

2012 US Workplace Injuries/Illnesses and Fatality Report

By: Justin Blackshear

Data Exploration:

The provided data provided information regarding workplace safety during the year 2012 within the United States. The data included the referenced state, it's number and rate of both fatalities and injuries/illnesses, the number of inspectors and the number of years to inspect each workplace once, and the states ranking based on their number of and rate of fatalities and injuries/illnesses.

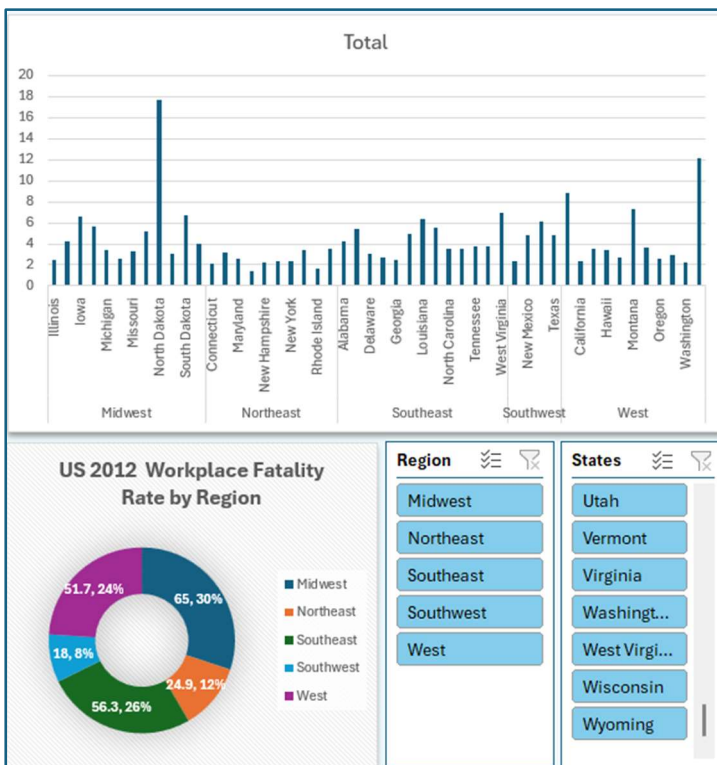
Developing the Data within Excel:

To make use of the data I turned the provided data into a chart. This allowed me to more easily discern aspects and create visuals for the data. To avoid errors and miscalculations I checked for any duplicate data and filled in any blank cells within the data. Columns with missing data included the "Number of Injuries/Illnesses 2012" column, the "Injuries/Illnesses 2012 Rate" column, and the "Inspector" column. Since the missing values were all numeric, I was able to use the median value of each column to fill in their missing values. I found the median of each column by using the 'descriptive statistics' function within the 'Data Analysis' toolkit.

To better utilize and understand state data, I used the trim function within Excel to remove the coordinates provided with each state in the "State" column. I named this new column "States" and used this new column to create visuals. I also thought there was room to develop the data to create more incites, so I categorized groups of states by their region within a new column named "Region". This allowed me to not only see how state injury/illness and fatality rates compared to each other, but also how different regions compared to each other.

Visualizing the Data within Excel:

I was very interested in how fatality rates differed and compared between states, the different regions within the United States, and whether the governing body had federal or state oversight. The bulk of my 'Dashboard' included visual representations of these relationships. I first included the following charts to give a general understanding of how fatality rates compared between states and regions.

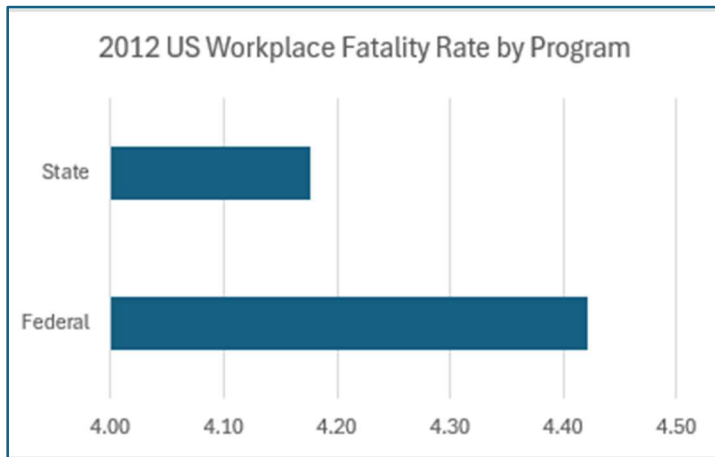


I included splicers with the bar chart so that its information could be customized. You can either see all the states and their regions, select a specific region to inspect, or even a specific state. The pie chart gives a concise depiction of how the fatality rates between different regions compare with the Midwest and Southwest having the highest and lowest rates respectively. I also provided a map of the United States displaying state by state fatality rates. This provides a geographic visual of how the fatality rates compare.

I used the rest of my 'Dashboard' to answer the following questions I had about the data:

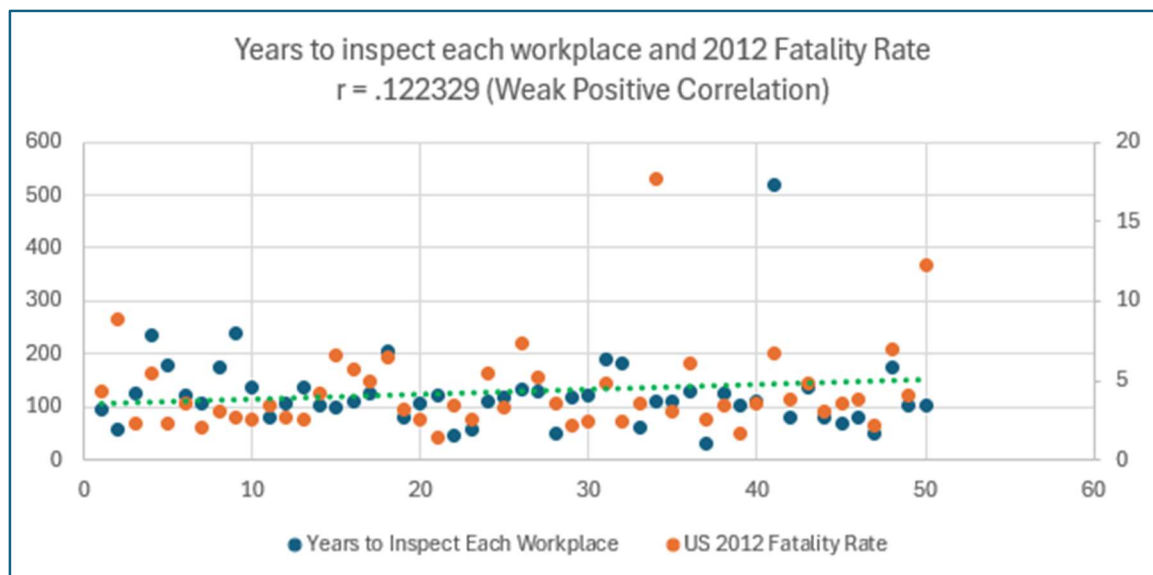
1. Which program, state or federal, has the highest rate of fatalities?

- a. As the visual below supports, Federal programs had a higher average fatality rate.



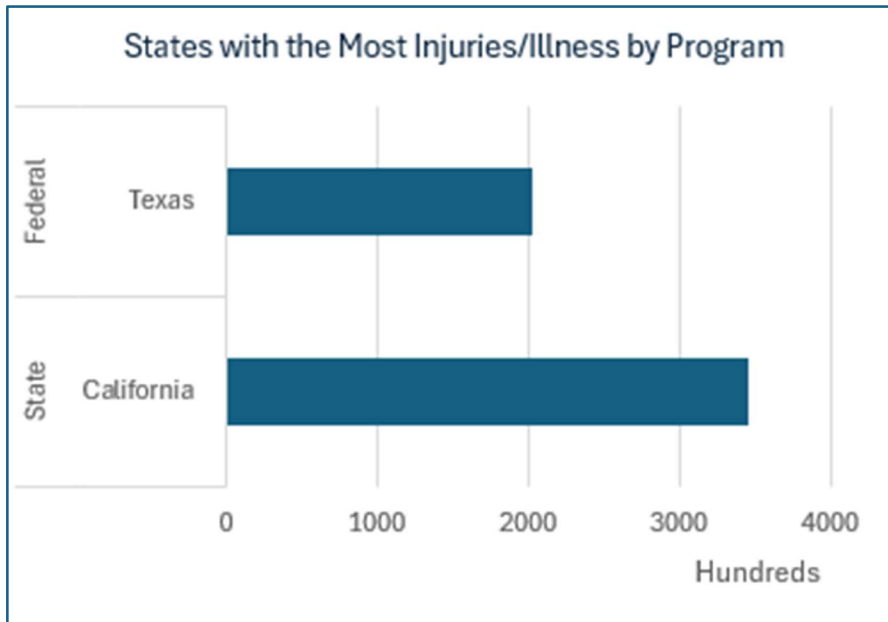
2. What is the relationship, if any, between "Average of Years to Inspect Each Workplace Once" and "Rate of Fatalities"?

- a. I used the r-value (correlation) between "Average of Years to Inspect Each Workplace Once" and "Rate of Fatalities" to determine their relationship. The r-value came out to .122329. Based on the data we have; this indicates a weak-positive correlation between the two sets of data but given more data it is possible that there is no relationship or perhaps even a stronger relationship than represented by the current data set.



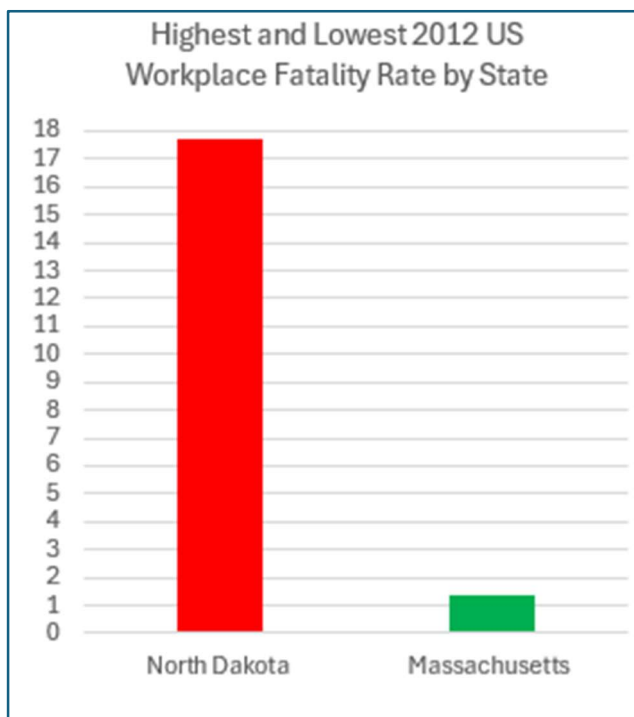
3. Which state with a state program has the highest number of injuries/illnesses?

- a. I was able to determine California had the highest number of injuries/illnesses for a state with a state program. I found it helpful to include that Texas had the highest number of injuries/illnesses for a state with a federal program because this suggests that it's possible these states have the highest number of injuries/illnesses due to their population. California is the most populous state in the United States and Texas is the second most populous state.



4. Which states have the highest and lowest workplace fatality rate in the United States?

- a. I was able to determine that North Dakota had the highest workplace fatality rate in the United States and Massachusetts had the lowest workplace fatality rate in the United States. It's also notable to point out that both of those states are under federal programs.



Conclusions, additional questions, and improvements:

I'm very interested in what factors contribute to state workplace fatality rates, but that question wasn't answered by the given data set. A common theme among states with higher fatality rates was their lack of population with the sparsely populated North Dakota, Wyoming, Alaska, and Montana with the four highest rates, but Texas has the 17th highest fatality rate, so population isn't the only contributing factor. It's possible that it is more about population density than overall population, but Georgia and Arizona (states that aren't particularly densely populated) have among the lowest fatality rates in the United States. Since we weren't provided the population or population density of states and there was at best a weak-positive correlation between "Years to Inspect Each Workplace Once" and "Rate of Fatalities, 2012" we would need to have additional data provided to determine factors with a higher correlation.

Additional questions I had were about why North Dakota and Wyoming (as shown in the box and whisker plot below) are such notable outliers regarding their workplace fatality rate.



I also have additional questions about the data set itself as it is not abundantly clear what the data represents. What is the injury/illness and fatality rate based on? What are the units of the rate? What is the "Year to Inspect Each Workplace Once" telling us? Are we referring to specific workplaces or all workplaces within a state? What does it mean for a state to have a federal versus a state program in this context? With so many questions about the data itself, it's practically impossible to come to any meaningful conclusions.

Ways I would make the project better in general would be to have more descriptive and clear categories of data as well as provide more categories to work with. To make my project, itself, better I could have added additional columns of data myself (population, population density, political affiliation, etc.) to try to determine factors that have a strong correlation with workplace fatality rates.