

Comparing Texas Emissions and Rainfall with Air Quality

<https://github.com/maevegr/BassettGRHowey>

Sara Bassett, Maeve Gualtieri-Reed, Xia Meng Howey

December 12, 2025

Contents

Rationale and Research Questions	5
Dataset Information	6
Exploratory Analysis	7
Analysis	8
Question 1: Is there a correlation between county level emission levels from power plants and county air quality?	8
How often is emissions data reported?	9
How are air quality stations spatially located compared to emissions reporting power plants?	10
Are annual county emissions correlated with annual county air quality?	11
Question 2: Is there a trend in air quality values and emissions over time in Harris county, Texas?	13
Question 3: Is there a relationship between precipitation and air quality between and within the counties of Harris and Dallas in Texas?	18
Summary and Conclusions	24
References	25

List of Tables

1	Linear Regression of Emissions vs. Ozone Air Quality	11
2	Linear Regression of Emissions vs. PM2.5 Air Quality	12
3	Linear Regression of Precipitation vs. Ozone Air Quality Harris County	21
4	Linear Regression of Precipitation vs. PM2.5 Air Quality Harris County	21
5	Linear Regression of Precipitation vs. Ozone Air Quality Dallas County	22
6	Linear Regression of Precipitation vs. PM2.5 Air Quality Dallas County	22
7	Dallas vs Harris Two Sample T test Mean Precipitation	23
8	Dallas vs Harris Two Sample T test Mean AQI Ozone	23
9	Dallas vs Harris Two Sample T test Mean AQI PM2.5	23

List of Figures

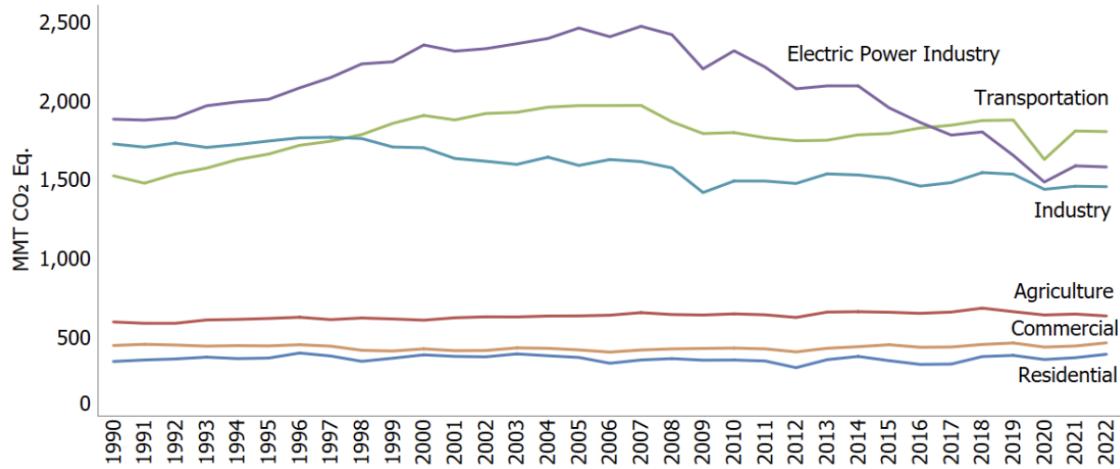
1	Listed and Emissions Reporting Texas Power Plants. 2019 - 2023 eGRID reports.	9
2	Air Quality Monitors & Annual Emissions from Power Plants and in Texas, 2019 - 2023. . . .	10
3	Average Daily Ozone AQI Value vs. Annual CO2 Equivalent Emissions (tons). Per Texas County, 2019 - 2023	11
4	Average Daily PM2.5 AQI Value vs. Annual CO2 Equivalent Emissions (tons). Per Texas County, 2019 - 2023.	12
5	Average daily PM2.5 AQI in Harris County, Texas, 2019-2023	14
6	Average daily ozone AQI in Harris County, Texas, 2019-2023	15
7	Annual CO2 equivalent in Harris County, Texas, 2019-2023	16
8	Annual methane emissions in Harris County, Texas, 2019-2023	17
9	Precipitation Harris County, 2019 - 2023	18
10	Precipitation Dallas County, 2019 - 2023	19
11	Ozone AQI Harris County, 2019 - 2023	19
12	Ozone AQI Dallas County, 2019 - 2023	20
13	PM2.5 AQI Harris County, 2019 - 2023	20
14	PM2.5 AQI Dallas County, 2019 - 2023	21

Rationale and Research Questions

This report examines how local conditions affect local air quality, focusing on power plant emissions and precipitation levels. While pollution is often studied at a global level, examining its contribution to climate change and warming temperatures, its affect on communities local to the emissions source is studied less frequently. This analysis considers other factors (precipitation) that may influence air quality and explores how air quality levels vary day-to-day. Texas was chosen as the study area for this analysis because of its number of power plants and counties. This report investigates the following three questions:

- 1) Is there a correlation between county level emission levels from power plants and county air quality in Texas?
- 2) Is there a trend in air quality values and emissions over time in Harris county, Texas?
- 3) Is there a relationship between precipitation and air quality in Harris county Texas?

Figure 2-13: U.S. Greenhouse Gas Emissions Allocated to Economic Sectors



Note: Emissions and removals from Land Use, Land-Use Change, and Forestry are excluded from figure above. Excludes U.S. Territories.

Dataset Information

This report analyzes three sources of data: EPA Air Quality, Emissions & Generation Resource Integrated Database (eGRID), and National Oceanic and Atmospheric Administration via National Centers for Environmental Information.

EPA Air Quality Data is collected at local monitors across the United States. It can be downloaded by pollutant type, region, and year. This report uses data for Texas between 2018 and 2023 for PM2.5 and Ozone. PM2.5 and Ozone are two criteria pollutants that can cause respiratory issues, especially for people with pre-existing health conditions, children, and seniors. The AQI based on PM2.5 and Ozone are recorded separately and range on a scale from 0-500, with higher values being more dangerous. Values 0-50 indicate "Good" AQ, 50 - 100 "Moderate", and increasingly poor from there. For more information about Air Quality Index and criteria pollutants, visit <https://www.epa.gov/outdoor-air-quality-data/air-data-basic-information>

eGRID data is released annually and contains information for all power plants across the United States. It contains information about generation, age of the plant, fuel type, and emissions, including CO2 Equivalent emissions. This report uses the eGRID 2018 - 2023 annual reports and focuses only on Texas power plants.

The National Oceanic and Atmospheric Administration provides public data via the National Centers for Environmental Information. The information pulled for this report included precipitation data in inches from 44 monitoring stations in Dallas county and 83 monitoring stations in Harris county. The dates spanned 2019 - 2023.

Exploratory Analysis

This report analyzes air quality data from 2019 - 2023 in Texas for both PM2.5 and Ozone. For each year, there are separate PM2.5 and Ozone files, but the structure is the same for all. Below is the structure of the 2019 Texas PM2.5 data.

```
## 'data.frame': 18138 obs. of 22 variables:
##   $ Date                  : chr "03/16/2019" "03/17/2019" "03/18/2019" "03/19/2019" ...
##   $ Source                : chr "AQS" "AQS" "AQS" "AQS" ...
##   $ Site.ID               : int 480271045 480271045 480271045 480271045 480271045 480271045 ...
##   $ POC                   : int 1 1 1 1 1 1 1 1 1 1 ...
##   $ Daily.Mean.PM2.5.Concentration: num 4.4 6.3 7 11.1 12.7 5.6 7.5 8.4 7.7 14.3 ...
##   $ Units                 : chr "ug/m3 LC" "ug/m3 LC" "ug/m3 LC" "ug/m3 LC" ...
##   $ Daily.AQI.Value       : int 24 35 39 55 58 31 42 47 43 61 ...
##   $ Local.Site.Name       : chr "Temple Georgia" "Temple Georgia" "Temple Georgia" "Temple Georgia" ...
##   $ Daily.Obs.Count       : int 1 1 1 1 1 1 1 1 1 1 ...
##   $ Percent.Complete      : num 100 100 100 100 100 100 100 100 100 100 ...
##   $ AQS.Parameter.Code    : int 88101 88101 88101 88101 88101 88101 88101 88101 88101 88101 ...
##   $ AQS.Parameter.Description: chr "PM2.5 - Local Conditions" "PM2.5 - Local Conditions" "PM2.5 - Local Conditions" ...
##   $ Method.Code            : int 209 209 209 209 209 209 209 209 209 209 ...
##   $ Method.Description     : chr "Met One BAM-1022 Mass Monitor w/ VSCC or TE-PM2.5C" "Met One BAM-1022 Mass Monitor w/ VSCC or TE-PM2.5C" ...
##   $ CBSA.Code              : int 28660 28660 28660 28660 28660 28660 28660 28660 28660 28660 ...
##   $ CBSA.Name              : chr "Killeen-Temple, TX" "Killeen-Temple, TX" "Killeen-Temple, TX" ...
##   $ State.FIPS.Code        : int 48 48 48 48 48 48 48 48 48 48 ...
##   $ State                  : chr "Texas" "Texas" "Texas" "Texas" ...
##   $ County.FIPS.Code       : int 27 27 27 27 27 27 27 27 27 27 ...
##   $ County                 : chr "Bell" "Bell" "Bell" "Bell" ...
##   $ Site.Latitude          : num 31.1 31.1 31.1 31.1 31.1 ...
##   $ Site.Longitude         : num -97.4 -97.4 -97.4 -97.4 -97.4 ...
```

Analysis

Question 1: Is there a correlation between county level emission levels from power plants and county air quality?

The eGRID data for each year was adjusted to ensure all numeric columns were imported as numeric for quantitative analysis. The year, state, plant name, county, latitude, longitude, and emissions data for all types of recorded emissions (including CO₂ equivalent emissions) were selected to a new data frame. These data frames were combined and further manipulated to have a data frame with the total county emissions for each year.

Relevant columns were selected for each of the PM2.5 and Ozone data frames, including date AQI, site name, AQ type (PM2.5 or Ozone), State, County, longitude and latitude. This was combined into two dataframes, one for PM2.5 and one for Ozone and further manipulated to have data frames with the average air quality value for each county each year.

The data frames were joined to have a data frame with, for each year and each county, total emissions and average AQI.

How often is emissions data reported?

The following map displays all the plants referenced in at least one of the eGRID reports between 2019 and 2023. Not all plants listed in eGRID report CO2 equivalence. The plants that reported at some point in the 5 years are in blue. Plants that are listed in a least one of the eGRID reports but never report CO2 emissions (or have zero emissions) are in black.

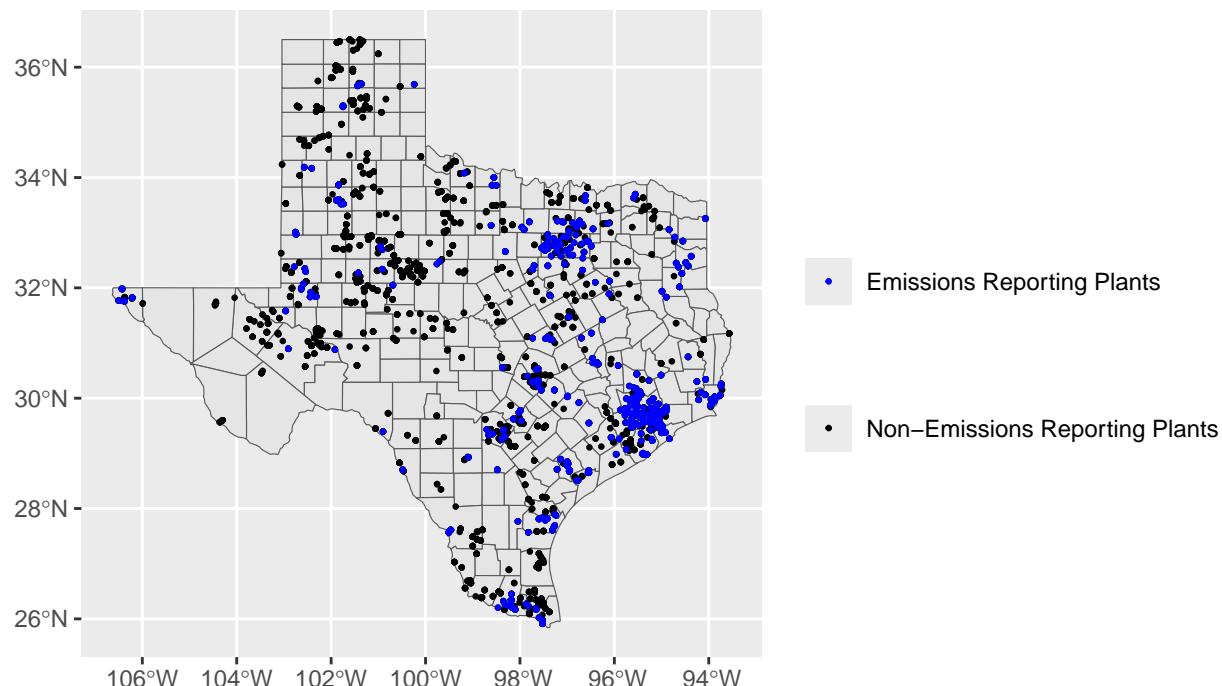


Figure 1: Listed and Emissions Reporting Texas Power Plants. 2019 - 2023 eGRID reports.

How are air quality stations spatially located compared to emissions reporting power plants?

The following map shows how plants and EPA air quality monitors are spatially related. Not all counties have both an air quality monitor and at least one CO₂ equivalence reporting plant. Only counties reporting both at some point between 2019 and 2023 are included in the statistical analysis below.

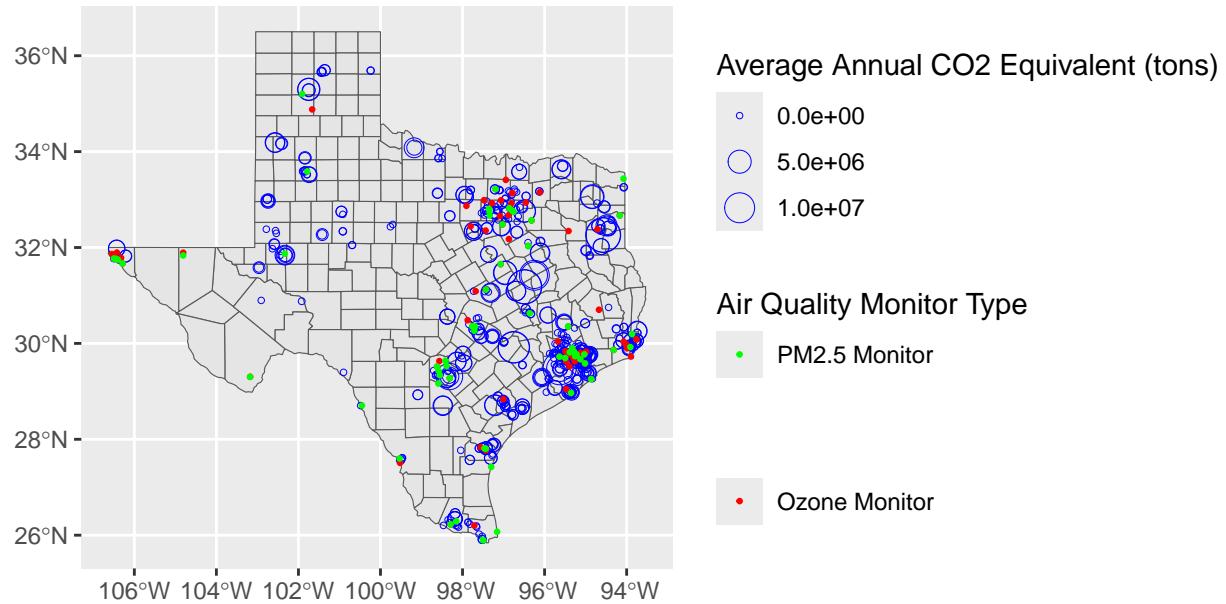


Figure 2: Air Quality Monitors & Annual Emissions from Power Plants and in Texas, 2019 - 2023.

Are annual county emissions correlated with annual county air quality?

The following graphs and tables show the relationship between total annual county emissions and average annual county air quality. A linear regression was run on the data to determine if the relationship has a significant correlation. Harris County data points are highlighted.

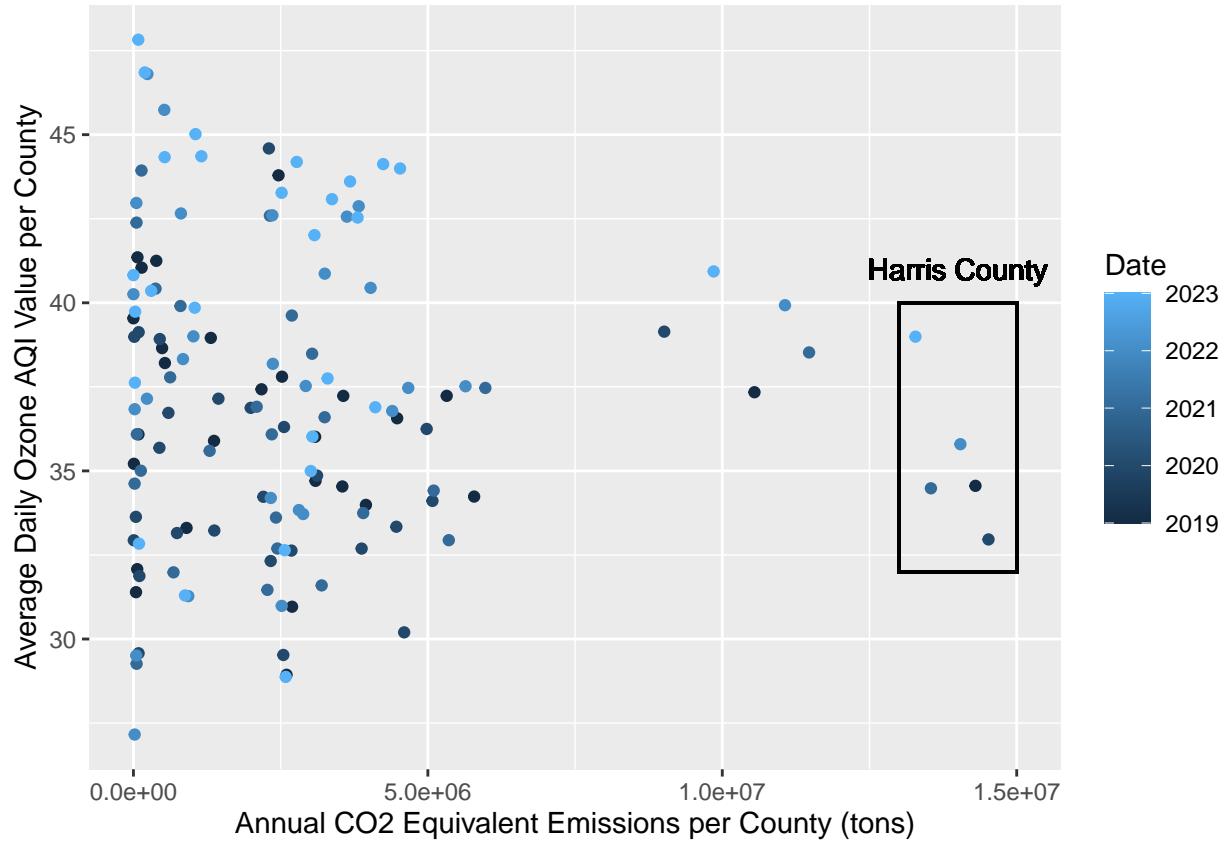


Figure 3: Average Daily Ozone AQI Value vs. Annual CO₂ Equivalent Emissions (tons). Per Texas County, 2019 - 2023

Table 1: Linear Regression of Emissions vs. Ozone Air Quality

Item	Value
P-value	0.5239
R-squared	-0.0043
Degrees of Freedom	136

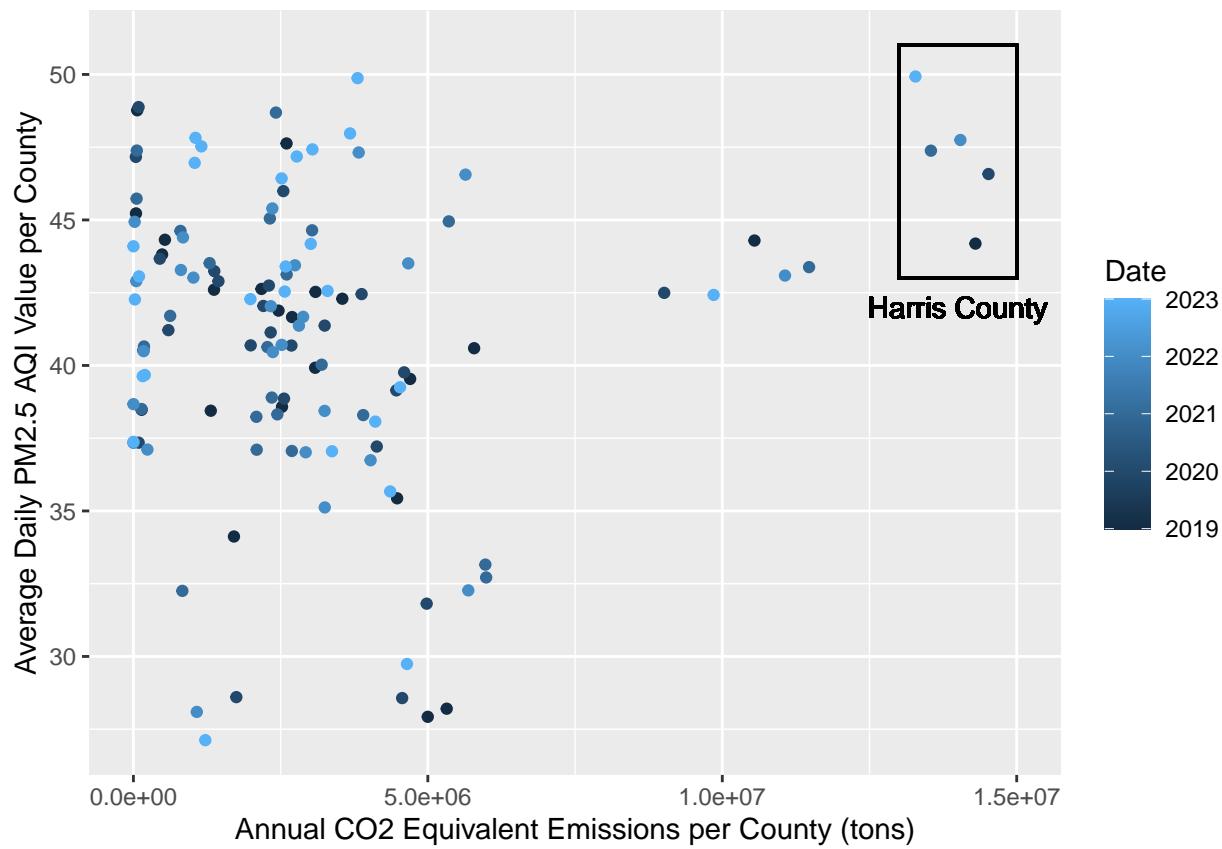


Figure 4: Average Daily PM2.5 AQI Value vs. Annual CO2 Equivalent Emissions (tons). Per Texas County, 2019 - 2023.

Table 2: Linear Regression of Emissions vs. PM2.5 Air Quality

Item	Value
P-value	0.4074
R-squared	-0.0025
Degrees of Freedom	125

Question 2: Is there a trend in air quality values and emissions over time in Harris county, Texas?

After analyzing the relationship between total CO₂ equivalent emissions and AQI in PM2.5 and Ozone, we found no statistical correlation between either. To further investigate from a different approach and look at temporal trends, we focused on one area, Harris county, due to its high number of both emissions and air quality monitor data. This is shown in the Figure 2 map as the concentrated area in the southeastern part of Texas. In addition to analyzing CO₂ equivalent emissions, methane emissions were analyzed. Unlike CO₂, SO₂, N₂O which are all summed to calculate the total CO₂ equivalent emissions, methane is a volatile organic compound (VOC) which can react with nitrogen oxides and sunlight to form tropospheric ozone. We wanted to test among these greenhouse gases, if methane specifically has a relationship with AQI.

Figure 5 following plot shows the PM2.5 AQI over time, in Harris County, Texas. There is an overall trend of increasing PM2.5 AQI over the five-year period.

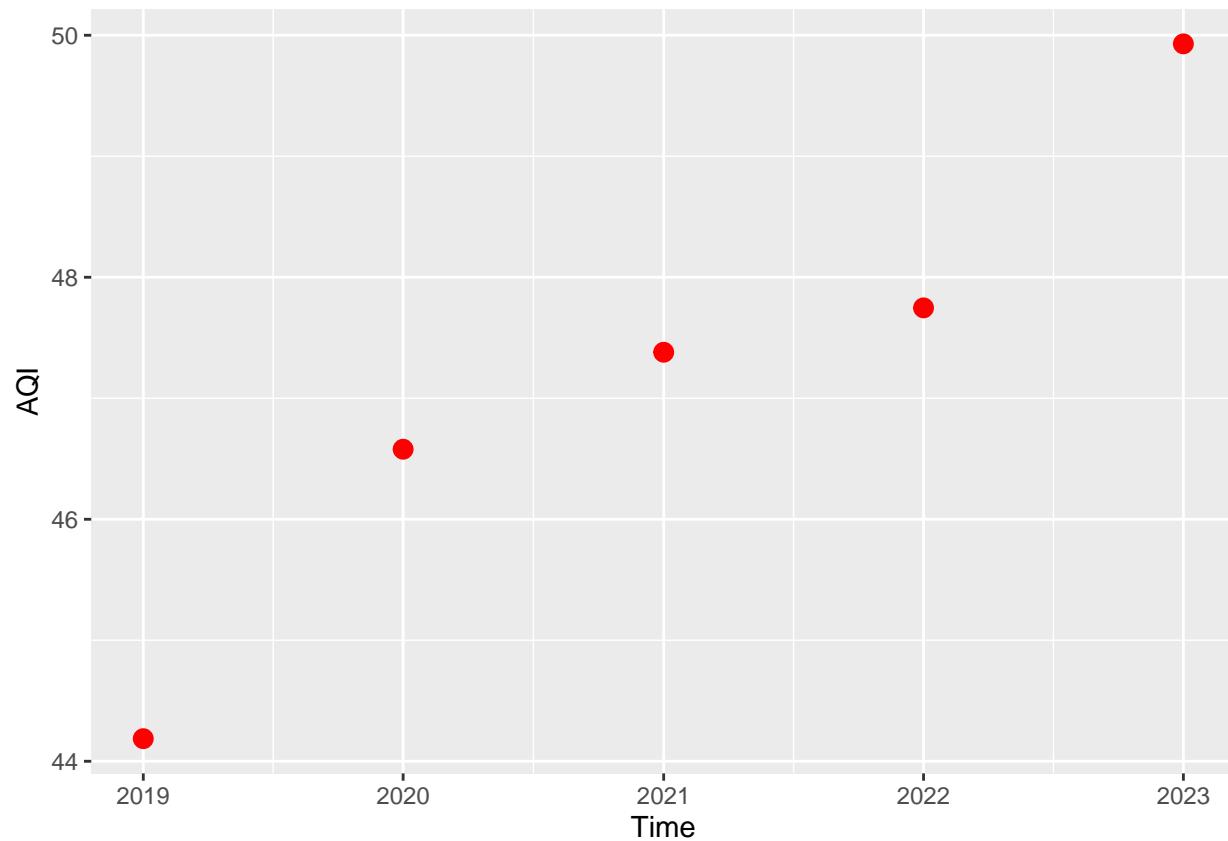


Figure 5: Average daily PM2.5 AQI in Harris County, Texas, 2019-2023

Figure 6 shows the ozone AQI over time in Harris County, Texas. There is an overall trend of increasing ozone AQI over the five-year period.

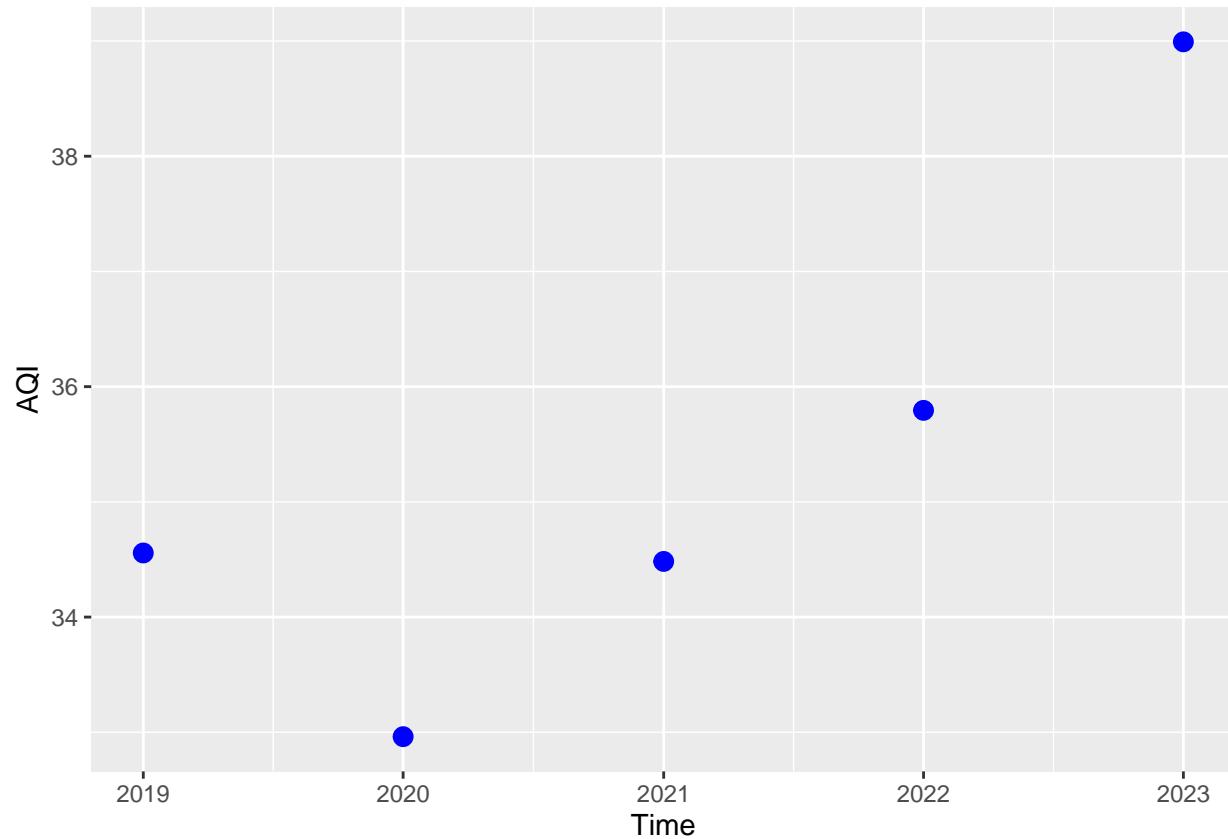


Figure 6: Average daily ozone AQI in Harris County, Texas, 2019-2023

Figure 7 shows the CO₂ equivalent emissions over time in Harris County, Texas. There is an overall trend of decreasing CO₂ equivalent emissions over the five-year period. The opposite trend of what occurred for PM_{2.5} and ozone AQI.

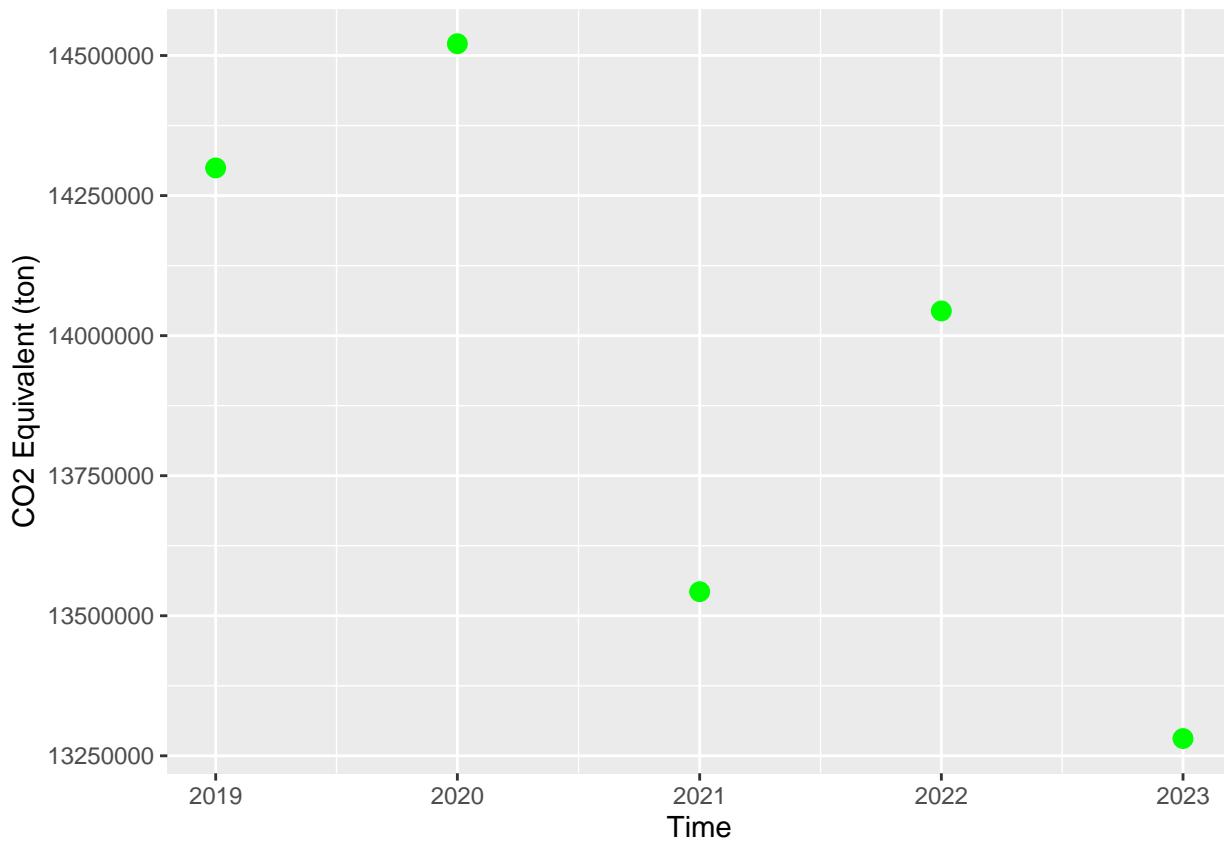


Figure 7: Annual CO₂ equivalent in Harris County, Texas, 2019-2023

Figure 8 shows the methane equivalent emissions over time in Harris County, Texas. There is an overall trend of decreasing methane emissions over the five-year period. The opposite trend of what occurred for PM2.5 and ozone AQI.

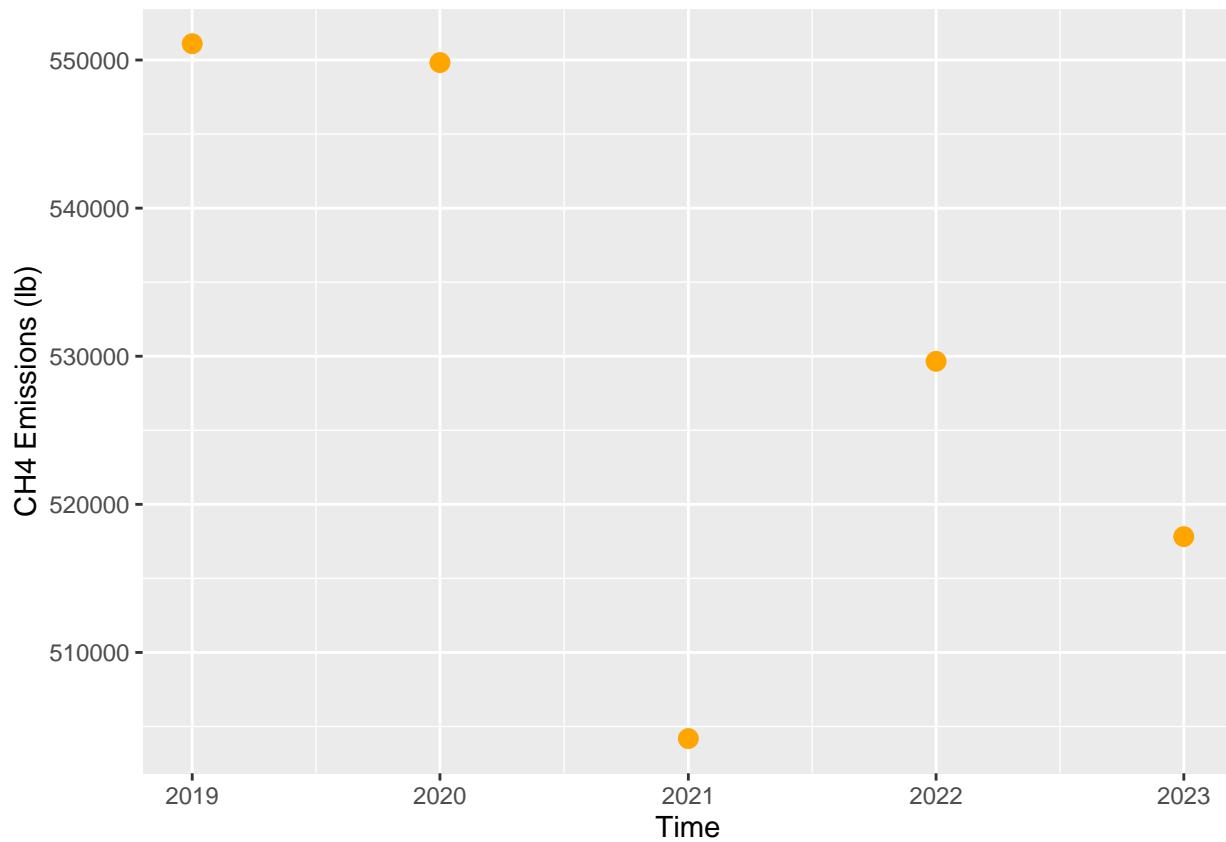


Figure 8: Annual methane emissions in Harris County, Texas, 2019-2023

Question 3: Is there a relationship between precipitation and air quality between and within the counties of Harris and Dallas in Texas?

Our third research question investigated the relationship between precipitation and air quality. Additionally, we looked spatially at two counties in Texas, Harris and Dallas, and provide comparative statistical analysis. From this data, we did not see a significant correlation between precipitation and AQI ozone nor precipitation and PM2.5 AQI. In comparing the two counties through two sample t-tests, no statistical difference in mean precipitation was seen at the 95% confidence level. However, mean Ozone and PM2.5 were seen to be different for the counties at the same confidence level.

To investigate this questions, data was wrangled for PM2.5 and Ozone to reflect the date and daily value for each year. It was then combined into a data set for the full date range (2019-2023). This was combined again with daily average precipitation data. This dataset was then filtered by county providing one data set with date, precipitation, ozone, and PM2.5 data for Harris and one data set for Dallas. These were combined again for the two sample t-tests.

Precipitation data was reported daily for most stations in the counties (83 in Harris and 44 in Dallas).

The graphs below provide a visual representation of the three considered environmental variables for Harris and Dallas County. A trend line is included to visualize general change over time.

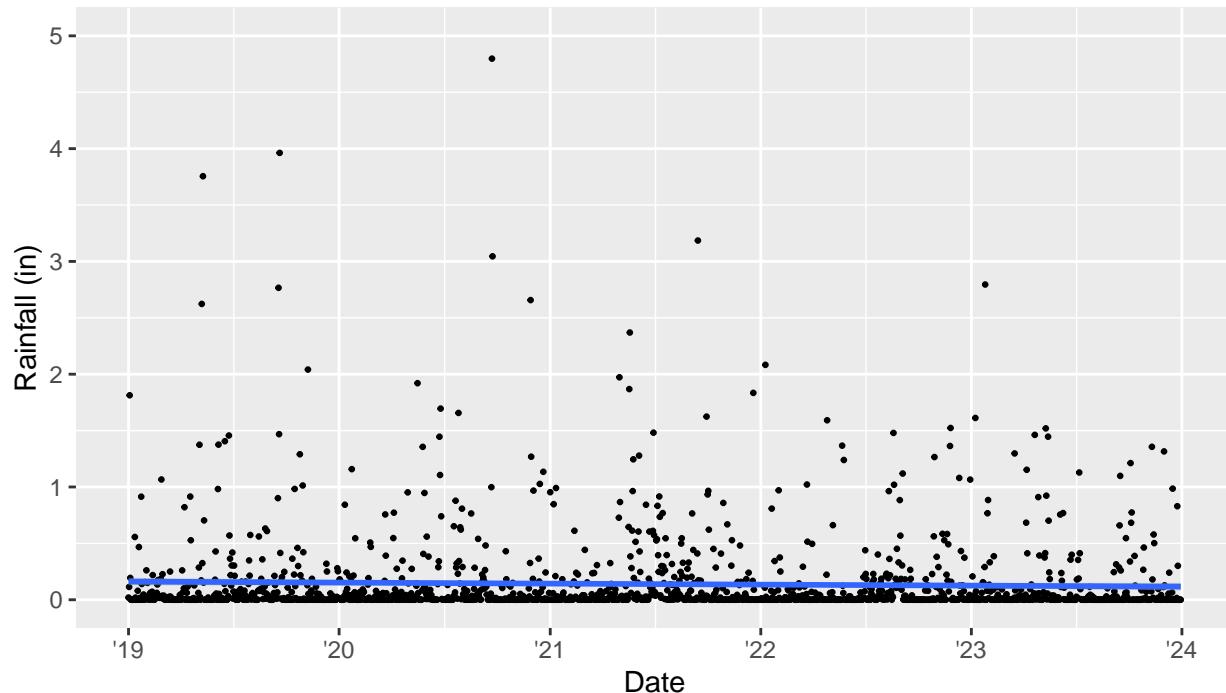


Figure 9: Precipitation Harris County, 2019 - 2023

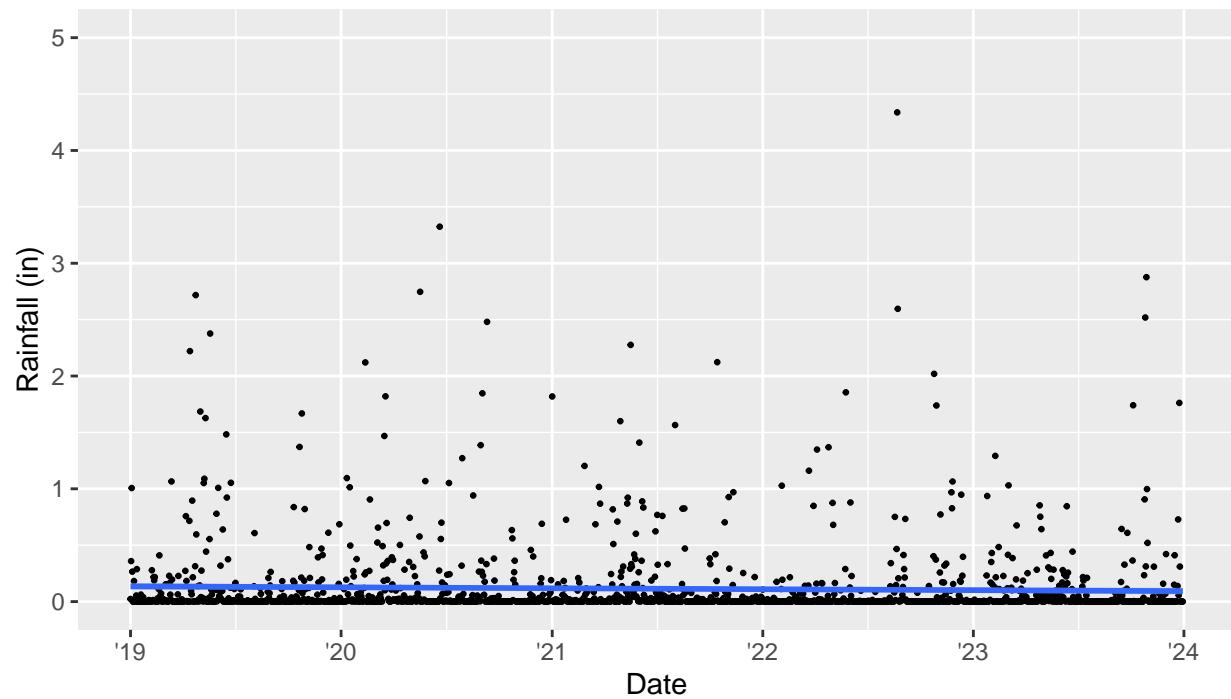


Figure 10: Precipitation Dallas County, 2019 - 2023

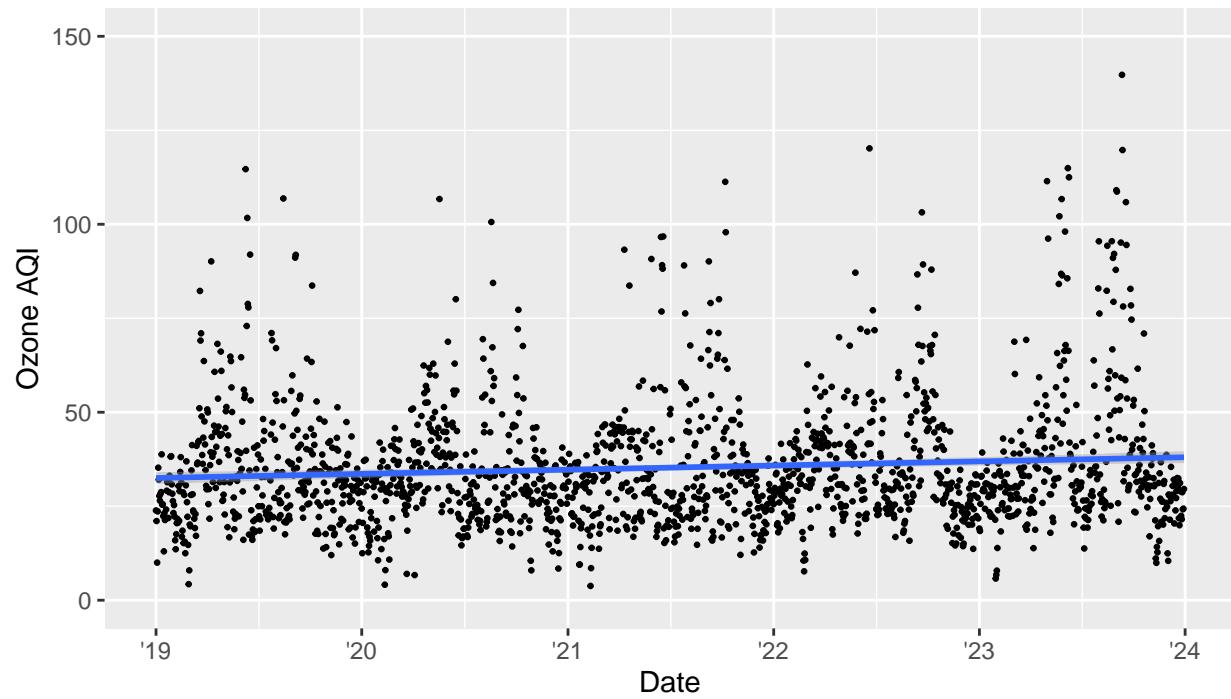


Figure 11: Ozone AQI Harris County, 2019 - 2023

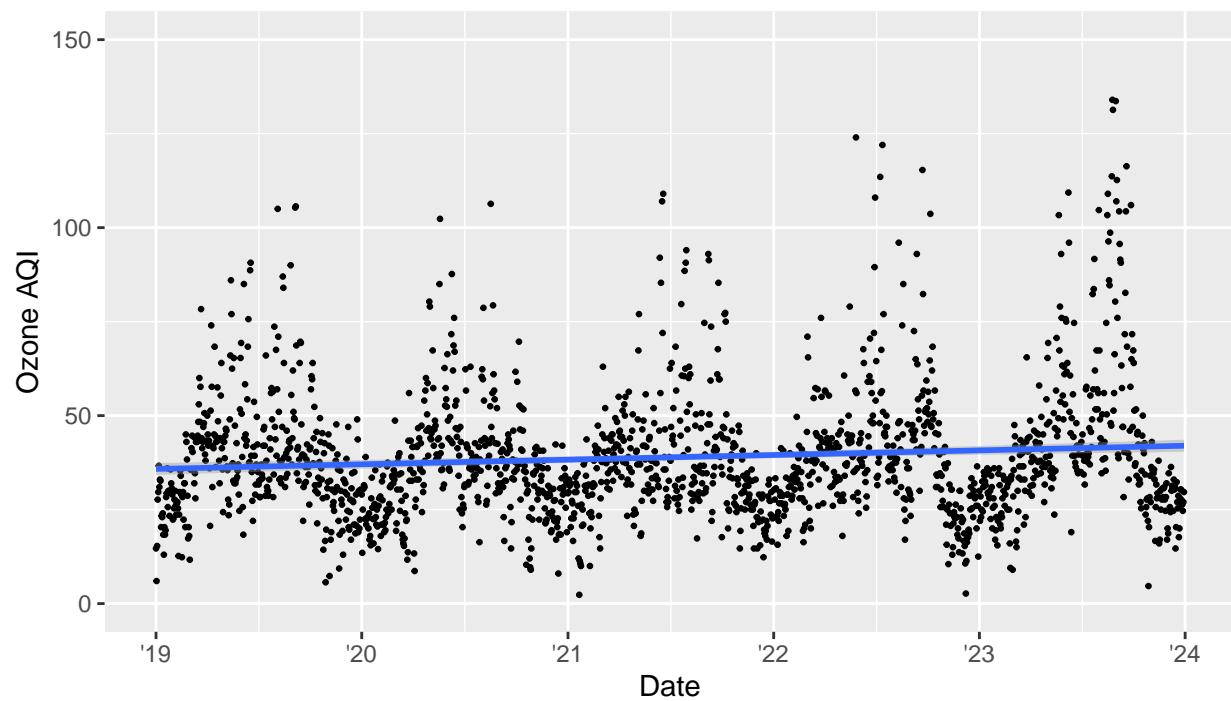


Figure 12: Ozone AQI Dallas County, 2019 - 2023

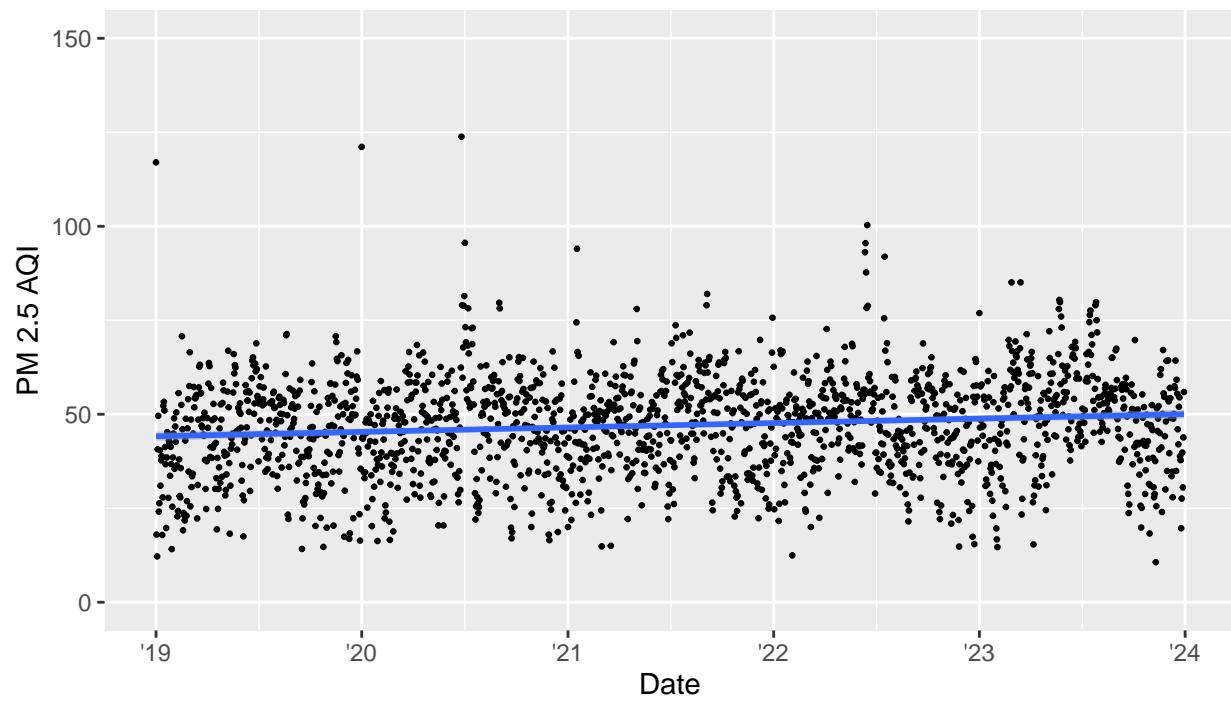


Figure 13: PM2.5 AQI Harris County, 2019 - 2023

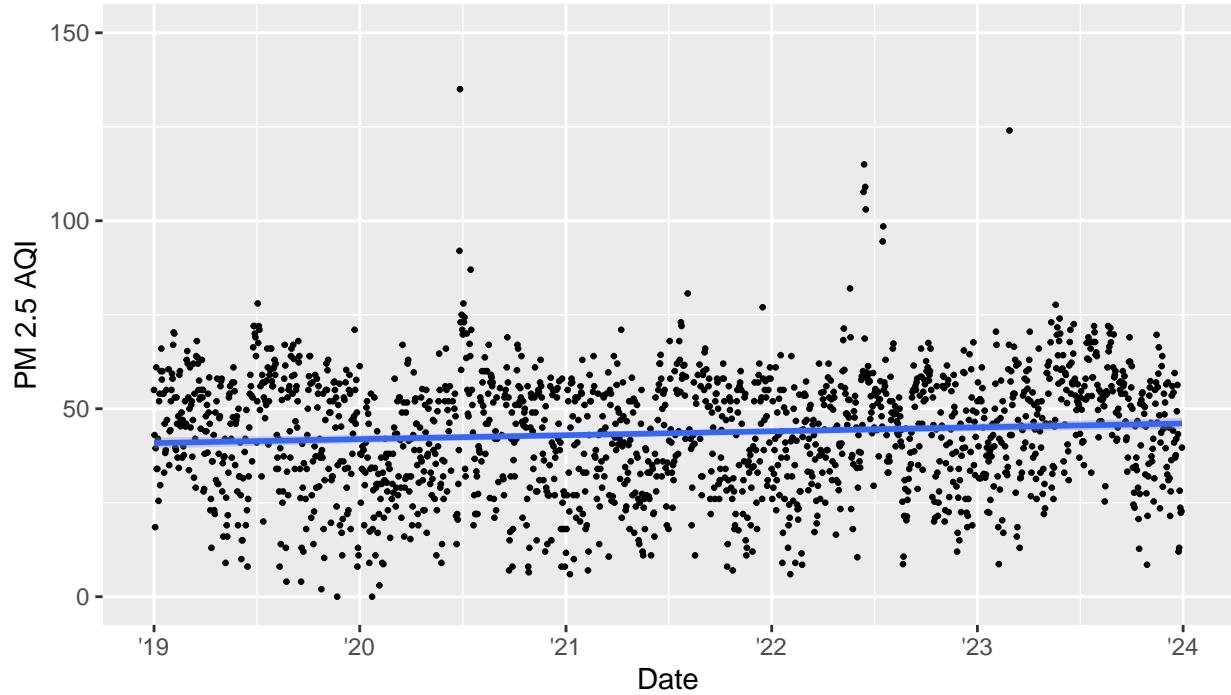


Figure 14: PM2.5 AQI Dallas County, 2019 - 2023

A separate linear regression was performed for Harris county and Dallas county for precipitation vs ozone AQI and precipitation vs PM2.5 AQI to assess precipitation as an explanatory variable for air quality.

Table 3: Linear Regression of Precipitation vs. Ozone Air Quality
Harris County

Item	Value
P-value	2.557e-07
R-squared	0.01396
Degrees of Freedom	1819

Table 4: Linear Regression of Precipitation vs. PM2.5 Air Quality
Harris County

Item	Value
P-value	2.2e-16
R-squared	0.06065
Degrees of Freedom	1819

Table 5: Linear Regression of Precipitation vs. Ozone Air Quality
 Dallas County

Item	Value
P-value	2.933e-07
R-squared	0.01382
Degrees of Freedom	1818

Table 6: Linear Regression of Precipitation vs. PM2.5 Air Quality
 Dallas County

Item	Value
P-value	2.2e-16
R-squared	0.0865
Degrees of Freedom	1778

The tables above show that each of the four regressions returned a low p-value and a low r-squared value. This means that precipitation is not a good predictor for ozone AQI nor PM2.5 AQI in these two counties and the investigated time span.

A two sample t-test was performed for Harris county and Dallas county to assess if the mean precipitation, ozone AQI, and PM2.5 AQI were the same or different for the two counties.

Null Hypothesis: μ_1 is equal to μ_2 Alternative Hypothesis μ_1 is not equal to μ_2

Table 7: Dallas vs Harris Two Sample T test Mean Precipitation

Item	Value
t-value	-2.1297
p-value	0.03326
Degrees of Freedom	3610

Table 8: Dallas vs Harris Two Sample T test Mean AQI Ozone

Item	Value
t-value	6.3494
p-value	2.429e-10
Degrees of Freedom	3632.9

Table 9: Dallas vs Harris Two Sample T test Mean AQI PM2.5

Item	Value
t-value	-7.3386
p-value	2.675e-13
Degrees of Freedom	3465.7

For all the variables (precipitation, ozone AQI, and PM2.5 AQI), the two sample t tests returned p-values less than 0.05. This supports the alternative hypothesis that there is a statistically significant difference in the means of these variables between the two counties.

Summary and Conclusions

The analysis of correlation between air quality and emissions at the county level per year did not result in significant results. There is not a significant correlation between annual county emission from power plants and average annual county AQI score from Ozone or PM2.5. Additionally, no county on average saw any poor or unsafe Air Quality Index levels. There are three potential factors influencing this lack of correlation. One is that air quality and emissions cannot be studied at localized scale. Likely emissions disperse out of the county it is based on local weather patterns. Conducting analysis of local air quality based on power plant emissions might be more accurate if wind patterns were taken into consideration to estimate where emissions may be concentrating. Another reason is that topography plays a role in where emissions and particulates accumulate. For example, pollution commonly gets trapped in valleys. Therefore, including topography into the analysis might elucidate more clear patterns. Lastly, there may not be a correlation because there are not enough data points in this analysis to see a trend. All plants do not report emissions and there are not air quality stations in all counties. To perform this analysis more accurately, data should be looked at country wide and there should be better standards for ensuring all active power plants report their emissions.

Ground level ozone is more likely to form in high heat and sunlight, when nitrogen oxides react with VOCs. This report mainly focused on total CO₂ equivalent emissions. The emissions that are summed include SO₂, CO₂, CH₄, and N₂O. Since all but methane are not VOCs, we thought it would be interesting to explore the relationship specifically between methane and ozone over time in one isolated location with abundant emissions and air quality data. As shown in Figure 6 and Figure 8, analysis of ozone AQI showed an increase from 2019-2023 in Harris county, Texas, while methane emissions decreased in that same time period. Our results did not indicate a positive relationship between methane and ozone.

In the case of precipitation, the data did not show this variable to be explanatory to the ozone or PM2.5 AQI at a statistically significant level for these counties and the time frame of interest. However, the mean precipitation, ozone AQI, and PM2.5 AQI were found to be different between Harris and Dallas at an alpha level of 0.05. This is surprising in some ways, as some pollutants can partition into water droplets as in rains and subsequently improve air quality. Additionally, rainfall can be associated with wind as a co-variable which may also be expected to improve air quality by diluting pollution through mixing. It would be interesting to investigate if these trends hold true on a larger analysis scope such as looking at more counties in Texas, looking at the state level nationally, or looking at a large time span. The difference in mean precipitation, ozone AQI, and PM2.5 AQI levels between the counties makes sense because Harris County is located on the gulf of Mexico while Dallas county is in northeast Texas further inland so it is expected to see differences in climate and air movement. This investigation does not provide insight as to whether location is a causative variable or if there is another reason for this difference, such as industrial production levels between the counties. This would be another interesting direction for future research.

References

<add references here if relevant, otherwise delete this section>