

Regional Analysis

Marc Los Huertos

10/1/2020

Introduction

Background

Air quality data varies seasonally, regionally, and with particular events. These create a distribution of values, some of which pose potential health effects.

Goals

We will collect EPA air quality data to compare with the local data we collect to assess if our data fall within the confidence intervals of the EPA database.

Rationale

Learning Outcomes

We have learned how to manipulate data in R (using the Rstudio) to analyze data before. In this case, we will

Grading – low stakes

As an assignment with relatively low stakes, 10 points, you will be assessed using the following criteria:

1. Did you collect 5 years of PM2.5 from a nearby air quality station.
2. Report the mean, standard deviation, and 95% confidence interval.
3. Plot 5-years of data and create a threshold line based on state and federal standards for PM2.5. Please include the following:
 - Rotated y-axis
 - Appropriate axis labels
 - Point graph, EPA or state health limits.
 - Estimate the mean, S.D. and 95% confidence intervals for each month.
4. Using the peer reviewed literature, discuss the implications of the results.

Assessing EPA Data

We will use the EPA website (<https://www.epa.gov/outdoor-air-quality-data/download-daily-data>) to collect the data

Upload to Rstudio Server Folder

Read Data

```
SCZ2020.csv <- "/home/CAMPUS/mwl04747/github/EJnPi/Air_Quality_Data_Analysis/ad_viz_plotval_data2020.csv"
SCZ2019.csv <- "/home/CAMPUS/mwl04747/github/EJnPi/Air_Quality_Data_Analysis/ad_viz_plotval_data2019.csv"
SCZ2018.csv <- "/home/CAMPUS/mwl04747/github/EJnPi/Air_Quality_Data_Analysis/ad_viz_plotval_data2018.csv"

SCZ2019 = read.csv(SCZ2019.csv); SCZ2020 = read.csv(SCZ2020.csv)
SCZ2018 = read.csv(SCZ2018.csv)
str(SCZ2020)[1:10]
SCZ=rbind(SCZ2018, SCZ2019, SCZ2020)

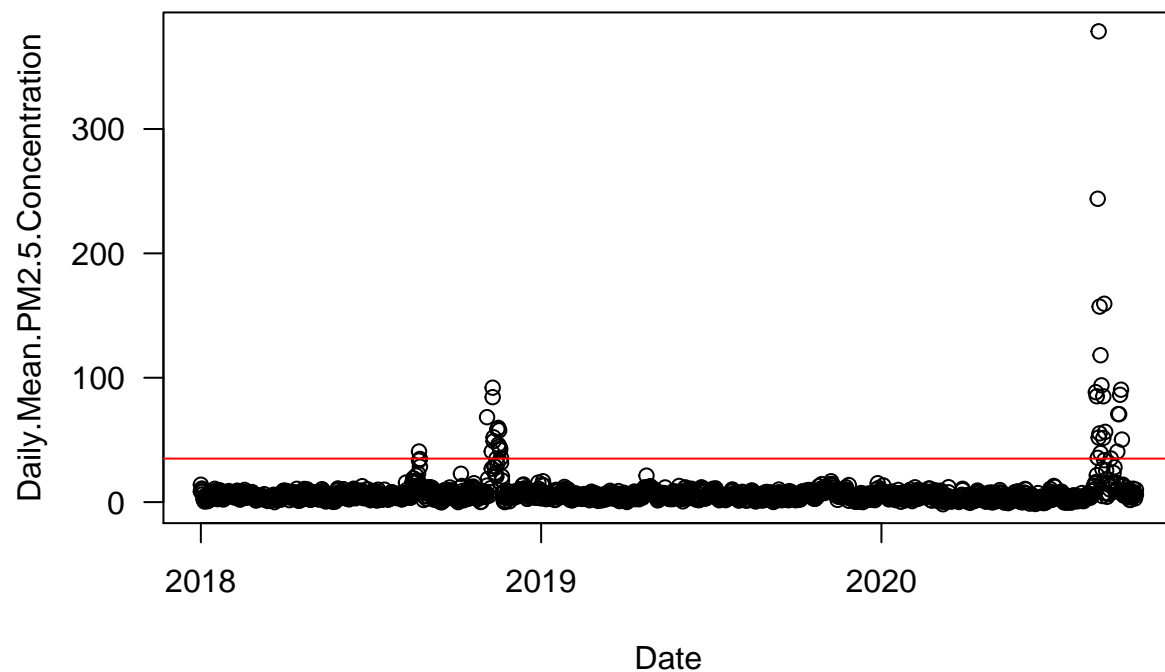
## Fix dates
Date.char = as.character(SCZ$Date)
# testing to make sure this works...
# as.Date(Date.char, format="%m/%d/%Y")
SCZ$Date = as.Date(Date.char, format="%m/%d/%Y")

str(SCZ)
names(SCZ)
```

Graphic Analysis

I have two stations together, might need to decide which one to focus on!

```
par(las=1)
plot(Daily.Mean.PM2.5.Concentration~Date, data=SCZ)
abline(h=35, col='red')
```



Estimate the Central Tendency for each Month

```
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(lubridate)
```

```
##
## Attaching package: 'lubridate'
## The following object is masked from 'package:base':
##
##   date
```

```
str(SCZ)
```

```
## 'data.frame':   1931 obs. of  20 variables:
##  $ Date          : Date, format: "2018-01-01" "2018-01-02" ...
```

```
## $ Source : Factor w/ 2 levels "AQS","AirNow": 1 1 1 1 1 1 1 1 1 1 ...
## $ Site.ID : int 60870007 60870007 60870007 60870007 60870007 60870007 60870007 60870007 60870007 60870007 ...
## $ POC : int 3 3 3 3 3 3 3 3 3 3 ...
## $ Daily.Mean.PM2.5.Concentration: num 8.5 9.9 6.2 4 3.6 4.2 5.7 4.5 1.2 6.4 ...
## $ UNITS : Factor w/ 1 level "ug/m3 LC": 1 1 1 1 1 1 1 1 1 1 ...
## $ DAILY_AQI_VALUE : int 35 41 26 17 15 18 24 19 5 27 ...
## $ Site.Name : Factor w/ 2 levels "San Lorenzo Valley Middle School",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ 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 : Factor w/ 1 level "PM2.5 - Local Conditions": 1 1 1 1 1 1 1 1 1 1 ...
## $ CBSA_CODE : int 42100 42100 42100 42100 42100 42100 42100 42100 42100 42100 ...
## $ CBSA_NAME : Factor w/ 1 level "Santa Cruz-Watsonville, CA": 1 1 1 1 1 1 1 1 1 1 ...
## $ STATE_CODE : int 6 6 6 6 6 6 6 6 6 6 ...
## $ STATE : Factor w/ 1 level "California": 1 1 1 1 1 1 1 1 1 1 ...
## $ COUNTY_CODE : int 87 87 87 87 87 87 87 87 87 87 ...
## $ COUNTY : Factor w/ 1 level "Santa Cruz": 1 1 1 1 1 1 1 1 1 1 ...
## $ SITE_LATITUDE : num 37 37 37 37 37 ...
## $ SITE_LONGITUDE : num -122 -122 -122 -122 -122 ...
```

```
df <- mutate(SCZ, Date = ymd(Date), Year = year(Date), Month = month(Date))
df <- group_by(df, Month, Year)
df <- summarise(df, result = mean(Daily.Mean.PM2.5.Concentration) )
```

```
df <- SCZ %>%
  mutate(Date = ymd(Date), Year = year(Date), Month = month(Date)) %>%
  group_by(Month) %>%
  summarise(mean = mean(Daily.Mean.PM2.5.Concentration), sd = sd(Daily.Mean.PM2.5.Concentration), N = 1)

#df$LCI95 =
```