Natalie Daugherty
Data Analytics and Visualization Bootcamp
11/10/2024

Data Analysis on “Workplace Fatalities by State”

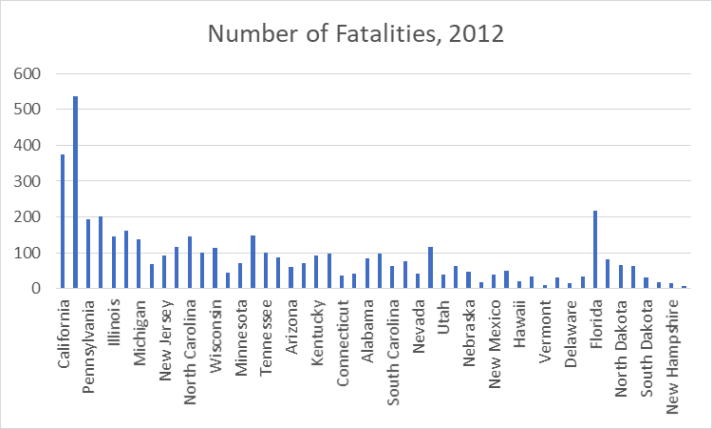
The objectives of this analysis were to determine if there were relationships between the rate of workplace fatalities between federal and state programs, years to inspect vs. rate of fatalities,and which state had the highest number of illnesses and injuries. Using simple charts and Pivot tables, I was able to demonstrate the following. Federal programs have a higher rate of fatalities than state programs. There is no set relationship between years to inspect and the rate of fatalities. Finally, California has the highest number of illnesses and injuries when the incidents are looked at by state.

This data was made available for the purpose of this analysis. Data was cleaned by searching for unneeded white space, formatting the data into a table, and addressing missing values. Missing values were addressed by taking the mean of the column variable and entering that as the entry in place of blank fields. No duplicate values were found. Microsoft Excel was utilized to clean and analyze the data. Pivot Tables, Charts, and Slicers used.

The dataset includes entries for all fifty states in the U.S. Variables related to injuries and fatalities in the workplace for 2012 are the focus of the table. Additional information to give context to the primary data is included such as the amount of years to inspect a workplace as well as if a state or federal program is in place. The data is numerical in nature and given in number form. Some fields are text based as well.

In analyzing this data, I am most interested in counts and averages. Counts allow for identifying metrics such as how many times a value occurs in a range and can help answer yes/no questions. Averages allows for demonstrating relationships between aspects of the table, such as years to inspect vs fatality rate.

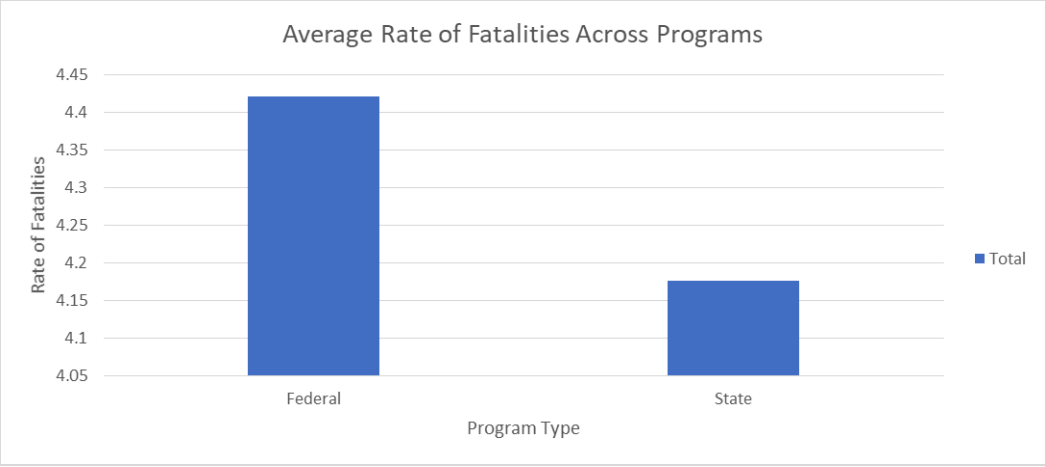
The dataset includes a wide range of values across fatalities vs injuries/illnesses. The maximum value for fatalities is 536, while the minimum is 8. This wide range gives an average of 92.34. Illnesses and injuries display a larger range, of almost 34,000. Notably, the median and average for both variables is the same.



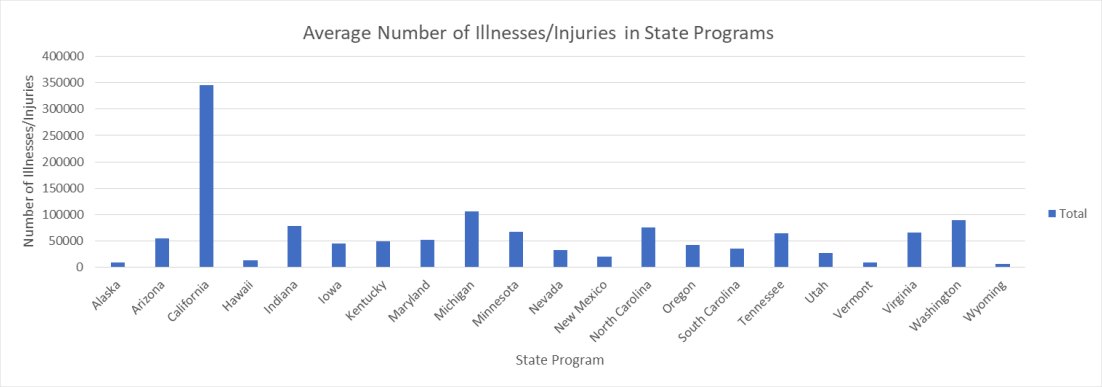
Eight states did not report any illnesses or injuries during 2012. These states were not excluded from the dataset since they did have reports for fatalities and penalties. Instead, the median value for that variable was entered for blank data.

To conduct the analysis, the data was formatted into a table using Excel. As stated earlier in the document, blank data was addressed and mean/median values found. Pivot Tables were drawn from the data to highlight the questions posed in the objective. Simple Charts were used to show the relationship between inspections and rate of fatalities. On the dashboard of the Excel document, the data was filtered using Slicers. State or Federal Program, individual state and years to inspect were used as different ways to sort the data.

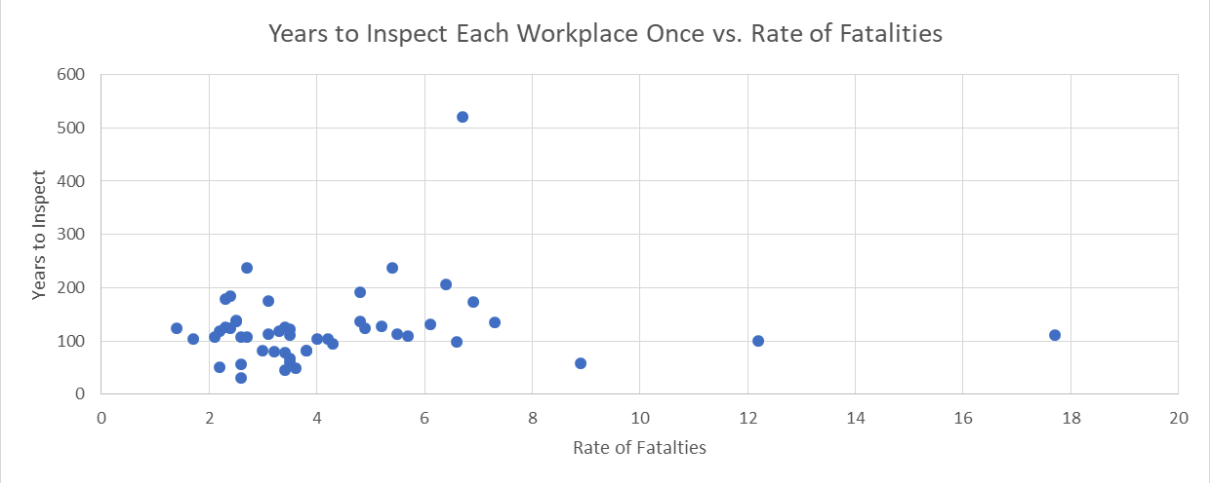
The visuals used in this analysis were derived from either the data tale as a whole or Pivot Tables. The first question asked what program, state or federal, had the highest rate of fatalities. I made a pivot table using the variables, program type and average of fatality rates. A column chart was then used to show that federal programs have the highest rates of fatalities in this dataset.



The second question asked which state, having a state program, had the highest number of illness/injury. This pivot table used states and average of illness/injuries with a filter for state/federal programs. This chart is also a column chart and shows that California has the highest number of illness and injuries while also having a state program.



The third question asked if there is a relationship between the years it takes to inspect a facility and the rate of fatalities. This chart, a scatter chart, shows a few outliers, then a well distributed dataset among the remaining values. This shows that there is no set relationship between the length of time to inspect a family once, vs the rate of fatalities.



A dashboard is left in the Excel document to display these charts along with their Slicers to further break down data. Additional questions were explored- what is the relationship to state rank of fatalities vs. the penalties rank, and which state has the highest number of fatalities regardless of program. The first analysis, using a scatter plot, shows there is no relationship between the ranks of fatalities or penalty rank. The second question shows that Texas has the highest rate of fatalities regardless of program.

Through analyzing this data, it is clear that state programs are linked with lower fatality rates as well as lower injury/illness rates. Interestingly, there does not seem to be a relationship

between the average of years to inspect a facility and the rate of fatalities. There is also no relationship between fatality ranks and penalty ranks. I would have hypothesized that the higher rate of fatality would have caused more penalties, therefore a higher rank. The larger states, California and Texas do have higher rates of fatality, which just by population size does make sense.

To improve my project in the future, I would include information on how the penalty amount relates to injury/illness and fatality rate. I would also experiment more with how deleting rows with empty fields might affect the dataset as a whole. Another step to improve the project would be to look at year over year data to determine larger trends.