



# Dirty Dozen Project

Air Alliance Houston

# Dirty Dozen Project

## Air Alliance Houston

### MOTIVATION

This work seeks to identify the worst industrial polluters in the Houston area by integrating several datasets including Toxic Release Inventory, EPA Enforcement and Compliance, TCEQ Enforcement, and Risk Screening Environmental Indicators (RSEI).

Industrial work is vital for economic growth and often comes at the cost of environmental degradation, and impact air quality and public health. The Dirty Dozen Project will use data analytics, cross-referencing techniques, and data visualization, to discover patterns, and correlations to ultimately identify the entities contributing significantly to air pollution in the Houston area.

The datasets used in this project provide multiple points of view of industrial activities, with data about releases of hazardous substances, environmental violations, and risk indicators. With data analysis and data visualization, the Dirty Dozen Project aims to provide a detailed description of the air pollution in Houston and help our readers identify sectors and industries that contribute the most to this problem.

Table of Contents

MOTIVATION.....2

TOXIC RELEASE INVENTORY (TRI) .....4

    TRI - Chemicals Releases ..... 6

TRI - Chemicals Releases & RSEI Hazard Score ..... 7

EPA GREENHOUSE GAS..... 9

    EPA Greenhouse – Carbon dioxide (CO2) ..... 11

    EPA Greenhouse – Methane (CH4) ..... 12

    EPA Greenhouse - Nitrous Oxide (N2O) ..... 13

TEXAS COMMISSION ON ENVIRONMENTAL QUALITY (TCEQ) -NOTICES OF  
ENFORCEMENT (NOE) ..... 14

EPA ENFORCEMENT AND COMPLIANCE ..... 16

    Penalties..... 16

    Compliance Status ..... 17

    Demographics..... 18

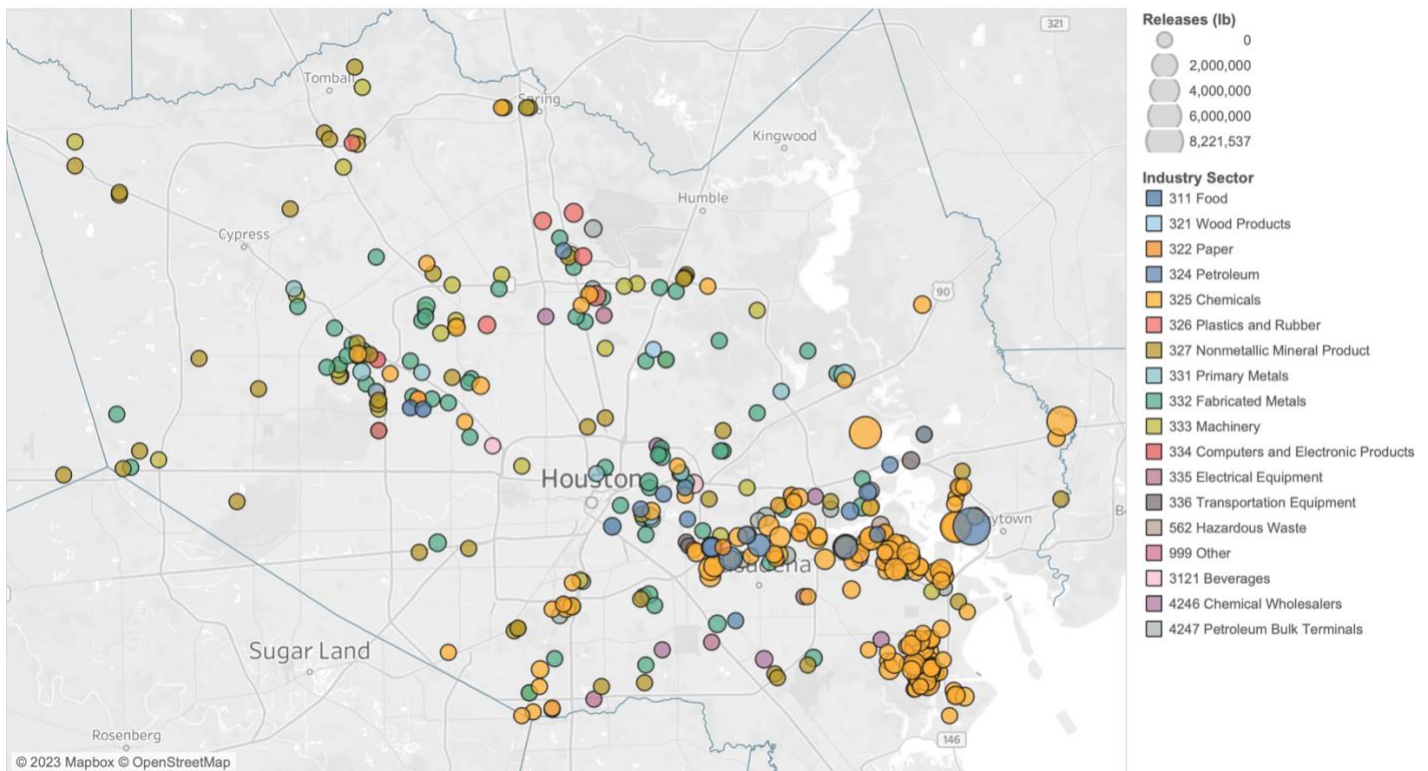
CONCLUSION..... 19

BIBLIOGRAPHY ..... 20

# TOXIC RELEASE INVENTORY (TRI)

The Toxic Release Inventory (TRI) is a resource for learning about toxic chemical releases and pollution prevention activities reported by industrial and federal facilities. Their database provides detailed information about the release and management of certain toxic chemicals by industrial facilities. It is administered by the Environmental Protection Agency (EPA). TRI was established under the Emergency Planning and Community Right-to-Know Act (EPCRA) and later expanded by the Pollution Prevention Act. Using data from 2018 – 2022 ((EPA) E. P., TRI Toxics Tracker, n.d.)

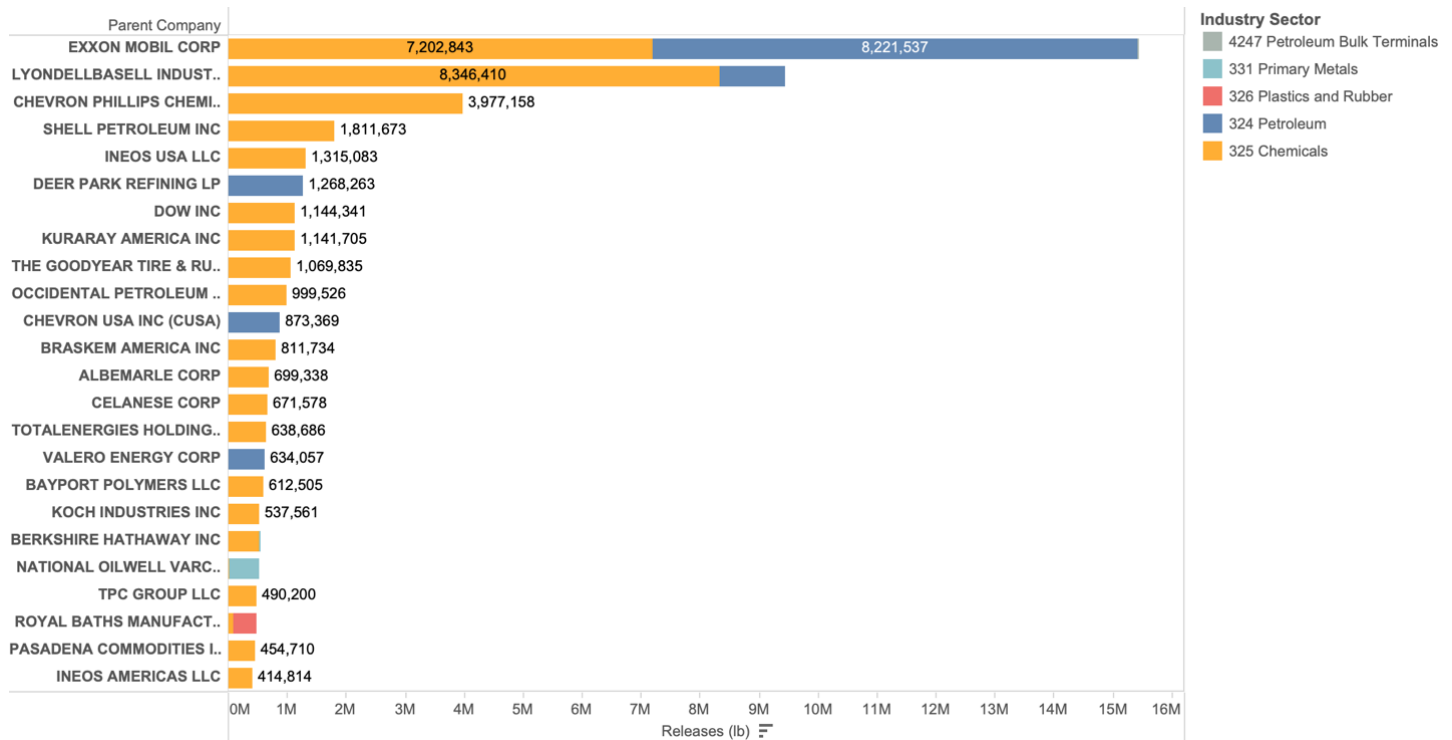
We discover some meaningful insights about the industries that contribute to the greatest number of chemical releases to the air. "Release" refers to different ways that toxic chemicals from industrial facilities enter the air. Releases include spilling, leaking, pumping, pouring, emitting, emptying, discharging, injecting, escaping, leaching, dumping, or disposing into the environment.



\_(Pacheco, Chemicals in TRI Data, 2023)

The most predominant industry in Harris County, is the petrochemical industry, as shown in the map above. The visualization allows us to identify clusters of these industrial polluters. There are many more of them in areas South-East, like Pasadena and Baytown, in comparison to any of the other area in Harris County.

Next, is a closer look at the top 25 out of 350 companies, sorted by the amount of chemicals released in the past 5 years (2018-2022)

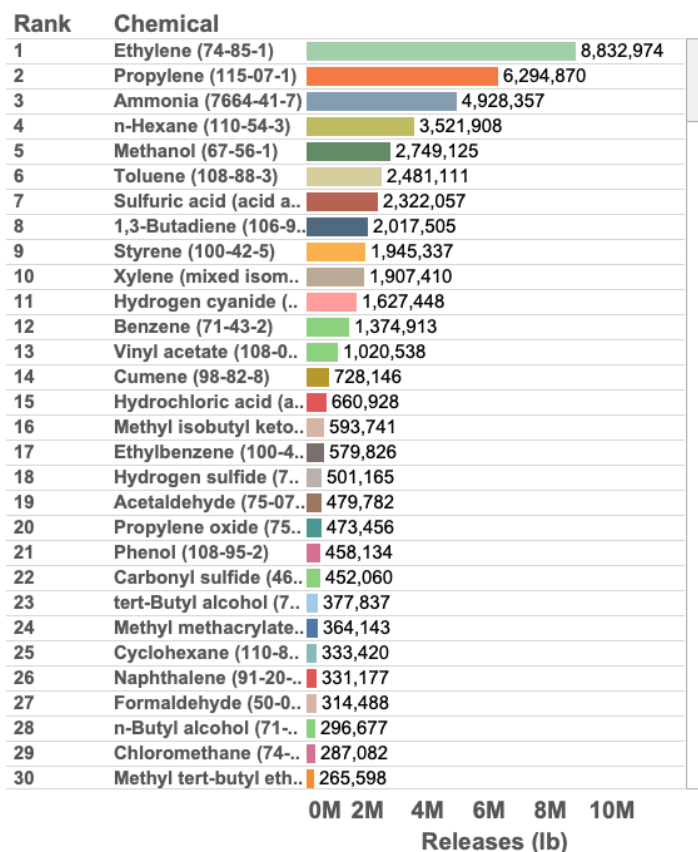


- **Exxon Mobil Corp**, one of the largest international oil & gas companies, also one of the largest refiners and marketers of petroleum products and chemicals where top on the list.
- **LyondellBasell**, a chemical company that manufactures and sells plastics, and petrochemicals, also refines crude oil to convert it into products such as diesel, gasoline, and jet fuel was number 2.
- **Chevron Phillips Chemical**, produces chemicals that are essential to the manufacturing of industrial products such as polyolefins, aromatics, styrene, specialty chemicals, and plastics.

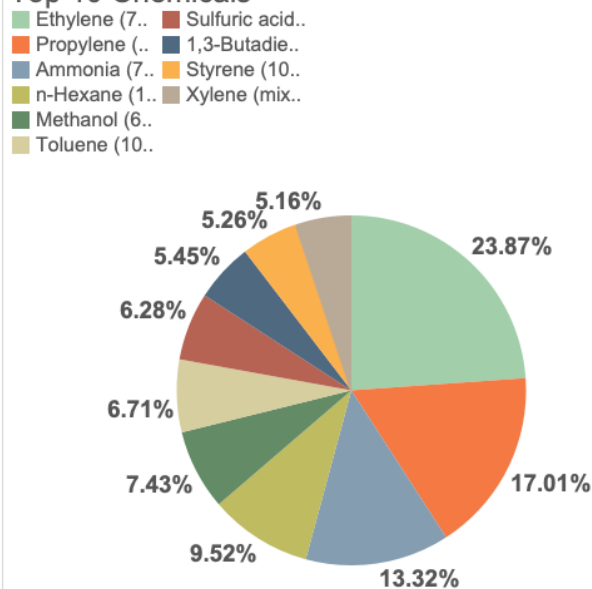
These 3 petrochemical companies are responsible for 60% of the amount of chemicals released in Harris County. Also, they are all part of the S&P 500, a stock market index that is meant to track the U.S. equity market. The index is made up of 500 of the largest public companies.

## TRI - Chemicals Releases

The TRI data set provided information about the different chemicals that go into the air. There were 272 different chemicals in our data set.



Top 10 Chemicals



We will explore more in-depth a few of the chemicals that appear at the top of this graph:

- **Ethylene:** This is widely used in the chemical industry, and much of this production goes toward creating polyethylene, which is a plastic containing polymer chains of ethylene units in various chain lengths. Ethylene is also an important natural plant hormone and is used in agriculture to accelerate the ripening of fruits. Metabolic studies in animals and man have revealed that ethylene is metabolized to ethylene oxide which is known to have carcinogenic and mutagenic effects. ((IARC), n.d.)
- **Propylene:** It is used in the production of many organic chemicals including resins, plastics, synthetic rubber, and gasoline. The primary human exposure to propylene is in the occupational setting. Consumer exposure to propylene is only likely from the low levels of heating gases and some anthropogenic sources such as combustion products of organic matter, motor vehicle exhaust, and cigarette smoke. ((IARC), n.d.)
- **1,3-Butadiene:** This is produced through the processing of petroleum and is mainly used in the production of synthetic rubber but is also found in smaller amounts in plastics and fuel. Acute low exposures may irritate the eyes, throat, nose, and lungs. Based on human and animal studies, the EPA has classified 1,3-butadiene as a known human carcinogen. ((IARC), n.d.)

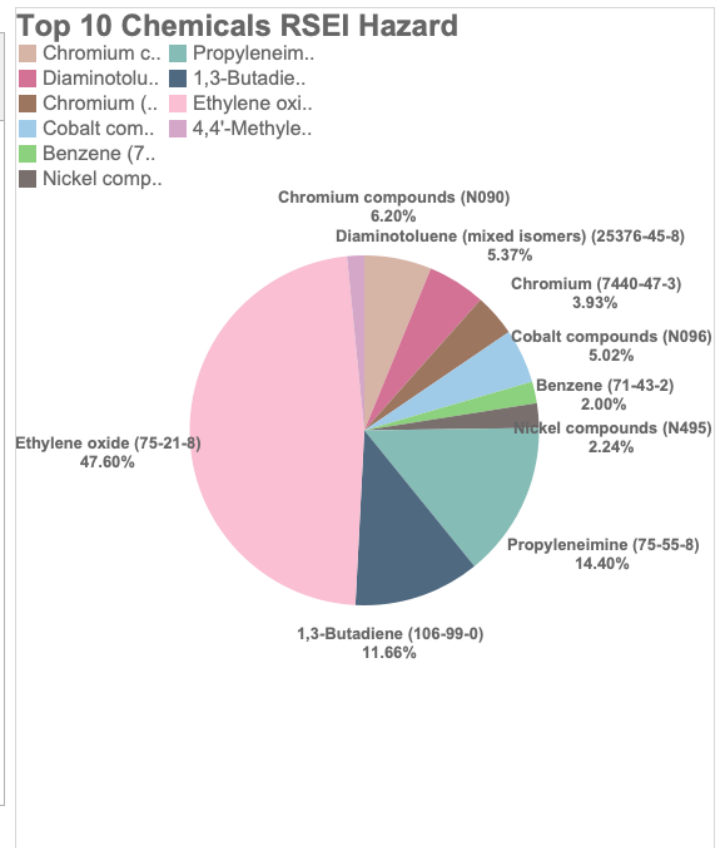
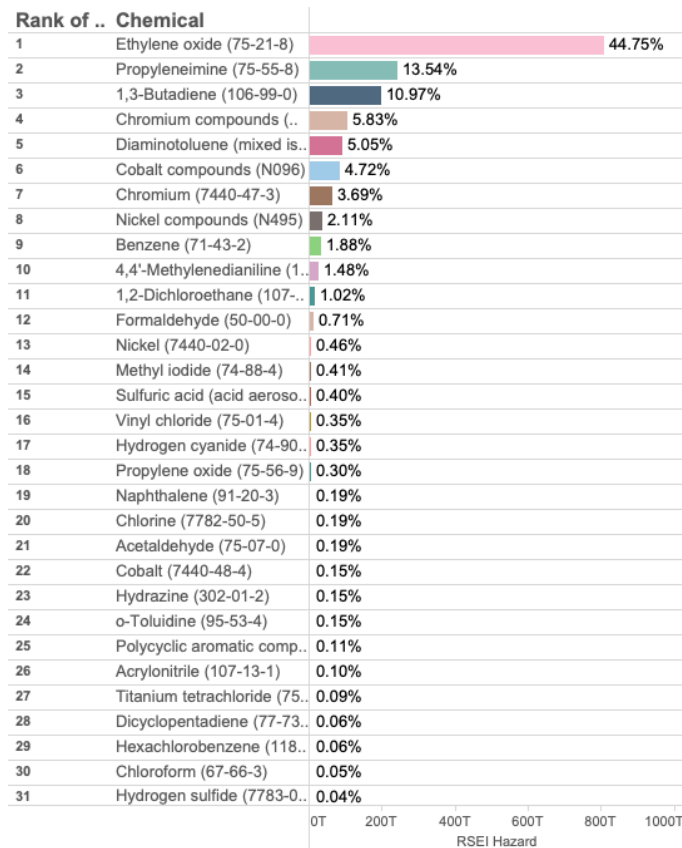


## TRI - Chemicals Releases & RSEI Hazard Score

We have explored the chemicals that are released the most, however with our data set and information from the TRI, we obtained an “RSEI Hazard Score”. This score is the waste management activity quantity (e.g., chemical quantity released to the environment or transferred off-site for further waste management) in pounds per year (TRI Pounds), multiplied by a chemical- and exposure route-specific toxicity weight ((EPA) E. P., Understanding RSEI Results, n.d.).

RSEI Hazard score can be calculated for any waste management activity reported to TRI. These scores account for the size of a chemical release, the size and locations of potentially exposed populations, and the relative toxicity.

Our graph below shows us the chemical with the highest RSEI Hazard score.

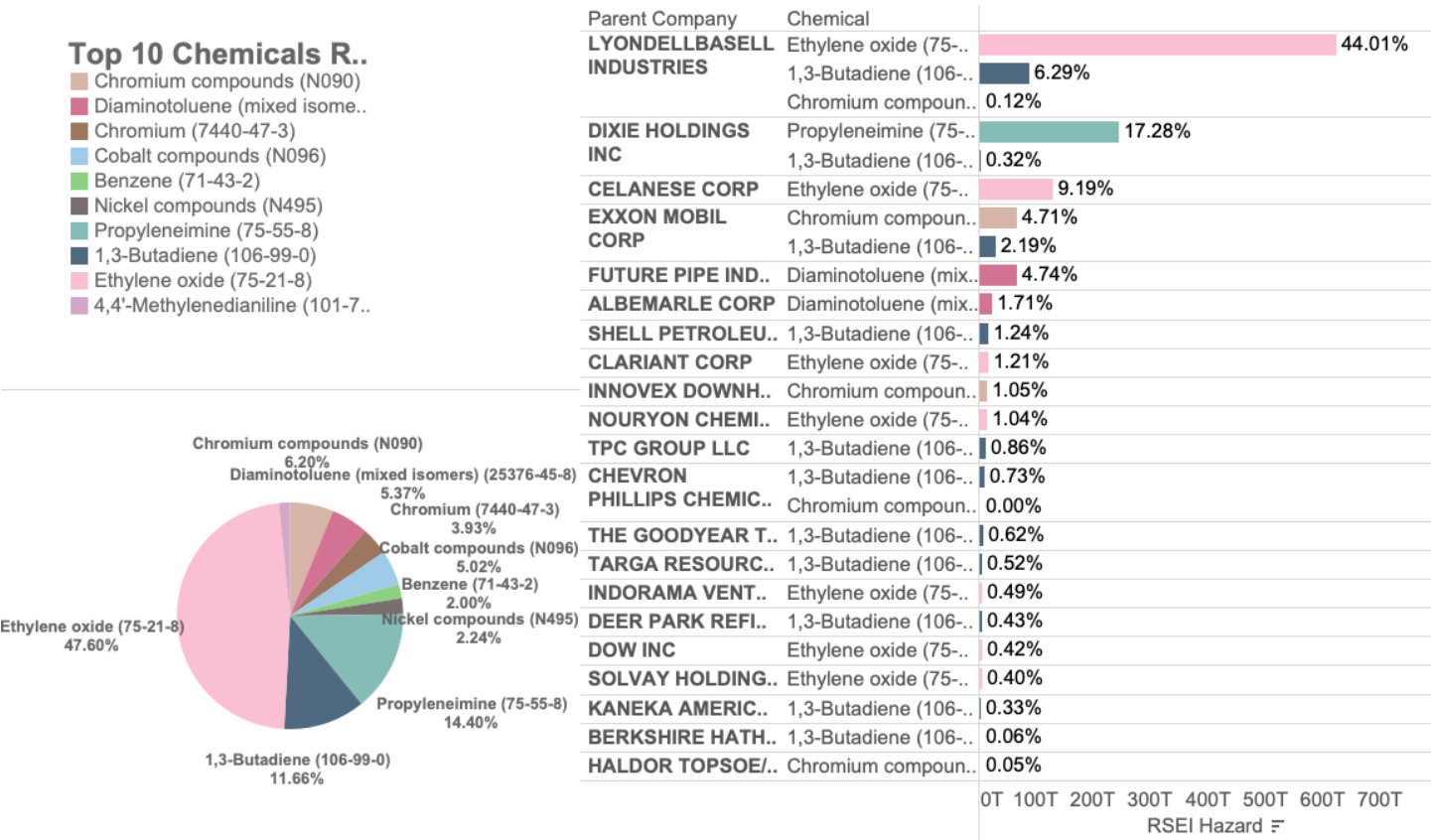


When we compare the results from the chemicals weighted just by the TRI release in pounds, and the chemicals by the RSEI Hazard score, there is a change in the chemicals shown.

For example, in the RSEI Hazard score, Ethylene oxide is number 1, which is a chemical that has been scientifically proven that chronic exposure is highly associated with the occurrence of cancer, reproductive effect, neurotoxicity, and mutagenic changes.

Propyleneimine is number 2, a potential occupational carcinogen. 1,3-Butadiene is number 3, which repeats in both graphs, with the difference that it occupies a higher position in the RSEI Hazard score, and the EPA has classified it as a known human carcinogen.

Here are the top companies, and their percentages on emissions of the top hazardous chemicals. Some of these names repeat in our graph of the top companies by number of releases. This is very important because it shows us that they are not only releasing large amounts, but they also release hazardous chemicals.



Ultimately the TRI data allows us to explore the different facilities and provides information about the industry type, amount of pounds released in a period, the RSEI Hazard score, and the list of chemicals emitted by these companies.

The most important part of the TRI dataset is the differences in results when comparing the chemicals and the pounds released versus the chemicals and the RSEI Hazard score. According to the EPA, the RSEI “scores” results are not to be judged only by high or low, however in this data set the top 10 chemicals by RSEI Hazard score, are mostly human carcinogens.

The TRI data set provides us with enough variables to visualize and understand cluster points, the size of the companies, the type of industry, and the most released chemical. With this information it is easier to distinguish the mega-polluters from the smaller companies.

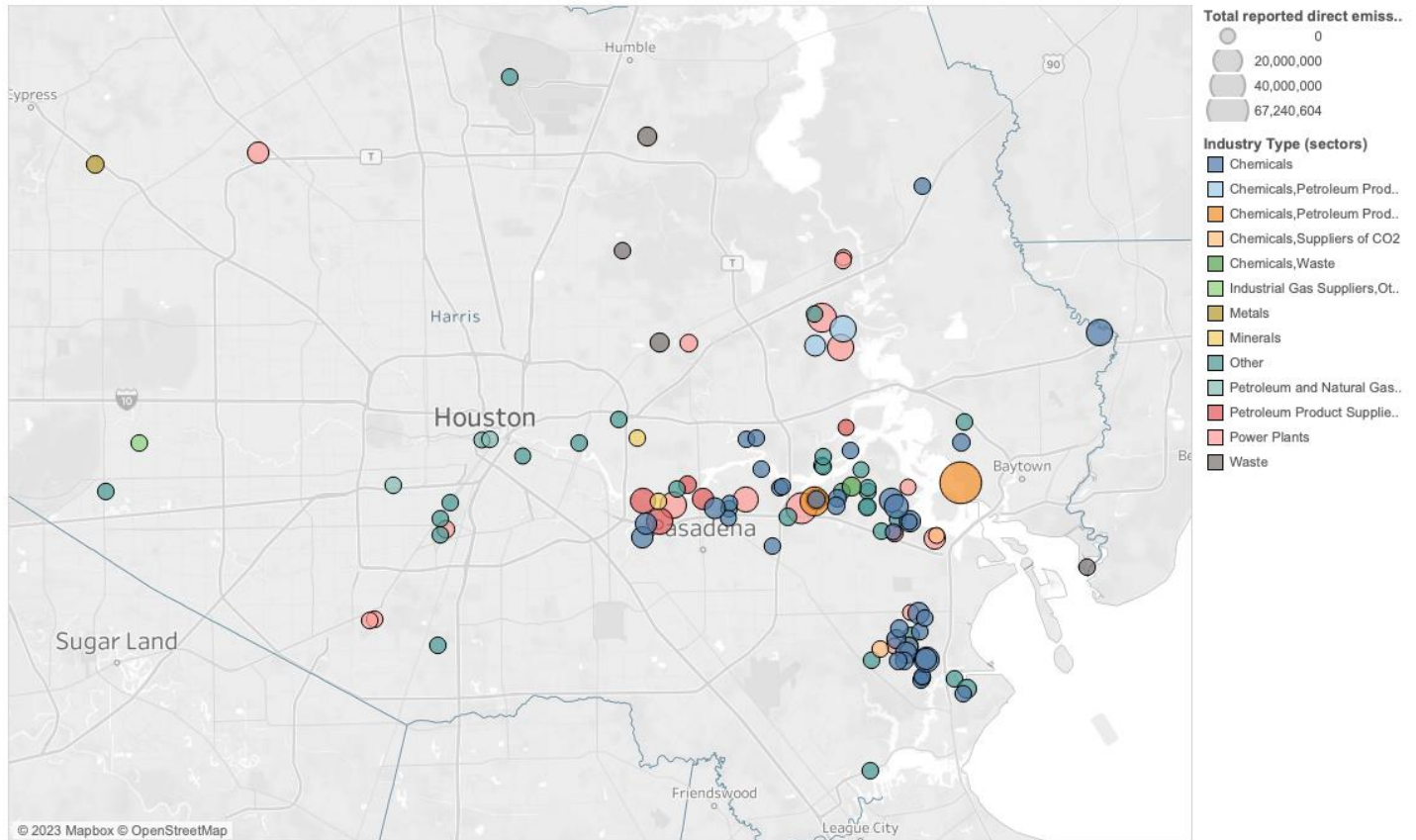


# EPA GREENHOUSE GAS

Greenhouse gases are the type of gases that trap heat in the atmosphere.

Transportation, electricity production, industry, commercial, residential, and agriculture, are the sectors that contribute the most to Greenhouse gases. For industry, the gases are produced by burning fossil fuels for energy, and gases are released from certain chemical reactions, when producing goods from raw materials.

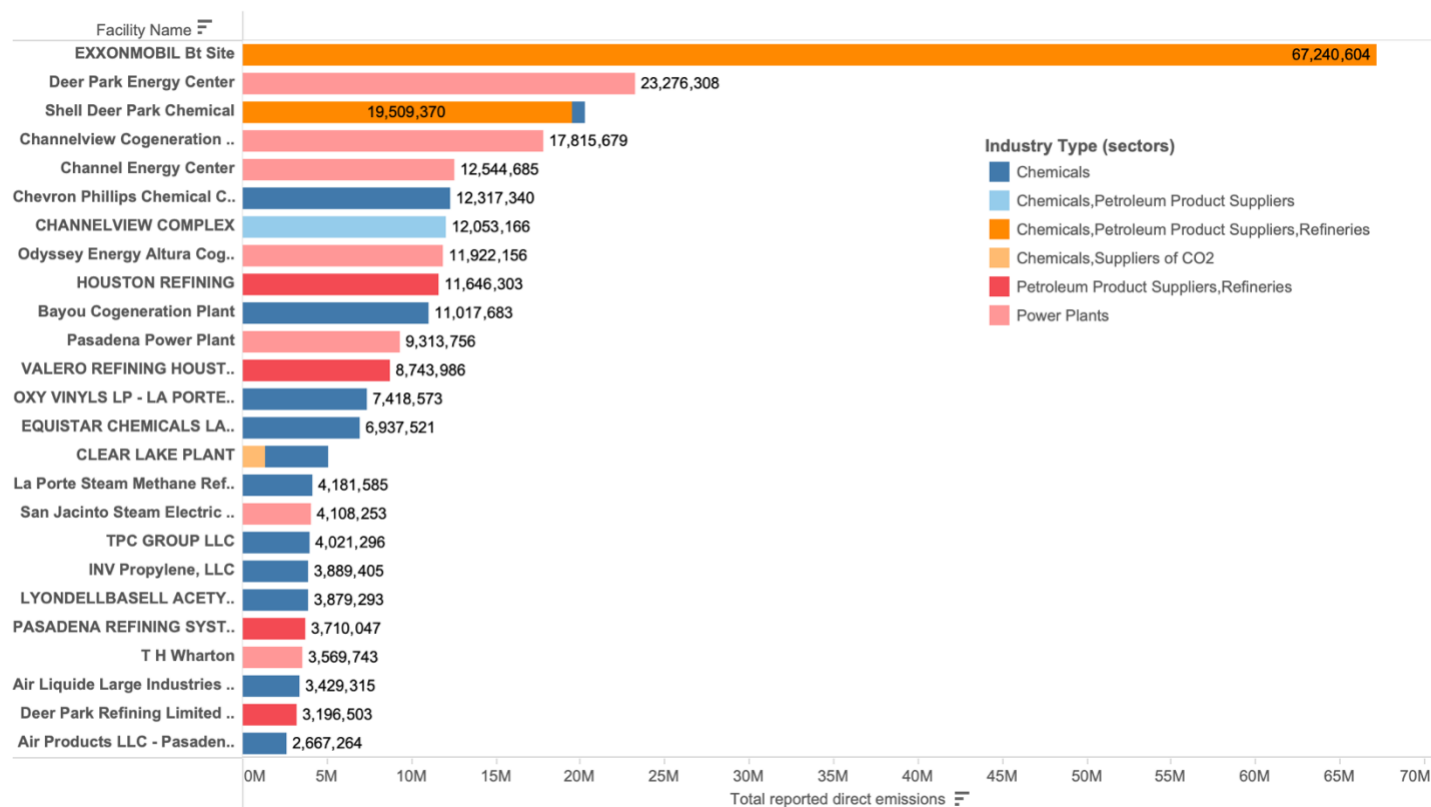
To explore the industry type, we used the EPA FLIGHT (Facility Level Information on Greenhouse Gases Tool) ((EPA) U. E., n.d.)



\_(Pacheco, Greenhouse Gases Visualization, 2023)

We gathered data from the past 5 years, from the 108 large facilities in Harris County. The dashboard allows us to locate the companies that have the most direct emissions of Greenhouse gases. Pasadena and Baytown have the most noticeable cluster of points because most of the facilities operate in the petrochemicals industry (petroleum product suppliers, refineries, suppliers of CO2, and waste).

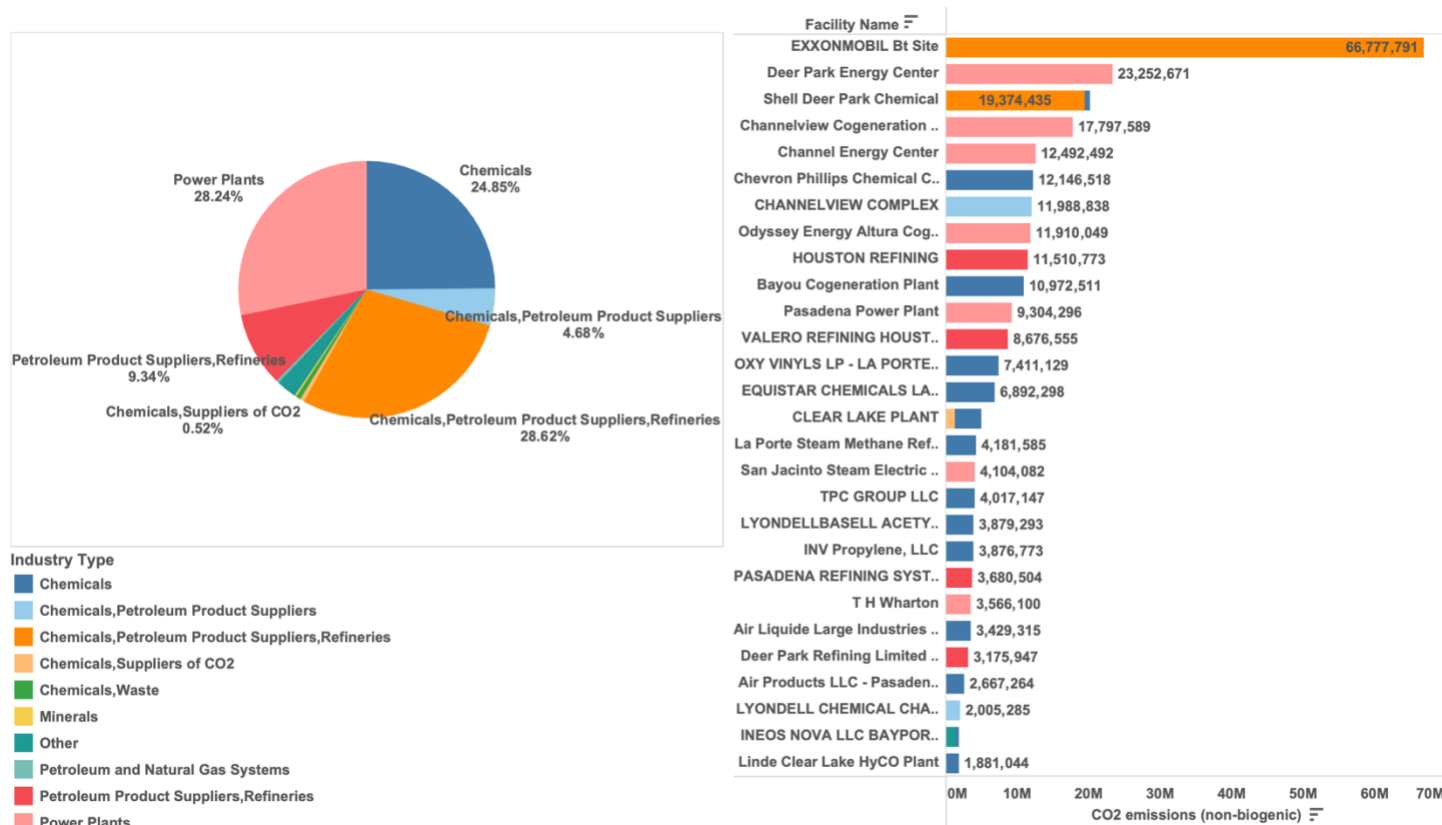
Here are the top 25 out of 108 companies, sorted by their number of direct emissions in the past 5 years (2018-2022)



- ExxonMobil Bt (Baytown) Site:** This ExxonMobil plant manufactures approximately 10 billion pounds of petrochemical products each year which are shipped via pipeline. Baytown operates eleven natural gas cogeneration units, used to generate electricity and high-pressure steam across the Baytown Complex. (ExxonMobil, 2022)
- Deer Park Energy Center:** Is the largest power plant in Calpine's fleet (America's largest generator of electricity from natural gas and geothermal resources). The plant captures CO<sub>2</sub>, supplies steam to Shell Chemical Company and generates electric power that is sold into the wholesale market. (Corporation, n.d.)
- Shell Deer Park Chemical:** It manufactures base chemicals or raw material chemicals and sells them to other chemical companies that turn them into thousands of consumer products. Most chemicals from Deer Park are shipped via pipelines, but the site also uses ships, barges, rail cars, and trucks to supply customers in the U.S. and foreign countries. (Shell)

## EPA Greenhouse – Carbon dioxide (CO<sub>2</sub>)

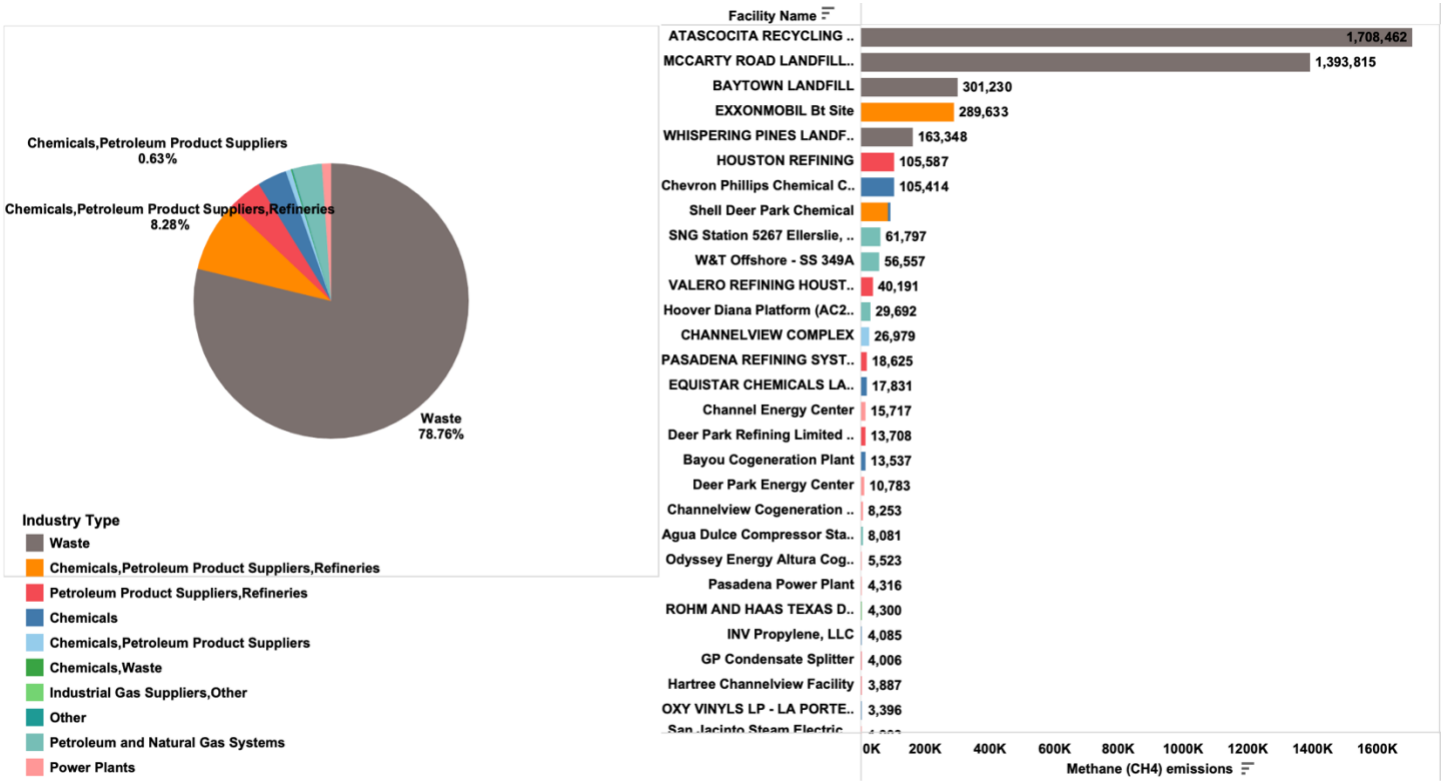
Carbon dioxide is the primary Greenhouse gas emitted. It enters the atmosphere through burning fossil fuels (coal, natural gas, and oil), solid waste, trees, and other biological materials, and from certain chemical reactions (e.g., cement production) (Agency, Overview of Greenhouse Gases, n.d.).



Above are the top companies with direct emissions of CO<sub>2</sub>, and their industry type. The companies that contribute the most to the direct emissions of CO<sub>2</sub> are mainly the Chemical and Power Plants industries. CO<sub>2</sub> can produce negative effects in humans, such as headaches, dizziness, restlessness, difficulty breathing, tiredness, elevated blood pressure, and increased rate, and ultimately high levels of CO<sub>2</sub> can cause poor air quality. (Power, n.d.)

# EPA Greenhouse – Methane (CH4)

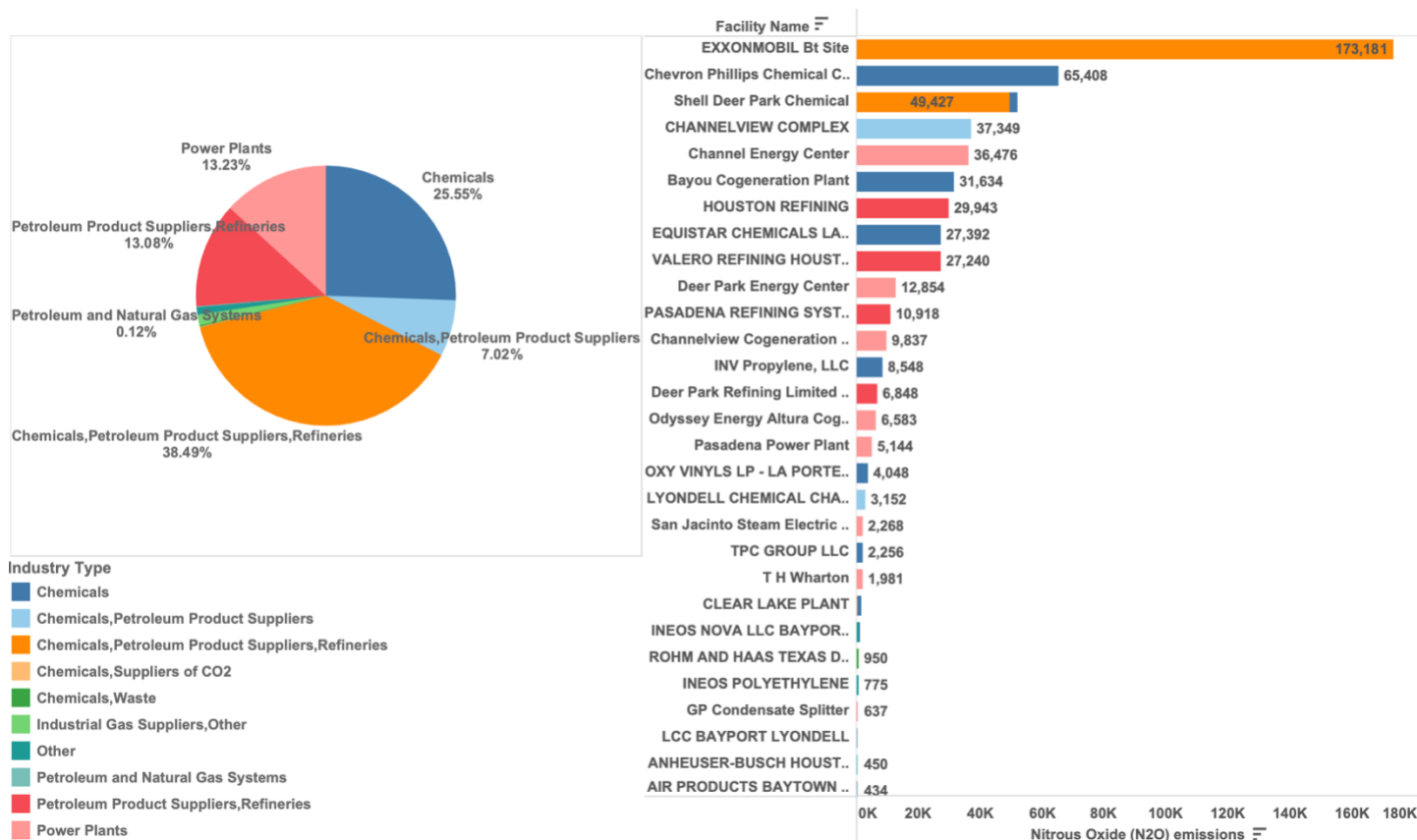
Methane is emitted during the production and transport of coal, natural gas, and oil. Methane emissions also result from livestock and other agricultural practices, land use, and the decay of organic waste in municipal solid waste landfills. One important part of information about CH<sub>4</sub> is that it’s lifetime in the atmosphere is much shorter than carbon dioxide (CO<sub>2</sub>), but CH<sub>4</sub> is more efficient at trapping radiation than CO<sub>2</sub>. Pound for pound, the comparative impact of CH<sub>4</sub> is 28 times greater than CO<sub>2</sub> over 100 years. (Agency, Overview of Greenhouse Gases, n.d.)



The industry type changes when it comes to Methane emissions, in this case, the “waste” industry is the first place by a large difference, contributing 78% of the total emissions of methane in Harris County.

## EPA Greenhouse - Nitrous Oxide (N<sub>2</sub>O)

Nitrous oxide is emitted during agricultural, land use, and industrial activities; combustion of fossil fuels and solid waste; as well as during treatment of wastewater. It is worthy of note that the impact of 1 pound of N<sub>2</sub>O on warming the atmosphere is 265 times that of 1 pound of carbon dioxide. (Agency, Overview of Greenhouse Gases, n.d.)



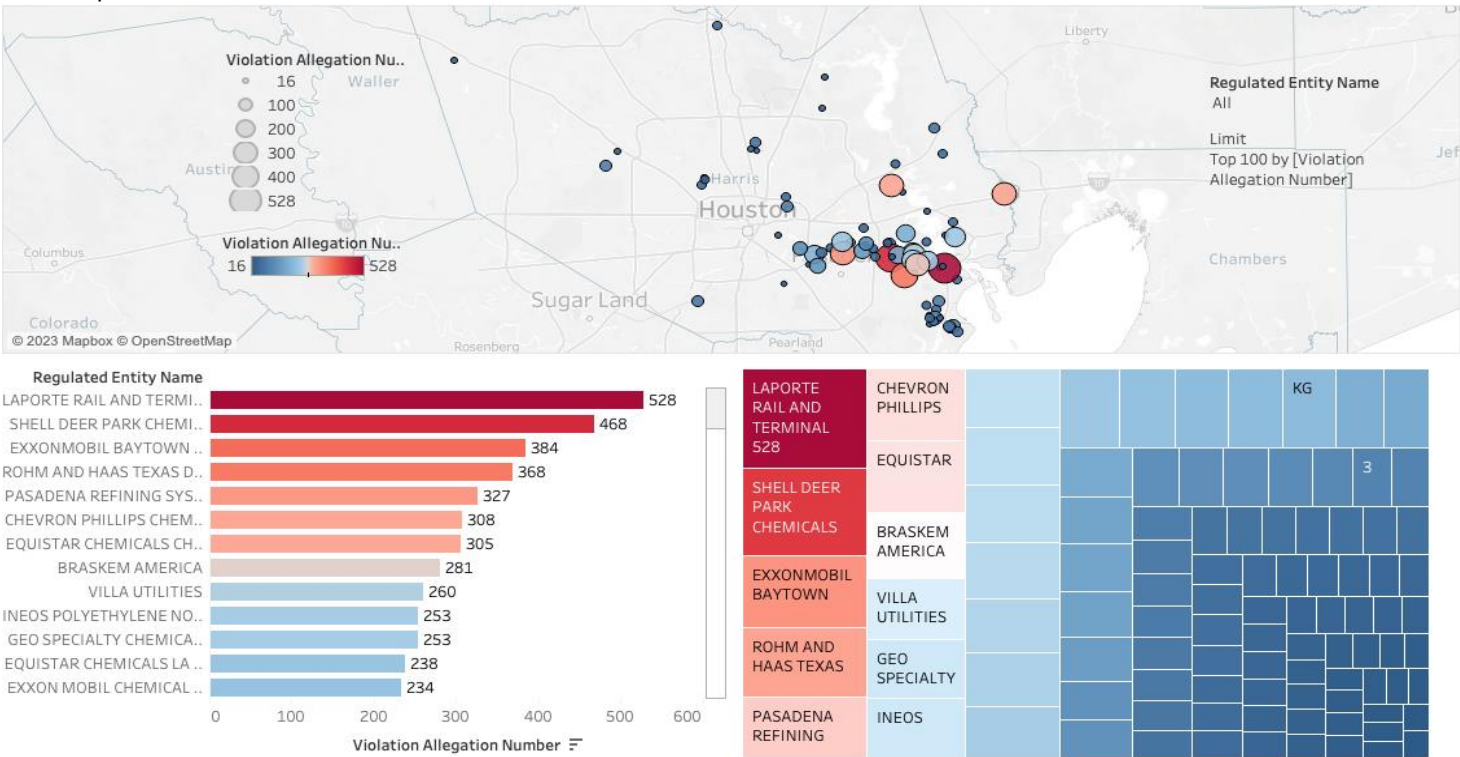
Chemicals, Petroleum Product Suppliers, and Refineries are the highest contributors to direct emissions of N<sub>2</sub>O in the past 5 years in Harris County. ExxonMobil Bt (Baytown) site wins first place again as the main direct emitter of Carbon Dioxide and Nitrous Oxide.

Overall, the EPA Greenhouse FLIGHT tool provides information about greenhouse gas emissions from large facilities in the U.S. with a focus on specific types of gases.

# TEXAS COMMISSION ON ENVIRONMENTAL QUALITY (TCEQ) – NOTICES OF ENFORCEMENT (NOE)

Next, we utilized the Texas Commission on Environmental Quality’s (TCEQ) records of Investigations resulting in a Notice of Enforcement (NOE) and the details associated with each violation of the different facilities. This database was the least user-friendly, accessing options to narrow and search for data of specific years, and county were not in plain sight. A query-building option was used to obtain the data needed. (Quality, n.d.)

The second limitation was that this data set contained only text. To extract some important information, we decided to count the number of violations the facilities had over the past five years and represent them in this dashboard:



\_(Pacheco, TCEQ Data Visualization, 2023)

**Laporte rail and terminal** was in first place. This is an industrial complex located along the Houston Ship Channel near the Fred Hartman Bridge. The facility offers chemical and industrial manufacturers leased land with access to numerous services. In addition, they offer rail car storage inside and outside the secured facility. (LPR)



Due to the lack of numerical data to perform data analysis, we decided to plot the most repeated words in the violation allegations reports on a word cloud.



This word cloud allows us to explore and understand what types of violations facilities in Harris County have committed in the past 5 years. The word cloud represents the words repeated the most in larger font size, for example, the words failure, failed, emissions, period, monitoring, and incident, among others, were the words that repeated the most in the violations reports.

# EPA ENFORCEMENT AND COMPLIANCE

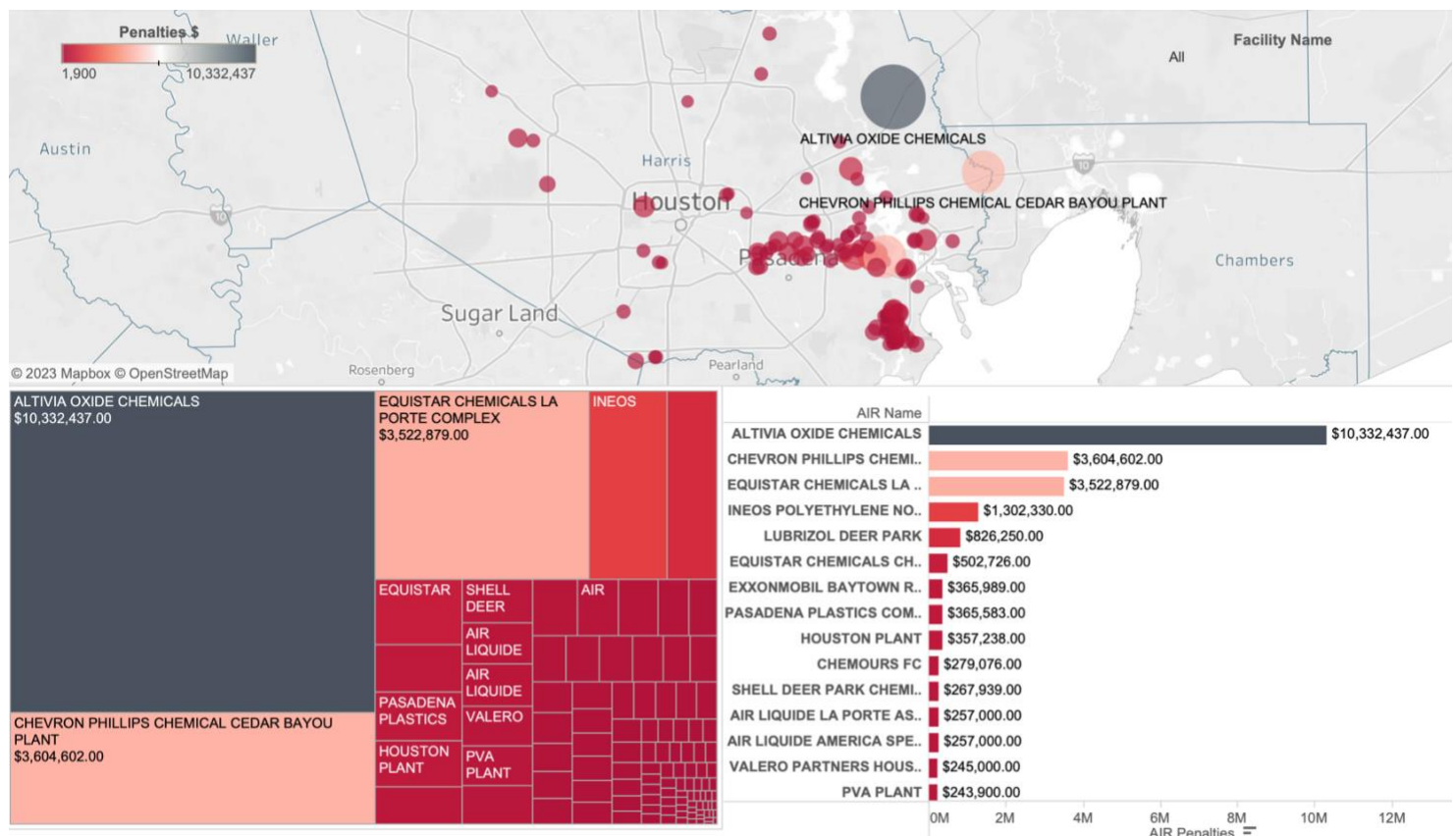
To obtain and assess historical data regarding compliance with environmental regulations for facilities, we used the ECHO (Enforcement and Compliance History Online) tool from the EPA (the United States Environmental Protection Agency). Since we are exploring the pollution in the Houston area, we focused on facilities by media program “Air (CAA)”, where we obtained data of the TRI facilities with enforcement and compliance data. (Agency, Environmental Compliance History Online (ECHO), n.d.)

## Penalties

We obtained data from 108 companies that shared their reports with the TRI and were found in the ECHO database. The dashboard below provides a visualization of 5 years (2018-2022) of compliance historical data.

The visualization allows us to quickly identify the companies that have paid the most amount of money in penalties.

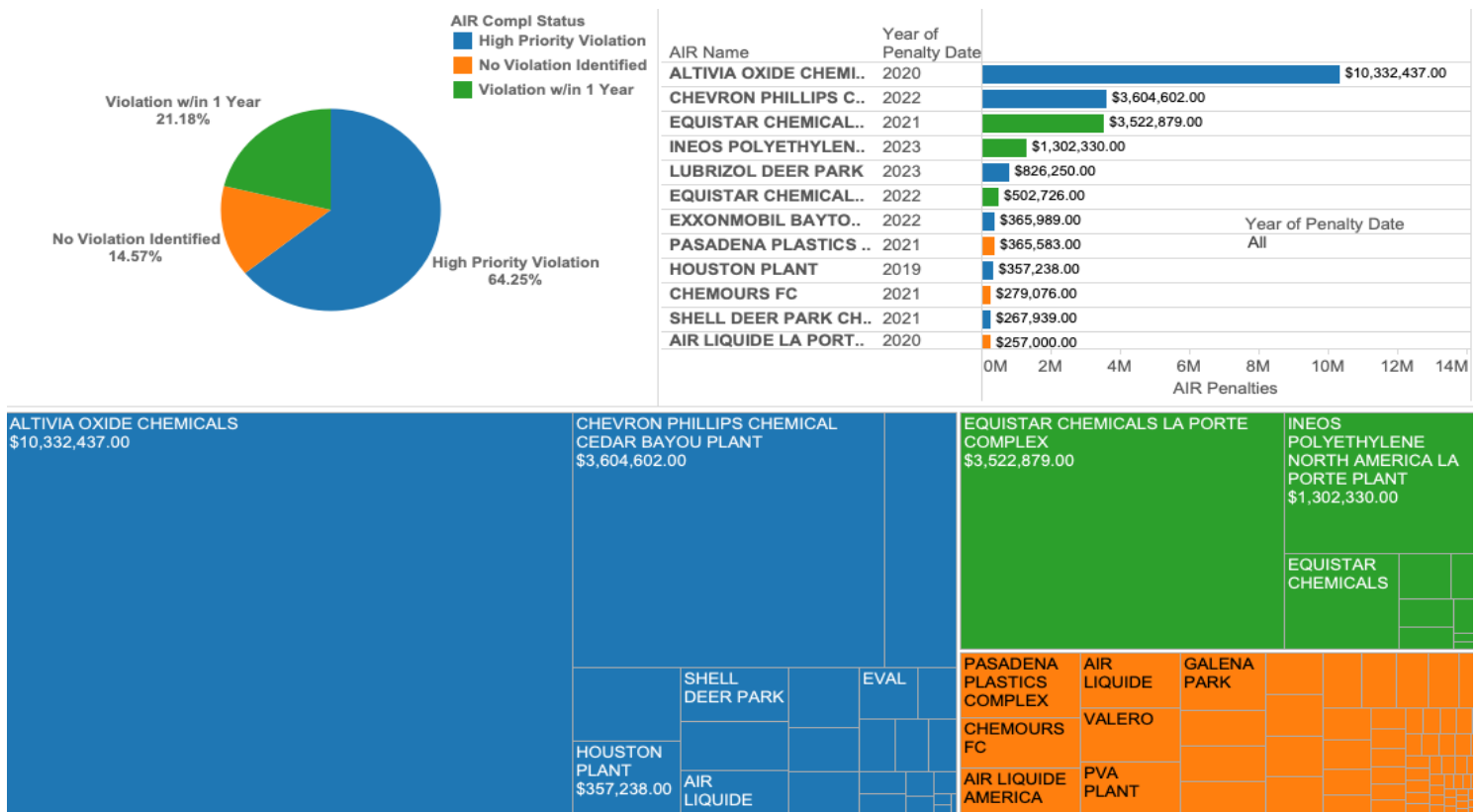
Again, the clusters are found in the Pasadena and Baytown areas with Altivia Oxide Chemicals, and ExxonMobil Baytown Refinery, leading the charts.



\_(Pacheco, EPA Penalties Data Visualization, 2023)

## Compliance Status

The EPA enforcement and compliance data also categorizes the type of violations that facilities have committed. The next graph reveals that the status of these violations had an impact on the amount of money the facilities were fined.



Most of the violations (64.25%) in the past 5 years, were “High Priority”, some examples are:

- Exceedance of a major stationary source annual emission threshold, as defined in the NSR regulations, by a synthetic minor stationary source.
- Failure to obtain a New Source Review (NSR) permit and/or install Best Available Control Technology or Lowest Available Emission Reductions for any new major stationary source or major modifications at a major stationary source.
- Violations of federally enforceable work practices, testing requirements, monitoring requirements, recordkeeping or reporting that substantially interferes with enforcement or determination of a facility's compliance requirements. (Agency, DFR Data Dictionary, n.d.)

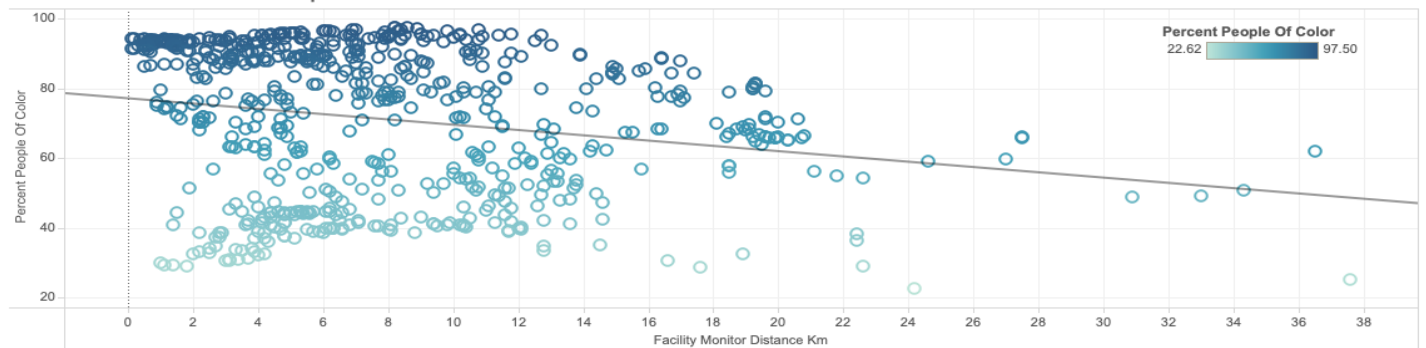
followed by 21.18% for violations within 1 year, this violation type can be explained more when utilizing the year filter to locate when the violation occurred. 14.57% where no violation was identified, however, even with no violation, some facilities still received monetary penalty fines. This could be due to remaining unsolved fines from penalties of past years.

Combined with the insight obtained from our previous data set, we can continue to see and understand the trends of air pollution in Harris County. Most of the fines imposed by the EPA are due to High Priority Violations, and considering the number of releases the top facilities have, is alarming, especially when we see almost the same companies at top of all the databases we explored.

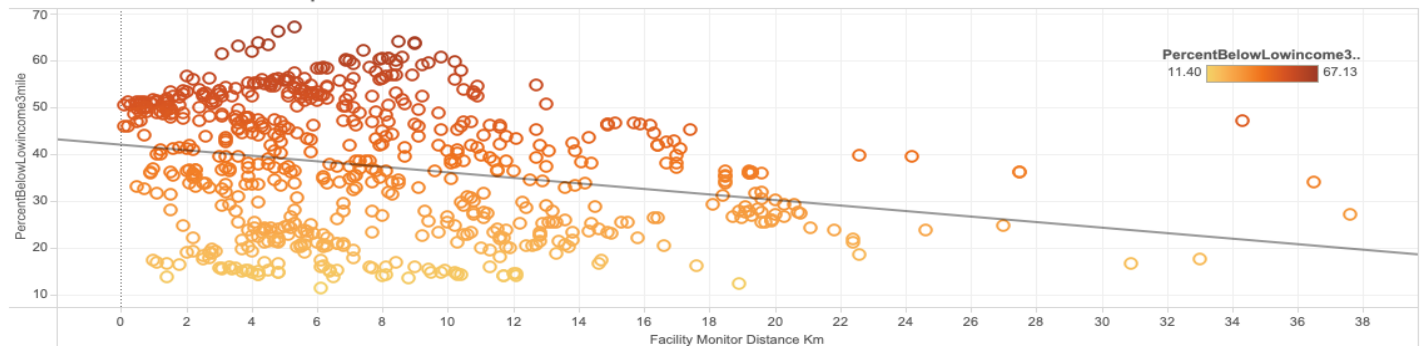
## Demographics

One of the concerning factors of the air pollution in Harris County is that most of the facilities are built around areas with high percentages of Hispanics, African Americans, low income, and people over 60 years old (Agency, n.d.). The next graph is a correlation plot of the facility air monitor distance in kilometers versus the percent of people of color, and the percent of people of low income.

Monitor distance vs Percent People of Color



Monitor distance vs Percent People Below Low Income



There is a negative correlation between these factors, with a downhill trend line in both correlation plots. The p-value was below 0.05, and R-square values were .045 (people of color), and .065 (people below income), making the demographical variables statistically significant to the variable for the facility monitor distance. This negative correlation means that when the facility monitor distance increases, the percentage of people of color, and the percent of people of below income, will decrease. Therefore, these facilities are concentrated, (as their facility monitor provides us with a location), near people of color and people of low income.

## CONCLUSION

Here are the “dirty dozen” (not in any particular order), industry polluters, according to the findings and patterns that these 5 datasets contained:

- ExxonMobil Baytown Site
- Chevron Phillips Chemical
- Shell Deer Park Chemical
- LyondellBassell Industry
- Altivia Oxide Chemicals
- Equistar Chemicals Channelview
- Atascocita Recycling Center
- McCarty Road Landfill
- Deerpark Energy Center
- Dixie Holdings Inc
- Ineos Oxide
- Celanese Clearlake

These companies rank among the top of one, or multiple, dashboards created with the data from our five resources. Also, if you look up on Google any of these facilities with the word “pollution” after, there is a large number of links, news, and reports on these companies. Lastly, these are the facilities that contribute the most to Harris County air pollution, with hazardous chemical releases, greenhouse gases, waste, and climate injustice.

The information presented is not effective without action, because all the data is self-reported by the facilities, there is no governmental entity that works to ensure the validity and veracity of the data reported. Even with that limitation, we can see the same patterns, trends, and the same group of companies always on top of the metrics, fines, releases, and air pollution.

The work that Air Alliance Houston does is powerful because combines information with action. They knock on doors of the communities that suffer climate injustice and translate their data and findings to multiple languages to reach more people, because they are not afraid to point out the wrongdoings of industry. They fight every day because **they believe that “everyone has a right to breathe clean air”**. (Houston, n.d.)

## BIBLIOGRAPHY

- (EPA), E. P. (n.d.). Retrieved from TRI Toxics Tracker:  
<https://edap.epa.gov/public/extensions/TRIToxicsTracker/TRIToxicsTracker.html>
- (EPA), E. P. (n.d.). *Understanding RSEI Results*. Retrieved from  
<https://www.epa.gov/rsei/understanding-rsei-results#:~:text=RSEI%20Scores%20are%20calculated%20as,and%20a%20chemical%27s%20relative%20toxicity>
- (EPA), U. E. (n.d.). *EPA Greenhouse Gas Reporting Program*. Retrieved from  
<https://ghgdata.epa.gov/ghgp/main.do#/facility/?q=Find%20a%20Facility%20or%20Location&st=&bs=&et=&fid=&sf=11001100&lowE=-20000&highE=23000000&g1=1&g2=1&g3=1&g4=1&g5=1&g6=0&g7=1&g8=1&g9=1&g10=1&g11=1&g12=1&s1=1&s2=1&s3=1&s4=1&s5=1&s6=1&s7=1&s8=1&s9=1&s10=1>
- (IARC), I. A. (n.d.). *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Volume 60 - Some Industrial Chemicals*. Retrieved from <https://monographs.iarc.who.int/wp-content/uploads/2018/06/mono60.pdf>
- Agency, U. E. (n.d.). *Study Finds Exposure to Air Pollution is Higher for People of Color Regardless of Region or Income*. Retrieved from <https://www.epa.gov/sciencematters/study-finds-exposure-air-pollution-higher-people-color-regardless-region-or-income>
- Agency, U. E. (n.d.). *DFR Data Dictionary*. Retrieved from <https://echo.epa.gov/help/reports/dfr-data-dictionary#compsum>
- Agency, U. E. (n.d.). *Environmental Compliance History Online (ECHO)*. Retrieved from <https://echo.epa.gov>
- Agency, U. E. (n.d.). *Overview of Greenhouse Gases*. Retrieved from <https://www.epa.gov/ghgemissions/overview-greenhouse-gases>
- Corporation, C. (n.d.). *Deer Park Energy Center*. Retrieved from Calpine :  
<https://www.calpine.com/deer-park-energy-center#:~:text=The%20Deer%20Park%20Energy%20Center,sold%20into%20the%20wholesale%20market.>
- ExxonMobil. (2022). *Baytown Olefins Plant 2022 Fact Sheet*. Retrieved from <https://corporate.exxonmobil.com/-/media/global/files/locations/united-states-operations/baytown/baytown-olefins-plant-2022-fact-sheet.pdf>
- Houston, A. A. (n.d.). Retrieved from Air Alliance Houston: <https://airalliancehouston.org/>
- LPR. (n.d.). *LPR Construction*. Retrieved from <https://www.lprandt.com/>
- Pacheco, O. (2023, 11). *Chemicals in TRI Data*. Retrieved from Tableau Public:  
<https://public.tableau.com/app/profile/oscar.pacheco4360/viz/chemicalsTRI/Story3>



- Pacheco, O. (2023, 11). *EPA Penalties Data Visualization*. Retrieved from Tableau Public:  
<https://public.tableau.com/app/profile/oscar.pacheco4360/viz/EPAPenalties/Story1>
- Pacheco, O. (2023, 11). *Greenhouse Gases Visualization*. Retrieved from Tableau Public:  
[https://public.tableau.com/app/profile/oscar.pacheco4360/viz/GH\\_Gases/Story1](https://public.tableau.com/app/profile/oscar.pacheco4360/viz/GH_Gases/Story1)
- Pacheco, O. (2023, 11). *TCEQ Data Visualization*. Retrieved from Tableau Public:  
<https://public.tableau.com/app/profile/oscar.pacheco4360/viz/tceq/Story1>
- Power, S. (n.d.). *Is Indoor Carbon Dioxide Really Harmful to Humans?* Retrieved from  
<https://www.soletairpower.fi/is-indoor-carbon-dioxide-really-harmful-to-humans/>
- Quality, e. C. (n.d.). *Texas Commission on Environmental Quality Notices*. Retrieved from  
[https://data.texas.gov/dataset/Texas-Commission-on-Environmental-Quality-Notices-/rua3-iswk/about\\_data](https://data.texas.gov/dataset/Texas-Commission-on-Environmental-Quality-Notices-/rua3-iswk/about_data)
- Shell. (n.d.). *About Shell Deer Park*. Retrieved from <https://www.shell.us/about-us/projects-and-locations/deer-park-manufacturing-site/about-shell-deer-park.html>