# Scatter plots

## DATA VISUALIZATION FOR EVERYONE

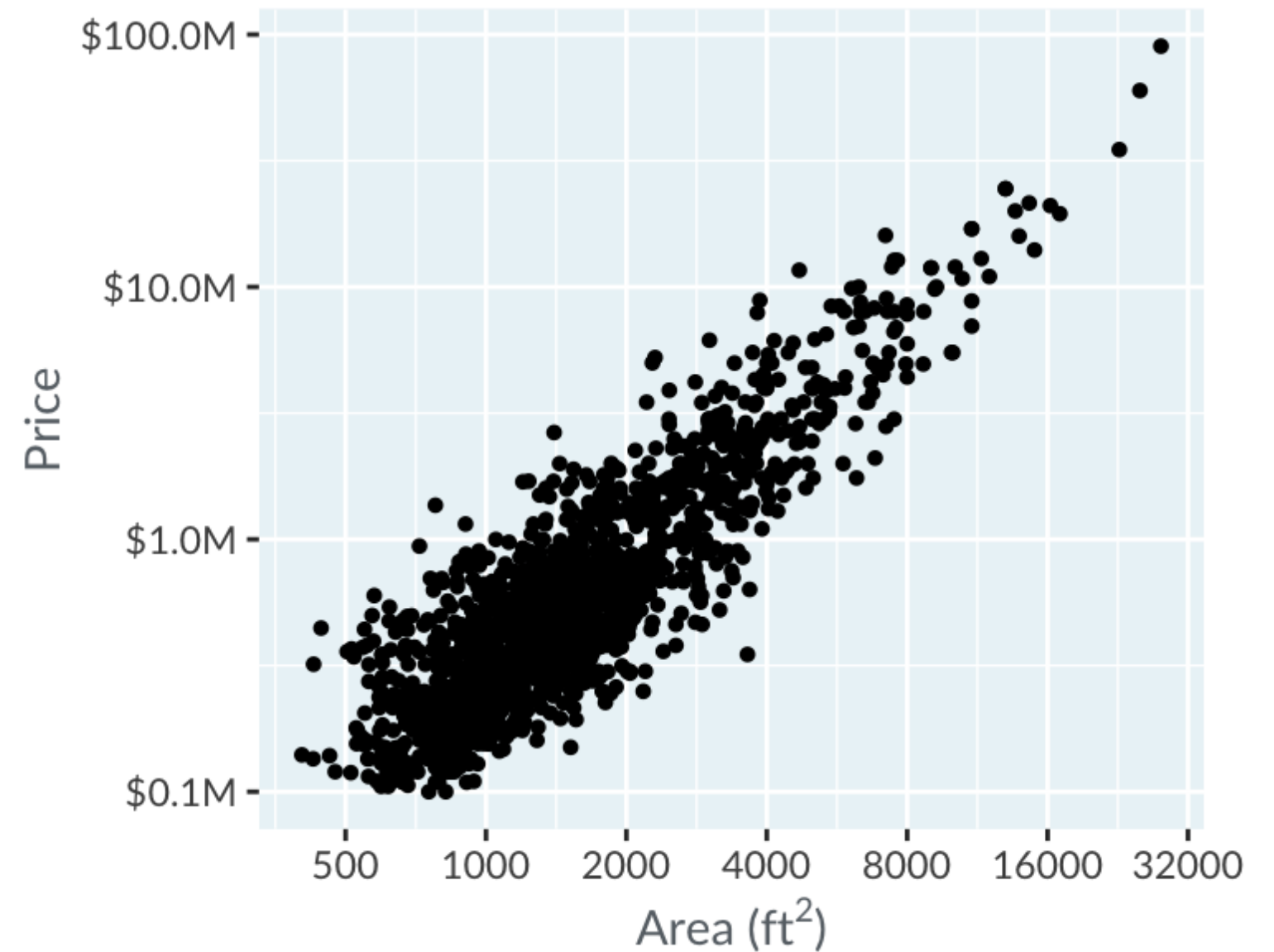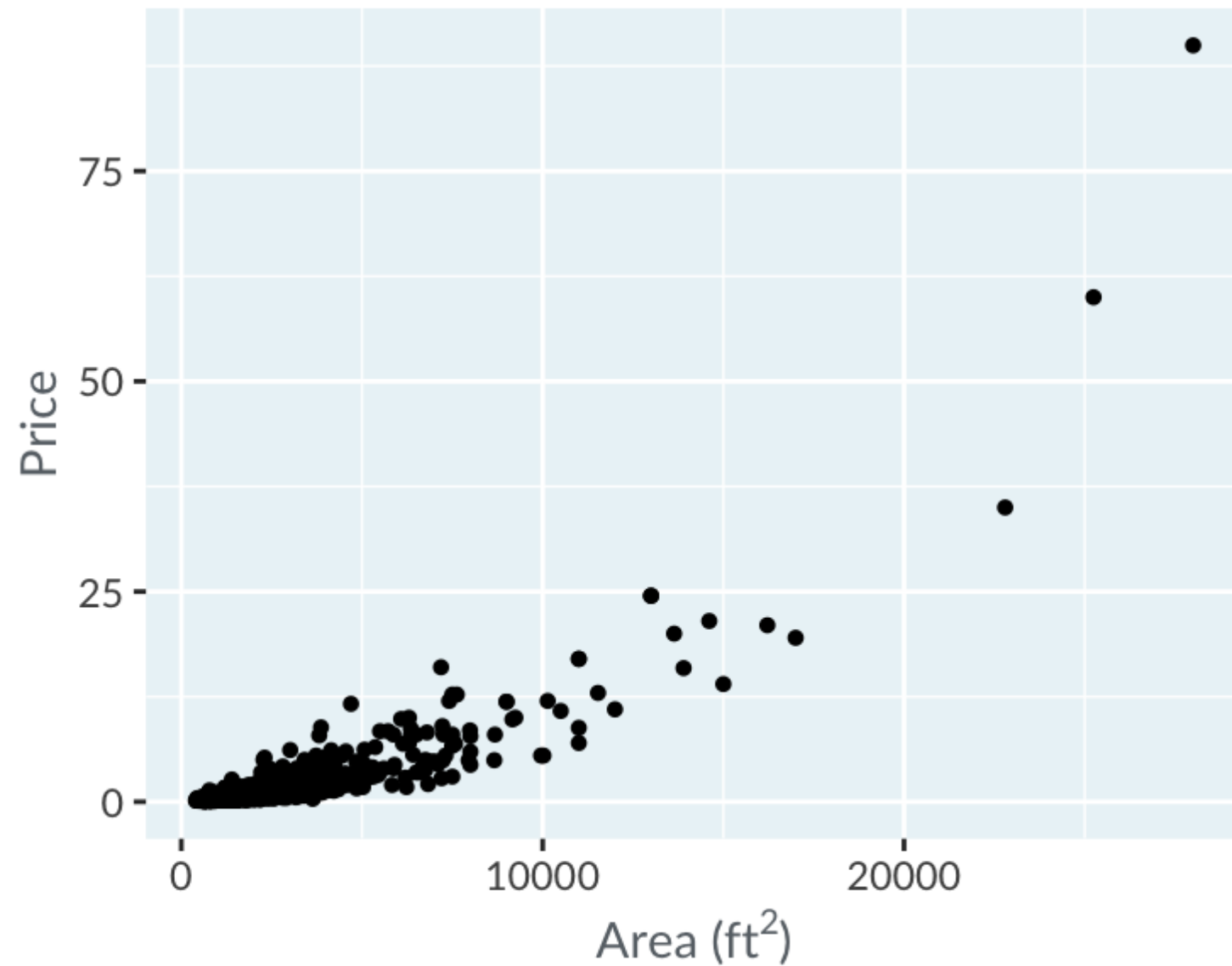**Richie Cotton**
Curriculum Architect at DataCamp

# When should you use a scatter plot?

1. You have two continuous variables.

2. You want to answer questions about the relationship between the two variables.
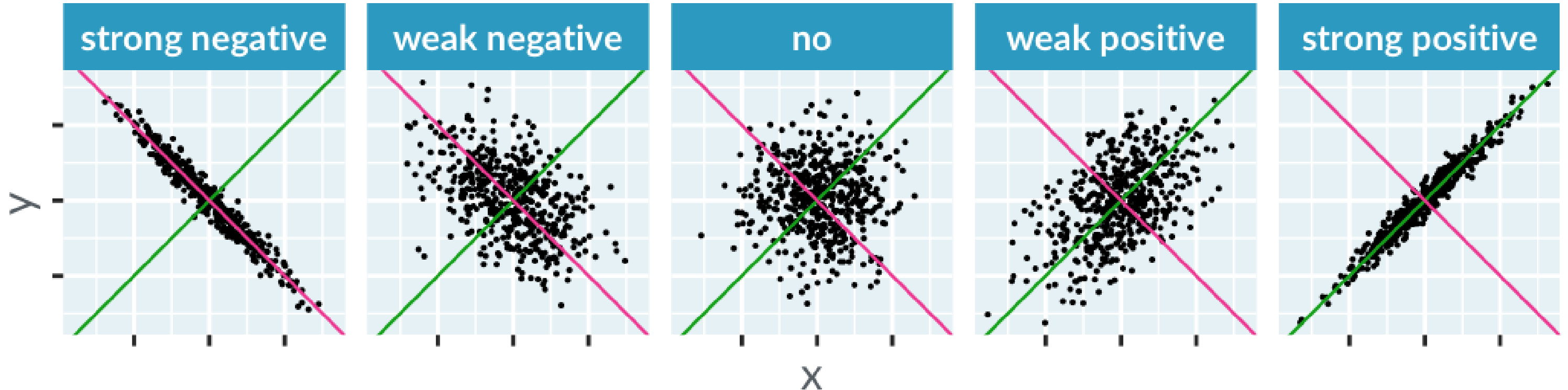
# Los Angeles County home prices

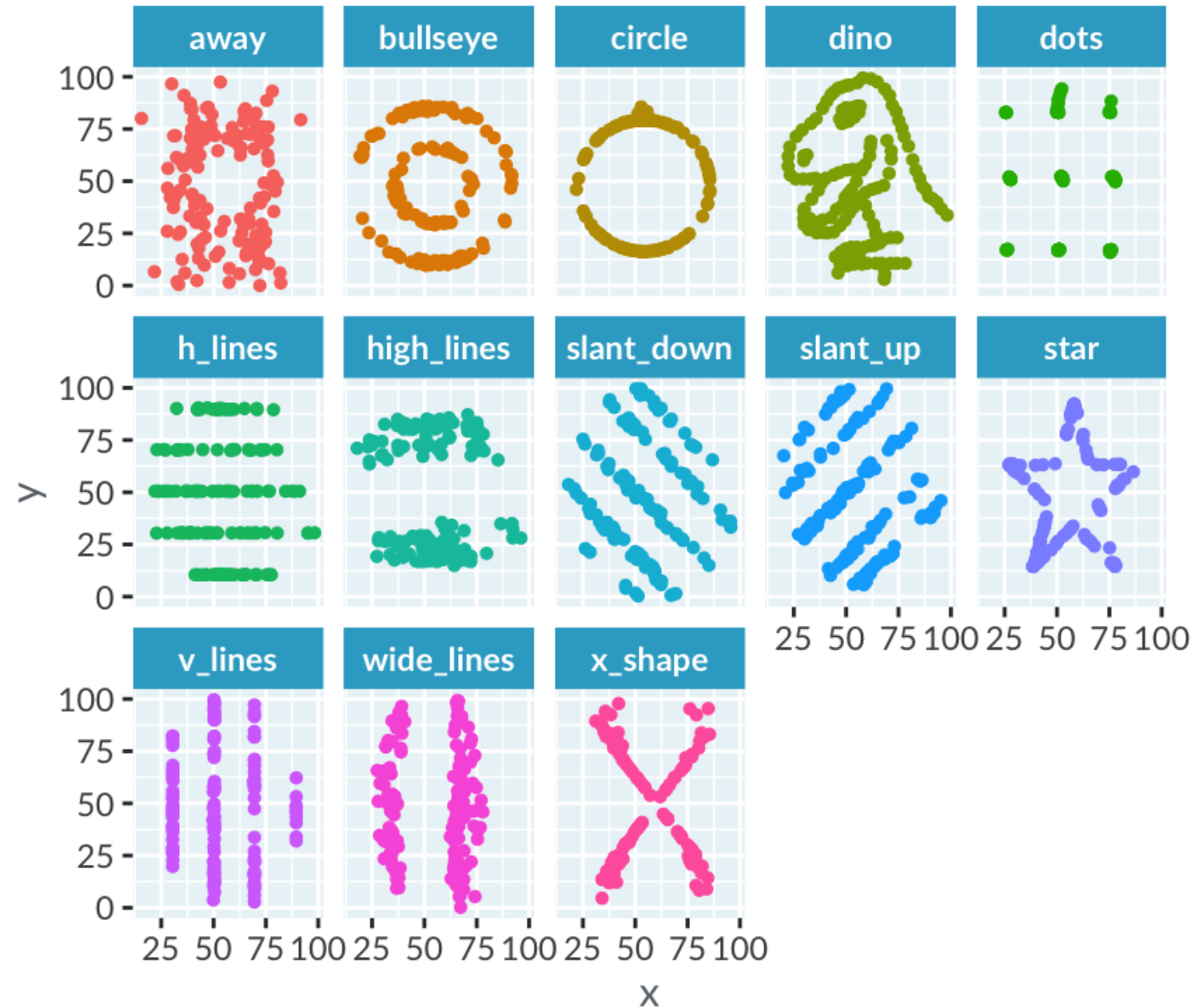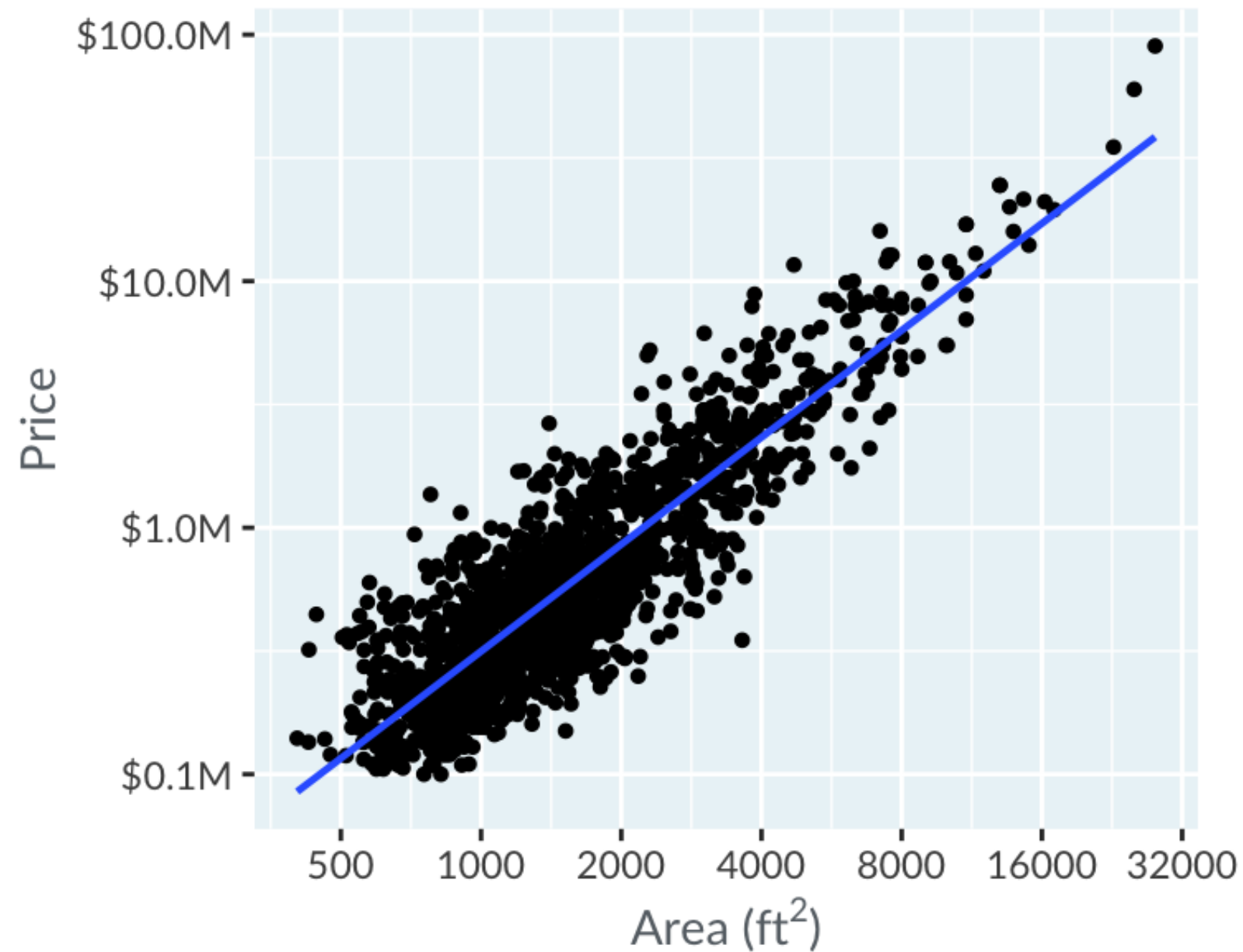| city | n_beds | price_musd | area_sqft |
|---|---|---|---|
| Long Beach | 1 | 0.3250 | 846 |
| Beverly Hills | 3 | 2.1950 | 2930 |
| Santa Monica | 2 | 0.5740 | 1037 |
| Santa Monica | 1 | 0.5990 | 576 |
| Beverly Hills | 5 | 3.9500 | 5600 |
| Long Beach | 4 | 0.2999 | 1571 |
| Westwood | 3 | 0.6950 | 1913 |

# Prices vs. area

# Correlation

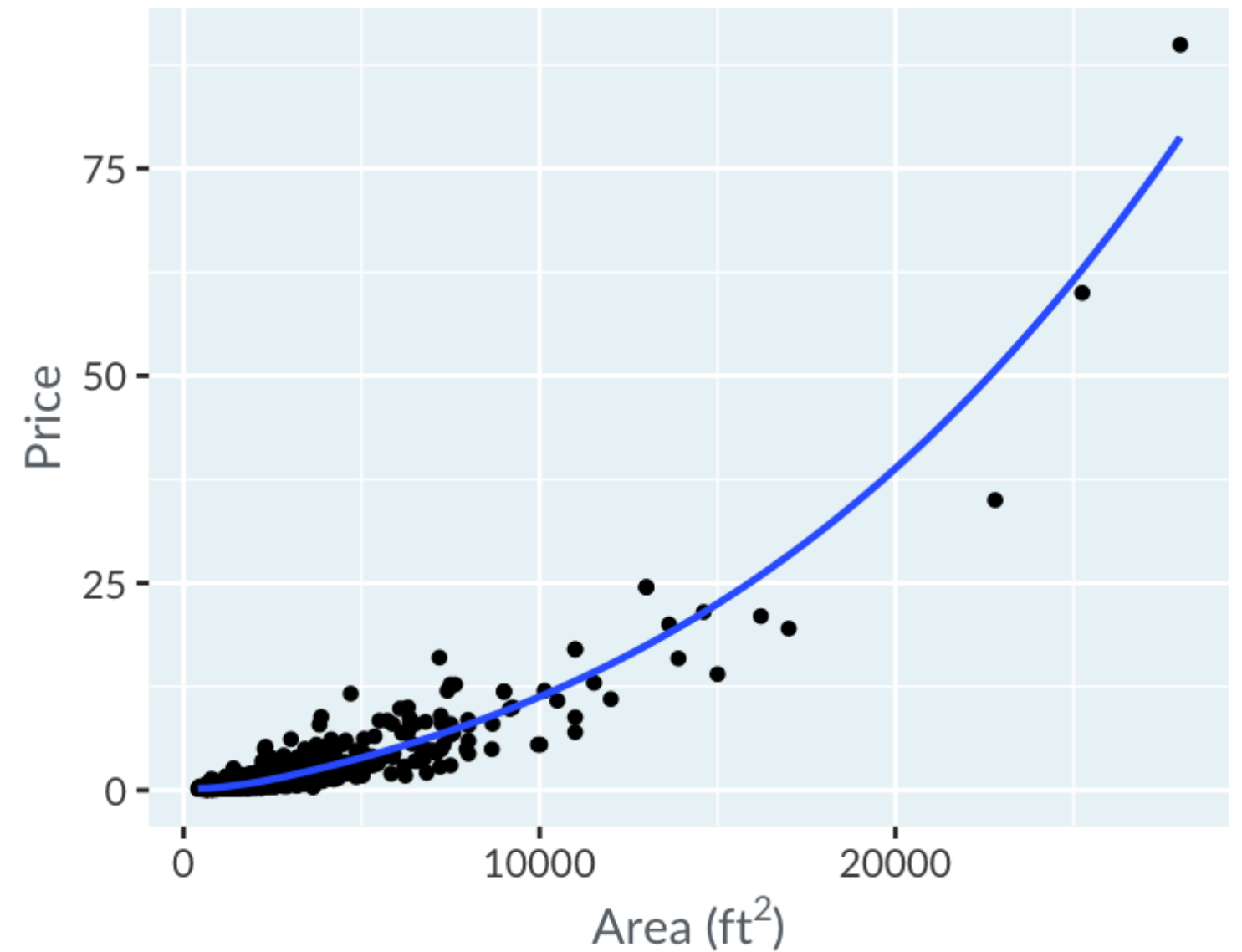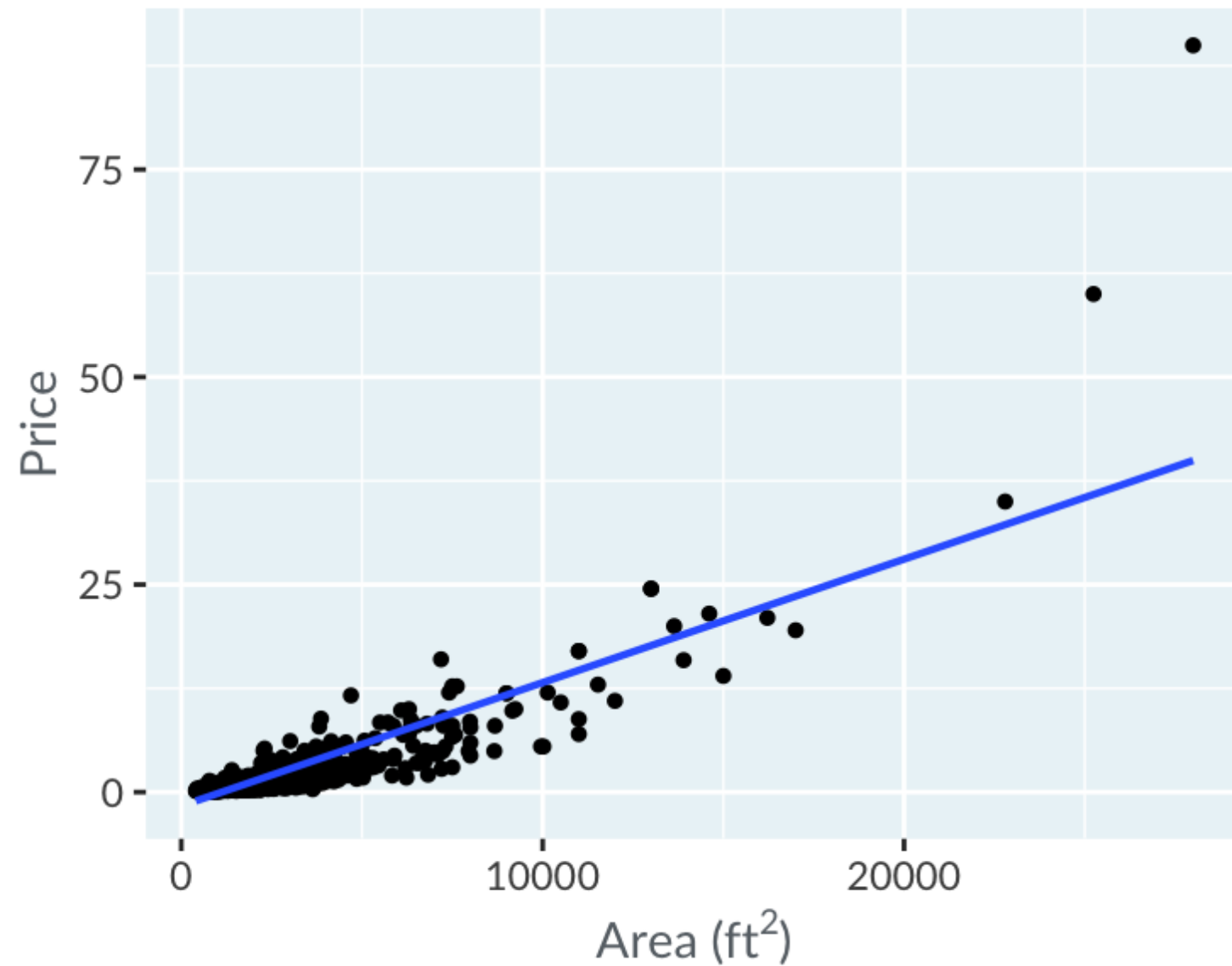How close are you to being able to fit a straight line through the points?

# Sometimes correlation isn't helpful

# Adding trend lines

# Adding smooth trend lines

# Let's practice!
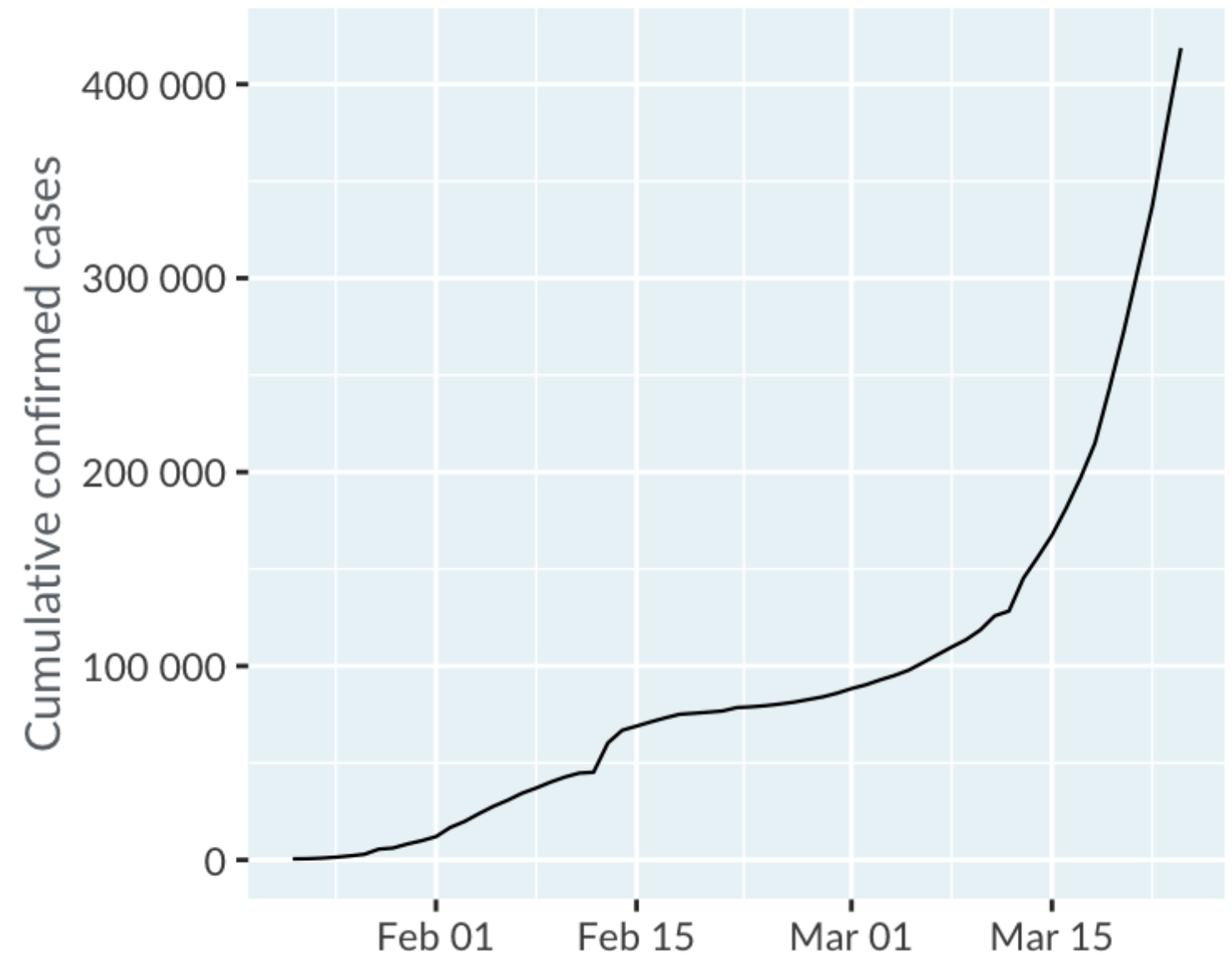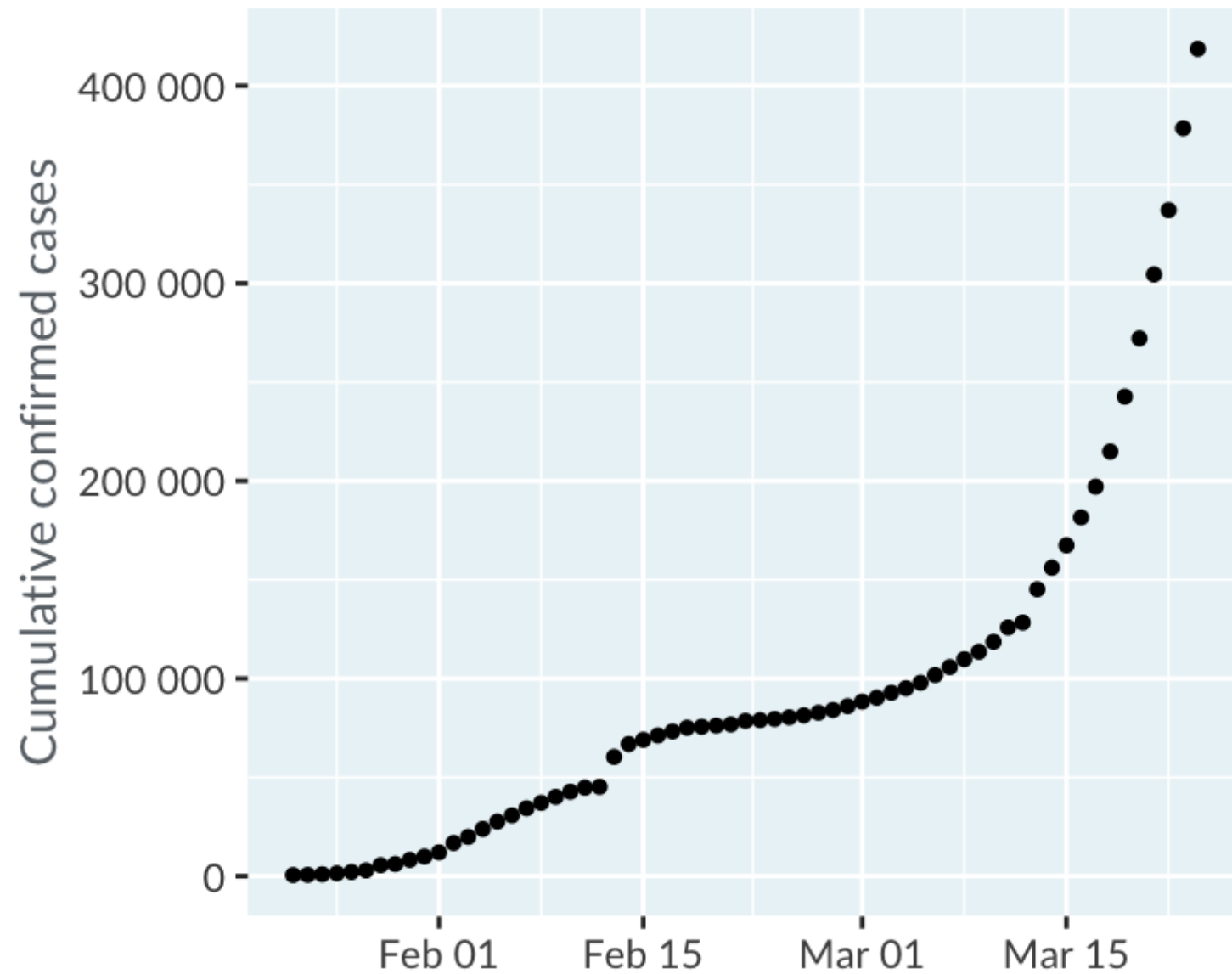
## DATA VISUALIZATION FOR EVERYONE

# Line plots

## DATA VISUALIZATION FOR EVERYONE

**Richie Cotton**
Curriculum Architect at DataCamp
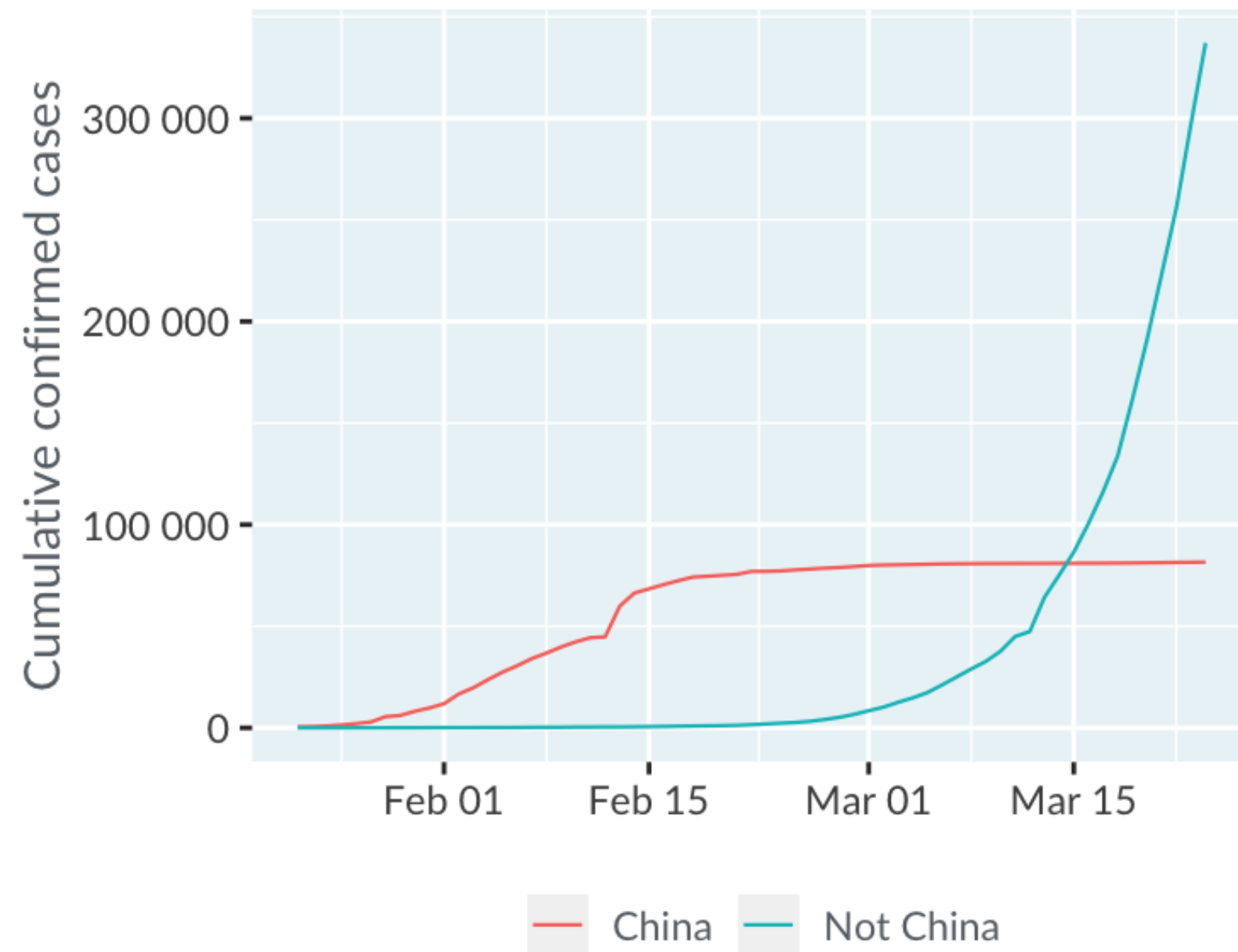
# Worldwide COVID-19 coronavirus cases
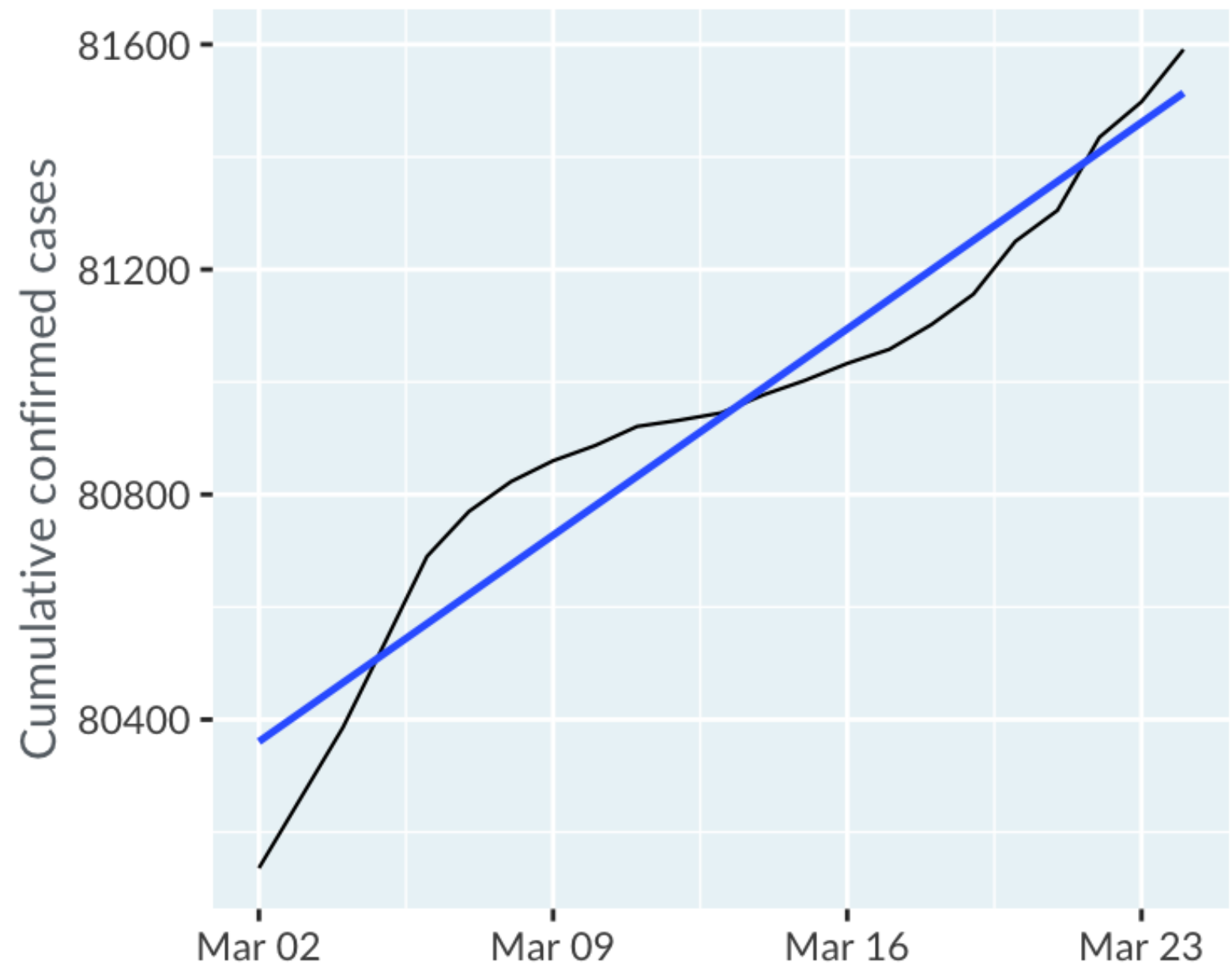
# When should you use a line plot?

1. You have two continuous variables.

2. You want to answer questions about their relationship.

3. Consecutive observations are connected somehow.
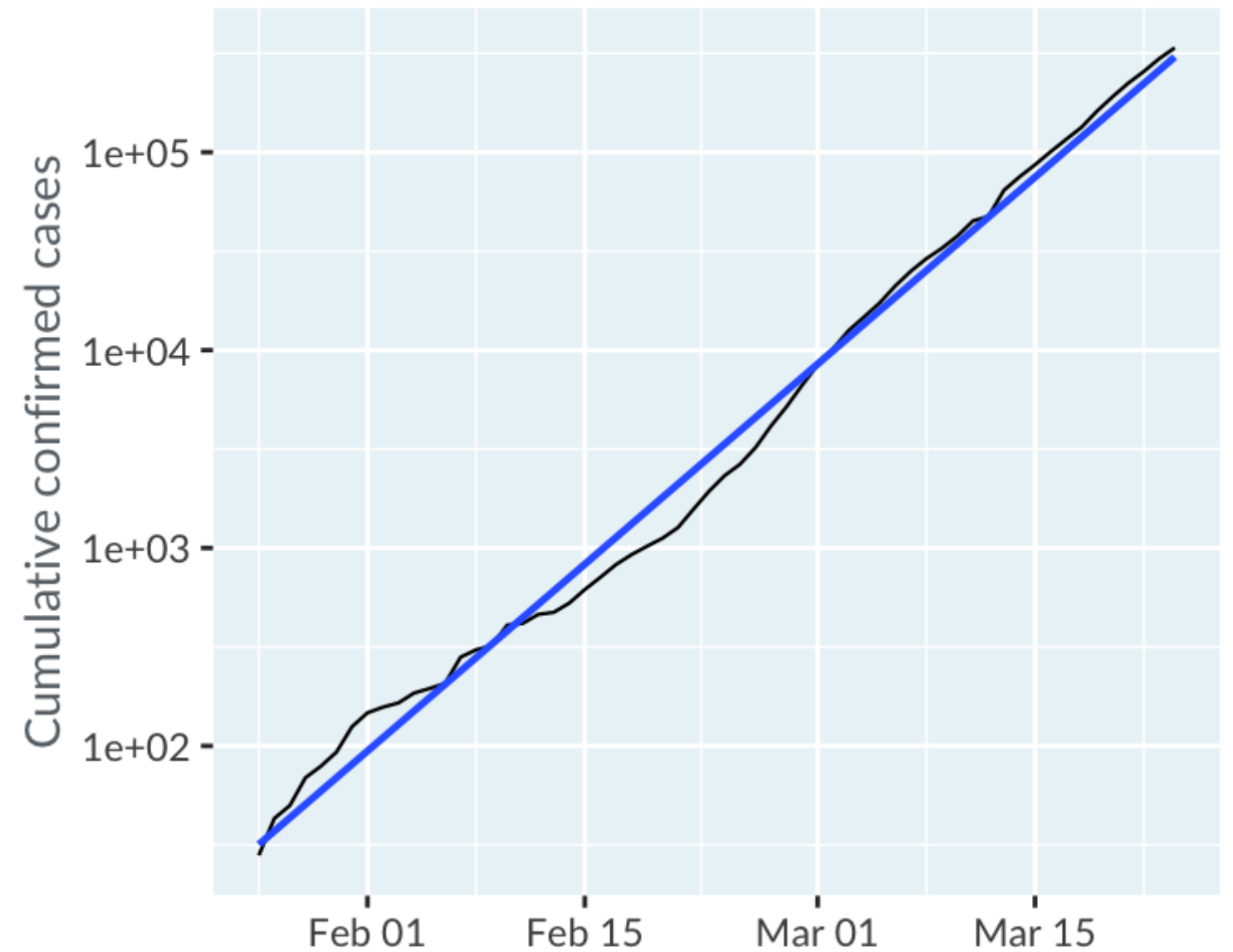
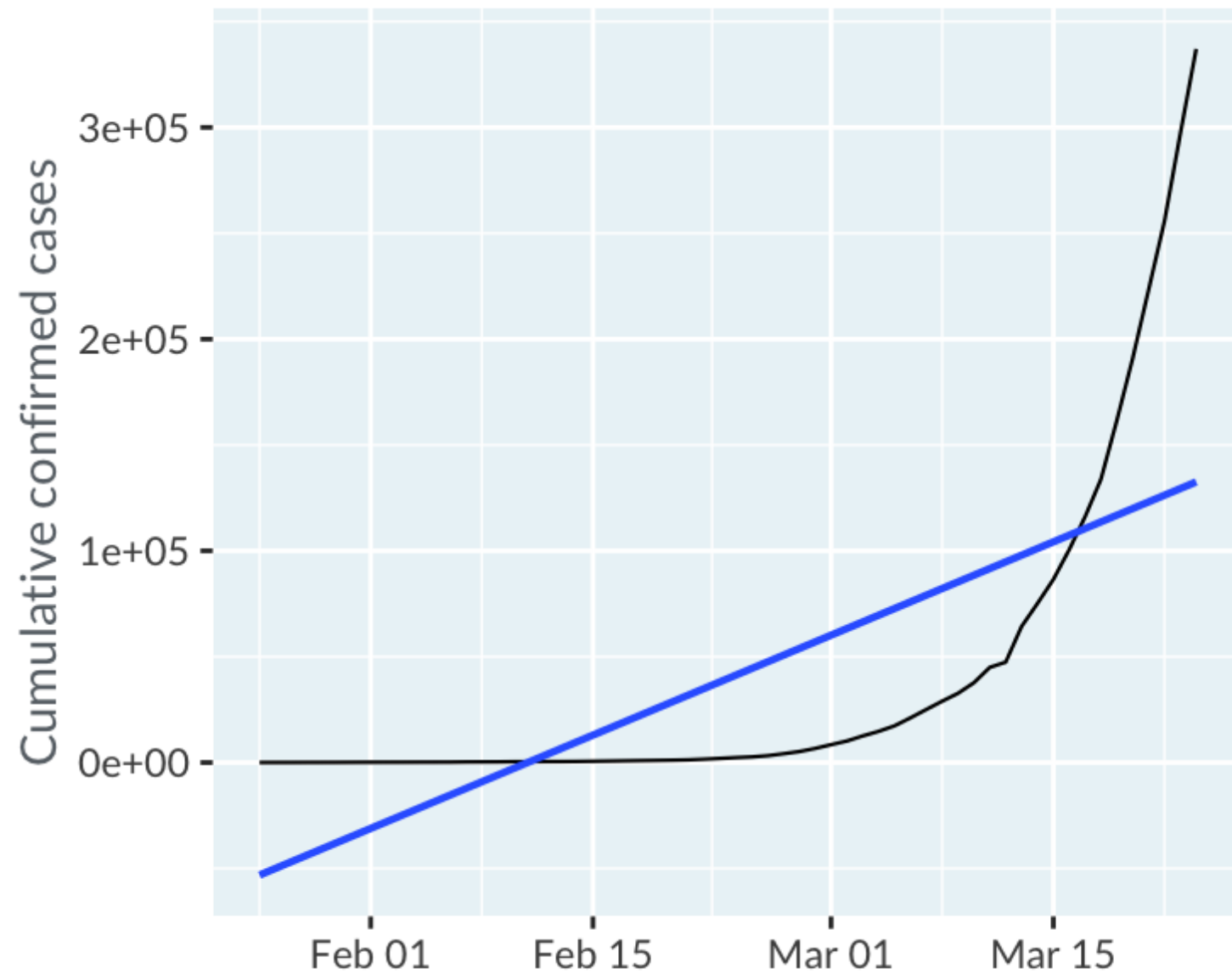Usually, but not always, the x-axis is dates or times.
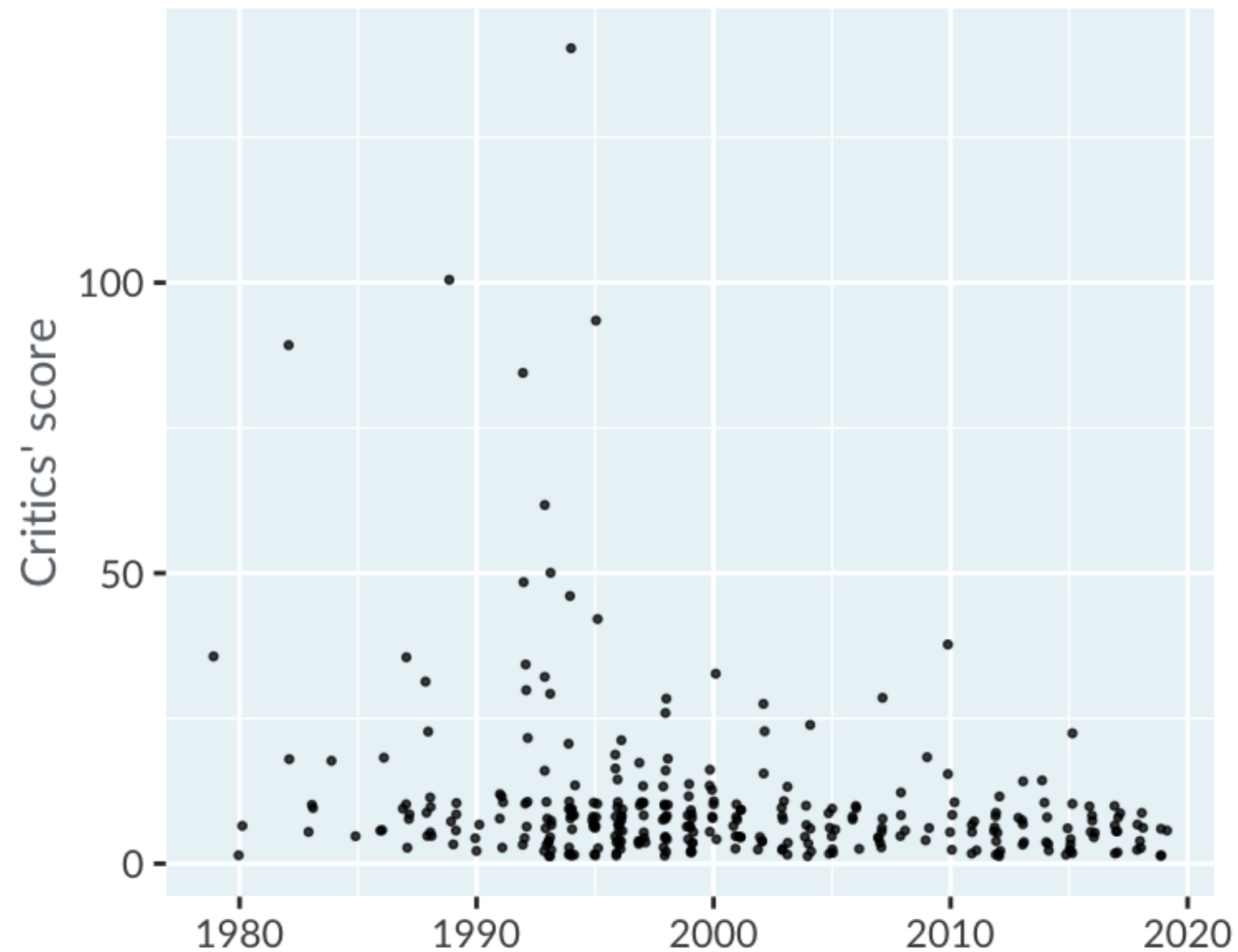
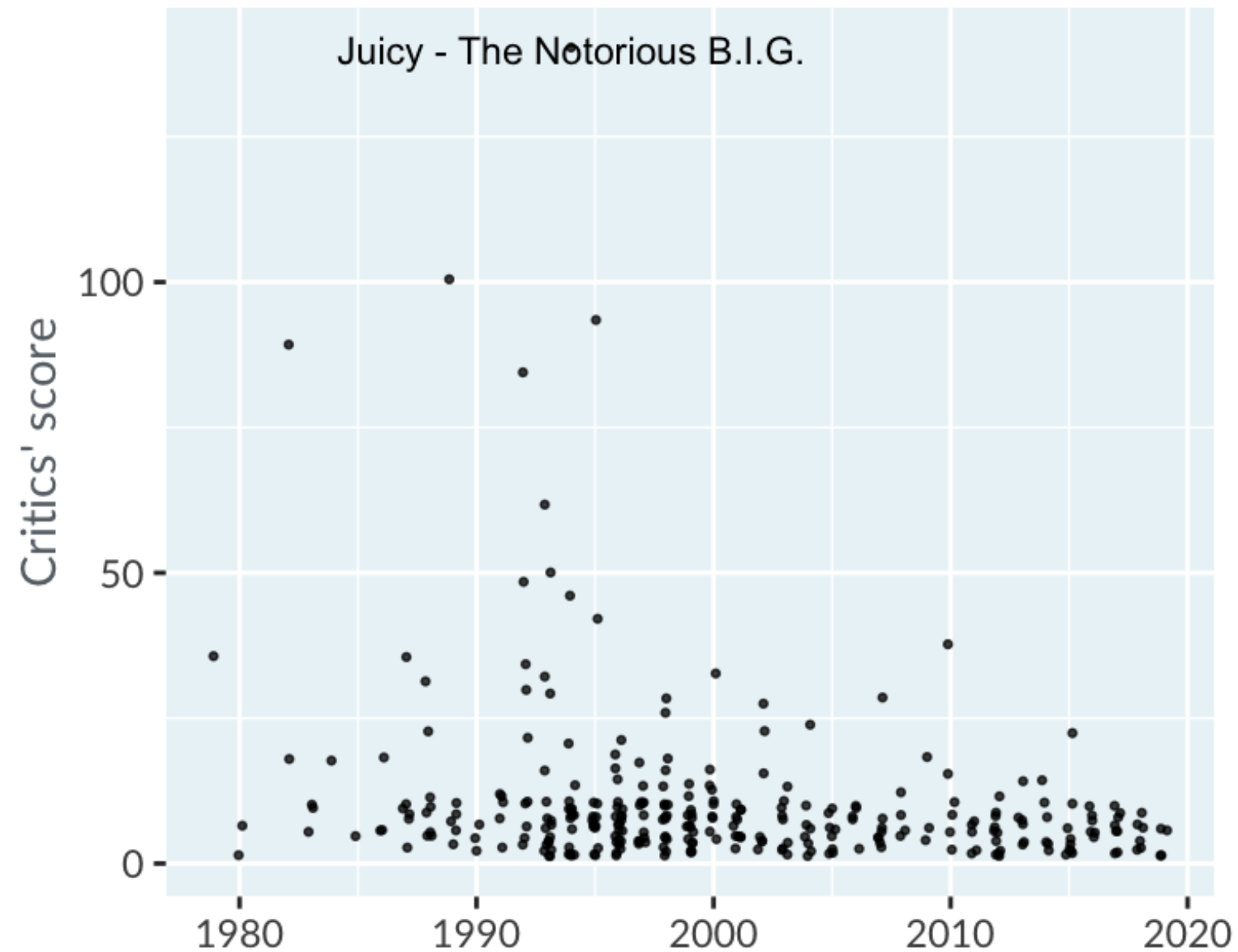# Comparing multiple lines

# Trend lines

# Trend lines + log scale

# Time x-axis doesn't always imply line plot

DATA VISUALIZATION FOR EVERYONE

# Time x-axis doesn't always imply line plot

datacamp

**DATA VISUALIZATION FOR EVERYONE**

# Time x-axis doesn't always imply line plot

datacamp

DATA VISUALIZATION FOR EVERYONE
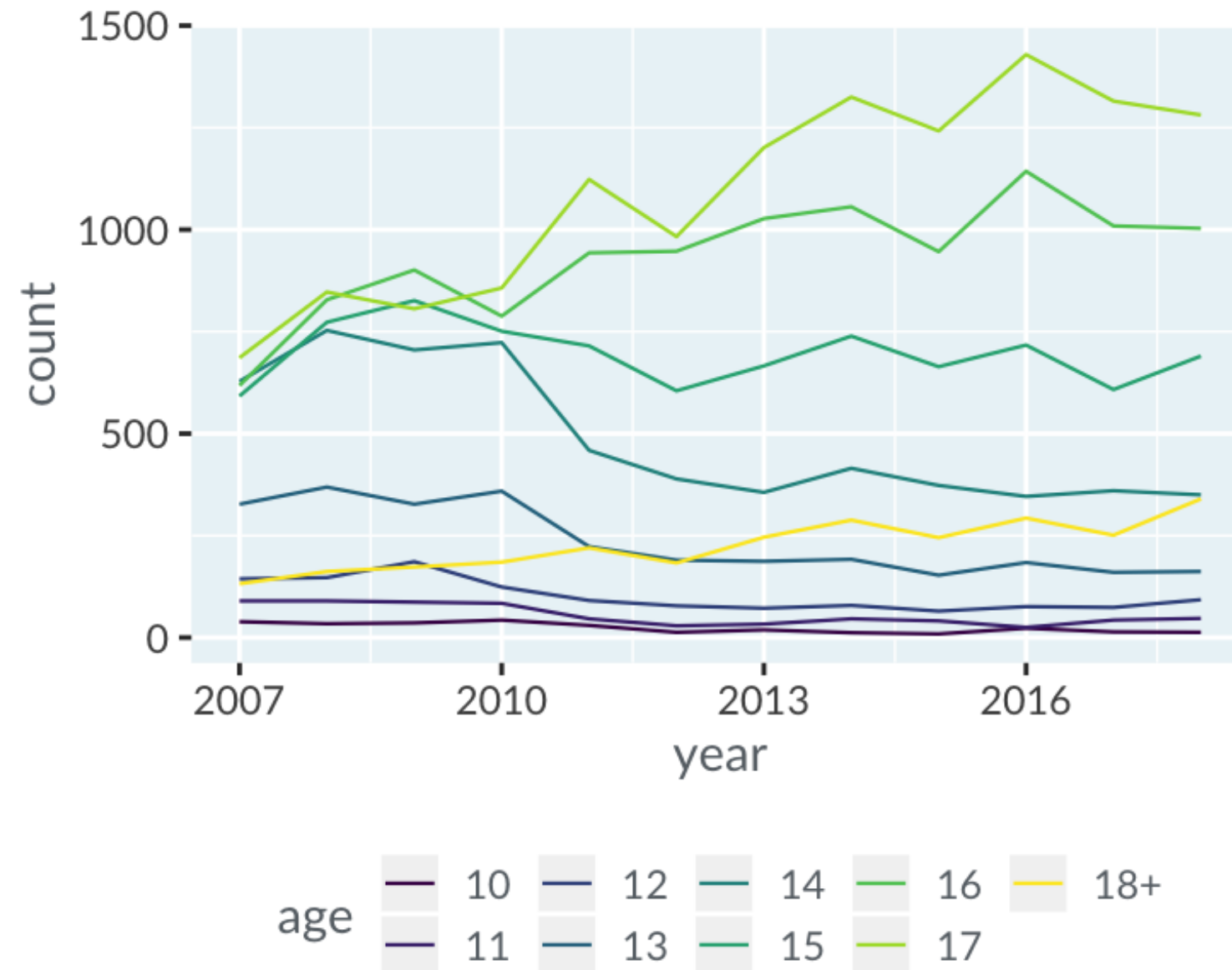
# Let's practice!

## DATA VISUALIZATION FOR EVERYONE

# Bar plots

## DATA VISUALIZATION FOR EVERYONE

**Richie Cotton**
Curriculum Architect at DataCamp

# When should you use a bar plot?

Most common cases:

1. You have a categorical variable.

2. You want counts or percentages for each category.

Occasionally:

1. You want another numeric score for each category, and need to include zero in the plot.

When will I, will I be famous?
I can't answer, I can't answer that
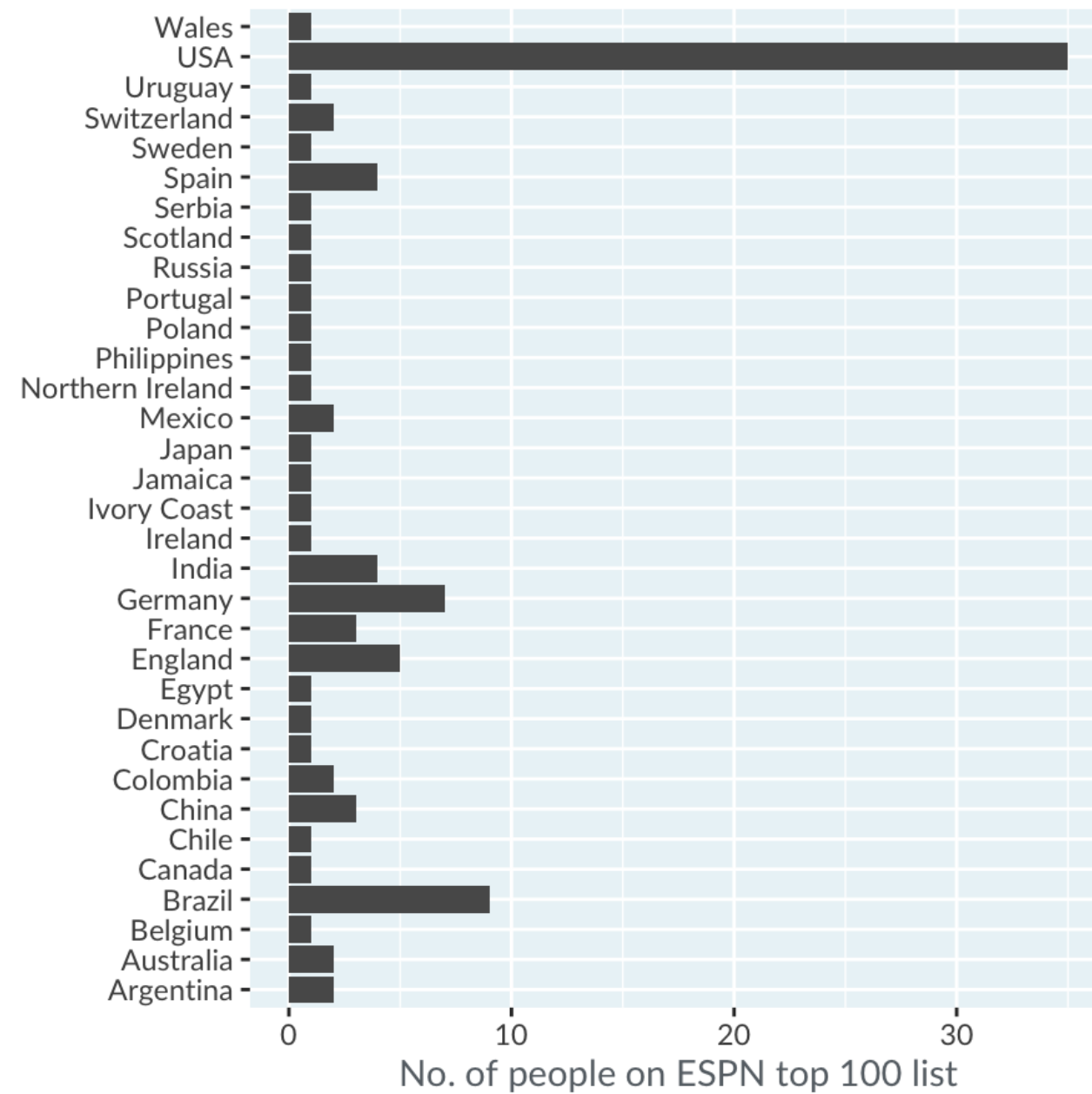
Bros – "When will I be famous?"
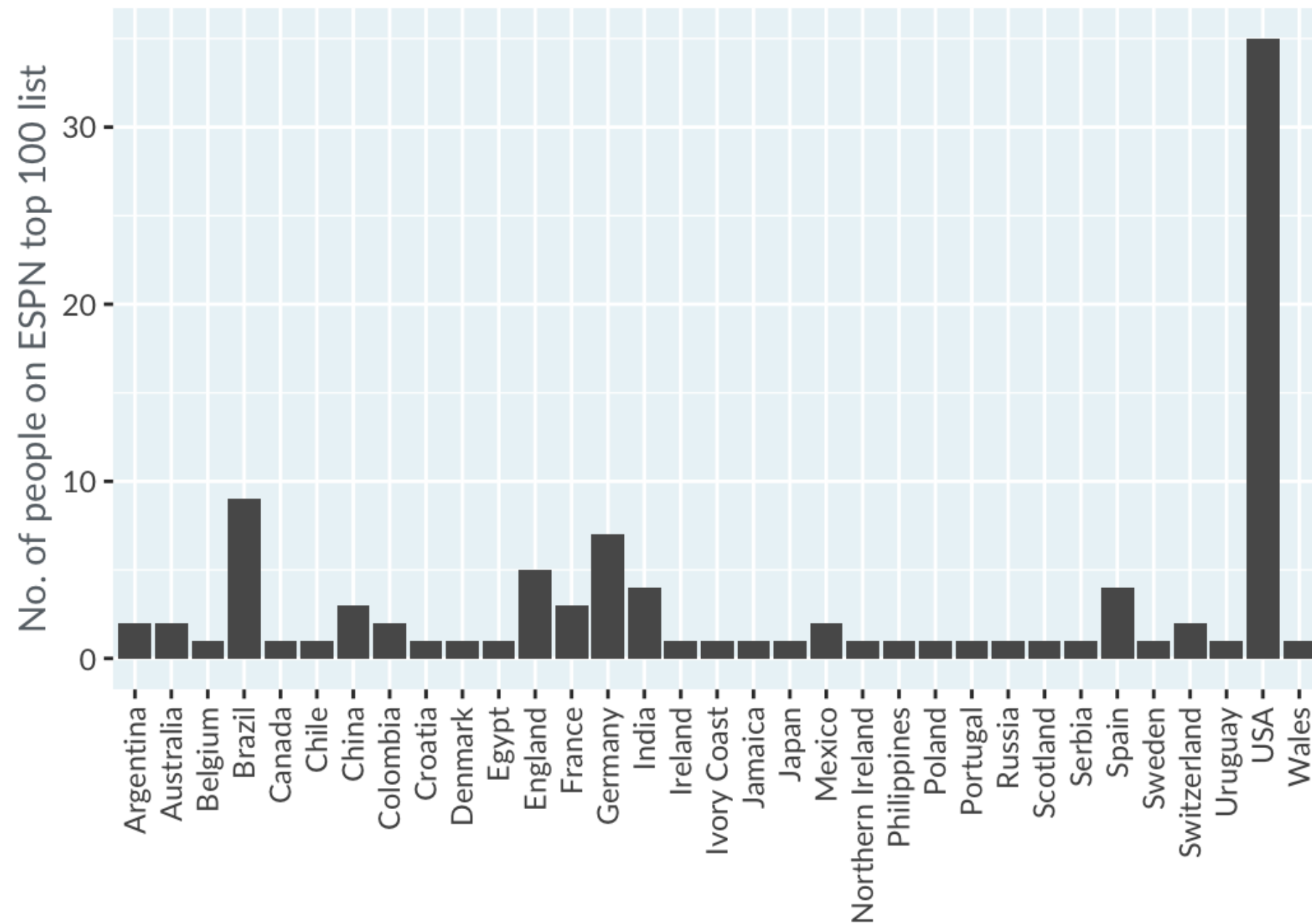
# ESPN 100 most famous athletes from 2017

| Rank | Last Name | First Name | Sport | Country |
|------|-----------|------------|-------|---------|
| 1 | Ronaldo | Cristiano | Soccer | Portugal |
| 2 | James | LeBron | Basketball | USA |
| 3 | Messi | Lionel | Soccer | Argentina |
| 4 | Federer | Roger | Tennis | Switzerland |
| 5 | Mickelson | Phil | Golf | USA |
| … | … | … | … | … |

datacamp

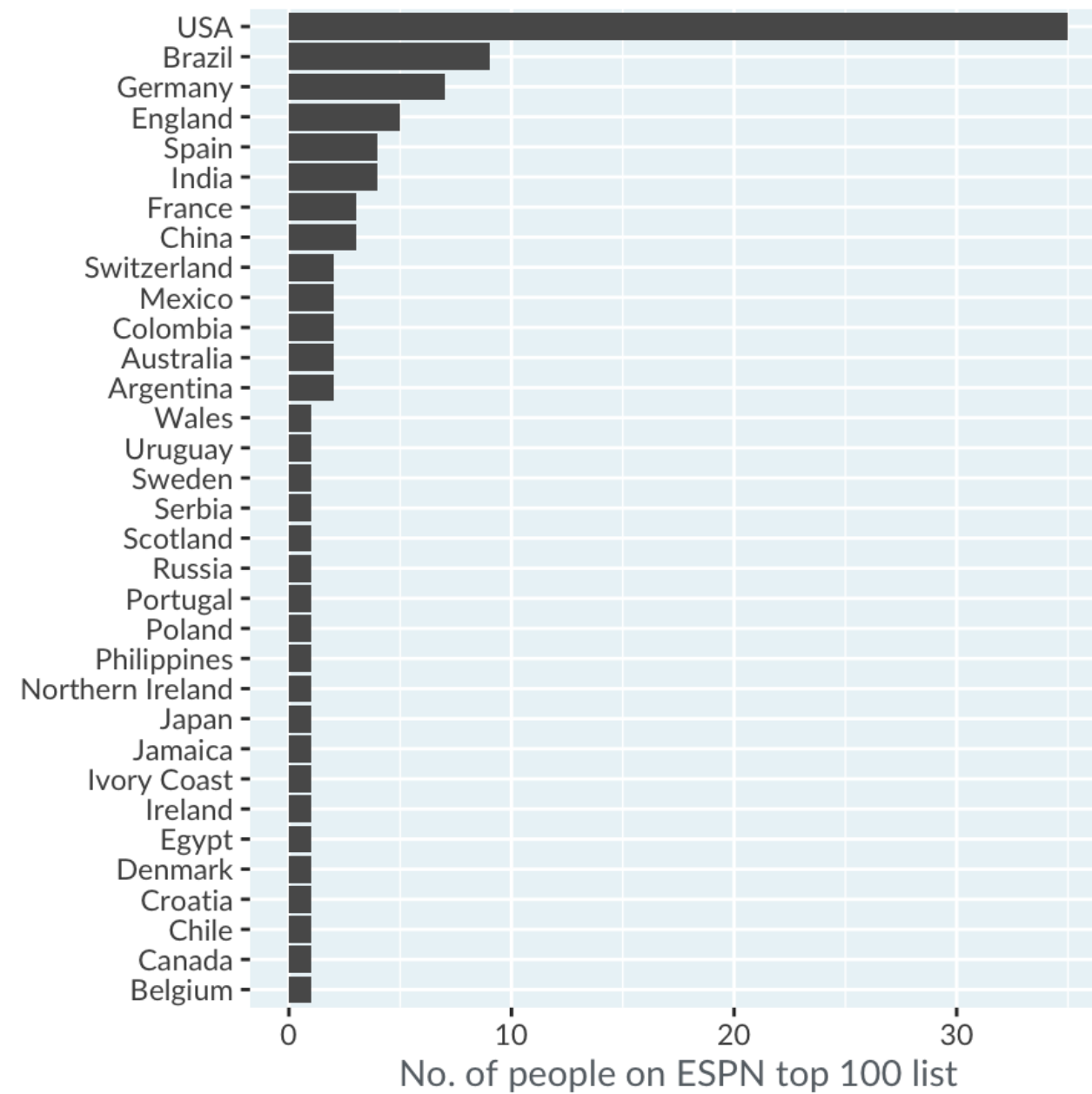**DATA VISUALIZATION FOR EVERYONE**

# Bar plot of counts by country

# Vertical bars

# Sorting by count

# Children's fruit and veg consumption

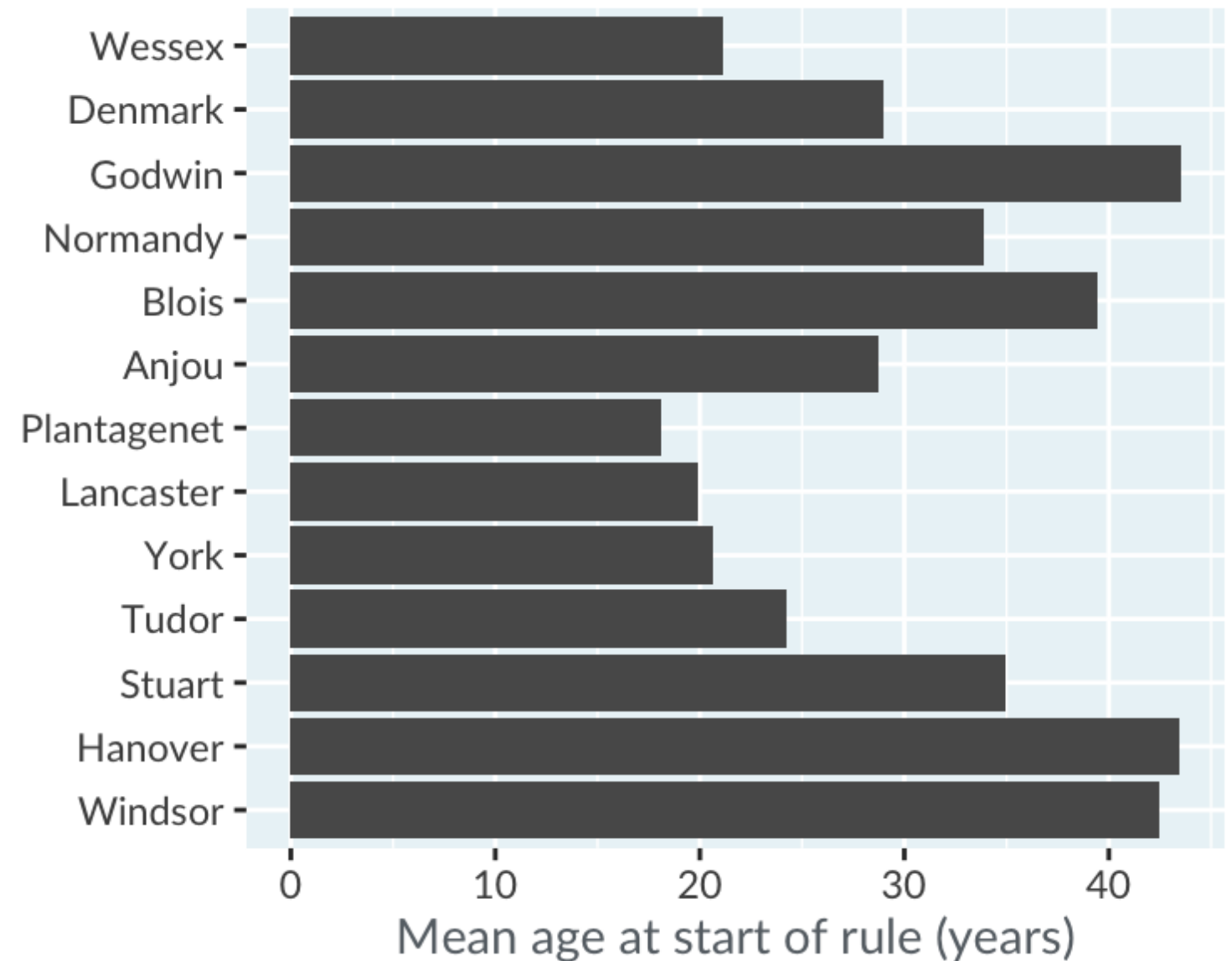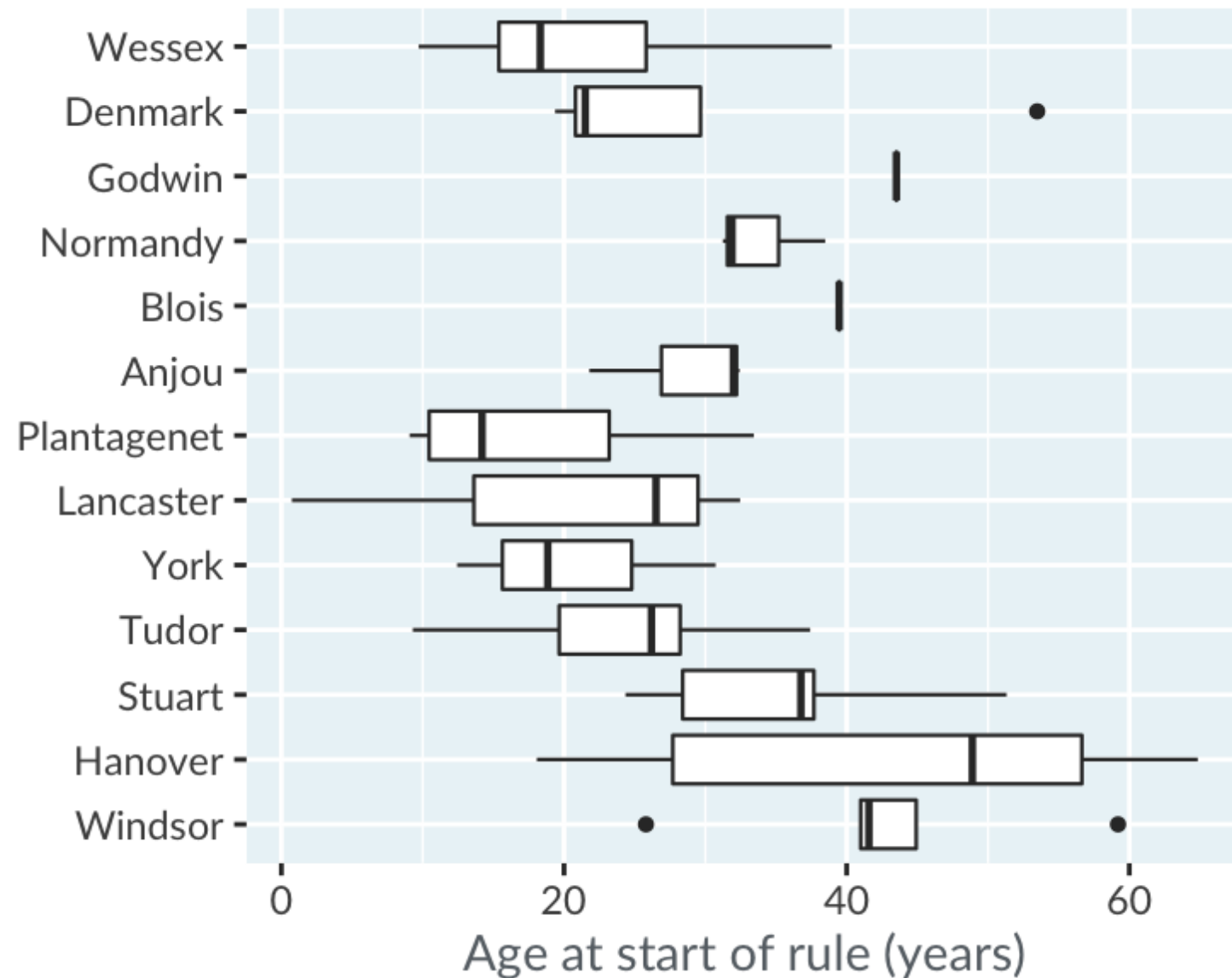| n_portions | year | pct_children |
|---|---|---|
| n = 0 | 2001 | 10.921779 |
| n < 1 | 2001 | 3.843093 |
| 1 <= n < 2 | 2001 | 23.659102 |
| ... | ... | ... |
| 4 <= n < 5 | 2018 | 12.28728 |
| 5 <= n | 2018 | 17.87497 |

# Stacking bars



DATA VISUALIZATION FOR EVERYONE
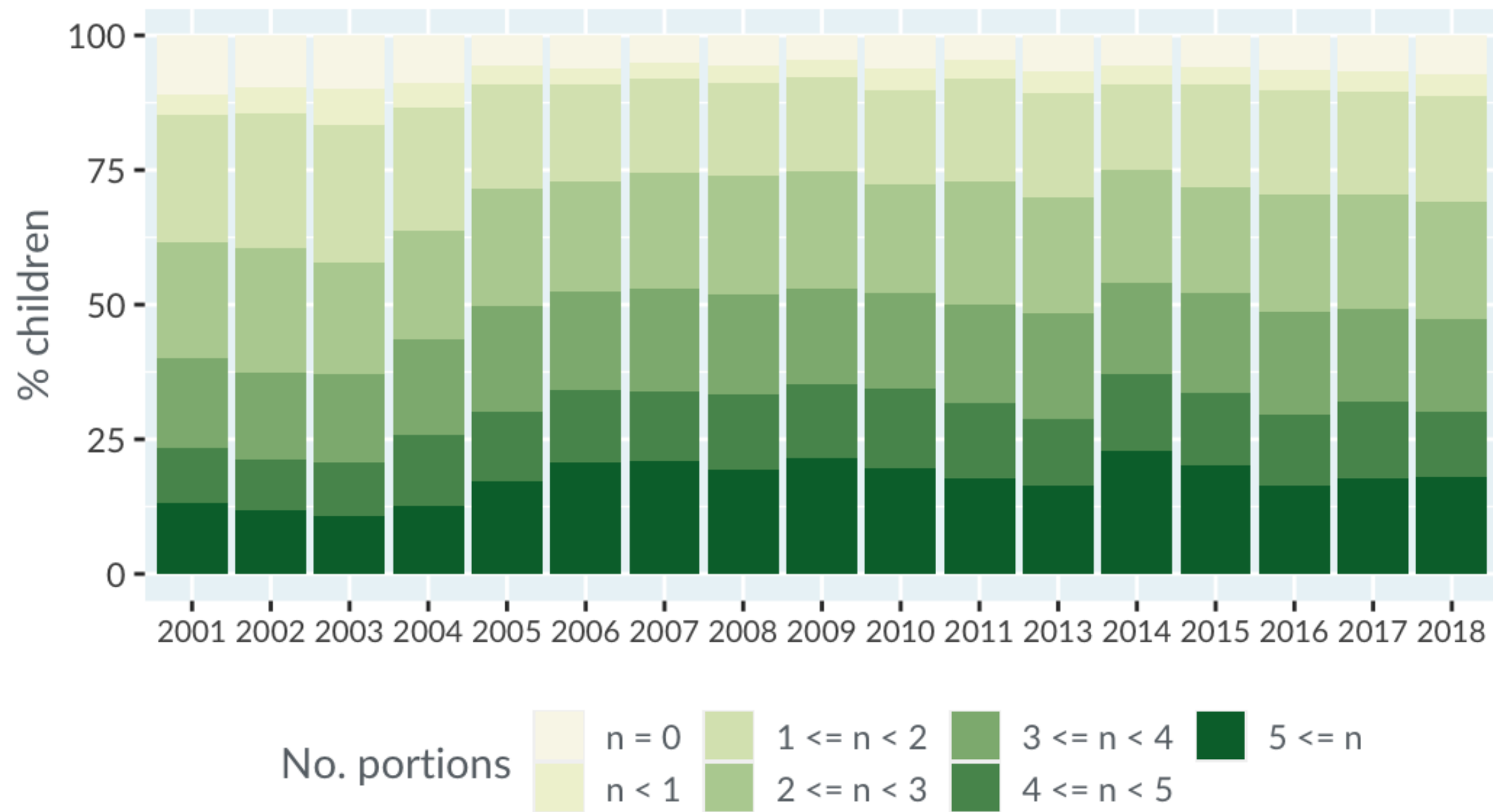
# Bar plots vs. box plots
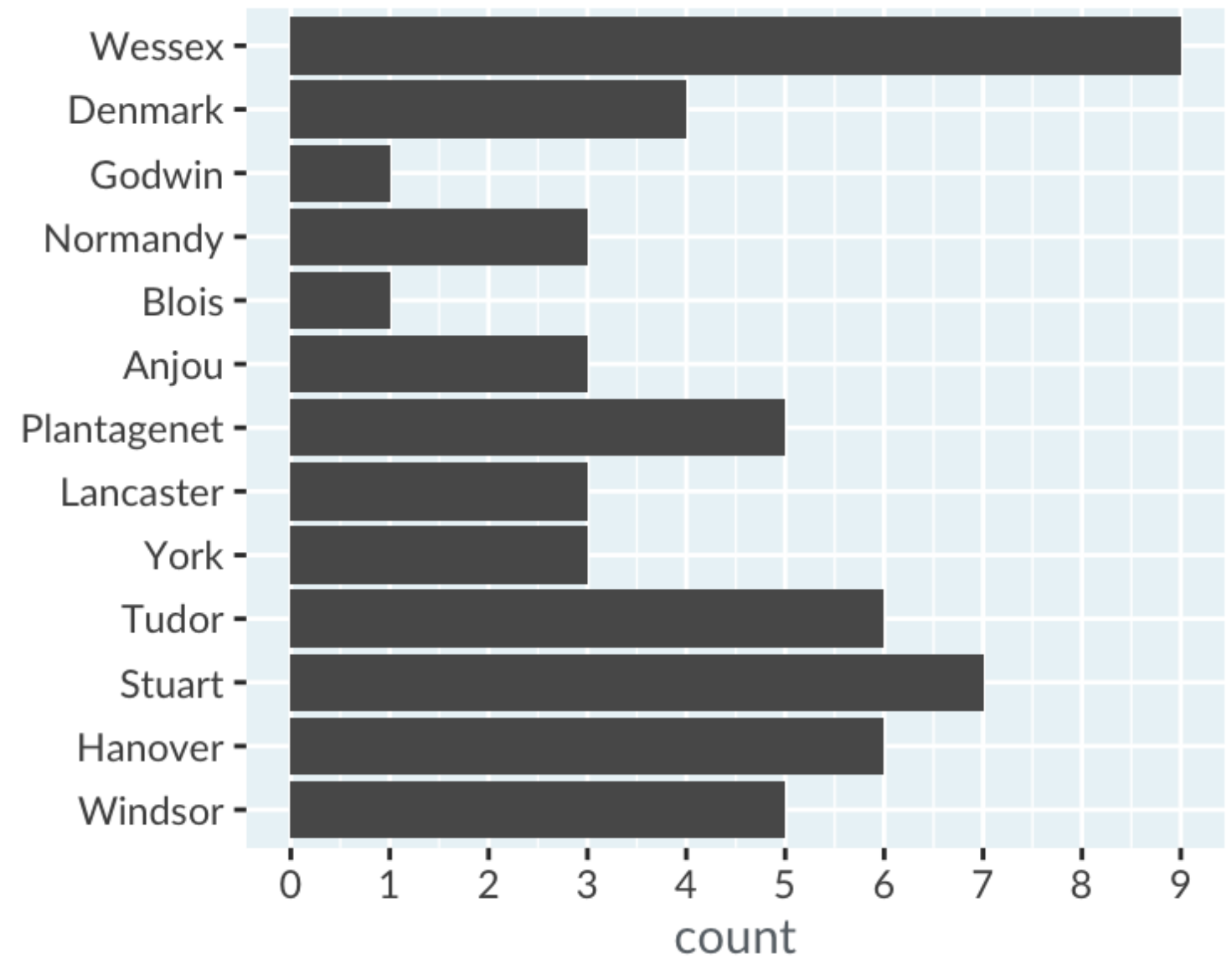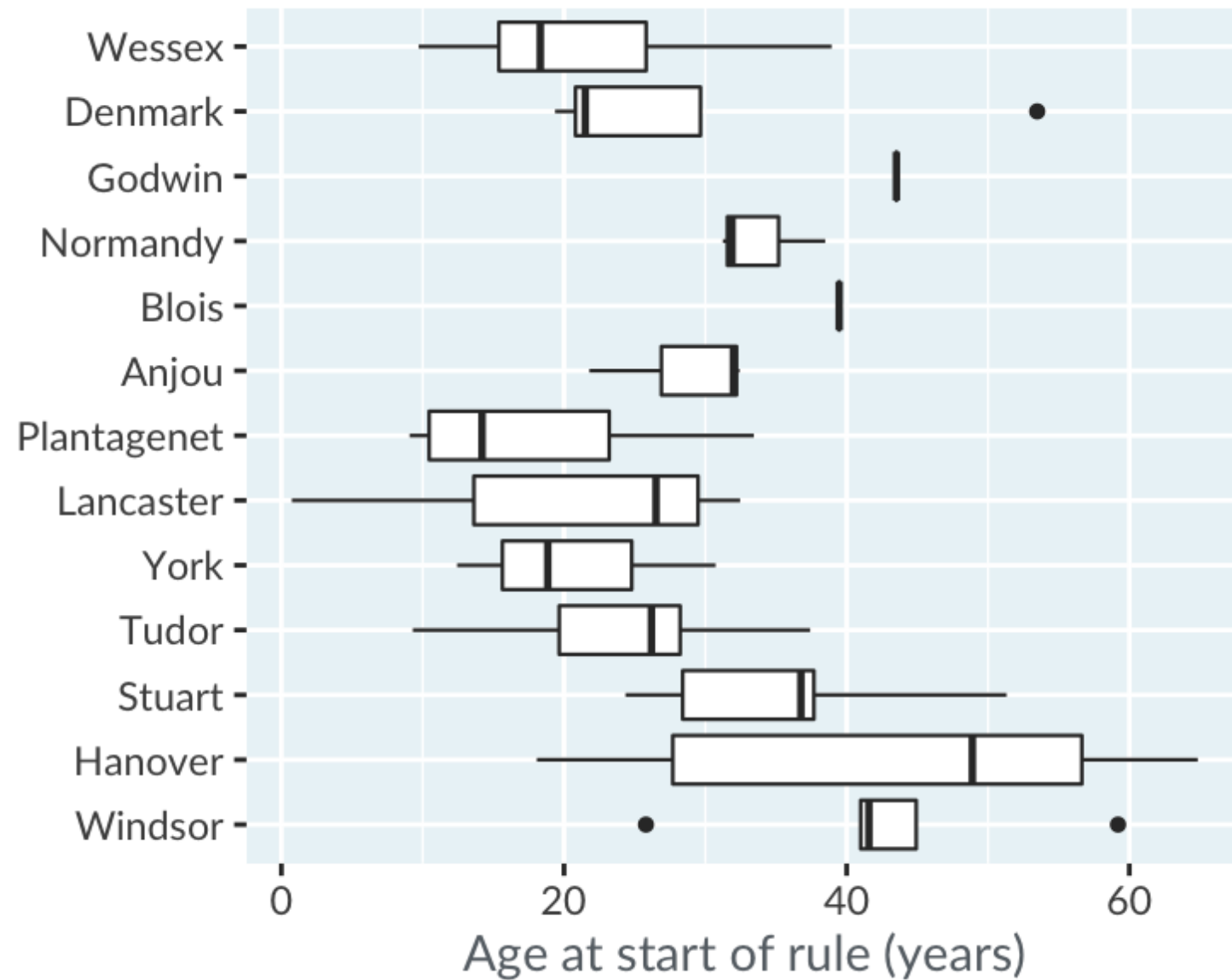
# Other metrics than counts

# Let's practice!

## DATA VISUALIZATION FOR EVERYONE

# Dot plots

## DATA VISUALIZATION FOR EVERYONE

**Richie Cotton**
Curriculum Architect at DataCamp

# When should you use a dot plot?

1. You have a categorical variable.

2. You want to display numeric scores for each category on a log scale, or

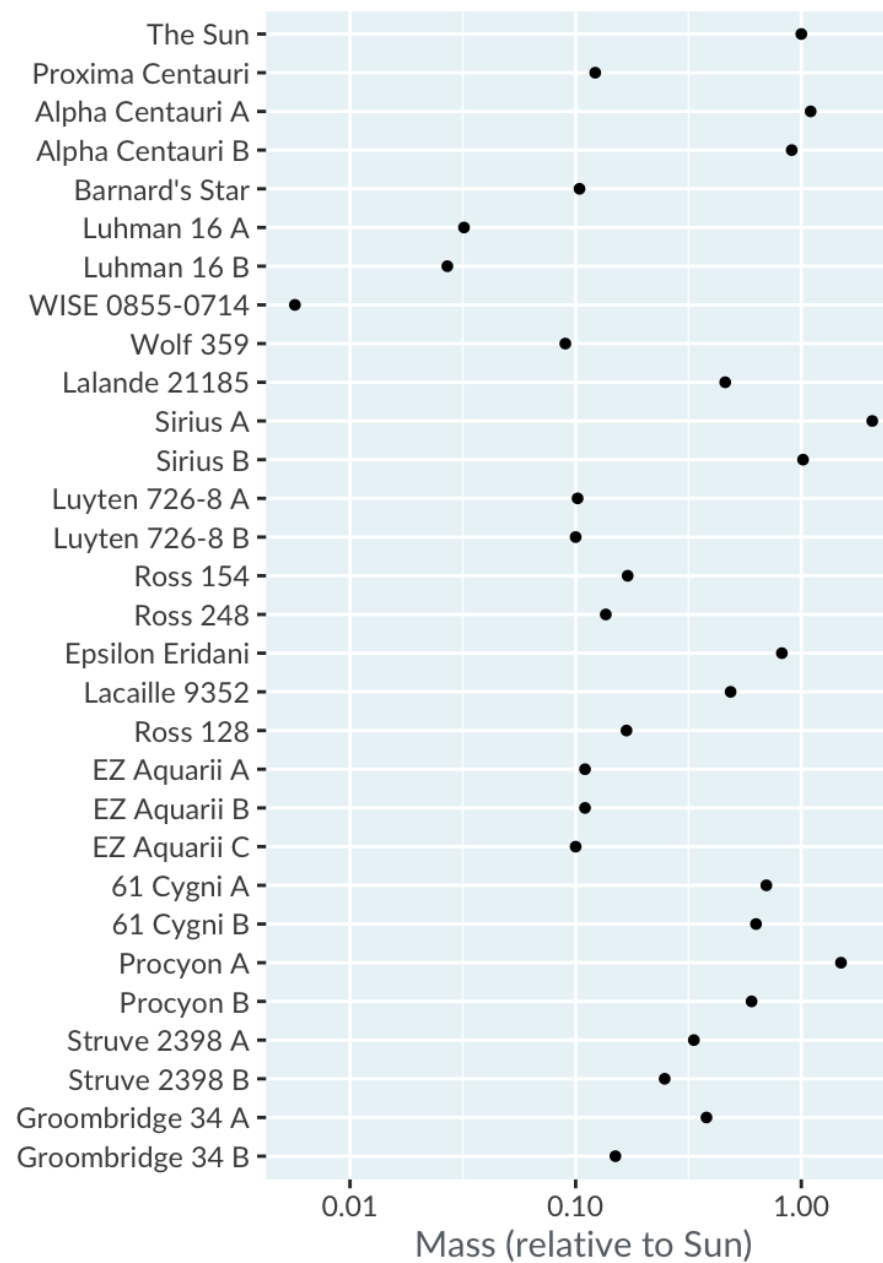3. You want to display multiple numeric scores for each category.

# Nearby stars and brown dwarfs

| star | distance_ly | mass_sm |
|---|---|---|
| The Sun | 0.0000158 | 1.0000 |
| Proxima Centauri | 4.2441000 | 0.1221 |
| Alpha Centauri A | 4.3650000 | 1.1000 |
| Alpha Centauri B | 4.3650000 | 0.9070 |
| Barnard's Star | 5.9577000 | 0.1040 |
| Luhman 16 A | 6.5029000 | 0.0320 |
| ... | ... | ... |

# Bar plot vs. dot plot
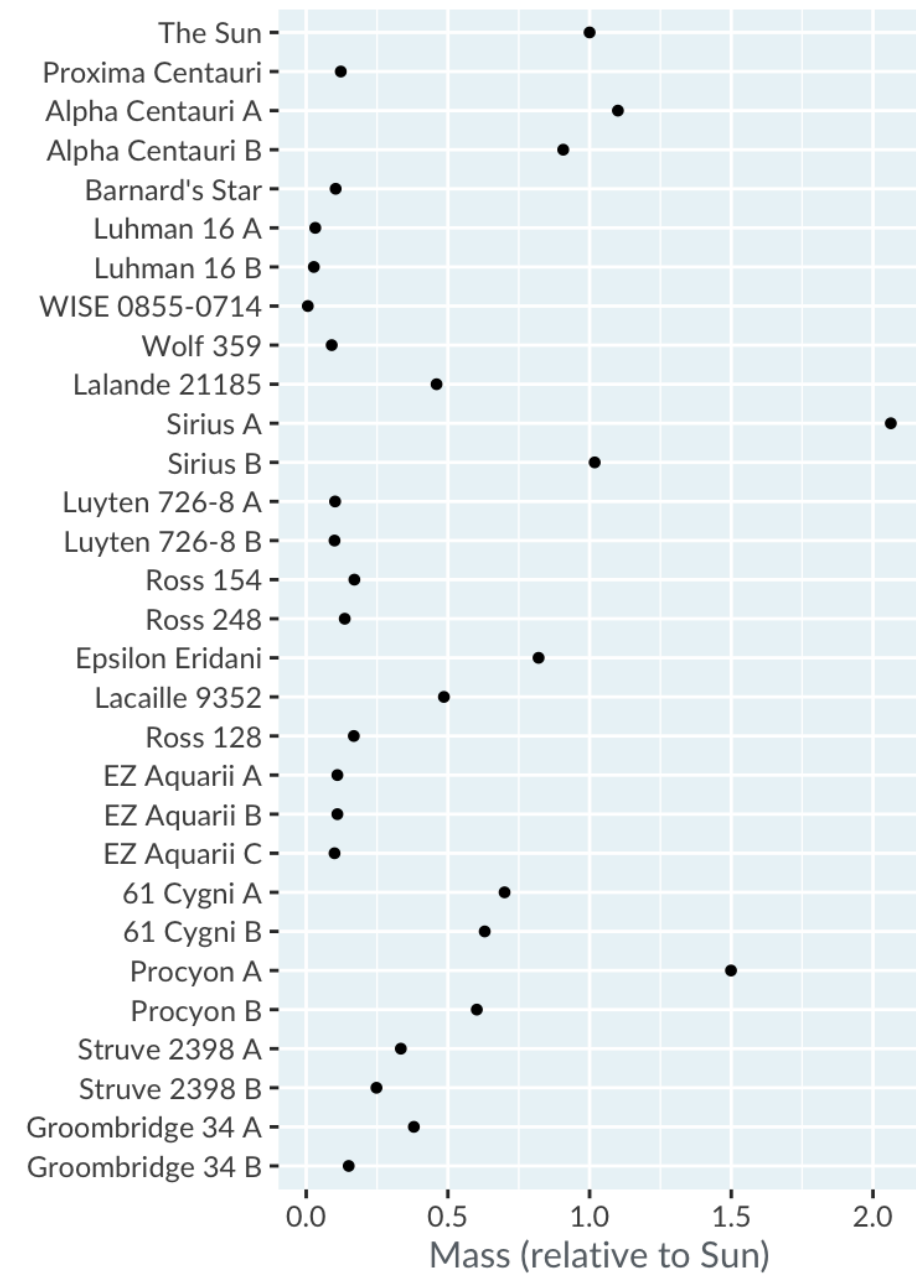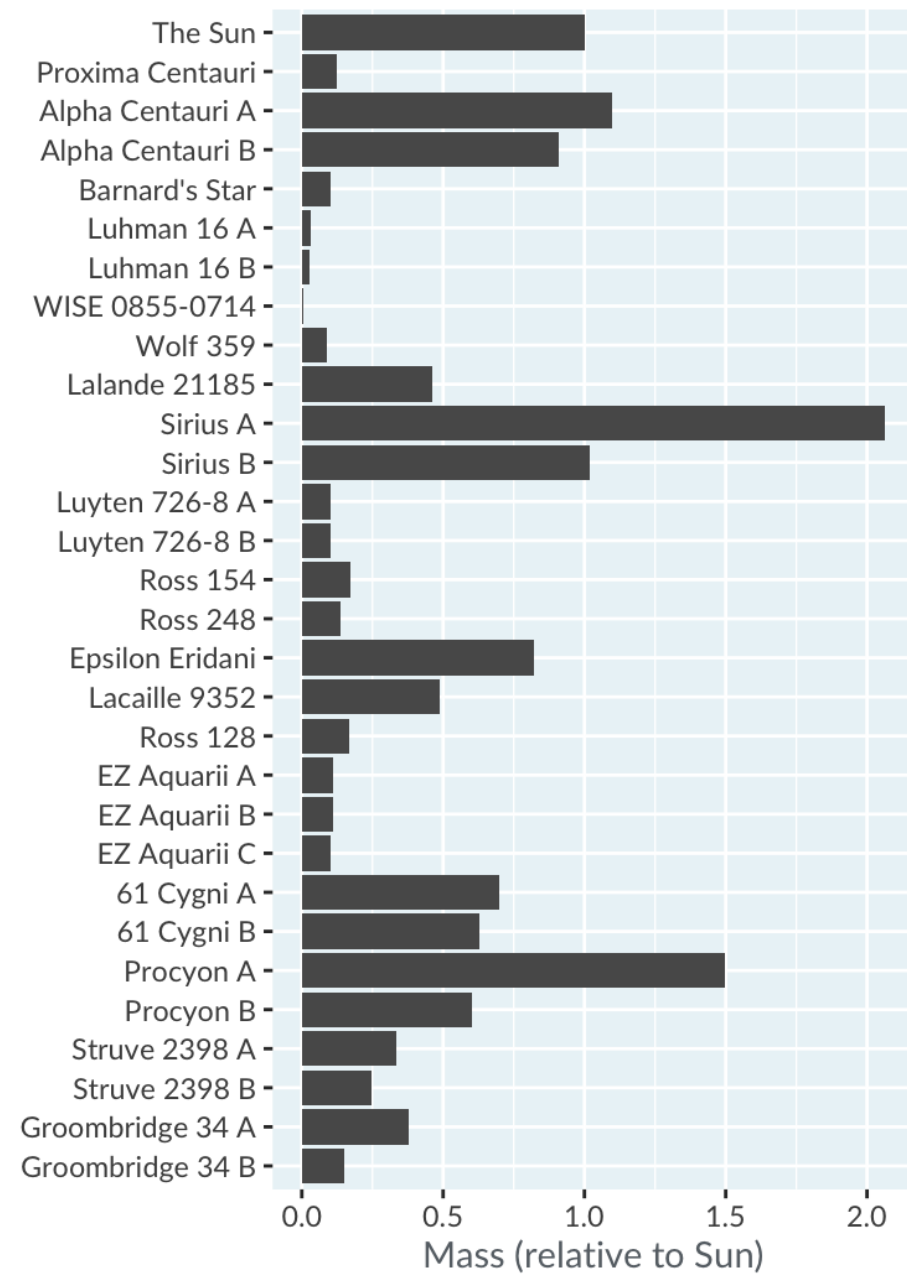
# Log scales

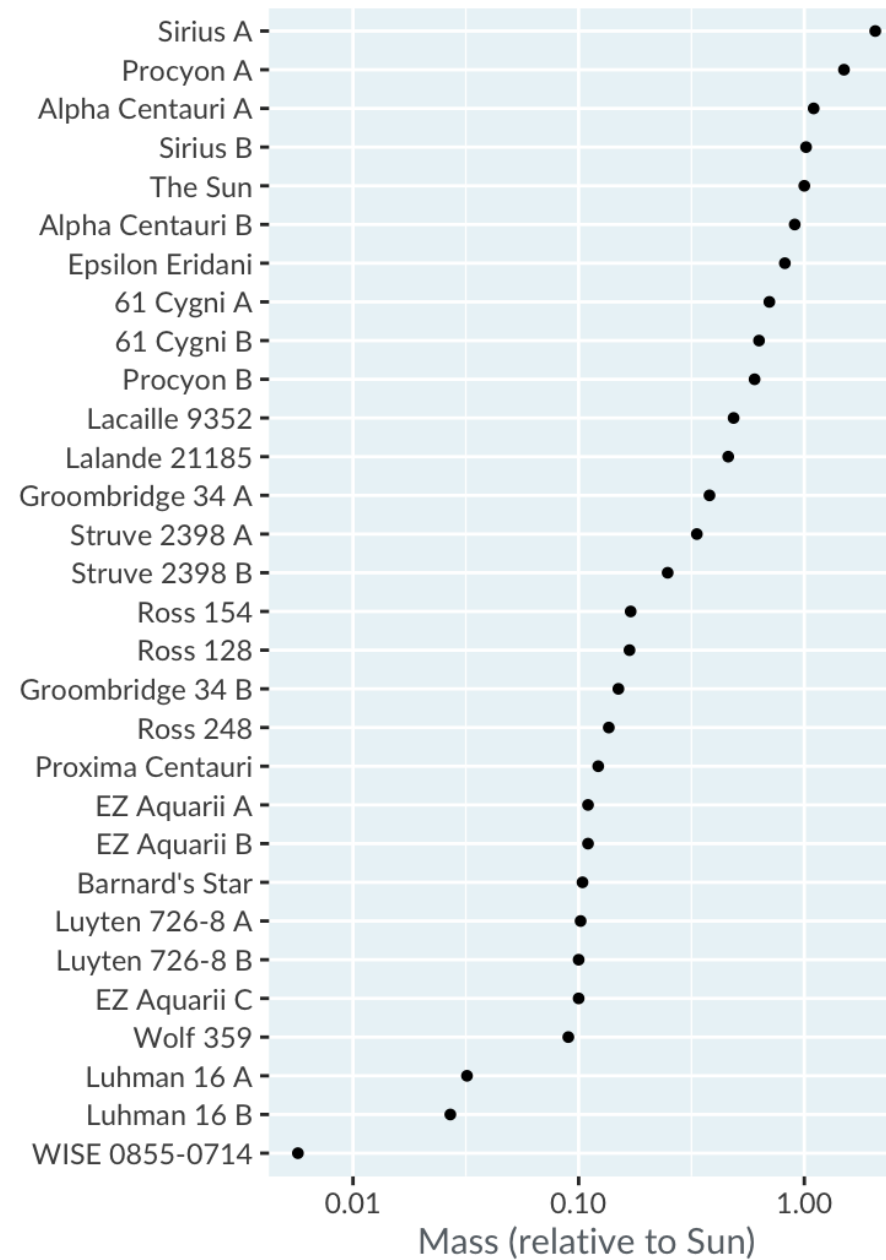# Sorting rows

# Let's practice!

## DATA VISUALIZATION FOR EVERYONE