Viz3: Dangerous Jobs: Shashwat Singh
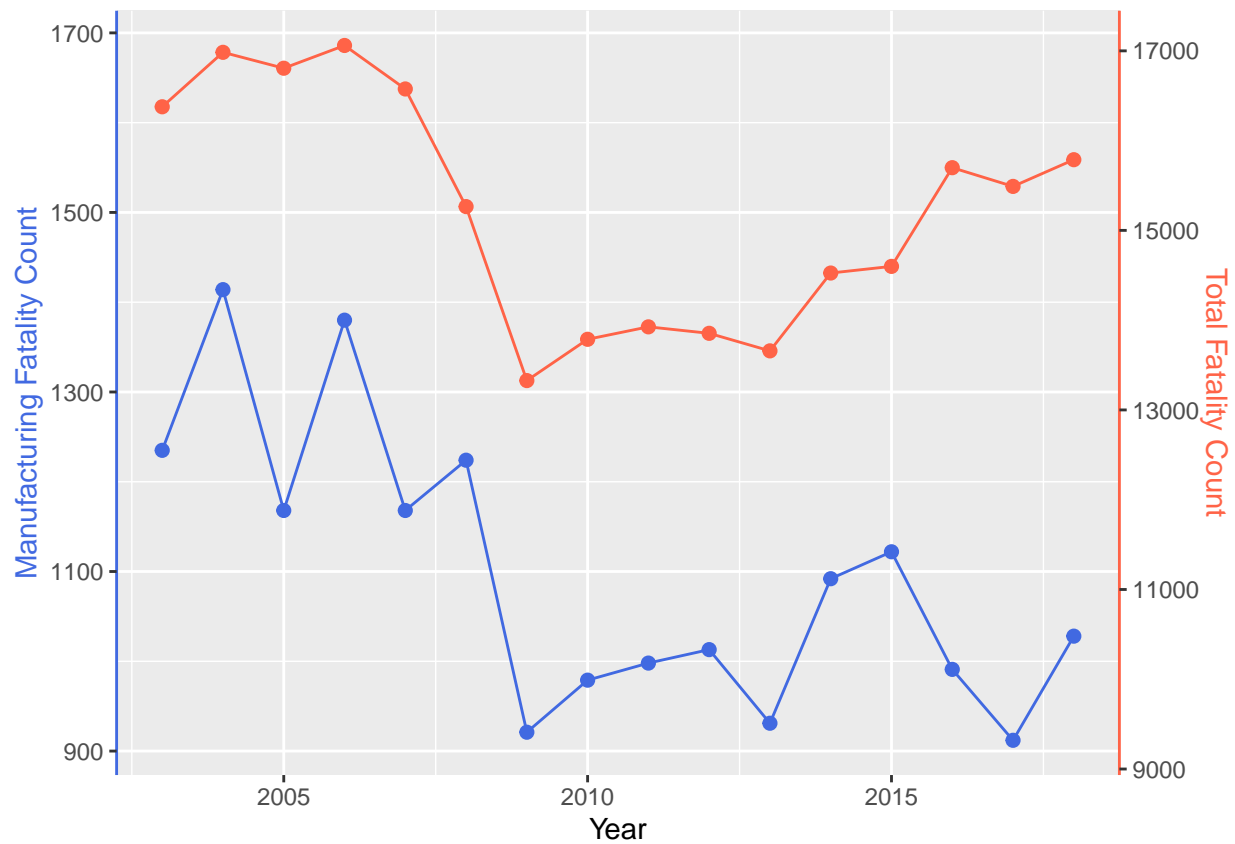

Figure 1. Figure 1. The above figure contains Line Plots with markers showing trends in the fatalities on jobs between 2003 and 2018. Red shows the Total Fatality count on all jobs, while Blue shows fatality counts on jobs in Manufacturing. The Total fatality count is scaled down by a factor of 20. The raw data was filtered to contain only "Value" rows from the Total or Value column, and Total counts were extracted using Excel. Any NA values were discarded for data integrity. Data source: U.S. Bureau of Labor Statistics' Census

Figure 2. The above figure contains a Line Plot with big triangle markers showing trends in the fatalities on jobs between 2003 and 2018. *Red*, shows the Total Fatility count while *Blue* shows fatality counts on jobs in Manufacturing.This plot contains major scaling issues. It is so off that at 2 points, the count of manufacturing fatalities is going above the total count (not possible). The raw data was filtered to contain only "Value" rows from the Total or Value column, and Total counts were extracted using Excel. Data source: U.S. Bureau of Labor Statistics' Census

Steps for Data Preprocessing

- Filtered the Total or Value column for "Value".
- For each year column, used the =SUM() formula to get the total count of fatalities, copied the data in a csv file "Assignment3_ProcessedData.csv"
- For Manufacturing role occupation, filter the Occupation column for "Manufacturing" and performed step 2 again
- Saved the excel file with the data.

Code Used To Make These Plots

```r
library(ggplot2)
fatality_data<-read.csv("Assignment3_ProcessedData.csv")

#Pretty Plot
ggplot(fatality_data, aes(x = Year)) +
  geom_line(aes(y = Count.Manufacturing), color = "royalblue") +
  geom_line(aes(y = Count / 20), color = "tomato") +
  geom_point(aes(y = Count.Manufacturing), color = "royalblue", size = 2) +   geom_point(aes(y = Count
  scale_y_continuous(name = "Manufacturing Fatality Count", sec.axis =   sec_axis(~(. + 20) * 10, name
  labs(x = "Year") +
  theme(axis.line.y.left = element_line(color = "royalblue"),
        axis.line.y.right = element_line(color = "tomato"),
        axis.title.y.left = element_text(color = "royalblue"),
        axis.title.y.right = element_text(color = "tomato"))
```

```r
#Illusion Plot

ggplot(fatality_data, aes(x = Year)) +
  geom_line(aes(y = Count.Manufacturing*25), color = "royalblue") +
  geom_line(aes(y = Count), color = "tomato") +
  geom_point(aes(y = Count.Manufacturing*25), color = "royalblue", size = 10, shape = 2) +
  geom_point(aes(y = Count), color = "tomato", size = 7, shape = 25) +
  scale_y_continuous(name = "Manufacturing Fatality Count", sec.axis = sec_axis(~(. + 20) * 10, name =
  labs(x = "Year") +
  theme(axis.line.y.left = element_line(color = "royalblue"),
        axis.line.y.right = element_line(color = "tomato"),
        axis.title.y.left = element_text(color = "royalblue"),
        axis.title.y.right = element_text(color = "tomato"))
```