Week 4 – Thursday – PREPARATION

Please try on your own before looking at recommended solutions!

Quantitative Reasoning 2020

Original data source

PM2.5 is an important air quality index (AQI): http://airnow.gov/index.cfm?action=aqibasics.aqi.The data for each city can be found here: https://www.airnow.gov/?city=New%20Delhi&country=IND

Air quality in five cities in India

This data set contains hourly PM2.5 measurements for the month of June, 2016, in 4 cities in India. PM2.5 is very small ("fine") particulate matter – see https://blissair.com/what-is-pm-2-5.htm, one of several common measurements used in studies of air quality.

BEFORE YOU START!

A good workflow with R is essential. Create a new folder or project (if you choose to use projects) just for this India Air Quality example. Put the CSV into that folder. Make a script in that folder for your work. Make sure you can do a `read.csv()` with the directory set properly (see earlier materials on this or ask someone for help if there is a problem).

Now begin:

```
x <- read.csv("India_AirQuality.csv")

dim(x)
```

```
## [1] 720   7
```

```
head(x)
```

```
##      DateTime Chennai Delhi Hyderabad Kolkata hour date
## 1 1/6/16 1:00      20    34        32      41    1    1
## 2 1/6/16 2:00      32    43        40      33    2    1
## 3 1/6/16 3:00      36    74        39      28    3    1
## 4 1/6/16 4:00      27    52        33      18    4    1
## 5 1/6/16 5:00      31    46        35      22    5    1
## 6 1/6/16 6:00      33    38        35      23    6    1
```

```
tail(x)
```

```
##          DateTime Chennai Delhi Hyderabad Kolkata hour date
## 715 30/6/16 19:00      35    41       -15      24   19   30
## 716 30/6/16 20:00      43    45        NA      18   20   30
## 717 30/6/16 21:00      27    39        NA      25   21   30
## 718 30/6/16 22:00      33    45        NA      29   22   30
## 719 30/6/16 23:00      32    44        NA      15   23   30
## 720 30/6/16 23:59      28    39        NA      17   24   30
```

```
str(x)
```

```
## 'data.frame':    720 obs. of  7 variables:
##  $ DateTime : Factor w/ 720 levels "1/6/16 1:00",..: 1 12 18 19 20 21 22 23 24 2 ...
```

```
##  $ Chennai  : int  20 32 36 27 31 33 27 28 31 27 ...
##  $ Delhi    : int  34 43 74 52 46 38 46 38 48 52 ...
##  $ Hyderabad: int  32 40 39 33 35 35 30 42 43 34 ...
##  $ Kolkata  : int  41 33 28 18 22 23 30 27 24 22 ...
##  $ hour     : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ date     : int  1 1 1 1 1 1 1 1 1 1 ...
```

Challenge 1: Consider Delhi. Calculate the following: the minimum, maximum, mean, median, and standard deviation of the Delhi PM2.5 measurements. This involves using functions: `min`, `max`, `mean`, `median`, and `sd`.

```
# YOUR WORK HERE
```

Challenge 2: Now consider one of the other cities. Are there any missing PM2.5 measurements for this city? If so, how many?

```
# YOUR WORK HERE
```

Challenge 3: Use `boxplot()` to explore with a graphic the PM2.5 measurements of Delhi. You might have to refer to the help page for boxplot by typing `?boxplot` in the console.

```
# YOUR WORK HERE
```

Challenge 4: Use `hist()` to explore with a graphic the PM2.5 measurements of Delhi. You might have to refer to the help page for boxplot by typing `?hist` in the console.

```
# YOUR WORK HERE
```

Challenge 5: Are any of Delhi's PM2.5 measurements below 0? In theory, the measurement should not contain any negative value.

```
# YOUR WORK HERE
```