# Summary of Wrangling

*Fadi*

*3 January 2018*

**Renaming:**

Most columns were aptly named, so there wasn't really much to be adjusted except the "Date Of Stop" and "Time Of Stop" columns; I went ahead and renamed them "Date" & "Time", respectively.

**NA & NONE Values:**

Many variables have null values, but not all of them are critical enough to weigh negatively on the analysis. On the Other hand, "Description" had six missing values, which I went ahead and filtered out.

"Make" also had many NA values, which indicates either that the car was unknown, or a vehicle was not involved in the violation. "Make" also had over 7000 thousand observations that was marked as **NONE**, which is most likely pedestrain violations, I am lead to such conclusion as many of these violations have been registered by "Foot Patrol"

```
Traffic_Violations %>%
  filter(Traffic_Violations$Make == "NONE") %>%
  select(Description, Make, `Arrest Type`)
```

```
## # A tibble: 7,032 x 3
##    Description                                          Make  `Arrest Typ~
##    <chr>                                                <chr> <chr>
##  1 PEDESTRIAN CROSSING ROADWAY BETWEEN ADJACENT INTERS~ NONE  O - Foot Pa~
##  2 PASSENGER AGE 16 OR MORE IN OUTBOARD FRONT SEAT OF ~ NONE  A - Marked ~
##  3 PEDESTRIAN FAIL TO OBEY UPRAISED HAND SIGNAL         NONE  B - Unmarke~
##  4 PEDESTRIAN CROSSING ROADWAY BETWEEN ADJACENT INTERS~ NONE  O - Foot Pa~
##  5 "PEDESTRIAN FAIL TO OBEY  \"UPRAISED HAND\" SIGNAL"  NONE  A - Marked ~
##  6 PEDESTRIAN FAIL TO OBEY UPRAISED HAND SIGNAL         NONE  A - Marked ~
##  7 PEDESTRIAN FAIL TO OBEY UPRAISED HAND SIGNAL         NONE  A - Marked ~
##  8 PEDESTRIAN FAIL TO OBEY UPRAISED HAND SIGNAL         NONE  B - Unmarke~
##  9 OPERATING A MOTOR SCOOTER ON HWY. W/O REQ. LICENSE ~ NONE  A - Marked ~
## 10 PEDESTRIAN CROSSING ROADWAY BETWEEN ADJACENT INTERS~ NONE  A - Marked ~
## # ... with 7,022 more rows
```

## NULL Geolocations

There are over 79,000 observations whose "Geolocation" values are **NA**. Many of those coordinates can be arduously extracted from google maps by cross-referencing their corrosponding "Location" values, but this process is tedious at best. I will be putting the resolution of these observations under consideration, for now.

## Other Variables

HAZMAT (Hazardous Material-related) accidents are very rare, and are more or less outliers in this analysis, so I have excluded these observations for the time being.

Article, Agency, SubAgency have inconsequential details, so I went ahead and disposed of these columns.