The Air Quality Dataset

Augusto Souza

03/07/2020

## Summary of Air Quality Dataset

This exercise will be working with the built-in air quality dataset.[[1]](#footnote-21) This dataset contains 154 daily air quality measurements in New York from May 1, 1973 ( *a Tuesday*) to September 30, 1973. The dataset contains 6 variables:

* **Ozone**: Mean ozone in parts per billion ( *ppb* ) from 1300 to 1500 hours at Roosevelt Island;
* **Solar.R**: Solar radiation in Langleys ( *lang* ) in the frequency band 4000–7700 Angstroms from 0800 to 1200 hours at Central Park;
* **Wind**: Average wind speed in miles per hour ( *mph*) at 0700 and 1000 hours at LaGuardia Airport;
* **Temp**: Maximum daily temperature in degrees Fahrenheit ( *oF*) at LaGuardia Airport;
* **Month**: numeric month ( *1-12*)
* **Day**: numeric Day of the month ( *1-31*)

### Table of Top of the Air Quality Dataset

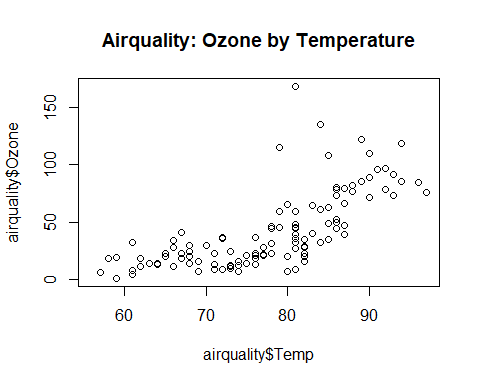
data("airquality")  
knitr::kable(head(airquality),   
 caption = "Top of the Air Quality Dataset")

Top of the Air Quality Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Ozone | Solar.R | Wind | Temp | Month | Day |
| 41 | 190 | 7.4 | 67 | 5 | 1 |
| 36 | 118 | 8.0 | 72 | 5 | 2 |
| 12 | 149 | 12.6 | 74 | 5 | 3 |
| 18 | 313 | 11.5 | 62 | 5 | 4 |
| NA | NA | 14.3 | 56 | 5 | 5 |
| 28 | NA | 14.9 | 66 | 5 | 6 |

### Plot of Ozone by Temperature –Air Quality Dataset

plot(airquality$Temp, airquality$Ozone,   
 main="Airquality: Ozone by Temperature")



1. Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) Graphical Methods for Data Analysis. Belmont, CA: Wadsworth. [↑](#footnote-ref-21)