**Draft R Plot**

The aim of this deliverable is to generate R plots for visualization so that it can easily help us interpret the results of our research questions. In order to create complex visually attractive visualizations following packages are installed in R; - “**ggplot2**”, “**extrafont**” and “**xkcd**”

The first plot tells us the ***state wise pollution in US***. Each bar plot represents the total releases present including onsite total releases, offsite total releases and POTW – Transfers for Release. Looking at the graph we can deduce which state is most polluted and also which state is least. **Figure-1** shows a snapshot of the code used to create this visualization in R using **xkcd package.**

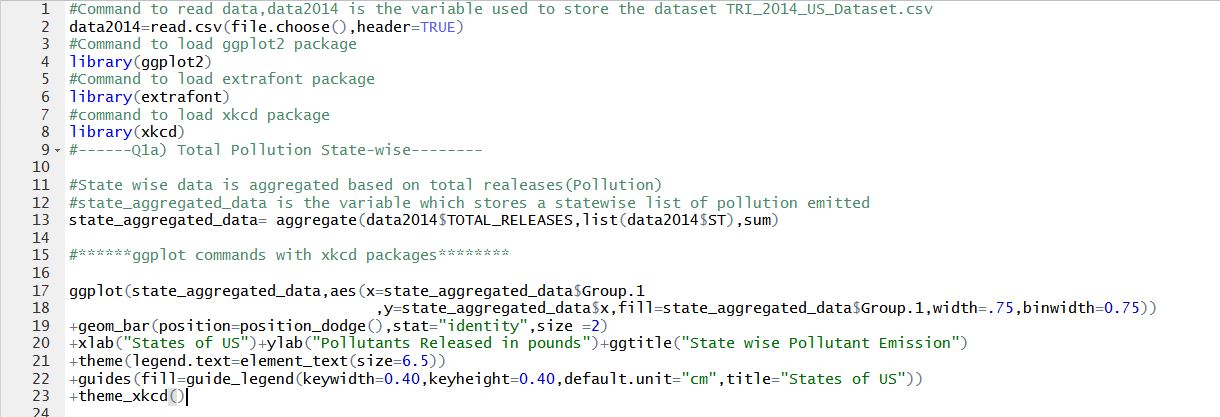


Figure 1

Figure2 represents the visualization for the same

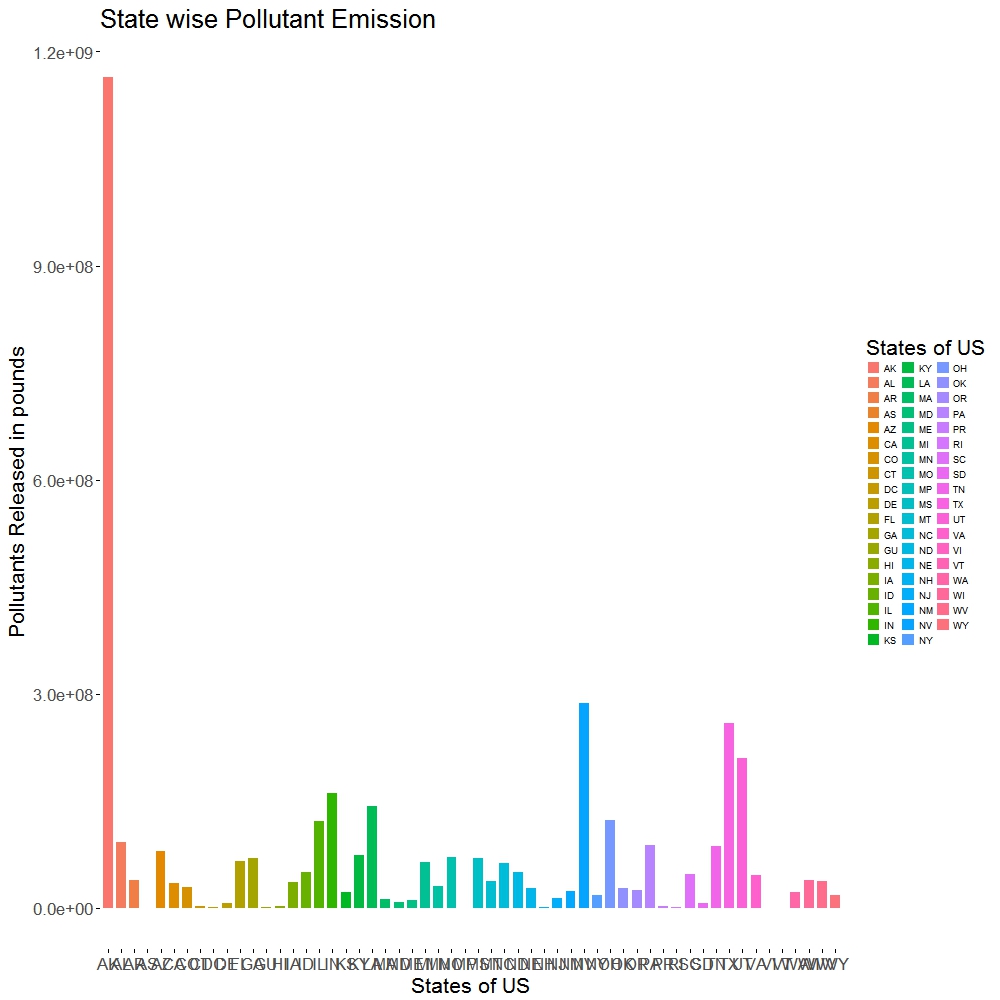


Figure 2

The ***x axis represents the states of US*** and ***y axis represents the total pollution content in each of these states.*** Looking at the graph derived from the data Alaska has the highest pollution .**Figure 3** represents snapshot of the code used to generate the same graph without xkcd package. And **Figure 4** is the visualization for the same.

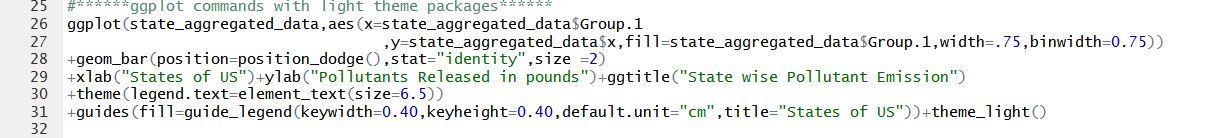


Figure 3

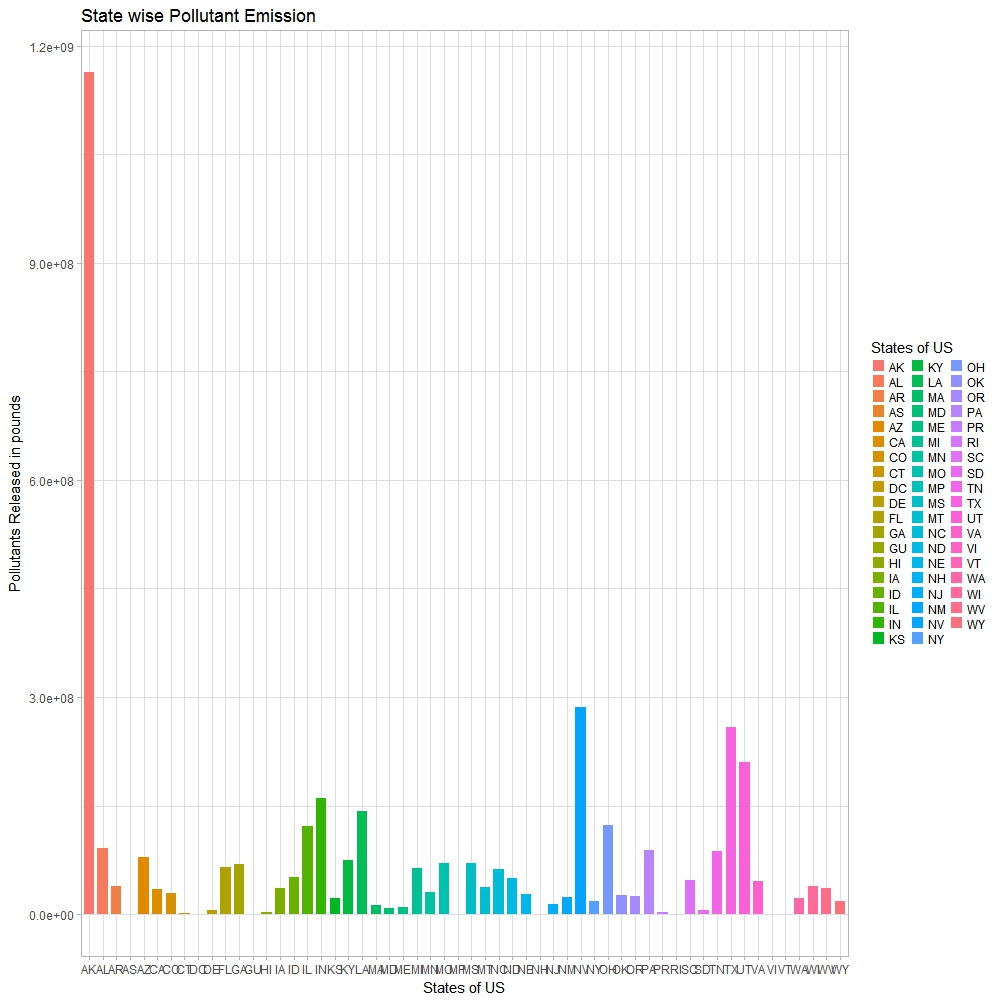


Figure 4

The following code and visualization shows ***which county in each state contributes to highest pollution***. Here we have represented these data only for one state however the same can be replicated for other states as well. **Figure 5** represents the code for which county in a state of Alaska is most polluted. **Figure 6** represents the visualization of county wise pollution distribution in Alaska.

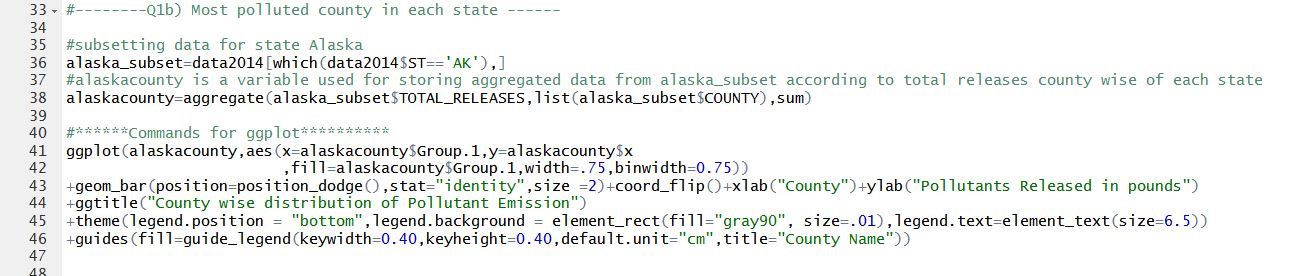


Figure 5

Looking at the visualization of the code mentioned in **Figure 5,** we realize that ***Northwest Arctic County*** is the most polluted county in Alaska, contributing to the maximum pollution for the State of Alaska. It released 1.105033e+09 pounds of pollutants in 2014 .The second most polluted county is ***Juneau County***. This county released 4.155704e+07 pounds of pollutants in 2014 There is a huge difference in pollution content of these two counties .Northwest Arctic County’s pollution is contributed due to ***metal mining industries***, the same is the case with Juneau County.

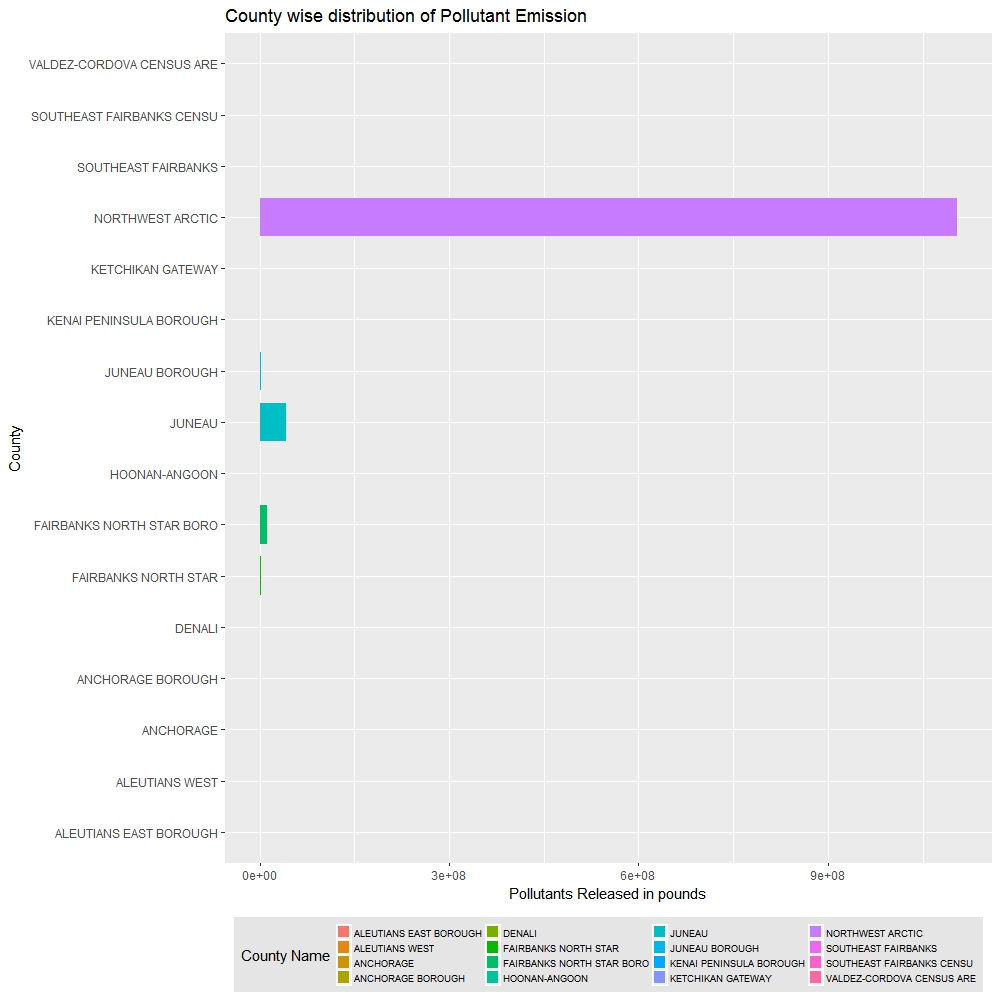


Figure 6

Next we try to find ***out which industry type is causing the most pollution including offsite and onsite releases.*Figure7** represents the code snap in R for the visualization and **Figure 8** represents the plot for the same.

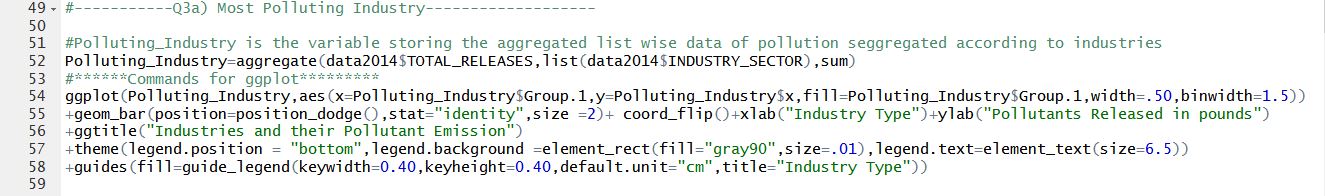


Figure 7

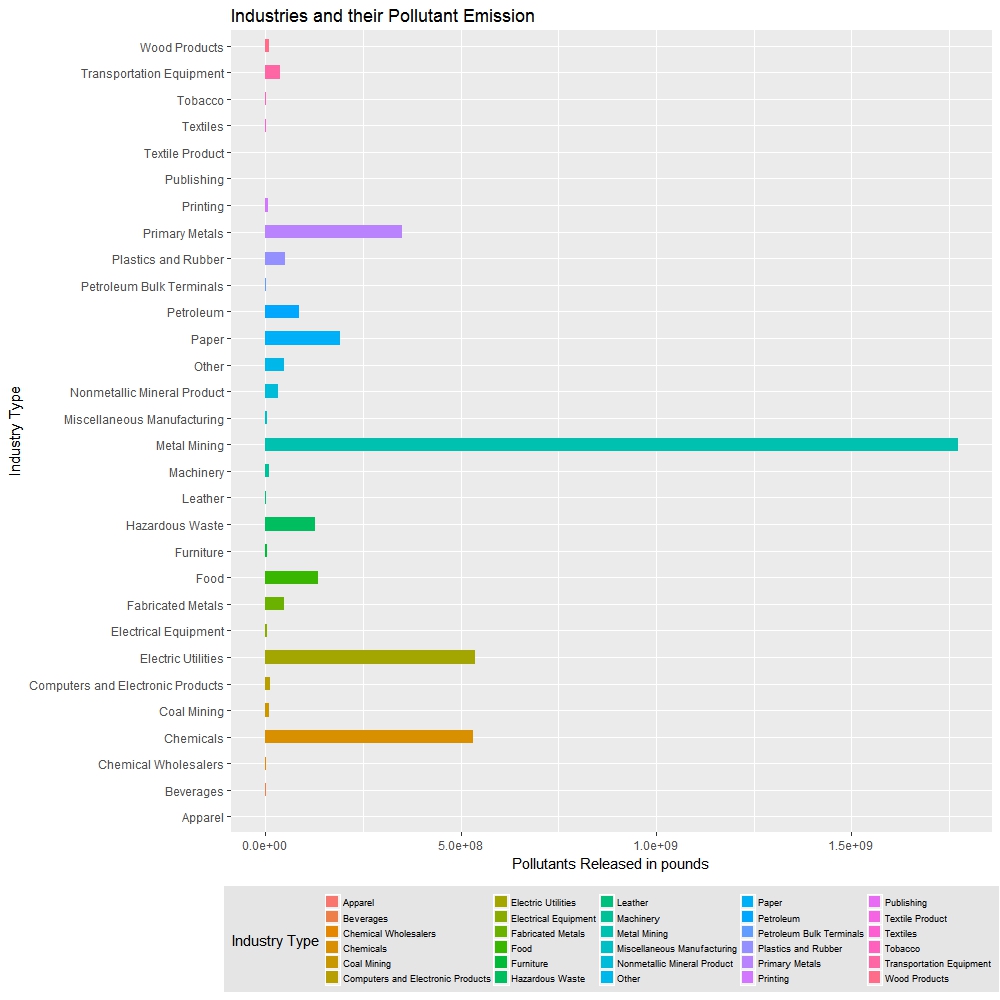


Figure 8

Looking at Figure 8 it can be deduced that ***metal mining industry*** (1.772263e+09 pounds of pollutants released in 2014) is the most polluted industry followed by ***electric utilities industry (***5.357731e+08 pounds of pollutants released in 2014)and ***chemicals industry (***5.308278e+08 pounds of wastes released in environment***).***

The next research question that we aim to answer is ***which chemical’s presence is substantially higher than others.*** **Figure 9** shows the code snapshot for finding the data and **Figure 10** shows the graph for the same.

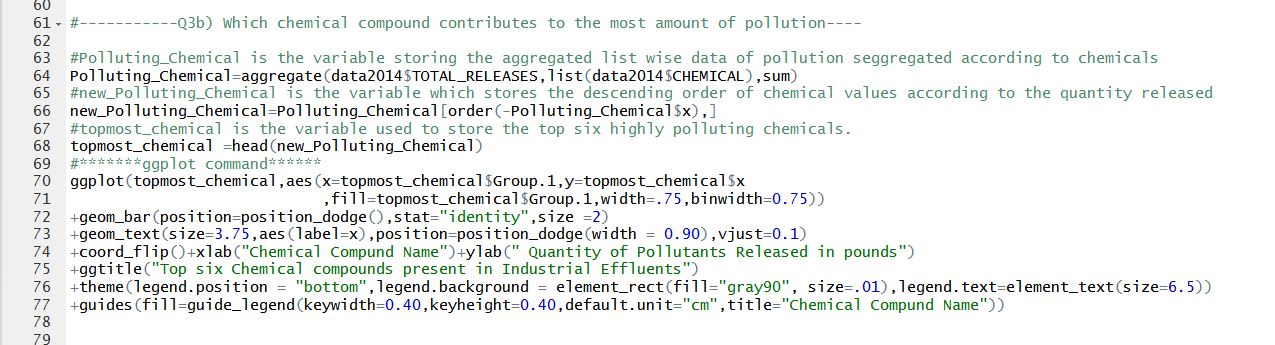


Figure 9

In this code we are only limiting our visualization for finding the top 6 most polluting chemicals present in industrial effluents.

Looking at Figure10 we realize that **Zinc Compound’s** presence is the highest (*980186311 pounds released in 2014)* followed by **Lead Compounds** *(730455131 pounds released in 2014).*

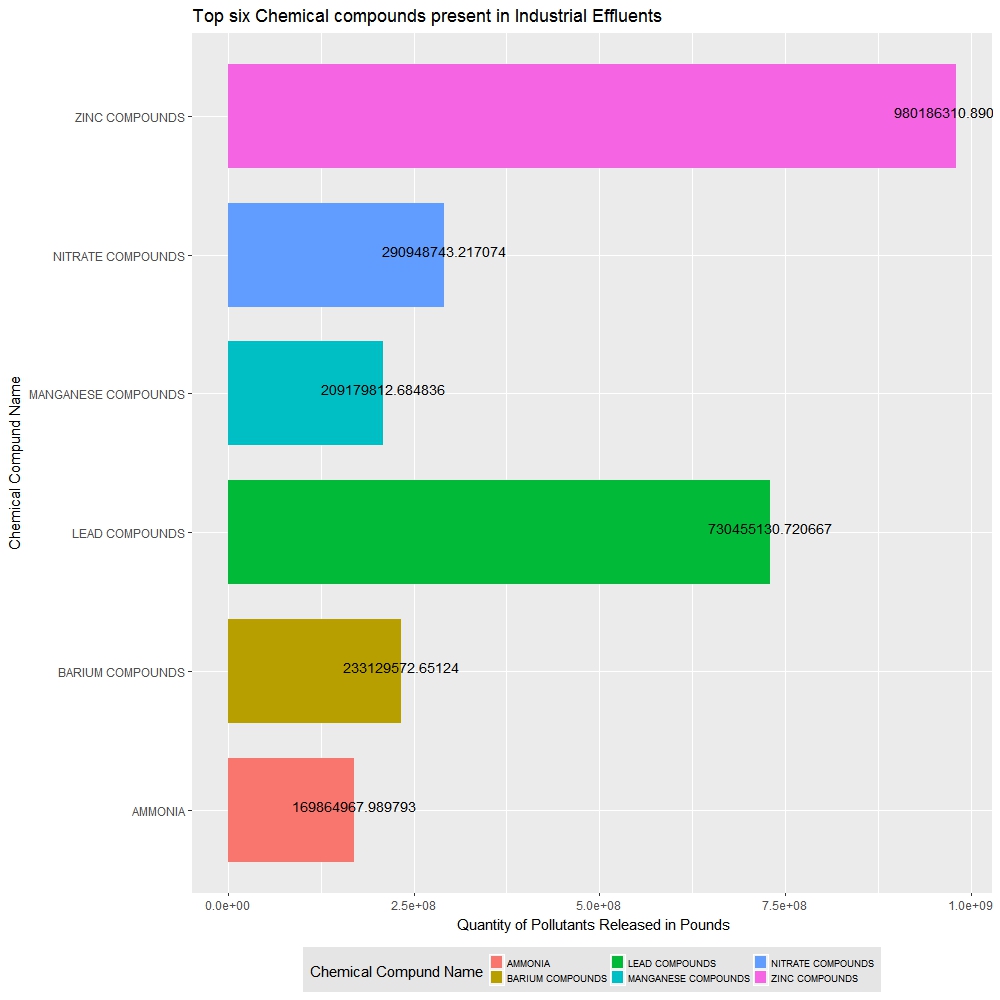


Figure 10

The following code snippets and graph tells us about ***which industry sector contributes to most carcinogenic compounds releases.* Figure11** shows the code and **Figure12** shows the visualization of industry wise carcinogenic releases..

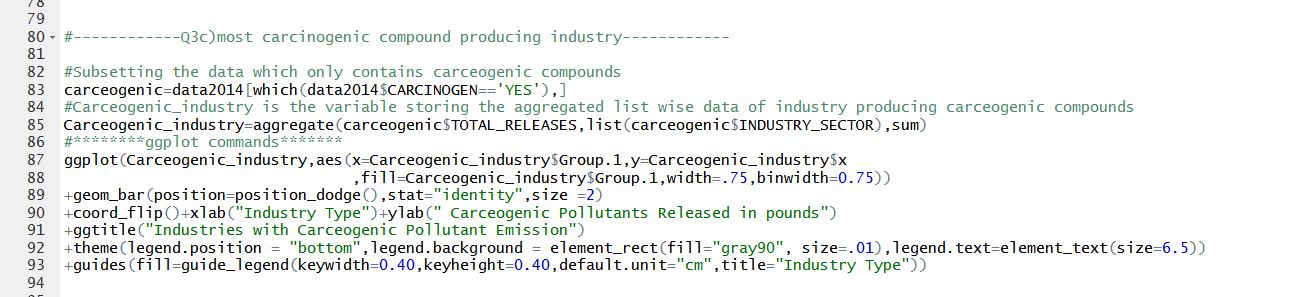


Figure 11

Looking at the graph (Figure12) we can deduce that **Chemical industry** is the *most carcinogenic* pollutant industry, producing 50914315.482pounds of carcinogenic wastes. The next most carcinogenic polluting industry is **Hazardous waste industry** producing 30071196.591 pounds of carcinogenic wastes.

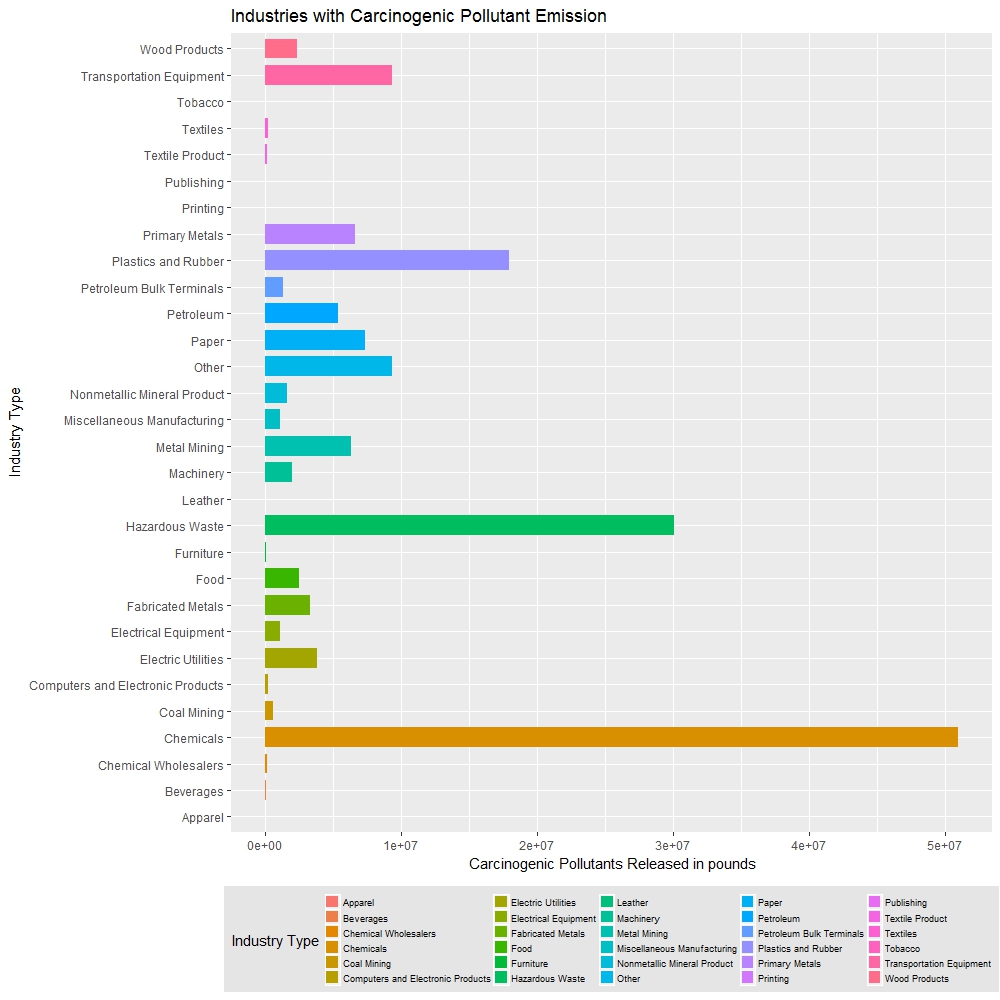


Figure 12

Next we aim to find ***which industry has the highest stack and fugitive emission***. Figure13 shows the code snippet for the visualization and Figure 14 shows the visualization.



Figure 13

**Figure14** shows the plot of industry wise **Stack emission**. **Electric Utilities** industries are highest in stack emission (181947411.63 pounds released in environment in 2014) followed by **Chemical industries** (111364441.93 pounds of pollutants released in environment in 2014) and then **Paper Industry** (98449632.72 pounds of pollutants released in environment in 2014).

**Figure15** shows the plot of industry wise **Fugitive emission**. **Chemical industries** are highest in stack emission (5.953391e+07 pounds released in environment in 2014) followed by **Paper Industries** (4.293918e+07 pounds of pollutants released in environment in 2014) and then **Food Industry** (1.560037e+07 pounds of pollutants released in environment in 2014).

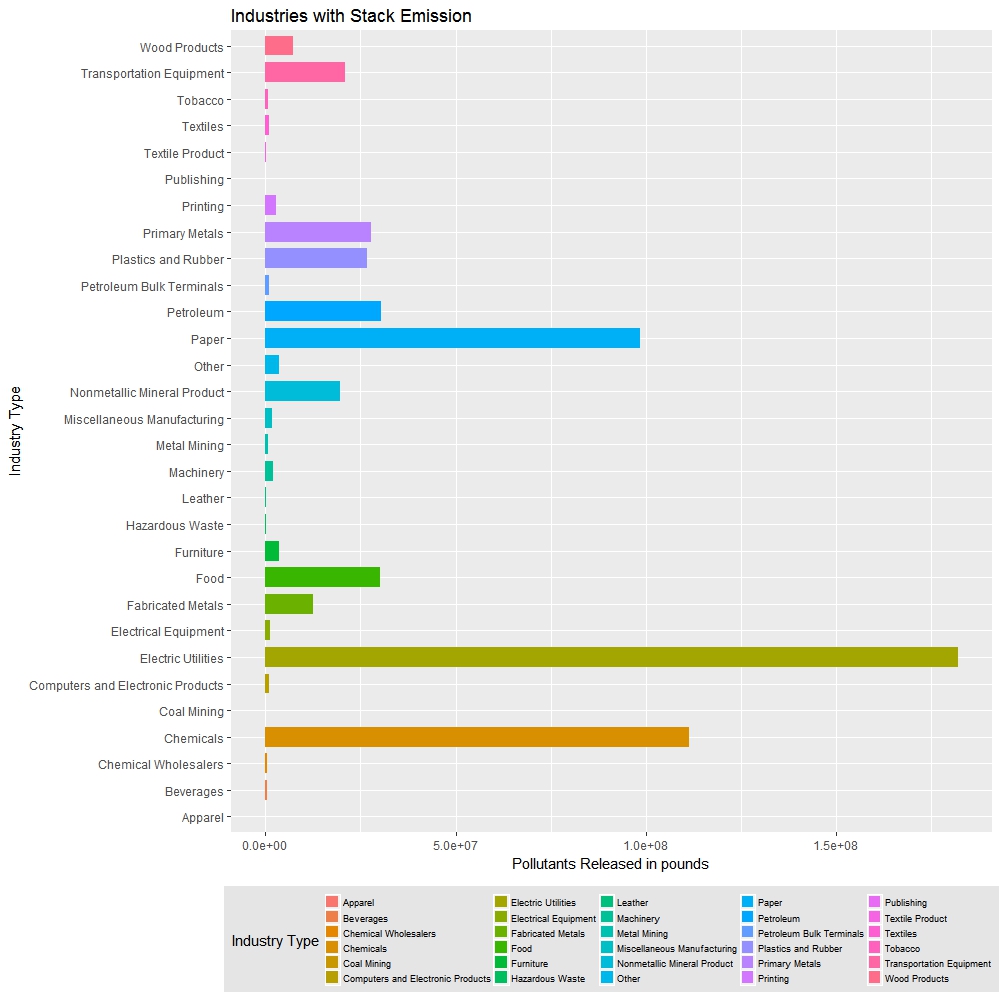


Figure 14

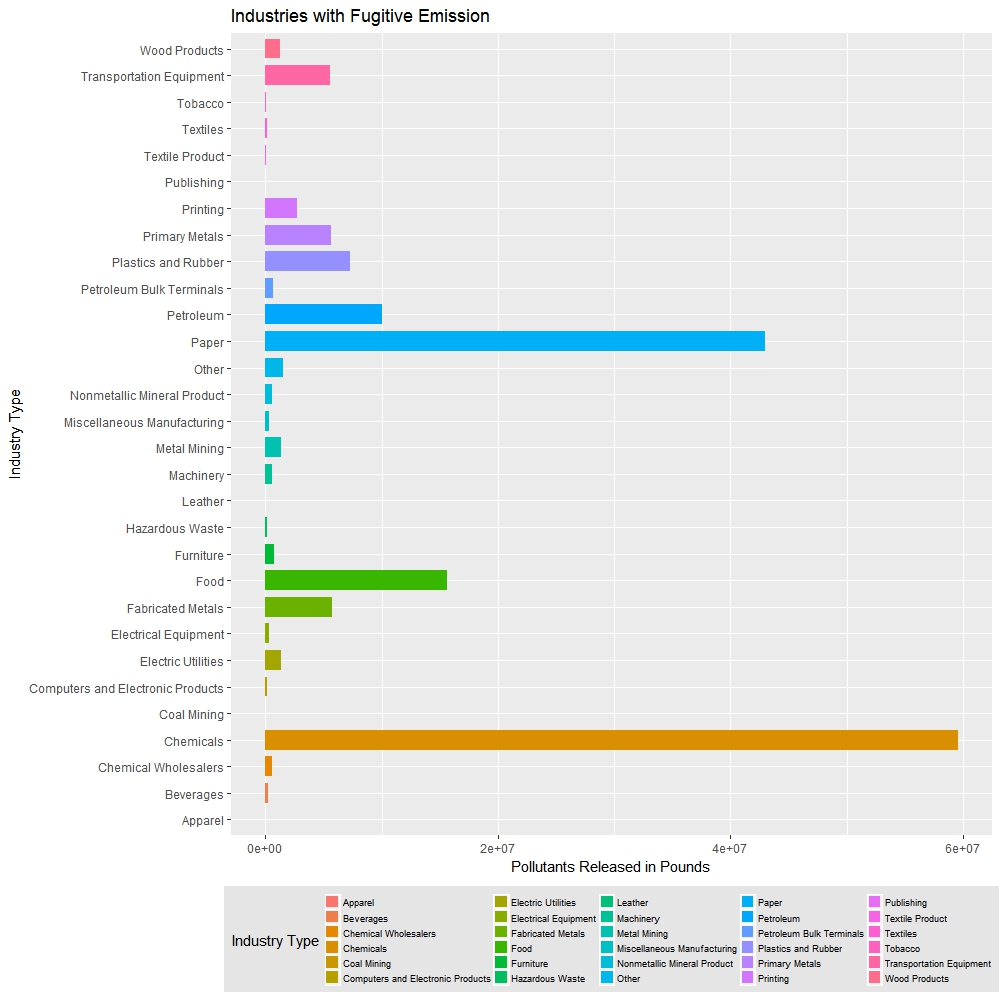
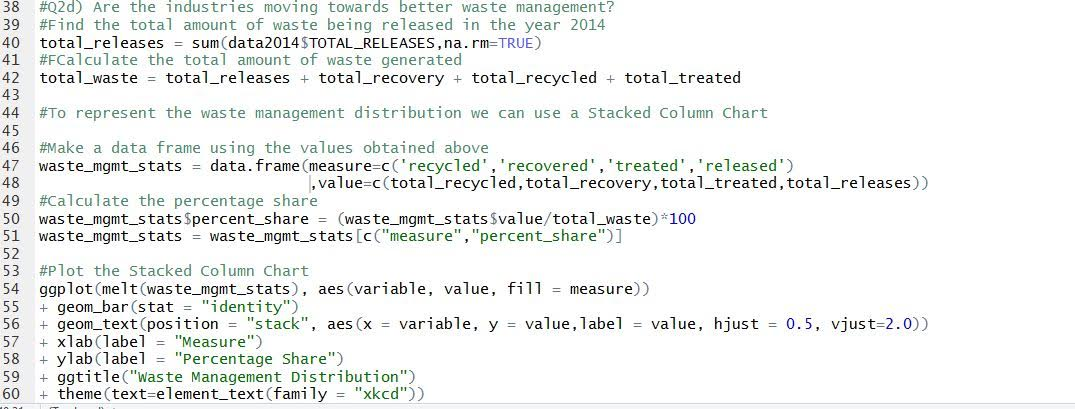
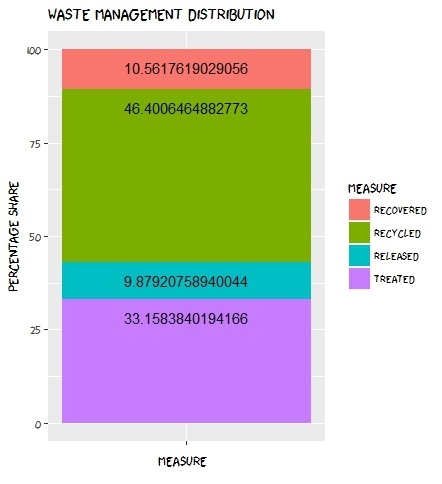


Figure 15

Question 4:   
In this question ***we measure the extent to which waste management is being implemented by all the industries***. We calculate how much of the total waste in pounds is being recycled, recovered, treated, and released.  
   


**Figure 16**



**Figure 17**

In **Figure 17**, we can see the percentage distribution of the waste management activities. From the graph se obtained, we understand that a majority (about 33%) of the waste is being treated before being released. And a small amount of waste (about 9.9%) is being released directly. One exciting conclusion we can arrive from this chart is that almost 46% of the waste is recycled and 10.5% is recovered for energy use. This shows that the industries are implementing good recycling strategies and are also reusing their waste for energy generation. Also, a large portion of the waste is being treated before being released. What the industries can aim to achieve soon is to minimize the waste being released un treated, rather than minimize, eliminate it. We need to identify he reasons why this waste goes untreated. Is it because of absence of a treatment plant in the vicinity, is it too costly to treat the waste, or is it difficult to treat such waste without generating other waste products in the process. The companies should also aim to recycle and recover more of the waste that is being treated.

Word Count: 958