Final Project Report 2

Zulin Zhou (zz69), Zhanhang Zhou (zz70)

10/29/2020

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. The secondart source chart, size 135,526X22 also sets 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

Project Goal

Find relationship between pollutants and weather/year. In the first two milestones we observe the increase/decrease of certain pollutants under effects of tempeature, rainfall and see the fluctuation over time

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)
library(readr)
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

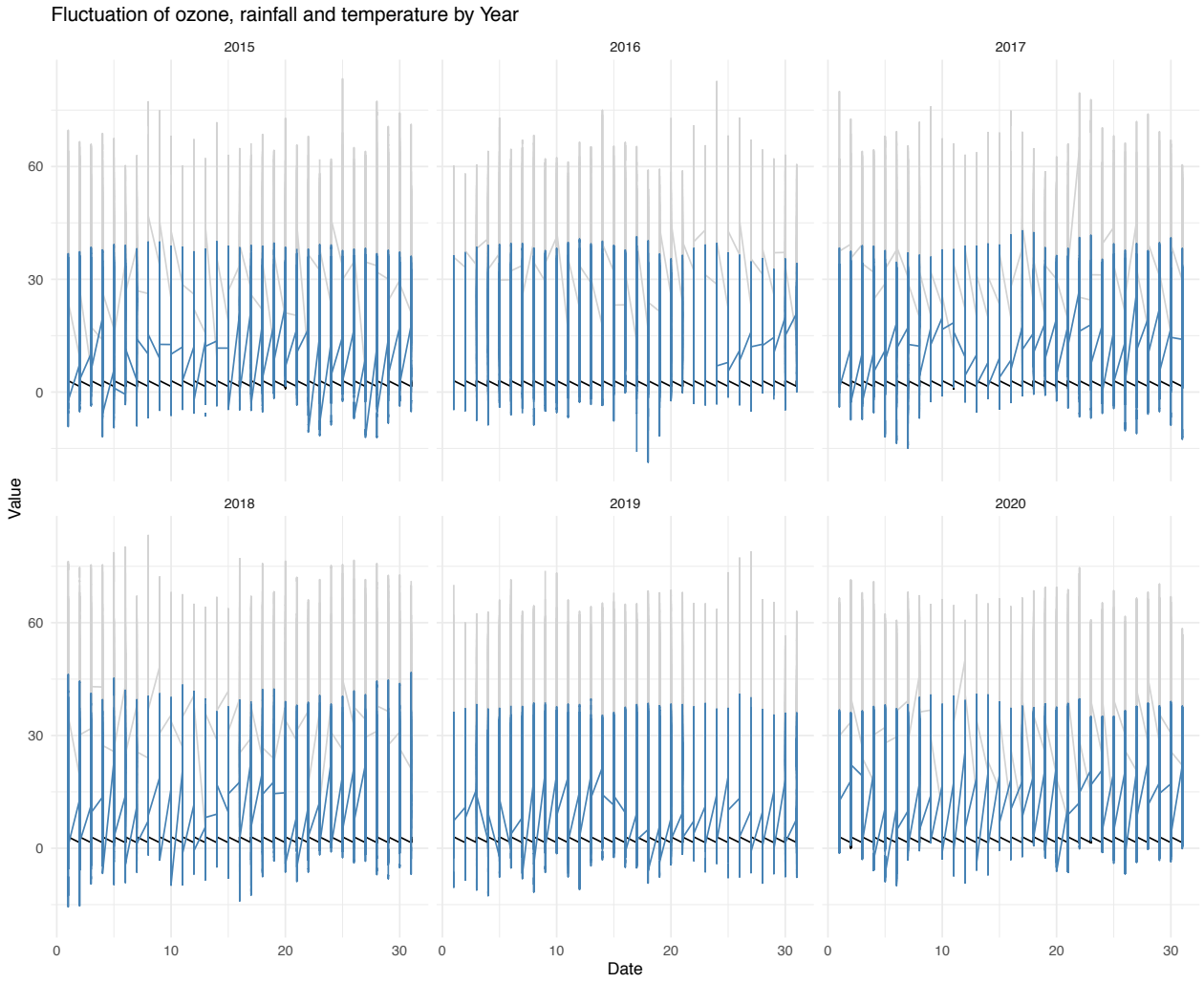
```
setTime <- strptime(Meteorological_Hourly$DATE_TIME, "%m/%d/%Y %H:%M:%S")
```

```
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 = "lightgray") +
  geom_line(aes(x = date, y = rainfall), color = "black") +
```

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



FeedBack from polt 1

Since there's too many observations in the same plots, it's hard to observe that all the values are on the same scale, and the data points are layered up by concertration of hours, so there's no apparent obervation. In order to get better visualization effect, we reduce our scope and radomly pick samples to represent in graph. Also updated project goal and picture theme according to the comments from last week.

Data elimination using RSQLite

```
library(RSQLite)
dcon <- dbConnect(SQLite(), dbname = "/Users/mac/Desktop/fall 20/605/meteor.db")
```

Insert secondary table. The main table was already inserted, see the last page for more information. We will combine those command lines later.

```
table <- read.csv(paste0("/Users/mac/Desktop/fall 20/605/Measurement.csv"))
dbWriteTable(conn = dcon, name = "measurements", table,
             append = TRUE, row.names = FALSE)
```

```
dbListTables(dcon)
```

```
## [1] "measurements" "meteor"
```

```
dbListFields(dcon,"meteor")
```

```
##  [1] "SITE_ID"             "DATE_TIME"           "TEMPERATURE"
##  [4] "TEMPERATURE_DELTA"   "RELATIVE_HUMIDITY"   "SOLAR_RADIATION"
##  [7] "OZONE"               "PRECIPITATION"       "WINDSPEED"
## [10] "WIND_DIRECTION"      "SIGMA_THETA"         "FLOW_RATE"
## [13] "WINDSPEED_SCALAR"    "WETNESS"             "SHELTER_TEMPERATURE"
## [16] "QA_CODE"             "UPDATE_DATE"
```

```
dbListFields(dcon,"measurements")
```

```
##  [1] "SITE_ID"       "TYPE"          "DATEON"        "DATEOFF"
##  [5] "TSO4"          "TNO3"          "TNH4"          "CA"
##  [9] "MG"            "NA."           "K"             "CL"
## [13] "NSO4"          "NHNO3"         "WSO2"          "WNO3"
## [17] "TOTAL_SO2"     "TOTAL_NO3"     "FLOW_VOLUME"   "VALID_HOURS"
## [21] "COMMENT_CODES" "STD2LOCAL_CF"  "TEMP_SOURCE"   "QA_CODE"
## [25] "UPDATE_DATE"
```

From the database, pick 1000 random rows then get the weather and ozone layer data. Later we can use the selected data to plot.

```
res <- dbSendQuery(conn = dcon, "
SELECT SITE_ID, DATE_TIME, TEMPERATURE, FLOW_RATE, OZONE
FROM meteor
WHERE SITE_ID IN
(SELECT SITE_ID
FROM meteor
WHERE TEMPERATURE IS NOT NULL and FLOW_RATE IS NOT NULL and OZONE IS NOT NULL
ORDER BY RANDOM()
LIMIT 1000);
")
mydf <- dbFetch(res, -1)
dbClearResult(res)
head(mydf)
```

```
##    SITE_ID             DATE_TIME TEMPERATURE FLOW_RATE OZONE
## 1 "EGB181" "08/05/1998 17:00:00"     "24.05"  "1.4904"    ""
## 2 "EGB181" "08/05/1998 18:00:00"      "23.5"  "1.4904"    ""
## 3 "EGB181" "08/05/1998 19:00:00"      "23.0"  "1.4904"    ""
## 4 "EGB181" "09/09/1998 10:00:00"      "13.1"  "1.4904"    ""
## 5 "EGB181" "09/09/1998 11:00:00"     "13.75"  "1.4904"    ""
## 6 "EGB181" "09/09/1998 12:00:00"      "14.5"  "1.4904"    ""
```

```
dbDisconnect(dcon)
```

```
dbDisconnect(dcon)
```

HW 7 Group

Zulin Zhou (zz69) Zhanhang Zhou(zz70)"

10/29/2020

```
library(RSQLite)
db <- dbConnect(SQLite(), dbname = "meteor.db")
field.types <- list(
  SITE_ID = "TEXT",
  DATE_TIME = ""
)
dbWriteTable(conn = db, name = "meteor", value = "Meteorological.csv", row.names = FALSE, header = TRUE
dbDisconnect(db)
```

```
db <- dbConnect(SQLite(), dbname = "meteor.db")
dbListTables(db)
```

```
## [1] "meteor"
```

```
dbListFields(db,"meteor")
```

```
##  [1] "SITE_ID"             "DATE_TIME"           "TEMPERATURE"
##  [4] "TEMPERATURE_DELTA"   "RELATIVE_HUMIDITY"   "SOLAR_RADIATION"
##  [7] "OZONE"               "PRECIPITATION"       "WINDSPEED"
## [10] "WIND_DIRECTION"      "SIGMA_THETA"         "FLOW_RATE"
## [13] "WINDSPEED_SCALAR"    "WETNESS"             "SHELTER_TEMPERATURE"
## [16] "QA_CODE"             "UPDATE_DATE"
```

```
dbDisconnect(db)
```