

Introduction to **statistical** graphics

The greatest
value of a
picture is when
it forces us to
notice what we
never expected
to see.

— John Tukey





Numbers have
an important
story to tell.
They rely on you
to give them a
clear and
convincing
voice.

— Stephen Few

Goals

- 1. Establish a set of simple graphical tools that allow you to make powerful insights about your data**
- 2. Establish a common vocabulary to discuss graphical findings**
- 3. Establish good habits**

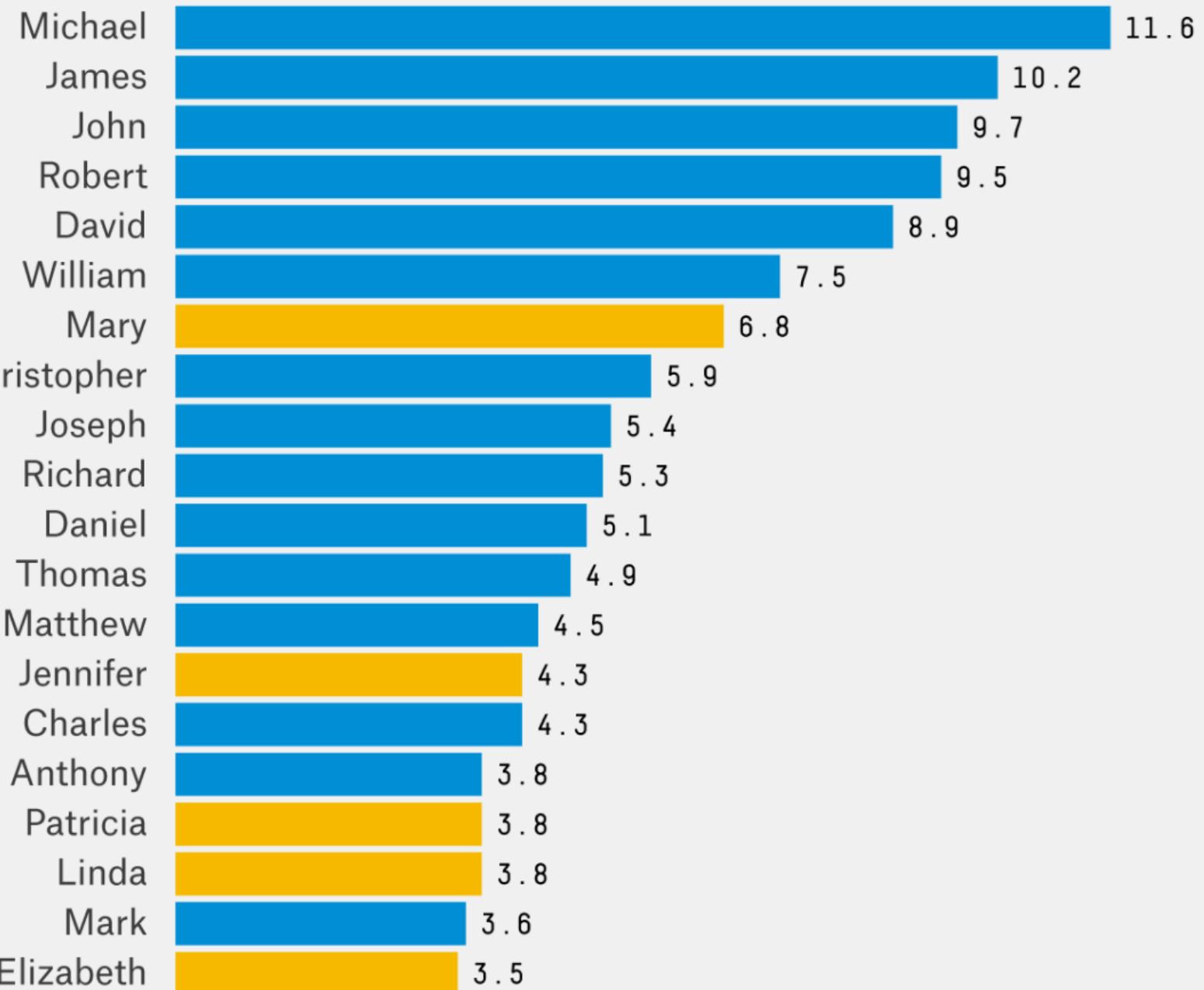
displaying a

distribution

Most Common First Names

Per 1,000 Americans as of 2013

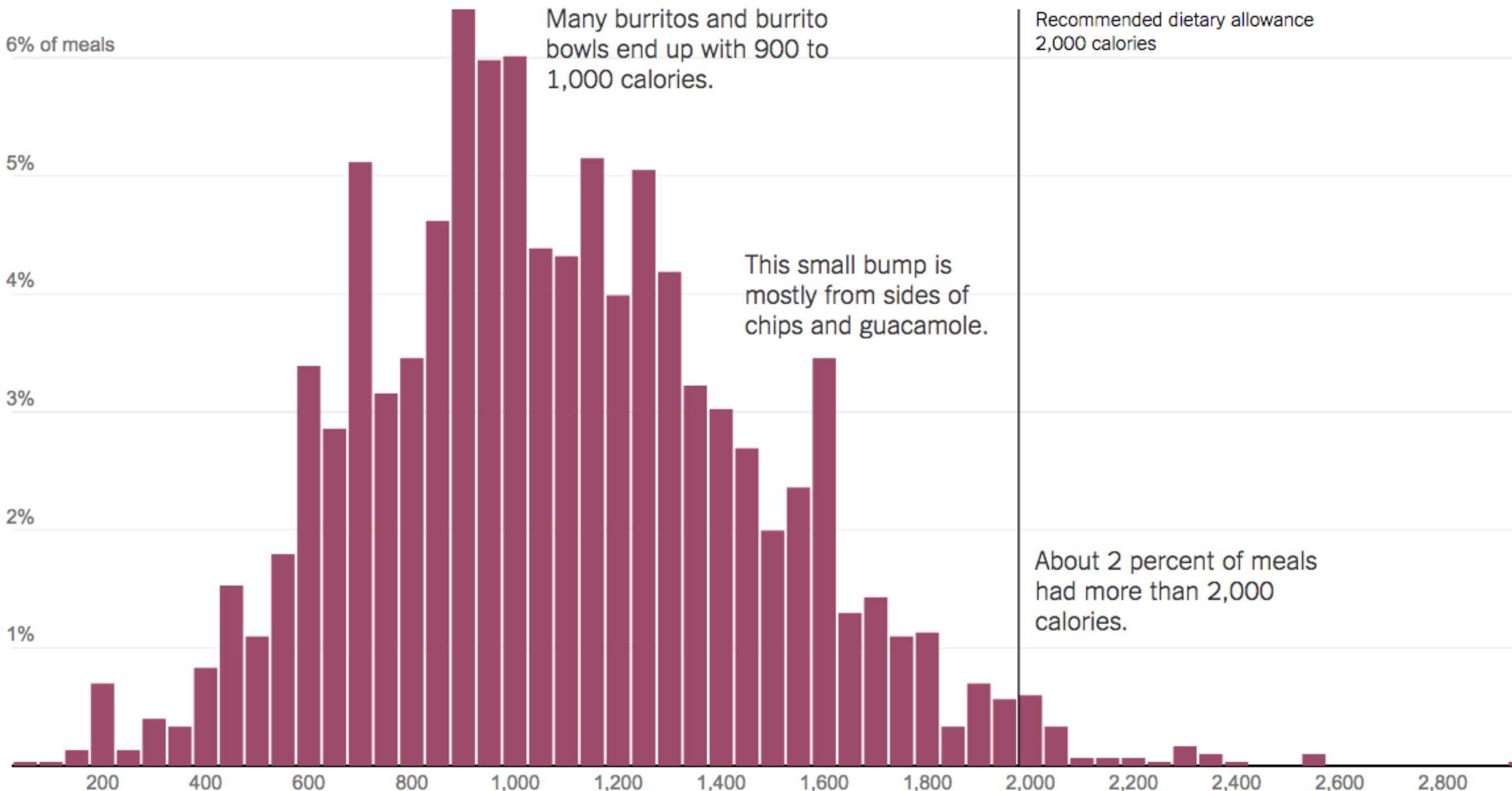
MALE FEMALE



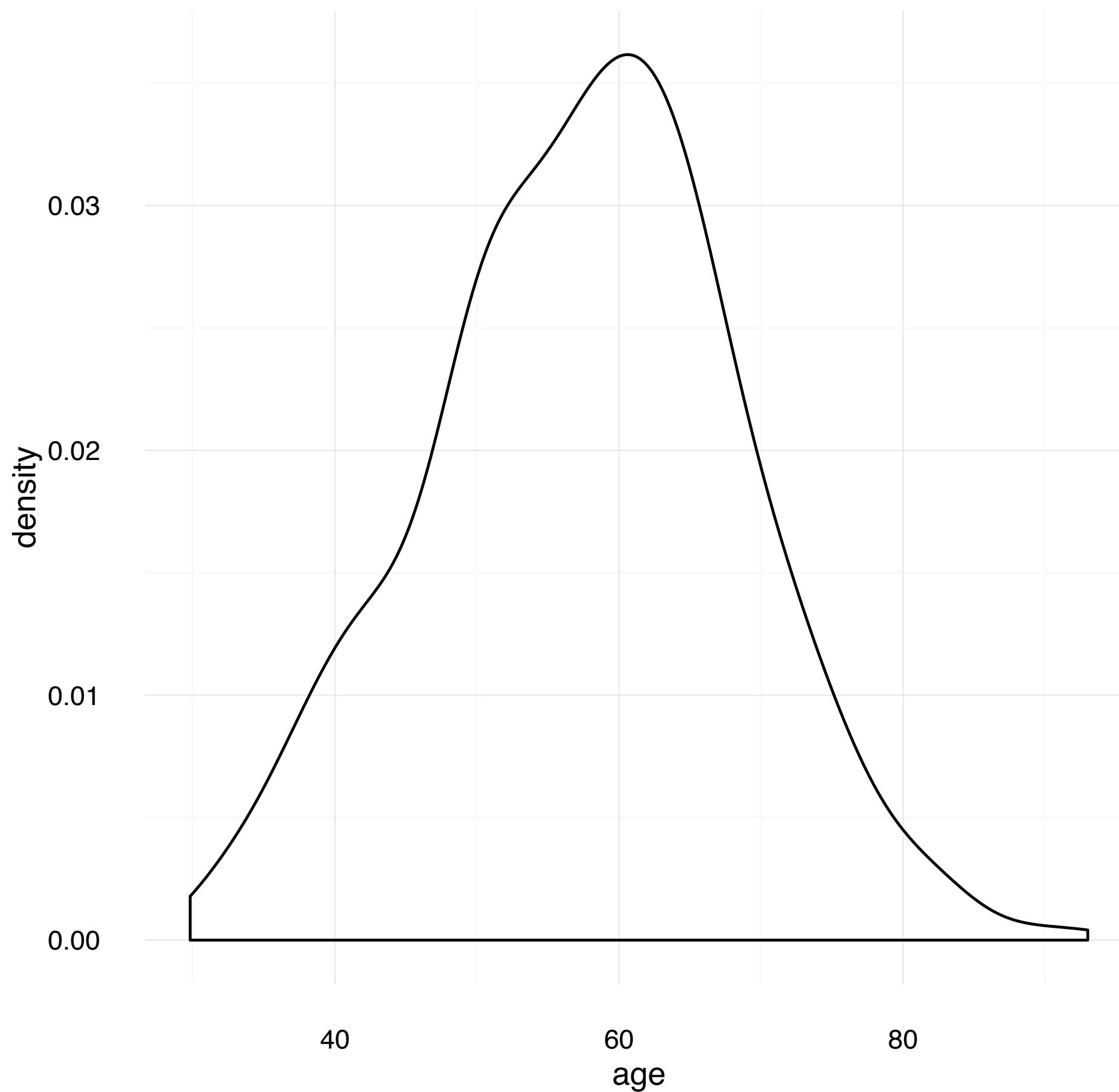
At Chipotle, How Many Calories Do People Really Eat?

By KEVIN QUEALY, AMANDA COX and JOSH KATZ FEB. 17, 2015

Most meals have more than 1,000 calories and almost a full day's worth of sodium.



How old was the 113th Congress?

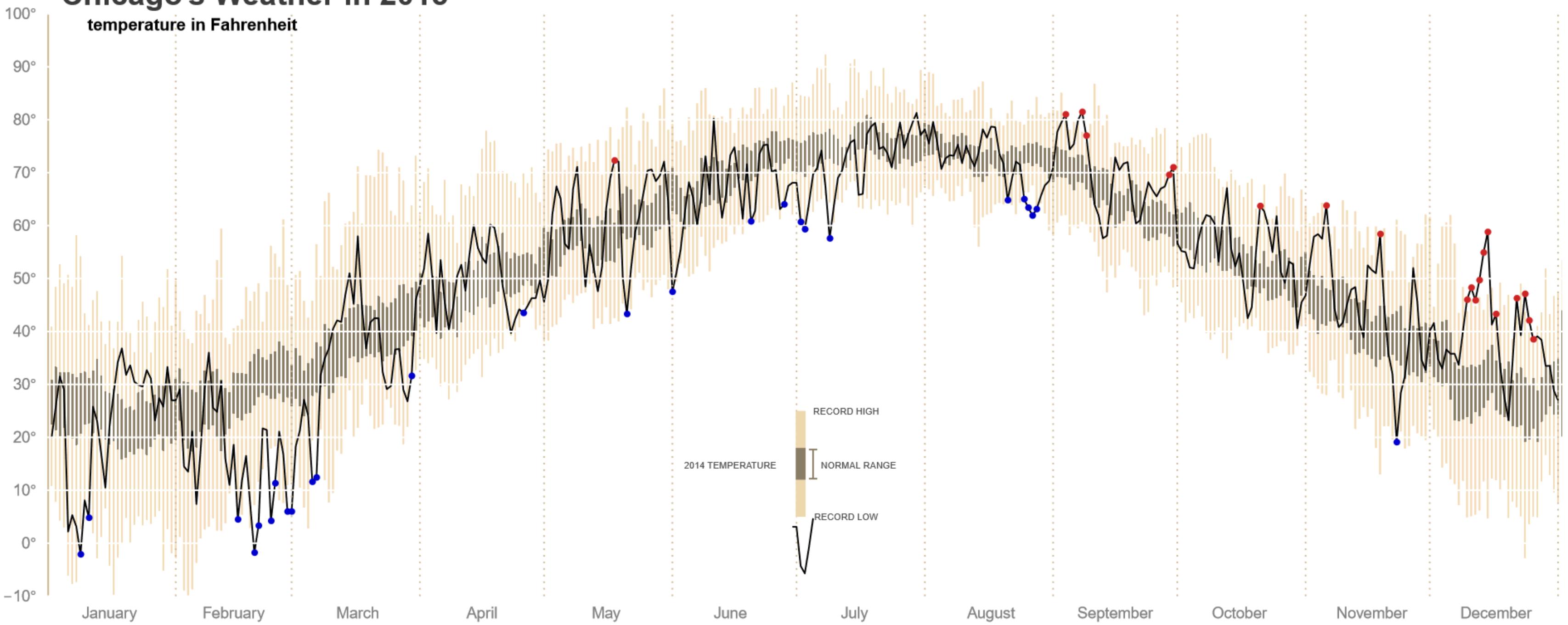


comparisons

over time

Chicago's Weather in 2015

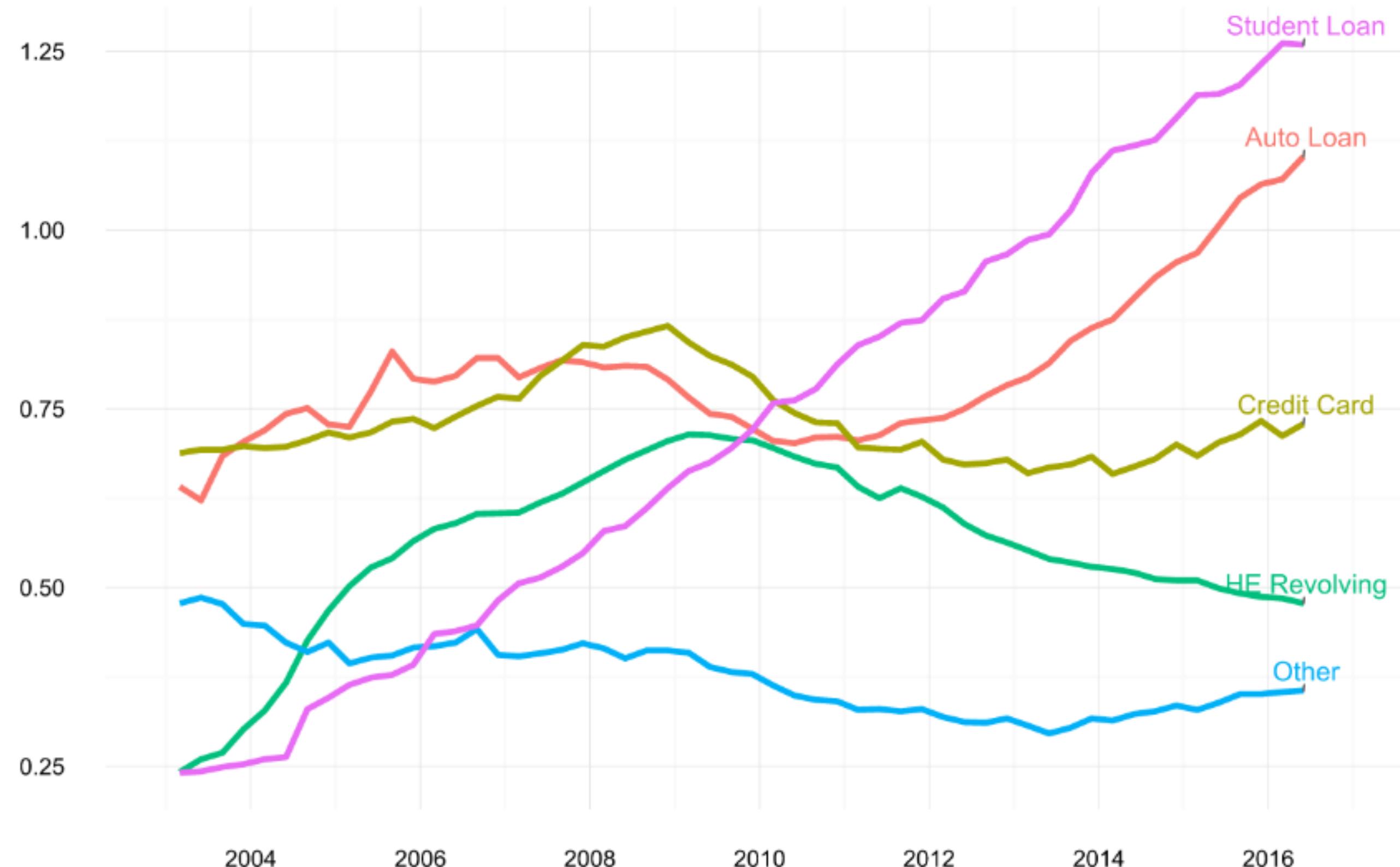
temperature in Fahrenheit



displaying
associations

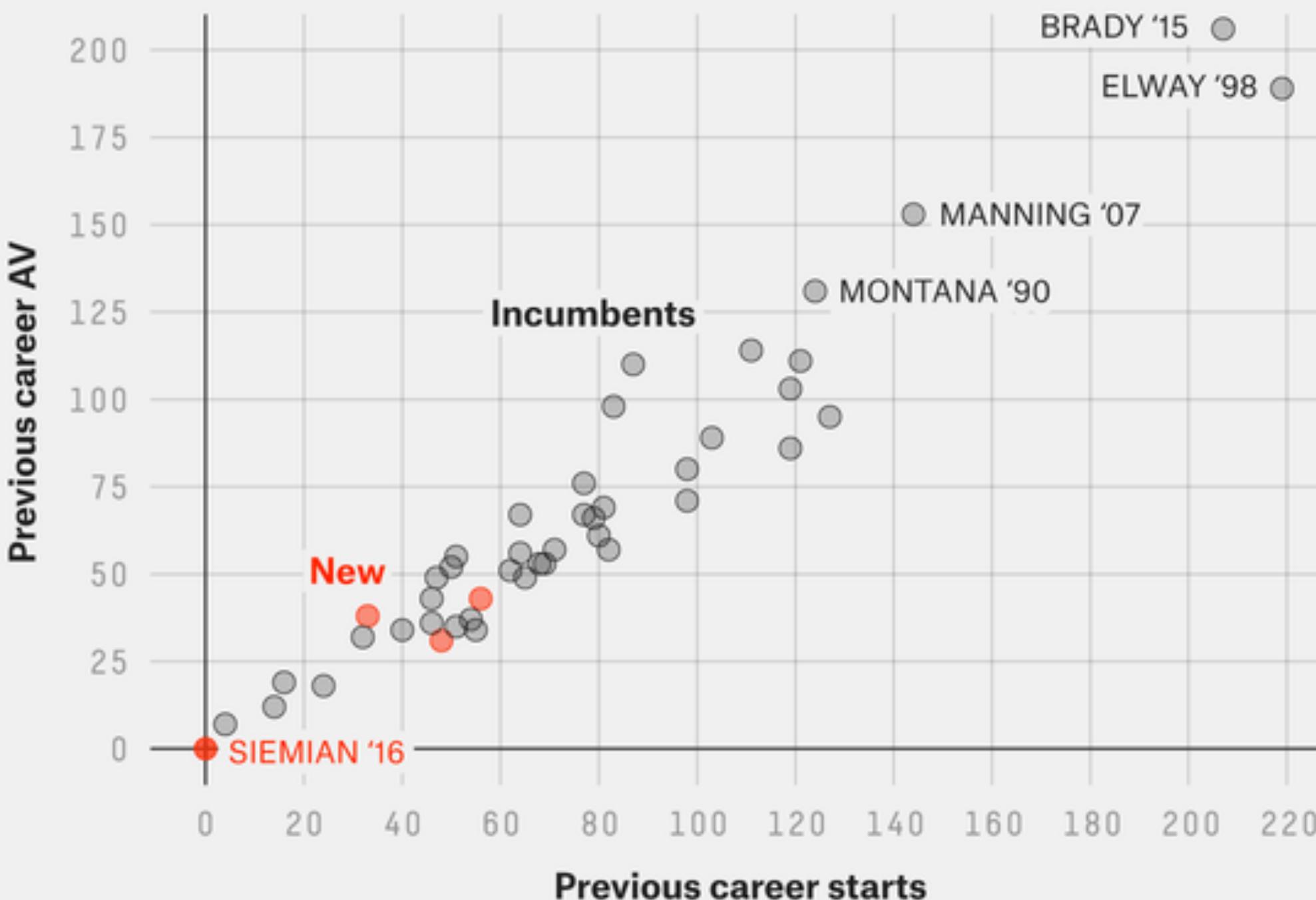
Household debt balances

Trillions of Dollars



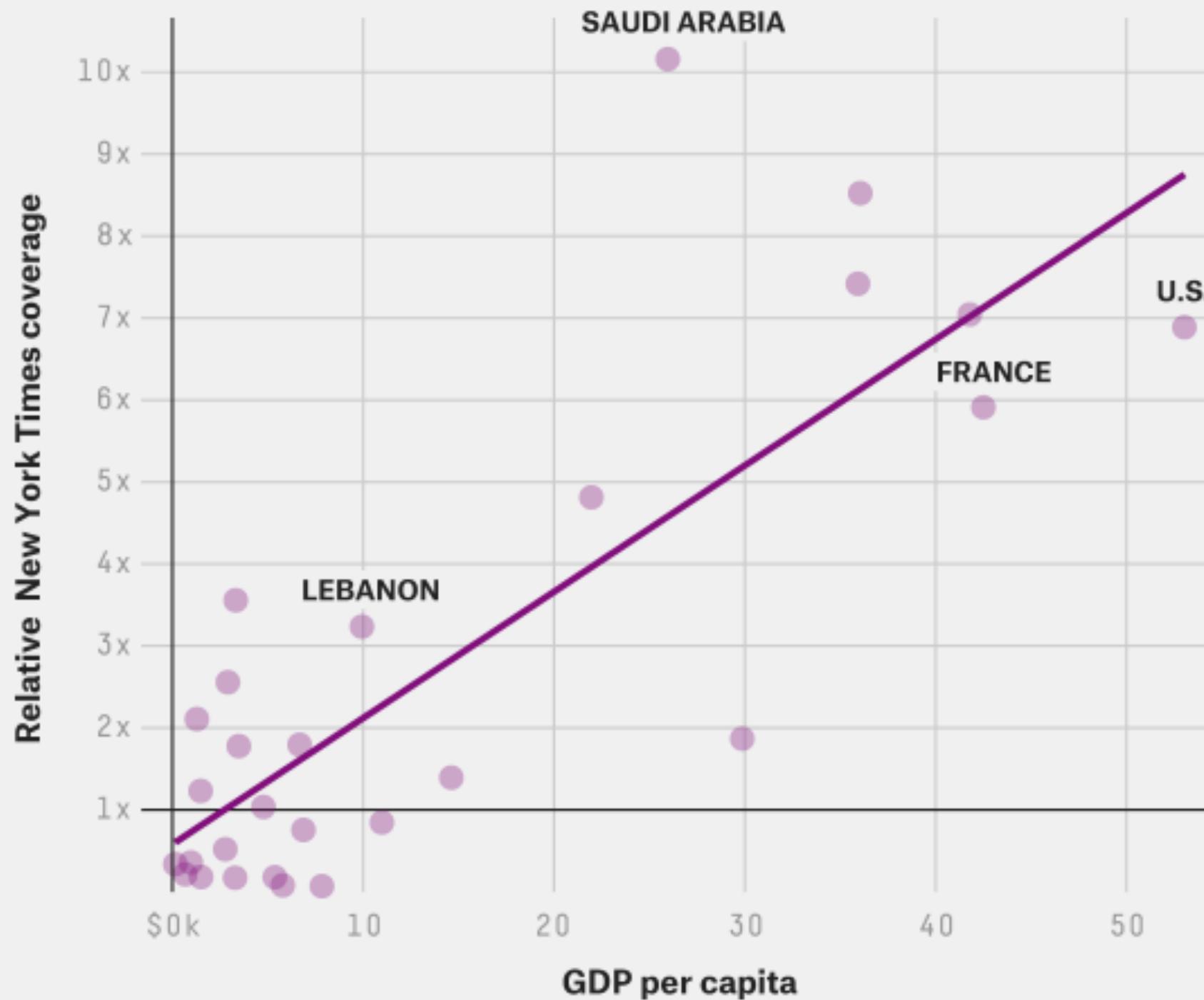
Week 1 experience for defending champs' QBs

Career Approximate Value (AV) and starts for defending champion's opening-day primary QB, 1967-2016

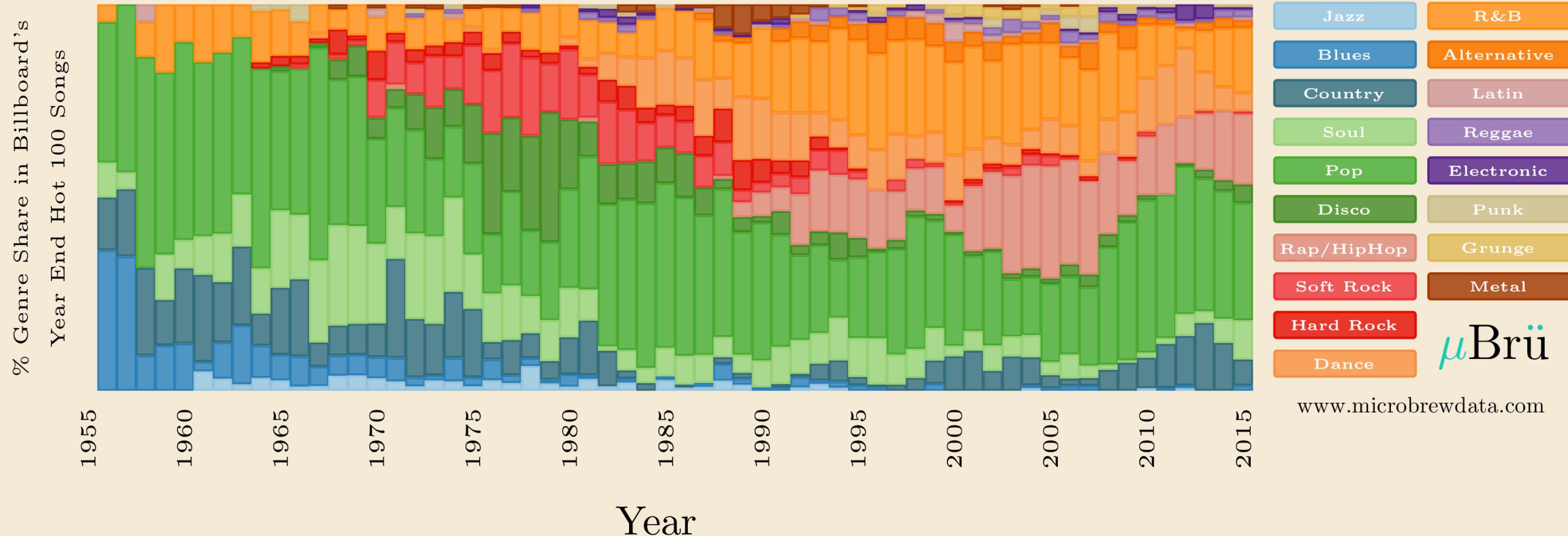


GDP per capita predicts terror coverage

Relative odds of a terrorist attack of a given magnitude being reported on by The New York Times vs. GDP per capita, by country



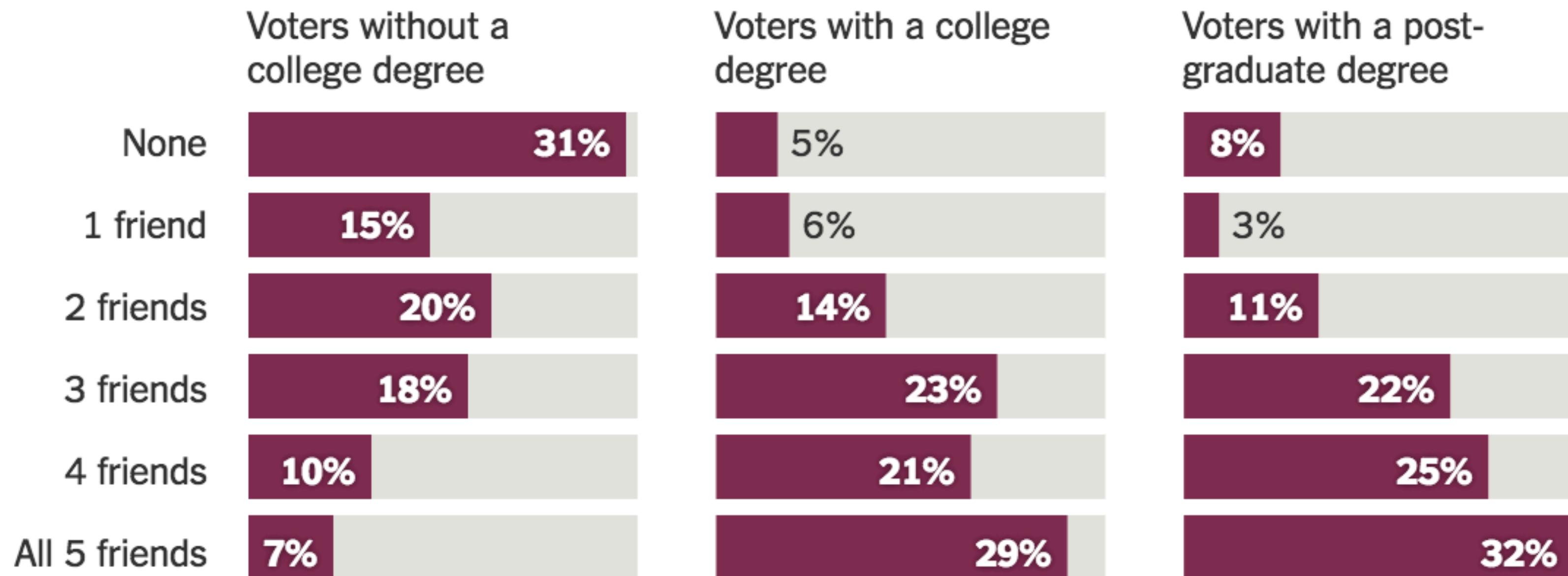
Music Genre Popularity Over the Years



μBrü

www.microbrewdata.com

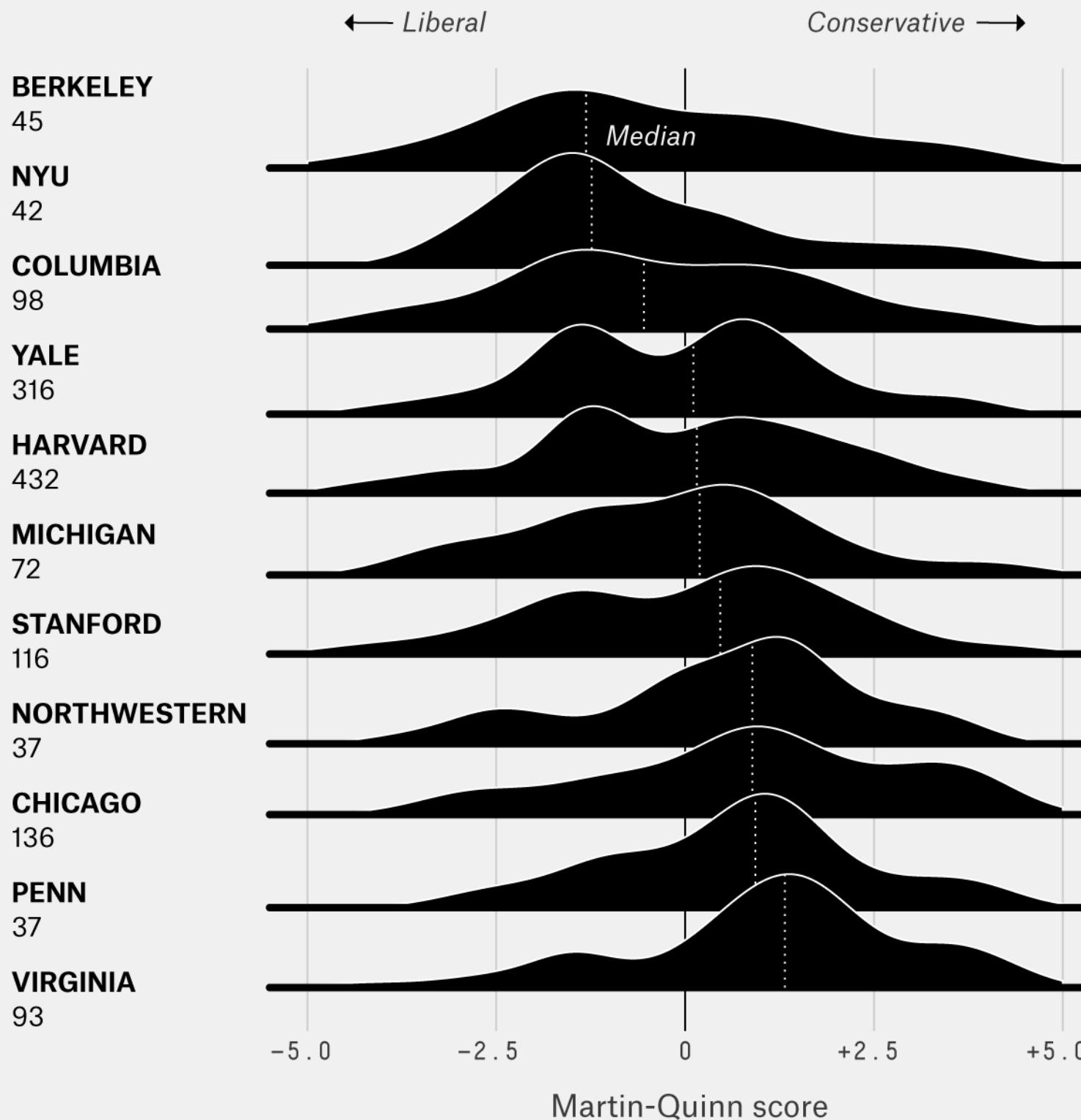
Registered voters were asked how many of their five closest friends have four-year college degrees. The divide was stark.



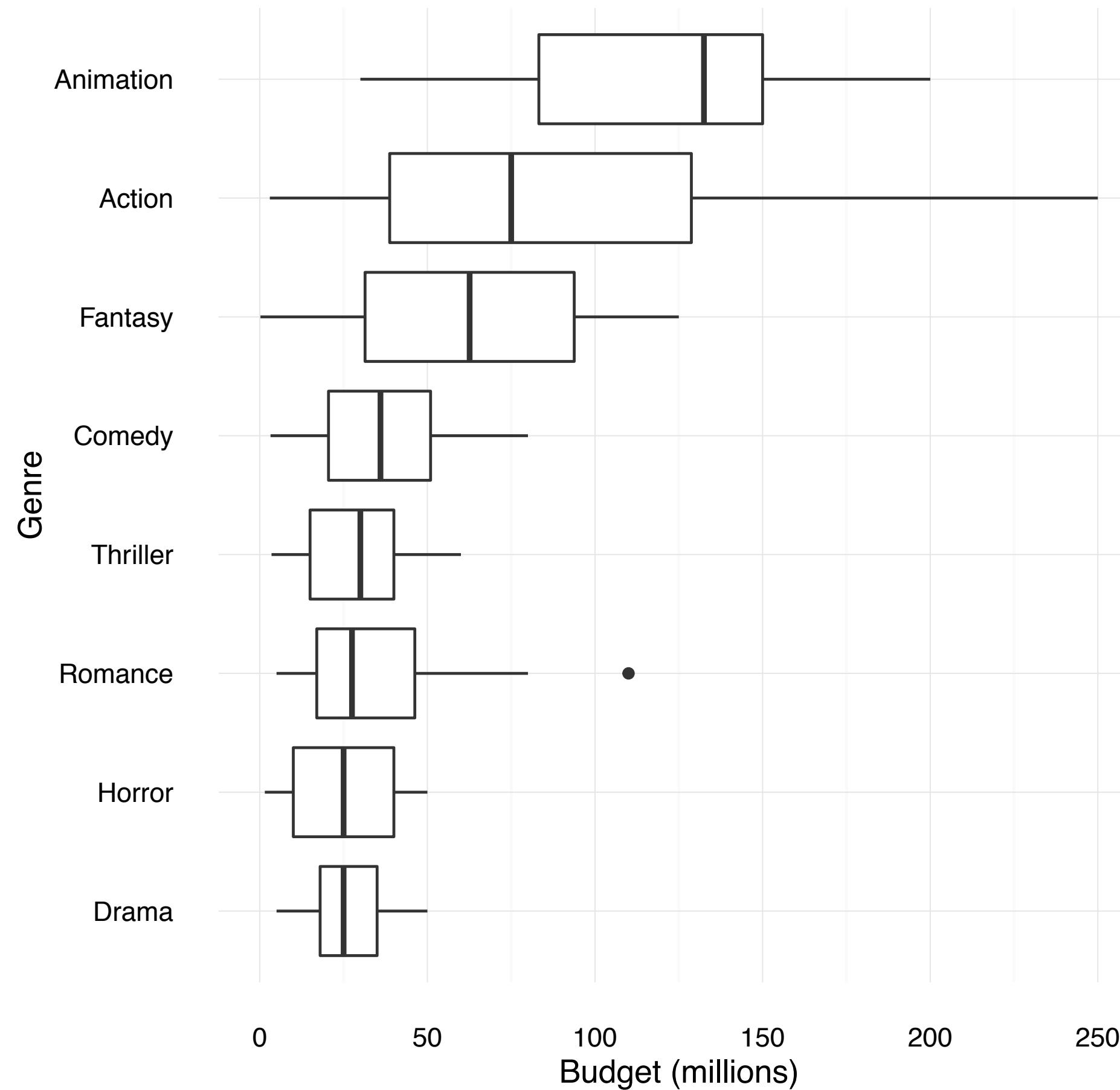
Source: Morning Consult

The Wide-Ranging Politics Of Elite Law Schools

Distributions of justice ideology for Supreme Court clerks, by school since 1937

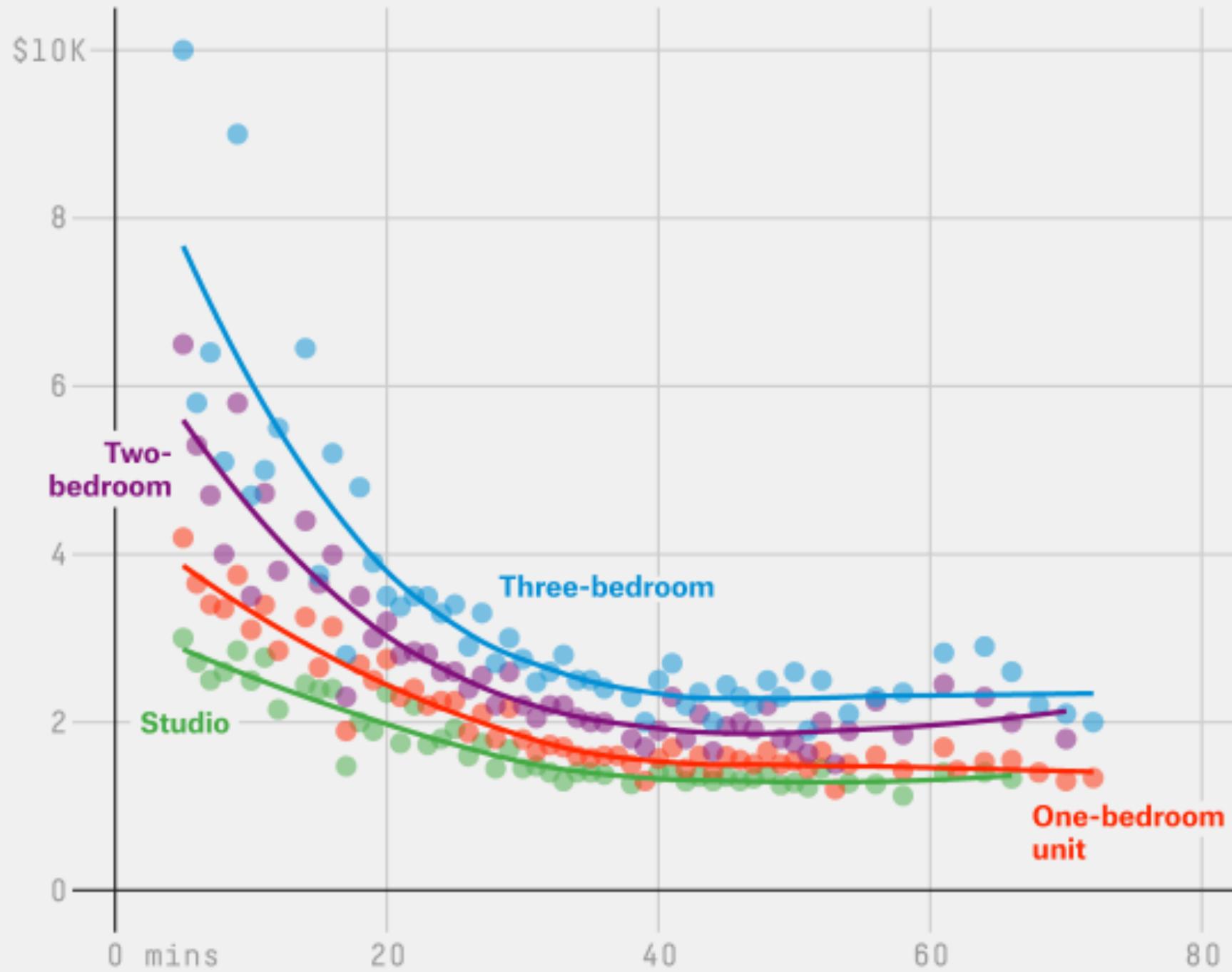


Hollywood Budgets in 2011



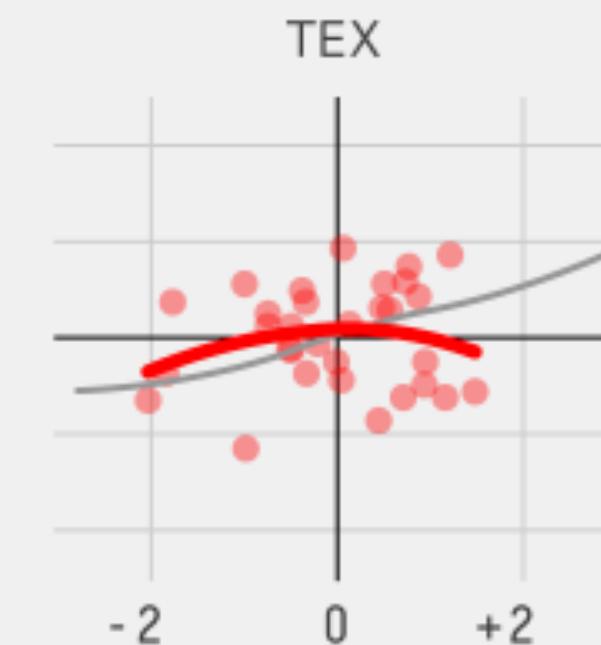
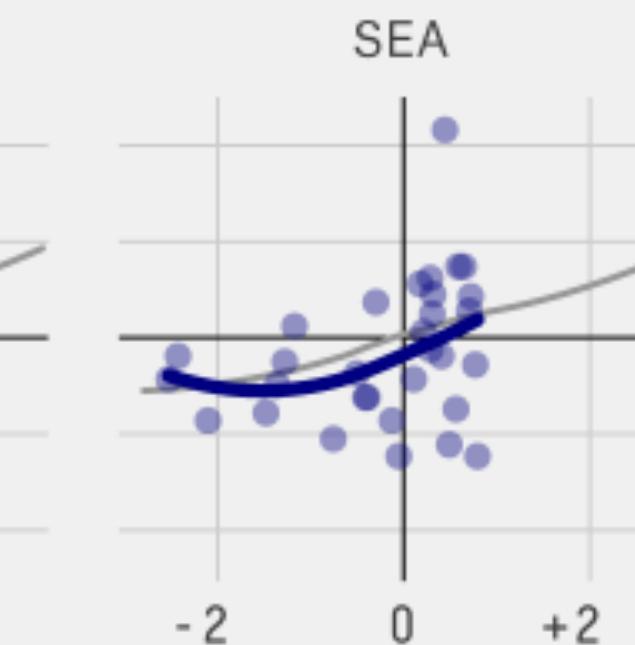
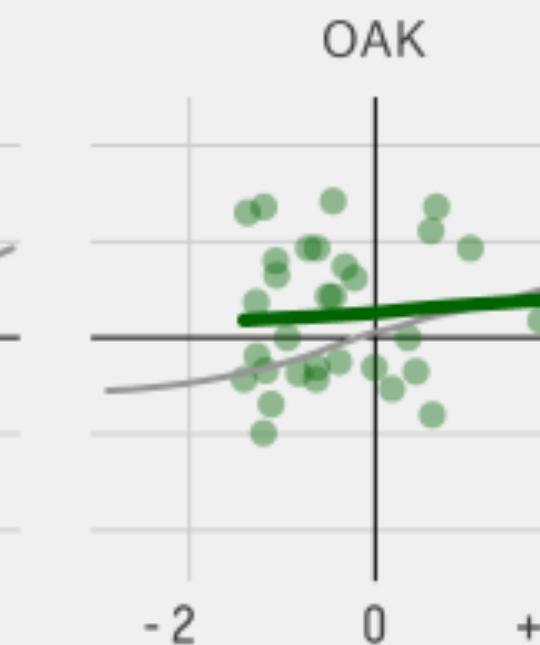
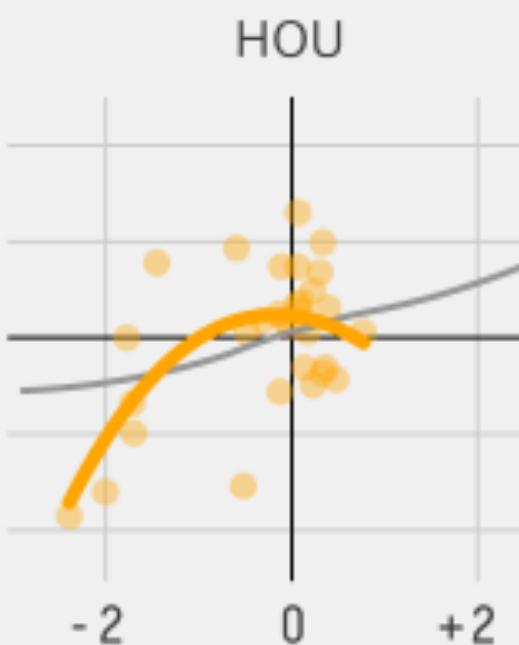
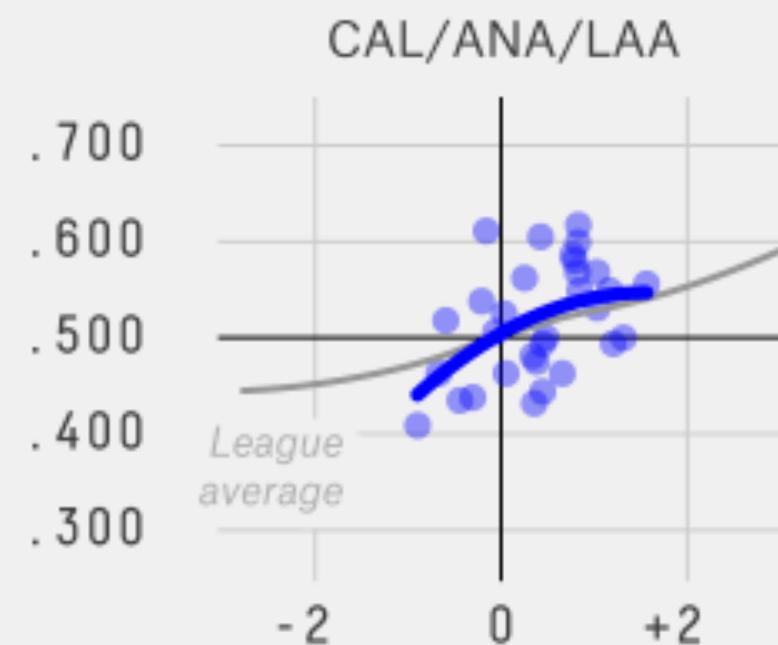
New Yorkers pay up for a shorter commute

Median monthly NYC rent in 2015 vs. commute time by subway

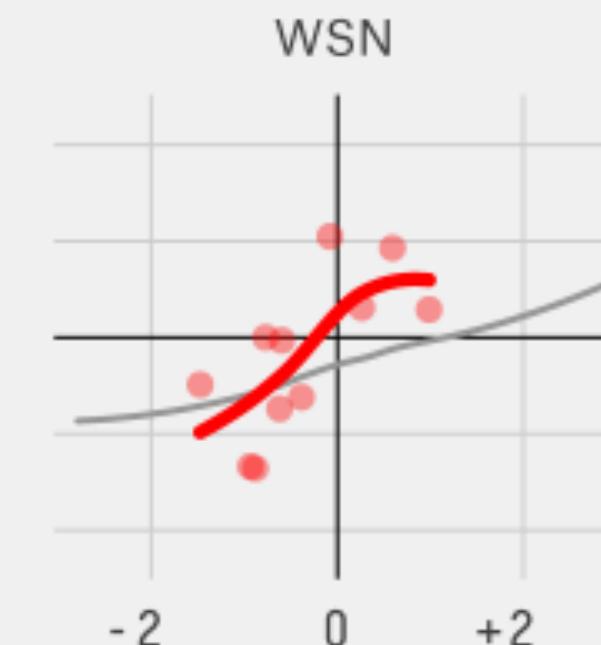
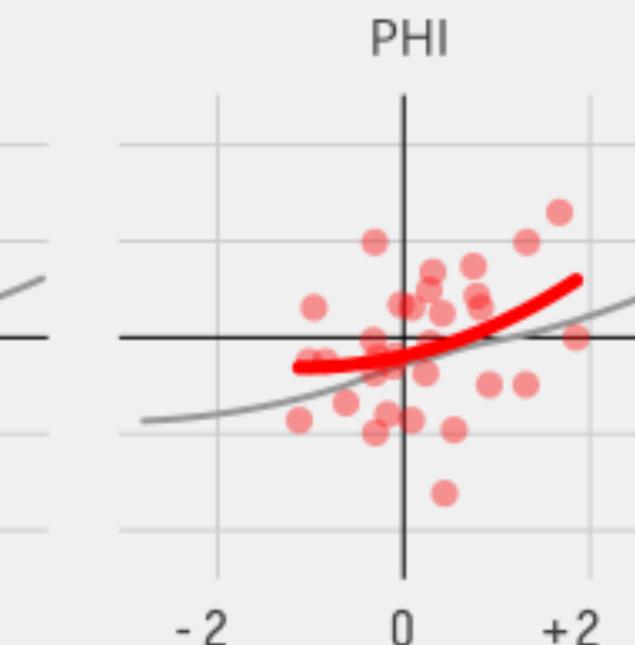
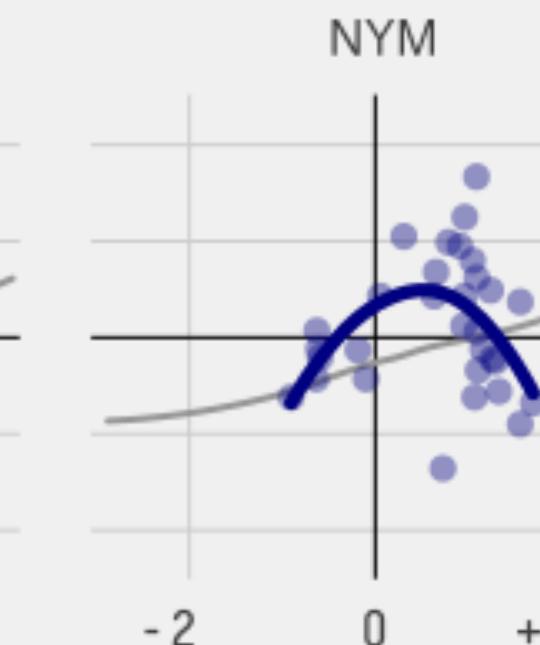
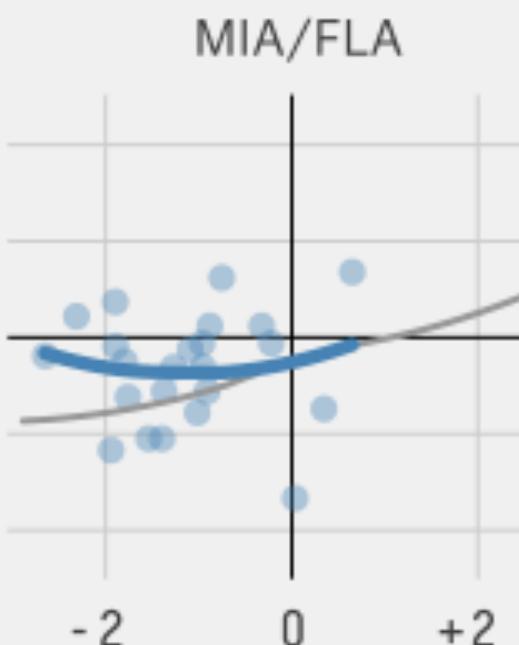
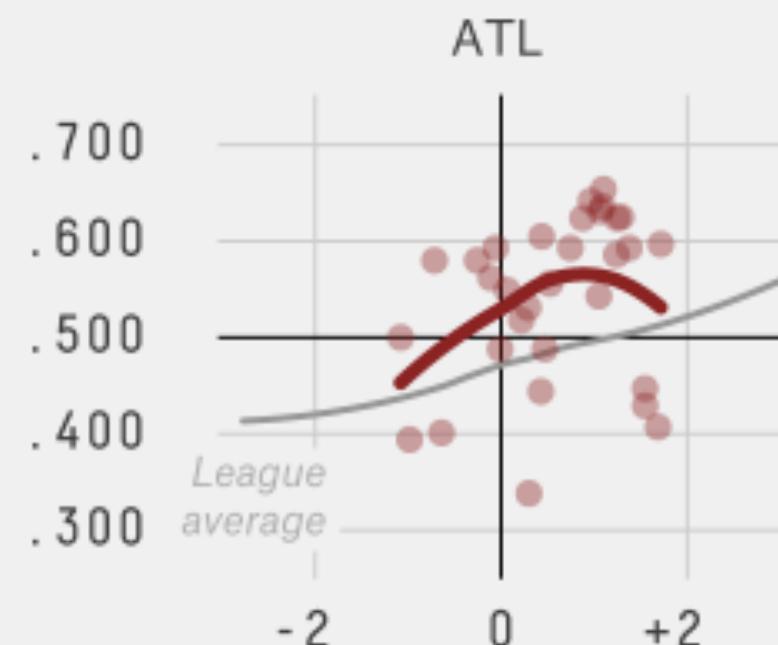


Rental prices are based on FiveThirtyEight's analysis of 2015 listings on StreetEasy. Commutes are calculated as the average time from the subway station nearest a home to the nearest 42nd Street and Chambers Street stations. Commutes without at least 10 listings are excluded.

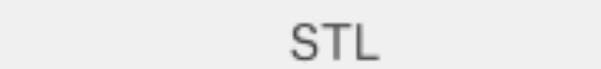
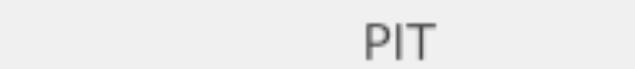
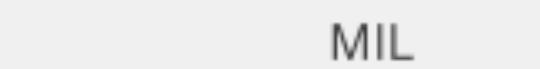
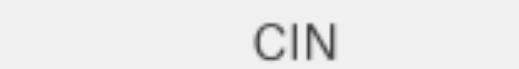
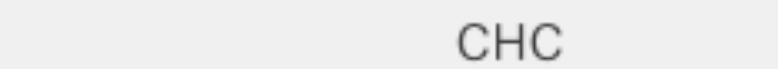
AMERICAN LEAGUE WEST



NATIONAL LEAGUE EAST



NATIONAL LEAGUE CENTRAL



good

practices

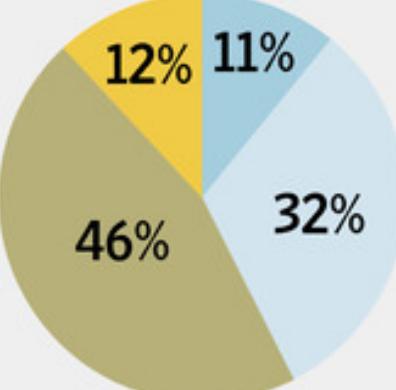
Don't use pie charts

Where Does the Time Go?

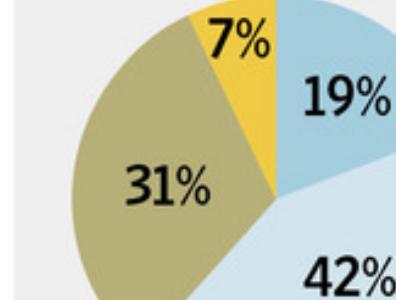
The amount of time spent on various tasks by surveyed nonmanagers in data-science positions

- Less than 1 hour a week
- 1 to 3 hours a day
- 1 to 4 hours a week
- 4 or more hours a day

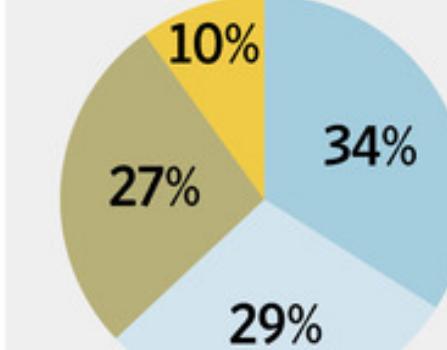
Basic exploratory data analysis



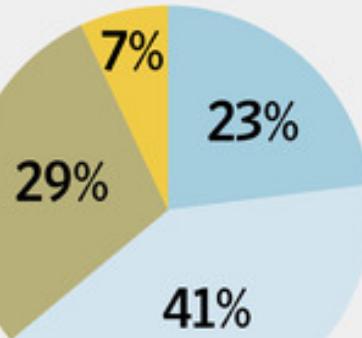
Data cleaning¹



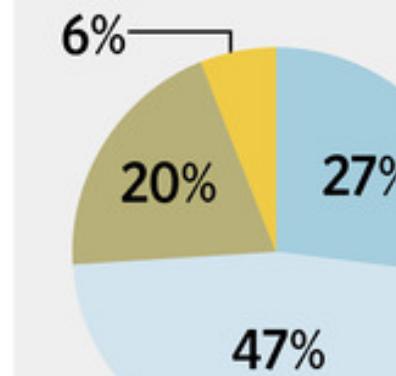
Machine learning, statistics²



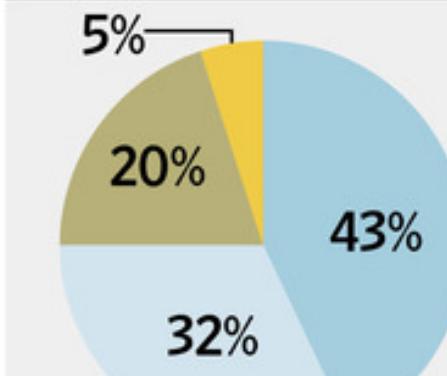
Creating visualizations



Presenting analysis



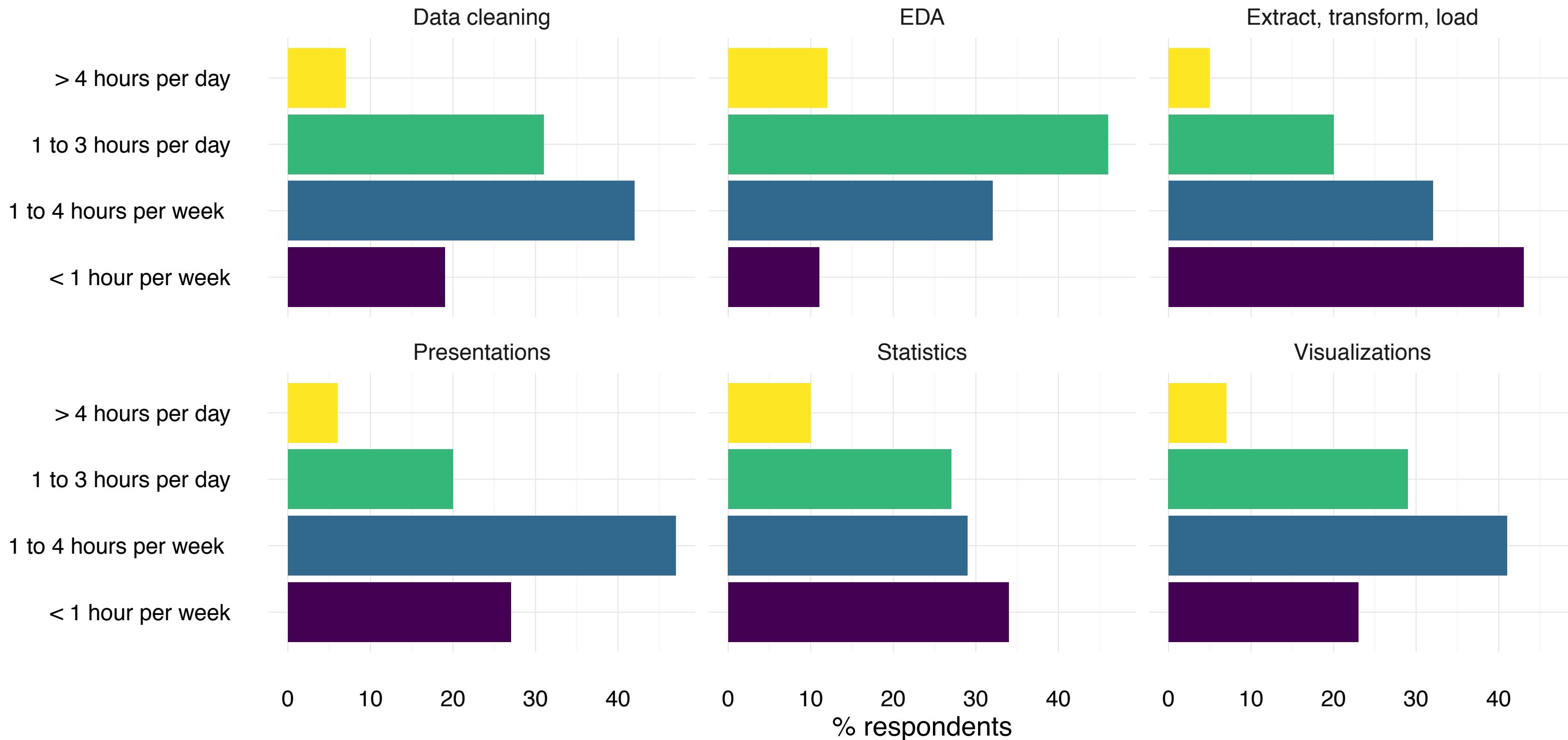
Extract, transform, load³



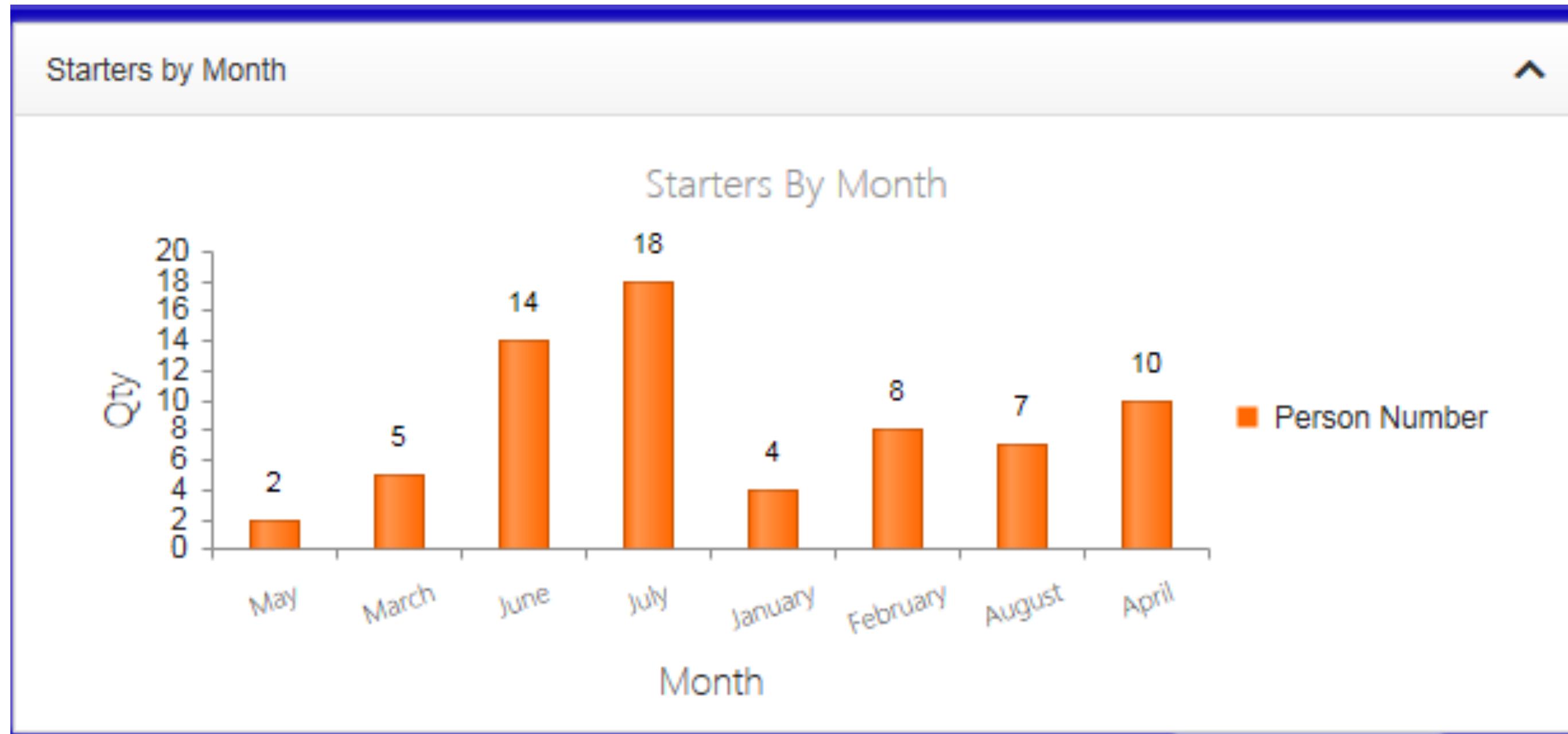
¹ Correcting or removing faulty data ² Creating computer models

³ Also known as ETL — moving information to a data warehouse

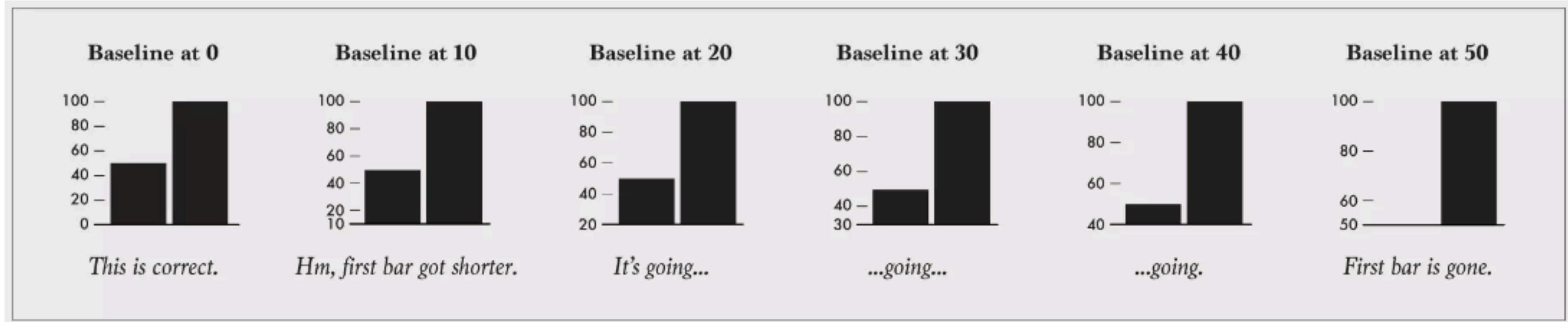
Where does a data scientists' time go?



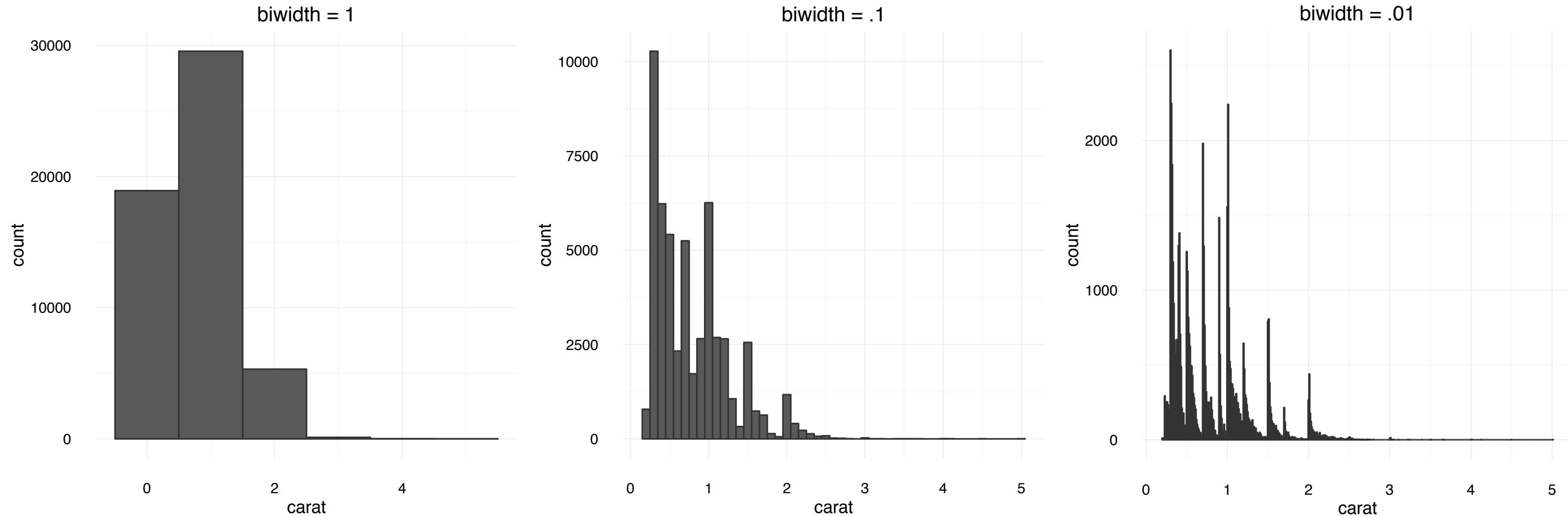
Meaningfully order your bar charts and boxplots



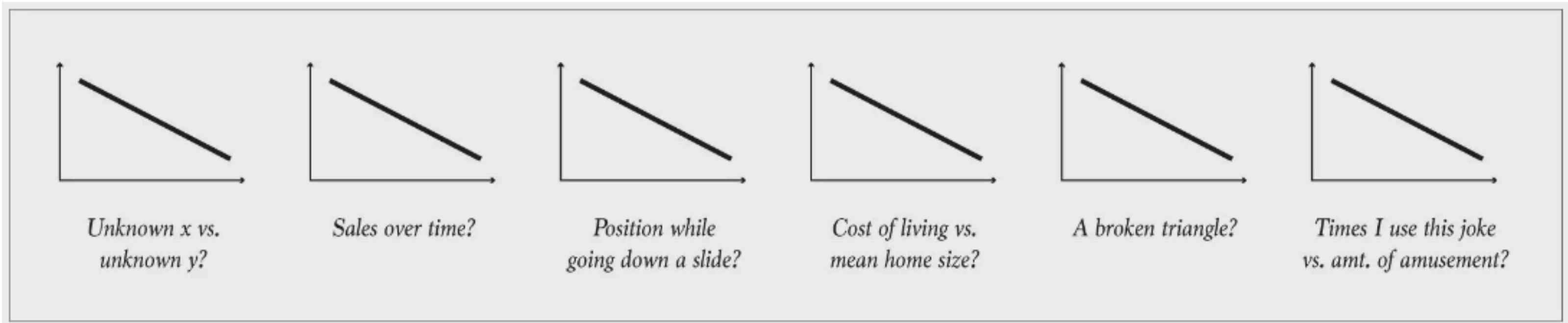
Bar charts start with a zero baseline



Experiment with histogram binwidth



Label your axes, including units



Source: flowingdata.com

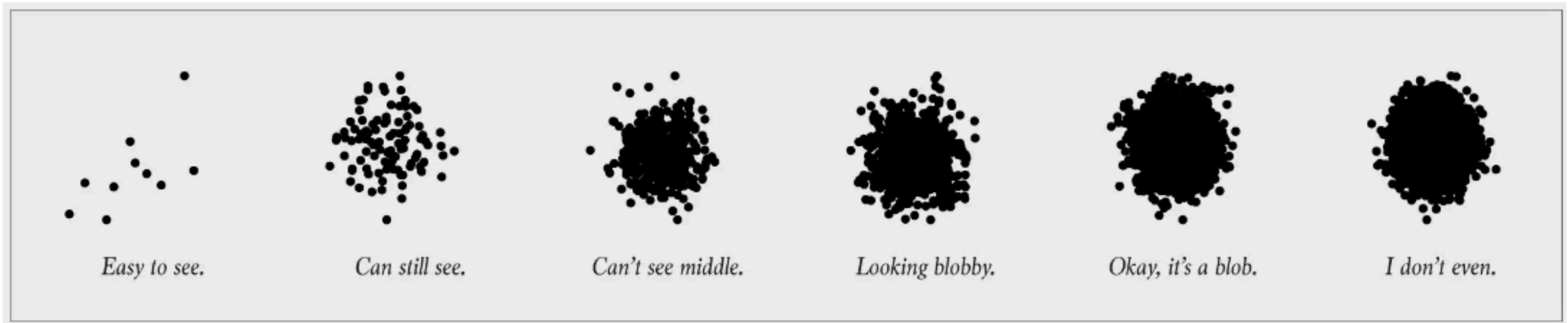
Use color strategically



63%

202,803,591 million Americans
are **represented by a climate
denier in Congress.**

Avoid overplotting



Source: flowingdata.com

Additional reading if you're interested

