Dataset to be Visualized

My dataset is about bad drivers in the US across different states. Is there some general pattern about bad drivers that I can explore? In addition, I want to explore which state has the worst drivers and how insurance company set the insurance premium across the states.

Schema of the bad-driver dataset:

bad-drivers1 State	bad-driver... Geo...	bad-drivers1 Fatal Pe...	bad-drivers1 Spee...	bad-drivers1 Alcohol...	bad-drivers1 Distr...	bad-drivers1 No Previous ...	bad-drivers1 Car Insuran...	bad-drivers1 Insurance Inc...
South Dakota	cw	19.4000	31	33	2	86	669.310	96.870
North Dakota	cw	23.9000	23	42	1	86	688.750	109.720
Ohio	cw	14.1000	28	34	1	82	697.730	133.520

State – US State names
Geotype – Every state has a geography category indicating which area the state belongs to. (cw – Central Western, ne – Northeastern, s – Southern, w – Western)
Fatal Per Billion Miles – Number of fatal collisions per billion miles in that state
Speeding - Percentage Of Drivers Involved In Fatal Collisions Who Were Speeding
Alcohol-Impaired - Percentage Of Drivers Involved In Fatal Collisions Who Were Alcohol-Impaired
Distracted - Percentage Of Drivers Involved In Fatal Collisions Who Were Distracted
No Previous Accidents - Percentage Of Drivers Involved In Fatal Collisions Who Had Not Been Involved In Any Previous Accidents
Car Insurance Premium – How much the insurance company will pay ($)
Insurance Incurred per Driver - Losses incurred by insurance companies for collisions per insured driver ($)

Visualization Techniques

I am using D3 to visualize my dataset and I will mainly use 3 kinds of graphs to present my raw data and my analysis. They are map, bar chart, and scatter plot.

Map

- Diverge color Scale mapping
- Size mapping
- Hover effect (Color change, svg:title)

Bar Charts

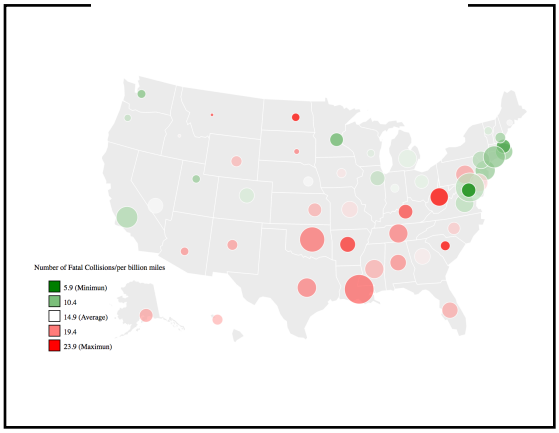
- Ordinal color encodings
- Selector to filter data
- Hover effect (Color change, reference line generated)

Scatter Plots

- Selector to filter data
- Automated regression generated (Trend line)
- Hover effect (Color change, State name reminder)
- Ordinal color encodings

Storyboard

Overall perception



The map can help us find some potential pattern of geographical distribution.

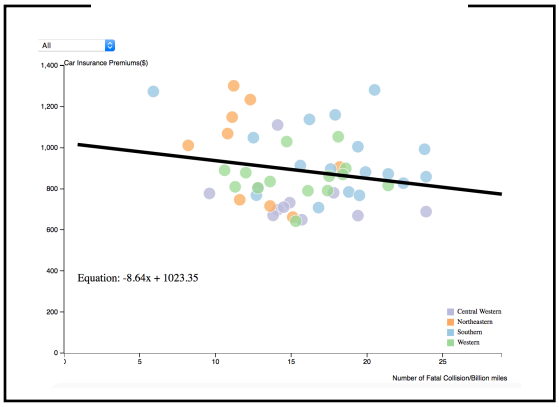
- I map the number of fatal collision per billion miles to a diverging color scale to see which state has above or below average number of fatal collisions. And we can easily get that there is an obvious geographical distribution pattern. Southern area seems to have more fatal collisions per billion miles.
- I also map the insurance incurred amount per driver to the size of the circles to see whether it has some geographical distribution pattern.
- There is a hover effect that when you hover the point, you can know the detailed statistics of that specific state.

Dig deeper into details

It seems there is a geographical distribution pattern about bad drivers. in order to better explore other patterns across the regions, I grouped the data by their geography regions and used different colors to encode the 4 regions, Central Western, Northeastern, Southern and Western.

- The left bar chart is used to explore the causes of fatal collisions across states. We can use a selector to check out the percentage of different causes of fatal collisions, including speeding, alcohol-impaired and distracted.
- The Right bar chart is used to present the percentage of drivers not having any previous accidents, also grouped by states.
- For both bar charts, there is a hover effect that will generate a reference line to help reader read the values and compare with other bars.

Regression Analysis



After having a better understanding what kind of bad drivers each state has, we can examine the relationship between the number of fatal collision and car insurance premium paid by insurance company.

- To better perceive difference between different regions, we can use the selector to filter data by a specific region, and the trend line and its equation will vary accordingly. Through this, reader can explore the regional phenomenon freely.
- Reader can brush the graph to get a closer look at a specific interval. In this case, reader can have a better understanding about the variance and outliers information.