**Draft R Plot**

**INTRODUCTION**

We have examined numerous companies in a variety of industries, and identified several ways each has individually disposed of waste in their respective states and counties. Honing in on data from the year 2014, we have also identified certain industries that were particularly egregious in their waste release, some emitted substantially higher amounts of chemicals than other industries listed. Along with the industries examined, we looked at the six most prominent chemicals in industrial effluents. Since carcinogenesis is a major component and unfortunate byproduct of industrial waste emissions, we decided that taking a closer look at certain industry sectors that contributed the most to carcinogenic compound releases would be judicious. Another important area of examination is how the waste from certain industries is readily emitted, whether through more traditional ways of stack emissions, or through unintentional or uncontrolled means known as fugitive emissions.

**RESEARCH QUESTIONS**

Following are the research questions which we aim to answer in this project as it helps in identifying areas of concern and will help environmental strategists to formulate better regulations.

1. Which state ranks the highest in pollution? Which county in a state contributes to highest pollution? We can scale up the data from county to state level and identify potential areas under threat from high levels of pollution?
2. How much waste is being recycled, how much waste is being converted to energy, how much waste is being treated?
3. Which industry type is causing the most amount of pollution (by volume/weight)? Which chemical compound contributes to the most amount of pollution?  Which industry is producing the most carcinogenic compound?
4. Which companies are more prone to stack (traditional means) emissions versus fugitive (escaped/leaked) emissions?

**ANALYSIS OF RESEARCH QUESTIONS**

The first question is related to finding out the most polluting state in US. Since in the given dataset all the data present is in county level; in order to find the most polluted State of US according to *total releases* we aggregate all the counties of a same state and add up their releases .Thus we get total emission data for all the 50 states and 6 union territories of USA .**Figure1** shows the code snippet for computing most polluted state in US. A detailed state wise pollution emission for the year 2014 can be found in **Appendix A.**

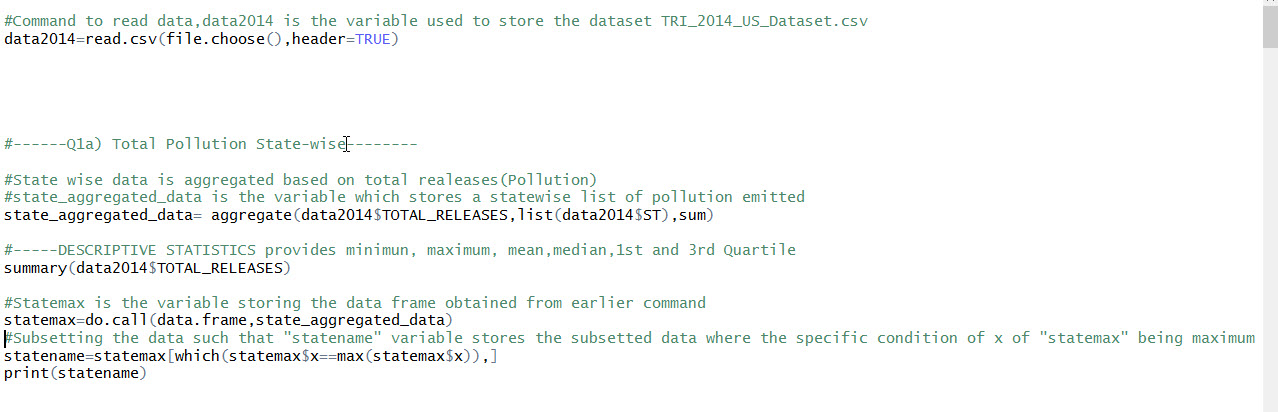


Figure 1

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-2** shows a snapshot of the code used to create this visualization in R using **xkcd package**

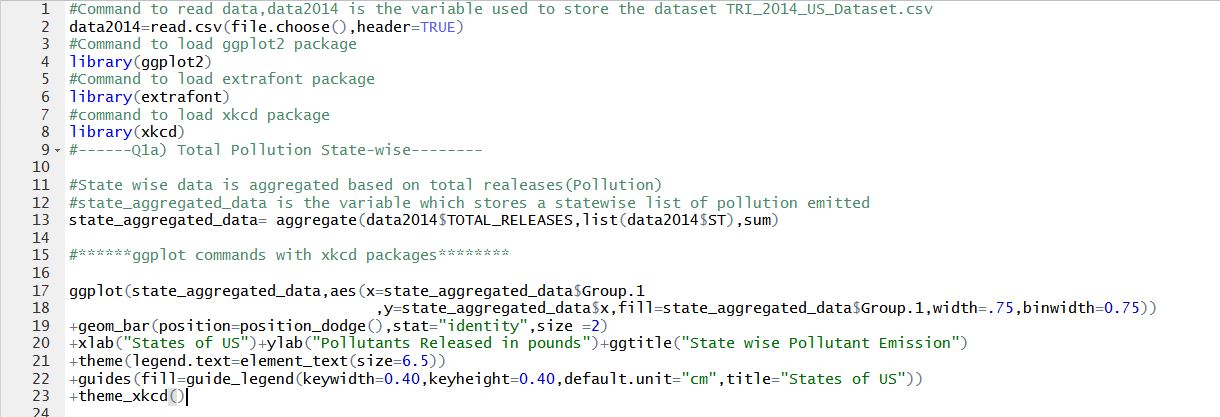


Figure 2

**Figure3** represents the visualization for the same

.



Figure 3

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 4** represents snapshot of the code used to generate the same graph without xkcd package. And **Figure 5** is the visualization for the same.



Figure 4



Figure 5

Next we find out all state wise counties in us with highest toxic releases. Figure 6 shows the counties in each state have highest toxic releases.

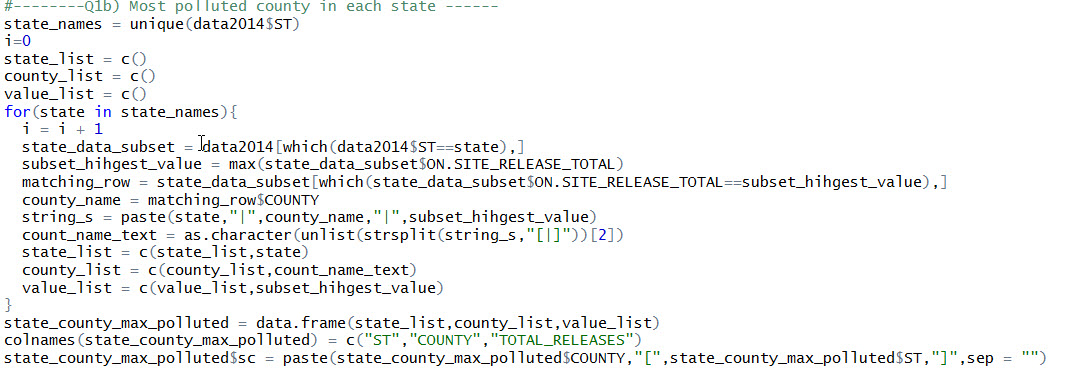


Figure 6

The following figure 7 show the visualization for the same

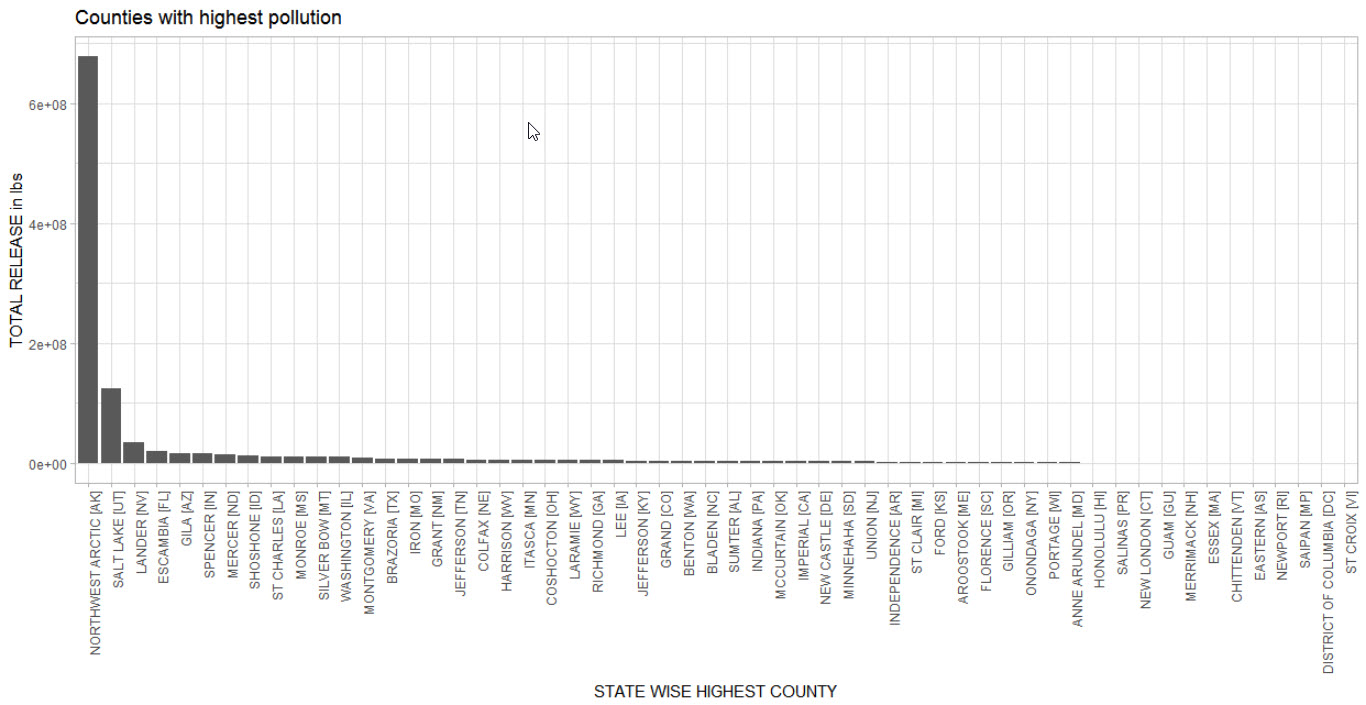


Figure 7

A detailed state wise highest polluting counties table can be found APPENDIX B

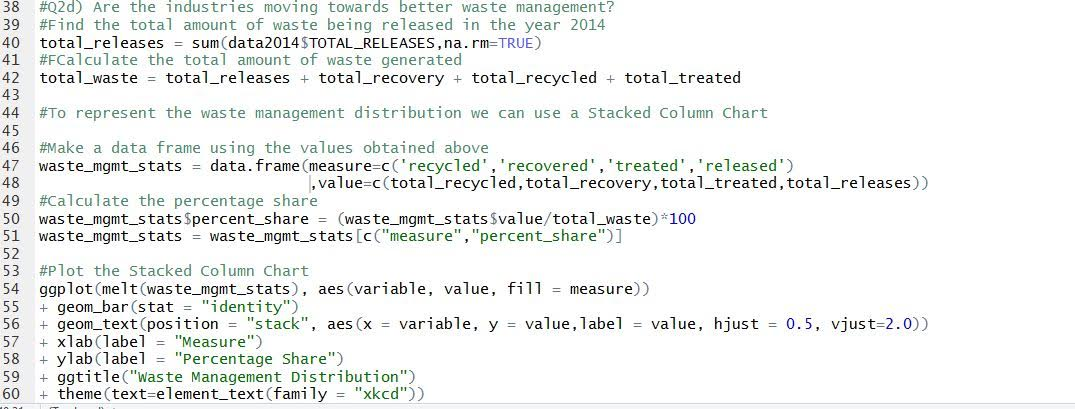
Our next research question aims at  ***measuring the extent to which waste management is being implemented by all the industries***. We calculate how much of the total waste in pounds is being recycled, recovered, treated, and released.  
   


Figure 8

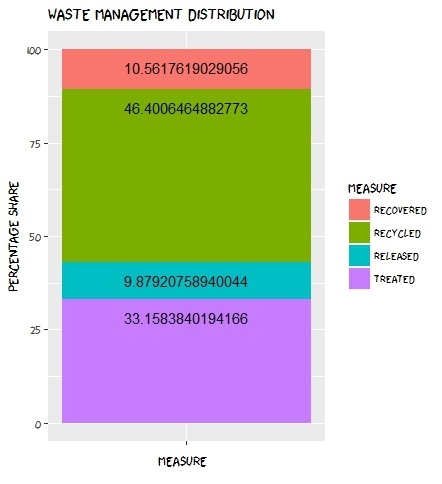


Figure 9

In **Figure 9**, 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.

Next we find ***which industries are most polluting*** according to the dataset and we aggregate the total release according to the industry type and compare in order to find the most polluting industry. **Figure 10** shows the code snipped in R used to find most polluting industry. **Appendix C** gives the complete list of industries ranked according to Toxic releases.

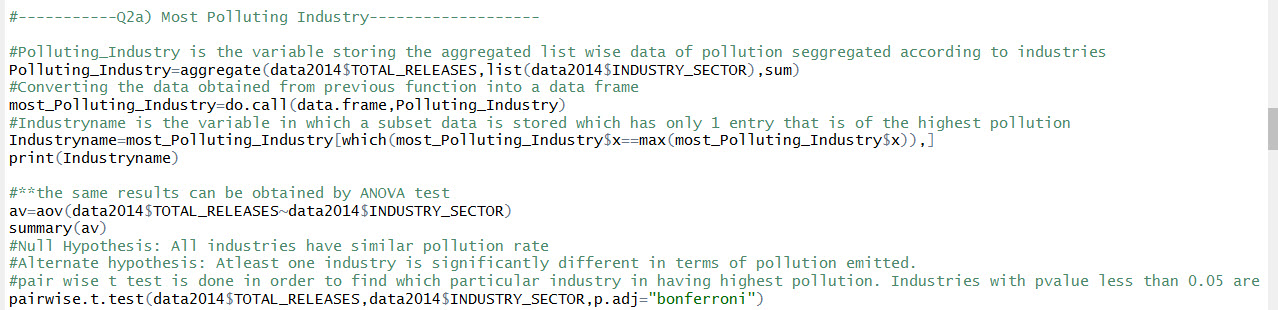


Figure 10

The method of aggregation of data is one way to find most polluting industry. The other method is to use ANOVA test wherein we assume that all industries equally pollute environment as null hypothesis. The alternate hypothesis is that at least one of the industry’s pollutant emissions is different than others.

Df Sum Sq Mean Sq F value Pr(>F)

data2014$INDUSTRY\_SECTOR 29 4.589e+15 1.582e+14 19.67 <2e-16 \*\*\*

Residuals 82251 6.617e+17 8.044e+12

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

As seen the p value is less than 0.05 hence we reject the null hypothesis. Thus there is a variation in pollution emission according to type of industry. To find which industry is having the most different pollution emission rate, we do Bonferroni test.

**Figure11** represents the code snap in R for the visualization and **Figure 12** represents the plot for the same.

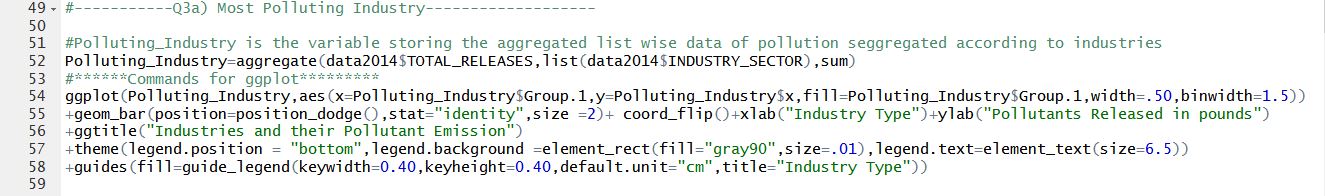


Figure 11

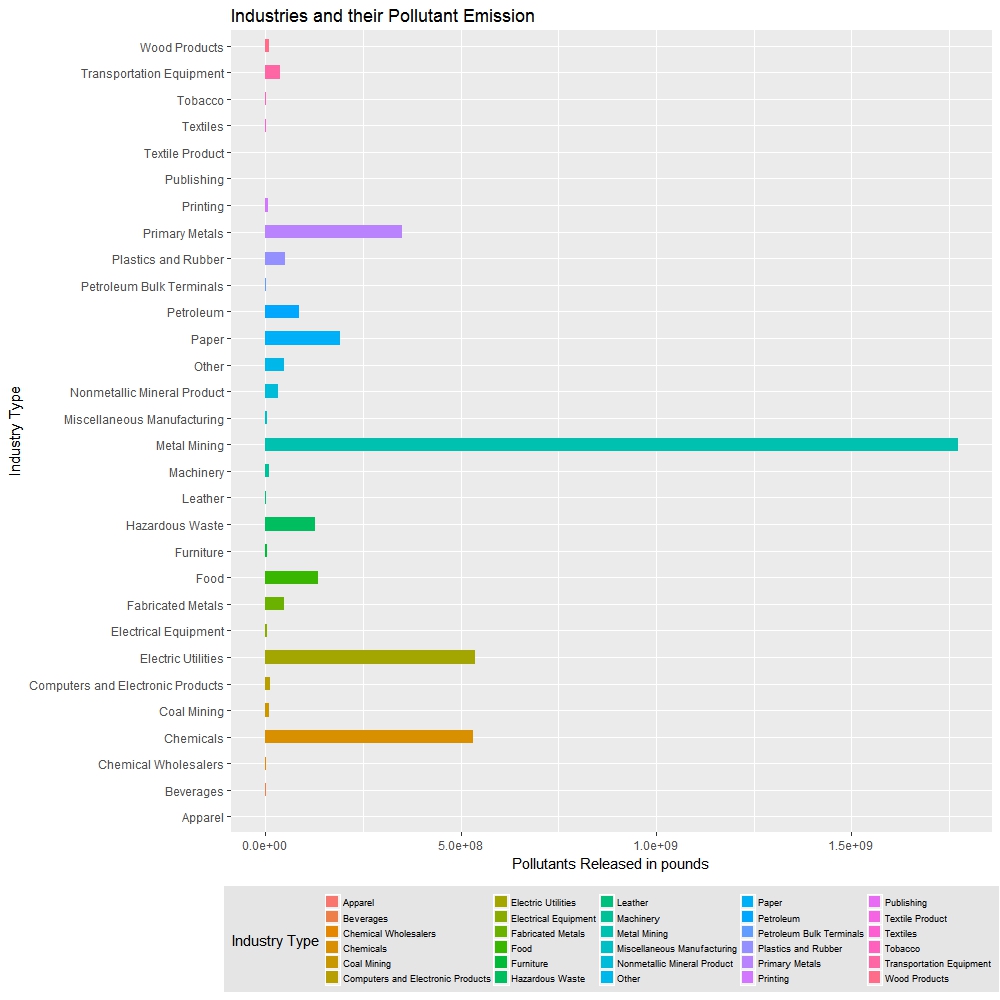


Figure 12

Looking at **Figure 12** 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 aims at answering which ***chemical’s presence is substantially higher than others.*** To answer this we aggregate onsite total release according to chemical composition and compare them with one another in order to get the chemicals which are highly present in industrial effluents. **Figure 13** shows code snippet to obtain highest chemical in effluents. Appendix D gives detailed quantity wise release of chemical compounds.

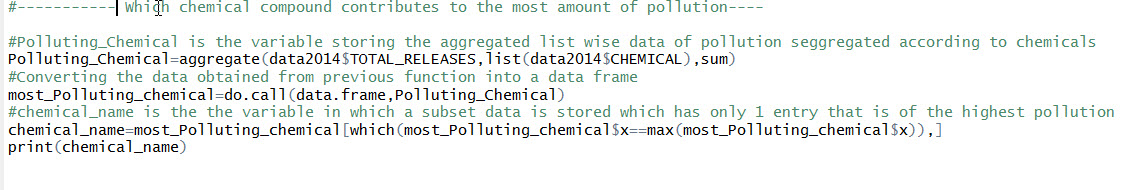


Figure 13

**Figure 14** shows the code snapshot for visualization and **Figure 15** shows the graph for the same.

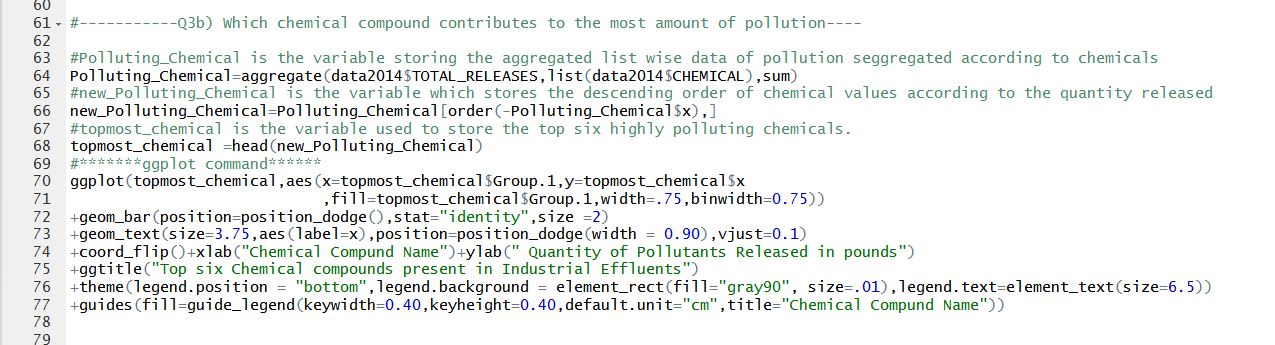


Figure 14

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

Looking at Figure15 we realize that **Zinc Compound’s** presence is the highest (*980186311 pounds released in 2014)* followed by **Lead Compounds** *(730455131 pounds released in 2014).*

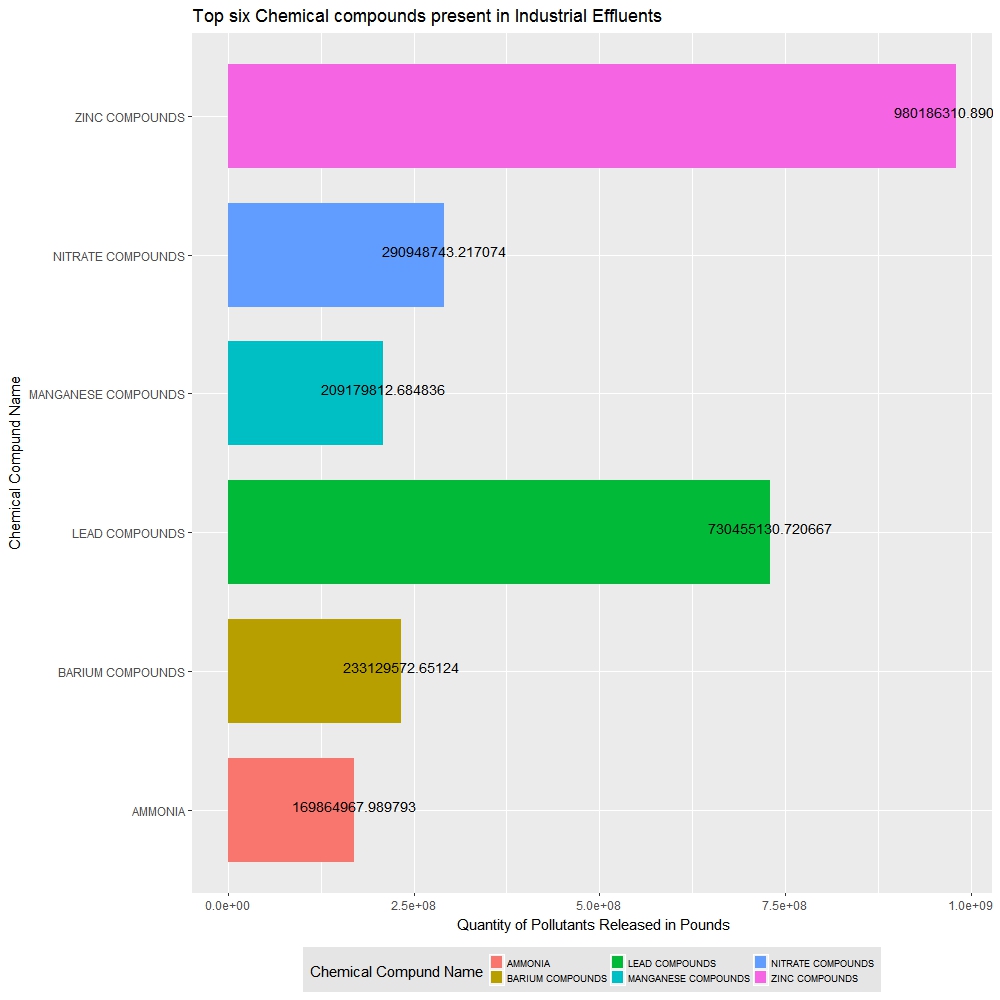


Figure 15

The following code snippet **Figure 16** tells us about ***which industry sector contributes to most carcinogenic compounds releases.*** Appendix E shows us most carcinogenic industries.

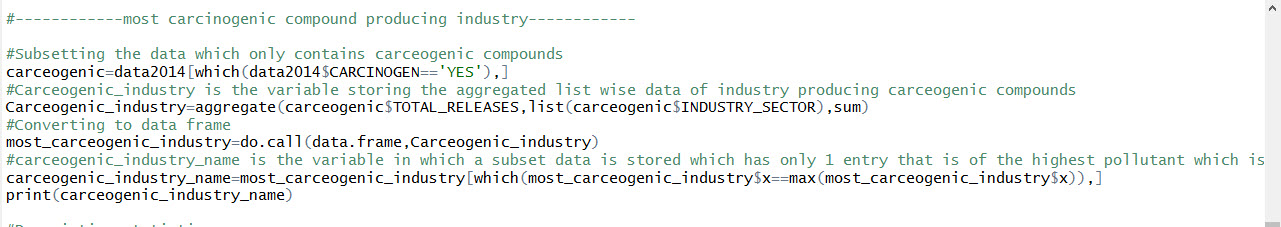
******

Figure 16

In this code we have first created a subset of data from the given data set only limiting data to cacogenic releases and then aggregated all the industrial effluents based on industries.

**Figure17** shows the code for visualization and **Figure18** shows the visualization of industry wise carcinogenic releases

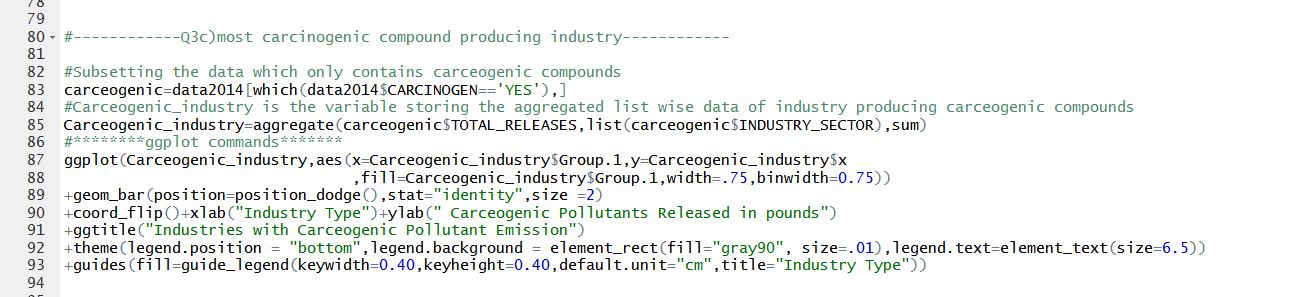


Figure 17

Looking at the graph (Figure18) 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.

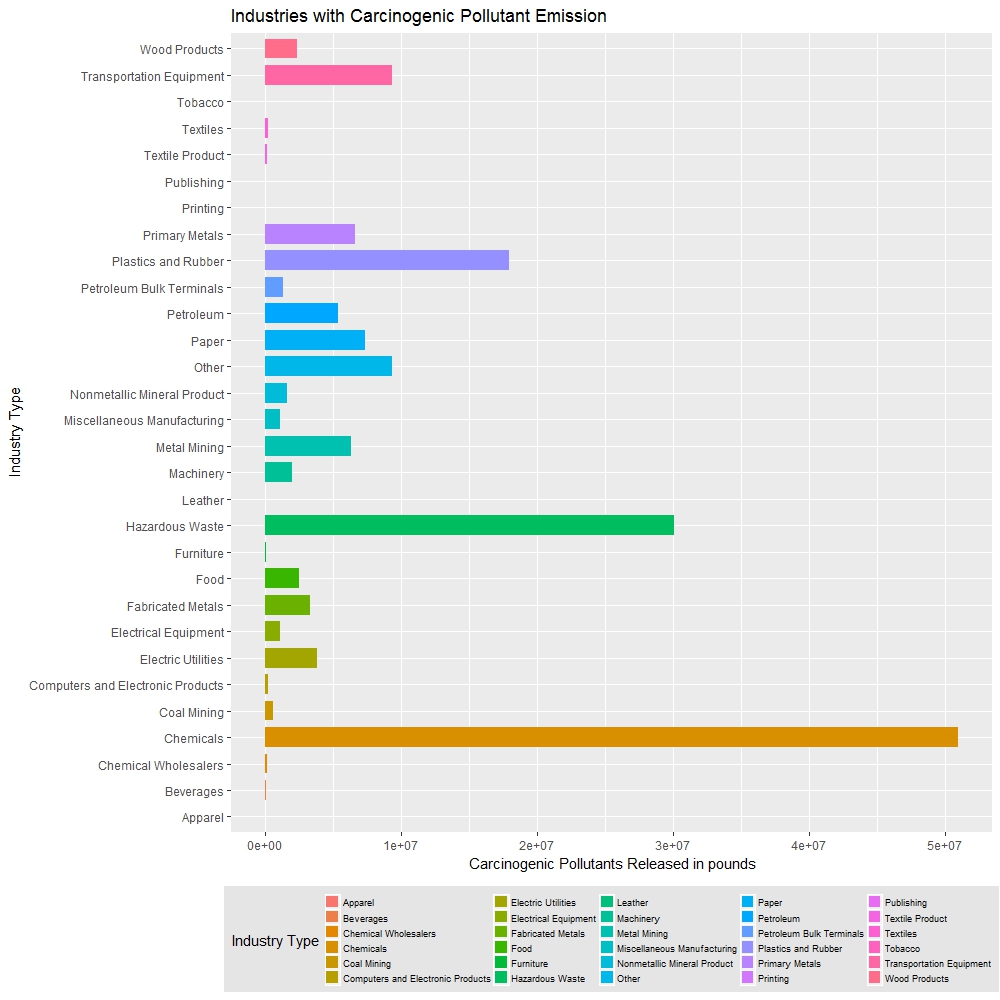


Figure 18

Next we aim to find ***which industry has the highest stack and fugitive emission***. Figure 19 shows the code snippet industries with highest stack emission .this obtained by sub setting the data and aggregating the total releases based on industries. Another Statistical test ;ANOVA test is used to validate our result. Appendix F shows all industries ranked according to fugitive and stack emission.

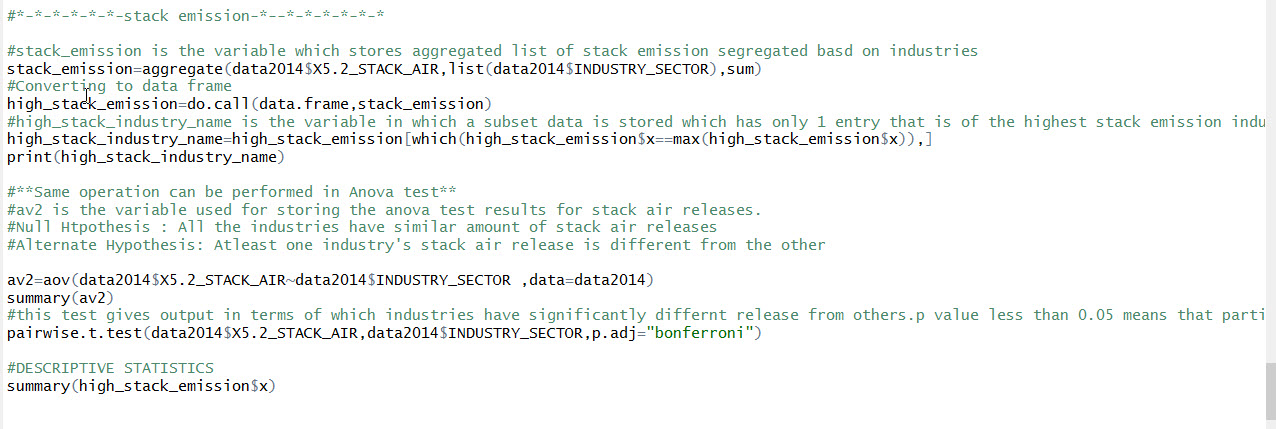


Figure 19

The next figure shows the code in R for industries with highest fugitive emission. **Figure 20** is the code snippet in R comprising of ANOVA tests and subset and aggregation of data. Also pairwise t test using Bonferroni is used for finding out industries with highest stack and fugitive emission.

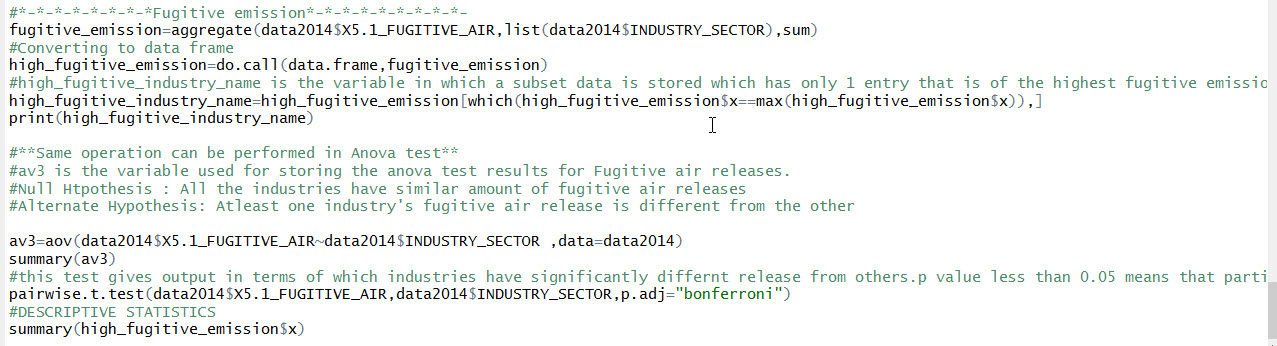


Figure 20

**Figure21** shows the code snippet for the visualization and **Figure 22** shows the visualization.

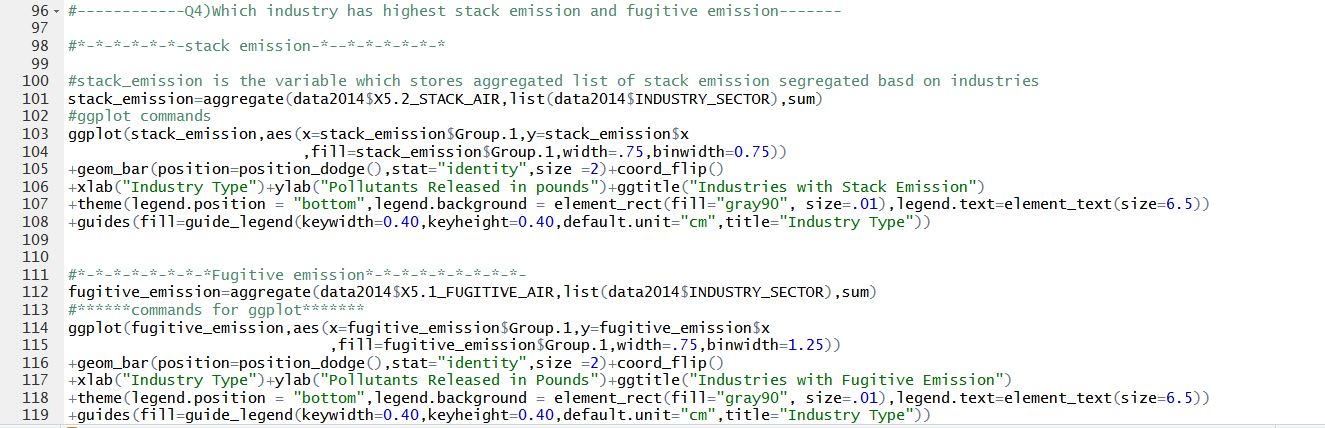


Figure 21

**Figure 22** 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).

**Figure 23** 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).

[Words-1579]

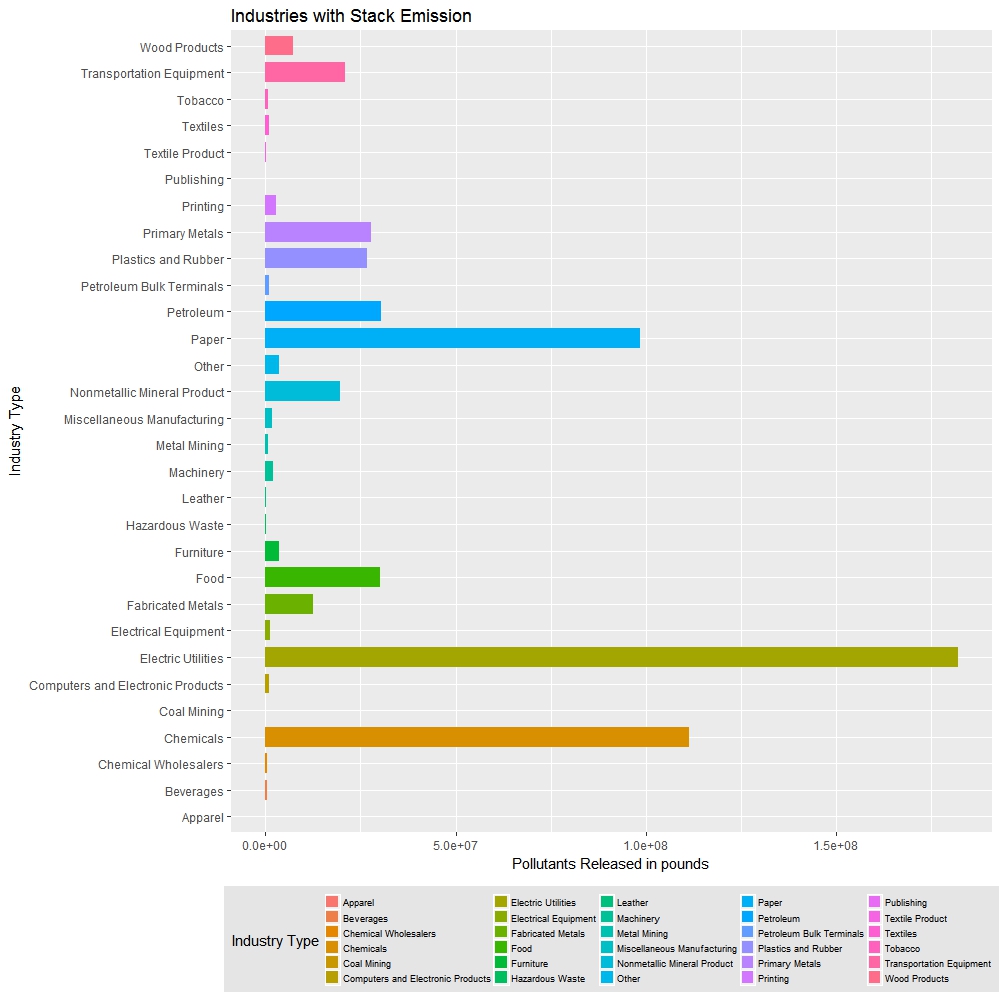


Figure 22

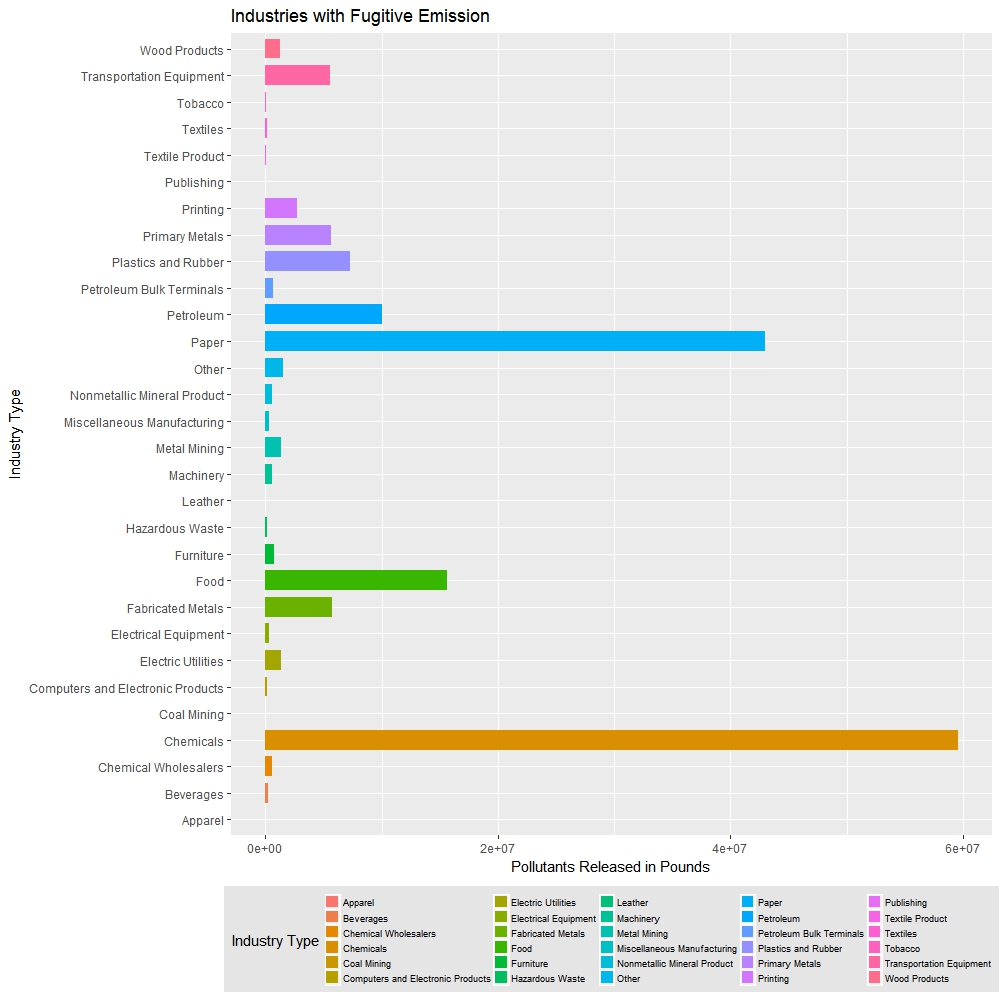


Figure 23

**APPENDIX A**

State wise Pollution emission

|  |  |  |
| --- | --- | --- |
| State wise Rank | State Name | Pollutants released in Pounds |
| 1 | AK | 1.16E+09 |
| **2** | NV | 2.87E+08 |
| **3** | TX | 2.59E+08 |
| **4** | UT | 2.10E+08 |
| **5** | IN | 1.60E+08 |
| **6** | LA | 1.43E+08 |
| **7** | OH | 1.23E+08 |
| **8** | IL | 1.21E+08 |
| **9** | AL | 9.19E+07 |
| **10** | PA | 8.83E+07 |
| **11** | TN | 8.72E+07 |
| **12** | AZ | 7.94E+07 |
| **13** | KY | 7.44E+07 |
| **14** | MO | 7.06E+07 |
| **15** | MS | 7.01E+07 |
| **16** | GA | 6.93E+07 |
| **17** | FL | 6.50E+07 |
| **18** | MI | 6.38E+07 |
| **19** | NC | 6.26E+07 |
| **20** | ID | 5.08E+07 |
| **21** | ND | 5.03E+07 |
| **22** | SC | 4.71E+07 |
| **23** | VA | 4.56E+07 |
| **24** | AR | 3.87E+07 |
| **25** | WI | 3.83E+07 |
| **26** | MT | 3.78E+07 |
| **27** | WV | 3.68E+07 |
| **28** | IA | 3.61E+07 |
| **29** | CA | 3.47E+07 |
| **30** | MN | 3.10E+07 |
| **31** | CO | 2.98E+07 |
| **32** | NE | 2.75E+07 |
| **33** | OK | 2.72E+07 |
| **34** | OR | 2.47E+07 |
| **35** | NM | 2.37E+07 |
| **36** | WA | 2.22E+07 |
| **37** | KS | 2.20E+07 |
| **38** | NY | 1.85E+07 |
| **39** | WY | 1.84E+07 |
| **40** | NJ | 1.37E+07 |
| **41** | MA | 1.24E+07 |
| **42** | ME | 1.04E+07 |
| **43** | MD | 8.27E+06 |
| **44** | DE | 6.24E+06 |
| **45** | SD | 6.21E+06 |
| **46** | PR | 2.80E+06 |
| **47** | HI | 2.67E+06 |
| **48** | CT | 2.29E+06 |
| **49** | NH | 7.75E+05 |
| **50** | DC | 5.88E+05 |
| **51** | RI | 5.52E+05 |
| **52** | GU | 4.92E+05 |
| **53** | VT | 3.09E+05 |
| **54** | AS | 1.00E+05 |
| **55** | VI | 3.48E+04 |
| **56** | MP | 1.56E+04 |
|  |  |  |

APPENDIX B

State wise counties with highest Toxic release

|  |  |  |  |
| --- | --- | --- | --- |
| **State** | **County** | **Pollutants released** |  |
|  |  |  |  |
| AK | NORTHWEST ARCTIC | 678041040 | NORTHWEST ARCTIC [AK] |
| UT | SALT LAKE | 124989494 | SALT LAKE [UT] |
| NV | LANDER | 34172792 | LANDER [NV] |
| FL | ESCAMBIA | 19432001 | ESCAMBIA [FL] |
| AZ | GILA | 16190575 | GILA [AZ] |
| IN | SPENCER | 15957848 | SPENCER [IN] |
| ND | MERCER | 14106000 | MERCER [ND] |
| ID | SHOSHONE | 11832188 | SHOSHONE [ID] |
| LA | ST CHARLES | 11018200 | ST CHARLES [LA] |
| MS | MONROE | 10862892 | MONROE [MS] |
| MT | SILVER BOW | 10110898 | SILVER BOW [MT] |
| IL | WASHINGTON | 10106087 | WASHINGTON [IL] |
| VA | MONTGOMERY | 8484928 | MONTGOMERY [VA] |
| TX | BRAZORIA | 7573795 | BRAZORIA [TX] |
| MO | IRON | 6752997 | IRON [MO] |
| NM | GRANT | 6430692 | GRANT [NM] |
| TN | JEFFERSON | 6113606.81 | JEFFERSON [TN] |
| NE | COLFAX | 5110320 | COLFAX [NE] |
| WV | HARRISON | 5100000 | HARRISON [WV] |
| MN | ITASCA | 5047375 | ITASCA [MN] |
| OH | COSHOCTON | 4800000 | COSHOCTON [OH] |
| WY | LARAMIE | 4759589 | LARAMIE [WY] |
| GA | RICHMOND | 4737933 | RICHMOND [GA] |
| IA | LEE | 4370250 | LEE [IA] |
| KY | JEFFERSON | 3959100 | JEFFERSON [KY] |
| CO | GRAND | 3750183.9 | GRAND [CO] |
| WA | BENTON | 3698998.5 | BENTON [WA] |
| NC | BLADEN | 3668587 | BLADEN [NC] |
| AL | SUMTER | 3648010 | SUMTER [AL] |
| PA | INDIANA | 3488468 | INDIANA [PA] |
| OK | MCCURTAIN | 2874005 | MCCURTAIN [OK] |
| CA | IMPERIAL | 2781116.3 | IMPERIAL [CA] |
| DE | NEW CASTLE | 2742685 | NEW CASTLE [DE] |
| SD | MINNEHAHA | 2718000 | MINNEHAHA [SD] |
| NJ | UNION | 2543100 | UNION [NJ] |
| AR | INDEPENDENCE | 2410600 | INDEPENDENCE [AR] |
| MI | ST CLAIR | 2303660 | ST CLAIR [MI] |
| KS | FORD | 2220105 | FORD [KS] |
| ME | AROOSTOOK | 2024822 | AROOSTOOK [ME] |
| SC | FLORENCE | 1833790.97 | FLORENCE [SC] |
| OR | GILLIAM | 1804369.4 | GILLIAM [OR] |
| NY | ONONDAGA | 1636722 | ONONDAGA [NY] |
| WI | PORTAGE | 1258820 | PORTAGE [WI] |
| MD | ANNE ARUNDEL | 716200.2 | ANNE ARUNDEL [MD] |
| HI | HONOLULU | 460000 | HONOLULU [HI] |
| PR | SALINAS | 291416 | SALINAS [PR] |
| CT | NEW LONDON | 275213 | NEW LONDON [CT] |
| GU | GUAM | 246168.95 | GUAM [GU] |
| NH | MERRIMACK | 245000 | MERRIMACK [NH] |
| MA | ESSEX | 187583 | ESSEX [MA] |
| VT | CHITTENDEN | 117743 | CHITTENDEN [VT] |
| AS | EASTERN | 98314 | EASTERN [AS] |
| RI | NEWPORT | 46963 | NEWPORT [RI] |
| MP | SAIPAN | 11253 | SAIPAN [MP] |
| DC | DISTRICT OF COLUMBIA | 10590 | DISTRICT OF COLUMBIA [DC] |
| VI | ST CROIX | 10459.77 | ST CROIX [VI] |

**APPENDIX C**

List of Industries ranked according to Pollution.

|  |  |  |
| --- | --- | --- |
| **State wise Rank** | **Industry Name** | **Pollutants released in Pounds** |
| **1** | Metal Mining | 1.77E+09 |
| **2** | Electric Utilities | 5.36E+08 |
| **3** | Chemicals | 5.31E+08 |
| **4** | Primary Metals | 3.51E+08 |
| **5** | Paper | 1.90E+08 |
| **6** | Food | 1.36E+08 |
| **7** | Hazardous Waste | 1.28E+08 |
| **8** | Petroleum | 8.74E+07 |
| **9** | Plastics and Rubber | 5.09E+07 |
| **10** | Fabricated Metals | 4.88E+07 |
| **11** | Other | 4.77E+07 |
| **12** | Transportation Equipment | 3.78E+07 |
| **13** | Nonmetallic Mineral Product | 3.12E+07 |
| **14** | Computers and Electronic Products | 1.14E+07 |
| **15** | Wood Products | 9.43E+06 |
| **16** | Coal Mining | 8.57E+06 |
| **17** | Machinery | 8.30E+06 |
| **18** | Printing | 5.97E+06 |
| **19** | Electrical Equipment | 4.96E+06 |
| **20** | Furniture | 4.57E+06 |
| **21** | Miscellaneous Manufacturing | 3.63E+06 |
| **22** | Textiles | 3.03E+06 |
| **23** | Petroleum Bulk Terminals | 2.90E+06 |
| **24** | Beverages | 2.76E+06 |
| **25** | Leather | 2.18E+06 |
| **26** | Tobacco | 2.07E+06 |
| **27** | Chemical Wholesalers | 1.39E+06 |
| **28** | Textile Product | 3.82E+05 |
| **29** | Apparel | 2.61E+04 |
| **30** | Publishing | 6.00E+03 |

**APPENDIX D**

Top 30 Chemicals released

|  |  |  |
| --- | --- | --- |
| **Rank** | **Chemical Compound** | **Released in Pounds** |
| **1** | ZINC COMPOUNDS | 980186310.9 |
| **2** | LEAD COMPOUNDS | 730455130.7 |
| **3** | NITRATE COMPOUNDS | 290948743.2 |
| **4** | BARIUM COMPOUNDS | 233129572.7 |
| **5** | MANGANESE COMPOUNDS | 209179812.7 |
| **6** | AMMONIA | 169864968 |
| **7** | METHANOL | 156285898 |
| **8** | ARSENIC COMPOUNDS | 139141739.8 |
| **9** | COPPER COMPOUNDS | 134104562.7 |
| **10** | HYDROCHLORIC ACID | 119068612.6 |
| **11** | SULFURIC ACID | 101264184 |
| **12** | CHROMIUM COMPOUNDS(EXCEPT CHROMITE ORE MINED IN THE TRANSVAAL REGION) | 48714251.88 |
| **13** | VANADIUM COMPOUNDS | 39892796.25 |
| **14** | N-HEXANE | 34124668.91 |
| **15** | NICKEL COMPOUNDS | 29023214.38 |
| **16** | STYRENE | 28766782.93 |
| **17** | HYDROGEN FLUORIDE | 27312313.18 |
| **18** | HYDROGEN SULFIDE | 25877924.29 |
| **19** | TOLUENE | 24941829.54 |
| **20** | FORMALDEHYDE | 19903762.8 |
| **21** | ALUMINUM OXIDE (FIBROUS FORMS) | 19531667.13 |
| **22** | NITRIC ACID | 17639408.82 |
| **23** | ASBESTOS (FRIABLE) | 17533178.47 |
| **24** | COPPER | 17142674.43 |
| **25** | CERTAIN GLYCOL ETHERS | 17032813.69 |
| **26** | XYLENE (MIXED ISOMERS) | 16402055.63 |
| **27** | ETHYLENE | 16074619.43 |
| **28** | FORMIC ACID | 15780919.38 |
| **29** | LEAD | 14695302.96 |
| **30** | CARBONYL SULFIDE | 14196086.6 |

APPENDIX E

Most Carcinogenic Industries

|  |  |  |
| --- | --- | --- |
| **Rank** | **Industry Name** | **Pollutants in Pounds** |
| **1** | Chemicals | 50914315.48 |
| **2** | Hazardous Waste | 30071196.59 |
| **3** | Plastics and Rubber | 17923020.36 |
| **4** | Other | 9323227.461 |
| **5** | Transportation Equipment | 9290782.066 |
| **6** | Paper | 7362645.565 |
| **7** | Primary Metals | 6617182.316 |
| **8** | Metal Mining | 6311325.16 |
| **9** | Petroleum | 5372693.726 |
| **10** | Electric Utilities | 3771930.275 |
| **11** | Fabricated Metals | 3286876.648 |
| **12** | Food | 2508156.885 |
| **13** | Wood Products | 2359646.227 |
| **14** | Machinery | 1959162.819 |
| **15** | Nonmetallic Mineral Product | 1605390.836 |
| **16** | Petroleum Bulk Terminals | 1332650.352 |
| **17** | Miscellaneous Manufacturing | 1061496.733 |
| **18** | Electrical Equipment | 1049116.231 |
| **19** | Coal Mining | 564948.447 |
| **20** | Textiles | 204030.218 |
| **21** | Computers and Electronic Products | 167627.799 |
| **22** | Chemical Wholesalers | 92385.113 |
| **23** | Textile Product | 91136.851 |
| **24** | Furniture | 71169.177 |
| **25** | Beverages | 44418.04 |
| **26** | Apparel | 8453 |
| **27** | Leather | 3250 |
| **28** | Printing | 3122.629 |
| **29** | Tobacco | 1217 |
| **30** | Publishing | 0.14 |

APPENDIX F

Stack Emission

|  |  |  |
| --- | --- | --- |
| **Rank** | **Industry** | **Pollutants released in pounds** |
| **1** | Electric Utilities | 181947411.6 |
| **2** | Chemicals | 111364441.9 |
| **3** | Paper | 98449632.72 |
| **4** | Petroleum | 30365795.31 |
| **5** | Food | 30136484.88 |
| **6** | Primary Metals | 27677382.05 |
| **7** | Plastics and Rubber | 26746566.29 |
| **8** | Transportation Equipment | 20875993.2 |
| **9** | Nonmetallic Mineral Product | 19507846.88 |
| **10** | Fabricated Metals | 12579662.74 |
| **11** | Wood Products | 7380874.69 |
| **12** | Furniture | 3730606.92 |
| **13** | Other | 3501279.15 |
| **14** | Printing | 2876975.99 |
| **15** | Machinery | 2031109.72 |
| **16** | Miscellaneous Manufacturing | 1724885.6 |
| **17** | Electrical Equipment | 1131474.97 |
| **18** | Computers and Electronic Products | 1057926.69 |
| **19** | Textiles | 873834.18 |
| **20** | Petroleum Bulk Terminals | 841882.48 |
| **21** | Metal Mining | 740285.33 |
| **22** | Tobacco | 695726.99 |
| **23** | Beverages | 467587.89 |
| **24** | Chemical Wholesalers | 464511.06 |
| **25** | Hazardous Waste | 233431.69 |
| **26** | Leather | 201422 |
| **27** | Textile Product | 113804.2 |
| **28** | Apparel | 13050 |
| **29** | Coal Mining | 11734.29 |
| **30** | Publishing | 260 |

Fugitive Emission

|  |  |  |
| --- | --- | --- |
| **Rank** | **Industries** | **Pollutants released in pound** |
| **1** | Chemicals | 5.95E+07 |
| **2** | Paper | 4.29E+07 |
| **3** | Food | 1.56E+07 |
| **4** | Petroleum | 1.00E+07 |
| **5** | Plastics and Rubber | 7.27E+06 |
| **6** | Fabricated Metals | 5.77E+06 |
| **7** | Primary Metals | 5.67E+06 |
| **8** | Transportation Equipment | 5.54E+06 |
| **9** | Printing | 2.69E+06 |
| **10** | Other | 1.55E+06 |
| **11** | Electric Utilities | 1.32E+06 |
| **12** | Metal Mining | 1.32E+06 |
| **13** | Wood Products | 1.22E+06 |
| **14** | Furniture | 7.65E+05 |
| **15** | Petroleum Bulk Terminals | 6.59E+05 |
| **16** | Machinery | 6.04E+05 |
| **17** | Chemical Wholesalers | 5.66E+05 |
| **18** | Nonmetallic Mineral Product | 5.47E+05 |
| **19** | Miscellaneous Manufacturing | 3.34E+05 |
| **20** | Electrical Equipment | 2.85E+05 |
| **21** | Beverages | 2.45E+05 |
| **22** | Computers and Electronic Products | 1.82E+05 |
| **23** | Textiles | 1.78E+05 |
| **24** | Hazardous Waste | 1.57E+05 |
| **25** | Textile Product | 6.56E+04 |
| **26** | Tobacco | 4.59E+04 |
| **27** | Leather | 1.36E+04 |
| **28** | Apparel | 1.45E+03 |
| **29** | Coal Mining | 8.31E+02 |
| **30** | Publishing | 1.40E-01 |