

HW_1

AUTHOR

Tyler Gallagher

Section 1

```
library(data.table)
library(tidyverse)
```

— Attaching core tidyverse packages — tidyverse 2.0.0 —

```
✓ dplyr      1.1.2    ✓ readr      2.1.4
✓ forcats    1.0.0    ✓ stringr    1.5.0
✓ ggplot2    3.4.3    ✓ tibble     3.2.1
✓ lubridate  1.9.2    ✓ tidyr      1.3.0
✓ purrr      1.0.2
```

— Conflicts — tidyverse_conflicts() —

```
* dplyr::between()    masks data.table::between()
* dplyr::filter()     masks stats::filter()
* dplyr::first()      masks data.table::first()
* lubridate::hour()   masks data.table::hour()
* lubridate::isoweek() masks data.table::isoweek()
* dplyr::lag()        masks stats::lag()
* dplyr::last()       masks data.table::last()
* lubridate::mday()   masks data.table::mday()
* lubridate::minute() masks data.table::minute()
* lubridate::month()  masks data.table::month()
* lubridate::quarter() masks data.table::quarter()
* lubridate::second() masks data.table::second()
* purrr::transpose()  masks data.table::transpose()
* lubridate::wday()   masks data.table::wday()
* lubridate::week()   masks data.table::week()
* lubridate::yday()   masks data.table::yday()
* lubridate::year()   masks data.table::year()
```

i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

```
epa22 <- fread("/Users/TylerGallagher13/Desktop/2022_EPA_Data.csv")
epa02 <- fread("/Users/TylerGallagher13/Desktop/2002_EPA_Data.csv")
```

```
dim(epa22)
```

```
[1] 57775    20
```

```
str(epa22)
```

```
Classes 'data.table' and 'data.frame': 57775 obs. of 20 variables:
 $ Date          : chr  "01/01/2022" "01/02/2022" "01/03/2022"
"01/04/2022" ...
 $ Source         : chr  "AQS" "AQS" "AQS" "AQS" ...
 $ Site ID        : int   60010007 60010007 60010007 60010007 60010007
60010007 60010007 60010007 60010007 60010007 ...
 $ POC            : int   3 3 3 3 3 3 3 3 3 3 ...
 $ Daily Mean PM2.5 Concentration: num 12.7 13.9 7.1 3.7 4.2 3.8 2.3 6.9 13.6 11.2 ...
 $ UNITS          : chr  "ug/m3 LC" "ug/m3 LC" "ug/m3 LC" "ug/m3 LC" ...
 $ DAILY_AQI_VALUE : int   52 55 30 15 18 16 10 29 54 47 ...
 $ Site Name      : chr  "Livermore" "Livermore" "Livermore" "Livermore"
...
 $ 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_DESC : chr  "PM2.5 - Local Conditions" "PM2.5 - Local
Conditions" "PM2.5 - Local Conditions" "PM2.5 - Local Conditions" ...
 $ CBSA_CODE        : int 41860 41860 41860 41860 41860 41860 41860 41860
41860 41860 ...
 $ CBSA_NAME         : chr  "San Francisco-Oakland-Hayward, CA" "San
San Francisco-Oakland-Hayward, CA" "San Francisco-Oakland-Hayward, CA" "San Francisco-
Oakland-Hayward, CA" ...
 $ STATE_CODE        : int 6 6 6 6 6 6 6 6 6 6 ...
 $ STATE             : chr  "California" "California" "California"
"California" ...
 $ COUNTY_CODE       : int 1 1 1 1 1 1 1 1 1 1 ...
 $ COUNTY            : chr  "Alameda" "Alameda" "Alameda" "Alameda" ...
 $ SITE_LATITUDE     : num 37.7 37.7 37.7 37.7 37.7 ...
 $ SITE_LONGITUDE    : num -122 -122 -122 -122 -122 ...
- attr(*, ".internal.selfref")=<externalptr>
```

```
head(epa22)
```

	Date	Source	Site ID	POC	Daily Mean PM2.5 Concentration	UNITS
1:	01/01/2022	AQS	60010007	3	12.7	ug/m3 LC
2:	01/02/2022	AQS	60010007	3	13.9	ug/m3 LC
3:	01/03/2022	AQS	60010007	3	7.1	ug/m3 LC
4:	01/04/2022	AQS	60010007	3	3.7	ug/m3 LC
5:	01/05/2022	AQS	60010007	3	4.2	ug/m3 LC
6:	01/06/2022	AQS	60010007	3	3.8	ug/m3 LC

	DAILY_AQI_VALUE	Site Name	DAILY_OBS_COUNT	PERCENT_COMPLETE
1:	52	Livermore	1	100
2:	55	Livermore	1	100
3:	30	Livermore	1	100
4:	15	Livermore	1	100
5:	18	Livermore	1	100
6:	16	Livermore	1	100

	AQS_PARAMETER_CODE	AQS_PARAMETER_DESC	CBSA_CODE
1:	88101	PM2.5 - Local Conditions	41860

2:	88101 PM2.5 - Local Conditions	41860
3:	88101 PM2.5 - Local Conditions	41860
4:	88101 PM2.5 - Local Conditions	41860
5:	88101 PM2.5 - Local Conditions	41860
6:	88101 PM2.5 - Local Conditions	41860

	CBSA_NAME	STATE_CODE	STATE	COUNTY_CODE	COUNTY
1:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
2:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
3:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
4:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
5:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
6:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda

	SITE_LATITUDE	SITE_LONGITUDE
1:	37.68753	-121.7842
2:	37.68753	-121.7842
3:	37.68753	-121.7842
4:	37.68753	-121.7842
5:	37.68753	-121.7842
6:	37.68753	-121.7842

```
tail(epa22)
```

	Date	Source	Site ID	POC	Daily Mean PM2.5 Concentration	UNITS
1:	12/01/2022	AQS	61131003	1	3.4 ug/m3	LC
2:	12/07/2022	AQS	61131003	1	3.8 ug/m3	LC
3:	12/13/2022	AQS	61131003	1	6.0 ug/m3	LC
4:	12/19/2022	AQS	61131003	1	34.8 ug/m3	LC
5:	12/25/2022	AQS	61131003	1	23.2 ug/m3	LC
6:	12/31/2022	AQS	61131003	1	1.0 ug/m3	LC

	DAILY_AQI_VALUE	Site Name	DAILY_OBS_COUNT	PERCENT_COMPLETE
1:	14	Woodland-Gibson Road	1	100
2:	16	Woodland-Gibson Road	1	100
3:	25	Woodland-Gibson Road	1	100
4:	99	Woodland-Gibson Road	1	100
5:	74	Woodland-Gibson Road	1	100
6:	4	Woodland-Gibson Road	1	100

	AQS_PARAMETER_CODE	AQS_PARAMETER_DESC	CBSA_CODE
1:	88101 PM2.5 - Local Conditions		40900
2:	88101 PM2.5 - Local Conditions		40900
3:	88101 PM2.5 - Local Conditions		40900
4:	88101 PM2.5 - Local Conditions		40900
5:	88101 PM2.5 - Local Conditions		40900
6:	88101 PM2.5 - Local Conditions		40900

	CBSA_NAME	STATE_CODE	STATE	COUNTY_CODE
1:	Sacramento--Roseville--Arden-Arcade, CA	6	California	113
2:	Sacramento--Roseville--Arden-Arcade, CA	6	California	113
3:	Sacramento--Roseville--Arden-Arcade, CA	6	California	113
4:	Sacramento--Roseville--Arden-Arcade, CA	6	California	113
5:	Sacramento--Roseville--Arden-Arcade, CA	6	California	113
6:	Sacramento--Roseville--Arden-Arcade, CA	6	California	113

	COUNTY	SITE_LATITUDE	SITE_LONGITUDE
1:	Yolo	38.66121	-121.7327
2:	Yolo	38.66121	-121.7327
3:	Yolo	38.66121	-121.7327
4:	Yolo	38.66121	-121.7327
5:	Yolo	38.66121	-121.7327
6:	Yolo	38.66121	-121.7327

```
summary(epa22$`Daily Mean PM2.5 Concentration`)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.200	4.200	7.000	8.574	10.900	302.500

A concentration below 0 does not seem possible. I will delete observations below 0 concentration.

```
epa22 <- epa22[epa22$`Daily Mean PM2.5 Concentration`>0]
summary(epa22$`Daily Mean PM2.5 Concentration`)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.100	4.200	7.000	8.626	10.900	302.500

Now, I will repeat these things for the other data table.

```
dim(epa02)
```

```
[1] 15976    20
```

```
str(epa02)
```

```
Classes 'data.table' and 'data.frame': 15976 obs. of 20 variables:
 $ Date          : chr  "01/05/2002" "01/06/2002" "01/08/2002"
"01/11/2002" ...
 $ Source         : chr  "AQS" "AQS" "AQS" "AQS" ...
 $ Site ID       : int   60010007 60010007 60010007 60010007 60010007
60010007 60010007 60010007 60010007 60010007 ...
 $ POC           : int   1 1 1 1 1 1 1 1 1 1 ...
 $ Daily Mean PM2.5 Concentration: num  25.1 31.6 21.4 25.9 34.5 41 29.3 15 18.8 37.9 ...
 $ UNITS         : chr   "ug/m3 LC" "ug/m3 LC" "ug/m3 LC" "ug/m3 LC" ...
 $ DAILY_AQI_VALUE : int   78 92 71 80 98 115 87 57 65 107 ...
 $ Site Name     : chr   "Livermore" "Livermore" "Livermore" "Livermore"
...
 $ 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 88101 ...
 $ AQS_PARAMETER_DESC : chr   "PM2.5 - Local Conditions" "PM2.5 - Local
Conditions" "PM2.5 - Local Conditions" "PM2.5 - Local Conditions" ...
 $ CBSA_CODE       : int  41860 41860 41860 41860 41860 41860 41860 41860 41860
41860 41860 ...
```

```

$ CBSA_NAME          : chr  "San Francisco-Oakland-Hayward, CA" "San
San Francisco-Oakland-Hayward, CA" "San Francisco-Oakland-Hayward, CA" "San Francisco-
Oakland-Hayward, CA" ...
$ STATE_CODE         : int  6 6 6 6 6 6 6 6 6 6 ...
$ STATE              : chr  "California" "California" "California"
"California" ...
$ COUNTY_CODE        : int  1 1 1 1 1 1 1 1 1 1 ...
$ COUNTY             : chr  "Alameda" "Alameda" "Alameda" "Alameda" ...
$ SITE_LATITUDE      : num  37.7 37.7 37.7 37.7 37.7 ...
$ SITE_LONGITUDE     : num  -122 -122 -122 -122 -122 ...
- attr(*, ".internal.selfref")=<externalptr>

```

```
head(epa02)
```

	Date	Source	Site ID	POC	Daily Mean PM2.5 Concentration	UNITS
1:	01/05/2002	AQS	60010007	1	25.1	ug/m3 LC
2:	01/06/2002	AQS	60010007	1	31.6	ug/m3 LC
3:	01/08/2002	AQS	60010007	1	21.4	ug/m3 LC
4:	01/11/2002	AQS	60010007	1	25.9	ug/m3 LC
5:	01/14/2002	AQS	60010007	1	34.5	ug/m3 LC
6:	01/17/2002	AQS	60010007	1	41.0	ug/m3 LC

	DAILY_AQI_VALUE	Site Name	DAILY_OBS_COUNT	PERCENT_COMPLETE
1:	78	Livermore	1	100
2:	92	Livermore	1	100
3:	71	Livermore	1	100
4:	80	Livermore	1	100
5:	98	Livermore	1	100
6:	115	Livermore	1	100

	AQS_PARAMETER_CODE	AQS_PARAMETER_DESC	CBSA_CODE
1:	88101	PM2.5 - Local Conditions	41860
2:	88101	PM2.5 - Local Conditions	41860
3:	88101	PM2.5 - Local Conditions	41860
4:	88101	PM2.5 - Local Conditions	41860
5:	88101	PM2.5 - Local Conditions	41860
6:	88101	PM2.5 - Local Conditions	41860

	CBSA_NAME	STATE_CODE	STATE	COUNTY_CODE	COUNTY
1:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
2:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
3:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
4:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
5:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
6:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda

	SITE_LATITUDE	SITE_LONGITUDE
1:	37.68753	-121.7842
2:	37.68753	-121.7842
3:	37.68753	-121.7842
4:	37.68753	-121.7842
5:	37.68753	-121.7842
6:	37.68753	-121.7842

```
tail(epa02)
```

	Date	Source	Site ID	POC	Daily Mean PM2.5 Concentration	UNITS
1:	12/10/2002	AQS	61131003	1	15	ug/m3 LC
2:	12/13/2002	AQS	61131003	1	15	ug/m3 LC
3:	12/22/2002	AQS	61131003	1	1	ug/m3 LC
4:	12/25/2002	AQS	61131003	1	23	ug/m3 LC
5:	12/28/2002	AQS	61131003	1	5	ug/m3 LC
6:	12/31/2002	AQS	61131003	1	6	ug/m3 LC

	DAILY_AQI_VALUE	Site Name	DAILY_OBS_COUNT	PERCENT_COMPLETE
1:	57	Woodland-Gibson Road	1	100
2:	57	Woodland-Gibson Road	1	100
3:	4	Woodland-Gibson Road	1	100
4:	74	Woodland-Gibson Road	1	100
5:	21	Woodland-Gibson Road	1	100
6:	25	Woodland-Gibson Road	1	100

	AQS_PARAMETER_CODE	AQS_PARAMETER_DESC	CBSA_CODE
1:	88101	PM2.5 - Local Conditions	40900
2:	88101	PM2.5 - Local Conditions	40900
3:	88101	PM2.5 - Local Conditions	40900
4:	88101	PM2.5 - Local Conditions	40900
5:	88101	PM2.5 - Local Conditions	40900
6:	88101	PM2.5 - Local Conditions	40900

	CBSA_NAME	STATE_CODE	STATE	COUNTY_CODE
1:	Sacramento--Roseville--Arden-Arcade, CA	6	California	113
2:	Sacramento--Roseville--Arden-Arcade, CA	6	California	113
3:	Sacramento--Roseville--Arden-Arcade, CA	6	California	113
4:	Sacramento--Roseville--Arden-Arcade, CA	6	California	113
5:	Sacramento--Roseville--Arden-Arcade, CA	6	California	113
6:	Sacramento--Roseville--Arden-Arcade, CA	6	California	113

	COUNTY	SITE_LATITUDE	SITE_LONGITUDE
1:	Yolo	38.66121	-121.7327
2:	Yolo	38.66121	-121.7327
3:	Yolo	38.66121	-121.7327
4:	Yolo	38.66121	-121.7327
5:	Yolo	38.66121	-121.7327
6:	Yolo	38.66121	-121.7327

```
summary(epa02$`Daily Mean PM2.5 Concentration`)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	7.00	12.00	16.12	20.50	104.30

Since the minimum concentration is zero, this is a logical reading. We can move on to next steps.

Section 2

```
library(dplyr)
epa22 <- epa22 %>%
  mutate(year = "2022")
```

```
epa02 <- epa02 %>%
  mutate(year = "2002")
```

```
epa0222 <- rbind(epa02, epa22)
```

```
head(epa0222)
```

	Date	Source	Site ID	POC	Daily Mean PM2.5 Concentration	UNITS
1:	01/05/2002	AQS	60010007	1	25.1	ug/m3 LC
2:	01/06/2002	AQS	60010007	1	31.6	ug/m3 LC
3:	01/08/2002	AQS	60010007	1	21.4	ug/m3 LC
4:	01/11/2002	AQS	60010007	1	25.9	ug/m3 LC
5:	01/14/2002	AQS	60010007	1	34.5	ug/m3 LC
6:	01/17/2002	AQS	60010007	1	41.0	ug/m3 LC

	DAILY_AQI_VALUE	Site Name	DAILY_OBS_COUNT	PERCENT_COMPLETE
1:	78	Livermore	1	100
2:	92	Livermore	1	100
3:	71	Livermore	1	100
4:	80	Livermore	1	100
5:	98	Livermore	1	100
6:	115	Livermore	1	100

	AQS_PARAMETER_CODE	AQS_PARAMETER_DESC	CBSA_CODE
1:	88101	PM2.5 - Local Conditions	41860
2:	88101	PM2.5 - Local Conditions	41860
3:	88101	PM2.5 - Local Conditions	41860
4:	88101	PM2.5 - Local Conditions	41860
5:	88101	PM2.5 - Local Conditions	41860
6:	88101	PM2.5 - Local Conditions	41860

	CBSA_NAME	STATE_CODE	STATE	COUNTY_CODE	COUNTY
1:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
2:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
3:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
4:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
5:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda
6:	San Francisco-Oakland-Hayward, CA	6	California	1	Alameda

	SITE_LATITUDE	SITE_LONGITUDE	year
1:	37.68753	-121.7842	2002
2:	37.68753	-121.7842	2002
3:	37.68753	-121.7842	2002
4:	37.68753	-121.7842	2002
5:	37.68753	-121.7842	2002
6:	37.68753	-121.7842	2002

```
summary(epa0222$year)
```

```
Length      Class      Mode
73416 character character
```

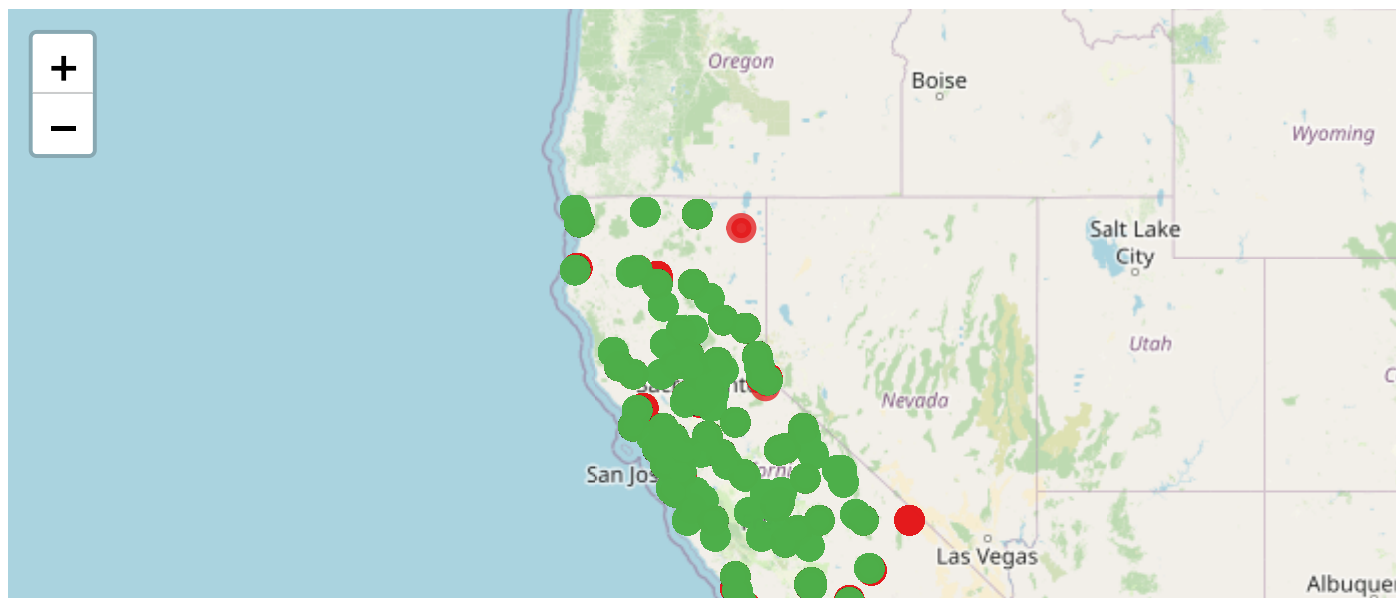
```
names(epa0222)[names(epa0222) == "Daily Mean PM2.5 Concentration"] <- "PM2.5"
summary(epa0222$PM2.5)
```

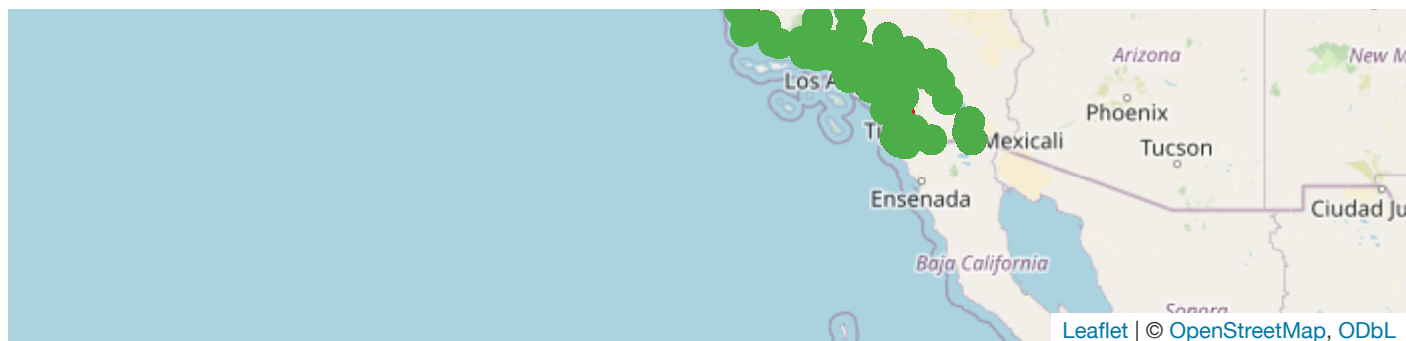
```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.00   4.60   7.70  10.26  12.50  302.50
```

Section 3

```
library(leaflet)
library(RColorBrewer)
```

```
my_map <- leaflet(data = epa0222) %>%
  addTiles()
color_mapping <- colorFactor(
  palette = "Set1",
  domain = epa0222$year
)
my_map <- my_map %>%
  addCircleMarkers(
    lng = ~SITE_LONGITUDE,
    lat = ~SITE_LATITUDE,
    color = ~color_mapping(year),
    radius = 5,
    fillOpacity = 0.7
  )
my_map
```





There are monitoring sites throughout California. There is a region in the Southeast of the state that is poorly covered by weather monitoring devices. There is a heavy concentration along the coastline, particularly in Los Angeles and San Francisco.

Section 4

In section 1, I already addressed the issue of implausible values for PM2.5 concentration (ones that were negative).

```
summary(epa0222$PM2.5)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	4.60	7.70	10.26	12.50	302.50

There do not appear to be any missing NA values.

Section 5

First, we will create summary statistics and exploratory plots for the state level.

```
table(epa0222$STATE)
```

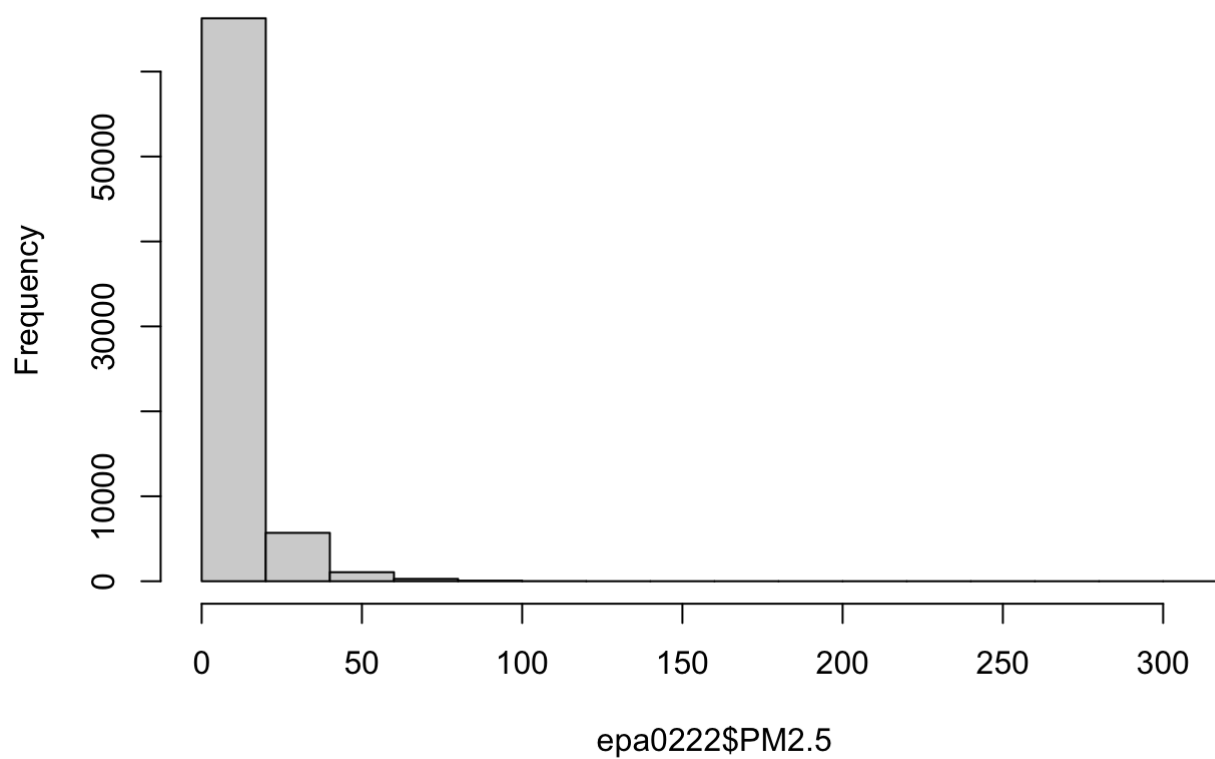
```
California
73416
```

```
summary(epa0222$PM2.5)
```

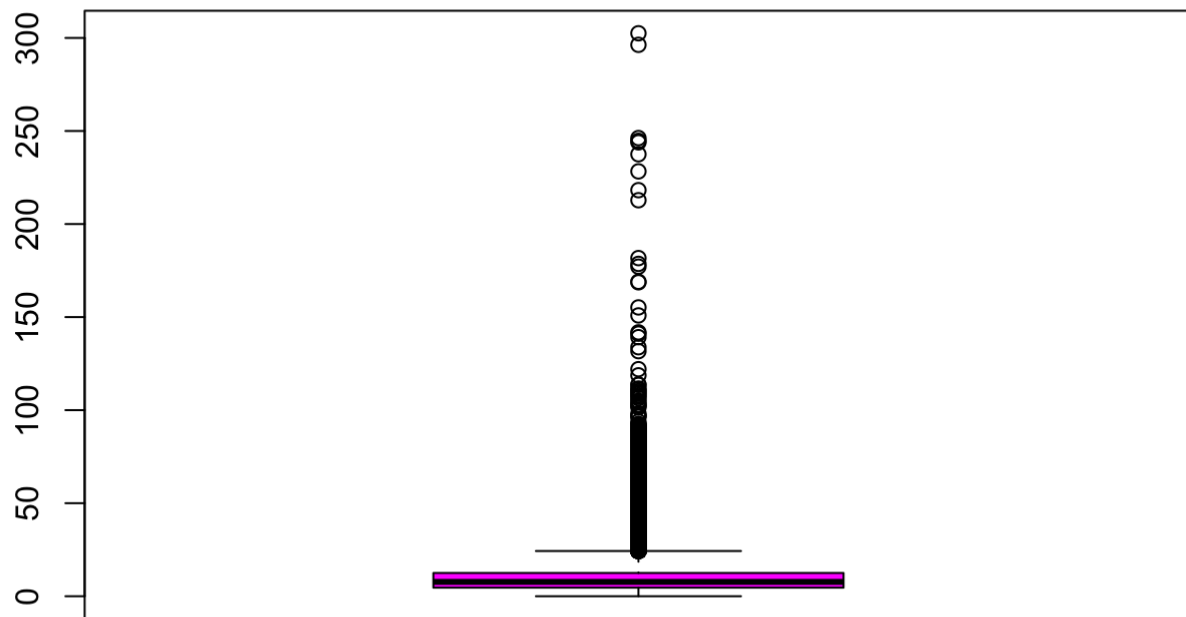
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	4.60	7.70	10.26	12.50	302.50

```
hist(epa0222$PM2.5)
```

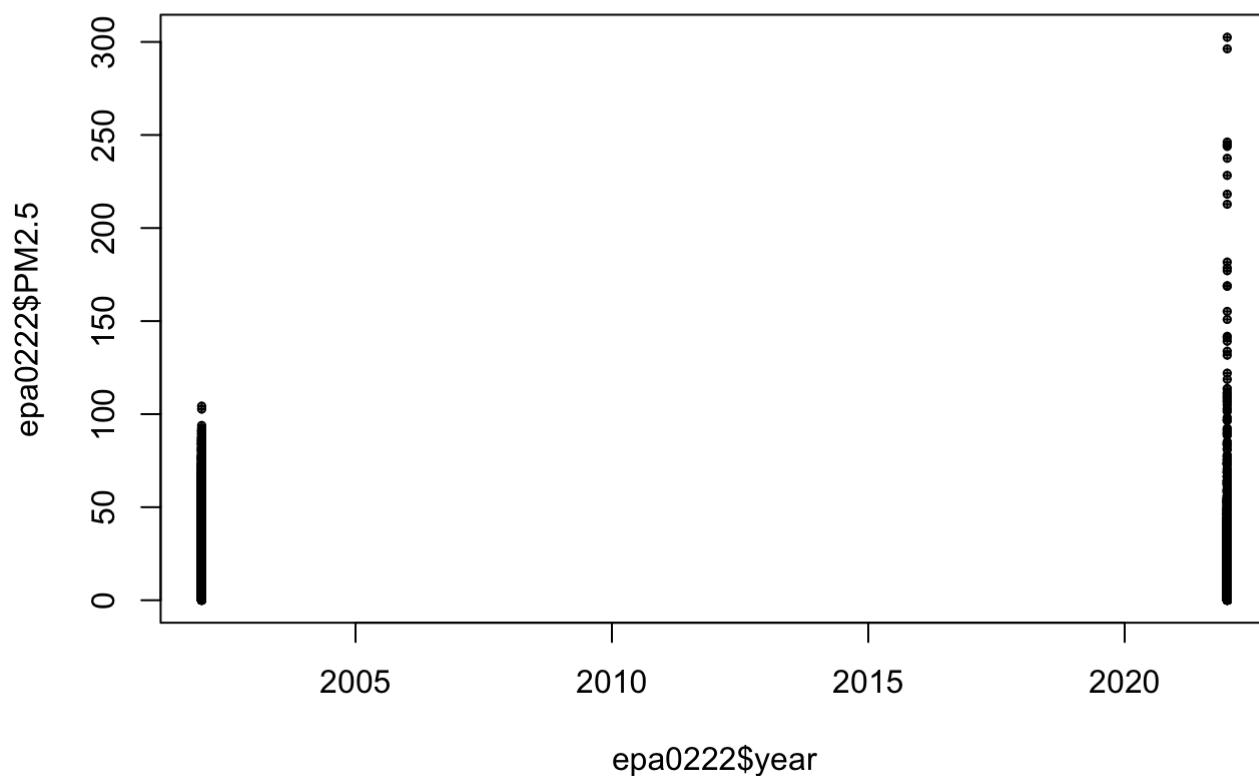
Histogram of epa0222\$PM2.5



```
boxplot(epa0222$PM2.5, col = "magenta")
```



```
plot(epa0222$year, epa0222$PM2.5, pch=10, cex=0.5)
```



It appears that in California generally, the data is right-skewed, meaning most days have a lower mean PM2.5 level, but there are many outliers at higher levels. There is a daily mean of 10.2, though highest readings are above 300. This trend is particularly accentuated in the boxplot. The scatterplot shows that there are many higher individual mean days in 2022 as opposed to 2002.

Now, we will create summary statistics and exploratory plots for the county level.

```
table(epa0222$COUNTY)
```

Alameda	Butte	Calaveras	Colusa	Contra Costa
1989	1580	414	496	1097
Del Norte	El Dorado	Fresno	Glenn	Humboldt
562	436	3533	328	175
Imperial	Inyo	Kern	Kings	Lake
1743	2157	4113	804	122
Los Angeles	Madera	Marin	Mariposa	Mendocino
7003	360	573	868	821
Merced	Modoc	Mono	Monterey	Nevada
808	2	805	1221	1232
Orange	Placer	Plumas	Riverside	Sacramento
1350	1837	1451	5192	3360
San Benito	San Bernardino	San Diego	San Francisco	San Joaquin
594	3730	4476	546	1158

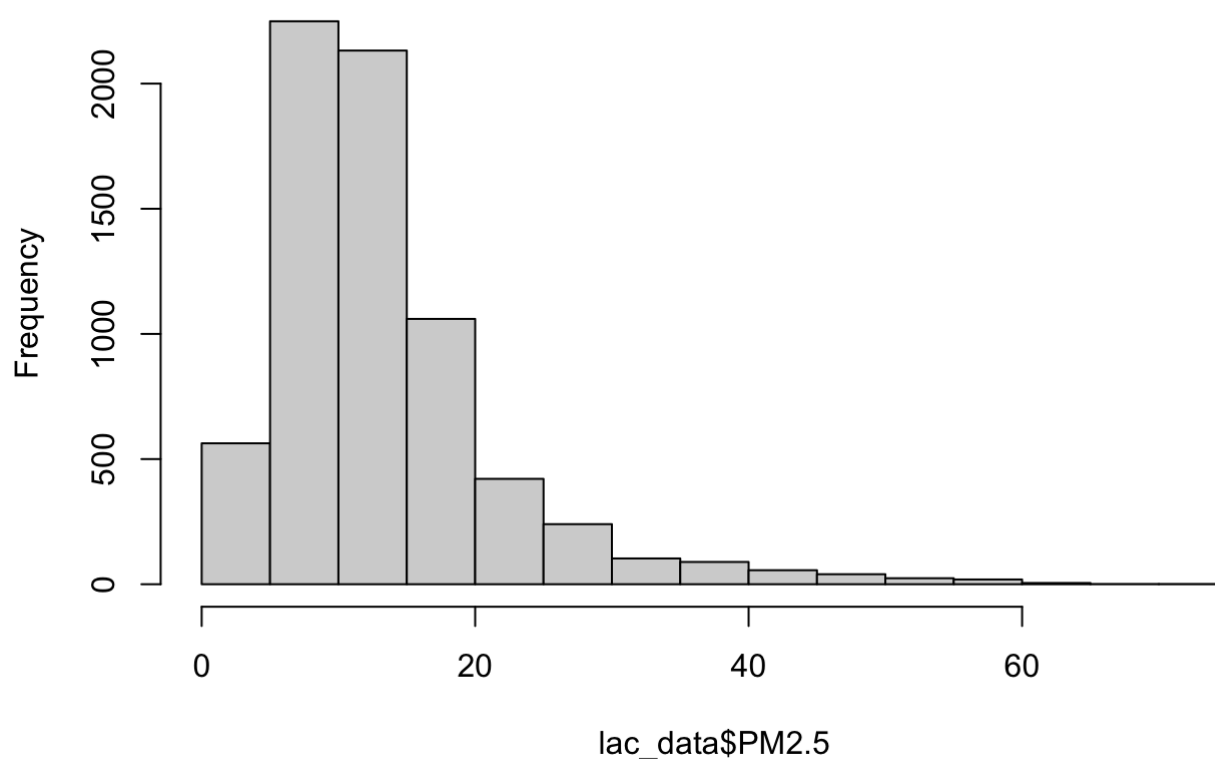
San Luis Obispo	San Mateo	Santa Barbara	Santa Clara	Santa Cruz
1598	449	1672	1651	752
Shasta	Siskiyou	Solano	Sonoma	Stanislaus
752	805	816	439	954
Sutter	Tehama	Trinity	Tulare	Ventura
829	347	489	1719	2691
Yolo				
517				

```
lac_data <- subset(epa0222, COUNTY == "Los Angeles")
summary(lac_data$PM2.5)
```

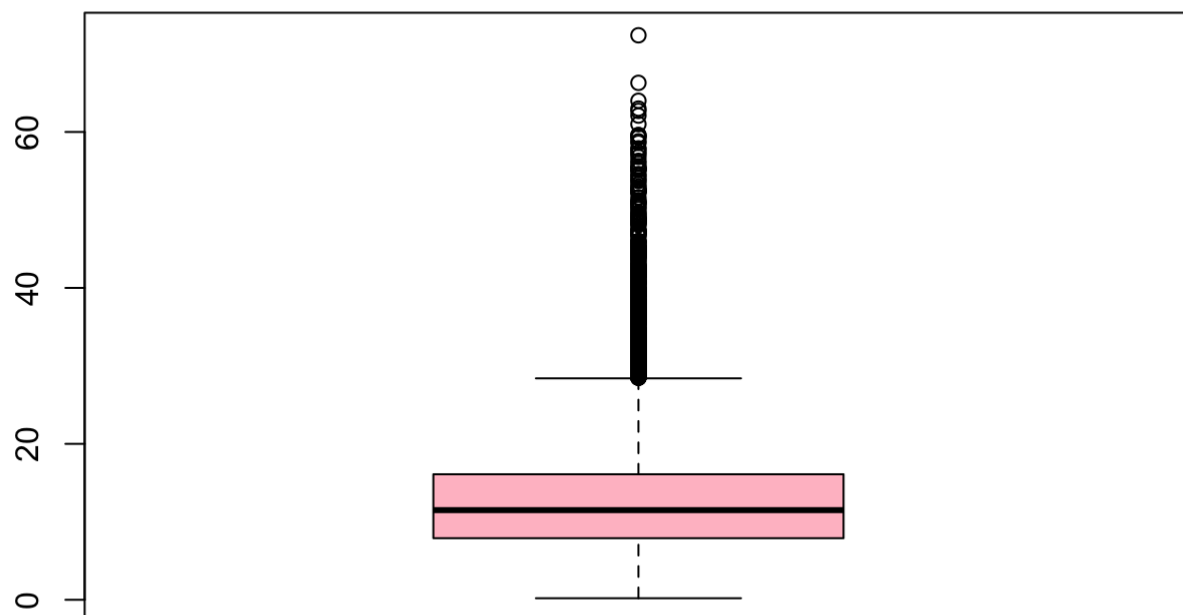
```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.20   7.90   11.50   13.37  16.10   72.40
```

```
hist(lac_data$PM2.5)
```

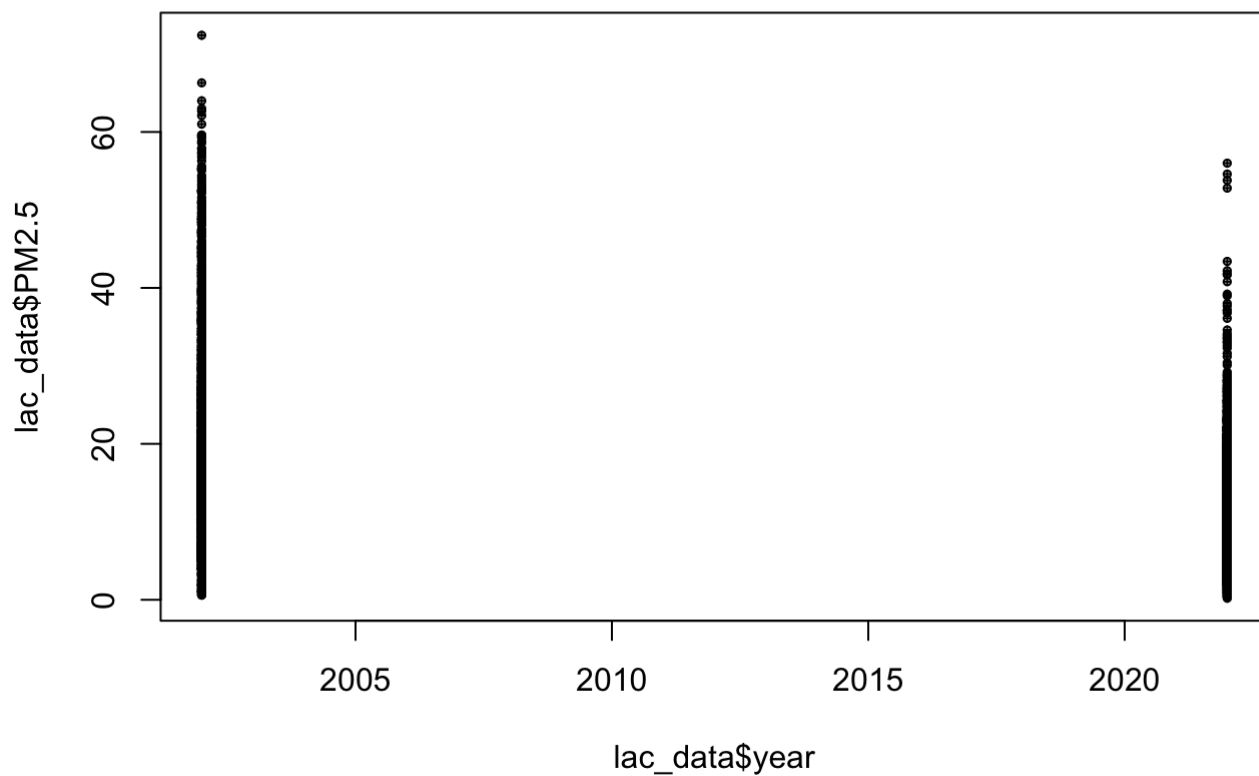
Histogram of lac_data\$PM2.5



```
boxplot(lac_data$PM2.5, col = "pink")
```



```
plot(lac_data$year, lac_data$PM2.5, pch=10, cex=0.5)
```



It appears that in Los Angeles county generally, the data is right-skewed, meaning most days have a lower mean PM2.5 level, but there are many outliers at higher levels. Mean PM2.5 level is 13.2 with highest numbers above 70. This trend is particularly accentuated in the boxplot. The scatterplot shows that the highest mean days of PM2.5 are in 2002 as opposed to 2020, which is different from the trends at the overall California level.

Now, we will create summary statistics and plots for the site in Los Angeles.

```
table(epa0222$'Site Name')
```

```

      823
29 Palms      414
3425 N FIRST ST, FRESNO 499
      Alpine    379
Alturas-Fourth St.      2
      Anaheim  1123
Aqua Tibia Wilderness

```

	168
Arroyo Grande CDF	
	359
Atascadero	
	350
Atascadero (original site)	
	116
Auburn-Atwood	
	350
Azusa	
	415
Bakersfield-Airport (Planz)	
	231
Bakersfield-California	
	2008
Bakersfield-Golden / M St	
	421
Banning Airport	
	365
Big Bear	
	450
Bishop Tribe EM0	
	354
Bliss SP	
	113
Brawley-220 Main Street	
	358
Brawley-401 Main St.	
	85
Burbank	
	122
Calexico-Ethel Street	
	613
Camp Pendleton	
	357
Campo Indian Reservation	
	362
Carmel Valley	
	343
Chester-222 First Ave	
	289
Chico-East Avenue	
	410
Chico-Manzanita Ave.	
	120
Chula Vista	
	218
Clovis-Villa	
	443
Colfax-City Hall	
	358

Colusa-Sunrise Blvd
454
Compton
723
Concord
739
Corcoran-Patterson
443
Cortina Indian Rancheria
42
Crescent City-Crescent Elk School
332
Crestline
178
Davis-UCD Campus
346
Death Valley NP - Park Village
115
Donovan
360
Echo Summit
2
El Cajon
416
El Cajon - Lexington Elementary School
491
El Centro-9th Street
452
El Rio-Rio Mesa School #2
469
Escondido
355
Eureka I Street
59
Folsom-Natoma St.
689
Fontana
298
Fremont - Chapel Way
105
Fresno - Garland
542
Fresno-Foundry
363
Fresno-Pacific
443
Gilroy
349
Glendora
361
Goleta

	359
Grass Valley-Litton Building	
	406
Hanford-Irwin	
	361
Hollister	
	356
Hoover Wilderness	
	228
Huron	
	360
Indio	
	240
Jacobs	
	116
Joshua Tree NP – Black Rock	
	215
Kaiser Wilderness	
	195
Kearny Mesa	
	111
Keeler	
	641
King City 2	
	348
Lake Elsinore	
	360
Lake Tahoe Community College	
	111
Lakeport-Lakeport Blvd.	
	61
Lakeport-S. Main Street	
	61
Lancaster-Division Street	
	458
Laney College	
	360
Lassen Volcanic NP – Manzanita Lake Fire Station	
	210
Lava Beds National Monument	
	216
Lebec	
	489
Lebec-Peace Valley Road	
	93
Lebec-Peace Valley/Frazier Park Roads	
	106
Lee Vining	
	363
Lincoln-2885 Moore Road	
	357

Livermore
439
Lompoc H Street
358
Lone Pine Paiute-Shoshone Reservation
302
Long Beach (North)
411
Long Beach (South)
243
Long Beach-Route 710 Near Road
625
Los Angeles-North Main Street
1276
Lynwood
122
Madera-City
360
Mammoth
214
Manteca
340
Merced-Coffee
359
Merced-M St
449
Mesa2
358
Mira Loma (Van Buren)
778
Mission Viejo
227
Modesto-14th Street
598
Mojave
100
Mojave - CA 58 Business
355
Morongo Air Monitoring Station
370
North Hollywood (NOH0)
365
Oakland
365
Oakland West
359
Ojai - East Ojai Ave
360
Ontario Fire Station
111
Ontario-Route 60 Near Road

	705
Pala Airpad	
	347
Palm Springs	
	239
Paradise – Theater	
	350
Pasadena	
	241
Pechanga	
	401
Pico Rivera #2	
	118
Pinnacles NP – Southwest of East Entrance Station	
	238
Piru – Pacific	
	475
Pleasanton – Owens Ct	
	361
Point Reyes NS Ranger Station	
	213
Porterville	
	356
Portola	
	552
Portola–161 Nevada Street	
	100
Quincy–N Church Street	
	510
Red Bluff–Walnut St. District Office	
	347
Redding – Buckeye	
	54
Redding – Health Department	
	429
Redding – Toyon	
	59
Redwood City	
	449
Redwood NP	
	230
Reseda	
	602
Ridgecrest–California Ave	
	104
Ridgecrest–Ward	
	355
Riverside (Magnolia)	
	115
Roseville–N Sunrise Ave	
	417

Rubidoux
1377
Sacramento Health Department–Stockton Blvd.
154
Sacramento–1309 T Street
826
Sacramento–Bercut Drive
359
Sacramento–Del Paso Manor
988
Salinas 3
530
San Andreas–Gold Strike Road
414
San Bernardino
235
San Diego – Kearny Villa Rd.
172
San Diego – Sherman Elementary School
437
San Diego –Rancho Carmel Drive
119
San Diego–12th Ave
352
San Francisco
546
San Jose
217
San Jose – 4th St.
141
San Jose – Jackson
586
San Jose – Knox Avenue
358
San Lorenzo Valley Middle School
345
San Luis Obispo–Marsh St.
52
San Pablo
358
San Rafael
360
San Rafael Wilderness
187
Santa Barbara
349
Santa Clarita
365
Santa Cruz
407
Santa Maria

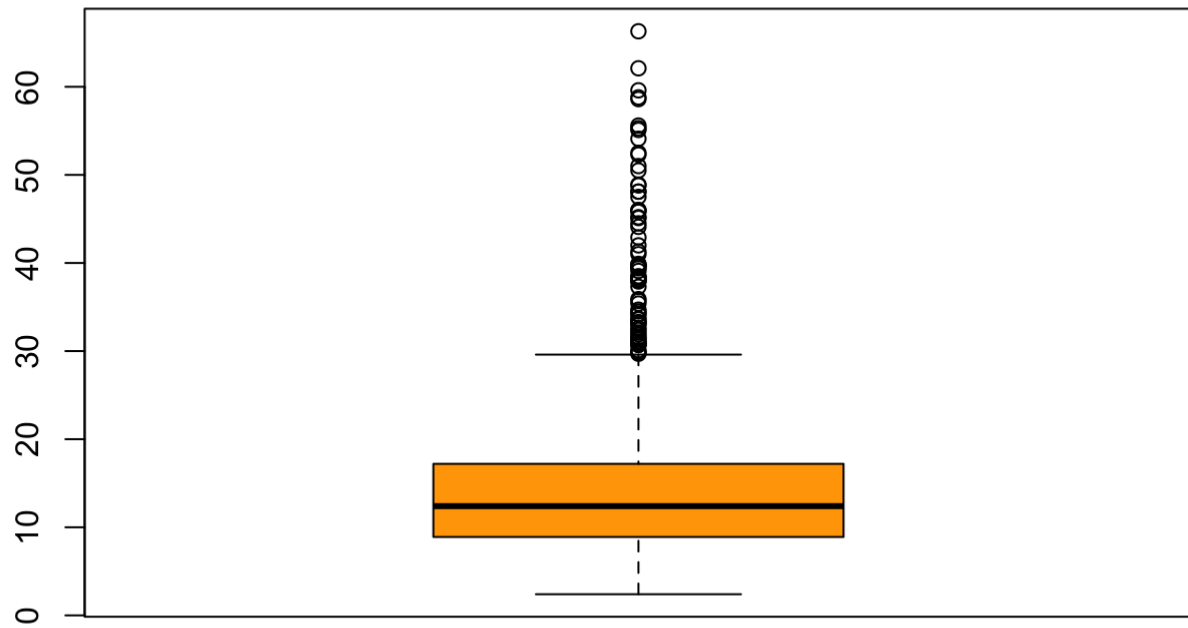
	60
Santa Rosa – 5th St	
	93
Sebastopol	
	346
Sequoia & Kings Canyon NPs – Ash Mountain	
	340
Sequoia NP–Ash Mountain	
	226
Signal Hill (LBSH)	
	289
Simi Valley–Cochran Street	
	873
SLO Roberto	
	363
Sloughhouse	
	344
South Lake Tahoe–Sandy Way	
	93
Stn.1 Big Pine Paiute site	
	325
Stockton – University Park	
	349
Stockton–Hazelton	
	124
Table Mountain Air Monitoring Site	
	339
Tahoe City–Fairway Drive	
	355
Temecula	
	365
Thousand Oaks	
	514
Torres Martinez Reservation	
	235
Tracy–Airport	
	345
TRAFFIC, RURAL PAVED ROAD	
	700
Tranquillity	
	349
Trinity	
	139
Truckee–Fire Station	
	826
Turlock	
	356
Ukiah–Library	
	470
Upland	
	365

Vallejo	
816	
Victorville-Park Avenue	
944	
Visalia-Church	
404	
Visalia-W. Ashland Avenue	
393	
Weaverville-Courthouse	
350	
White Mountain Research Center - Owens Valley Lab	
420	
Willits-Blosser Lane	
351	
Willows-Colusa Street	
328	
Woodland-Gibson Road	
171	
Yosemite NP - Turtleback Dome	
340	
Yosemite NP-Yosemite Village Visitor Center	
528	
Yreka	
589	
Yuba City	
829	

```
los_angeles_data <- subset(epa0222, `Site Name` == "Los Angeles-North Main Street")
summary(los_angeles_data$PM2.5)
```

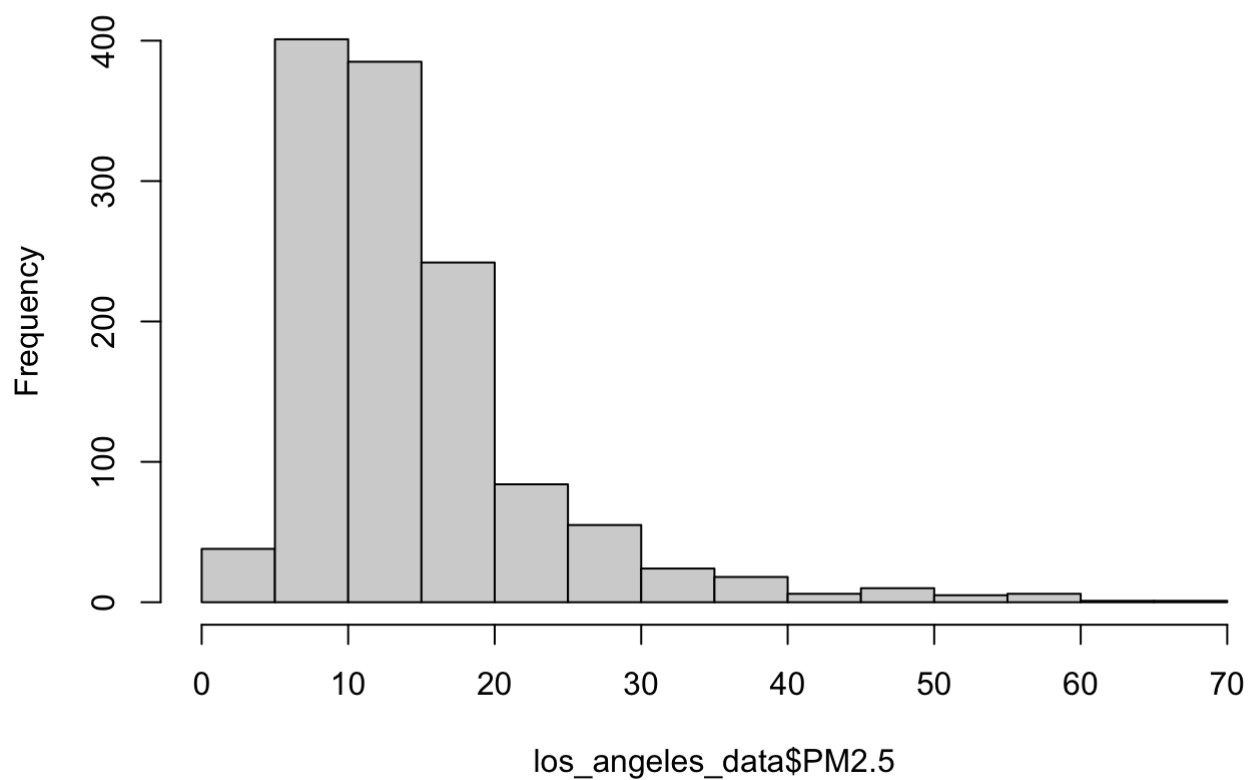
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.40	8.90	12.40	14.56	17.20	66.30

```
boxplot(los_angeles_data$PM2.5, col = "orange")
```

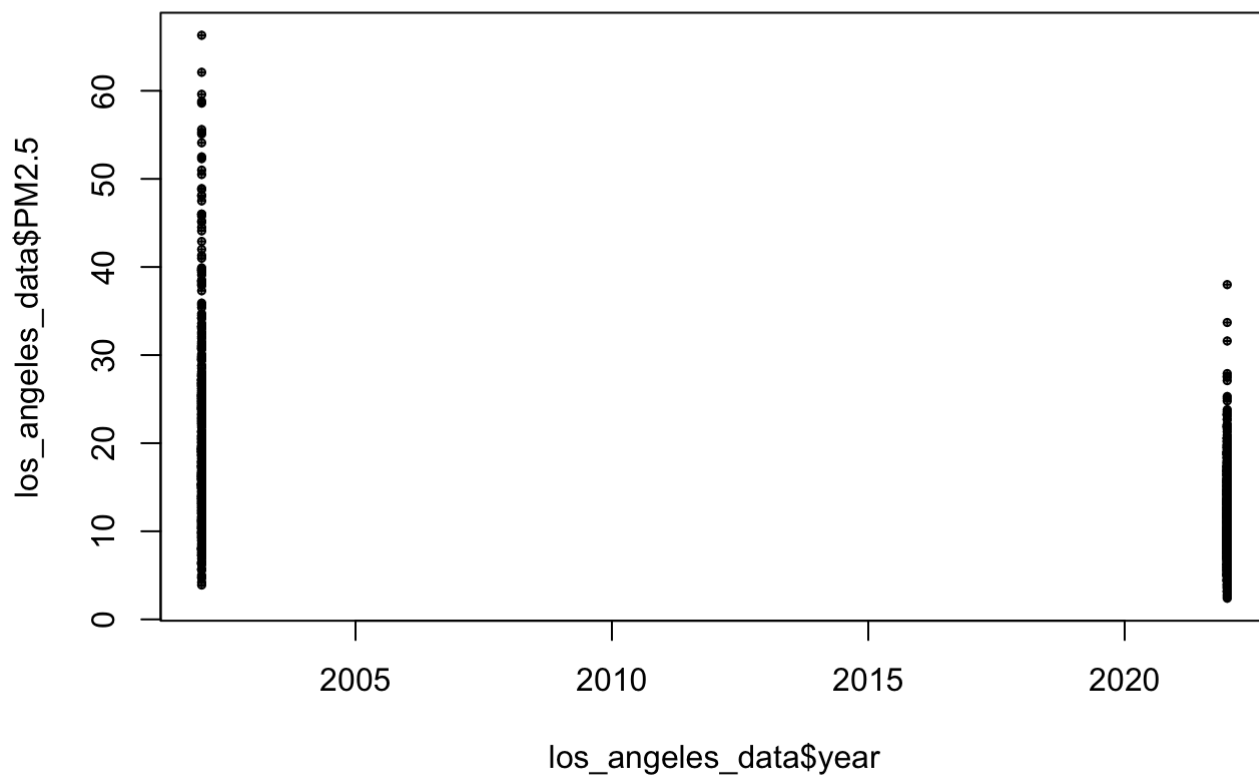


```
hist(los_angeles_data$PM2.5)
```


Histogram of los_angeles_data\$PM2.5



```
plot(los_angeles_data$year, los_angeles_data$PM2.5, pch=10, cex=0.5)
```



It appears that in Los Angeles city, the data is right-skewed, meaning most days have a lower mean PM2.5 level, but there are many outliers at higher levels. Mean PM2.5 level is 14.6 with highest numbers above 60. This trend is particularly accentuated in the boxplot. The scatterplot shows that the highest mean days of PM2.5 are in 2002 as opposed to 2020, which is different from the trends at the overall California level.