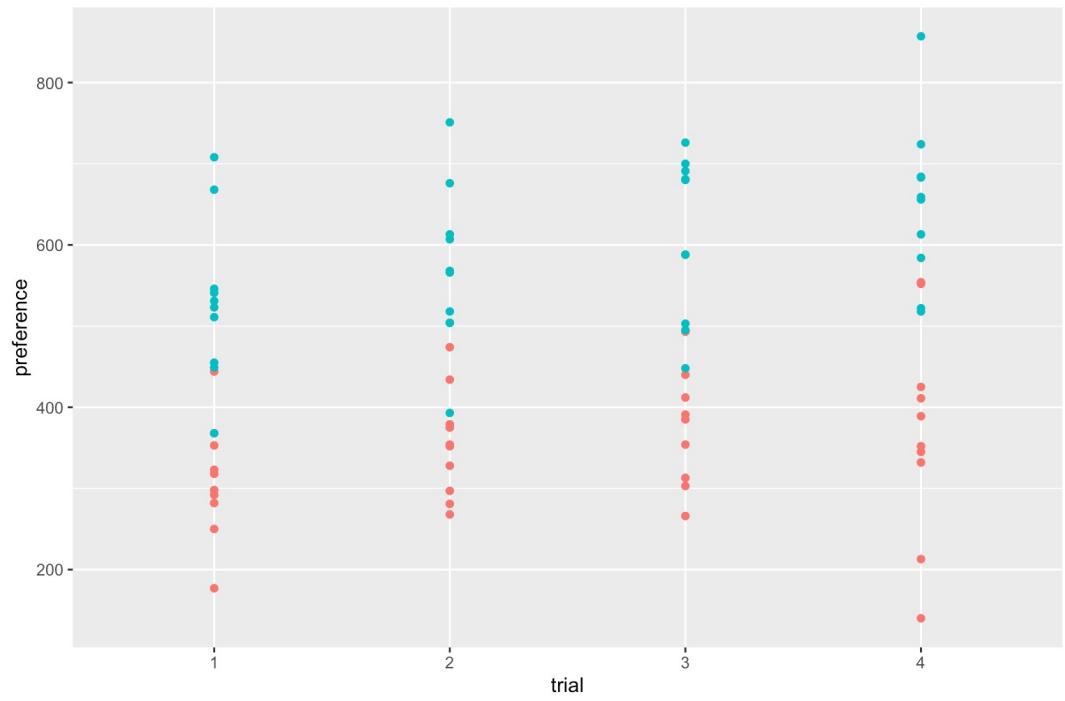
```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_point(aes(color = "hotpink", size = 5, shape = "triangle"))
```

size  
● 5

colour  
● hotpink

shape  
● triangle

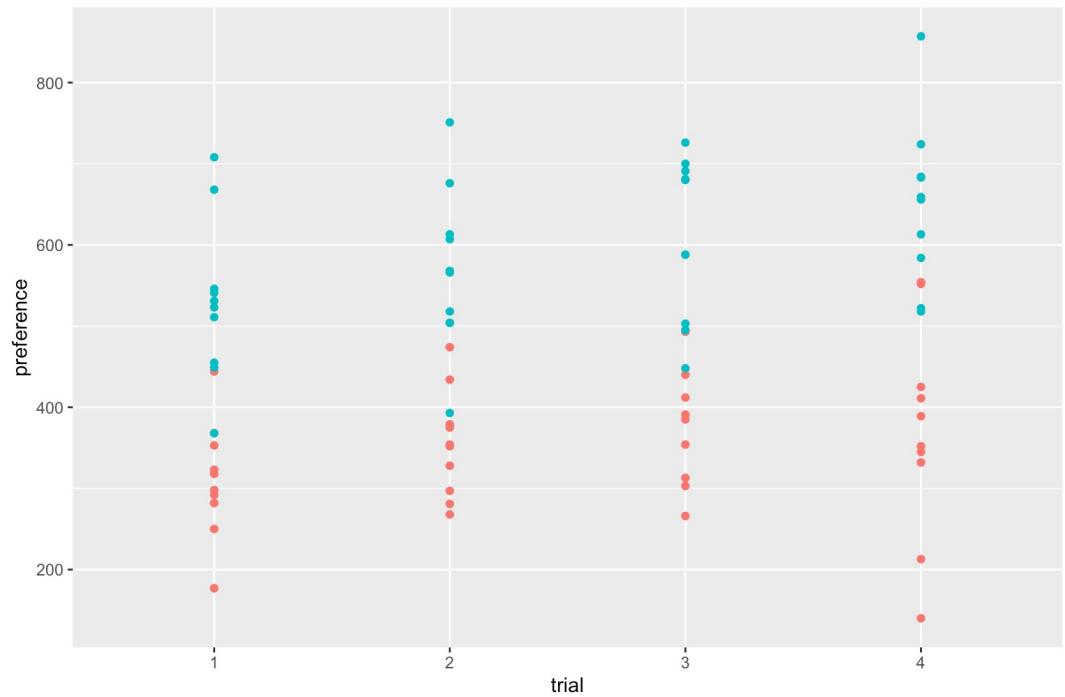
if you accidentally put constants inside the aesthetic,  
the behavior is not what you would want



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_point(aes(color = group))
```

as before

what if you want lines illustrating your data, like one line per subject going across trials?



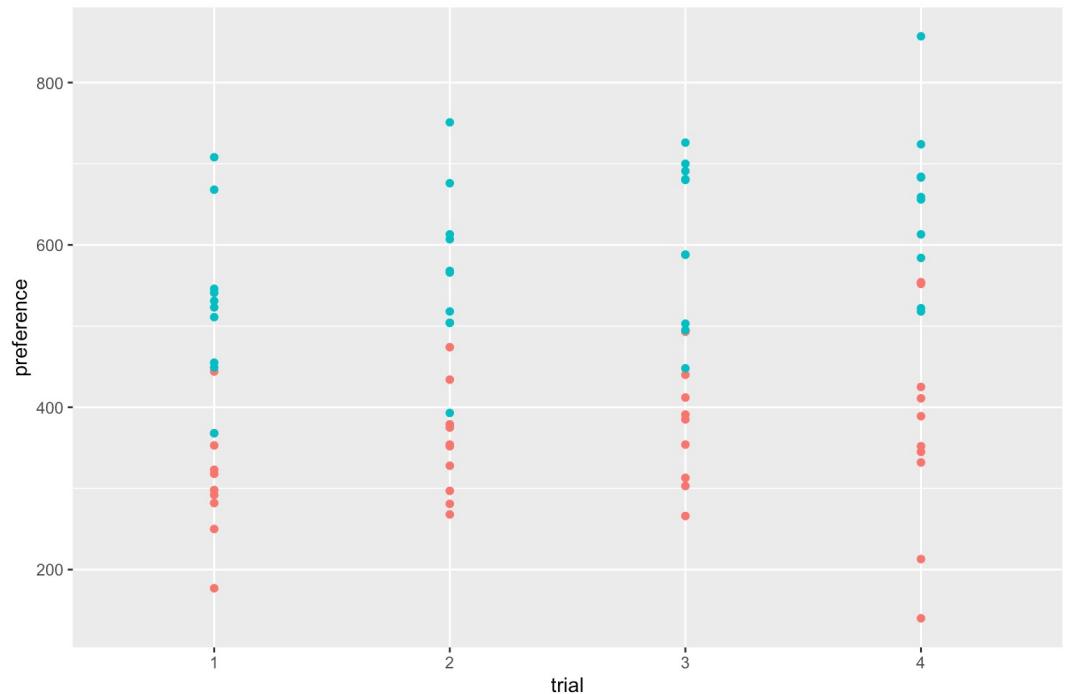
```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_point(aes(color = group))
```

as before

what if you want lines illustrating your data, like one line per subject going across trials?

```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_line(aes(color = group))
```

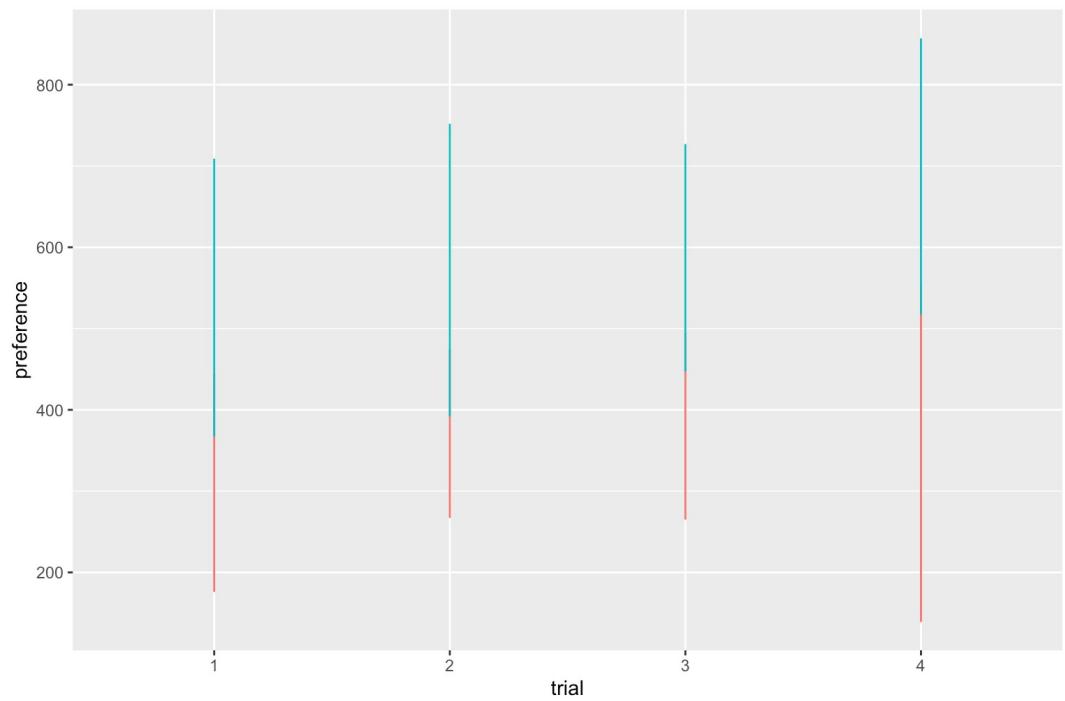
just use `geom_line()` instead of `geom_point()` !



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_point(aes(color = group))
```

as before

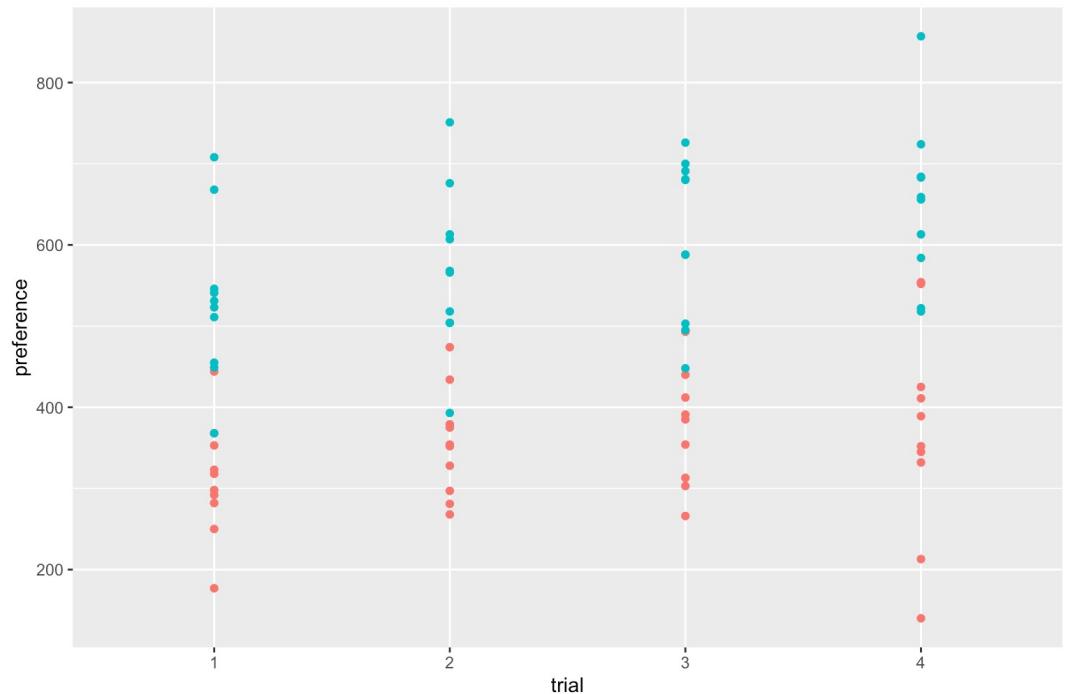
what if you want lines illustrating your data, like one line per subject going across trials?



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_line(aes(color = group))
```

just use `geom_line()` instead of `geom_point()` !

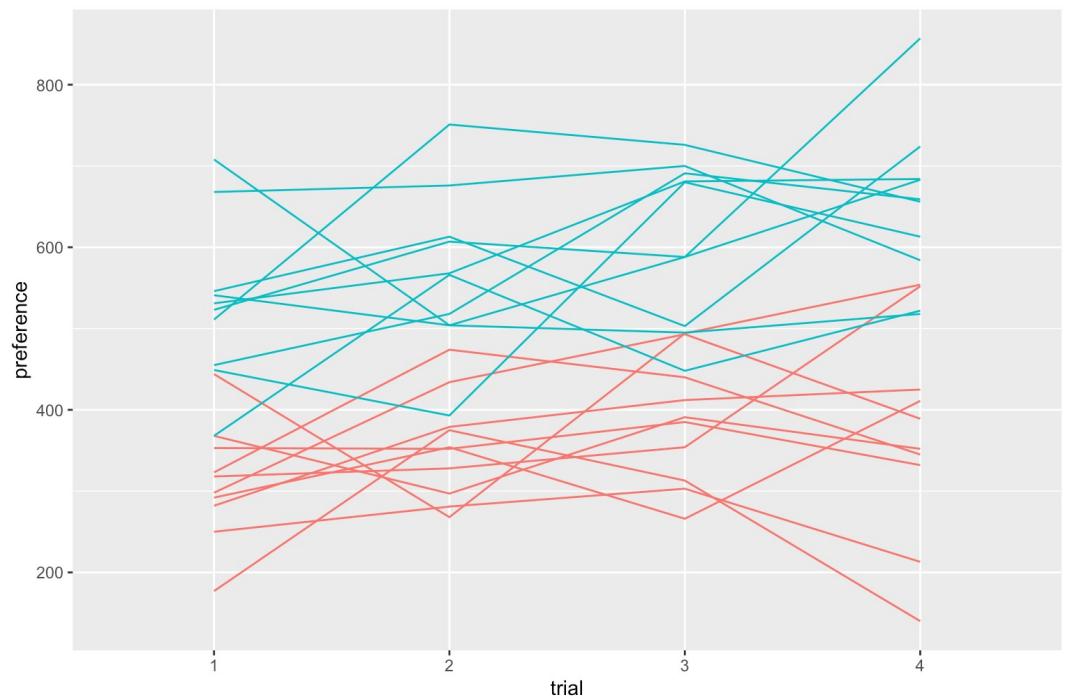
oh no that's no good



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_point(aes(color = group))
```

as before

what if you want lines illustrating your data, like one line per subject going across trials?



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_line(aes(color = group, group = subject_ID))
```

this is one option: `geom_line()` takes a "group" aesthetic

since your data specify which points come from which subject, you can use that to tell it which data points to connect with a line

note "group" is both an aesthetic and a variable!

# Some plotting tips: stat\_summary

make a summary data frame and graph it (fine)

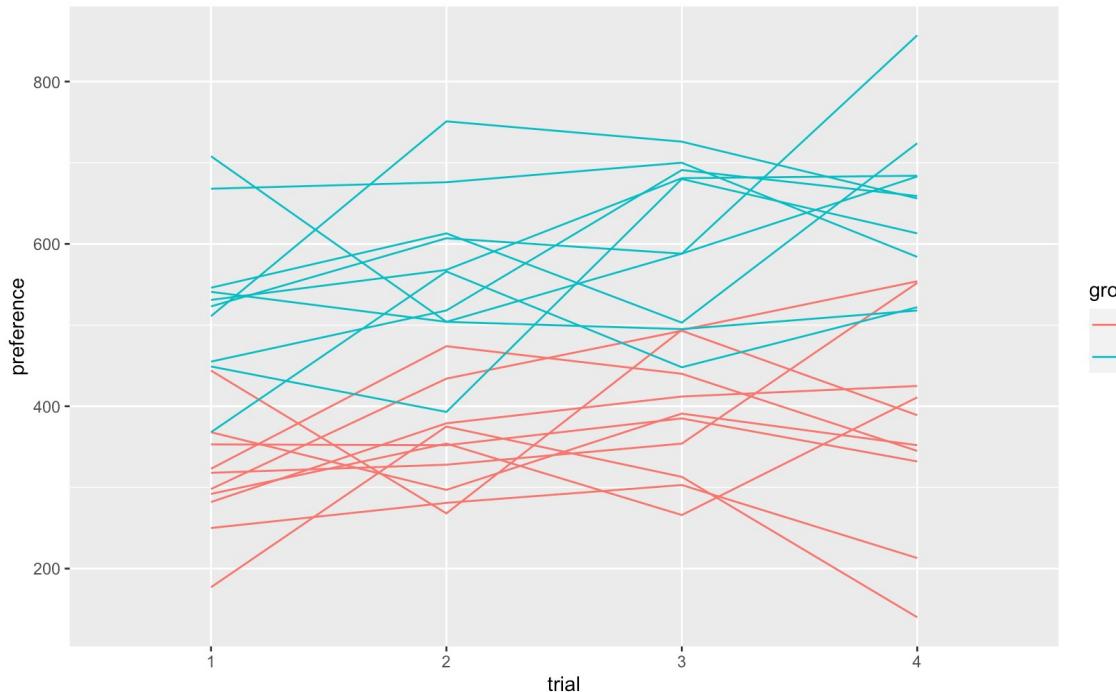
```
hw4_means <- hw4_data %>% group_by(group, trial) %>% summarize(pref_mean = mean(preference)) %>% ungroup()  
  
hw4_means %>% ggplot(aes(x = trial, y = pref_mean, color = group, fill = group)) +  
  geom_col(position = "dodge")
```

use original data and allow stat\_summary to generate a statistic from the data (better)

```
hw4_data %>% ggplot(aes(x=trial,y=preference, color=group, fill = group)) +  
  geom_line(aes(group=subject_ID)) +  
  stat_summary(fun = mean, geom = "col", position = "dodge")
```

```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_line(aes(color = group, group = subject_ID)) +  
  stat_summary(geom = "line", fun = "mean", color = "black")
```

add an overall summary line

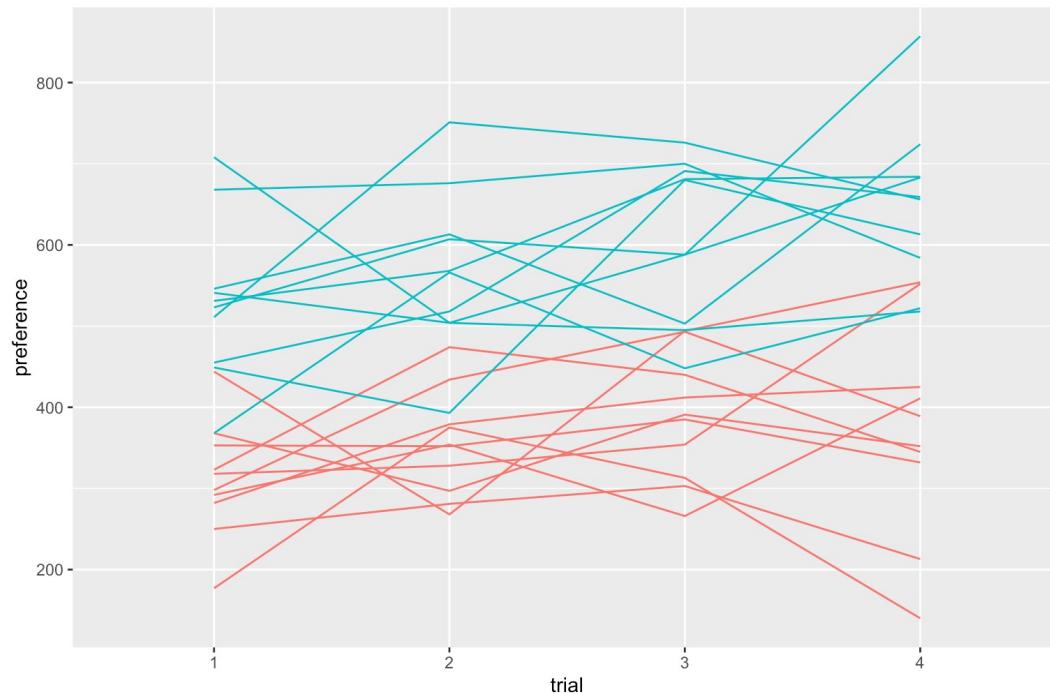


```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_line(aes(color = group, group = subject_ID)) +  
  stat_summary(geom = "line", fun = "mean", color = "black")
```

add an overall summary line

... nothing happens and you get this error

```
> hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
+   geom_line(aes(color = group, group = subject_ID)) +  
+   stat_summary(geom = "line", fun = "mean", color = "black")  
geom_path: Each group consists of only one observation. Do you need to adjust the group aesthetic?
```

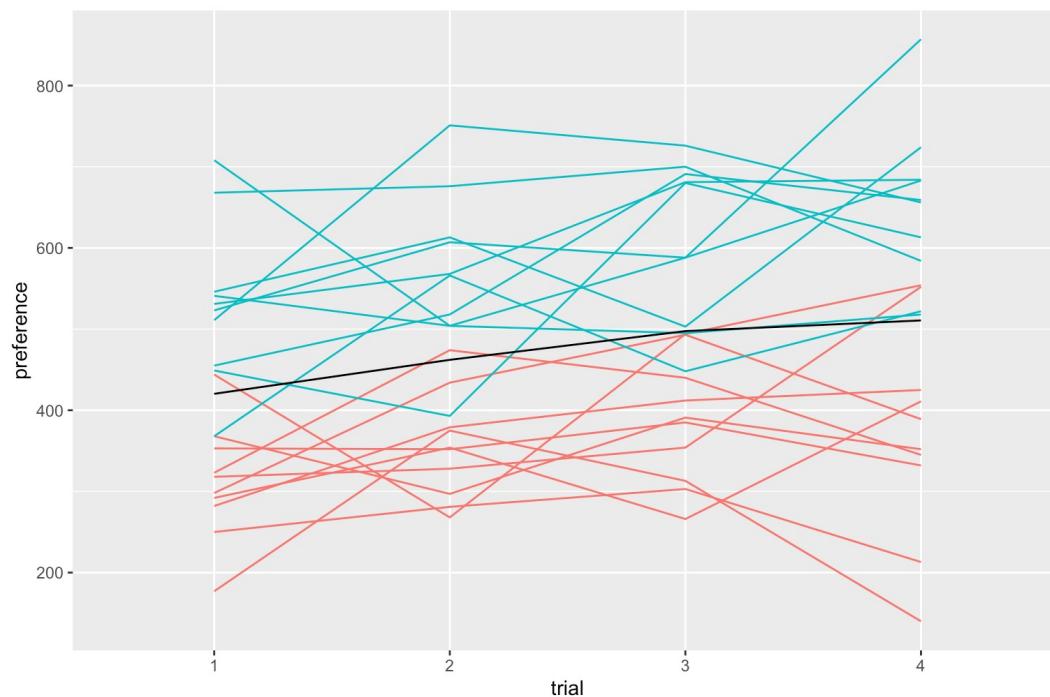


group  
control  
drug

```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_line(aes(color = group, group = subject_ID)) +  
  stat_summary(geom = "line", fun = "mean", color = "black")
```

add an overall summary line

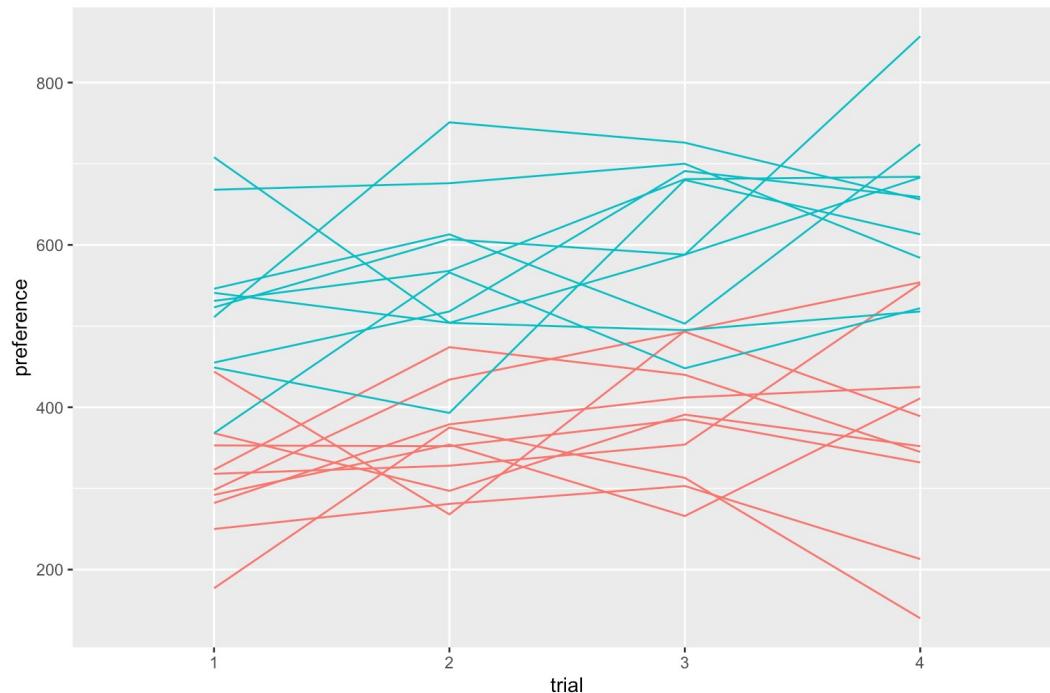
... nothing happens and you get this error



group  
control  
drug

```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_line(aes(color = group, group = subject_ID)) +  
  stat_summary(geom = "line", fun = "mean", color = "black", aes(group = 1))
```

now you have an overall summary line

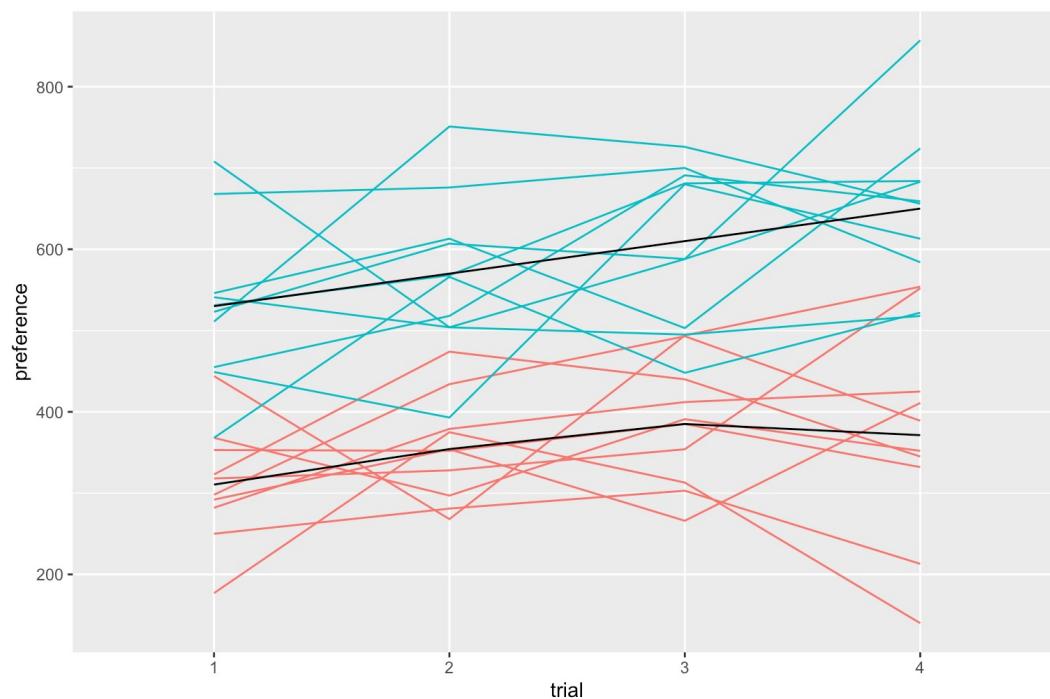


```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +
  geom_line(aes(color = group, group = subject_ID)) +
  stat_summary(geom = "line", fun = "mean", color = "black")
```

group  
control  
drug

add an overall summary line

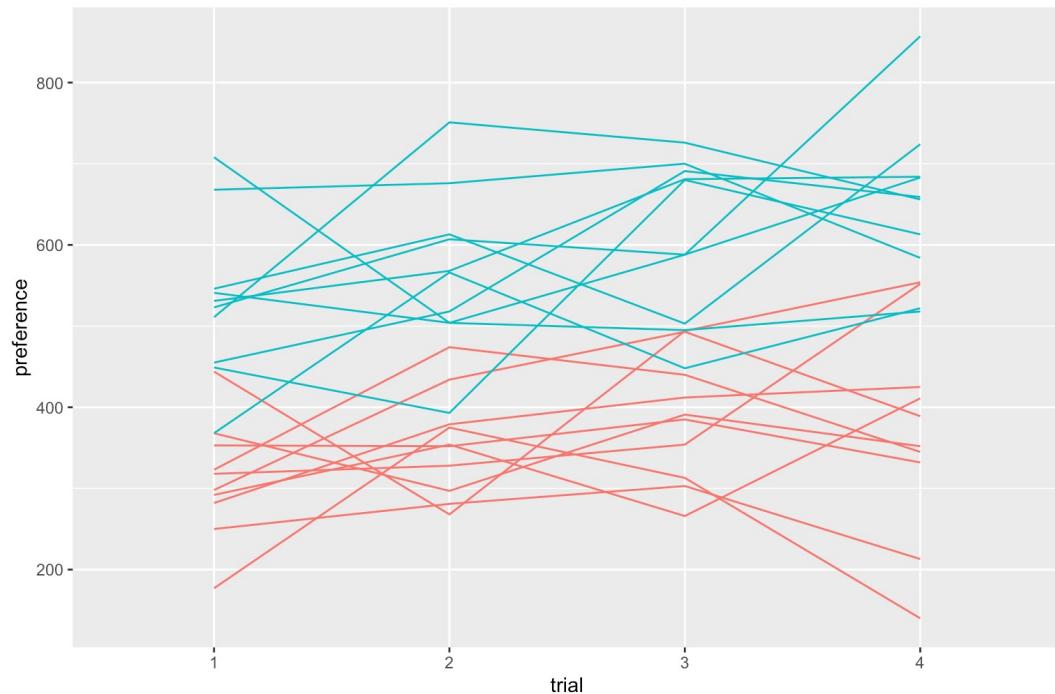
... nothing happens and you get this error



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +
  geom_line(aes(color = group, group = subject_ID)) +
  stat_summary(geom = "line", fun = "mean", color = "black", aes(group = group))
```

group  
control  
drug

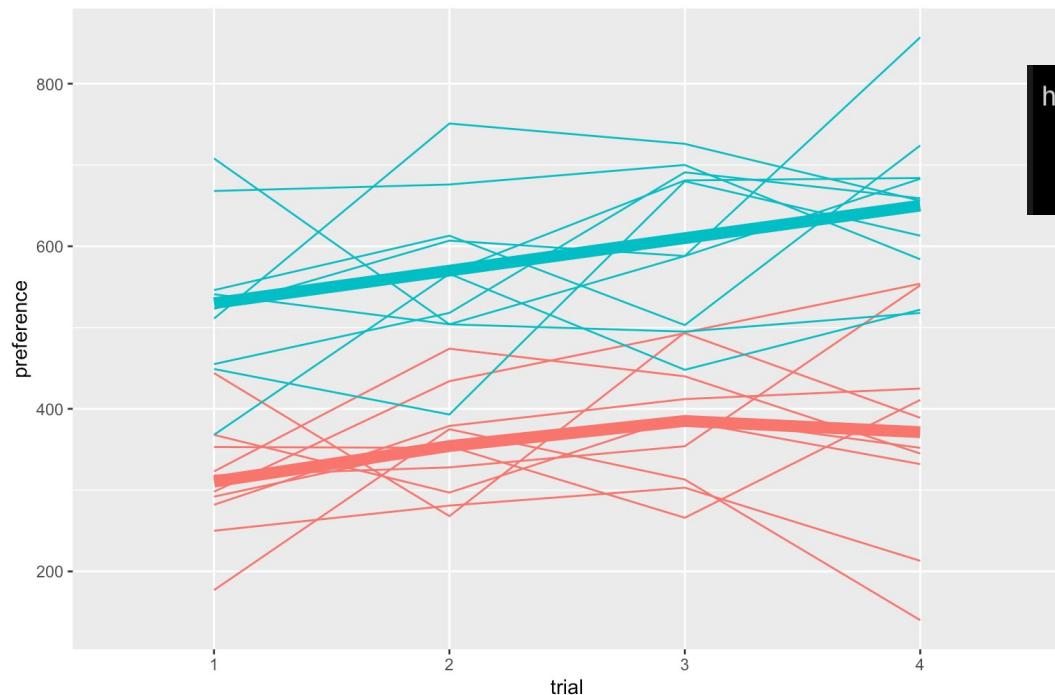
change grouping to "group" gives a separate line for each group (but they're both black)



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +
  geom_line(aes(color = group, group = subject_ID)) +
  stat_summary(geom = "line", fun = "mean", color = "black")
```

add an overall summary line

... nothing happens and you get this error



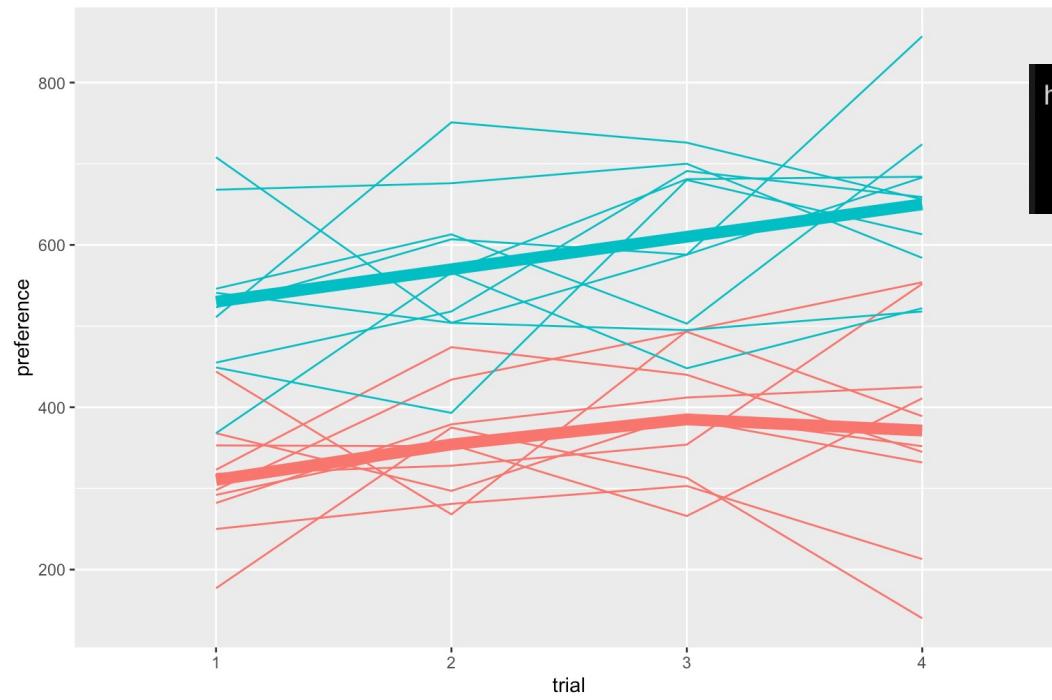
```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +
  geom_line(aes(color = group, group = subject_ID)) +
  stat_summary(geom = "line", fun = "mean", size = 3, aes(group = group, color = group))
```

group  
control  
drug

put "color" inside aesthetic so color varies by group

add "size = 3" so you can tell the mean lines apart from the raw data

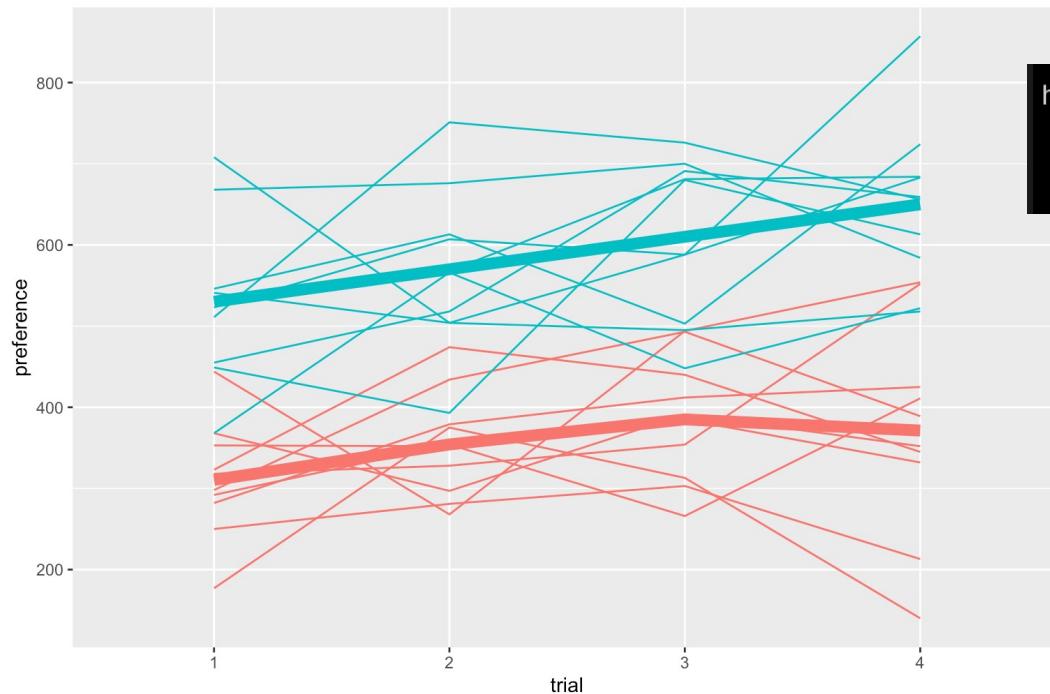
lovely!



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +  
  geom_line(aes(color = group, group = subject_ID)) +  
  stat_summary(geom = "line", fun = "mean", size = 3, aes(group = group, color = group))
```

group  
control  
drug

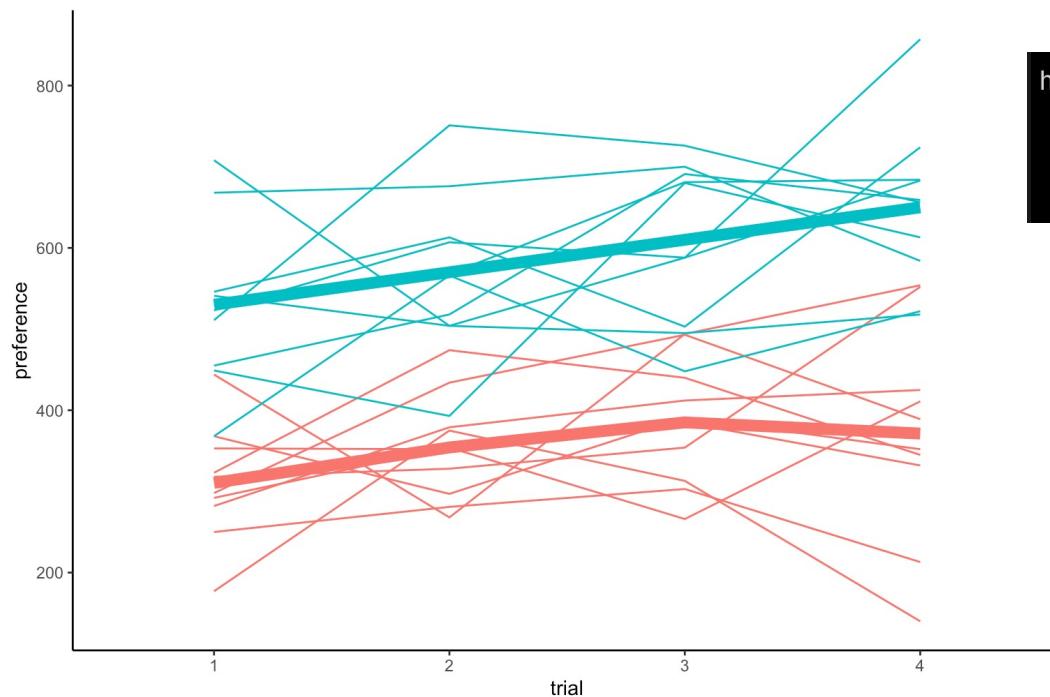
that gray background grid though



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +
  geom_line(aes(color = group, group = subject_ID)) +
  stat_summary(geom = "line", fun = "mean", size = 3, aes(group = group, color = group))
```

group  
control  
drug

that gray background grid though



```
hw4_data %>% ggplot(aes(x=trial,y=preference)) +
  geom_line(aes(color = group, group = subject_ID)) +
  stat_summary(geom = "line", fun = "mean", size = 3, aes(group = group, color = group)) +
  theme_classic()
```

group  
control  
drug

**themes** – many preset options to change visual properties of the graph (axes, grid, background)

these can all be adjusted manually as well

# Themes

<https://twitter.com/adamjnafa/status/1477092932561866752?s=20>

```
# Custom theme for data visualizations
plot_theme <- function(...) {
  theme_bw() + theme(
    # Set the outer margins of the plot to 1/5 of an inch on all sides
    #plot.margin = margin(0.2, 0.2, 0.2, 0.2, "in"),
    # Specify the default settings for the plot title
    plot.title = element_text(
      size = 22,
      face = "bold",
      family = "serif"
    ),
    # Specify the default settings for caption text
    plot.caption = element_text(
      size = 12,
      family = "serif"
    ),
    # Specify the default settings for subtitle text
    plot.subtitle = element_text(
      size = 16,
      family = "serif"
    ),
    # Specify the default settings for axis titles
    axis.title = element_text(
      size = 18,
      face = "bold",
      family = "serif"
    ),
    # Specify the default settings for axis labels
    axis.label = element_text(
      size = 16,
      family = "serif"
    ),
    # Specify the default settings for legend titles
    legend.title = element_text(
      size = 16,
      family = "serif"
    ),
    # Specify the default settings for legend labels
    legend.text = element_text(
      size = 14,
      family = "serif"
    )
  )
}
```

# Themes

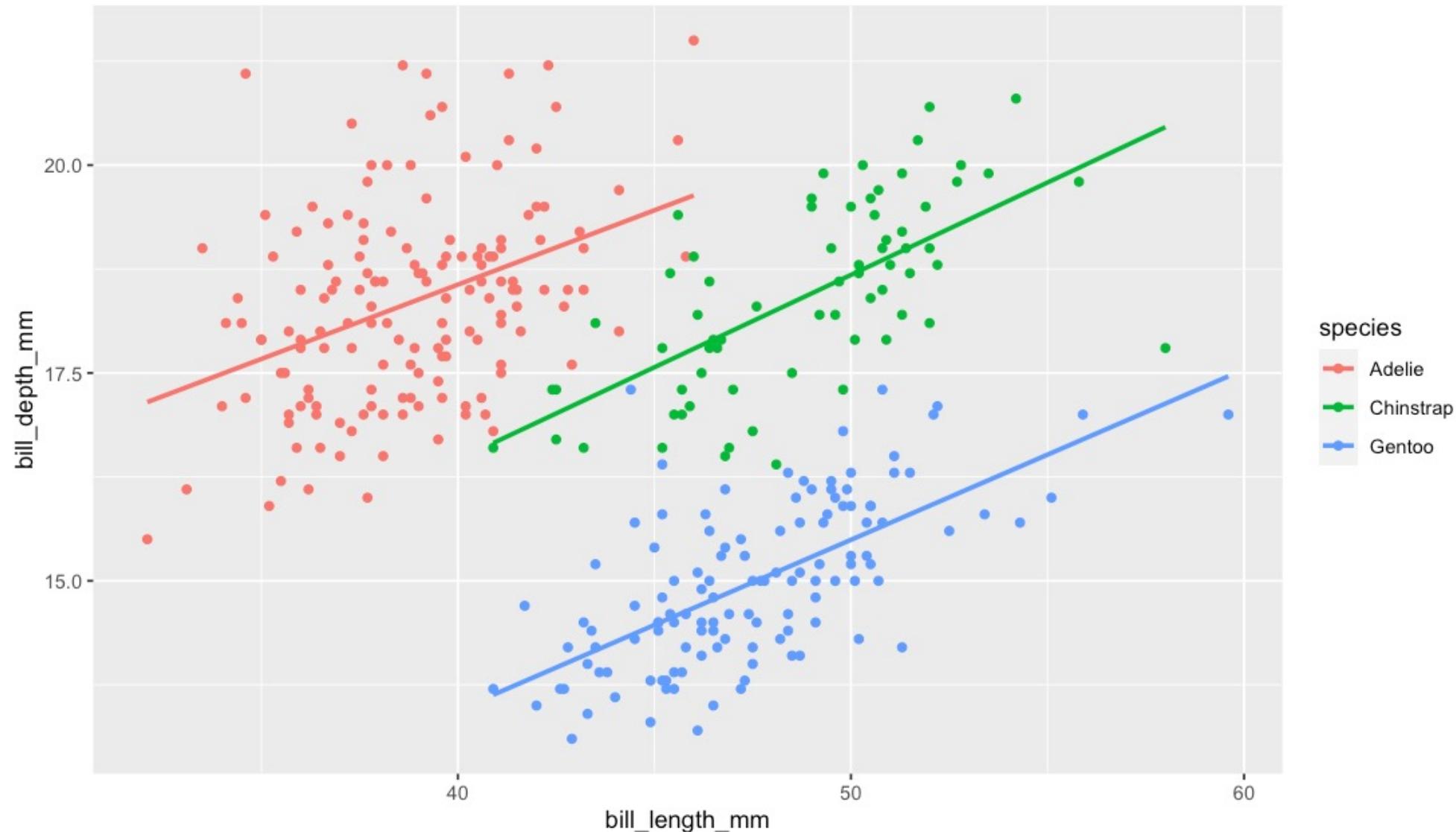
```
plot_theme <- function(...) {
  theme_classic() + theme(
    plot.title = element_text(size = 22),
    axis.title = element_text(face = "bold"),
    ...
  )
}

plot_stuff <- list(geom_point(), scale_color_viridis_c())

ggplot(aes(x,y) + plot_stuff + plot_theme()
```

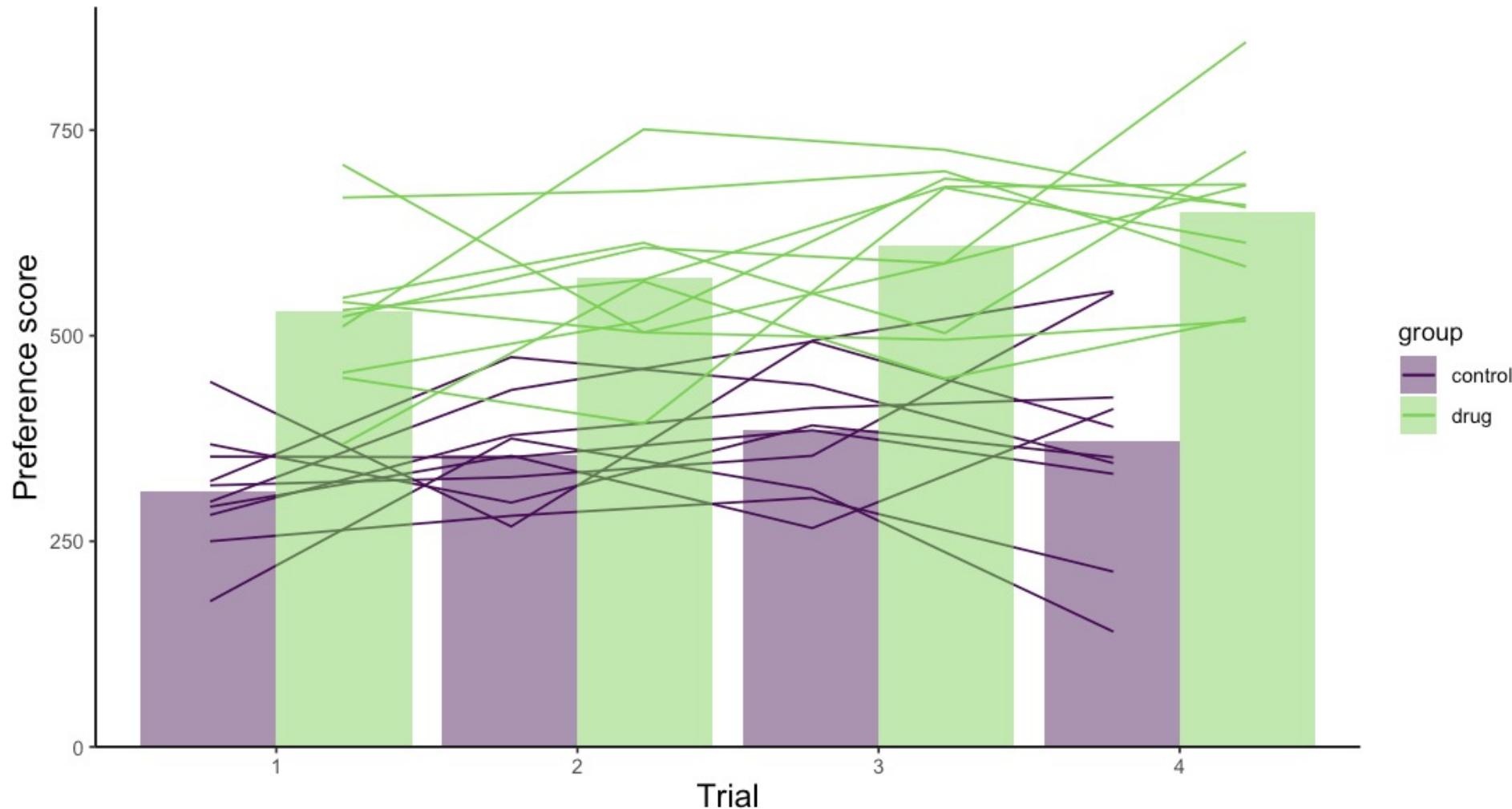
may help conserve code when making many similar graphs

geom\_smooth(method = "lm") will add regression lines



options for adjusting color scales (preset or make your own)

## Homework 4 Data



# Choice of colors

- About ~4.5% of US population has some degree of color blindness
  - Various collections of "safe" colors available for R
  - `prismatic`, `dichromat`, `colorblindr`, `colorBlindness` packages

replace the color of a plot to safe color

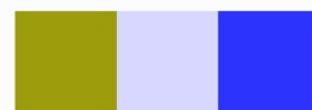
If you have figures which is not safe to color blindness people, you can try to replace the colors by `replacePlotColor` function.

```
cvdPlot(replacePlotColor(displayColors(c("Red", "Green", "blue")))))
```

### **normal vision**



**deutanopia (6%)**



**protanopia (2%)**



**desaturated (BW)**



# Multiple graphs in a frame

- cowplot
- gridExtra
- patchwork

# Output to graphics devices

- <https://bookdown.org/rdpeng/exdata/graphics-devices.html>
- <https://rforbiochemists.blogspot.com/2015/05/export-high-resolution-graph-for.html>
- may be part of your workflow
- (or screenshot and clean up in Photoshop / Illustrator / Pixelmator Pro)
- vector vs bitmap formats (be careful)

# Terrifying graphs

- <https://twitter.com/biogeobiochem/status/1172547846479831040>

