Introduction

The release of pollutants into the atmosphere is what causes air pollution.It’s impact is not only on the environment but also impacts public and indivitual health.Among the pollutants like carbon dioxide and methane,Particulate Matter(PM) has adverse effects on the environment and health.It is formed as a result of chemical reaction between different pollutants.

The variables particulate matter 2.5(PM2.5) and Air Quality Index(AQI) for 2018 are used from this dataset to study the trend and patterns of air pollution.

The annual AQI by county dataset has 1056 observations and 19 variables.

The Daily vevel of PM 2.5 concentration dataset has 502607 observations and 29 variables.

References

The data is sourced from the Environmental Protection Agency of the United States(EPA): https://aqs.epa.gov/aqsweb/airdata/download_files.html (https://aqs.epa.gov/aqsweb/airdata/download_files.html)

The data sources for the selected variables are:

-Annual AQI by county 2018

-Daily level of PM 2.5 concentration 2018

Data Preprocessing

```
##
## 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
```

```
##  Attaching packages                                      tidyverse 1.3.1
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## ── Conflicts ──────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
## Loading required package: flexmix
```

```
## Loading required package: lattice
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
## Loading required package: maps
```

```
##
## Attaching package: 'maps'
```

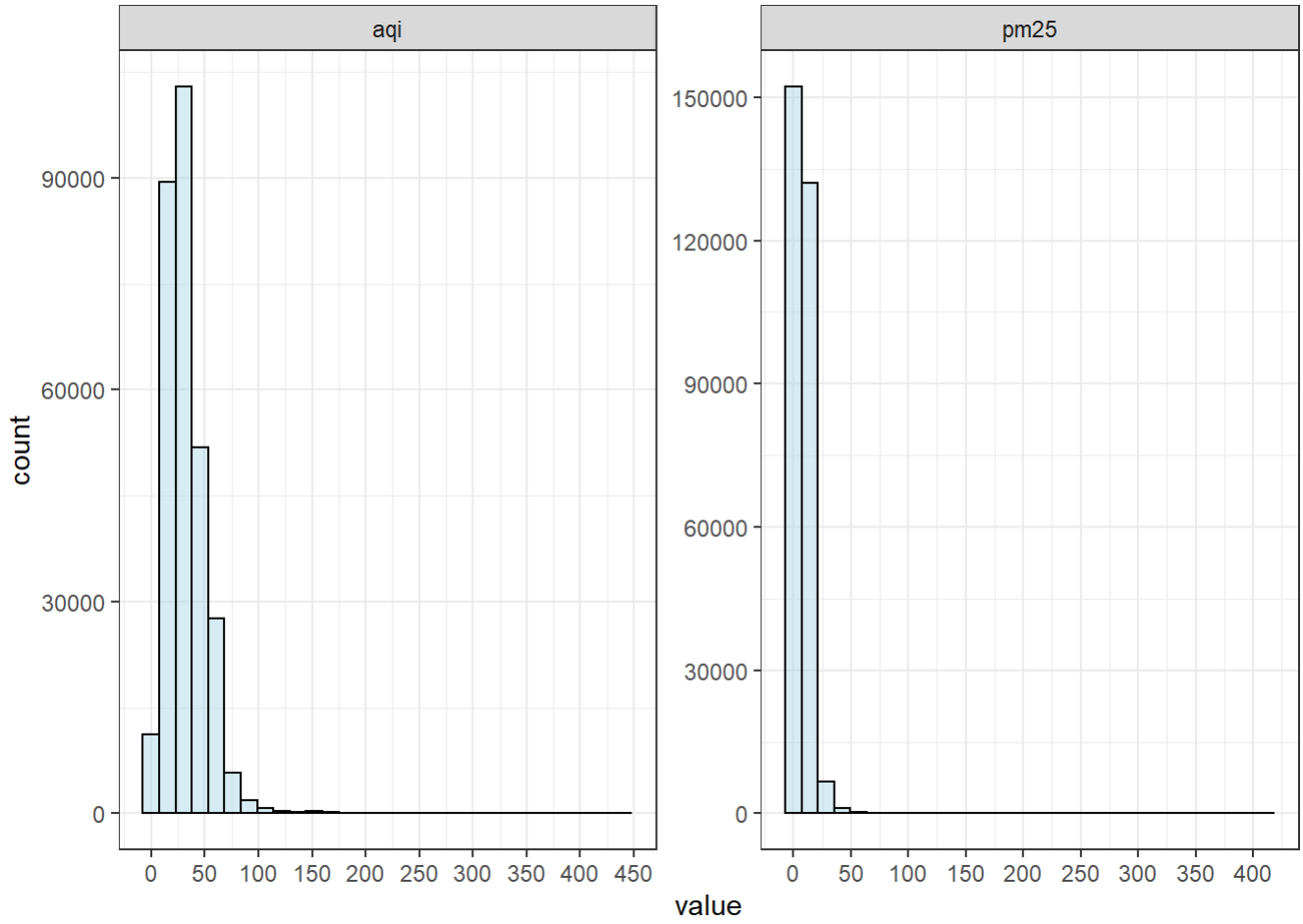
```
## The following object is masked from 'package:purrr':
##
##     map
```

```
##
## ── Column specification ────────────────────────────────────────────────────────
## cols(
##   State = col_character(),
##   County = col_character(),
##   Year = col_double(),
##   `Days with AQI` = col_double(),
##   `Good Days` = col_double(),
##   `Moderate Days` = col_double(),
##   `Unhealthy for Sensitive Groups Days` = col_double(),
##   `Unhealthy Days` = col_double(),
##   `Very Unhealthy Days` = col_double(),
##   `Hazardous Days` = col_double(),
##   `Max AQI` = col_double(),
##   `90th Percentile AQI` = col_double(),
##   `Median AQI` = col_double(),
##   `Days CO` = col_double(),
##   `Days NO2` = col_double(),
##   `Days Ozone` = col_double(),
##   `Days SO2` = col_double(),
##   `Days PM2.5` = col_double(),
##   `Days PM10` = col_double()
## )
```

```
##
## ── Column specification ────────────────────────────────────────────────────────
## cols(
##   .default = col_character(),
##   `Parameter Code` = col_double(),
##   POC = col_double(),
##   Latitude = col_double(),
##   Longitude = col_double(),
##   `Date Local` = col_date(format = ""),
##   `Observation Count` = col_double(),
##   `Observation Percent` = col_double(),
##   `Arithmetic Mean` = col_double(),
##   `1st Max Value` = col_double(),
##   `1st Max Hour` = col_double(),
##   AQI = col_double(),
##   `Method Code` = col_double(),
##   `Date of Last Change` = col_date(format = "")
## )
## i Use `spec()` for the full column specifications.
```

Visualization 1

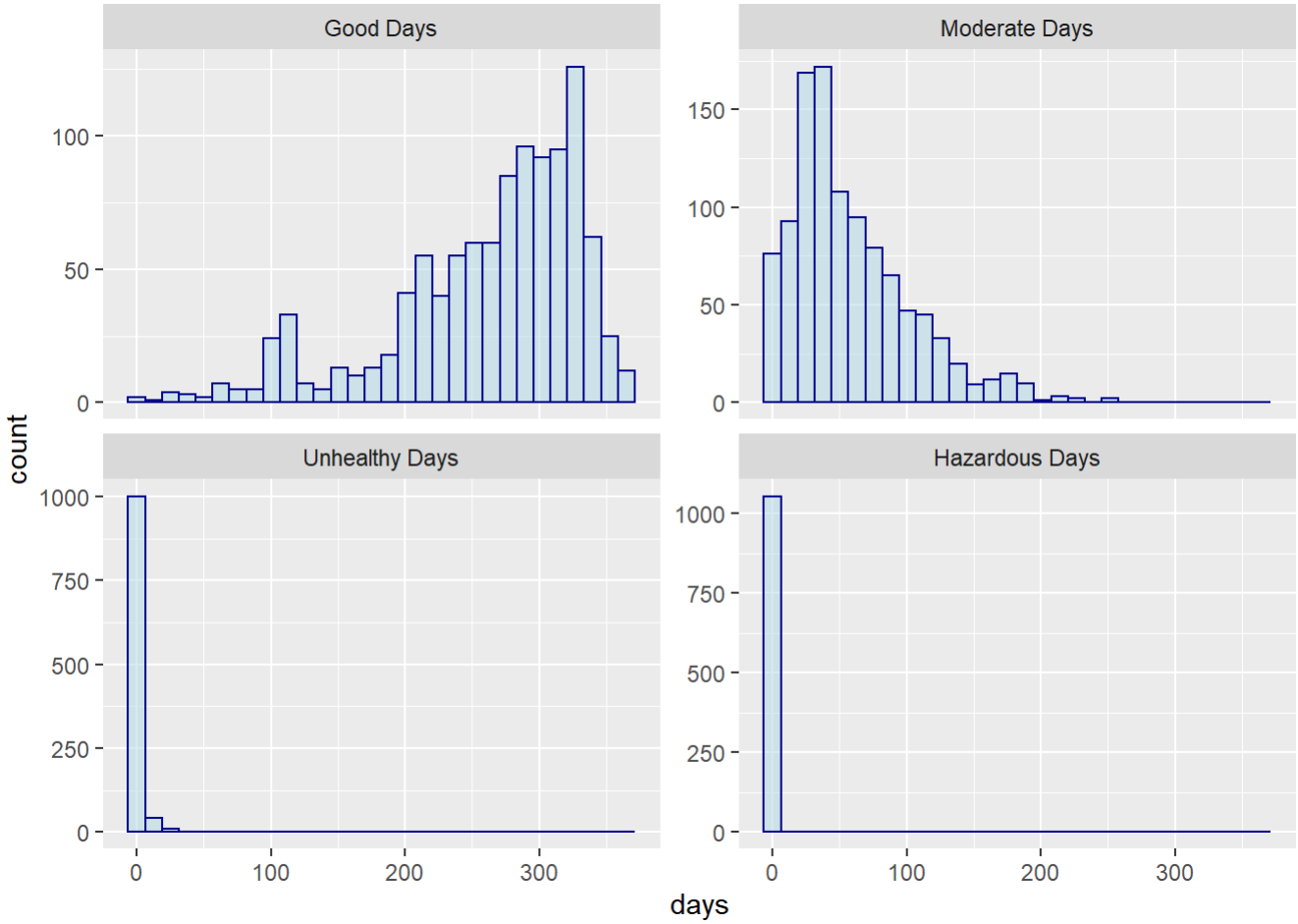
The below graph shows distributions of PM2.5 and AQI at station or site levels which is more detailed and includes local conditions of air pollution. The graphs show positive skewness with AQI concentrated around 50 and pm25 under 25. The values for both variables are quite high which rarely occured during the year.



Visualization 2

The below graphs shows level of air pollution in terms of four categories which are Good days,Moderate days,Unhealthy days and Hazardous days across counties. It can be inferred from the histograms that moderate days indicates a right skewness which suggests there were moderate days of air quality for most

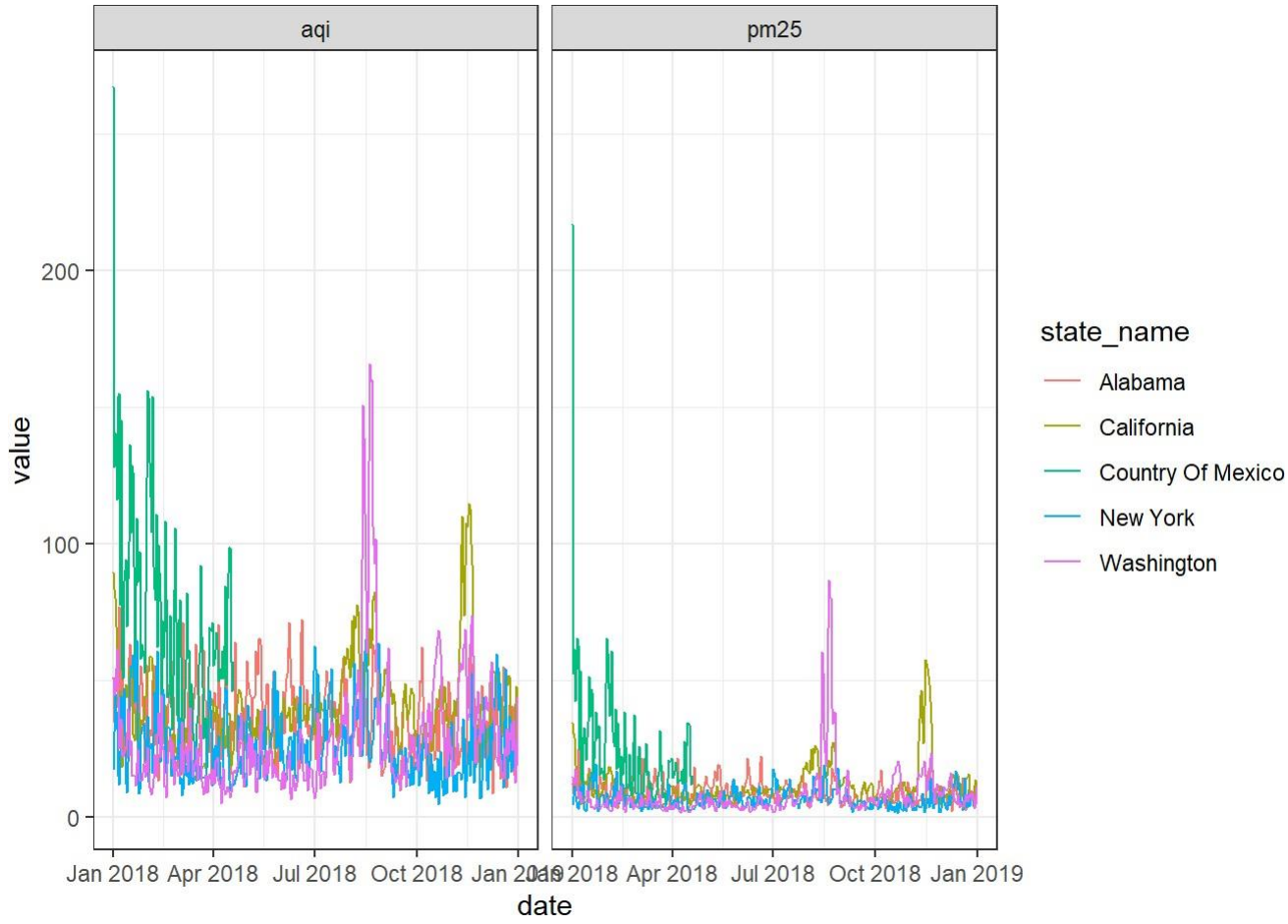
counties in 2018.The graph for Good days shows opposite skewness which means the air quality wasn't consistent that year and narrow distribution levels at 0 for unhealthy and hazardous days.



Visualization 3

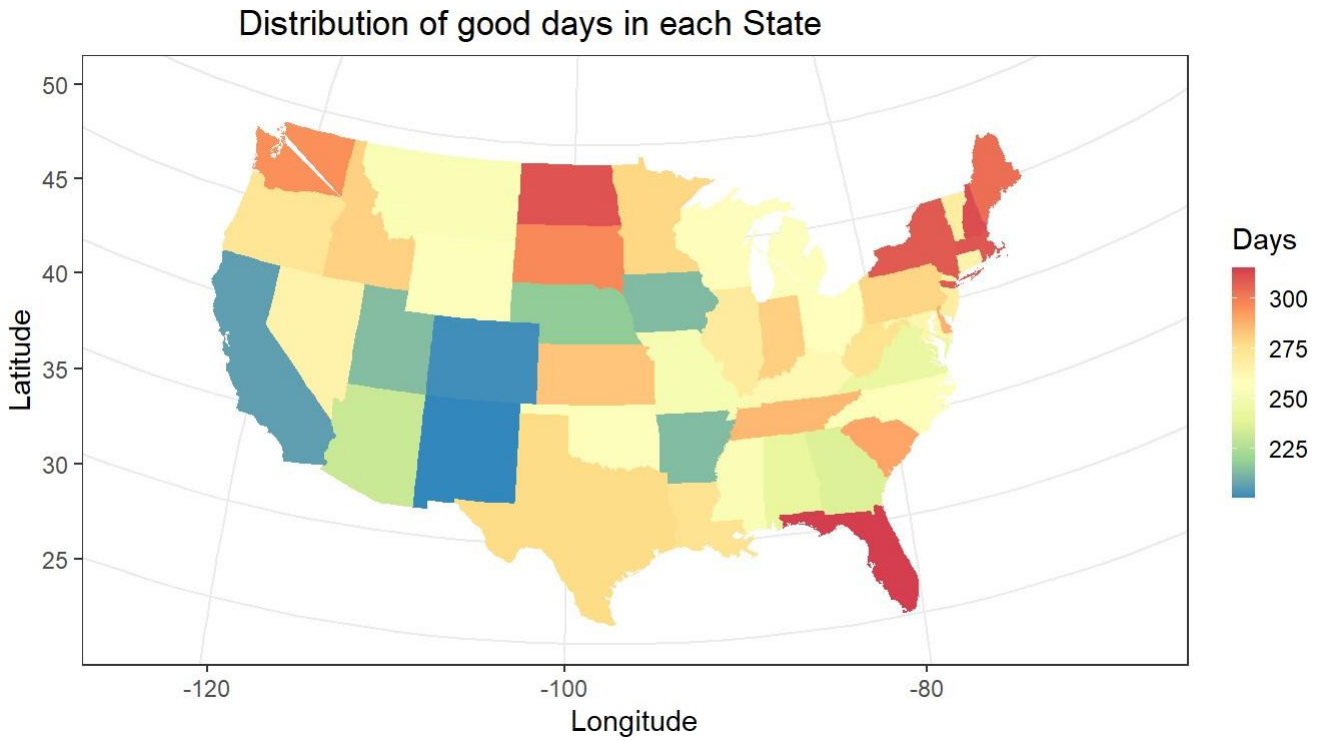
The below graph shows the temporal variation along with trend of PM 2.5 and AQI across five major states situated in various geographical locations of the country. It can be inferred from the plots that Country of Mexico which is in the south of USA experiences high levels of air pollution in the beginning of the year followed by Washington mid year and California towards the end which are towards the eastern and western region of America respectively.

```
## `summarise()` has grouped output by 'state_name'. You can override using the `.groups` arg
ument.
```



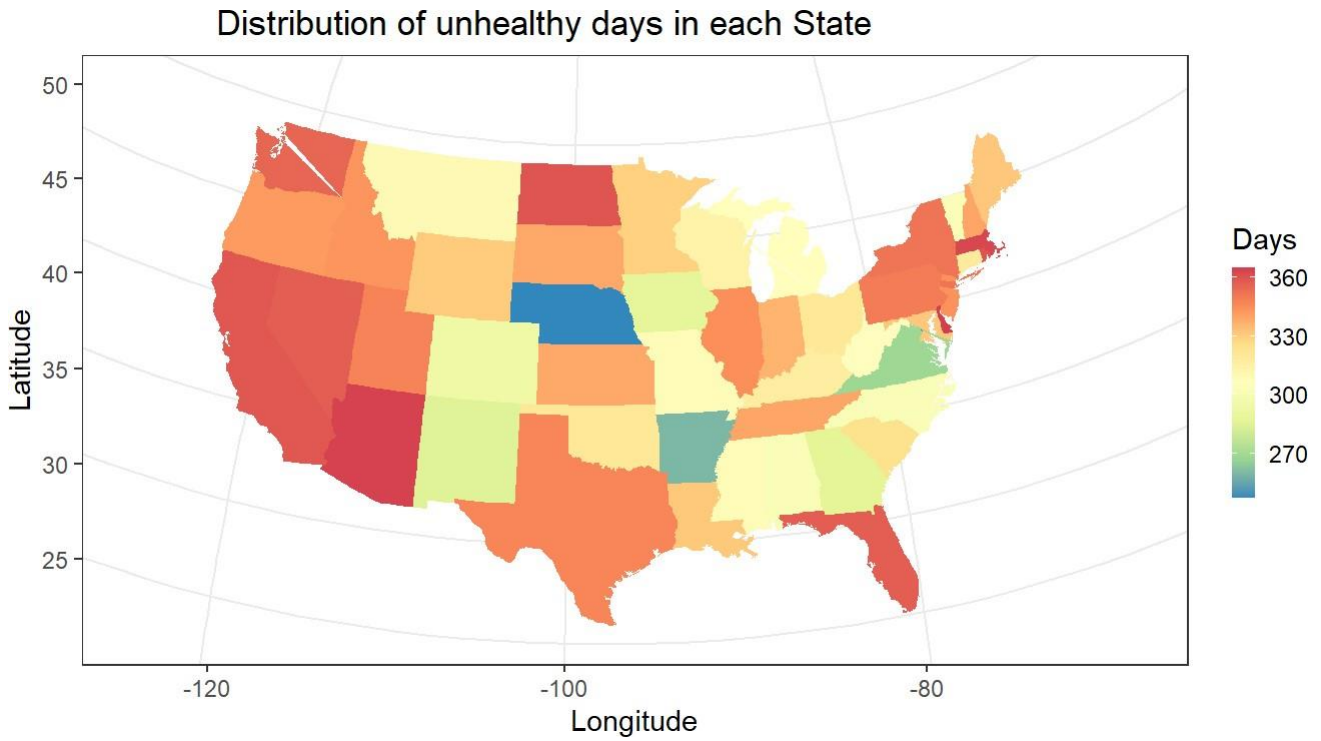
Visualization 4

The below graph shows the spatial variation of air quality in each state and the number of good air quality days for each state in 2018. The averages of the variable ‘Good Days’ are depicted by gradient colours in the map. The northern and eastern parts of the country seem to have higher quality of air than other states.



Visualization 5

The below graph shows the spatial variation of air quality in each state and the number of unhealthy days for each state in 2018.The below map confirms the observations from Visualization 4 that the northern and eastern parts have better air quality than the other states.



Conclusion

On exploration of data and plotting graphs it can be interpreted that PM 2.5 and AQI have high temporal and spatial variations. The states in the north and east seem to have better air quality than the states in the west and south.