Analysis: NYPD Shooting Incident

Kohav, J.

08 January 2022

Import (Data)

```
nypd_data <- read_csv("https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?access
Type=DOWNLOAD")
```

```
## Rows: 23585 Columns: 19
```

```
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (10): OCCUR_DATE, BORO, LOCATION_DESC, PERP_AGE_GROUP, PERP_SEX, PERP_R...
## dbl   (7): INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, X_COORD_CD, Y_COORD_CD...
## lgl   (1): STATISTICAL_MURDER_FLAG
## time  (1): OCCUR_TIME
```

```
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
nypd_data
```

```
## # A tibble: 23,585 × 19
##    INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO     PRECINCT JURISDICTION_CODE
##           <dbl> <chr>      <time>     <chr>       <dbl>              <dbl>
##  1     24050482 08/27/2006 05:35      BRONX          52                  0
##  2     77673979 03/11/2011 12:03      QUEENS        106                  0
##  3    203350417 10/06/2019 01:09      BROOKLYN       77                  0
##  4     80584527 09/04/2011 03:35      BRONX          40                  0
##  5     90843766 05/27/2013 21:16      QUEENS        100                  0
##  6     92393427 09/01/2013 04:17      BROOKLYN       67                  0
##  7     73057167 06/05/2010 21:16      BROOKLYN       77                  0
##  8    211362213 03/20/2020 21:27      BROOKLYN       81                  0
##  9    137564752 07/04/2014 00:25      QUEENS        101                  0
## 10    147024011 10/18/2015 01:33      QUEENS        106                  0
## # … with 23,575 more rows, and 13 more variables: LOCATION_DESC <chr>,
## #   STATISTICAL_MURDER_FLAG <lgl>, PERP_AGE_GROUP <chr>, PERP_SEX <chr>,
## #   PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>, VIC_RACE <chr>,
## #   X_COORD_CD <dbl>, Y_COORD_CD <dbl>, Latitude <dbl>, Longitude <dbl>,
## #   Lon_Lat <chr>
```

```
summary(nypd_data)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME           BORO
##  Min.   :  9953245   Length:23585       Length:23585      Length:23585
##  1st Qu.: 55322804   Class :character   Class1:hms        Class :character
##  Median : 83435362   Mode  :character   Class2:difftime   Mode  :character
##  Mean   :102280741                      Mode  :numeric
##  3rd Qu.:150911774
##  Max.   :230611229
##
##     PRECINCT      JURISDICTION_CODE LOCATION_DESC      STATISTICAL_MURDER_FLAG
##  Min.   :  1.00   Min.   :0.000     Length:23585       Mode :logical
##  1st Qu.: 44.00   1st Qu.:0.000     Class :character   FALSE:19085
##  Median : 69.00   Median :0.000     Mode  :character   TRUE :4500
##  Mean   : 66.21   Mean   :0.333
##  3rd Qu.: 81.00   3rd Qu.:0.000
##  Max.   :123.00   Max.   :2.000
##                   NA's   :2
##  PERP_AGE_GROUP       PERP_SEX          PERP_RACE         VIC_AGE_GROUP
##  Length:23585       Length:23585       Length:23585       Length:23585
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    VIC_SEX             VIC_RACE           X_COORD_CD        Y_COORD_CD
##  Length:23585       Length:23585       Min.   : 914928   Min.   :125757
##  Class :character   Class :character   1st Qu.: 999925   1st Qu.:182539
##  Mode  :character   Mode  :character   Median :1007654   Median :193470
##                                        Mean   :1009379   Mean   :207300
##                                        3rd Qu.:1016782   3rd Qu.:239163
##                                        Max.   :1066815   Max.   :271128
##
##     Latitude       Longitude        Lon_Lat
##  Min.   :40.51   Min.   :-74.25   Length:23585
##  1st Qu.:40.67   1st Qu.:-73.94   Class :character
##  Median :40.70   Median :-73.92   Mode  :character
##  Mean   :40.74   Mean   :-73.91
##  3rd Qu.:40.82   3rd Qu.:-73.88
##  Max.   :40.91   Max.   :-73.70
##
```

Clean (Data)

```
# Remove: Columns
nypd_data_clean <- nypd_data %>% select(-c(X_COORD_CD:Lon_Lat))
nypd_data_clean <- nypd_data_clean %>% select(-c(PRECINCT, JURISDICTION_CODE))

# Change: Format
nypd_data_clean <- nypd_data_clean %>% mutate(OCCUR_DATE = mdy(OCCUR_DATE))
nypd_data_clean <- nypd_data_clean %>% mutate(OCCUR_TIME = hms(OCCUR_TIME))
```

Analysis (Preliminary)

```
nypd_data_clean %>% group_by(LOCATION_DESC) %>% summarize()
```

```
## # A tibble: 40 × 1
##    LOCATION_DESC
##    <chr>
##  1 ATM
##  2 BANK
##  3 BAR/NIGHT CLUB
##  4 BEAUTY/NAIL SALON
##  5 CANDY STORE
##  6 CHAIN STORE
##  7 CHECK CASH
##  8 CLOTHING BOUTIQUE
##  9 COMMERCIAL BLDG
## 10 DEPT STORE
## # … with 30 more rows
```

```
nypd_data_clean %>% group_by(VIC_AGE_GROUP) %>% summarize()
```

```
## # A tibble: 6 × 1
##   VIC_AGE_GROUP
##   <chr>
## 1 <18
## 2 18-24
## 3 25-44
## 4 45-64
## 5 65+
## 6 UNKNOWN
```

```
nypd_data_clean %>% group_by(VIC_RACE) %>% summarize()
```

```
## # A tibble: 7 × 1
##   VIC_RACE
##   <chr>
## 1 AMERICAN INDIAN/ALASKAN NATIVE
## 2 ASIAN / PACIFIC ISLANDER
## 3 BLACK
## 4 BLACK HISPANIC
## 5 UNKNOWN
## 6 WHITE
## 7 WHITE HISPANIC
```

```
nypd_data_clean %>% group_by(BORO) %>% summarize()
```

```
## # A tibble: 5 × 1
##   BORO
##   <chr>
## 1 BRONX
## 2 BROOKLYN
## 3 MANHATTAN
## 4 QUEENS
## 5 STATEN ISLAND
```

Analysis

```
# Summary: (Setting: Cases)
nypd_data_clean %>% group_by(LOCATION_DESC) %>% summarize(Cases = n())
```

```
## # A tibble: 40 × 2
##    LOCATION_DESC     Cases
##    <chr>             <int>
##  1 ATM                   1
##  2 BANK                  1
##  3 BAR/NIGHT CLUB      562
##  4 BEAUTY/NAIL SALON   100
##  5 CANDY STORE           6
##  6 CHAIN STORE           5
##  7 CHECK CASH            1
##  8 CLOTHING BOUTIQUE    14
##  9 COMMERCIAL BLDG     234
## 10 DEPT STORE            5
## # … with 30 more rows
```

```
# Summary: (Setting: Age)
nypd_data_clean %>% group_by(LOCATION_DESC) %>% summarize ("<18" = sum(VIC_AGE_GROUP ==
"<18"), "18-24" = sum(VIC_AGE_GROUP == "18-24"), "25-44" = sum(VIC_AGE_GROUP == "25-44"
), "45-64" = sum(VIC_AGE_GROUP == "45-64"), "65+" = sum(VIC_AGE_GROUP == "65+"))
```

```
## # A tibble: 40 × 6
##    LOCATION_DESC     `<18` `18-24` `25-44` `45-64` `65+`
##    <chr>             <int>   <int>   <int>   <int> <int>
##  1 ATM                   0       1       0       0     0
##  2 BANK                  0       0       1       0     0
##  3 BAR/NIGHT CLUB       11     212     313      23     0
##  4 BEAUTY/NAIL SALON     6      21      65       7     1
##  5 CANDY STORE           1       3       2       0     0
##  6 CHAIN STORE           0       1       4       0     0
##  7 CHECK CASH            0       1       0       0     0
##  8 CLOTHING BOUTIQUE     0       3       7       2     2
##  9 COMMERCIAL BLDG      16      84     113      20     1
## 10 DEPT STORE            1       0       0       4     0
## # … with 30 more rows
```

```
# Summary: (Setting: Race)
nypd_data_clean %>% group_by(LOCATION_DESC) %>% summarize ("NATIVE" = sum(VIC_RACE == "A
MERICAN INDIAN/ALASKAN NATIVE"), "ASIAN" = sum(VIC_RACE == "ASIAN / PACIFIC ISLANDER"),
"BLACK" = sum(VIC_RACE == "BLACK"), "BLACK (Hispanic)" = sum(VIC_RACE == "BLACK HISPANI
C"), "WHITE" = sum(VIC_RACE == "WHITE"), "WHITE (Hispanic))" = sum(VIC_RACE == "WHITE HI
SPANIC"))
```

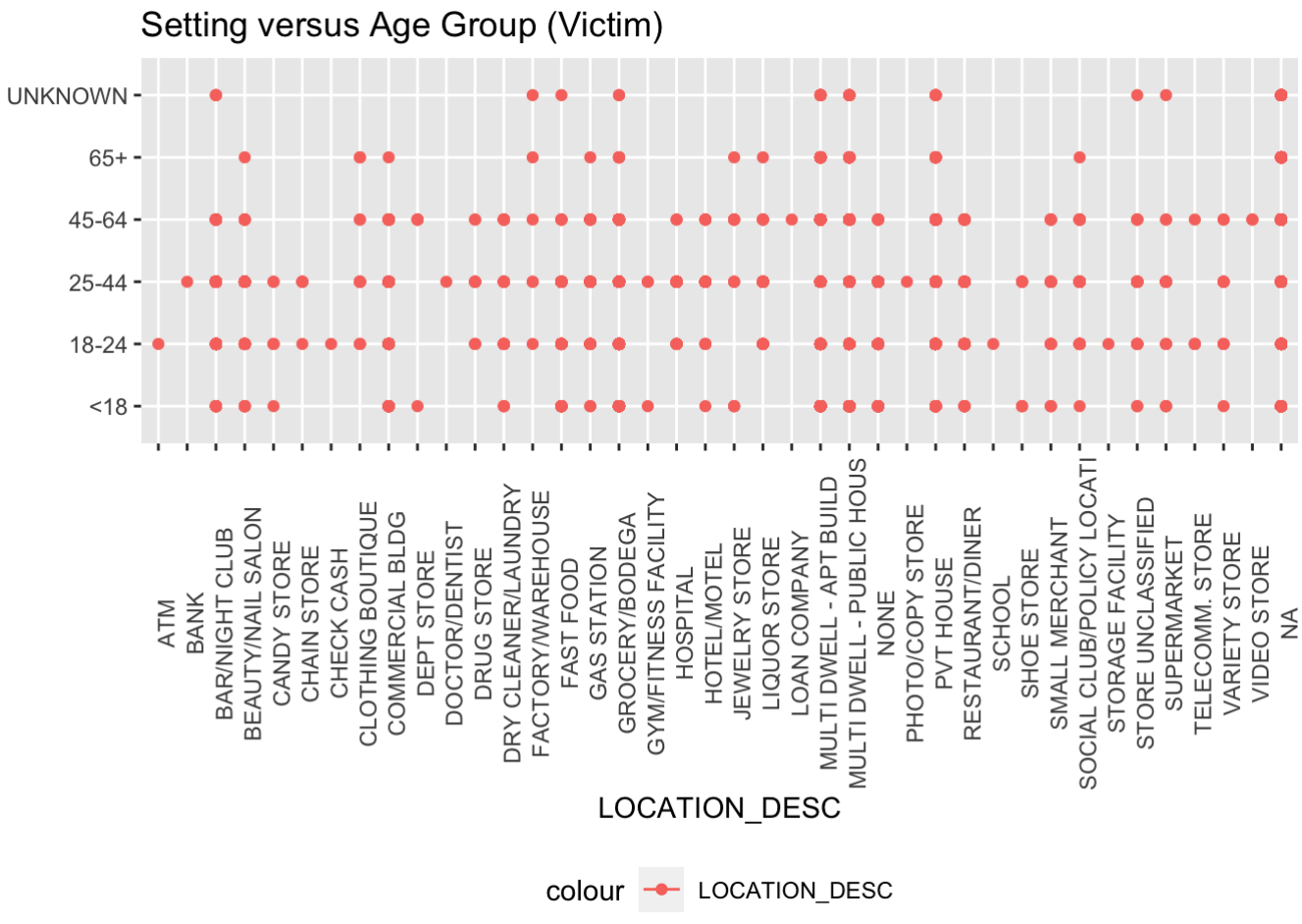
```
## # A tibble: 40 × 7
##    LOCATION_DESC     NATIVE ASIAN BLACK `BLACK (Hispanic… WHITE `WHITE (Hispani…
##    <chr>              <int> <int> <int>             <int> <int>            <int>
##  1 ATM                    0     0     1                 0     0                0
##  2 BANK                   0     0     0                 0     0                1
##  3 BAR/NIGHT CLUB         0     9   380                62    22               88
##  4 BEAUTY/NAIL SALON      0     1    71                 8     3               17
##  5 CANDY STORE            0     1     5                 0     0                0
##  6 CHAIN STORE            0     1     4                 0     0                0
##  7 CHECK CASH             0     0     1                 0     0                0
##  8 CLOTHING BOUTIQUE      0     0     9                 1     1                3
##  9 COMMERCIAL BLDG        0     9   169                17    13               26
## 10 DEPT STORE             0     1     1                 0     3                0
## # … with 30 more rows
```

```
# Summary: (Setting: Gender)
nypd_data_clean %>% group_by(LOCATION_DESC) %>% summarize ("MALE" = sum(VIC_SEX == "M"),
"FEMALE" = sum(VIC_SEX == "F"))
```

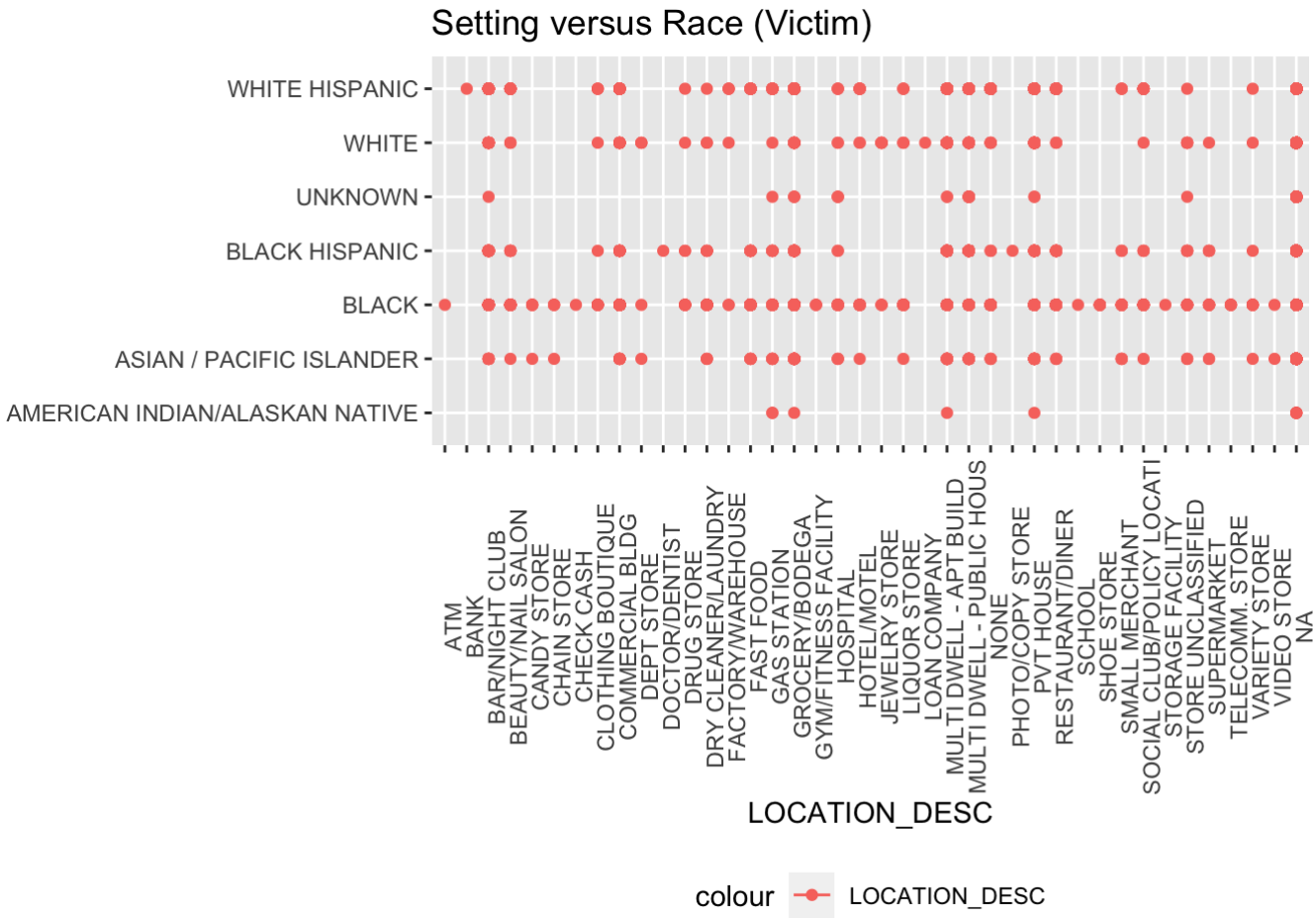
```
## # A tibble: 40 × 3
##    LOCATION_DESC      MALE FEMALE
##    <chr>             <int>  <int>
##  1 ATM                   1      0
##  2 BANK                  1      0
##  3 BAR/NIGHT CLUB      494     68
##  4 BEAUTY/NAIL SALON    89     11
##  5 CANDY STORE           6      0
##  6 CHAIN STORE           5      0
##  7 CHECK CASH            1      0
##  8 CLOTHING BOUTIQUE    13      1
##  9 COMMERCIAL BLDG     207     27
## 10 DEPT STORE            4      1
## # … with 30 more rows
```

Visualization (Initial)

```
# Visualize (Setting: Age)
nypd_data_clean %>% ggplot(aes(x = LOCATION_DESC, y = VIC_AGE_GROUP)) +
  geom_line(aes(color = "LOCATION_DESC")) +
  geom_point(aes(color = "LOCATION_DESC")) +
  theme(legend.position="bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "Setting versus Age Group (Victim)", y = NULL)
```



```
# Visualize (Setting: Race)
nypd_data_clean %>% ggplot(aes(x = LOCATION_DESC, y = VIC_RACE)) +
  geom_line(aes(color = "LOCATION_DESC")) +
  geom_point(aes(color = "LOCATION_DESC")) +
  theme(legend.position="bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "Setting versus Race (Victim)", y = NULL)
```




```
# Visualize (Setting: Gender)
nypd_data_clean %>% ggplot(aes(x = LOCATION_DESC, y = VIC_SEX)) +
  geom_line(aes(color = "LOCATION_DESC")) +
  geom_point(aes(color = "LOCATION_DESC")) +
  theme(legend.position="bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "Setting versus Gender (Victim)", y = NULL)
```

Transform (Data: Post-visualization/analysis) (i.e. analysis, additional)

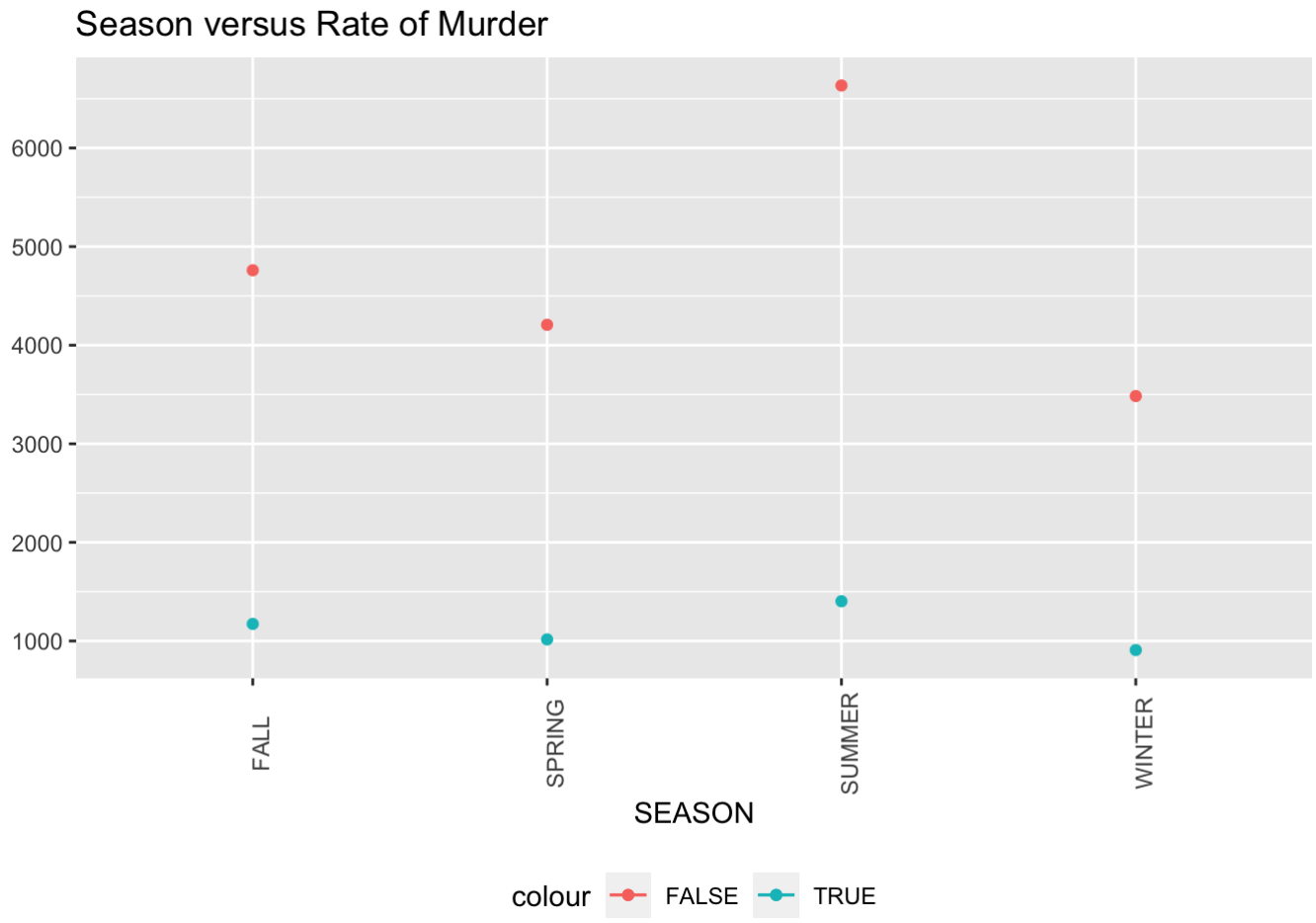
```
# Extract (Month)
nypd_data_transformed <- nypd_data_clean %>% mutate (OCCUR_MONTH = as.integer((month(OCC
UR_DATE))))

# Calculate (Season)
nypd_data_transformed <- nypd_data_transformed %>% mutate (SEASON = ifelse(OCCUR_MONTH >
=3 & OCCUR_MONTH <= 5, "SPRING", ifelse(OCCUR_MONTH >=6 & OCCUR_MONTH <= 8, "SUMMER",ife
lse(OCCUR_MONTH >=9 & OCCUR_MONTH <= 11, "FALL", ifelse(OCCUR_MONTH == 12 | OCCUR_MONTH
 <= 2, "WINTER", ""))))

# Group (By: Season)
nypd_data_grouped_season <- nypd_data_transformed %>% group_by(SEASON) %>% summarize(STA
TISTICAL_MURDER_FLAG_TRUE = sum(ifelse(STATISTICAL_MURDER_FLAG == TRUE, 1, 0)), STATISTI
CAL_MURDER_FLAG_FALSE = sum(ifelse(STATISTICAL_MURDER_FLAG == FALSE, 1, 0)))

nypd_data_grouped_season
```

```
## # A tibble: 4 × 3
##   SEASON STATISTICAL_MURDER_FLAG_TRUE STATISTICAL_MURDER_FLAG_FALSE
##   <chr>                         <dbl>                         <dbl>
## 1 FALL                           1173                          4760
## 2 SPRING                         1016                          4207
## 3 SUMMER                         1403                          6634
## 4 WINTER                          908                          3484
```

Visualization (Post-analysis; post-transformation)

```
nypd_data_grouped_season %>% ggplot(aes(x = SEASON, y = STATISTICAL_MURDER_FLAG_TRUE)) +
  geom_line(aes(color = "TRUE")) +
  geom_point(aes(color = "TRUE")) +
  geom_line(aes(y = STATISTICAL_MURDER_FLAG_FALSE, color = "FALSE")) +
  geom_point(aes(y = STATISTICAL_MURDER_FLAG_FALSE, color = "FALSE")) +
  theme(legend.position="bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "Season versus Rate of Murder", y = NULL)
```

```
## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
```



Identification (Bias)

Sources of bias in the data include: (1.) Errors in the data's collection (stemming from potential perceived bias of the recorder). (2.) Errors in the data's reporting (for example: a certain area is more likely to report the a crime which may go unreported in another area).

Bias (Personal): (1.) Personal bias includes assumptions made by the researcher about the data such as the characteristics of its source, including its validity. (2.) Assumptions about the data's meaning (for example, race may be measured subjectively, rather than on a scientifically-based classification).

Summary and Conclusion

In this analysis, we are able to tell whether the setting of a crime is related to the victim's age, race, or gender. After graphing these results, it is possible to see that some specific ages are more likely to be involved in a crime in a certain setting. Similarly, specific races and genders are more likely to be involved in a a crime in a certain setting.

Following this, we are able to use to data to determine the season during which the criminal activities took place. With this information, we once more graph the data and see that the type of crime is potentially related to the season in which it took place.

In conclusion, the NYPD data indicates that a victim's age, race, or gender may be related to the crime's setting, while the type of crime may be related to the season in which it took place.