Career Simulation 1 - Alex Munger

This dataset encapsulates workplace deaths, injuries, illnesses, and inspections in the United States. The data is split by state. Because the simulation relied heavily on geography, one of the first steps in enhancing visualizations was getting the latitude and longitude of each state in order to plot heatmap overlays on an actual map of the United States. After this was accomplished, a number of rate metrics were calculated for state-by-state comparison. Among these were Fatalities per Inspector, Injuries per Inspection, and (dollar) Penalties per Inspection.

The purpose of this analysis was to simply “audit” the states’ workplace safety: whether they were over or under-performing in preventing workplace injury or death. Clearly, the main deterrent for workplace accidents are inspectors and inspections, at the state and federal level. For each state, the data reported was either state or federal, so one of the PowerBI slides allows users to split the two and observe any biases that they may find in state vs. federal inspection and reporting.

Map visuals, bar graphs, and line graphs were all used in the making of this report. The line graph was made for the metric “Penalties per Inspection”, and an appropriate trendline was added. Naturally, there is a positive correlation here. The more inspections a state received, generally speaking, the more penalties occurred. The usefulness of this graph is in its ability to identify which states may be “over” or “under” inspected. For example, South Dakota was inspected the most by almost a standard deviation, but accrued below average penalties per inspection, indicating they were “over-inspected”. With limited resources, this is useful information to the federal inspectors that were reporting for South Dakota. California and Ohio, on the other hand, accrued penalties at a high rate per inspection, meaning, if they were inspected more, there may be less violations as employers would be more incentivized to remain compliant.

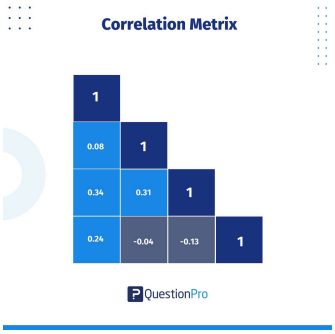
In general, states with federal inspections had a higher fatality rate (see PowerBI tab “Fatality Rates State/Federal”). The fatality rates were 4.42 for federal inspection states and 4.18 for state inspection states.

California, with a state program, had the highest number of injuries/illnesses (345,000). Wyoming, also with a state program, had the lowest number of injuries/illnesses (6,500). State or federal program injuries and illnesses can be split by table and map in the Power BI tab “Inspections vs. Injuries/Illnesses”.

The PowerBI features used for this analysis include heatmaps (by state), tables with conditional formatting, gauges with summary metrics, line graphs detailing the relationship between two quantitative variables, column charts to visualize differences by state, and slicers to separate federal from state programs in nearly every visual.

This analysis attempts to tell a story through variety in presentation, succinct and concise conclusions, disclaimers on gaps in data, colors, conditional formats, and spatial parametrics like maps.

Future improvement of this project would certainly begin with the data source. Some entries are blank, mainly in reports on the number of injuries and illnesses. As this is a core variable in the analysis, it hampers the overall ability to present findings somewhat. Secondly, instead of different tabs each with a map, I could try to incorporate one map tab with a dropdown of different metrics to heatmap, allowing the map and the corresponding table to be updated and sliced dynamically. Third, and I have yet to master this in PowerBI, I would present a Correlation Matrix, which is an *n x n* matrix of correlation coefficients, with 1.00s down the square diagonal, since two variables correlate perfectly with themselves. Below is an example matrix to show the visual, from the web:



This is almost trivial to do in Excel, but has presented some difficulties for me in PowerBI. In R, you can do it even better, where each correlation plot is shown along with the coefficient value.

Overall, I hope this analysis gleaned some insight into the data, its trends, usefulness, and lack of usefulness in certain aspects.