Weather Patterns X COVID-19

Final Project Documentation

Uhuru Kamau, Shuyang Lu, Yifan Zhao, Chenhao Zhao, Zhaofeng Liu

Contents


```
# Load Required Packages
library(tidyverse)
library(kableExtra)
library(readr)
library(gridExtra)
library(knitr)
```

```
knitr::include_graphics(path = "QBS181_1.png")
```

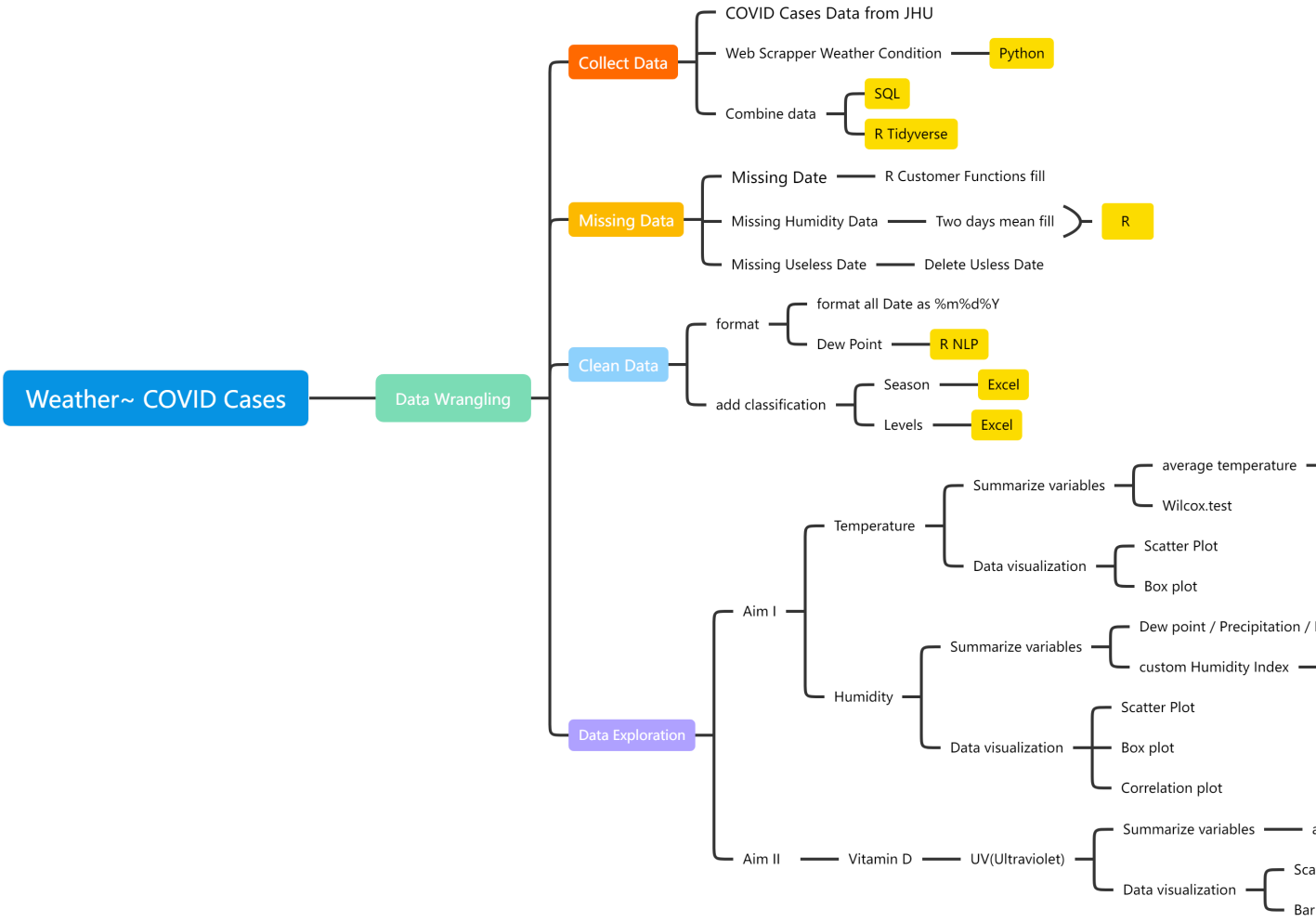

Figure 1: Highlighting the Keys to our Relational Database

Data Acquisition

1. New York City COVID-19 Data Archive

- Source: NYC OpenData
- Acquisation Method
  - Download .csv file
- Purpose:
  - We will use this time series data to track changes in the incidence of COVID-19.

2. New York City Weather Data

- Source: Weather Underground - Weather Archive
- Acquisition Method
  - Webscraping/ API Tool
- Purpose:
  - Merge time series weather data with timeseries Covid-19 data and investigate potential associations

3. Daily UV Index Scores - New York City

- Source: Central New York's Live Weather Source
- Acquisition Method
  - UV index values are presented as tables (see figure)
  - Copy tables and paste into Microsoft Excel
  - Save as .csv file
- Purpose
  - Sunlight and Vitamin-D absorbtion
    - It is generally accepted that there is a positive association between exposure to sunlight and absorbtion of vitamin-D.
    - It is also generally accepted that there is a positive association between vitamin-D absorbtion and immune system capacity.
  - We will us UV-Index as a proxy for exposure to sunlight at the population level and test for associations between UV Index and the incidence of Covid-19.

Relational Schema

```
knitr::include_graphics(path = "images/Relational_Schema.png")
```

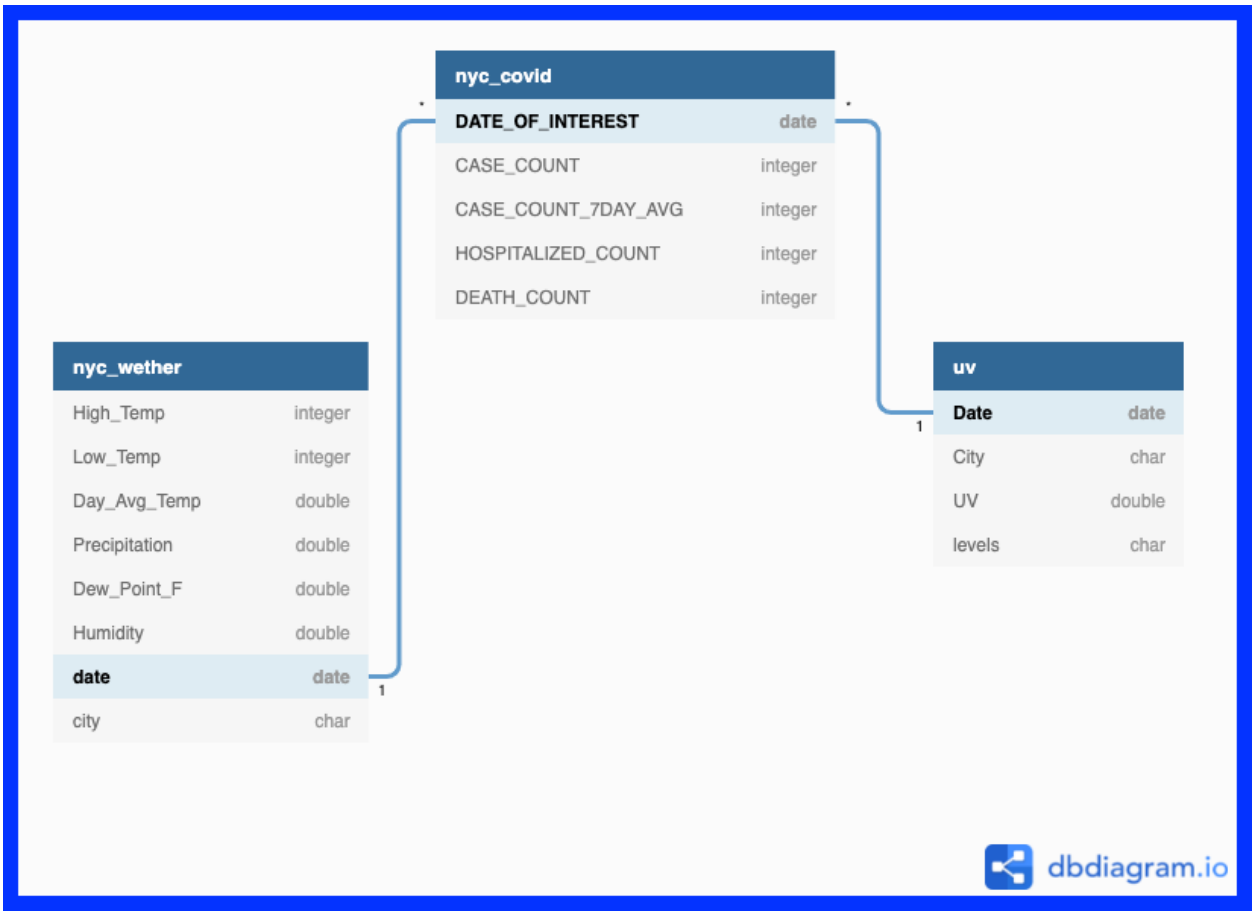

Figure 2: Highlighting the Keys to our Relational Database

Data Cleaning

1. New York City COVID-19 Data Archive

2. New York City Weather Data

A. Read-in File from Raw Data File

- The raw file has an issue with the column headers.
  - Several Headers include symbols that don't work with the interpretor
    - eg. $Low\_Temp(^oF)$, $High\_Temp(^oF)$
- Solution: Update column names while reading in the file!

```
# vector with acceptable column names
headers <- c("High.Temp", "Low.Temp", "Avg.Temp", "Precip", "Dew.Point", "Humidity",
    "Date", "City")

# read in the raw data file
weather.raw <- read.csv(file = "data/Raw Data/nyc_weather_raw.csv", header = TRUE,
    col.names = headers)
```

B. Remove the "City" Variable

- Every observation is is "new york city"

```
weather.clean <- weather.raw %>%
    select(-City)
```

C. Display

```
y = weather.raw[1:5, ]
knitr::kable(x = y, digits = 2, align = "c")
```

High.Temp	Low.Temp	Avg.Temp	Precip	Dew.Point	Humidity	Date	City
44	26	35.46	0.00	13.67	41.83	2020/3/1 0:00	new york city
56	38	48.17	0.00	30.46	51.12	2020/3/2	new york city
58	48	52.41	0.01	44.59	75.47	2020/3/3	new york city
57	46	50.52	0.28	28.52	44.76	2020/3/4	new york city
52	40	44.75	0.00	25.38	48.50	2020/3/5	new york city

D. Write the Processed Data to a new .csv file

```
write.csv(x = weather.clean, file = "data/Processed Data/nyc_weather.csv")
```

3. Daily UV Index Scores - New York City