Midterm Project

Mohan Wang

2025-05-29

Data analysis

Huge amount of data with year, month, day, and hour.

START 3.1 avery year

```
data <- read.csv("PRSA_Data_Tiantan_20130301-20170228.csv")

daydata <- data %>%
  mutate(datetime = make_datetime(year, month, day, hour)) %>%
  select(No, datetime, PM2.5, PM10)

daily_pm25 <- daydata %>%
  mutate(date = as_date(datetime)) %>%
  group_by(date) %>%
  summarise(pm25_daily = mean(PM2.5, na.rm = TRUE)) %>%
  ungroup()

daily_pm10 <- daydata %>%
  mutate(date = as_date(datetime)) %>%
  group_by(date) %>%
  summarise(pm10_daily = mean(PM10, na.rm = TRUE)) %>%
  ungroup()


head(daily_pm25)
```

```
## # A tibble: 6 x 2
##   date       pm25_daily
##   <date>          <dbl>
## 1 2013-03-01       8.62
## 2 2013-03-02      31.7
## 3 2013-03-03      98.0
## 4 2013-03-04      22.3
## 5 2013-03-05     142.
## 6 2013-03-06     194.
```

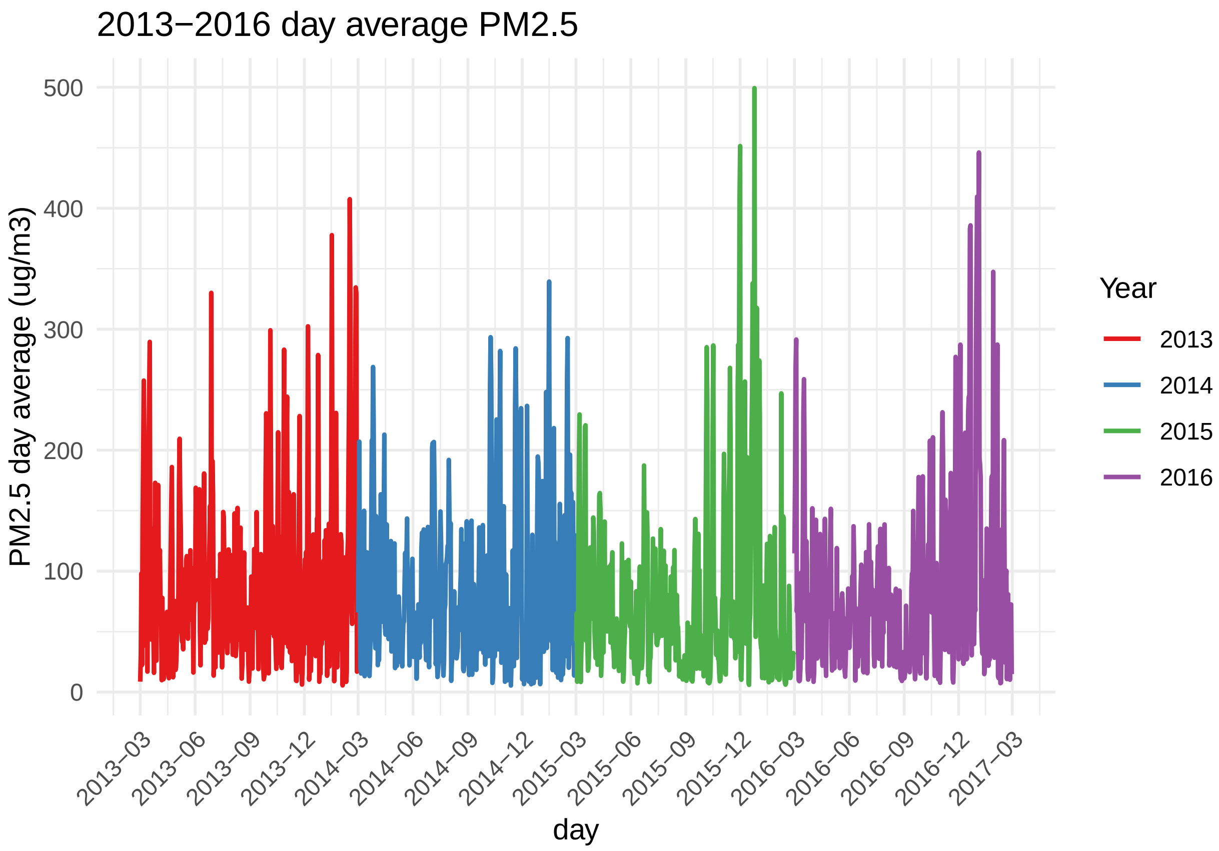
```
start_2013 <- ymd_hms("2013-03-01 00:00:00")
end_2013   <- ymd_hms("2014-03-01 00:00:00")

start_2014 <- ymd_hms("2014-03-01 00:00:00")
end_2014   <- ymd_hms("2015-03-01 00:00:00")

start_2015 <- ymd_hms("2015-03-01 00:00:00")
end_2015   <- ymd_hms("2016-03-01 00:00:00")

start_2016 <- ymd_hms("2016-03-01 00:00:00")
```

```r
end_2016   <- ymd_hms("2017-03-01 00:00:00")


daydata_2013 <- daydata %>%
  filter(datetime >= start_2013, datetime <  end_2013)

daydata_2014 <- daydata %>%
  filter(datetime >= start_2014, datetime <  end_2014)

daydata_2015 <- daydata %>%
  filter(datetime >= start_2015, datetime <  end_2015)

daydata_2016 <- daydata %>%
  filter(datetime >= start_2016, datetime <  end_2016)

daily_pm25_2013 <- daily_pm25 %>%
  filter(date >= as_date(start_2013), date < as_date(end_2013))

daily_pm25_2014 <- daily_pm25 %>%
  filter(date >= as_date(start_2014), date < as_date(end_2014))

daily_pm25_2015 <- daily_pm25 %>%
  filter(date >= as_date(start_2015), date < as_date(end_2015))

daily_pm25_2016 <- daily_pm25 %>%
  filter(date >= as_date(start_2016), date < as_date(end_2016))

daily_pm10_2013 <- daily_pm10 %>%
  filter(date >= as_date(start_2013), date < as_date(end_2013))

daily_pm10_2014 <- daily_pm10 %>%
  filter(date >= as_date(start_2014), date < as_date(end_2014))

daily_pm10_2015 <- daily_pm10 %>%
  filter(date >= as_date(start_2015), date < as_date(end_2015))

daily_pm10_2016 <- daily_pm10 %>%
  filter(date >= as_date(start_2016), date < as_date(end_2016))

all_pm25 <- bind_rows(
  daily_pm25_2013 %>% mutate(year = 2013),
  daily_pm25_2014 %>% mutate(year = 2014),
  daily_pm25_2015 %>% mutate(year = 2015),
  daily_pm25_2016 %>% mutate(year = 2016)
)
```

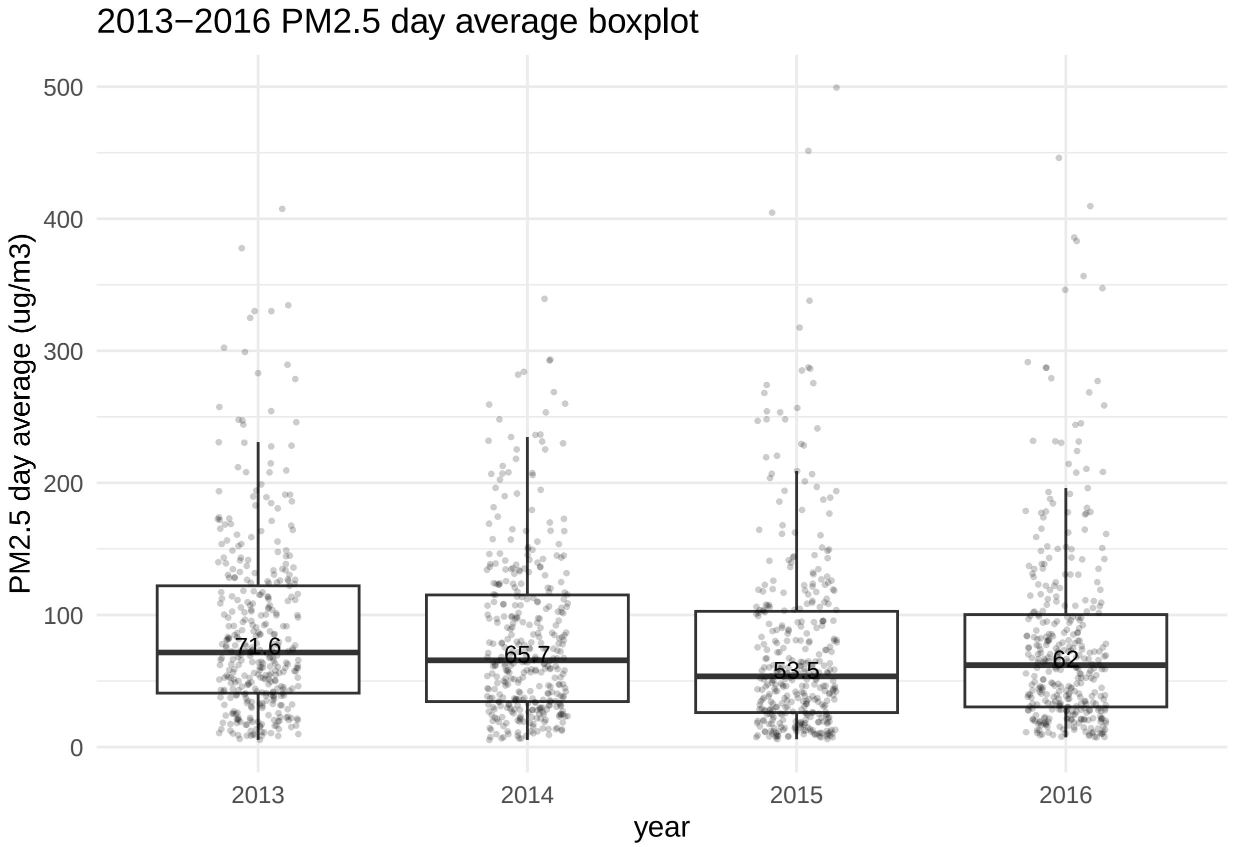
```r
all_pm10 <- bind_rows(
  daily_pm10_2013 %>% mutate(year=2013),
  daily_pm10_2014 %>% mutate(year=2014),
  daily_pm10_2015 %>% mutate(year=2015),
  daily_pm10_2016 %>% mutate(year=2016)
) %>%
  mutate(
    day = as.integer(date - as.Date(paste0(year, "-03-01")))
  ) %>%
  filter(day >= 0, day <= 365)
```

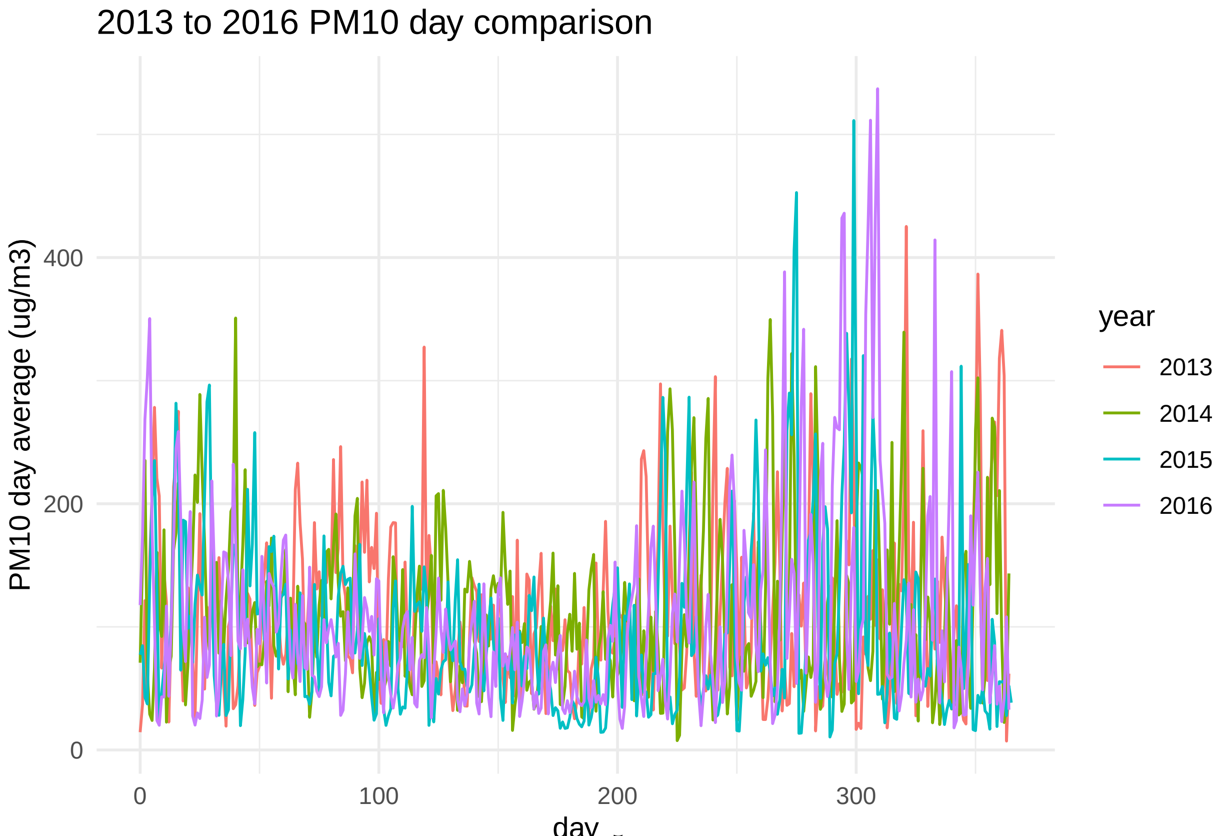
Average day data plot

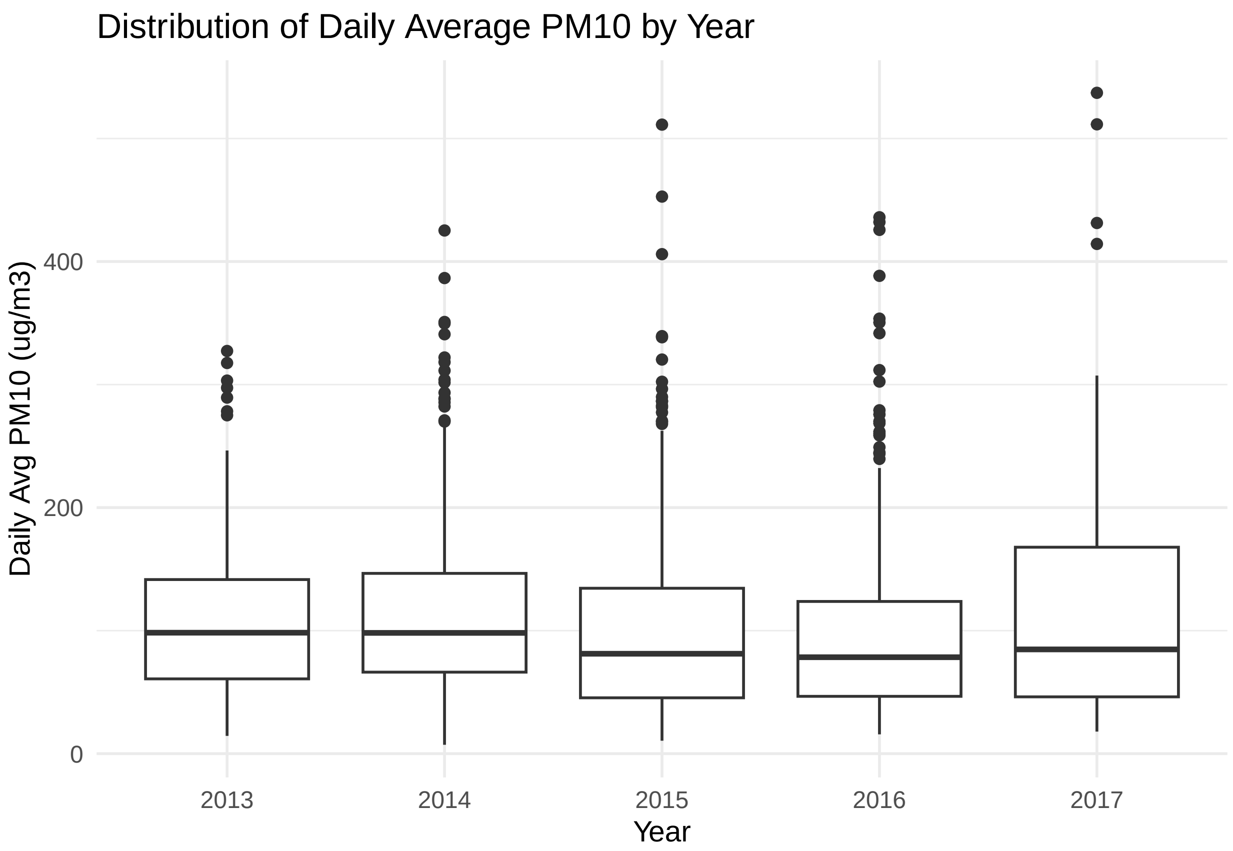
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Daily average data comparison

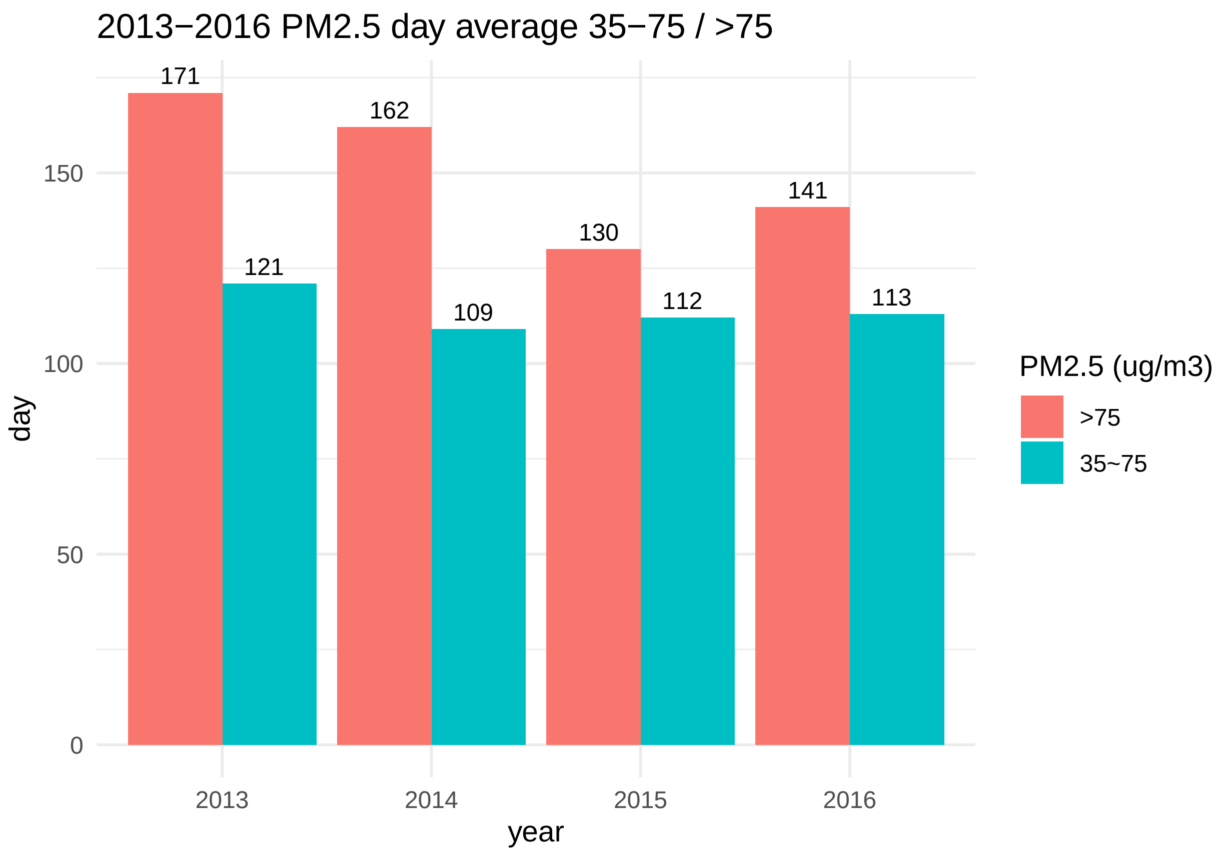

2013−2016 PM2.5 day average boxplot

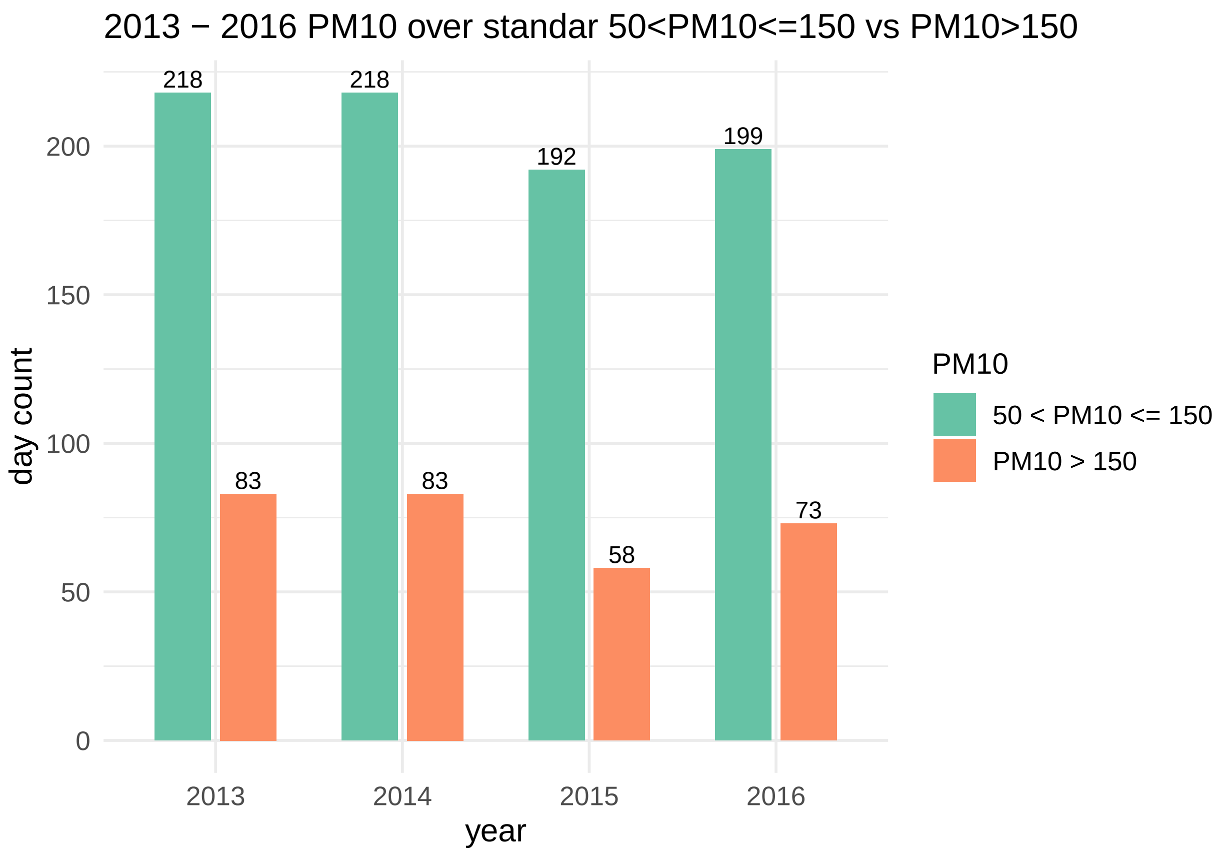

2013 to 2016 PM10 day comparison



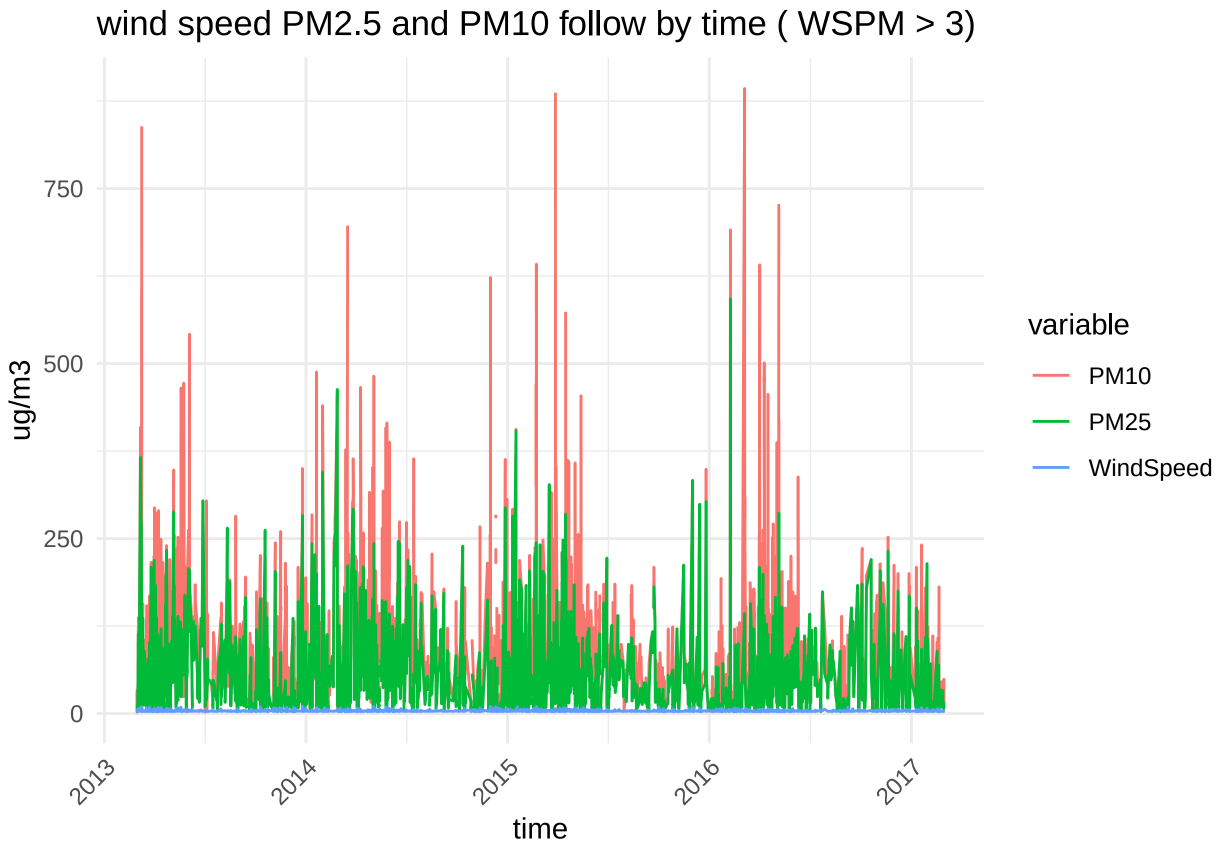
Daily average data over standard

China air quility standart in 2012 daily average value level 1 PM2.5 > 35 PM10 > 50 level 2 PM2.5 > 75 PM10 > 150


2013−2016 PM2.5 day average 35−75 / >75
171
121
162
109
130
112
141
113
150
100
50
0
day
2013
2014
2015
2016
year
PM2.5 (ug/m3)
>75
35~75

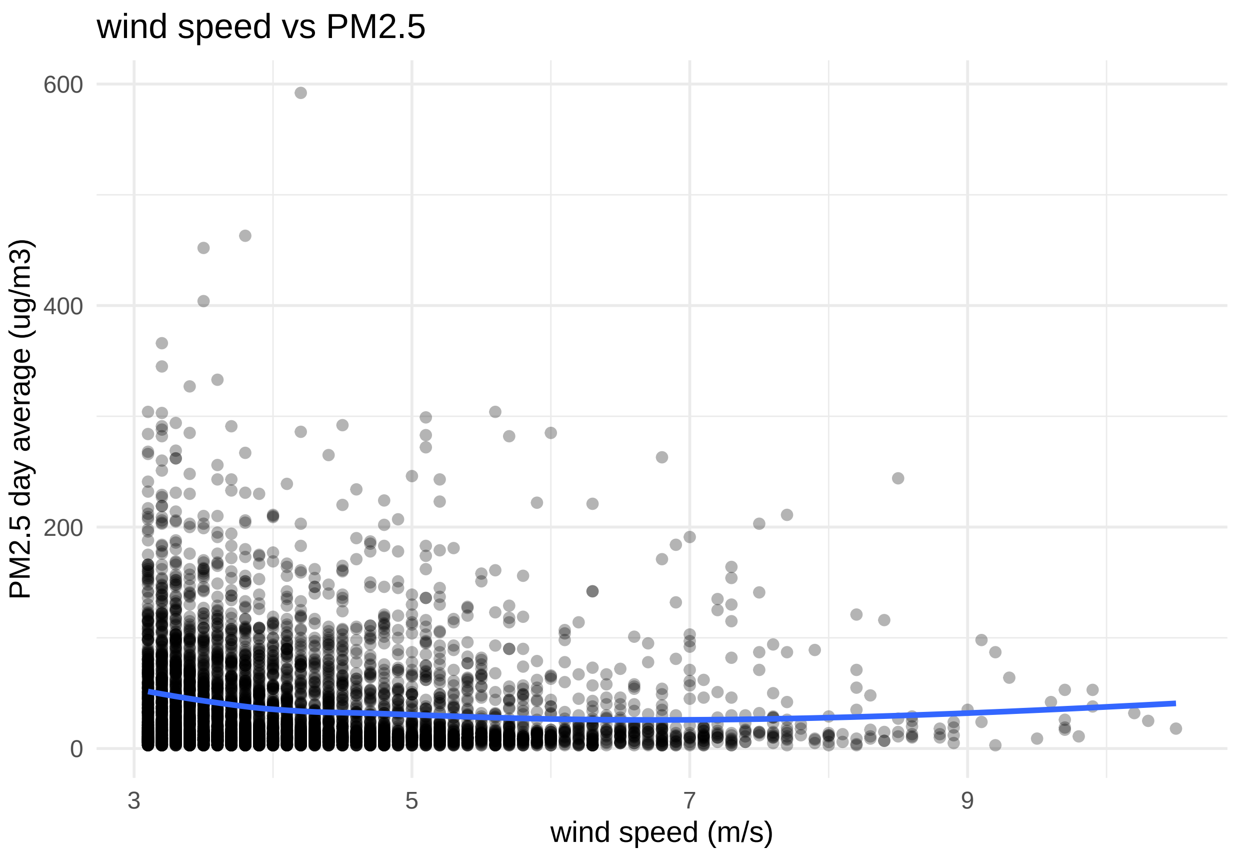

2013 − 2016 PM10 over standar 50<PM10<=150 vs PM10>150
day count
200
150
100
50
0
218
83
218
83
192
58
199
73
2013
2014
2015
2016
year
PM10
50 < PM10 <= 150
PM10 > 150

wind speed

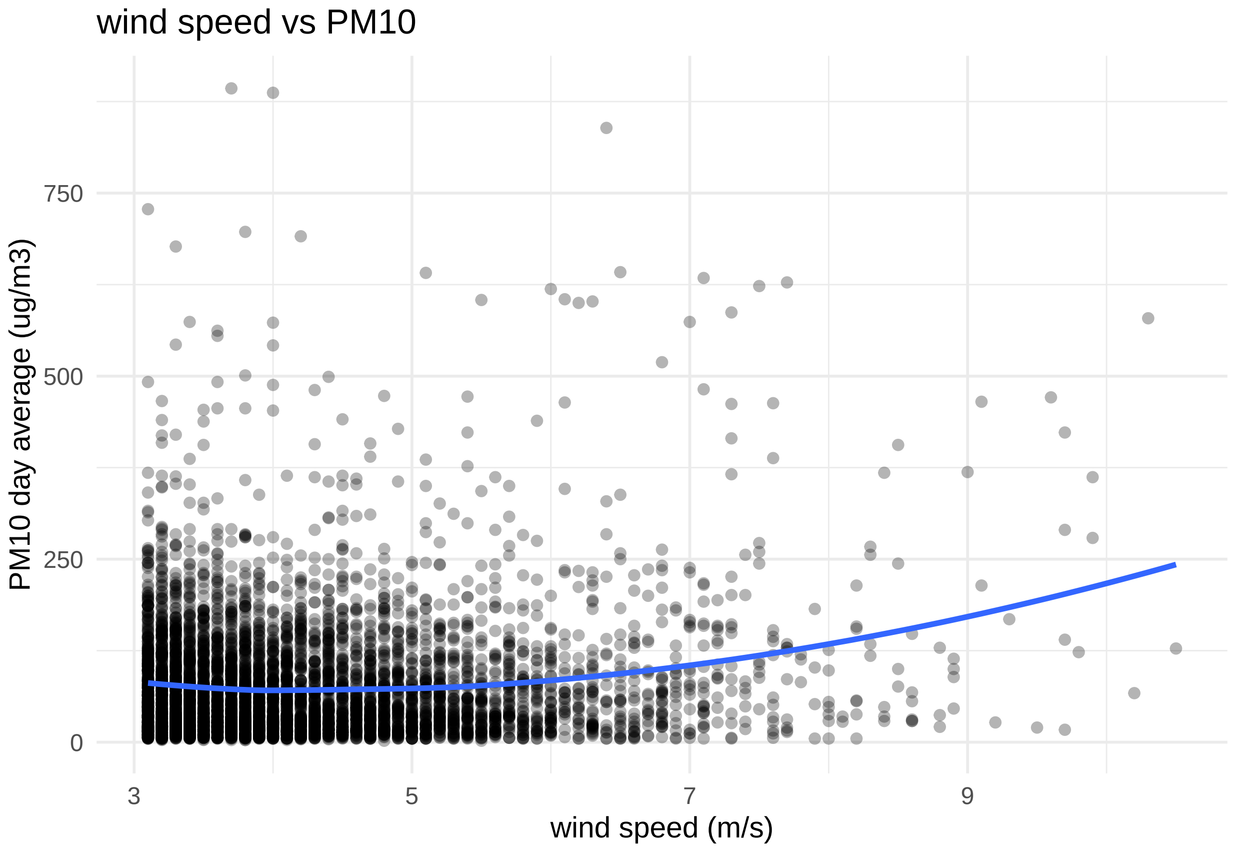


```
##      cor_PM25   cor_PM10
## 1 -0.1318061 0.09963948
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
## `geom_smooth()` using formula = 'y ~ x'
```


wind speed vs PM10
PM10 day average (ug/m3)
750
500
250
0
3
5
7
9
wind speed (m/s)