hw1

AUTHOR
Andre Gala-Garza

Disclaimer: Generative AI was used to assist with templating and writing code in this assignment; however, this code was checked manually and edited by hand to ensure accuracy.

Source: OpenAI. (2026). *ChatGPT (GPT-5.2 Thinking)* [Large language model]. https://chatgpt.com/.

1. Exploratory data analysis

Define question

We will work with air pollution data from the U.S. Environmental Protection Agency (EPA). The EPA has a national monitoring network of air pollution sites that measure particulate matter (PM) concentrations. The primary question we will answer is whether daily concentrations of PM (particulate matter air pollution with aerodynamic diameter less than 2.5 μm) decreased in Michigan over the 20 years spanning from 2005 to 2025.

EDA Step 1: Download and read data

The 2005 and 2025 PM2.5 air quality data for all sites in Michigan were downloaded from the BIOSTAT 620 GitHub repository, as MI_05.csv and MI_25.csv.

```
pm_2005 <- read.csv("MI_05.csv")
pm_2025 <- read.csv("MI_25.csv")
```

EDA Step 2: Check data size

```
dim(pm_2005)
```

```
[1] 5280   22
```

```
dim(pm_2025)
```

```
[1] 9934   22
```

EDA Step 3: Examine variables and their types

```r
names(pm_2005)
```

```
 [1] "Date"                            "Source"                         
 [3] "Site.ID"                         "POC"                            
 [5] "Daily.Mean.PM2.5.Concentration" "Units"                          
 [7] "Daily.AQI.Value"                 "Local.Site.Name"                
 [9] "Daily.Obs.Count"                 "Percent.Complete"               
[11] "AQS.Parameter.Code"              "AQS.Parameter.Description"      
[13] "Method.Code"                     "Method.Description"             
[15] "CBSA.Code"                       "CBSA.Name"                      
[17] "State.FIPS.Code"                 "State"                          
[19] "County.FIPS.Code"                "County"                         
[21] "Site.Latitude"                   "Site.Longitude"                 
```

```r
names(pm_2025)
```


```
 [1] "Date"                            "Source"                         
 [3] "Site.ID"                         "POC"                            
 [5] "Daily.Mean.PM2.5.Concentration" "Units"                          
 [7] "Daily.AQI.Value"                 "Local.Site.Name"                
 [9] "Daily.Obs.Count"                 "Percent.Complete"               
[11] "AQS.Parameter.Code"              "AQS.Parameter.Description"      
[13] "Method.Code"                     "Method.Description"             
[15] "CBSA.Code"                       "CBSA.Name"                      
[17] "State.FIPS.Code"                 "State"                          
[19] "County.FIPS.Code"                "County"                         
[21] "Site.Latitude"                   "Site.Longitude"                 
```


```r
str(pm_2005)
```

```
'data.frame':	5280 obs. of  22 variables:
 $ Date                          : chr  "01/01/2005" "01/02/2005" "01/03/2005" "01/04/2005" ...
 $ Source                        : chr  "AQS" "AQS" "AQS" "AQS" ...
 $ Site.ID                       : int  260050003 260050003 260050003 260050003 260050003
260050003 260050003 260050003 260050003 260050003 ...
 $ POC                           : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Daily.Mean.PM2.5.Concentration: num  9.3 7 15.7 13.4 5.3 4 12.6 29 15.5 13.2 ...
 $ Units                         : chr  "ug/m3 LC" "ug/m3 LC" "ug/m3 LC" "ug/m3 LC" ...
 $ Daily.AQI.Value               : int  51 39 63 59 29 22 58 88 63 59 ...
 $ Local.Site.Name               : chr  "966 32nd St (Holland)" "966 32nd St (Holland)" "966 32nd
St (Holland)" "966 32nd St (Holland)" ...
 $ 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" "PM2.5 - Local Conditions" ...
 $ Method.Code                   : int  118 118 118 118 118 118 118 118 118 118 ...
 $ Method.Description            : chr  "R & P Model 2025 PM2.5 Sequential w/WINS" "R & P Model
```

```
2025 PM2.5 Sequential w/WINS" "R & P Model 2025 PM2.5 Sequential w/WINS" "R & P Model 2025 PM2.5
Sequential w/WINS" ...
 $ CBSA.Code                     : int  26090 26090 26090 26090 26090 26090 26090 26090 26090
26090 ...
 $ CBSA.Name                     : chr  "Holland, MI" "Holland, MI" "Holland, MI" "Holland, MI"
...
 $ State.FIPS.Code               : int  26 26 26 26 26 26 26 26 26 26 ...
 $ State                         : chr  "Michigan" "Michigan" "Michigan" "Michigan" ...
 $ County.FIPS.Code              : int  5 5 5 5 5 5 5 5 5 5 ...
 $ County                        : chr  "Allegan" "Allegan" "Allegan" "Allegan" ...
 $ Site.Latitude                 : num  42.8 42.8 42.8 42.8 42.8 ...
 $ Site.Longitude                : num  -86.1 -86.1 -86.1 -86.1 -86.1 ...
```

```r
str(pm_2025)
```

```
'data.frame':	9934 obs. of  22 variables:
 $ Date                          : chr  "01/01/2025" "01/02/2025" "01/03/2025" "01/04/2025" ...
 $ Source                        : chr  "AQS" "AQS" "AQS" "AQS" ...
 $ Site.ID                       : int  260050003 260050003 260050003 260050003 260050003
260050003 260050003 260050003 260050003 260050003 ...
 $ POC                           : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Daily.Mean.PM2.5.Concentration: num  3.7 4.5 3.8 3.3 4.6 3.3 3 2.8 7.7 17.7 ...
 $ Units                         : chr  "ug/m3 LC" "ug/m3 LC" "ug/m3 LC" "ug/m3 LC" ...
 $ Daily.AQI.Value               : int  21 25 21 18 26 18 17 16 43 67 ...
 $ Local.Site.Name               : chr  "966 32nd St (Holland)" "966 32nd St (Holland)" "966 32nd
St (Holland)" "966 32nd St (Holland)" ...
 $ 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" "PM2.5 - Local Conditions" ...
 $ Method.Code                   : int  170 170 170 170 170 170 170 170 170 170 ...
 $ Method.Description            : chr  "Met One BAM-1020 Mass Monitor w/VSCC" "Met One BAM-1020
Mass Monitor w/VSCC" "Met One BAM-1020 Mass Monitor w/VSCC" "Met One BAM-1020 Mass Monitor
w/VSCC" ...
 $ CBSA.Code                     : int  26090 26090 26090 26090 26090 26090 26090 26090 26090
26090 ...
 $ CBSA.Name                     : chr  "Holland, MI" "Holland, MI" "Holland, MI" "Holland, MI"
...
 $ State.FIPS.Code               : int  26 26 26 26 26 26 26 26 26 26 ...
 $ State                         : chr  "Michigan" "Michigan" "Michigan" "Michigan" ...
 $ County.FIPS.Code              : int  5 5 5 5 5 5 5 5 5 5 ...
 $ County                        : chr  "Allegan" "Allegan" "Allegan" "Allegan" ...
 $ Site.Latitude                 : num  42.8 42.8 42.8 42.8 42.8 ...
 $ Site.Longitude                : num  -86.1 -86.1 -86.1 -86.1 -86.1 ...
```

EDA Step 4: Look at top/bottom of data

```r
head(pm_2005)
```

```
        Date Source   Site.ID POC Daily.Mean.PM2.5.Concentration    Units
1 01/01/2005    AQS 260050003   1                            9.3 ug/m3 LC
2 01/02/2005    AQS 260050003   1                            7.0 ug/m3 LC
3 01/03/2005    AQS 260050003   1                           15.7 ug/m3 LC
4 01/04/2005    AQS 260050003   1                           13.4 ug/m3 LC
5 01/05/2005    AQS 260050003   1                            5.3 ug/m3 LC
6 01/06/2005    AQS 260050003   1                            4.0 ug/m3 LC
  Daily.AQI.Value       Local.Site.Name Daily.Obs.Count Percent.Complete
1              51 966 32nd St (Holland)               1              100
2              39 966 32nd St (Holland)               1              100
3              63 966 32nd St (Holland)               1              100
4              59 966 32nd St (Holland)               1              100
5              29 966 32nd St (Holland)               1              100
6              22 966 32nd St (Holland)               1              100
  AQS.Parameter.Code AQS.Parameter.Description Method.Code
1              88101  PM2.5 - Local Conditions         118
2              88101  PM2.5 - Local Conditions         118
3              88101  PM2.5 - Local Conditions         118
4              88101  PM2.5 - Local Conditions         118
5              88101  PM2.5 - Local Conditions         118
6              88101  PM2.5 - Local Conditions         118
                       Method.Description CBSA.Code   CBSA.Name
1 R & P Model 2025 PM2.5 Sequential w/WINS     26090 Holland, MI
2 R & P Model 2025 PM2.5 Sequential w/WINS     26090 Holland, MI
3 R & P Model 2025 PM2.5 Sequential w/WINS     26090 Holland, MI
4 R & P Model 2025 PM2.5 Sequential w/WINS     26090 Holland, MI
5 R & P Model 2025 PM2.5 Sequential w/WINS     26090 Holland, MI
6 R & P Model 2025 PM2.5 Sequential w/WINS     26090 Holland, MI
  State.FIPS.Code    State County.FIPS.Code  County Site.Latitude
1              26 Michigan                5 Allegan      42.76779
2              26 Michigan                5 Allegan      42.76779
3              26 Michigan                5 Allegan      42.76779
4              26 Michigan                5 Allegan      42.76779
5              26 Michigan                5 Allegan      42.76779
6              26 Michigan                5 Allegan      42.76779
  Site.Longitude
1      -86.14858
2      -86.14858
3      -86.14858
4      -86.14858
5      -86.14858
6      -86.14858
```

```r
tail(pm_2005)
```

```
           Date Source   Site.ID POC Daily.Mean.PM2.5.Concentration    Units
5275 09/01/2005    AQS 261639901   1                           12.8 ug/m3 LC
```

```
5276 09/07/2005    AQS 261639901   1                          30.4 ug/m3 LC
5277 09/13/2005    AQS 261639901   1                          46.5 ug/m3 LC
5278 09/25/2005    AQS 261639901   1                          24.4 ug/m3 LC
5279 10/01/2005    AQS 261639901   1                          24.0 ug/m3 LC
5280 10/07/2005    AQS 261639901   1                           2.9 ug/m3 LC
     Daily.AQI.Value Local.Site.Name Daily.Obs.Count Percent.Complete
5275              58                               1              100
5276              91                               1              100
5277             128                               1              100
5278              80                               1              100
5279              79                               1              100
5280              16                               1              100
     AQS.Parameter.Code             AQS.Parameter.Description Method.Code
5275              88502 Acceptable PM2.5 AQI & Speciation Mass         781
5276              88502 Acceptable PM2.5 AQI & Speciation Mass         781
5277              88502 Acceptable PM2.5 AQI & Speciation Mass         781
5278              88502 Acceptable PM2.5 AQI & Speciation Mass         781
5279              88502 Acceptable PM2.5 AQI & Speciation Mass         781
5280              88502 Acceptable PM2.5 AQI & Speciation Mass         781
                 Method.Description CBSA.Code                   CBSA.Name
5275 SINGLE-FILTR WINS 2.5UM IMPACT     19820 Detroit-Warren-Dearborn, MI
5276 SINGLE-FILTR WINS 2.5UM IMPACT     19820 Detroit-Warren-Dearborn, MI
5277 SINGLE-FILTR WINS 2.5UM IMPACT     19820 Detroit-Warren-Dearborn, MI
5278 SINGLE-FILTR WINS 2.5UM IMPACT     19820 Detroit-Warren-Dearborn, MI
5279 SINGLE-FILTR WINS 2.5UM IMPACT     19820 Detroit-Warren-Dearborn, MI
5280 SINGLE-FILTR WINS 2.5UM IMPACT     19820 Detroit-Warren-Dearborn, MI
     State.FIPS.Code    State County.FIPS.Code County Site.Latitude
5275              26 Michigan              163  Wayne      42.25004
5276              26 Michigan              163  Wayne      42.25004
5277              26 Michigan              163  Wayne      42.25004
5278              26 Michigan              163  Wayne      42.25004
5279              26 Michigan              163  Wayne      42.25004
5280              26 Michigan              163  Wayne      42.25004
     Site.Longitude
5275      -83.19993
5276      -83.19993
5277      -83.19993
5278      -83.19993
5279      -83.19993
5280      -83.19993
```

```r
head(pm_2025)
```

```
        Date Source   Site.ID POC Daily.Mean.PM2.5.Concentration    Units
1 01/01/2025    AQS 260050003   1                            3.7 ug/m3 LC
2 01/02/2025    AQS 260050003   1                            4.5 ug/m3 LC
3 01/03/2025    AQS 260050003   1                            3.8 ug/m3 LC
4 01/04/2025    AQS 260050003   1                            3.3 ug/m3 LC
5 01/05/2025    AQS 260050003   1                            4.6 ug/m3 LC
6 01/06/2025    AQS 260050003   1                            3.3 ug/m3 LC
```

```
  Daily.AQI.Value       Local.Site.Name Daily.Obs.Count Percent.Complete
1              21 966 32nd St (Holland)               1              100
2              25 966 32nd St (Holland)               1              100
3              21 966 32nd St (Holland)               1              100
4              18 966 32nd St (Holland)               1              100
5              26 966 32nd St (Holland)               1              100
6              18 966 32nd St (Holland)               1              100
  AQS.Parameter.Code AQS.Parameter.Description Method.Code
1              88101  PM2.5 - Local Conditions         170
2              88101  PM2.5 - Local Conditions         170
3              88101  PM2.5 - Local Conditions         170
4              88101  PM2.5 - Local Conditions         170
5              88101  PM2.5 - Local Conditions         170
6              88101  PM2.5 - Local Conditions         170
                  Method.Description CBSA.Code   CBSA.Name State.FIPS.Code
1 Met One BAM-1020 Mass Monitor w/VSCC     26090 Holland, MI              26
2 Met One BAM-1020 Mass Monitor w/VSCC     26090 Holland, MI              26
3 Met One BAM-1020 Mass Monitor w/VSCC     26090 Holland, MI              26
4 Met One BAM-1020 Mass Monitor w/VSCC     26090 Holland, MI              26
5 Met One BAM-1020 Mass Monitor w/VSCC     26090 Holland, MI              26
6 Met One BAM-1020 Mass Monitor w/VSCC     26090 Holland, MI              26
     State County.FIPS.Code  County Site.Latitude Site.Longitude
1 Michigan                5 Allegan      42.76779      -86.14858
2 Michigan                5 Allegan      42.76779      -86.14858
3 Michigan                5 Allegan      42.76779      -86.14858
4 Michigan                5 Allegan      42.76779      -86.14858
5 Michigan                5 Allegan      42.76779      -86.14858
6 Michigan                5 Allegan      42.76779      -86.14858
```

```r
tail(pm_2025)
```

```
           Date Source   Site.ID POC Daily.Mean.PM2.5.Concentration    Units
9929 12/26/2025 AirNow 261630100   1                            5.2 ug/m3 LC
9930 12/27/2025 AirNow 261630100   1                            5.4 ug/m3 LC
9931 12/28/2025 AirNow 261630100   1                           10.7 ug/m3 LC
9932 12/29/2025 AirNow 261630100   1                            2.4 ug/m3 LC
9933 12/30/2025 AirNow 261630100   1                            3.4 ug/m3 LC
9934 12/31/2025 AirNow 261630100   1                            6.7 ug/m3 LC
     Daily.AQI.Value      Local.Site.Name Daily.Obs.Count Percent.Complete
9929              29 Military Park (GHIB)               1              100
9930              30 Military Park (GHIB)               1              100
9931              54 Military Park (GHIB)               1              100
9932              13 Military Park (GHIB)               1              100
9933              19 Military Park (GHIB)               1              100
9934              37 Military Park (GHIB)               1              100
     AQS.Parameter.Code AQS.Parameter.Description Method.Code
9929              88101  PM2.5 - Local Conditions          NA
9930              88101  PM2.5 - Local Conditions          NA
9931              88101  PM2.5 - Local Conditions          NA
9932              88101  PM2.5 - Local Conditions          NA
```

```
9933              88101  PM2.5 - Local Conditions          NA
9934              88101  PM2.5 - Local Conditions          NA
     Method.Description CBSA.Code                   CBSA.Name State.FIPS.Code
9929                        19820 Detroit-Warren-Dearborn, MI              26
9930                        19820 Detroit-Warren-Dearborn, MI              26
9931                        19820 Detroit-Warren-Dearborn, MI              26
9932                        19820 Detroit-Warren-Dearborn, MI              26
9933                        19820 Detroit-Warren-Dearborn, MI              26
9934                        19820 Detroit-Warren-Dearborn, MI              26
        State County.FIPS.Code County Site.Latitude Site.Longitude
9929 Michigan              163  Wayne      42.31208      -83.10347
9930 Michigan              163  Wayne      42.31208      -83.10347
9931 Michigan              163  Wayne      42.31208      -83.10347
9932 Michigan              163  Wayne      42.31208      -83.10347
9933 Michigan              163  Wayne      42.31208      -83.10347
9934 Michigan              163  Wayne      42.31208      -83.10347
```

EDA Step 5: Visualize the distribution of PM_2.5

```
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(lubridate)
```

```
Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

```
df <- pm_2005

df2 <- df %>%
  mutate(
    Date = mdy(Date),
```

```
    PM_2_5 = as.numeric(Daily.Mean.PM2.5.Concentration)
  )
```

```
# Numeric summary
summary(df2$PM_2_5)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00    6.20   11.00   13.77   18.70   81.60 
```

```
quantile(df2$PM_2_5, probs = c(0, .01, .05, .10, .25, .50, .75, .90, .95, .99, 1), na.rm = TRUE)
```

```
  0%   1%   5%  10%  25%  50%  75%  90%  95%  99% 100% 
 0.0  1.5  2.7  3.7  6.2 11.0 18.7 27.2 32.5 52.3 81.6 
```

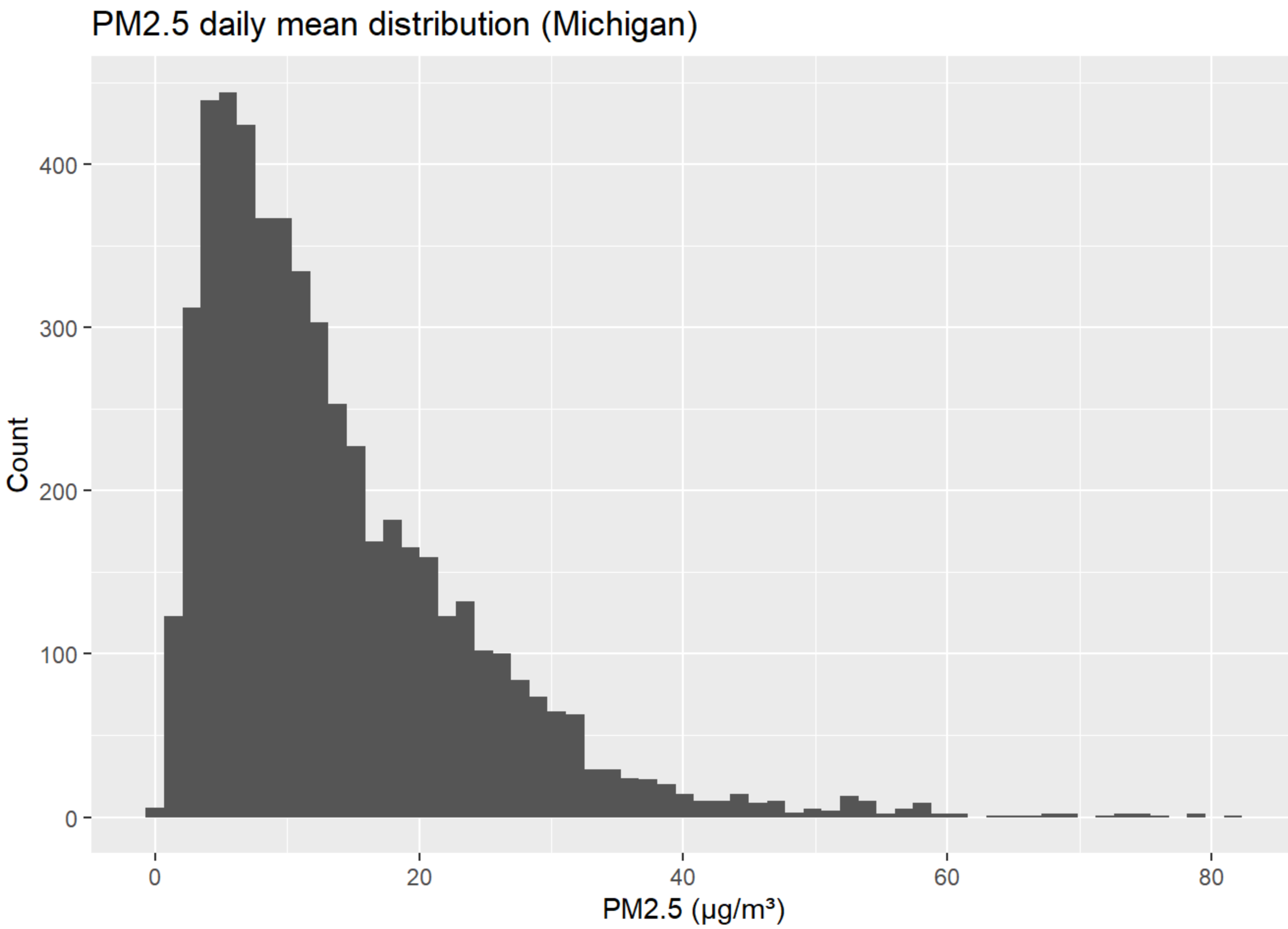
```
mean(df2$PM_2_5, na.rm = TRUE)
```

```
[1] 13.77182
```

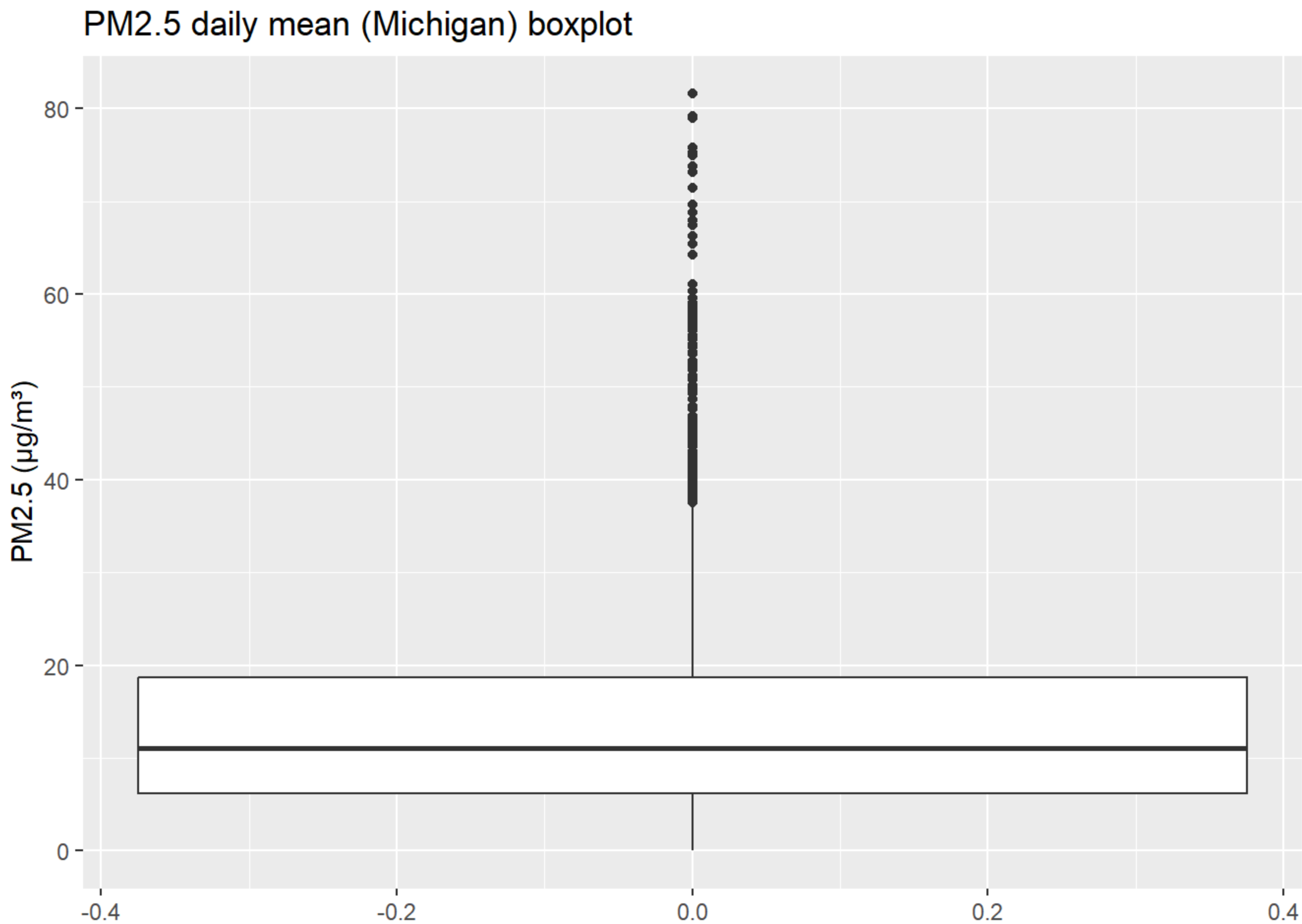
```
sd(df2$PM_2_5, na.rm = TRUE)
```

```
[1] 10.41329
```

```
ggplot(df2, aes(x = PM_2_5)) +
  geom_histogram(bins = 60) +
  labs(title = "PM2.5 daily mean distribution (Michigan)", x = "PM2.5 (µg/m³)", y = "Count")
```




```
# Boxplot
ggplot(df2, aes(y = PM_2_5)) +
  geom_boxplot() +
  labs(title = "PM2.5 daily mean (Michigan) boxplot", y = "PM2.5 (µg/m³)")
```



EDA Summary

After reading the data into R, we find that both the 2005 and 2025 PM2.5 data have 22 variables (columns), which have the same names between the datasets. The 2005 data have 5,280 records (rows), while the 2025 data have 9,934 records.

The variables in the data include important information such as state, county, latitude, longitude, date, units of measurement, and the most important variable: the PM2.5 measurement, recorded as "Daily.Mean.PM2.5.Concentration". Some variables encode the same element in different ways, such as "State.FIPS.Code" and "State", although they are both the "chr" data type in this case. Other variables like latitude, longitude, and "Daily.Mean.PM2.5.Concentration" have the "num" datatype to store floating-point numbers.

From examining the top and bottom of each dataset, many of the values are grouped into similar sets, so the values at the top and bottom are frequently the same. The exception is "Daily.Mean.PM2.5.Concentration", which always sees significant changes between records.

From observing the histogram of the distribution of PM2.5, we find that the distribution is right-skewed, with the mean located at roughly 8 µg/m^3. The frequency of PM2.5 concentrations decreases sharply after this point, with a very small number of records surpassing 50 µg/m^3. The boxplot of Michigan records confirms that any records above 40 µg/m^3 are considered outliers.

2. Combine data frames into one

```
library(dplyr)
library(lubridate)

clean_pm <- function(df) {
  df %>%
    mutate(
      Date = mdy(Date),
      year = as.integer(format(Date, "%Y")),
      pm25 = as.numeric(Daily.Mean.PM2.5.Concentration),
      state = State,
      county = County,
      site = Site.ID
    ) %>%
    select(state, county, site, year, Date, pm25,
           Site.Latitude, Site.Longitude, Method.Code, Method.Description, Units) %>%
    arrange(state, county, site, Date)
}

pm_all <- bind_rows(
  clean_pm(pm_2005),
  clean_pm(pm_2025)
)
```

```
dplyr::count(pm_all, year)
```

```
  year    n
1 2005 5280
2 2025 9934
```

```
summary(pm_all$pm25)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    0.0     4.5     7.4    10.2    12.5   116.0 
```

```
str(pm_all)
```

```
'data.frame':	15214 obs. of  11 variables:
 $ state             : chr  "Michigan" "Michigan" "Michigan" "Michigan" ...
 $ county            : chr  "Allegan" "Allegan" "Allegan" "Allegan" ...
 $ site              : int  260050003 260050003 260050003 260050003 260050003 260050003 260050003
260050003 260050003 260050003 ...
 $ year              : int  2005 2005 2005 2005 2005 2005 2005 2005 2005 2005 ...
 $ Date              : Date, format: "2005-01-01" "2005-01-02" ...
 $ pm25              : num  9.3 7 15.7 13.4 12.4 5.3 4 12.6 29 15.5 ...
 $ Site.Latitude     : num  42.8 42.8 42.8 42.8 42.8 ...
 $ Site.Longitude    : num  -86.1 -86.1 -86.1 -86.1 -86.1 ...
```

```
 $ Method.Code       : int  118 118 118 118 810 118 118 118 118 118 ...
 $ Method.Description: chr  "R & P Model 2025 PM2.5 Sequential w/WINS" "R & P Model 2025 PM2.5
Sequential w/WINS" "R & P Model 2025 PM2.5 Sequential w/WINS" "R & P Model 2025 PM2.5 Sequential
w/WINS" ...
 $ Units             : chr  "ug/m3 LC" "ug/m3 LC" "ug/m3 LC" "ug/m3 LC" ...
```

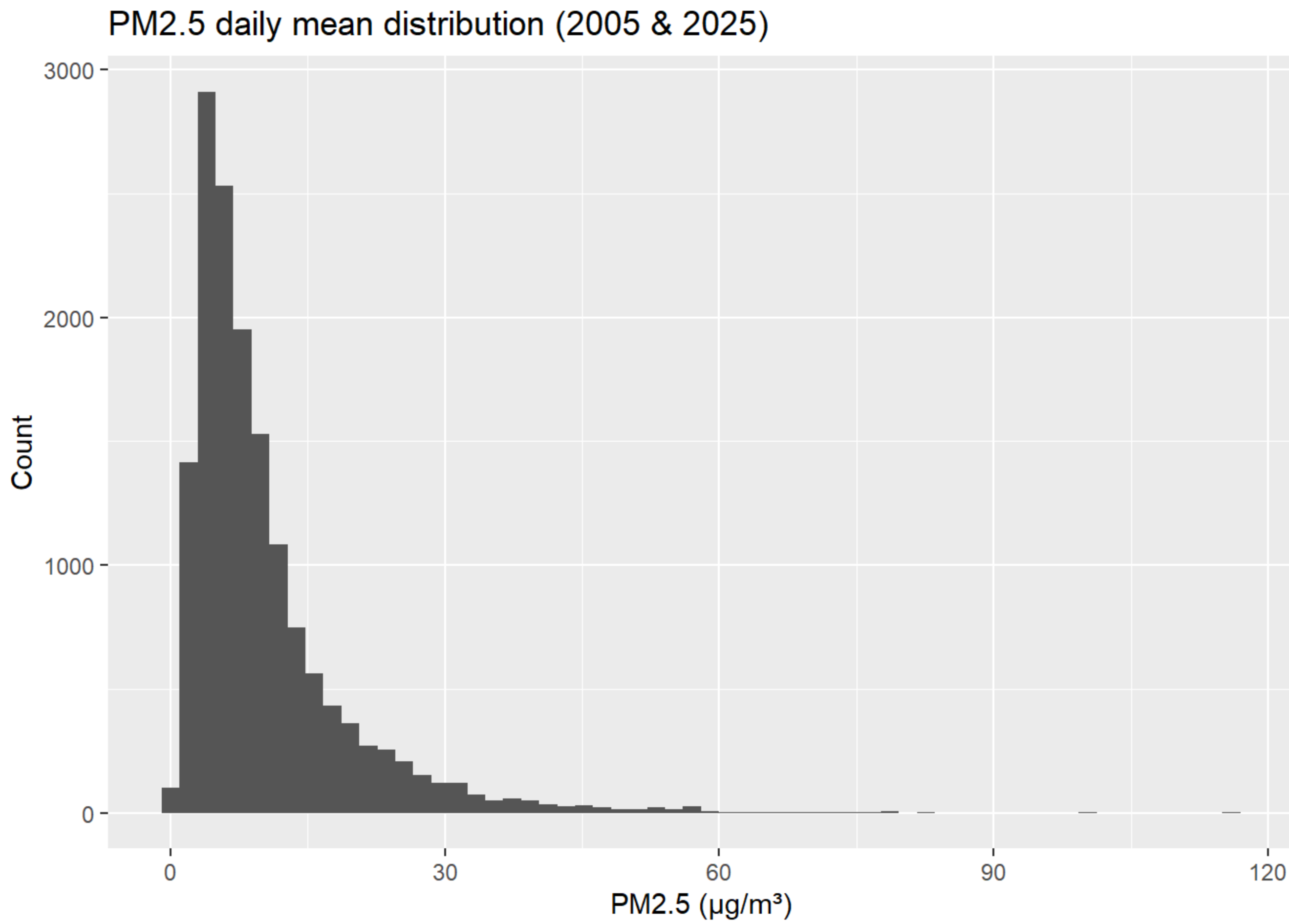
```r
head(pm_all)
```

```
     state  county      site year       Date pm25 Site.Latitude Site.Longitude
1 Michigan Allegan 260050003 2005 2005-01-01  9.3      42.76779      -86.14858
2 Michigan Allegan 260050003 2005 2005-01-02  7.0      42.76779      -86.14858
3 Michigan Allegan 260050003 2005 2005-01-03 15.7      42.76779      -86.14858
4 Michigan Allegan 260050003 2005 2005-01-04 13.4      42.76779      -86.14858
5 Michigan Allegan 260050003 2005 2005-01-04 12.4      42.76779      -86.14858
6 Michigan Allegan 260050003 2005 2005-01-05  5.3      42.76779      -86.14858
  Method.Code                       Method.Description    Units
1         118 R & P Model 2025 PM2.5 Sequential w/WINS ug/m3 LC
2         118 R & P Model 2025 PM2.5 Sequential w/WINS ug/m3 LC
3         118 R & P Model 2025 PM2.5 Sequential w/WINS ug/m3 LC
4         118 R & P Model 2025 PM2.5 Sequential w/WINS ug/m3 LC
5         810             Met One SASS/SuperSASS Teflon ug/m3 LC
6         118 R & P Model 2025 PM2.5 Sequential w/WINS ug/m3 LC
```

```r
tail(pm_all)
```

```
         state county      site year       Date pm25 Site.Latitude
15209 Michigan  Wayne 261630100 2025 2025-12-26  5.2      42.31208
15210 Michigan  Wayne 261630100 2025 2025-12-27  5.4      42.31208
15211 Michigan  Wayne 261630100 2025 2025-12-28 10.7      42.31208
15212 Michigan  Wayne 261630100 2025 2025-12-29  2.4      42.31208
15213 Michigan  Wayne 261630100 2025 2025-12-30  3.4      42.31208
15214 Michigan  Wayne 261630100 2025 2025-12-31  6.7      42.31208
      Site.Longitude Method.Code Method.Description    Units
15209      -83.10347          NA                    ug/m3 LC
15210      -83.10347          NA                    ug/m3 LC
15211      -83.10347          NA                    ug/m3 LC
15212      -83.10347          NA                    ug/m3 LC
15213      -83.10347          NA                    ug/m3 LC
15214      -83.10347          NA                    ug/m3 LC
```

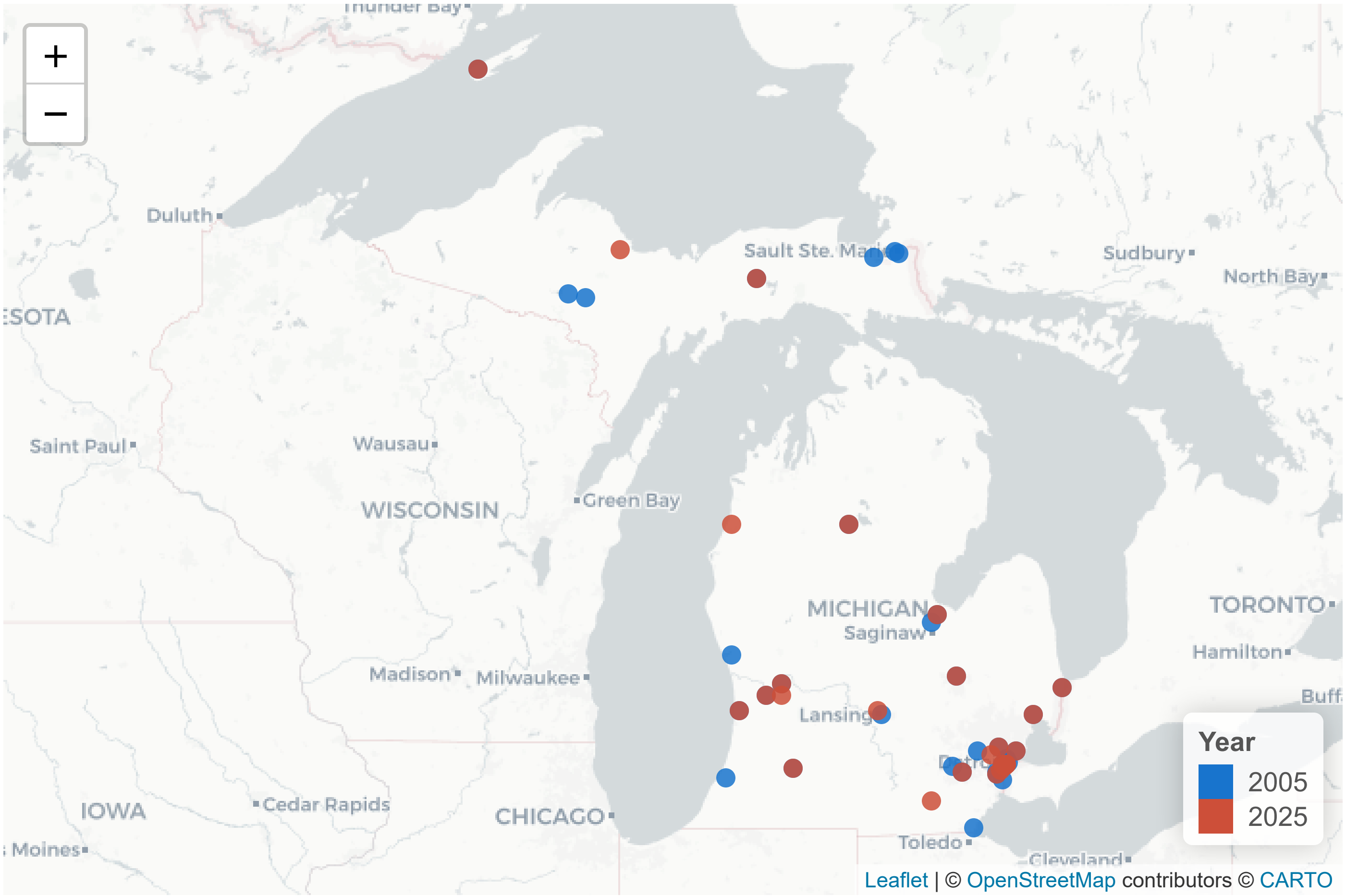
```r
ggplot(pm_all, aes(x = pm25)) +
  geom_histogram(bins = 60) +
  labs(title = "PM2.5 daily mean distribution (2005 & 2025)", x = "PM2.5 (µg/m³)", y = "Count")
```



3. Create leaflet map to show monitoring site locations

```
library(dplyr)
library(leaflet)

# Filter to Michigan data
df <- pm_all %>% filter(state == "Michigan")

# One row per site per year
sites_year <- df %>%
  filter(!is.na(Site.Latitude), !is.na(Site.Longitude)) %>%
  distinct(year, state, county, site, Site.Latitude, Site.Longitude)

pal <- colorFactor(palette = c("dodgerblue3", "tomato3"), domain = sort(unique(sites_year$year)))

m <- leaflet(sites_year) %>%
  addProviderTiles(providers$CartoDB.Positron) %>%
  addCircleMarkers(
    lng = ~Site.Longitude, lat = ~Site.Latitude,
    radius = 5, stroke = FALSE, fillOpacity = 0.85,
    color = ~pal(year),
    popup = ~paste0(
```

```
      "<b>Year:</b> ", year, "<br/>",
      "<b>Site:</b> ", state, "-", county, "-", site, "<br/>",
      "<b>Lat/Lon:</b> ", round(Site.Latitude, 4), ", ", round(Site.Longitude, 4)
    )
  ) %>%
  addLegend(
    position = "bottomright",
    pal = pal, values = ~year,
    title = "Year", opacity = 1
  )

m
```




```
# Counts by year
sites_year %>%
  count(year, name = "n_sites")
```

```
  year n_sites
1 2005      34
2 2025      27
```

```
# Bounding box by year (rough geographic footprint)
sites_year %>%
  group_by(year) %>%
```

```
  summarize(
    n_sites = n(),
    lat_min = min(Site.Latitude), lat_max = max(Site.Latitude),
    lon_min = min(Site.Longitude), lon_max = max(Site.Longitude),
    .groups = "drop"
  )
```

```
# A tibble: 2 × 6
   year n_sites lat_min lat_max lon_min lon_max
  <int>   <int>   <dbl>   <dbl>   <dbl>   <dbl>
1  2005      34    41.8    47.9   -89.2   -82.5
2  2025      27    42.0    47.9   -89.2   -82.5
```

```
# Sites present in both years vs. only one year
site_id <- sites_year %>%
  mutate(site_id = paste(state, county, site, sep = "-")) %>%
  distinct(year, site_id)

site_id %>%
  count(site_id) %>%
  count(n, name = "n_sites") %>%
  mutate(present_in_years = n) %>%
  arrange(desc(present_in_years))
```

```
  n n_sites present_in_years
1 2      17                2
2 1      27                1
```

From the map, we find the the monitoring sites are not evenly spread across Michigan. Instead, they have the highest concentration in Southeeast Michigan (the Detroit - Ann Arbor - Downriver area) for both 2005 and 2025. This makes sense, as this would be the area in Michigan where the population - and thus the emission source - is the highest. There are additional clusters of sites in central/east-central Michigan, such as around Lansing and the Saginaw / Bay City / Flint region. Meanwhile, there are very few sites in western and northern Michigan (which were only introduced as of 2025), and a small amount in the Upper Peninsula. Overall, while the number of monitoring sites increased from 2005 to 2025, the spatial distribution of the sites across Michigan remained mostly unchanged.

4. Check for data issues

Identify missing or implausible values

```
library(dplyr)

df <- pm_all %>%
  mutate(
    year = as.integer(year),
```

```
    pm25 = as.numeric(pm25)
  )
```

```
# 1) Missingness
df %>%
  summarize(
    n = n(),
    pm25_missing = sum(is.na(pm25)),
    pm25_missing_pct = mean(is.na(pm25)) * 100
  )
```

```
      n pm25_missing pm25_missing_pct
1 15214            0                0
```

```
df %>%
  group_by(year) %>%
  summarize(
    n = n(),
    pm25_missing = sum(is.na(pm25)),
    pm25_missing_pct = mean(is.na(pm25)) * 100,
    .groups = "drop"
  )
```

```
# A tibble: 2 × 4
   year     n pm25_missing pm25_missing_pct
  <int> <int>        <int>            <dbl>
1  2005  5280            0                0
2  2025  9934            0                0
```

We find that there are no values in the dataset that are completely missing.

```
# 2) Implausible values (negative, extremely large)
df %>%
  summarize(
    n_neg = sum(pm25 < 0, na.rm = TRUE),
    n_zero = sum(pm25 == 0, na.rm = TRUE),
    n_gt_500 = sum(pm25 > 500, na.rm = TRUE),
    n_gt_1000 = sum(pm25 > 1000, na.rm = TRUE)
  )
```

```
  n_neg n_zero n_gt_500 n_gt_1000
1     0      4        0         0
```

We find that there are only four measurements in the data that are exactly zero, and none that are negative or implausibly large. Since these measurements appear to be reasonable, we do not remove any records from the dataset.

```
# 3) Distribution checks by year
df %>%
  group_by(year) %>%
  summarize(
    n = sum(!is.na(pm25)),
    min = min(pm25, na.rm = TRUE),
    p1 = quantile(pm25, 0.01, na.rm = TRUE),
    p5 = quantile(pm25, 0.05, na.rm = TRUE),
    median = median(pm25, na.rm = TRUE),
    p95 = quantile(pm25, 0.95, na.rm = TRUE),
    p99 = quantile(pm25, 0.99, na.rm = TRUE),
    max = max(pm25, na.rm = TRUE),
    mean = mean(pm25, na.rm = TRUE),
    sd = sd(pm25, na.rm = TRUE),
    .groups = "drop"
  )
```

```
# A tibble: 2 × 11
   year     n   min    p1    p5 median   p95   p99   max  mean    sd
  <int> <int> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1  2005  5280     0   1.5   2.7   11    32.5  52.3  81.6 13.8  10.4 
2  2025  9934     0   1     2      6.3  20.6  41.6 116    8.30  7.42
```

From this table, we see an overall level shift since the mean and median significantly decreased from 2005 to 2025; the mean decreased from 13.77 to 8.3, and the median decreased from 11.0 to 6.3. The highest percentiles are also lower in 2025 compared to 2005, which suggests that the overall PM2.5 concentrations in the air were less in 2025. There are also outliers of very high maximum concentrations in both years (81.6 and 116.0 respectively), which could be caused by an extreme event such as a wildfire.

```
# 4) Missing days per site-year
site_days <- df %>%
  filter(!is.na(pm25)) %>%
  group_by(state, county, site, year) %>%
  summarize(
    n_days = n_distinct(Date),
    .groups = "drop"
  )

site_days %>%
  group_by(year) %>%
  summarize(
    n_site_years = n(),
    min_days = min(n_days),
    p25_days = quantile(n_days, 0.25),
    median_days = median(n_days),
    p75_days = quantile(n_days, 0.75),
    max_days = max(n_days),
    .groups = "drop"
  )
```

```
# A tibble: 2 × 7
   year n_site_years min_days p25_days median_days p75_days max_days
  <int>        <int>    <int>    <dbl>       <dbl>    <dbl>    <int>
1  2005           34       15     113.         117      119      362
2  2025           27       29     359          363      365      365
```

This shows that 2025 has roughly the same number of site-year pairs as 2005, with more frequent coverage per day in 2025 (higher median and percentiles). Both years are covered across most of the days in the calendar year for at least one site, since the maximum number of days per site-year pair for both 2005 and 2025 are close to 365.

Check methods used for data collection

```r
unique(pm_all$"Method.Description")
```

```
[1] "R & P Model 2025 PM2.5 Sequential w/WINS"
[2] "Met One SASS/SuperSASS Teflon"
[3] "IMPROVE Module A with Cyclone Inlet-Teflon Filter, 2.2 sq. cm."
[4] "SINGLE-FILTR WINS 2.5UM IMPACT"
[5] "Met One BAM-1020 Mass Monitor w/VSCC"
[6] ""
[7] "R & P Model 2025 PM-2.5 Sequential Air Sampler w/VSCC"
[8] "Teledyne T640X at 16.67 LPM w/Network Data Alignment enabled"
[9] "Teledyne T640 at 5.0 LPM w/Network Data Alignment enabled"
```

```r
library(dplyr)

df <- pm_all %>%
  mutate(
    year = as.integer(year),
    method_code = Method.Code,
    method_name = Method.Description
  )

# 1) Find counts of method codes
# 2005
methods <- df %>%
  filter(year == 2005) %>%
  group_by(method_code) %>%
  summarize(
    n = n(),
  ) %>%
  arrange(desc(n))

methods
```

```
# A tibble: 4 × 2
  method_code     n
```

```
        <int> <int>
1         118  4355
2         810   607
3         707   274
4         781    44
```

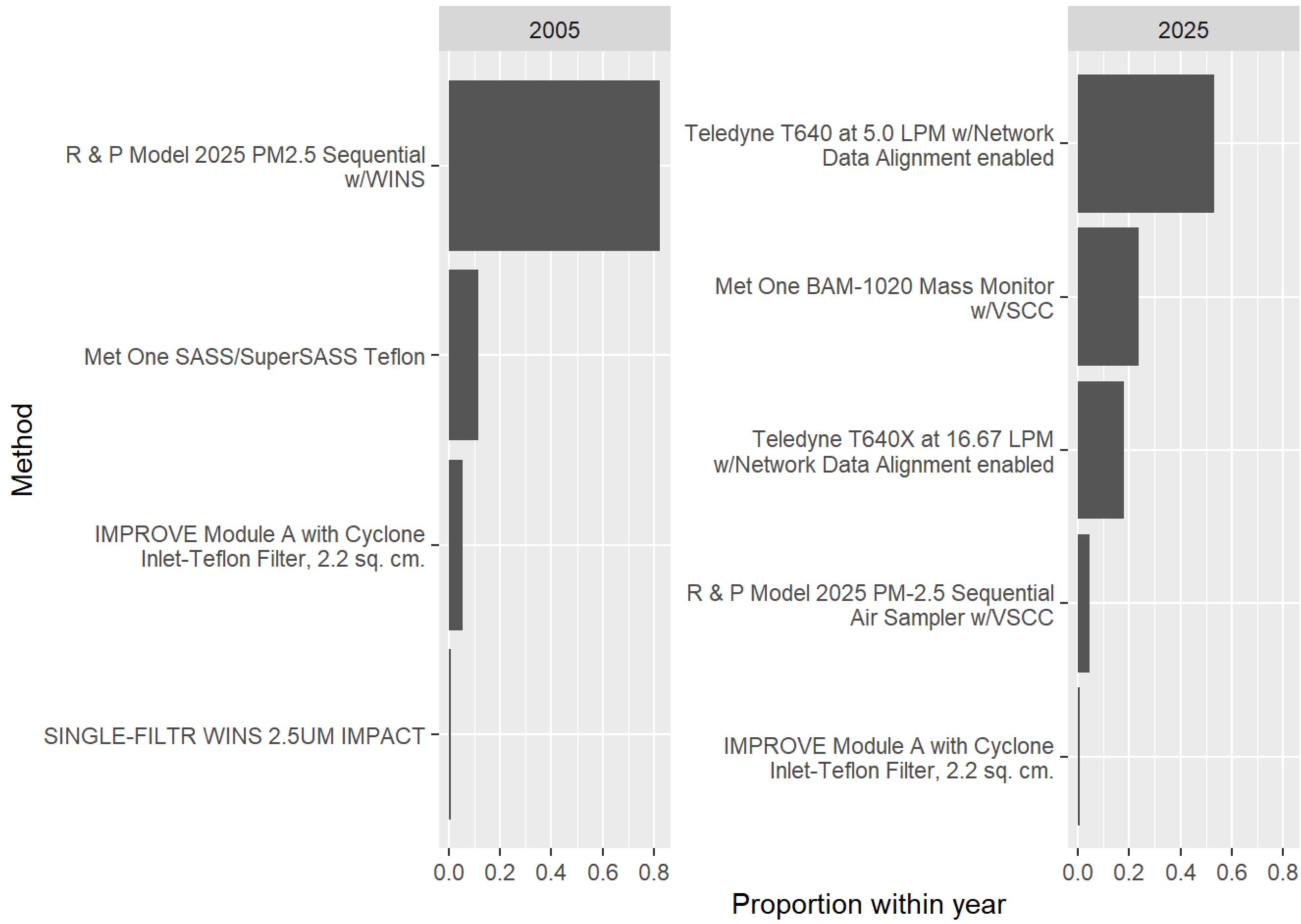
```r
# 2025
methods <- df %>%
  filter(year == 2025) %>%
  group_by(method_code) %>%
  summarize(
    n = n(),
  ) %>%
  arrange(desc(n))

methods
```

```
# A tibble: 6 × 2
  method_code     n
        <int> <int>
1         636  3941
2          NA  2543
3         170  1741
4         638  1336
5         145   324
6         707    49
```

We note that the number of observations with a missing method code was 0 in 2005, but this count increased to 2,543 observations with missing codes in 2025.

```r
library(stringr)

# 2) Distribution of Method.Name within each year (proportions among non-missing)
method_dist <- df %>%
  filter(!is.na(method_code)) %>%
  count(year, method_name, name = "n") %>%
  group_by(year) %>%
  mutate(prop = n / sum(n)) %>%
  arrange(year, desc(n)) %>%
  ungroup()

# 3) Top 10 method codes per year
top10 <- method_dist %>%
  group_by(year) %>%
  slice_max(n, n = 10, with_ties = FALSE) %>%
  ungroup() %>%
  mutate(method_wrap = str_wrap(method_name, width = 35))

ggplot(top10, aes(x = prop, y = reorder(method_wrap, prop))) +
  geom_col() +
```

```
  facet_wrap(~ year, scales = "free_y") +
  labs(x = "Proportion within year", y = "Method")
```



From the [EPA PM2.5 Codetable](), we find that the most common method names in 2005 correspond to codes 118 and 810 ("R & P Model 2025 PM2.5 Sequential w/WINS" and "Met One SASS/SuperSASS Teflon"), respectively, while the most common codes in 2025 are 636 and 170 ("Teledyne T640 at 5.0 LPM w/Network Data Alignment enabled" and "Met One BAM-1020 Mass Monitor w/VSCC"), respectively. However, the more helpful information is the proportion with respect to method names.

From 2005 to 2025, there is a big shift in the dominant measurement technology - instead of the overwhelming presence of gravimetric, filter-based samplers, there are more continuous and automated instrumenets (such as Teledyne for broadband spectroscopy and Met One for beta attenuation). Overall, there is a more diverse set of instruments being used in 2025 compared to in 2005, with the most frequently used method in 2025 being a smaller proportion (about 50% of the time) compared to in 2005 (over 80% of the time). These temporal patterns are important to analyze because they reflect how methodologies have been modernized over the past 20 years to include new types of instruments. In this case, the new instruments can measure PM2.5 concentrations in real time instead of only collecting samples at scheduled times.

However, the sharp increase in the presence of observations made with missing methods also highlights the importance of properly recording the methods used to make measurements, so that these types of trends can be discerned in the first place.

5. Visualize and summarize daily concentrations of PM2.5

```
# Standardize, restrict to Michigan, and keep only years of interest
library(dplyr)
library(ggplot2)
library(stringr)

pm_mi <- pm_all %>%
  mutate(
    state  = if ("state"  %in% names(.)) state  else State.Code,
    county = if ("county" %in% names(.)) county else County.Code,
    site   = if ("site"   %in% names(.)) site   else Site.Num,
    pm25   = if ("pm25"   %in% names(.)) pm25   else as.numeric(Arithmetic.Mean),
    Date   = if ("Date"   %in% names(.)) as.Date(Date) else as.Date(Date.Local),
    year   = if ("year"   %in% names(.)) year   else as.integer(format(Date, "%Y")),
    county_name = if ("County.Name" %in% names(.)) County.Name else NA_character_,
    city_name   = if ("City.Name"   %in% names(.)) City.Name   else NA_character_,
    lat = as.numeric(if ("Latitude"  %in% names(.)) Latitude  else NA),
    lon = as.numeric(if ("Longitude" %in% names(.)) Longitude else NA)
  ) %>%
  filter(state == "Michigan", year %in% c(2005, 2025))
```

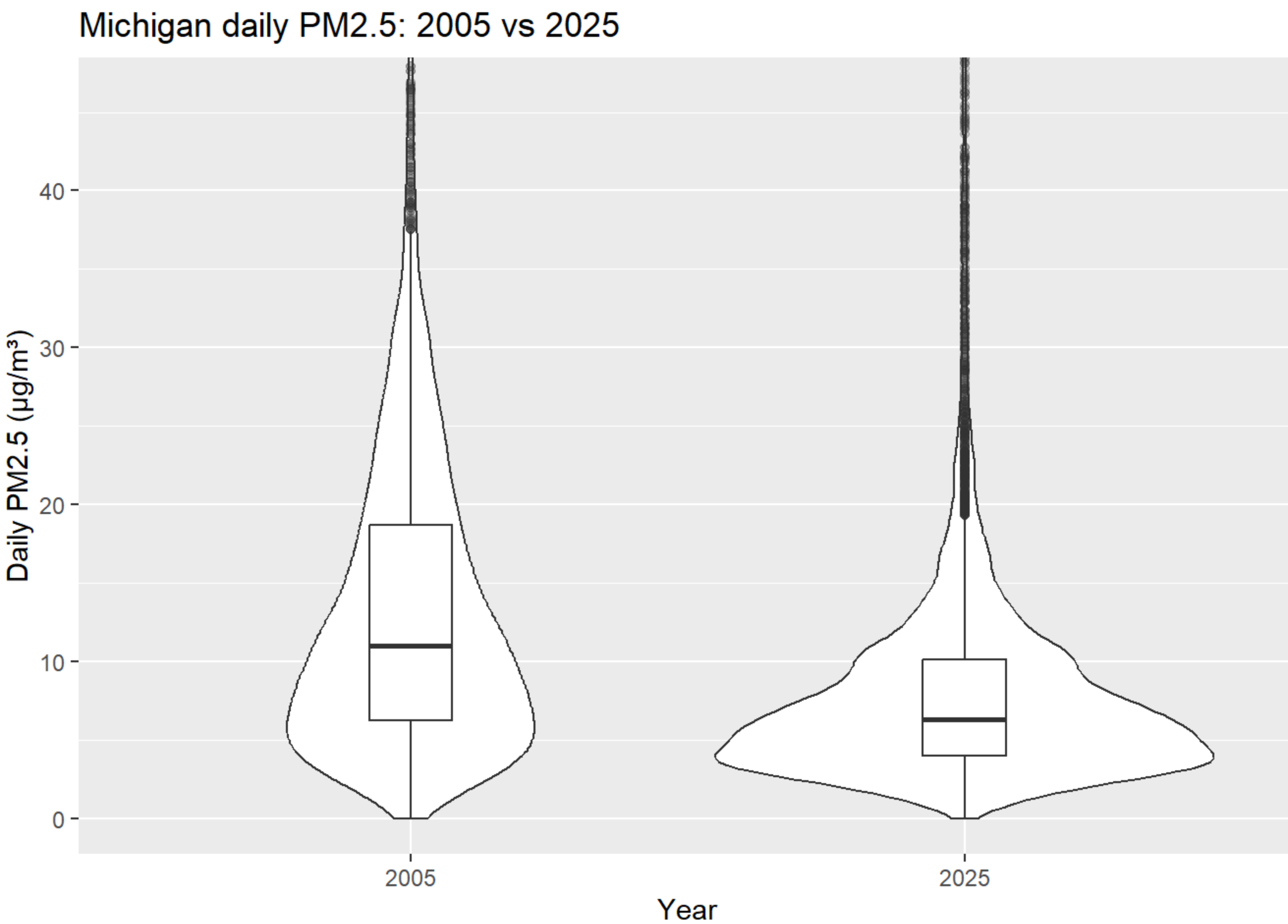
Level 1: PM2 concentrations in Michigan overall

```
state_summary <- pm_mi %>%
  group_by(year) %>%
  summarize(
    n_days = sum(!is.na(pm25)),
    mean = mean(pm25, na.rm = TRUE),
    median = median(pm25, na.rm = TRUE),
    p25 = quantile(pm25, 0.25, na.rm = TRUE),
    p75 = quantile(pm25, 0.75, na.rm = TRUE),
    p95 = quantile(pm25, 0.95, na.rm = TRUE),
    max = max(pm25, na.rm = TRUE),
    .groups = "drop"
  )

state_summary
```

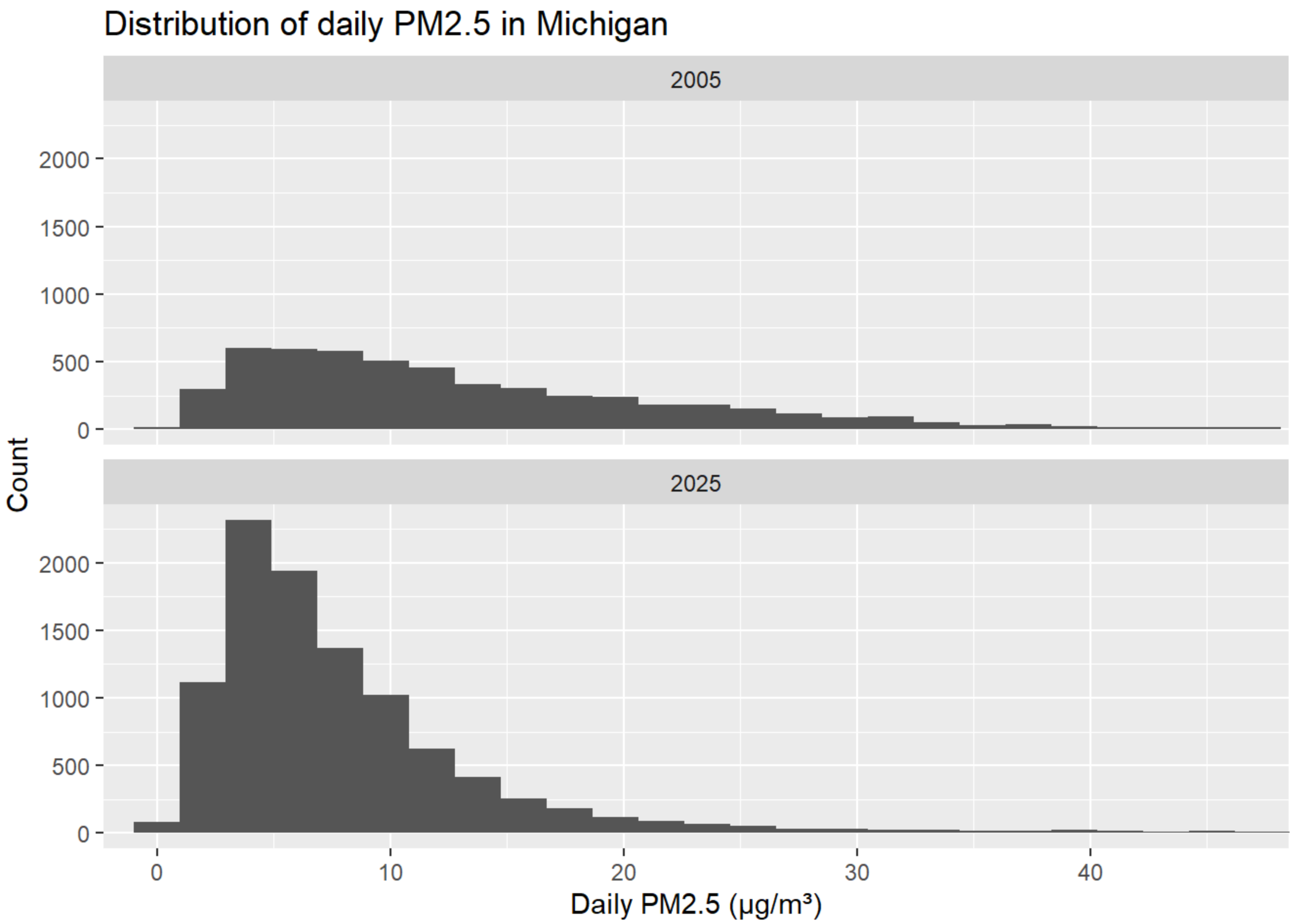
```
# A tibble: 2 × 8
   year n_days  mean median   p25   p75   p95   max
  <int>  <int> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1  2005   5280 13.8     11     6.2  18.7  32.5  81.6
2  2025   9934  8.30    6.3   4     10.1  20.6 116
```

```
ggplot(pm_mi, aes(x = factor(year), y = pm25)) +
  geom_violin(trim = TRUE) +
```

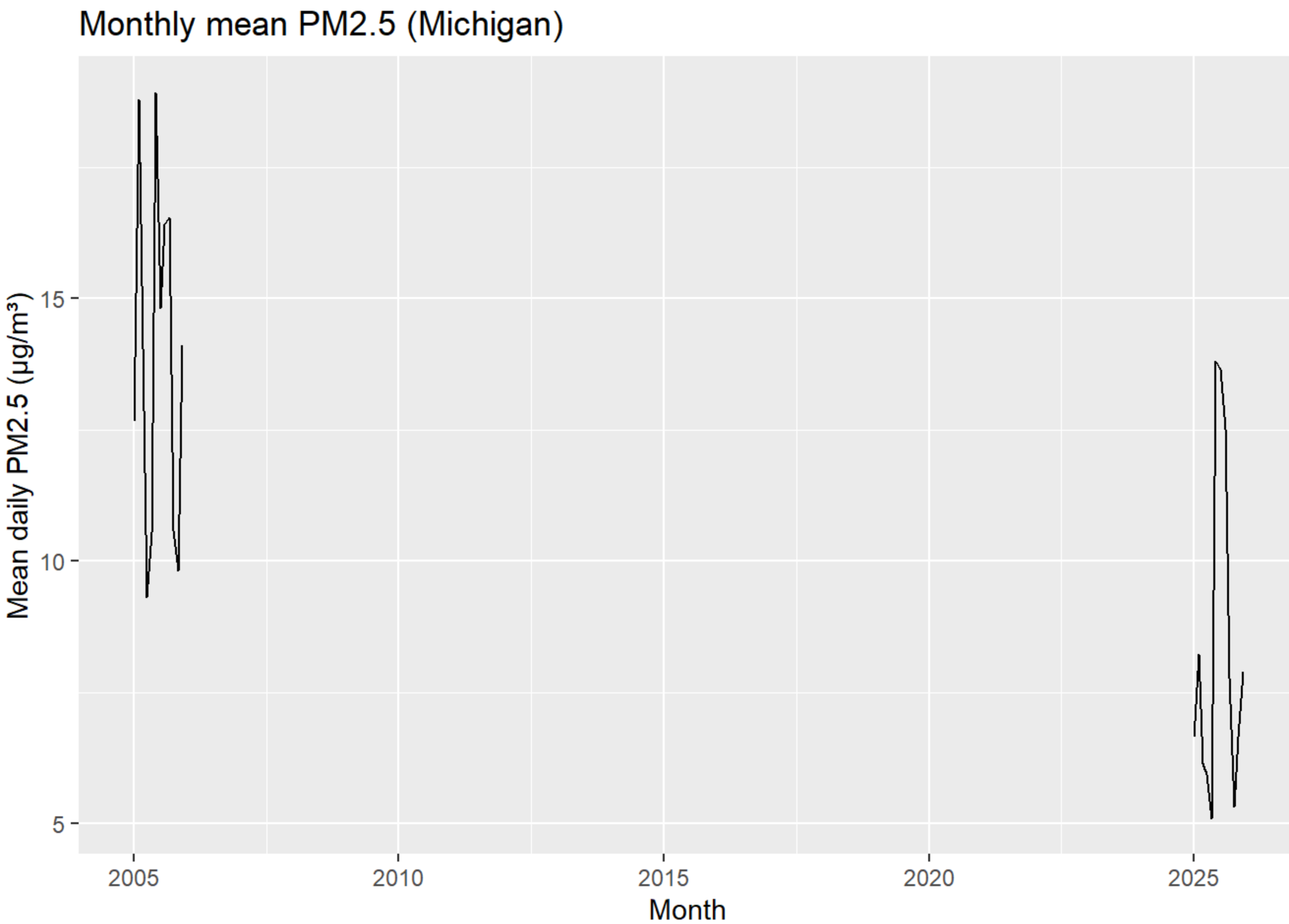
```
  geom_boxplot(width = 0.15, outlier.alpha = 0.2) +
  labs(x = "Year", y = "Daily PM2.5 (µg/m³)", title = "Michigan daily PM2.5: 2005 vs 2025") +
  coord_cartesian(ylim = c(0, quantile(pm_mi$pm25, 0.99, na.rm = TRUE)))
```




```
ggplot(pm_mi, aes(x = pm25)) +
  geom_histogram(bins = 60) +
  facet_wrap(~ year, ncol = 1) +
  labs(x = "Daily PM2.5 (µg/m³)", y = "Count", title = "Distribution of daily PM2.5 in Michigan")
  coord_cartesian(xlim = c(0, quantile(pm_mi$pm25, 0.99, na.rm = TRUE)))
```




```
pm_mi_month <- pm_mi %>%
  mutate(month = as.Date(format(Date, "%Y-%m-01"))) %>%
  group_by(year, month) %>%
  summarize(pm25_mean = mean(pm25, na.rm = TRUE), .groups = "drop")

ggplot(pm_mi_month, aes(x = month, y = pm25_mean, group = factor(year))) +
  geom_line() +
  labs(x = "Month", y = "Mean daily PM2.5 (µg/m³)", title = "Monthly mean PM2.5 (Michigan)")
```



The summary statistics show that the mean and all of the percentiles (except for the max) of the daily PM2.5 concentration per month have decreased from 2005 to 2025; for example, the mean concentration decreased from 13.77 to 8.3 µg/m^3. The side-by-side box/violin plots and histograms also illustrate this well by showing a high frequency of values around 5 µg/m^3 in 2025, while concentrations of 20 µg/m^3 or higher (which were considered evenly distrbuted in 2005) are considered outliers in 2025. In general, the violin plot and line plot show that the distribution of concentrations has shifted down.

Level 2: PM2 concentrations by Michigan county

```
county_summary <- pm_mi %>%
  filter(!is.na(county)) %>%
  group_by(year, county) %>%
  summarize(
    n = sum(!is.na(pm25)),
    mean = mean(pm25, na.rm = TRUE),
    median = median(pm25, na.rm = TRUE),
    p95 = quantile(pm25, 0.95, na.rm = TRUE),
    .groups = "drop"
  )

county_summary
```

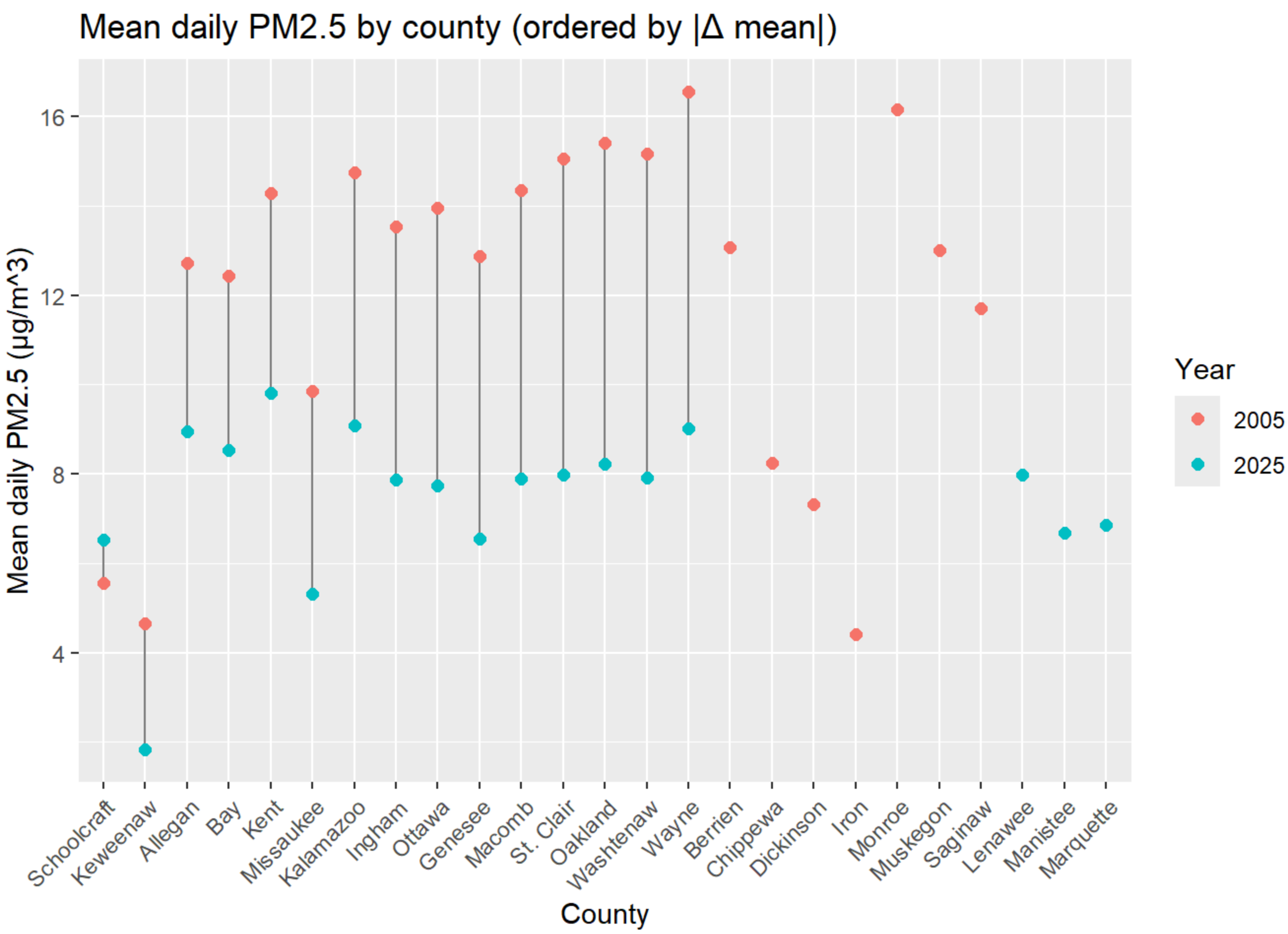
```
# A tibble: 40 × 6
    year county        n  mean median   p95
   <int> <chr>     <int> <dbl>  <dbl> <dbl>
 1  2005 Allegan     421 12.7    10.4 31.7
 2  2005 Bay         117 12.4     9.7 28.2
 3  2005 Berrien     118 13.1    10.8 31.3
 4  2005 Chippewa    378  8.25    6   21.1
 5  2005 Dickinson    25  7.32    6.9 13.2
 6  2005 Genesee     119 12.9    10.7 27.4
 7  2005 Ingham      120 13.5    11.2 30.0
 8  2005 Iron         25  4.40    3.5  9.36
 9  2005 Kalamazoo   228 14.7    13   31.1
10  2005 Kent        475 14.3    11.3 36.0
# ℹ 30 more rows
```

```
county_change <- county_summary %>%
  select(year, county, mean, median, p95) %>%
  tidyr::pivot_wider(names_from = year, values_from = c(mean, median, p95)) %>%
  mutate(
    d_mean   = mean_2025   - mean_2005,
    d_median = median_2025 - median_2005,
    d_p95    = p95_2025    - p95_2005
  )

county_change %>% arrange(d_mean)
```

```
# A tibble: 25 × 10
   county   mean_2005 mean_2025 median_2005 median_2025 p95_2005 p95_2025 d_mean
   <chr>        <dbl>     <dbl>       <dbl>       <dbl>    <dbl>    <dbl>  <dbl>
 1 Wayne        16.5       9.01        13.7        7.1      36.0     20.7  -7.53
 2 Washten…     15.2       7.91        12          5.9      33.7     19.5  -7.25
 3 Oakland      15.4       8.21        12.3        6.1      36.1     21.5  -7.19
 4 St. Cla…     15.0       7.97        12          5.95     36.6     21.5  -7.08
 5 Macomb       14.4       7.88        10.8        5.7      32.6     19.9  -6.47
 6 Genesee      12.9       6.54        10.7        5.25     27.4     16.3  -6.34
 7 Ottawa       13.9       7.74        11.9        5.6      32.1     21.2  -6.21
 8 Ingham       13.5       7.87        11.2        6        30.0     20.1  -5.67
 9 Kalamaz…     14.7       9.07        13          7.2      31.1     23.8  -5.66
10 Missauk…      9.86      5.32         7.9        3.6      25.7     16.8  -4.54
# ℹ 15 more rows
# ℹ 2 more variables: d_median <dbl>, d_p95 <dbl>
```

```
county_levels <- county_change %>%
  arrange(abs(d_mean)) %>%
  pull(county)

plot_df <- county_summary %>%
  mutate(county = factor(county, levels = county_levels))

ggplot(plot_df, aes(x = county, y = mean, color = factor(year), group = county)) +
```

```
  geom_line(color = "grey50") +
  geom_point(size = 2) +
  labs(x = "County", y = "Mean daily PM2.5 (µg/m^3)", color = "Year",
       title = "Mean daily PM2.5 by county (ordered by |Δ mean|)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

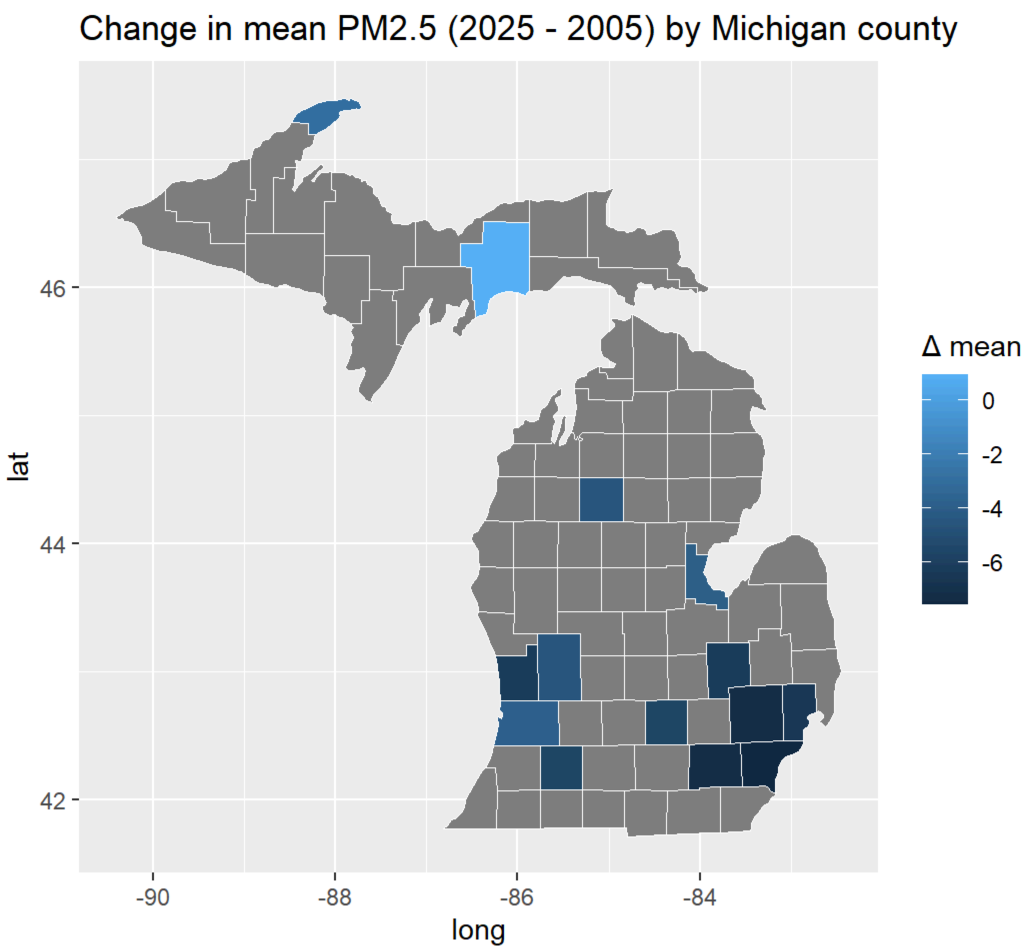


From this plot, we can see that out of every county in Michigan where measurements were recorded in both 2005 and 2025, all of the measurements from 2025 are significantly smaller than those in 2005 (except for Schoolcraft where the reverse is true). According to the summary statistics, the median and 95th percentile also uniformly decrease from 2005 to 2025. The decreases in concentrations range from -2.81 (in Keweenaw) to -7.53 µ/m^3 (in Wayne), meaning the concentrations decreased by at least 4 µ/m^3 and up to almost 8 µ/m^3 by county from 2005 to 2025.

We also see an overall decrease when comparing the 7 measurements for counties only taken in 2005 compared to the 3 measurements for counties only taken in 2025, since none of the measurements from counties only recorded in 2025 are higher than 8 µg/m^3.

```
library(maps)

mi_map <- map_data("county") %>%
  filter(region == "michigan") %>%
  mutate(county = str_to_title(subregion))
```

```r
county_map_df <- mi_map %>%
  left_join(county_change, by = "county")

ggplot(county_map_df, aes(long, lat, group = group, fill = d_mean)) +
  geom_polygon(color = "white", linewidth = 0.2) +
  coord_quickmap() +
  labs(title = "Change in mean PM2.5 (2025 - 2005) by Michigan county",
       fill = "Δ mean")
```



This choropleth shows the changes in mean PM2.5 concentrations per Michigan county from 2005 to 2025. We can see that the largest decreases (the deepest blue colors) are in counties in the southeast region of Michigan, which (as discussed previously) is where Metro Detroit is located. This suggests that the decreases are concentrated in metro/industrial areas, and the decreases are less pronounced in more rural counties closer to the north side of Michigan.

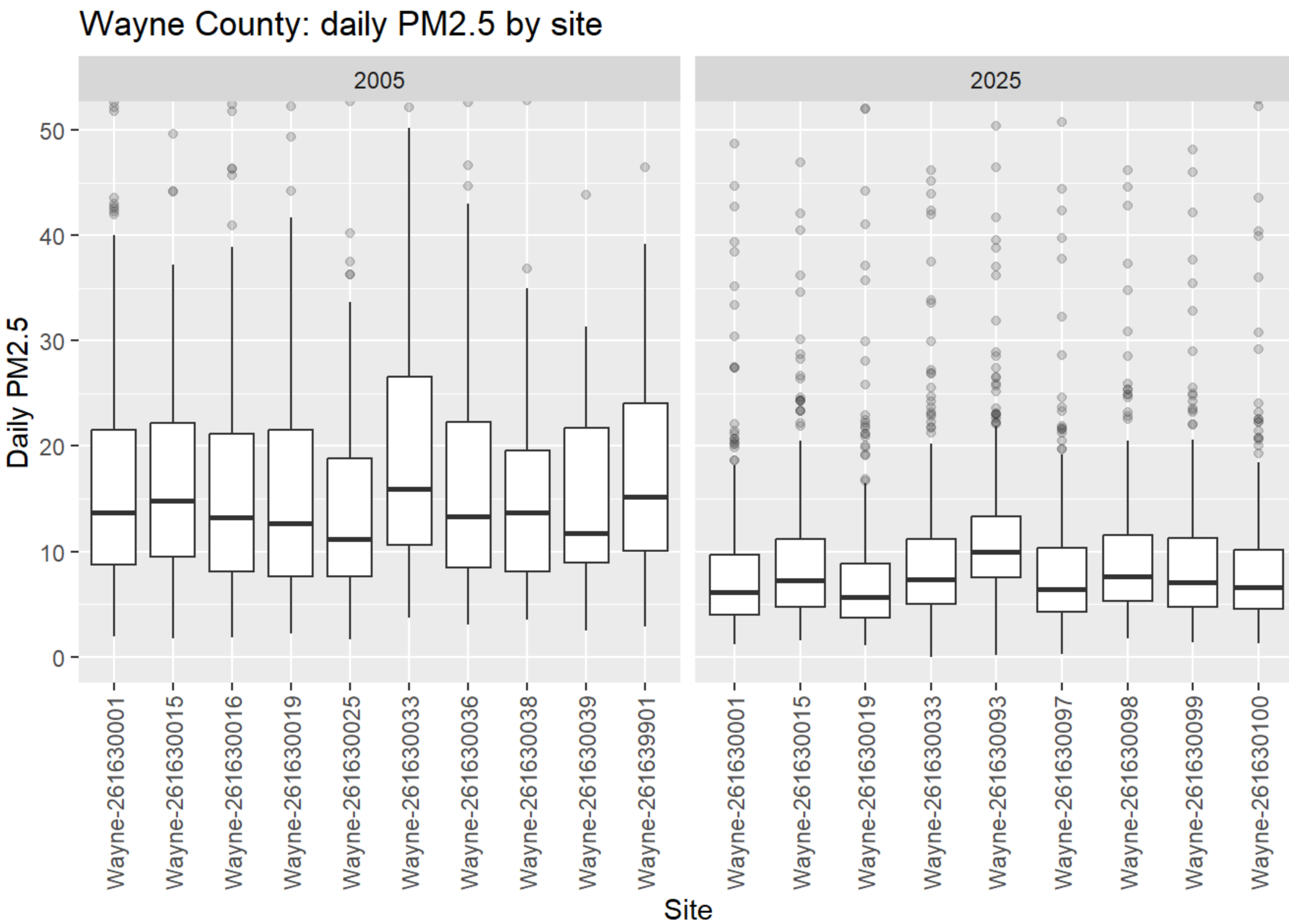
Level 3: PM2 concentrations by site in Wayne County

```r
pm_wayne <- pm_mi %>%
  filter(county == "Wayne") %>%
  mutate(site_id = paste0(county, "-", site))

site_summary <- pm_wayne %>%
```

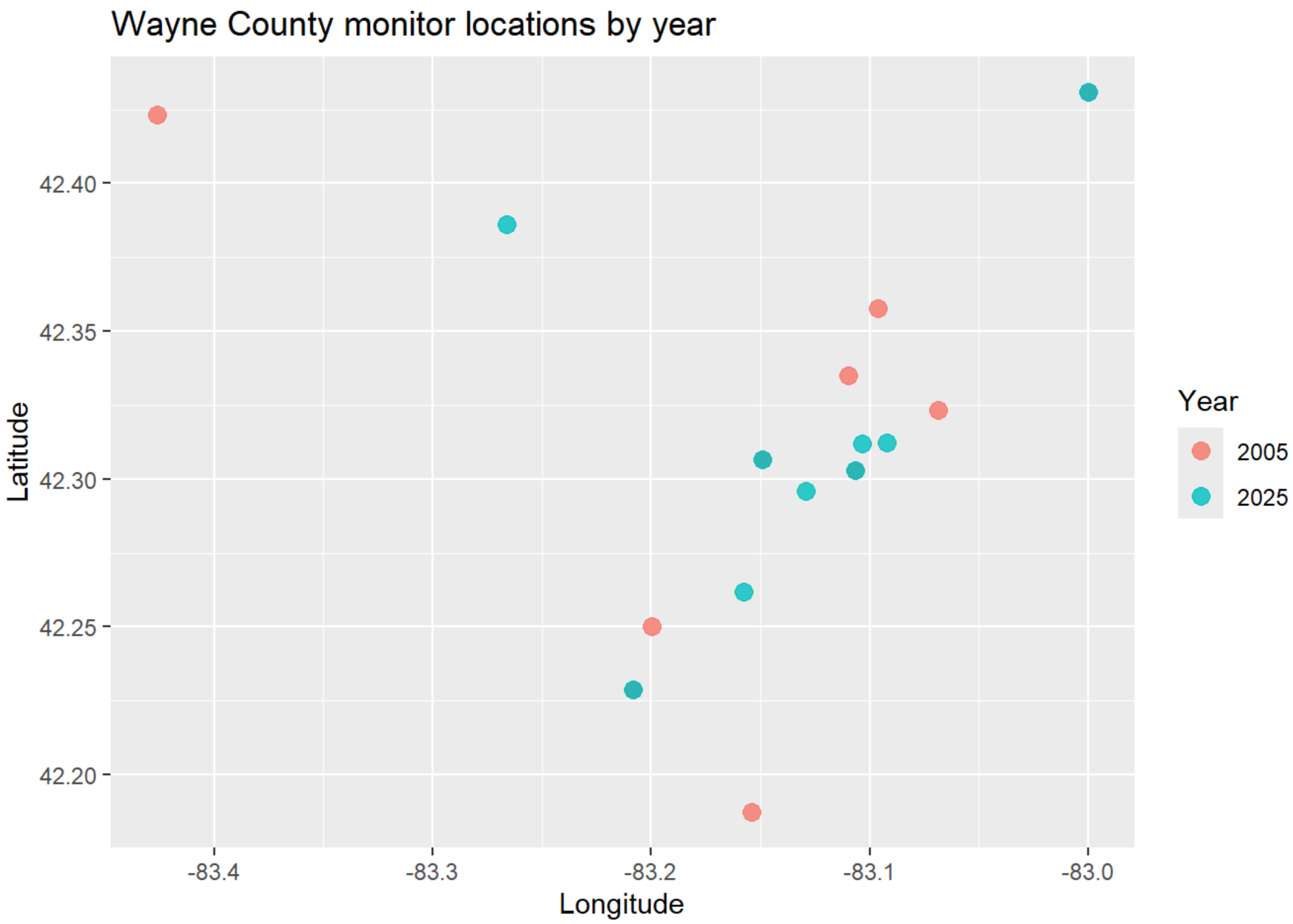
```r
  group_by(year, site_id, city_name) %>%
  summarize(
    n = sum(!is.na(pm25)),
    mean = mean(pm25, na.rm = TRUE),
    median = median(pm25, na.rm = TRUE),
    p95 = quantile(pm25, 0.95, na.rm = TRUE),
    .groups = "drop"
  )

site_summary
```

```
# A tibble: 19 × 7
    year site_id          city_name     n  mean median   p95
   <int> <chr>            <chr>     <int> <dbl>  <dbl> <dbl>
 1  2005 Wayne-261630001 <NA>        523 16.3    13.7  34.1
 2  2005 Wayne-261630015 <NA>        114 17.2    14.8  34.2
 3  2005 Wayne-261630016 <NA>        338 16.0    13.2  35.4
 4  2005 Wayne-261630019 <NA>        117 16.4    12.7  41.5
 5  2005 Wayne-261630025 <NA>        114 14.9    11.2  34.6
 6  2005 Wayne-261630033 <NA>        170 18.9    16.0  36.3
 7  2005 Wayne-261630036 <NA>        113 16.4    13.3  34.2
 8  2005 Wayne-261630038 <NA>         75 16.4    13.7  35.6
 9  2005 Wayne-261630039 <NA>         35 15.0    11.7  28.8
10  2005 Wayne-261639901 <NA>         44 19.6    15.2  45.4
11  2025 Wayne-261630001 <NA>        446  7.96    6.1  18.5
12  2025 Wayne-261630015 <NA>        461  9.23    7.2  20.5
13  2025 Wayne-261630019 <NA>        364  7.57    5.7  19.8
14  2025 Wayne-261630033 <NA>        465  9.35    7.3  21.6
15  2025 Wayne-261630093 <NA>        357 11.6    10     23.0
16  2025 Wayne-261630097 <NA>        365  8.33    6.4  19.6
17  2025 Wayne-261630098 <NA>        363  9.67    7.6  20.5
18  2025 Wayne-261630099 <NA>        365  9.14    7.1  20.6
19  2025 Wayne-261630100 <NA>        363  8.43    6.6  19.9
```

```r
ggplot(pm_wayne, aes(x = site_id, y = pm25)) +
  geom_boxplot(outlier.alpha = 0.2) +
  facet_wrap(~ year, scales = "free_x") +
  labs(x = "Site", y = "Daily PM2.5", title = "Wayne County: daily PM2.5 by site") +
  coord_cartesian(ylim = c(0, quantile(pm_wayne$pm25, 0.99, na.rm = TRUE))) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5))
```




```
wayne_sites <- pm_wayne %>%
  filter(!is.na(Site.Latitude), !is.na(Site.Longitude)) %>%
  distinct(year, site_id, city_name, Site.Latitude, Site.Longitude)

ggplot(wayne_sites, aes(x = Site.Longitude, y = Site.Latitude, color = factor(year))) +
  geom_point(size = 3, alpha = 0.8) +
  labs(x = "Longitude", y = "Latitude", color = "Year",
       title = "Wayne County monitor locations by year") +
  coord_quickmap()
```



From the series of boxplots that show the distribution of daily PM2.5 concentrations in all of the Wayne County sites, we see that the medians have uniformly decreased from 2005 to 2025; in fact, according to the summary statistics, the lowest mean and median concentrations in 2005 are still higher than the highest mean and median concentrations in 2025. The boxplots also show (much like the broad, state-wide analysis from before) that the concentrations above 20 µg/m^3 are within the 75th percentile in 2005, but outliers in 2025. Finally, the spatial plot of site locations colored by year shows that the sites in 2005 (red points) are more spread out across Wayne County, but in 2025 (blue points), they are more clustered in one area between latitude 42.30 and longitude -83.15, with only three sites located away from that cluster.

Therefore, we can finally conclude that on the level of the entire state of Michigan, on the level of each county within Michigan, and the level of each observation site within Wayne County, the daily concentrations of PM2.5 have decreased in Michigan over the 20 years spanning from 2005 to 2025.