Emissions and AQI

Maeve

2025-12-03

```
#Read in each year of eGRID data selecting relevant columns and adjusting
#data types for each column.
eGRID23_PLNT <- read.csv(here("Data/egrid2023_data_rev2_PLNT.csv"),
                         na.strings = c("", " ", "0")) %>%
  slice(-1) %>%
  mutate(Data.Year=as.Date(paste0(Data.Year, "-01-01"))) %>%
  mutate(Plant.latitude=as.numeric(Plant.latitude)) %>%
  mutate(Plant.longitude=as.numeric(Plant.longitude)) %>%
  mutate(Plant.annual.NOx.emissions..tons.=as.numeric(
    Plant.annual.NOx.emissions..tons.)) %>%
  mutate(Plant.annual.SO2.emissions..tons.=as.numeric(
    Plant.annual.SO2.emissions..tons.)) %>%
  mutate(Plant.annual.CO2.emissions..tons.=as.numeric(
    Plant.annual.CO2.emissions..tons.)) %>%
  mutate(Plant.annual.CH4.emissions..lbs.=as.numeric(
    Plant.annual.CH4.emissions..lbs.)) %>%
  mutate(Plant.annual.N2O.emissions..lbs.=as.numeric(
    Plant.annual.N2O.emissions..lbs.)) %>%
  mutate(Plant.annual.CO2.equivalent.emissions..tons.=as.numeric(
    Plant.annual.CO2.equivalent.emissions..tons.)) %>%
  mutate(Plant.annual.Hg.emissions..lbs.=as.numeric(
    Plant.annual.Hg.emissions..lbs.)) %>%
  select("Data.Year", "Plant.state.abbreviation", "Plant.name",
         "Plant.county.name", "Plant.latitude", "Plant.longitude",
         "Plant.primary.fuel","Plant.annual.NOx.emissions..tons.",
         "Plant.annual.SO2.emissions..tons.", "Plant.annual.CO2.emissions..tons.",
         "Plant.annual.CH4.emissions..lbs.","Plant.annual.N2O.emissions..lbs.",
         "Plant.annual.CO2.equivalent.emissions..tons.",
         "Plant.annual.Hg.emissions..lbs.")

eGRID22_PLNT <- read.csv(here("Data/egrid2022_data_PLNT.csv"),
                         na.strings = c("", " ", "0")) %>%
  slice(-1) %>%
  mutate(Data.Year=as.Date(paste0(Data.Year, "-01-01"))) %>%
  mutate(Plant.latitude=as.numeric(Plant.latitude)) %>%
  mutate(Plant.longitude=as.numeric(Plant.longitude)) %>%
  mutate(Plant.annual.NOx.emissions..tons.=as.numeric(
    Plant.annual.NOx.emissions..tons.)) %>%
  mutate(Plant.annual.SO2.emissions..tons.=as.numeric(
    Plant.annual.SO2.emissions..tons.)) %>%
  mutate(Plant.annual.CO2.emissions..tons.=as.numeric(
    Plant.annual.CO2.emissions..tons.)) %>%
```

```
  mutate(Plant.annual.CH4.emissions..lbs.=as.numeric(
    Plant.annual.CH4.emissions..lbs.)) %>%
  mutate(Plant.annual.N2O.emissions..lbs.=as.numeric(
    Plant.annual.N2O.emissions..lbs.)) %>%
  mutate(Plant.annual.CO2.equivalent.emissions..tons.=as.numeric(
    Plant.annual.CO2.equivalent.emissions..tons.)) %>%
  mutate(Plant.annual.Hg.emissions..lbs.=as.numeric(
    Plant.annual.Hg.emissions..lbs.)) %>%
  select("Data.Year", "Plant.state.abbreviation", "Plant.name",
         "Plant.county.name", "Plant.latitude", "Plant.longitude",
         "Plant.primary.fuel","Plant.annual.NOx.emissions..tons.",
         "Plant.annual.SO2.emissions..tons.", "Plant.annual.CO2.emissions..tons.",
         "Plant.annual.CH4.emissions..lbs.","Plant.annual.N2O.emissions..lbs.",
         "Plant.annual.CO2.equivalent.emissions..tons.",
         "Plant.annual.Hg.emissions..lbs.")

eGRID21_PLNT <- read.csv(here("Data/eGRID2021_data_PLNT.csv"),
                         na.strings = c("", " ", "0")) %>%
  slice(-1) %>%
  mutate(Data.Year=as.Date(paste0(Data.Year, "-01-01"))) %>%
    mutate(Plant.latitude=as.numeric(Plant.latitude)) %>%
    mutate(Plant.longitude=as.numeric(Plant.longitude)) %>%
    mutate(Plant.annual.NOx.emissions..tons.=as.numeric(
      Plant.annual.NOx.emissions..tons.)) %>%
    mutate(Plant.annual.SO2.emissions..tons.=as.numeric(
      Plant.annual.SO2.emissions..tons.)) %>%
    mutate(Plant.annual.CO2.emissions..tons.=as.numeric(
      Plant.annual.CO2.emissions..tons.)) %>%
    mutate(Plant.annual.CH4.emissions..lbs.=as.numeric(
      Plant.annual.CH4.emissions..lbs.)) %>%
    mutate(Plant.annual.N2O.emissions..lbs.=as.numeric(
      Plant.annual.N2O.emissions..lbs.)) %>%
    mutate(Plant.annual.CO2.equivalent.emissions..tons.=as.numeric(
      Plant.annual.CO2.equivalent.emissions..tons.)) %>%
    mutate(Plant.annual.Hg.emissions..lbs.=as.numeric(
      Plant.annual.Hg.emissions..lbs.)) %>%
  select("Data.Year", "Plant.state.abbreviation", "Plant.name",
         "Plant.county.name", "Plant.latitude", "Plant.longitude",
         "Plant.primary.fuel","Plant.annual.NOx.emissions..tons.",
         "Plant.annual.SO2.emissions..tons.", "Plant.annual.CO2.emissions..tons.",
         "Plant.annual.CH4.emissions..lbs.","Plant.annual.N2O.emissions..lbs.",
         "Plant.annual.CO2.equivalent.emissions..tons.",
         "Plant.annual.Hg.emissions..lbs.")


eGRID20_PLNT <- read.csv(here("Data/eGRID2020_Data_v2_PLNT.csv"),
                         na.strings = c("", " ", "0")) %>%
  slice(-1) %>%
  mutate(Data.Year=as.Date(paste0(Data.Year, "-01-01"))) %>%
  mutate(Plant.latitude=as.numeric(Plant.latitude)) %>%
  mutate(Plant.longitude=as.numeric(Plant.longitude)) %>%
  mutate(Plant.annual.NOx.emissions..tons.=as.numeric(
    Plant.annual.NOx.emissions..tons.)) %>%
```

```
  mutate(Plant.annual.SO2.emissions..tons.=as.numeric(
    Plant.annual.SO2.emissions..tons.)) %>%
  mutate(Plant.annual.CO2.emissions..tons.=as.numeric(
    Plant.annual.CO2.emissions..tons.)) %>%
  mutate(Plant.annual.CH4.emissions..lbs.=as.numeric(
    Plant.annual.CH4.emissions..lbs.)) %>%
  mutate(Plant.annual.N2O.emissions..lbs.=as.numeric(
    Plant.annual.N2O.emissions..lbs.)) %>%
  mutate(Plant.annual.CO2.equivalent.emissions..tons.=as.numeric(
    Plant.annual.CO2.equivalent.emissions..tons.)) %>%
  mutate(Plant.annual.Hg.emissions..lbs.=as.numeric(
    Plant.annual.Hg.emissions..lbs.)) %>%
  select("Data.Year", "Plant.state.abbreviation", "Plant.name",
         "Plant.county.name", "Plant.latitude", "Plant.longitude",
         "Plant.primary.fuel","Plant.annual.NOx.emissions..tons.",
         "Plant.annual.SO2.emissions..tons.", "Plant.annual.CO2.emissions..tons.",
         "Plant.annual.CH4.emissions..lbs.","Plant.annual.N2O.emissions..lbs.",
         "Plant.annual.CO2.equivalent.emissions..tons.",
         "Plant.annual.Hg.emissions..lbs.")

eGRID19_PLNT <- read.csv(here("Data/egrid2019_data_PLNT.csv"), na.strings = c("", " ", "0")) %>%
  slice(-1) %>%
  mutate(Data.Year=as.Date(paste0(Data.Year, "-01-01"))) %>%
  mutate(Plant.latitude=as.numeric(Plant.latitude)) %>%
  mutate(Plant.longitude=as.numeric(Plant.longitude)) %>%
  mutate(Plant.annual.NOx.emissions..tons.=as.numeric(
    Plant.annual.NOx.emissions..tons.)) %>%
  mutate(Plant.annual.SO2.emissions..tons.=as.numeric(
    Plant.annual.SO2.emissions..tons.)) %>%
  mutate(Plant.annual.CO2.emissions..tons.=as.numeric(
    Plant.annual.CO2.emissions..tons.)) %>%
  mutate(Plant.annual.CH4.emissions..lbs.=as.numeric(
    Plant.annual.CH4.emissions..lbs.)) %>%
  mutate(Plant.annual.N2O.emissions..lbs.=as.numeric(
    Plant.annual.N2O.emissions..lbs.)) %>%
  mutate(Plant.annual.CO2.equivalent.emissions..tons.=as.numeric(
    Plant.annual.CO2.equivalent.emissions..tons.)) %>%
  mutate(Plant.annual.Hg.emissions..lbs.=as.numeric(
    Plant.annual.Hg.emissions..lbs.)) %>%
  select("Data.Year", "Plant.state.abbreviation", "Plant.name",
         "Plant.county.name", "Plant.latitude", "Plant.longitude",
         "Plant.primary.fuel","Plant.annual.NOx.emissions..tons.",
         "Plant.annual.SO2.emissions..tons.", "Plant.annual.CO2.emissions..tons.",
         "Plant.annual.CH4.emissions..lbs.","Plant.annual.N2O.emissions..lbs.",
         "Plant.annual.CO2.equivalent.emissions..tons.",
         "Plant.annual.Hg.emissions..lbs.")

#Combine all eGRID years into one dataframe
eGRID_PLNT <- rbind (eGRID23_PLNT, eGRID22_PLNT, eGRID21_PLNT, eGRID20_PLNT,
                     eGRID19_PLNT)

#Texas plants with reported CO2 emissions
eGRID_TX_indiv_CO2eq <- eGRID_PLNT %>%
```

```
  filter(Plant.state.abbreviation == "TX") %>%
  select("Plant.name", "Plant.county.name", "Plant.latitude", "Plant.longitude","Plant.annual.CO2.equiva
  na.omit()

#All Texas plants, location only
eGRID_TX_indiv_location_only <- eGRID_PLNT %>%
  filter(Plant.state.abbreviation == "TX") %>%
  group_by(Plant.latitude, Plant.longitude) %>%
  select("Plant.latitude", "Plant.longitude") %>%
  na.omit()

#Locations of Texas plants that have reported emissions
eGRID_TX_locations <- eGRID_TX_indiv_CO2eq %>%
  group_by(Plant.latitude, Plant.longitude) %>%
  summarize(AVG.CO2 = mean(Plant.annual.CO2.equivalent.emissions..tons.)) %>%
  na.omit()
```

```
## `summarise()` has grouped output by 'Plant.latitude'. You can override using
## the `.groups` argument.
```

```
#Total emissions per county per year in Texas
eGRID_PLNT_TX <- eGRID_PLNT %>%
  group_by(Data.Year,Plant.state.abbreviation,Plant.county.name) %>%
  summarize(CO2_Equiv = sum(Plant.annual.CO2.equivalent.emissions..tons.)) %>%
  na.omit() %>%
  filter(Plant.state.abbreviation == "TX" ) %>%
  rename(Date = Data.Year) %>%
  rename(County = Plant.county.name)
```

```
## `summarise()` has grouped output by 'Data.Year', 'Plant.state.abbreviation'.
## You can override using the `.groups` argument.
```

```
#Texas PM2.5 for each year. Formatted into yearly rather than daily data
TXPM2523 <- read.csv(here("Data/TXPM2523.csv")) %>%
  mutate(Date=mdy(Date)) %>%
  select("Date", "Daily.AQI.Value", "Local.Site.Name", "AQS.Parameter.Description", "State", "County",
  mutate(Year = year(Date)) %>%
  mutate(Date=as.Date(paste0(Year, "-01-01")))

TXPM2522 <- read.csv(here("Data/TXPM2522.csv")) %>%
  mutate(Date=mdy(Date)) %>%
  select("Date", "Daily.AQI.Value", "Local.Site.Name", "AQS.Parameter.Description", "State", "County",
  mutate(Year = year(Date)) %>%
  mutate(Date=as.Date(paste0(Year, "-01-01")))

TXPM2521 <- read.csv(here("Data/TXPM2521.csv")) %>%
  mutate(Date=mdy(Date)) %>%
  select("Date", "Daily.AQI.Value", "Local.Site.Name", "AQS.Parameter.Description", "State", "County",
  mutate(Year = year(Date)) %>%
  mutate(Date=as.Date(paste0(Year, "-01-01")))

TXPM2520 <- read.csv(here("Data/TXPM2520.csv")) %>%
```

```
  mutate(Date=mdy(Date)) %>%
  select("Date", "Daily.AQI.Value", "Local.Site.Name", "AQS.Parameter.Description", "State", "County",
  mutate(Year = year(Date)) %>%
  mutate(Date=as.Date(paste0(Year, "-01-01")))

TXPM2519 <- read.csv(here("Data/TXPM2519.csv")) %>%
  mutate(Date=mdy(Date)) %>%
  select("Date", "Daily.AQI.Value", "Local.Site.Name", "AQS.Parameter.Description", "State", "County",
  mutate(Year = year(Date)) %>%
  mutate(Date=as.Date(paste0(Year, "-01-01")))

#Average annual PM2.5 in Texas by County
TXPM25 <- rbind(TXPM2519, TXPM2520, TXPM2521, TXPM2522, TXPM2523)
TXPM25_CNTY <- TXPM25 %>%
  group_by(Date, County) %>%
  summarize(Avg_AQI_PM25 = mean(Daily.AQI.Value))
```

```
## `summarise()` has grouped output by 'Date'. You can override using the
## `.groups` argument.
```

```
#Texas Ozone for each year. Formatted into yearly rather than daily data
TXOzone23 <- read.csv(here("Data/TXOzone23.csv")) %>%
  mutate(Date=mdy(Date)) %>%
  select("Date", "Daily.AQI.Value", "Local.Site.Name", "AQS.Parameter.Description", "State", "County",
  mutate(Year = year(Date)) %>%
  mutate(Date=as.Date(paste0(Year, "-01-01")))

TXOzone22 <- read.csv(here("Data/TXOzone22.csv")) %>%
  mutate(Date=mdy(Date)) %>%
  select("Date", "Daily.AQI.Value", "Local.Site.Name", "AQS.Parameter.Description", "State", "County",
  mutate(Year = year(Date)) %>%
  mutate(Date=as.Date(paste0(Year, "-01-01")))

TXOzone21 <- read.csv(here("Data/TXOzone21.csv")) %>%
  mutate(Date=mdy(Date)) %>%
  select("Date", "Daily.AQI.Value", "Local.Site.Name", "AQS.Parameter.Description", "State", "County",
  mutate(Year = year(Date)) %>%
  mutate(Date=as.Date(paste0(Year, "-01-01")))

TXOzone20 <- read.csv(here("Data/TXOzone20.csv")) %>%
  mutate(Date=mdy(Date)) %>%
  select("Date", "Daily.AQI.Value", "Local.Site.Name", "AQS.Parameter.Description", "State", "County",
  mutate(Year = year(Date)) %>%
  mutate(Date=as.Date(paste0(Year, "-01-01")))

TXOzone19 <- read.csv(here("Data/TXOzone19.csv")) %>%
  mutate(Date=mdy(Date)) %>%
  select("Date", "Daily.AQI.Value", "Local.Site.Name", "AQS.Parameter.Description", "State", "County",
  mutate(Year = year(Date)) %>%
  mutate(Date=as.Date(paste0(Year, "-01-01")))

#Average annual Ozone in Texas by County
TXOzone <- rbind(TXOzone19, TXOzone20, TXOzone21, TXOzone22, TXOzone23)
```

```r
TXOzone_CNTY <- TXOzone %>%
  group_by(Date, County) %>%
  summarize(Avg_AQI_Ozone = mean(Daily.AQI.Value))
```

```
## `summarise()` has grouped output by 'Date'. You can override using the
## `.groups` argument.
```

```r
#Annual average AQI (ozone and PM2.5) and total emissions per county
TXEmissions_PM25 <- left_join(TXPM25_CNTY, eGRID_PLNT_TX,
                              by = c("Date", "County")) %>%
  na.omit()

TXEmissions_Ozone <- left_join(TXOzone_CNTY, eGRID_PLNT_TX,
                              by = c("Date", "County")) %>%
  na.omit()
```

```r
#Map of Texas counties
Texas_Counties <- st_read(here('Data/cb_2018_us_county_20m.shp')) %>%
  filter(STATEFP == 48)
```

```
## Reading layer `cb_2018_us_county_20m' from data source
##   `/home/guest/BassettGRHowey/Data/cb_2018_us_county_20m.shp'
##   using driver `ESRI Shapefile'
## Simple feature collection with 3220 features and 9 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -179.1743 ymin: 17.91377 xmax: 179.7739 ymax: 71.35256
## Geodetic CRS:  NAD83
```

```r
#Locations of all Texas plants (reporting and not)
ALLPlant_Locations <- st_as_sf(eGRID_TX_indiv_location_only,
    coords = c('Plant.longitude','Plant.latitude'),
    crs = 4326)

#Locations of reporting plants only
Plant_Locations <- st_as_sf(eGRID_TX_locations,
    coords = c('Plant.longitude','Plant.latitude'),
    crs = 4326)

#Locations of Ozone monitors
TXOzone_locations <- TXOzone %>%
  group_by(Site.Latitude, Site.Longitude) %>%
  summarize(AVGAQI.O3 = mean(Daily.AQI.Value))
```

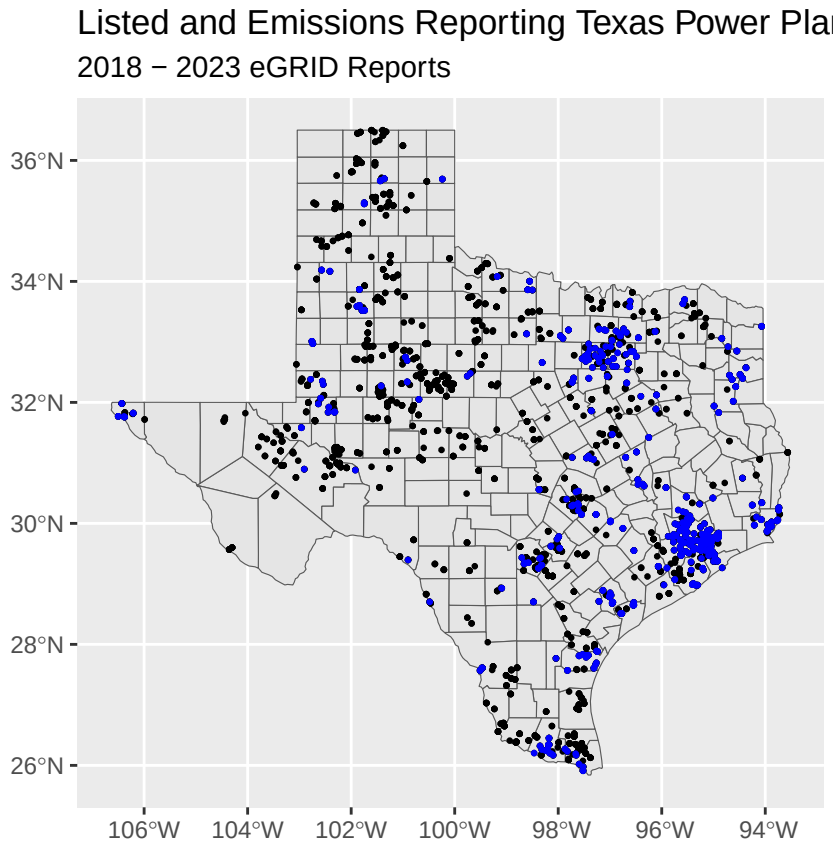
```
## `summarise()` has grouped output by 'Site.Latitude'. You can override using the
## `.groups` argument.
```

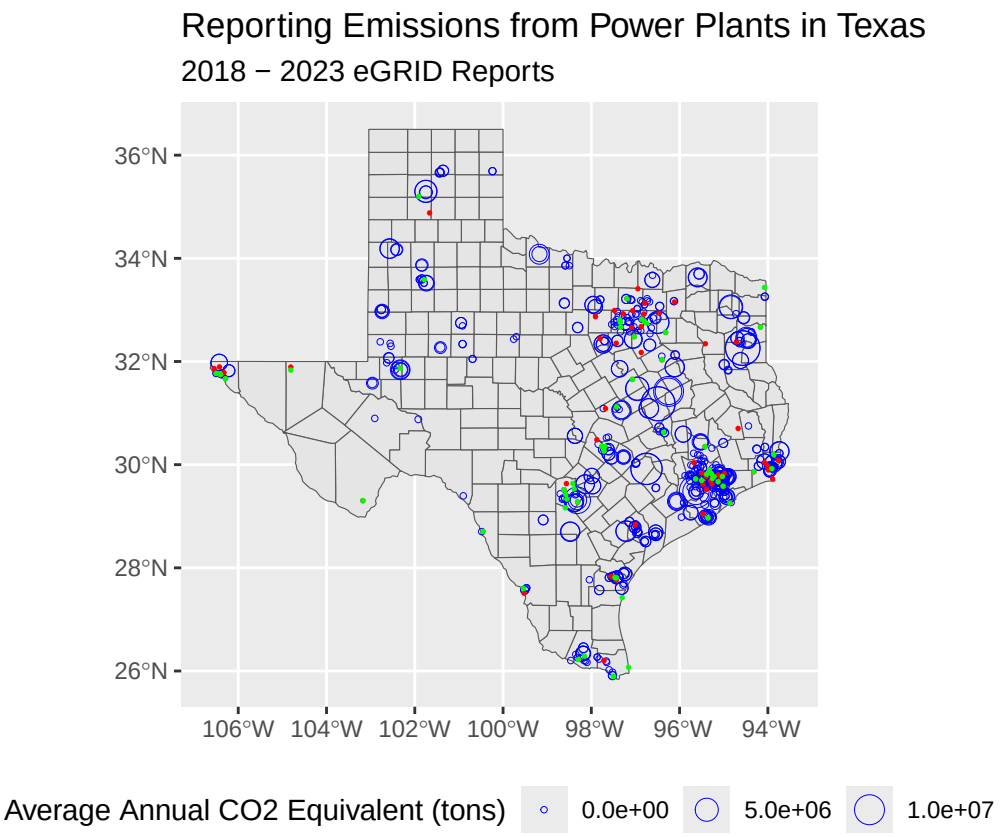
```r
Ozone_Locations <- st_as_sf(TXOzone_locations,
    coords = c('Site.Longitude','Site.Latitude'),
    crs = 4326)
```

```
#Locations of PM2.5 monitors
TXPM25_locations <- TXPM25 %>%
  group_by(Site.Latitude, Site.Longitude) %>%
  summarize(AVGAQI.PM25 = mean(Daily.AQI.Value))
```

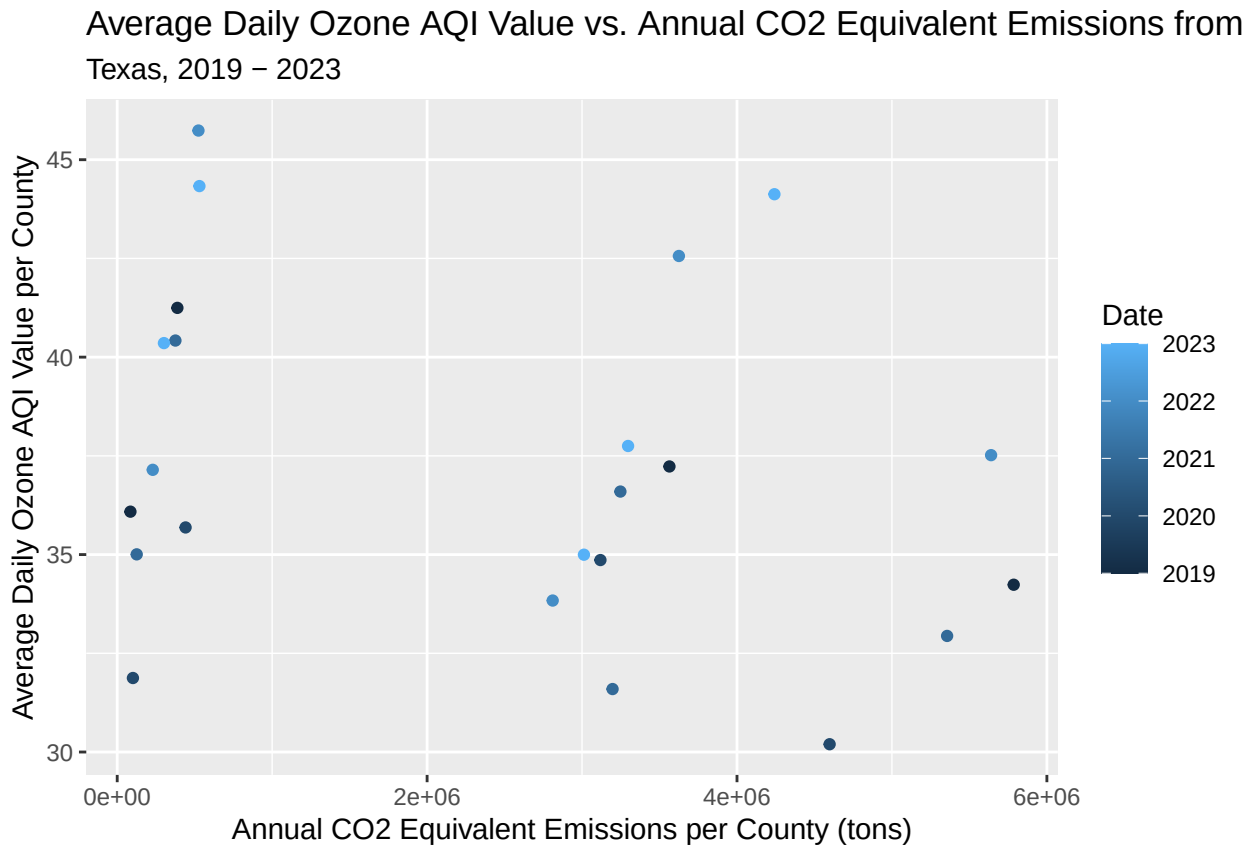
```
## `summarise()` has grouped output by 'Site.Latitude'. You can override using the
## `.groups` argument.
```

```
PM25_Locations <- st_as_sf(TXPM25_locations,
    coords = c('Site.Longitude','Site.Latitude'),
    crs = 4326)
```





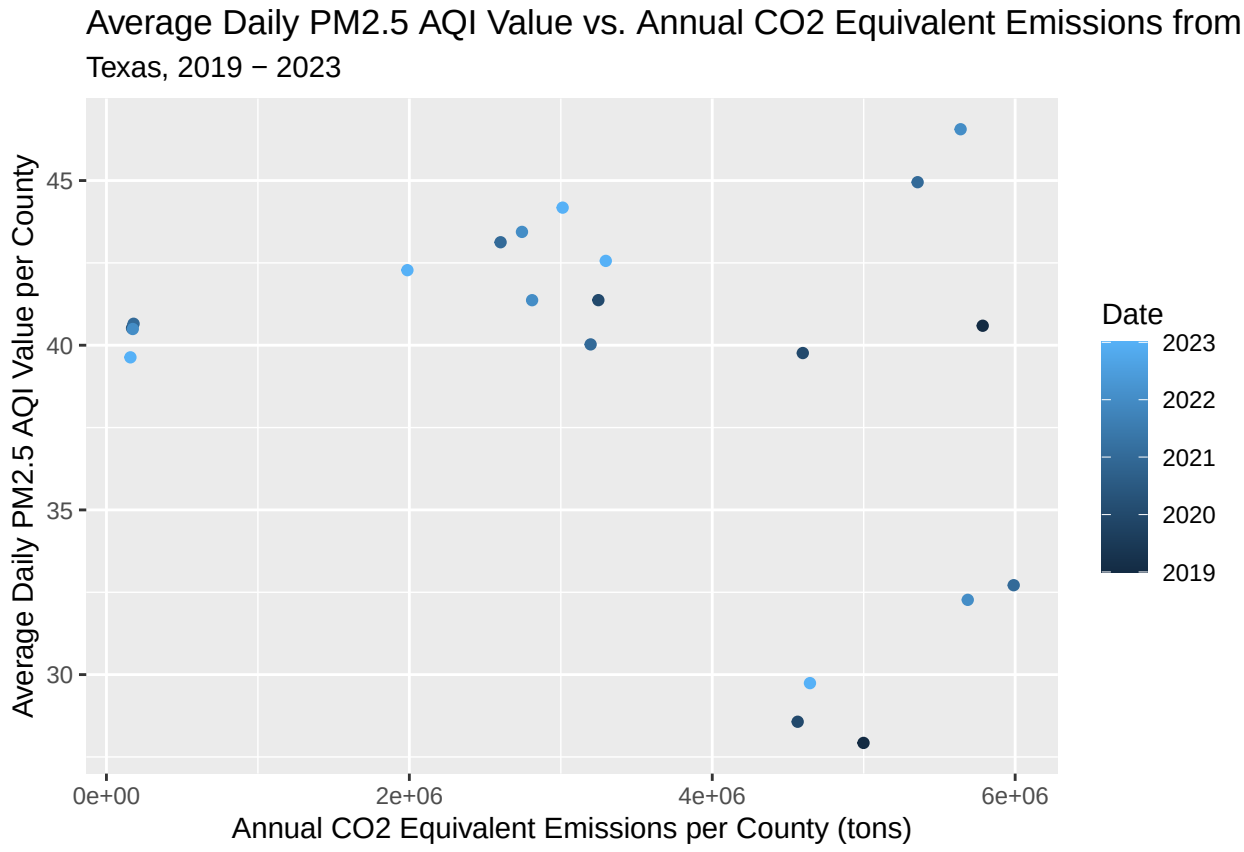

```
#Graph of emissions vs ozone
Emissions_O3 <- TXEmissions_Ozone %>%
  ggplot(aes(x = CO2_Equiv, y = Avg_AQI_Ozone))+
  geom_point(aes(color = Date))+
  xlab("Annual CO2 Equivalent Emissions per County (tons)")+
  ylab("Average Daily Ozone AQI Value per County")+
  ggtitle("Average Daily Ozone AQI Value vs. Annual CO2 Equivalent Emissions from Reporting Power Plants
Emissions_O3
```




```
#Linear regression of emissions vs ozone
O3byEmissionsRegression <- lm(
  TXEmissions_Ozone$Avg_AQI_Ozone ~ TXEmissions_Ozone$CO2_Equiv)
summary(O3byEmissionsRegression)
```

```
##
## Call:
## lm(formula = TXEmissions_Ozone$Avg_AQI_Ozone ~ TXEmissions_Ozone$CO2_Equiv)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7444 -2.6331 -0.9136  2.1236  8.0329
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   3.868e+01  1.373e+00   28.17   <2e-16 ***
## TXEmissions_Ozone$CO2_Equiv -6.102e-07  4.455e-07   -1.37    0.185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.201 on 21 degrees of freedom
## Multiple R-squared:  0.082,  Adjusted R-squared:  0.03828
## F-statistic: 1.876 on 1 and 21 DF,  p-value: 0.1853
```

```r
#Graph of emissions vs PM2.5
Emissions_PM25<-TXEmissions_PM25 %>%
  ggplot(aes(x = CO2_Equiv, y = Avg_AQI_PM25))+
  geom_point(aes(color = Date))+
  xlab("Annual CO2 Equivalent Emissions per County (tons)")+
  ylab("Average Daily PM2.5 AQI Value per County")+
  ggtitle("Average Daily PM2.5 AQI Value vs. Annual CO2 Equivalent Emissions from Reporting Power Plants
Emissions_PM25
```




```r
#Linear regression of emissions vs PM2.5
PM25byEmissionsRegression <- lm(
  TXEmissions_PM25$Avg_AQI_PM25 ~ TXEmissions_PM25$CO2_Equiv)
summary(PM25byEmissionsRegression)
```

```
##
## Call:
## lm(formula = TXEmissions_PM25$Avg_AQI_PM25 ~ TXEmissions_PM25$CO2_Equiv)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.766 -2.483  1.675  3.314  9.451
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 4.226e+01  2.333e+00  18.114 1.91e-13 ***
```

```
## TXEmissions_PM25$CO2_Equiv -9.138e-07  5.998e-07  -1.524    0.144
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.322 on 19 degrees of freedom
## Multiple R-squared:  0.1089, Adjusted R-squared:  0.06196
## F-statistic: 2.321 on 1 and 19 DF,  p-value: 0.1441
```