7. Confidence intervals, figure layouts, and interactivity

Lecture learning goals

By the end of the lecture you will be able to:

1. Create and understand how to interpret confidence intervals and confidence bands.
2. Layout plots in panels of a figure grid.
3. Create selections within a plot in Altair
4. Link selections between plots to highlight and select data.

Required activities

Before class:

- This 30 min video on confidence intervals and figure layouts (it starts a bit abruptly).

After class:

- Review the lecture notes.
- Section 16 on visualizing uncertainty (some of this will be repetition from 552).

Visualizing uncertainty

Predictions from 2016 presidential election

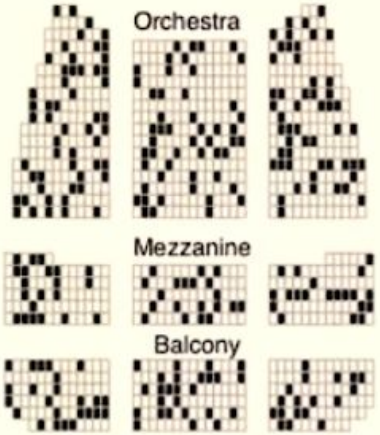
[Justin H. Gross, Washington Post, http://wapo.st/2fCYvDW]

FiveThirtyEight	NYT Upshot	HuffPo Pollster
28%	15%	2%

Predictions from 2016 presidential election

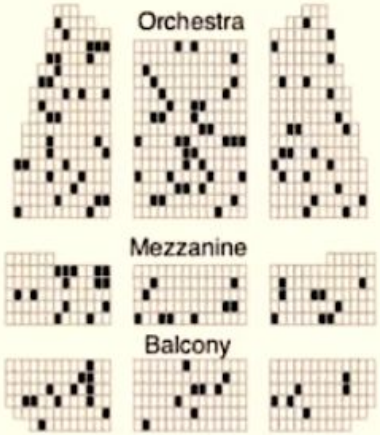
[Justin H. Gross, Washington Post, http://wapo.st/2fCYvDW]

FiveThirtyEight



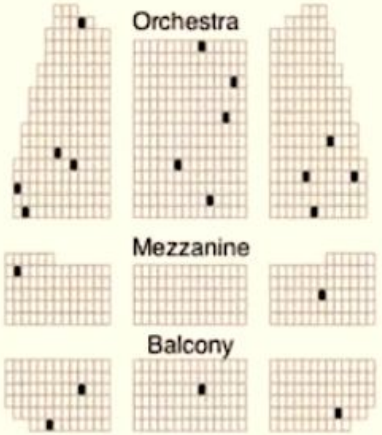
286 cases in 1,000

NYT Upshot



150 cases in 1,000

HuffPo Pollster



20 cases in 1,000

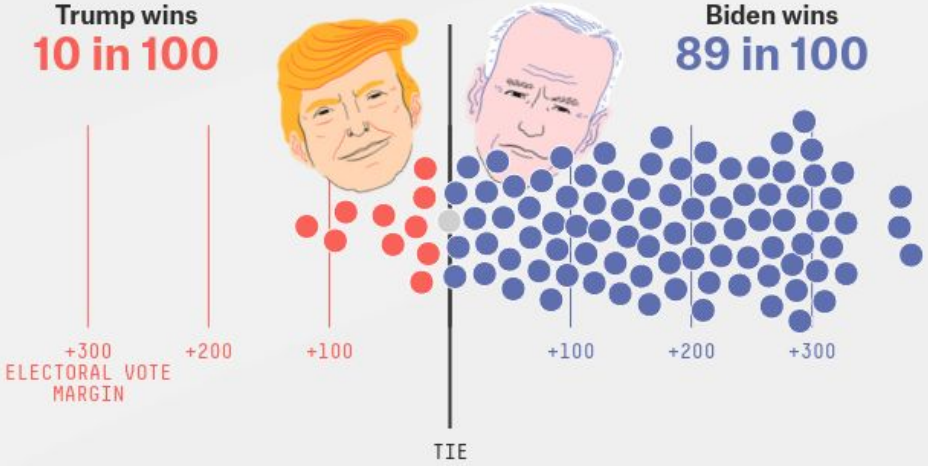
Frequency framing

"When called upon to judge probability, people actually judge something else and believe they have judged probability."

—Daniel Kahneman

Biden is *favored* to win the election

We simulate the election 40,000 times to see who wins most often. The sample of 100 outcomes below gives you a good idea of the range of scenarios our model thinks is possible.

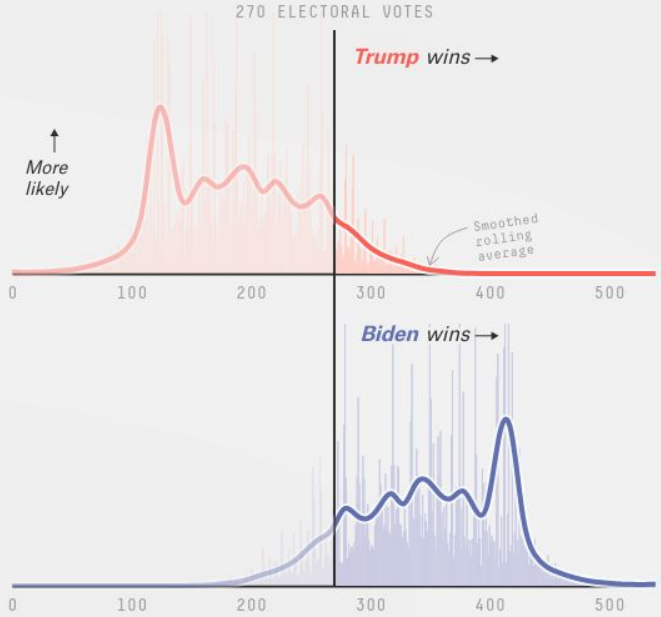


Trump win Biden win

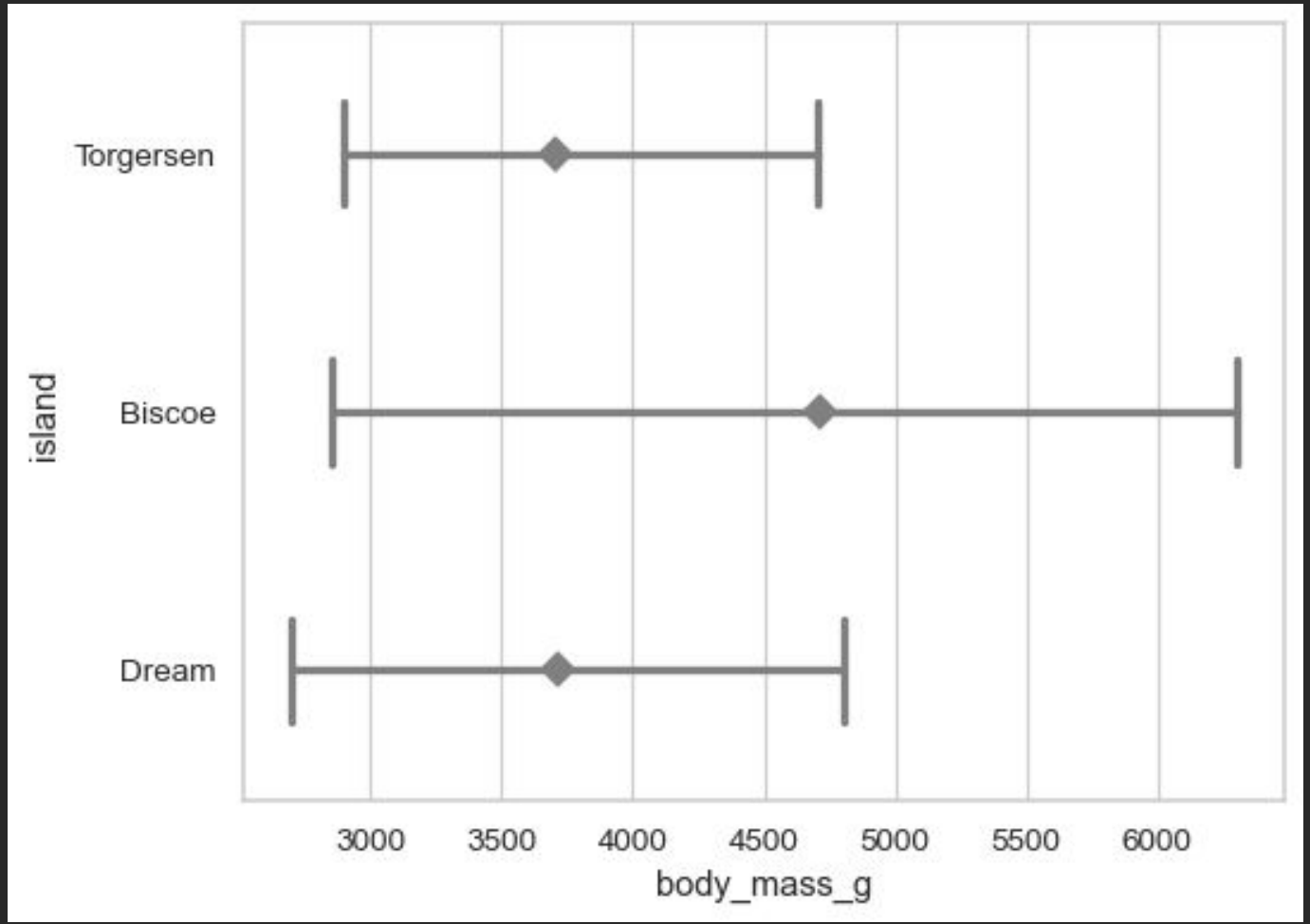
No Electoral College majority, House decides election

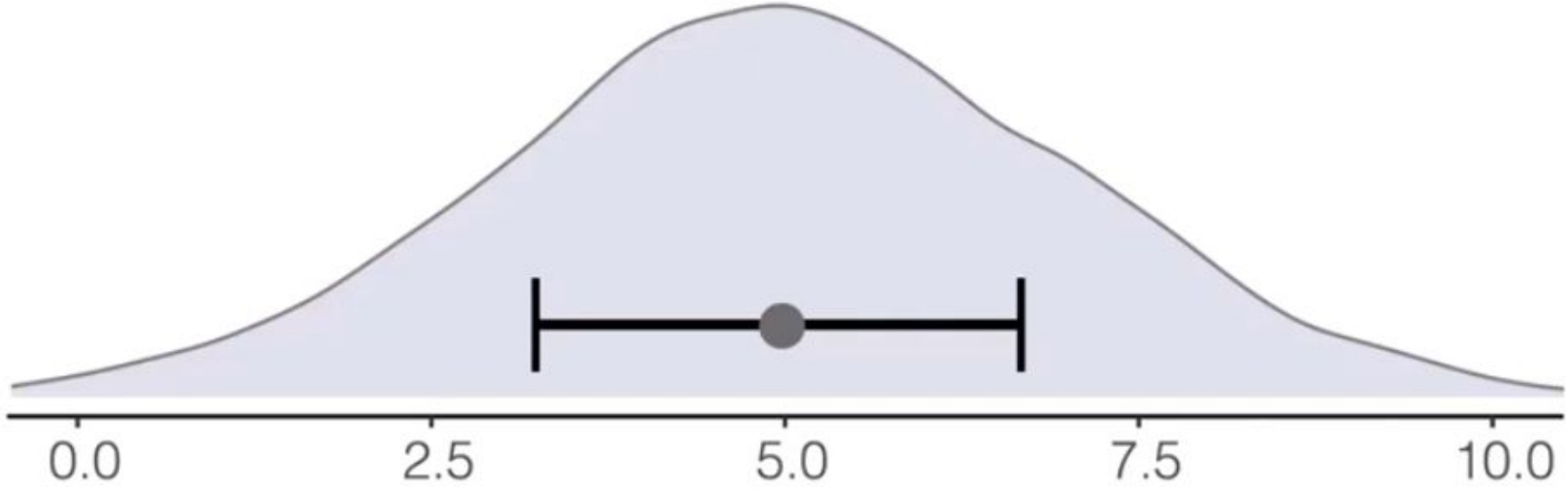
Every outcome in our simulations

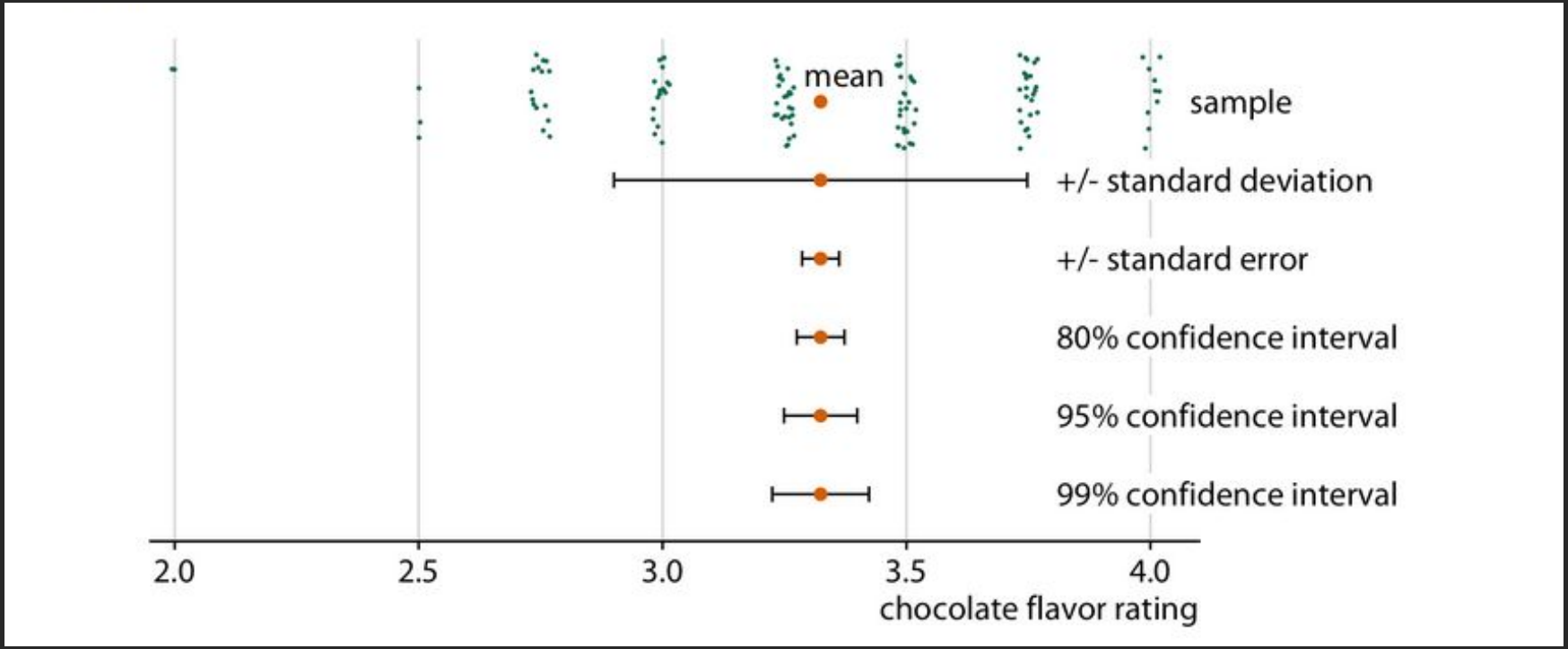
All possible Electoral College outcomes for each candidate, with higher bars showing outcomes that appeared more often in our 40,000 simulations

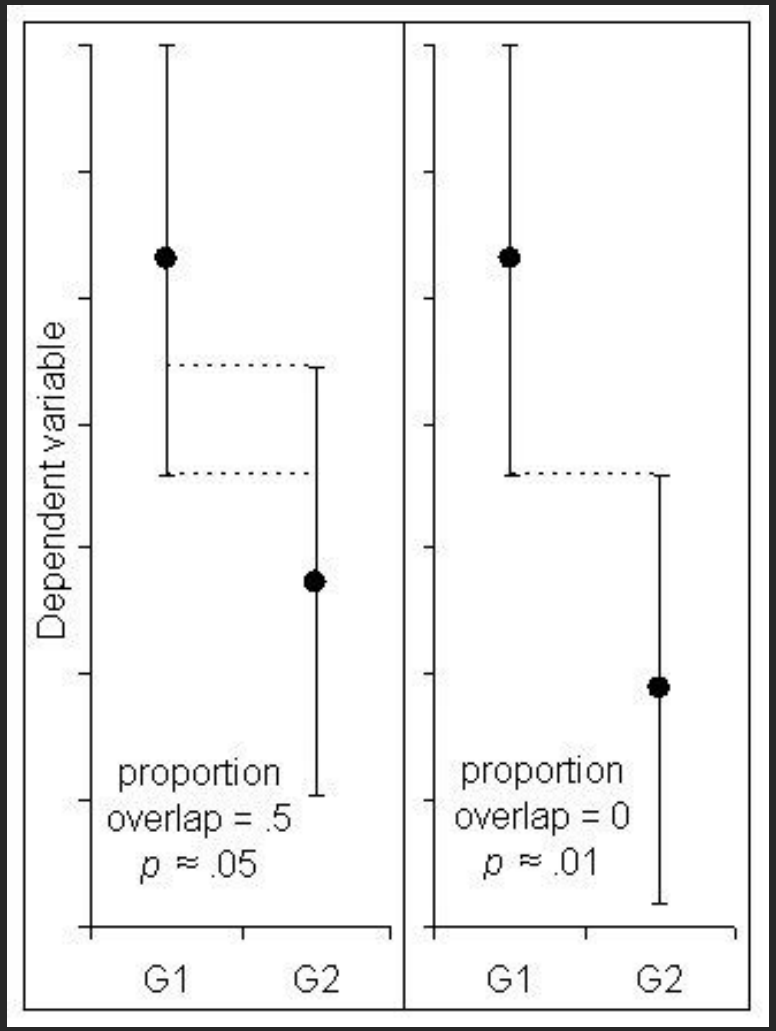


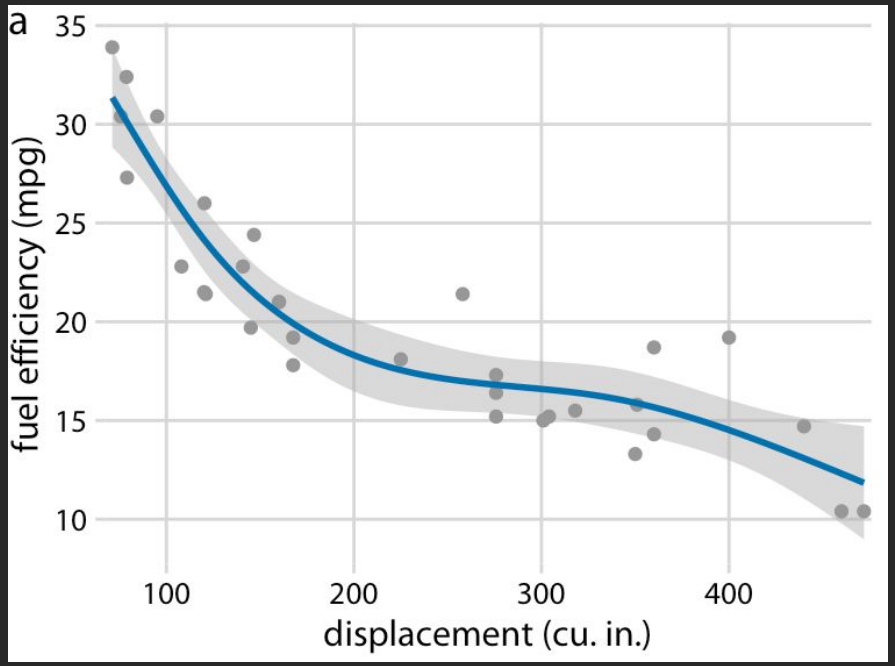
Error bars/areas

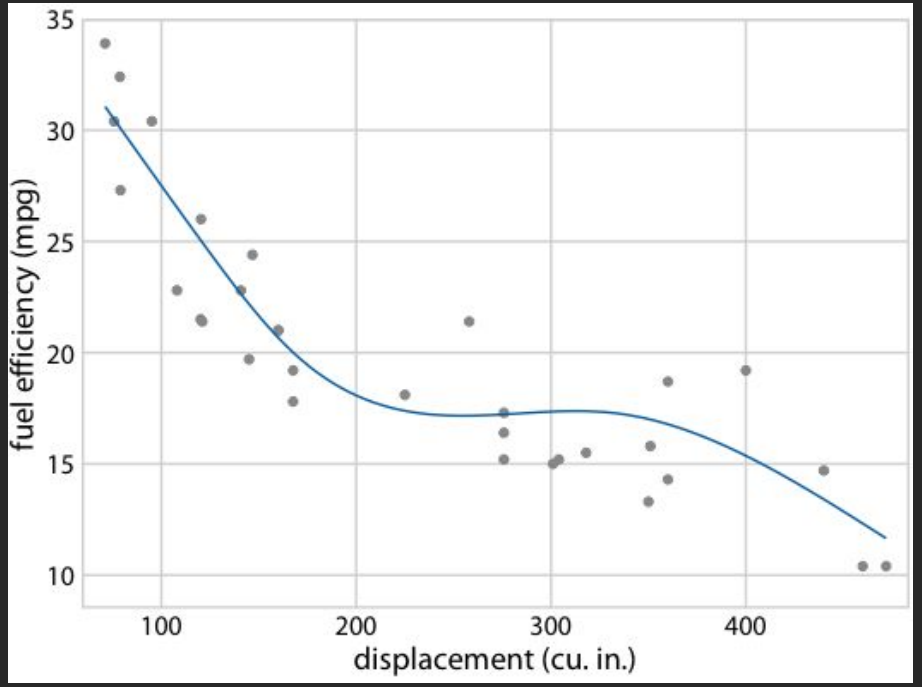

Torgersen
Biscoe
Dream
island
3000
3500
4000
4500
5000
5500
6000
body_mass_g

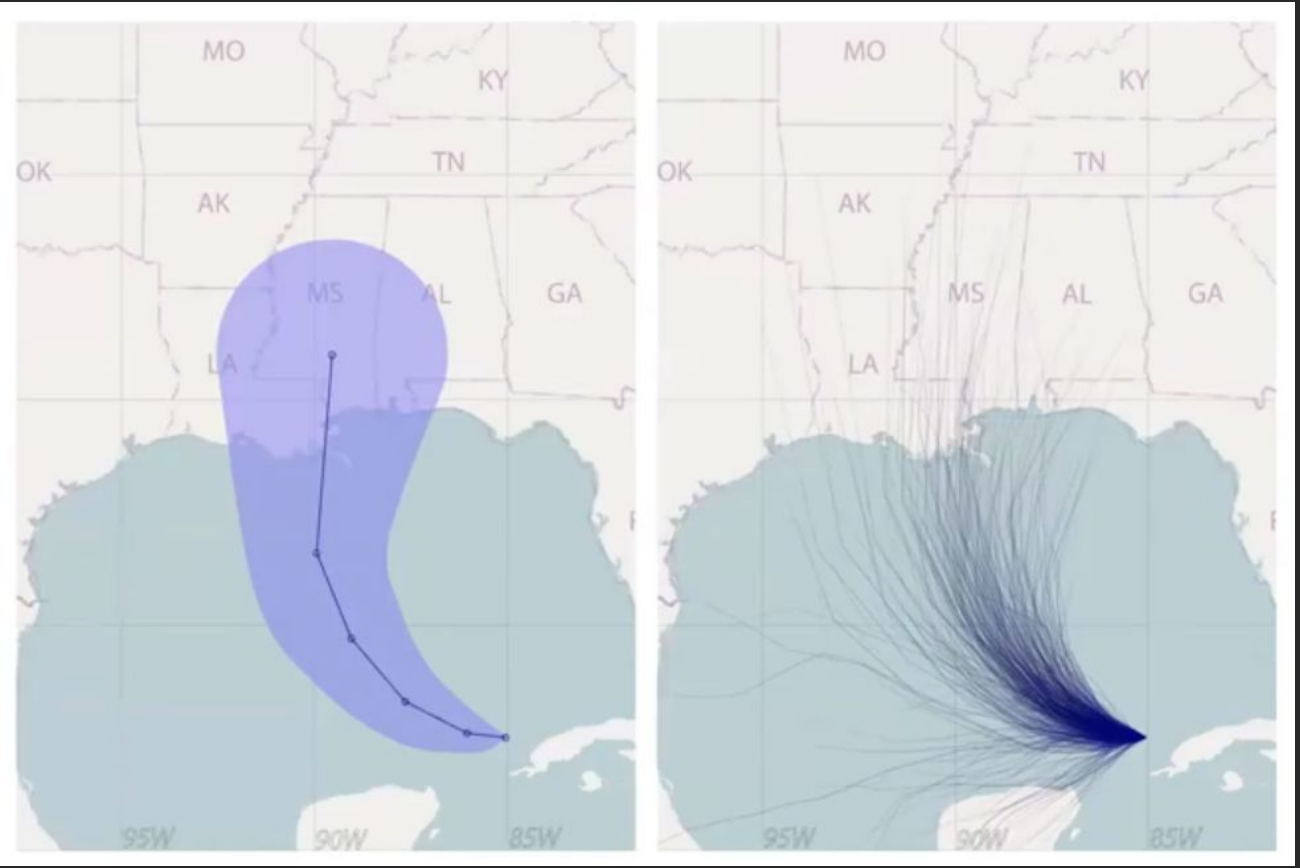

0.0
2.5
5.0
7.5
10.0


mean
sample
+/- standard deviation
+/- standard error
80% confidence interval
95% confidence interval
99% confidence interval
2.0
2.5
3.0
3.5
4.0
chocolate flavor rating


Dependent variable
proportion
overlap = .5
p ≈ .05
G1
G2
proportion
overlap = 0
p ≈ .01
G1
G2

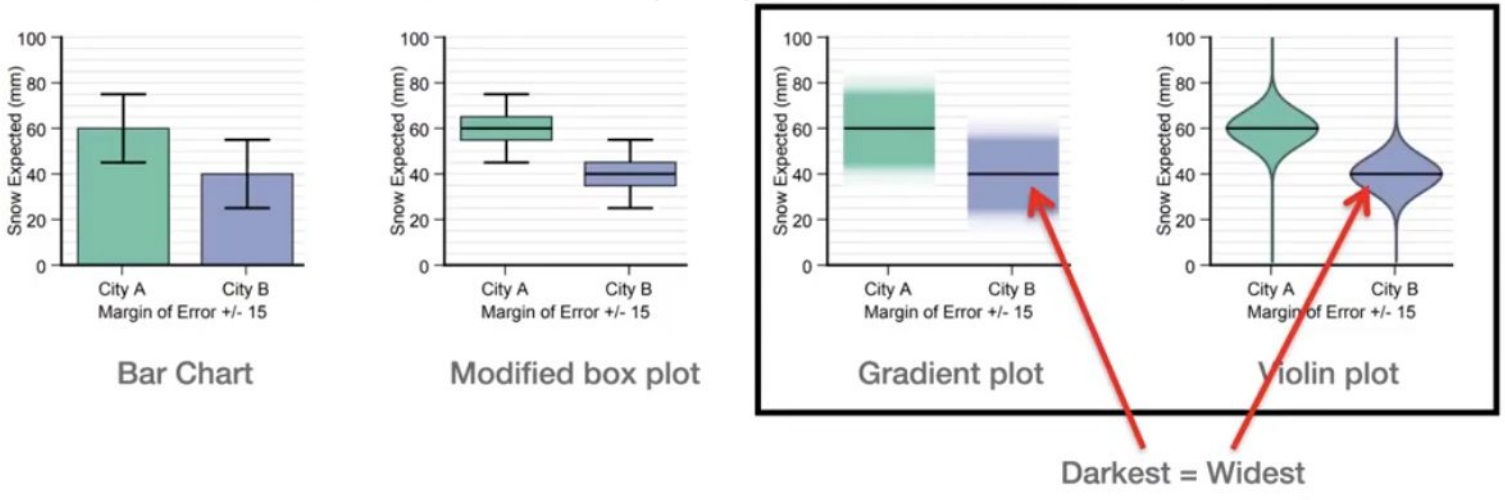

a
35
30
25
20
15
10
fuel efficiency (mpg)
100
200
300
400
displacement (cu. in.)


35
30
25
20
15
10
fuel efficiency (mpg)
100
200
300
400
displacement (cu. in.)

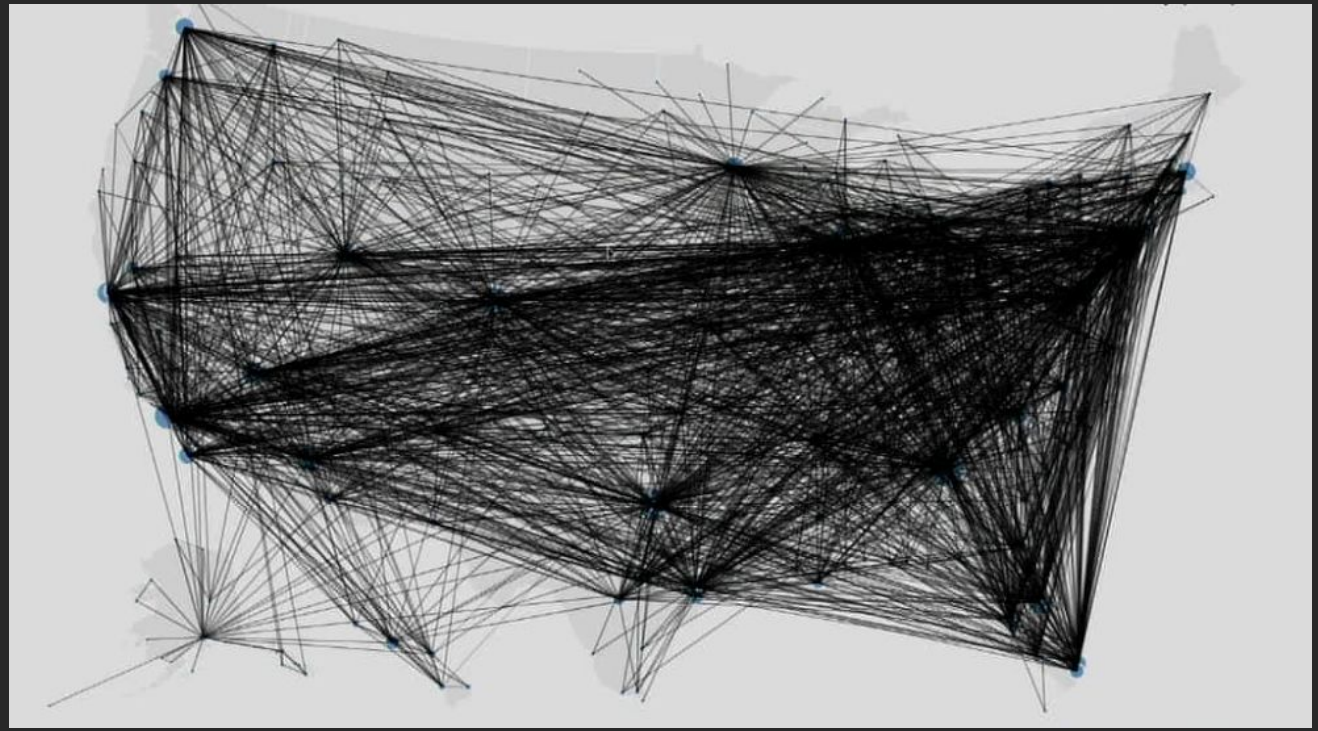

MO
KY
OK
TN
AK
MS
AL
GA
LA
95W
90W
85W
MO
KY
OK
TN
AK
MS
AL
GA
LA
95W
90W
85W

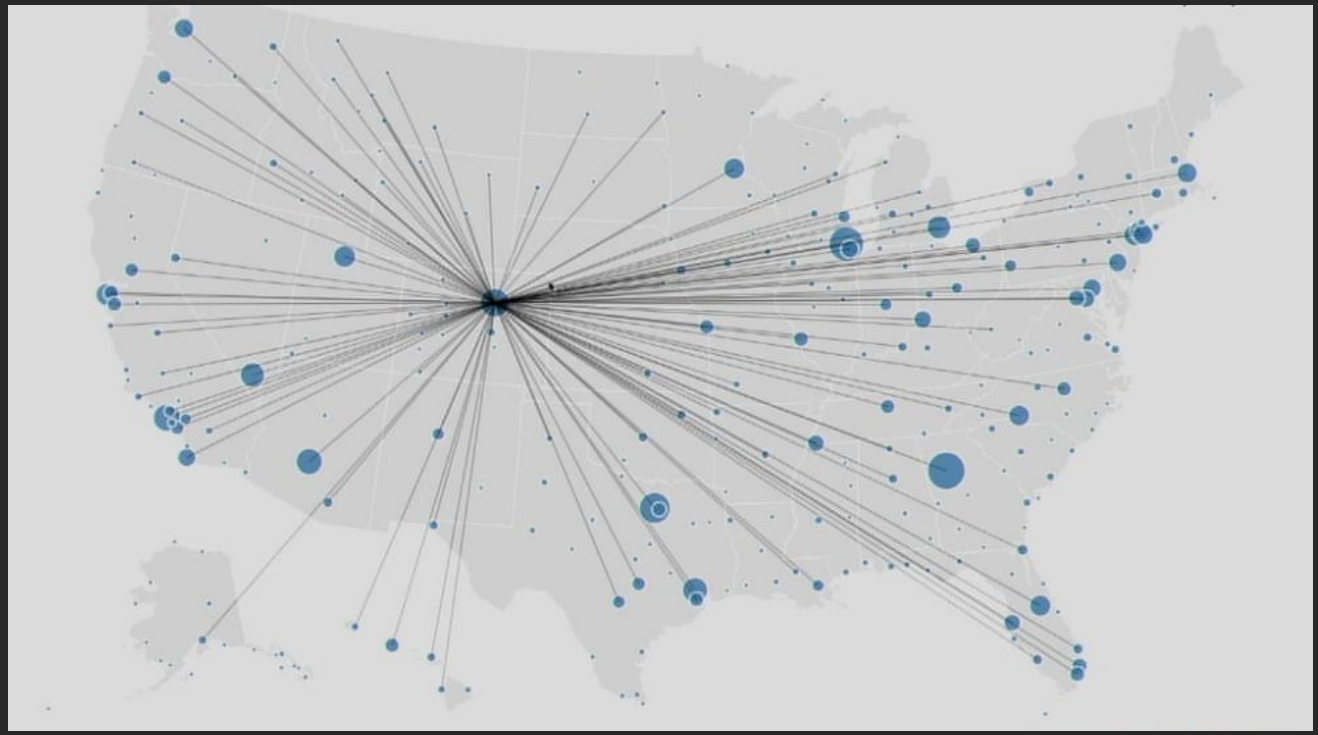
Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error

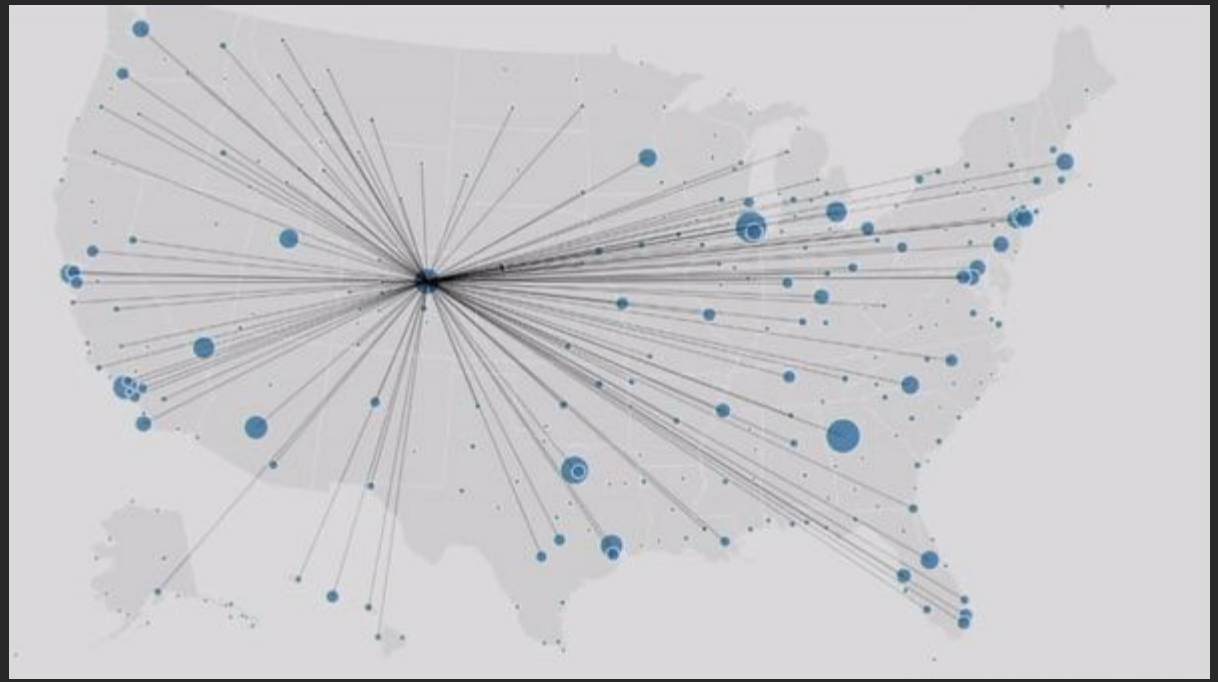
Michael Correll *Student Member, IEEE*, and Michael Gleicher *Member, IEEE*



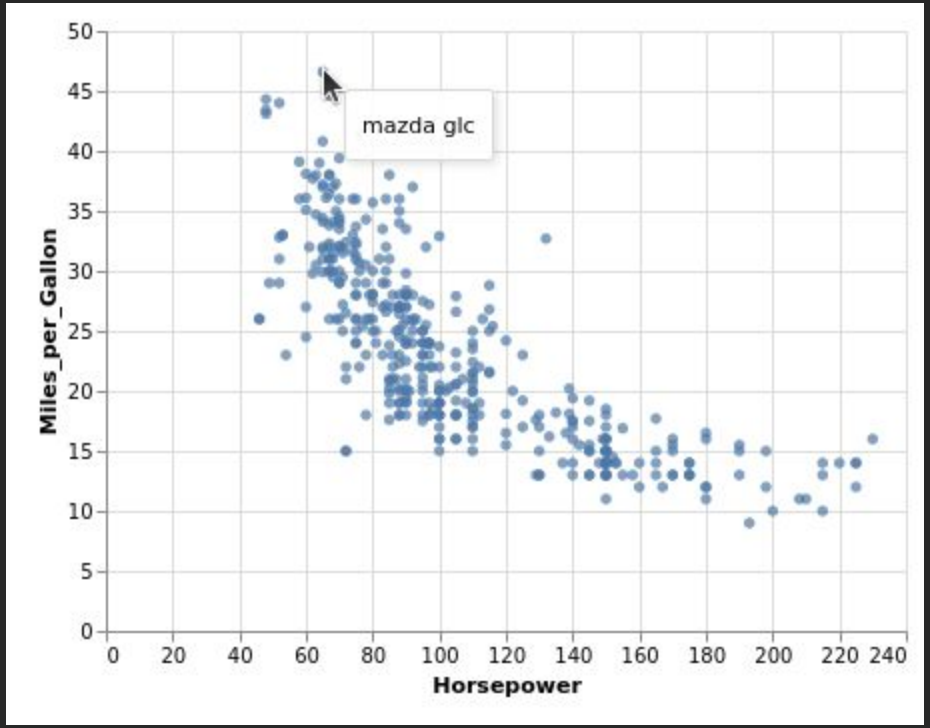
Interactive visualizations

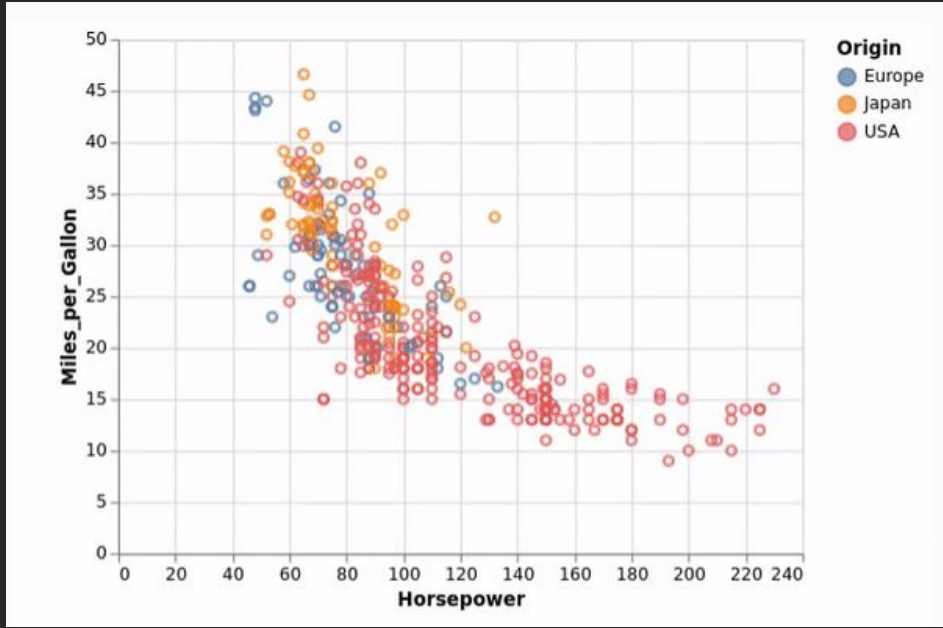


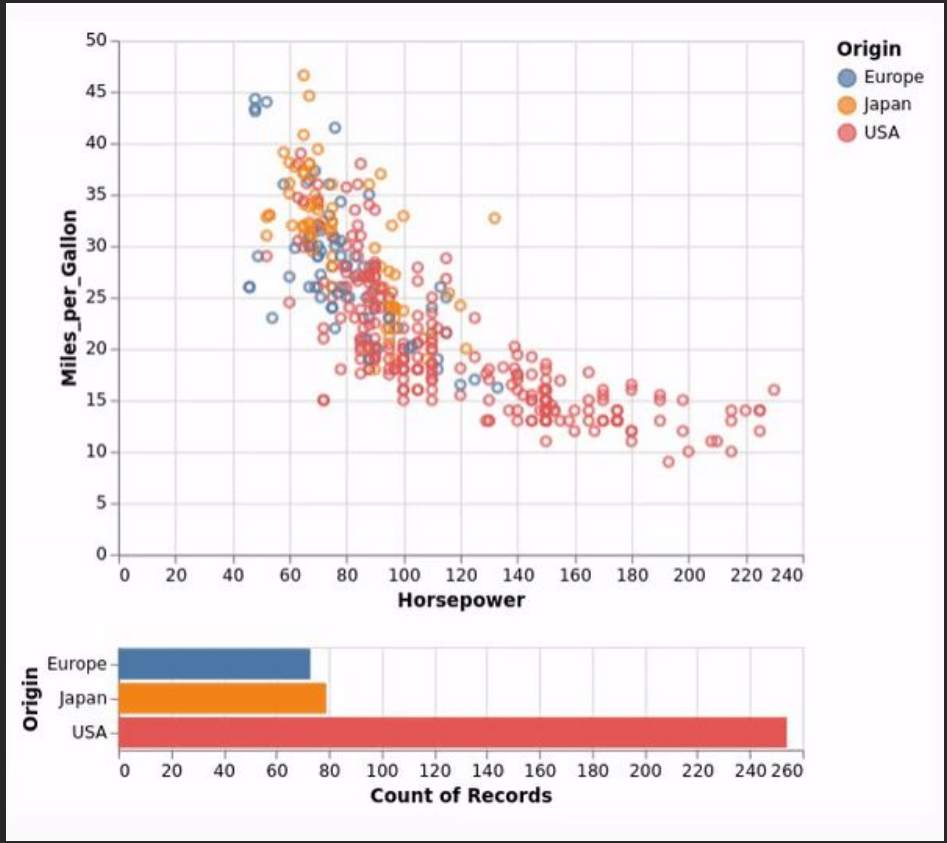


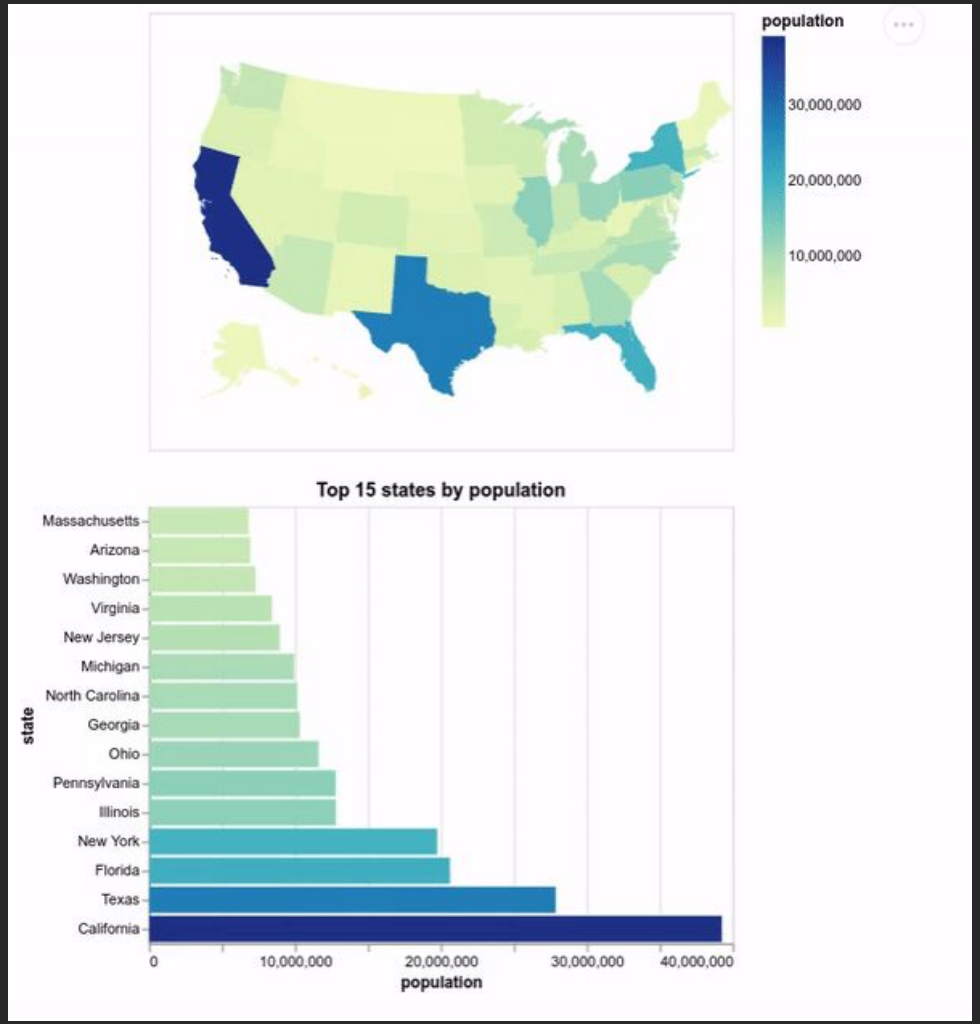


1. Panning and zooming
2. Details on demand
3. Highlighting
4. Filtering


mazda glc
Miles_per_Gallon
Horsepower
0
5
10
15
20
25
30
35
40
45
50
0
20
40
60
80
100
120
140
160
180
200
220
240


Origin
Europe
Japan
USA
Miles_per_Gallon
Horsepower
0
5
10
15
20
25
30
35
40
45
50
0
20
40
60
80
100
120
140
160
180
200
220
240


Origin
Europe
Japan
USA
Miles_per_Gallon
Horsepower
Origin
Count of Records


population
30,000,000
20,000,000
10,000,000
Top 15 states by population
Massachusetts
Arizona
Washington
Virginia
New Jersey
Michigan
North Carolina
Georgia
Ohio
Pennsylvania
Illinois
New York
Florida
Texas
California
state
0
10,000,000
20,000,000
30,000,000
40,000,000
population