Final Project First Plot

Zulin Zhou (zz69), Zhanhang Zhou (zz70)

Thursday October 22, 3:30pm

Introduction

We want to track pollutants levels in the air over US, especially in Texas, and observe if they are effected by other weather conditions. The main pollutant in the air is O3(ozone), so we will be tracking on that. Other pollutants level might have differnet importance in different districts so we are also planning to research on those factors

Datasets

We are using data from the official EPA website https://www.epa.gov/castnet. The main table is 22,290,625X17. The table takes weather conditions and ozone rates into consideration by scale of hour. We also have side charts indicating similar data by day, week and season. And there's side charts that set down other pollutants including SO2, HNO3, ammonium and so on weekly scale.

The main table includes the following factors: 1. SITE_ID, which stands for Site identification code, in form of CHAR 2. DATE_TIME, in form of STRING 3. WIND_DIRECTION, in form of NUMBER 4. WINDSPEED_SCALAR, which stands for scalar wind speed, in form of NUMBER 5. OZONE, in form of NUMBER 6. SOLAR_RADIATION, in form of NUMBER 7. FLOW_RATE, in form of NUMBER 8. SHELTER_TEMPERATURE, in form of NUMBER 9. QA_CODE, which stand for Quality assurance level of the record, in form of CHAR 10. SIGMA_THETA, which stands for Standard deviation of wind direction, in form of NUMBER 11. WETNESS, in form of NUMBER 12. TEMPERATURE, in form of NUMBER 13. UPDATE_DATE, in form of STRING 14. WINDSPEED, which stands for vector wind speed, in form of NUMBER 15. TEMPERATURE_DELTA, which stands for Temperature difference between 9m and 2m probes, in form of NUMBER 16. PRECIPITATION, in form of NUMBER 17. RELATIVE_HUMIDITY, in form of NUMBER

First Plot

Showing the percipitation (reflected by flow rate), ozone level and temperature for Texas in Year 2015-2019. Data subsetted through webpage API.

First load the subsetted data

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.0.2
```

```
Meteorological_Hourly <- read_csv("Meteorological - Hourly.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
```

```
##   SITE_ID = col_character(),
##   DATE_TIME = col_character(),
##   TEMPERATURE = col_double(),
##   TEMPERATURE_DELTA = col_logical(),
##   RELATIVE_HUMIDITY = col_logical(),
##   SOLAR_RADIATION = col_logical(),
##   OZONE = col_double(),
##   PRECIPITATION = col_logical(),
##   WINDSPEED = col_logical(),
##   WIND_DIRECTION = col_logical(),
##   SIGMA_THETA = col_logical(),
##   FLOW_RATE = col_double(),
##   WINDSPEED_SCALAR = col_logical(),
##   WETNESS = col_logical(),
##   SHELTER_TEMPERATURE = col_double(),
##   QA_CODE = col_double(),
##   UPDATE_DATE = col_character()
## )
```

```
## Warning: 529646 parsing failures.
##  row               col           expected actual                          file
## 4465 RELATIVE_HUMIDITY 1/0/T/F/TRUE/FALSE  10.0  'Meteorological - Hourly.csv'
## 4465 SOLAR_RADIATION   1/0/T/F/TRUE/FALSE  2.0   'Meteorological - Hourly.csv'
## 4465 PRECIPITATION     1/0/T/F/TRUE/FALSE  0.0   'Meteorological - Hourly.csv'
## 4465 WINDSPEED         1/0/T/F/TRUE/FALSE  3.3   'Meteorological - Hourly.csv'
## 4465 WIND_DIRECTION    1/0/T/F/TRUE/FALSE  303.0 'Meteorological - Hourly.csv'
## .... ................. ................... ...... ..............................
## See problems(...) for more details.
```

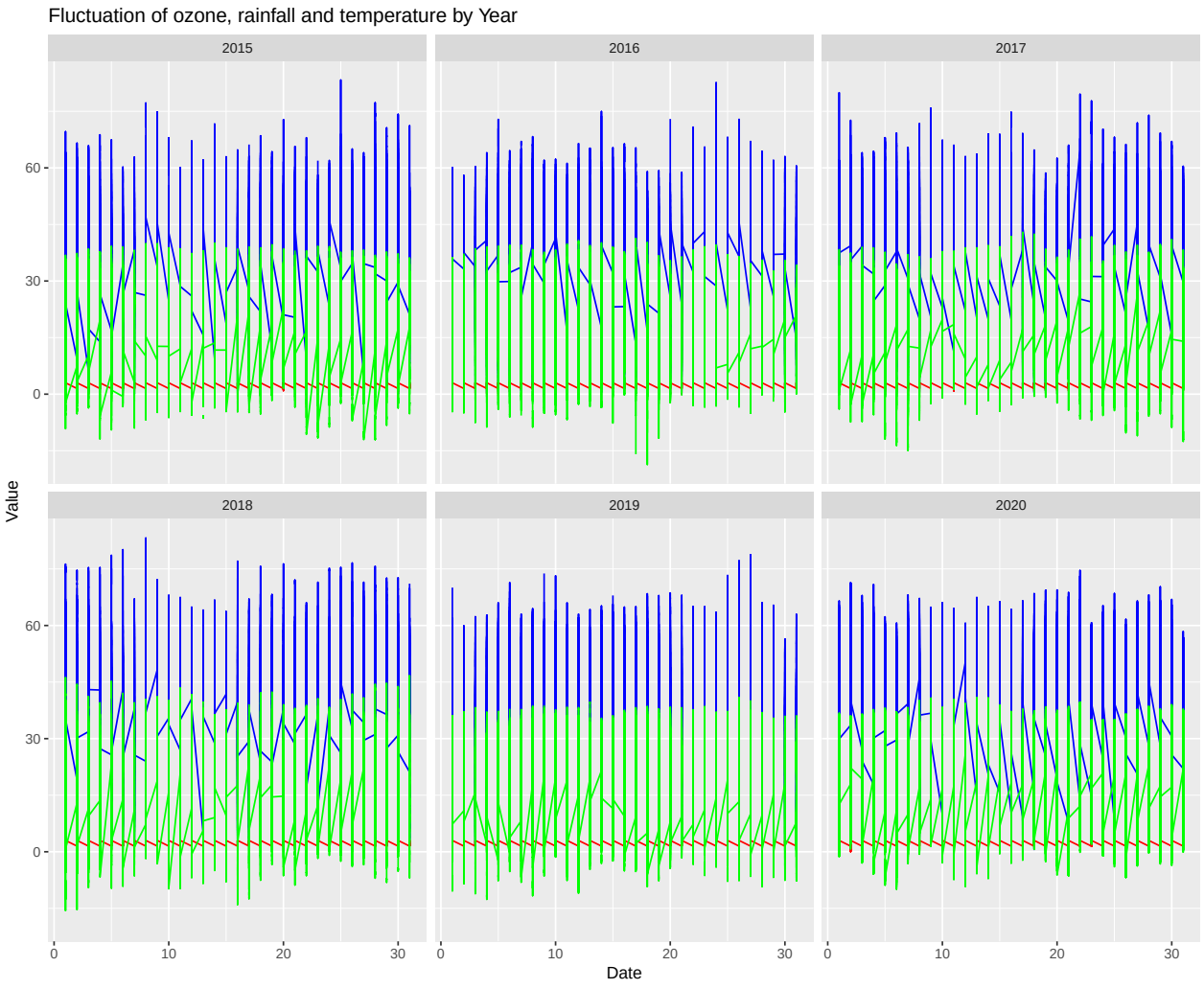
Strip date and related variables

```
setTime <- strptime(Meteorological_Hourly$DATE_TIME, "%m/%d/%Y %H:%M:%S")
```

```
year <- as.numeric(format(setTime,'%Y'))
date <- as.numeric(format(setTime,'%d'))
temperature <- Meteorological_Hourly$TEMPERATURE
ozone <- Meteorological_Hourly$OZONE
rainfall <- Meteorological_Hourly$FLOW_RATE
plotData <- data.frame(rainfall,ozone,temperature,year,date)
```

Plot and facet_wrap by year

```
ggplot(plotData) +
  geom_line(aes(x = date, y = ozone), color = "blue") +
  geom_line(aes(x = date, y = rainfall), color = "red") +
  geom_line(aes(x = date, y = temperature), color = "green") +
  facet_wrap(~year,ncol = 3) +
  labs(title = "Fluctuation of ozone, rainfall and temperature by Year",
       x = "Date",
       y = "Value")
```


Fluctuation of ozone, rainfall and temperature by Year
2015
2016
2017
2018
2019
2020
Value
Date
60
30
0
0
10
20
30