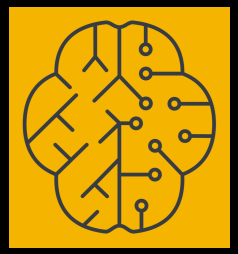


3 quick ways to lie 🤫 with your graph

Zezhen (Michael) Wu

The Agency Fund Research Mindsets Series

Learning Blitz Session

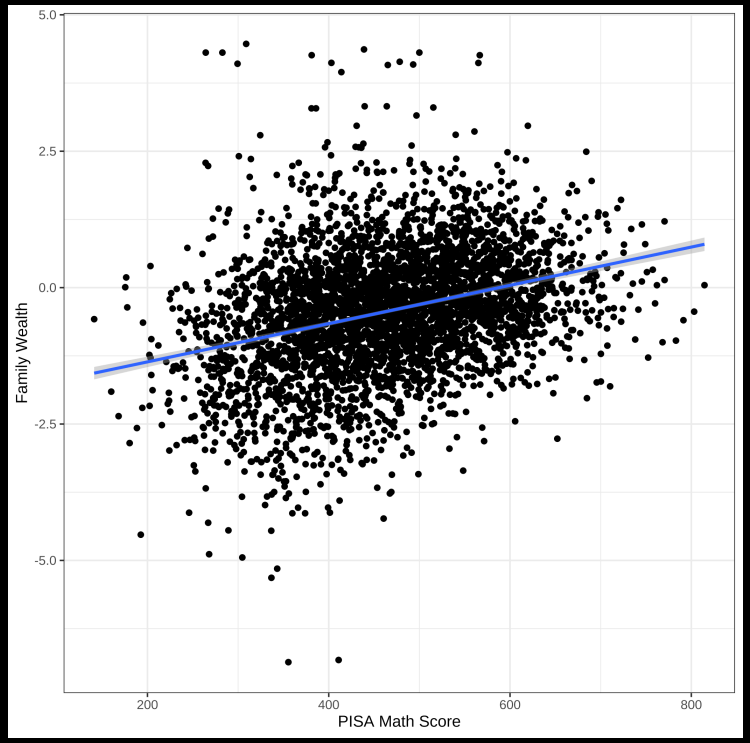
How to lie with your graph?

1. Trick the aspect ratio
2. Play with the axis
3. Muddy the plot

1. Trick the aspect ratio

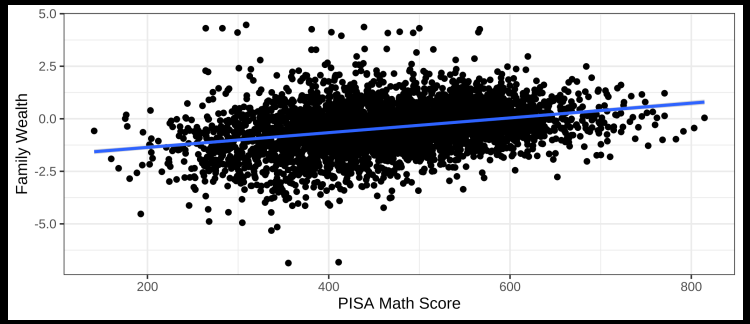
What story does this tell?

```
# Load PISA 2018 data

library(learningtower)
data(student_subset_2018)

# Plot the relationship between
# math score and family wealth
# (r = 0.31)

ggplot(student_subset_2018,
       aes(x = math, y = wealth)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_bw() +
  labs(x = "PISA Math Score",
       y = "Family Wealth")
```



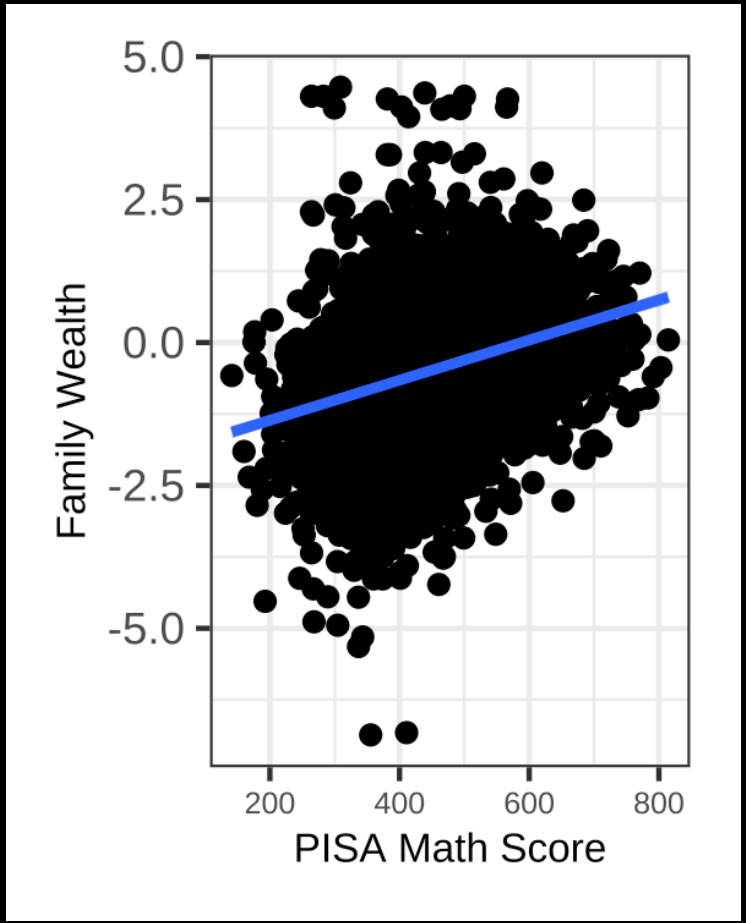
What about this one?

```r
# Load PISA 2018 data

library(learningtower)
data(student_subset_2018)

# Plot the relationship between
# math score and family wealth
# (r = 0.31)

ggplot(student_subset_2018,
       aes(x = math, y = wealth)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_bw() +
  labs(x = "PISA Math Score",
       y = "Family Wealth")
```

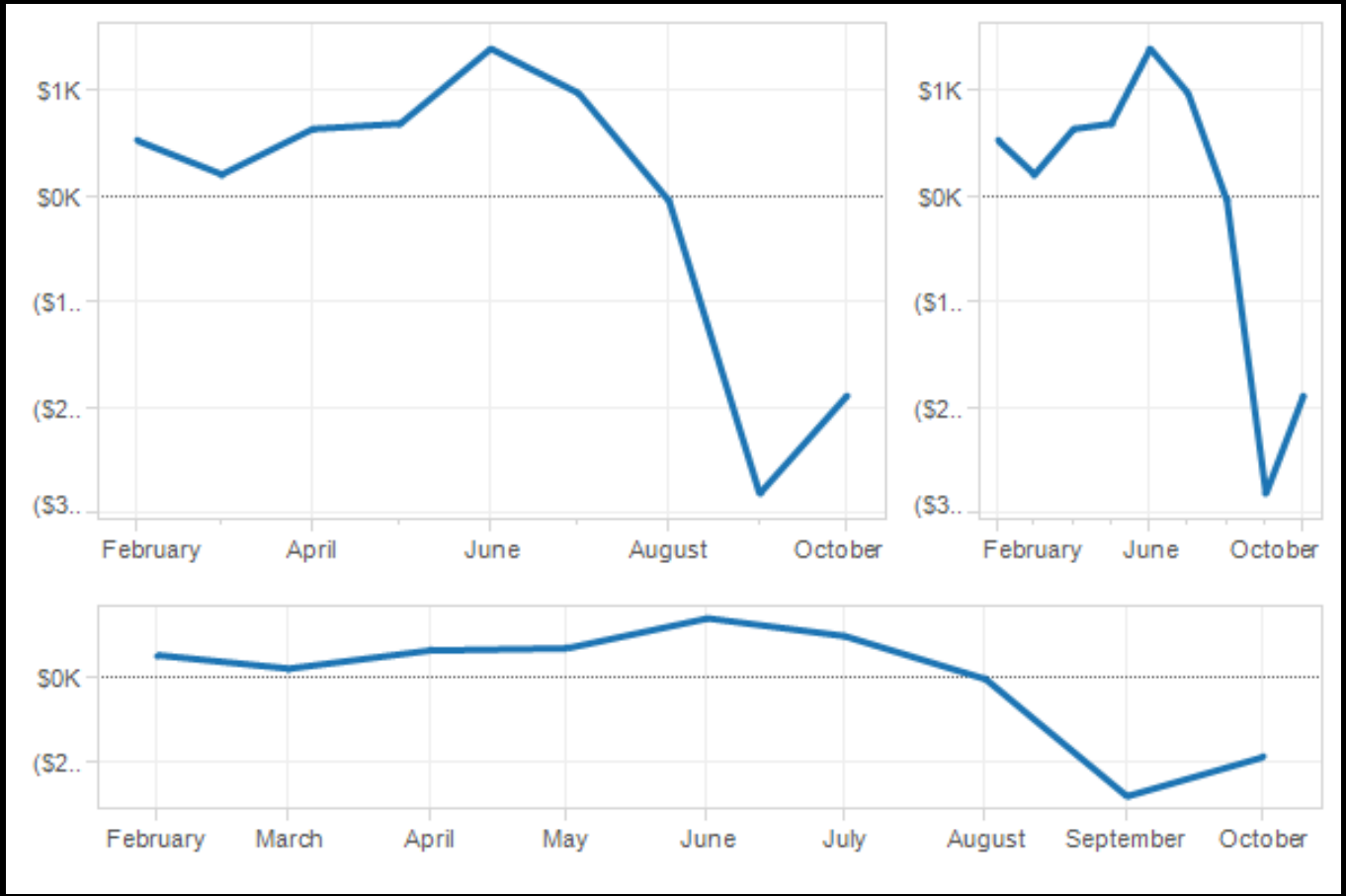


What about this one?

```r
# Load PISA 2018 data

library(learningtower)
data(student_subset_2018)

# Plot the relationship between
# math score and family wealth
# (r = 0.31)

ggplot(student_subset_2018,
       aes(x = math, y = wealth)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_bw() +
  labs(x = "PISA Math Score",
       y = "Family Wealth")
```



Solution: Banking to a "good" degree



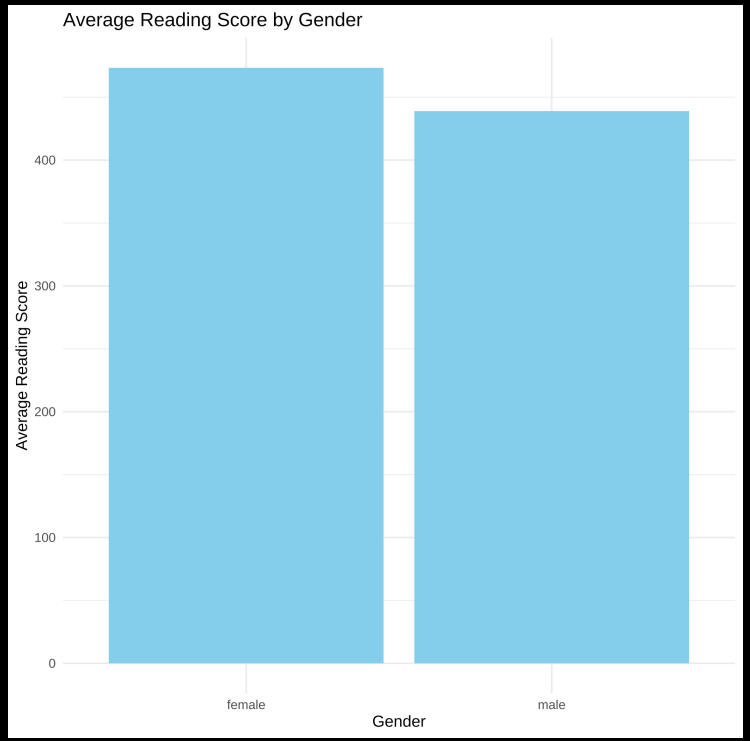
Source: https://eagereyes.org/blog/2013/banking-45-degrees

Cleveland, W. S., McGill, M. E., & McGill, R. (1988). The shape parameter of a two-variable graph. Journal of the American Statistical Association, 83(402), 289-300.

2. Play with the axis

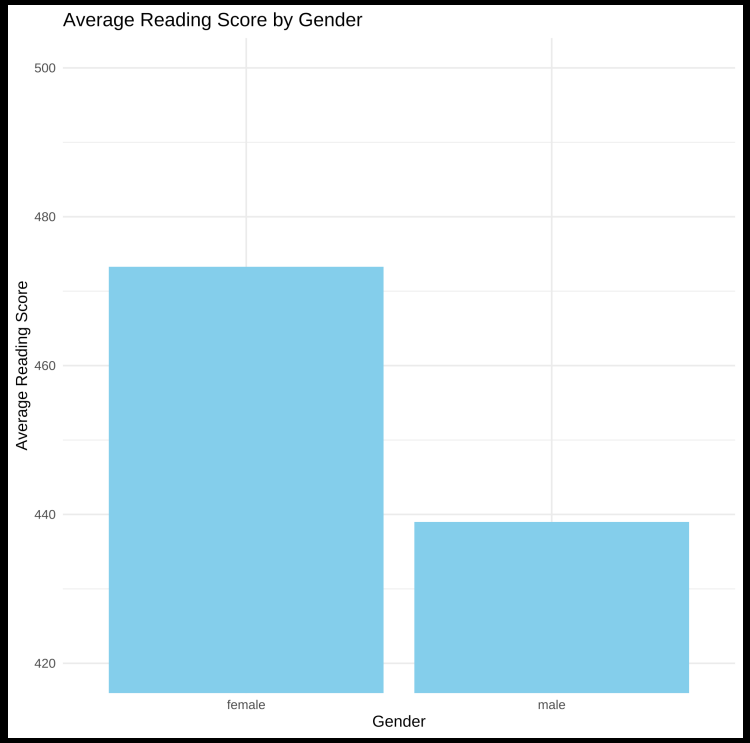
What story does this tell?

```
# Group by 'book' and calculate t
df <- student_subset_2018 %>%
  group_by(gender) %>%
  summarise(mean_read = mean(read
                             na.r
  na.omit()

# Create the bar plot
ggplot(df, aes(x = gender,
               y = mean_read)) +
  geom_bar(stat = "identity",
           fill = "skyblue") +
  labs(title = "Average Reading S
       x = "Gender",
       y = "Average Reading Score
  theme_minimal()
```



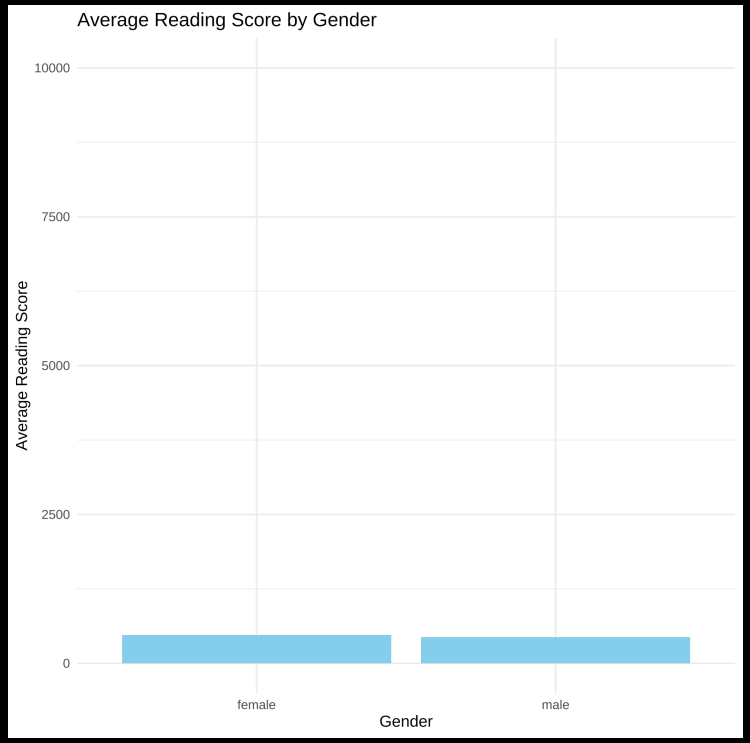
What about this one?

```
# Create the bar plot
ggplot(df, aes(x = gender,
               y = mean_read)) +
  geom_bar(stat = "identity",
           fill = "skyblue") +
  labs(title = "Average Reading S
       x = "Gender",
       y = "Average Reading Score
  coord_cartesian(ylim=c(420, 500
  theme_minimal()
```

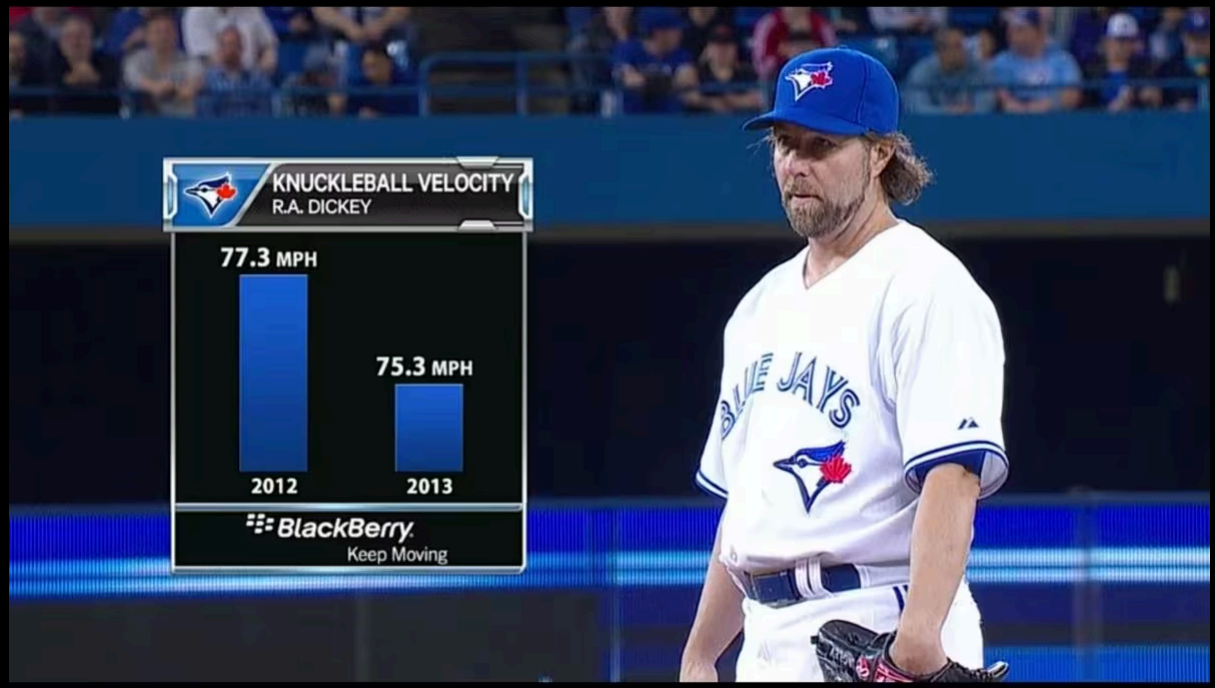


What about this one?

```
# Create the bar plot
ggplot(df, aes(x = gender,
               y = mean_read)) +
  geom_bar(stat = "identity",
           fill = "skyblue") +
  labs(title = "Average Reading S
       x = "Gender",
       y = "Average Reading Score
  coord_cartesian(ylim=c(0, 10000
  theme_minimal()
```

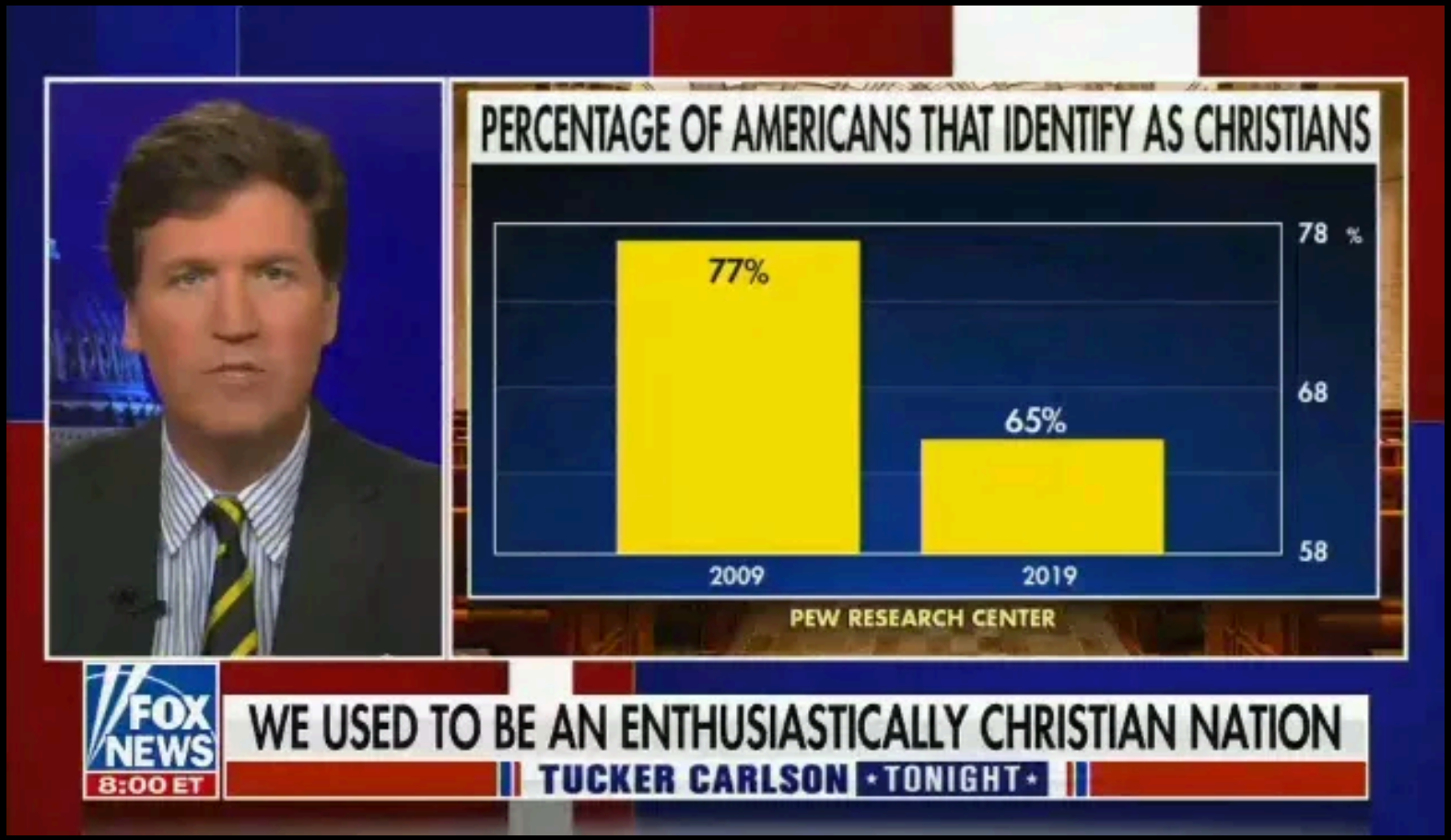


This is more common than you think!



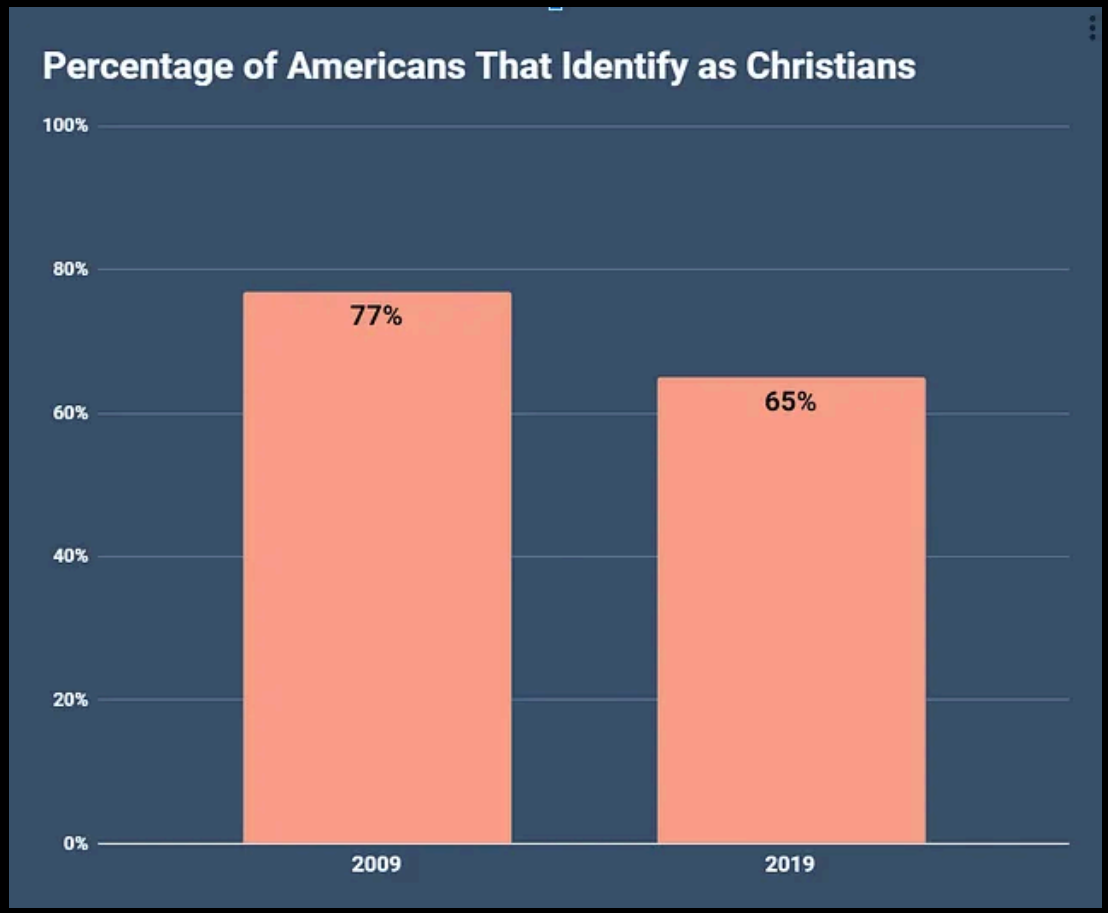
Source: https://www.heap.io/blog/how-to-lie-with-data-visualization

Another "good" one!



Source: https://medium.com/@thomas.ellyatt/bad-data-visualisation-real-life-examples-out-there-in-the-wild-eb5032329aeb

Solution: Be honest and informative!

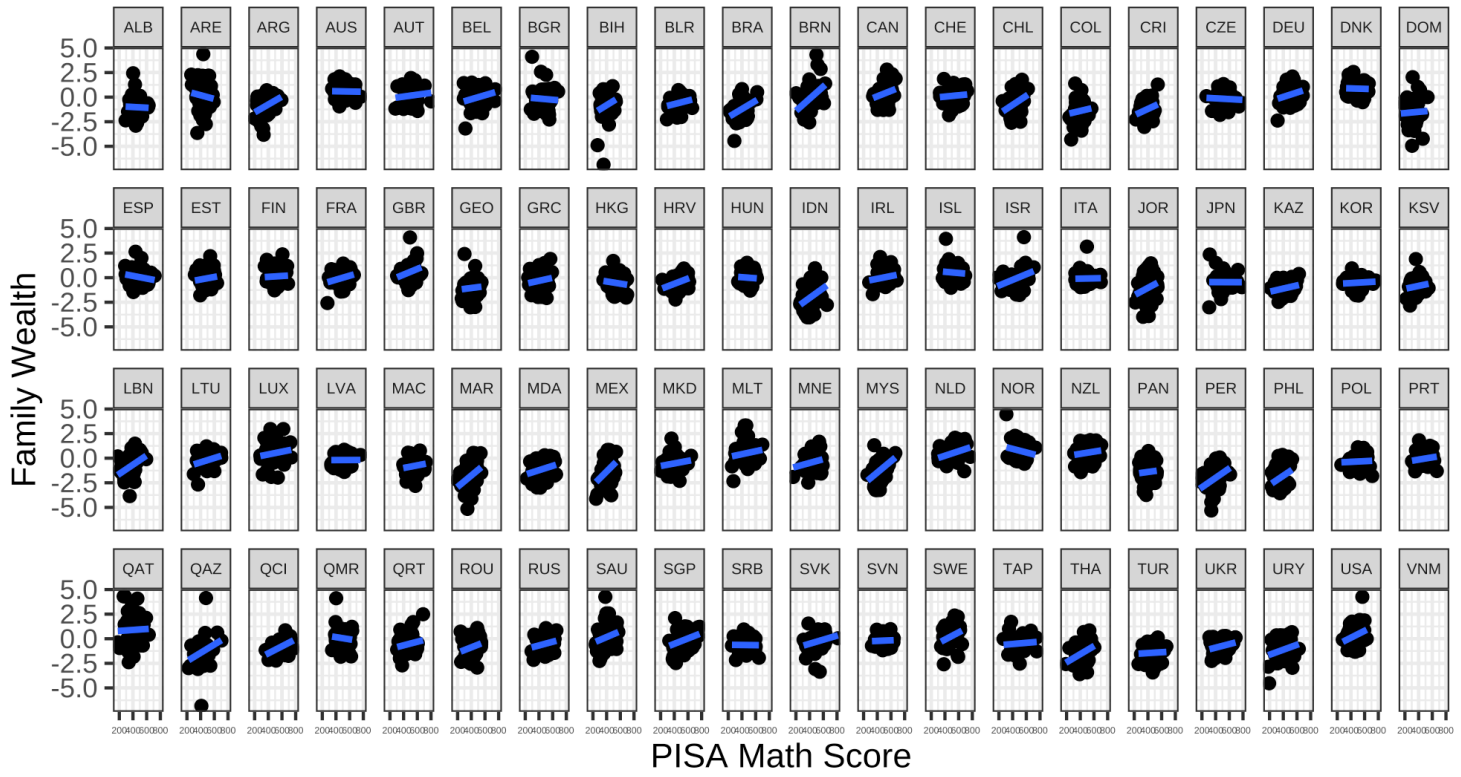


Source: https://medium.com/@thomas.ellyatt/bad-data-visualisation-real-life-examples-out-there-in-the-wild-eb5032329aeb]

3. Muddy the plot

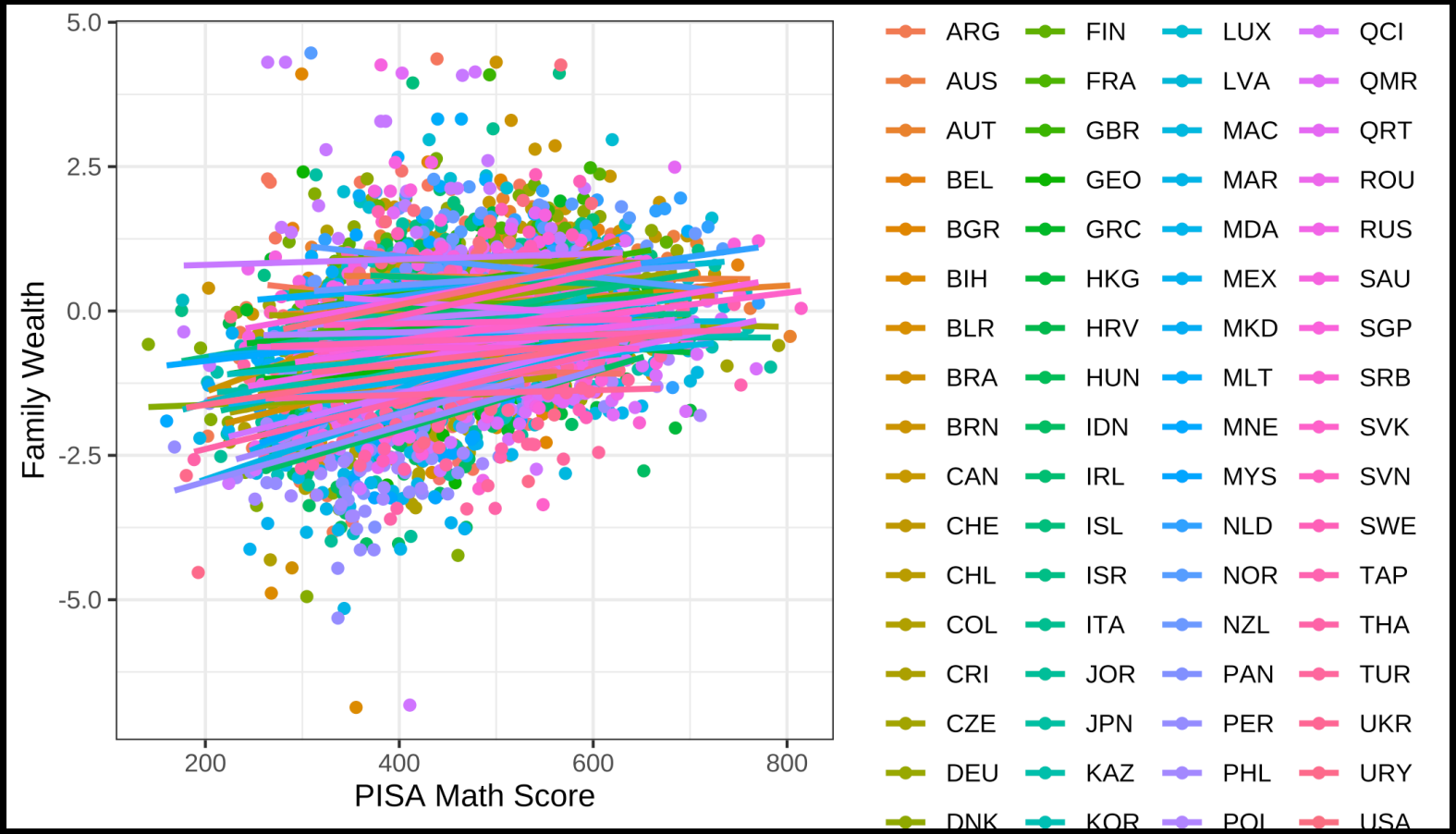
What story does this tell?

`facet_wrap(~country)`:



What about this?

`aes(x = math, y = wealth, color = country)`:



Solution: Think about data-ink ratio!

$$\text{Data-ink ratio} = \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

$$= \text{proportion of a graphic's ink devoted to the non-redundant display of data-information}$$

$$= 1.0 - \text{proportion of a graphic that can be erased}$$

Source: https://infovis-wiki.net/wiki/Data-Ink_Ratio

Thank you!