data exploration

Caitlyn Clendenin

11/29/2020

Initial Police Data Exploration

For my own purposes, the list of variable names and their meaning are below: - `CCR` : incident number - `AGE` : age of suspect - `GENDER` : gender of suspect - `RACE` : race of suspect - `ARREST_DATE` : date of arrest - `ARREST_TIME` : time of arrest - `ARRESTLOCATION` : location of arrest - `OFFENSES` : all offenses listed for the incident - `INCIDENTLOCATION` : incident location - `NEIGHBORHOOD` : neighborhood - `ZONE` : police zone where incident occurred - `X` : The geocoded X coordinate of the incident location - `Y` : The geocoded Y coordinate of the incident location

Summary Statistics

```
head(policedata)
```

```
##        PK      CCR AGE GENDER RACE          ARRESTTIME
## 1 1975272 16158872  42      F    B 2016-08-24T12:20:00
## 2 1974456 16144120  31      M    W 2016-08-03T14:55:00
## 3 1974466 16144165  63      F    B 2016-08-03T16:45:00
## 4 1974550 16145257  25      F    W 2016-08-05T02:36:00
## 5 1974596 16145962  25      M    B 2016-08-06T02:00:00
## 6 1974556 16144301  45      M    W 2016-08-15T13:30:00
##                                   ARRESTLOCATION
## 1        4700 Block Centre AV Pittsburgh, PA 15213
## 2 4200 Block Steubenville PKE Pittsburgh, PA 15205
## 3       900 Block Freeport RD Fox Chapel, PA 15238
## 4      Foreland ST & Cedar AV Pittsburgh, PA 15212
## 5        900 Block Woodlow ST Pittsburgh, PA 15205
## 6            600 Block 1st AV Pittsburgh, PA 15219
##
## 1
## 2
## 3
## 4
## 5 2702 Aggravated Assault. / 2705 Recklessy Endangering Another Person. / 3925 Receiving Stolen Prope
## 6
##                                   INCIDENTLOCATION INCIDENTNEIGHBORHOOD
## 1        4700 Block Centre AV Pittsburgh, PA 15213           Bloomfield
## 2 4200 Block Steubenville PKE Pittsburgh, PA 15205         Outside City
## 3       900 Block Freeport RD Fox Chapel, PA 15238             Westwood
## 4      Foreland ST & Cedar AV Pittsburgh, PA 15212       East Allegheny
## 5        900 Block Woodlow ST Pittsburgh, PA 15205      Crafton Heights
## 6      800 Block Hazelwood AV Pittsburgh, PA 15217           Greenfield
##   INCIDENTZONE INCIDENTTRACT COUNCIL_DISTRICT PUBLIC_WORKS_DIVISION         X
## 1            5           804                8                     2 -79.94928
```

```
## 2          OSC          5599               NA                    NA -80.08802
## 3            5          2811                9                     2 -79.89180
## 4            1          2304                1                     1 -80.00194
## 5            5          2814                2                     5 -80.05220
## 6            4          1517                5                     3 -79.92924
##          Y
## 1 40.45255
## 2 40.44014
## 3 40.48662
## 4 40.45408
## 5 40.44590
## 6 40.41970
```

```
summary(policedata)
```

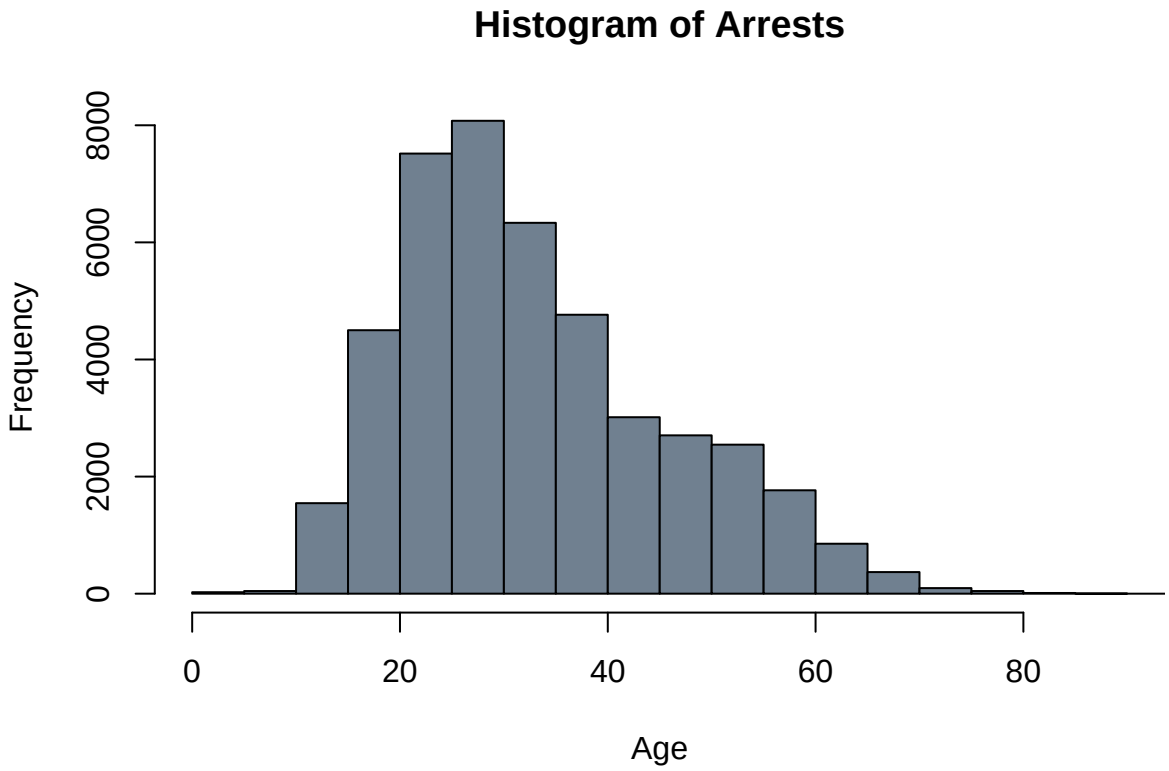
```
##        PK                CCR                  AGE          GENDER    RACE
##  Min.   :1974435   Min.   :  2186691   Min.   :  0.00   F:11822   A:  194
##  1st Qu.:1988515   1st Qu.: 17121470   1st Qu.: 24.00   M:32671   B:26976
##  Median :2002803   Median : 18112188   Median : 31.00   U:   48   H:  421
##  Mean   :2002874   Mean   : 18132479   Mean   : 33.47             I:   11
##  3rd Qu.:2017124   3rd Qu.: 19133843   3rd Qu.: 41.00             O:  498
##  Max.   :2031989   Max.   :180752926   Max.   :999.00             U:  317
##                                        NA's   :324                W:16124
##                ARRESTTIME                                        ARRESTLOCATION
##  2020-05-30T22:00:00:   42   600 Block 1st AV Pittsburgh, PA 15219   : 4191
##  2017-03-19T08:20:00:   17   900 Block 2nd AV Pittsburgh, PA 15219   : 2552
##  2017-08-08T21:45:00:   16   600 Block 1ST AV PITTSBURGH, PA 15219   : 1320
##  2018-04-03T18:36:00:   14   900 Block Second AV Pittsburgh, PA 15219:  844
##  2018-10-26T14:44:00:   13   Zone 2                                  :  774
##  2019-10-23T09:25:00:   13   900 Block 2ND AV Pittsburgh, PA 15219   :  698
##  (Other)            :44426   (Other)                                 :34162
##                                                                                OFFENSES
##  2701 Simple Assault.                                                             : 2959
##  9501 Bench Warrant                                                               : 2273
##  9015 Failure To Appear/Arrest on Attachment Order                                : 2154
##  13(a)(31) Marijuana: Possession Small Amount                                     : 1051
##  13(a)(32) Paraphernalia - Use or Possession                                      :  953
##  13(a)(16) Possession of Controlled Substance / 13(a)(32) Paraphernalia - Use or Possession:  758
##  (Other)                                                                          :34393
##                                INCIDENTLOCATION
##  Zone 1                                  :  350
##  Zone 3                                  :  318
##  Zone 2                                  :  275
##  Zone 5                                  :  273
##  300 Block Cedar AV Pittsburgh, PA 15212:  254
##  6200 Block Penn AV Pittsburgh, PA 15206:  243
##  (Other)                                 :42828
##                 INCIDENTNEIGHBORHOOD  INCIDENTZONE   INCIDENTTRACT
##  Central Business District: 2723     1      :10073   Min.   : 103
##  South Side Flats         : 2306     5      : 8922   1st Qu.:1113
##                           : 1671     3      : 8864   Median :1702
##  Carrick                  : 1615     2      : 7170   Mean   :1649
##  East Allegheny           : 1512     4      : 4692   3rd Qu.:2304
##  Homewood South           : 1446     6      : 4455   Max.   :9822
##  (Other)                  :33268     (Other):  365   NA's   :1649
```

```
##  COUNCIL_DISTRICT PUBLIC_WORKS_DIVISION       X               Y
##  Min.   :1.000    Min.   :0.000         Min.   :-80.49   Min.   : 0.00
##  1st Qu.:2.000    1st Qu.:2.000         1st Qu.:-80.00   1st Qu.:40.42
##  Median :5.000    Median :3.000         Median :-79.98   Median :40.45
##  Mean   :4.799    Mean   :2.953         Mean   :-72.81   Mean   :36.82
##  3rd Qu.:7.000    3rd Qu.:5.000         3rd Qu.:-79.92   3rd Qu.:40.46
##  Max.   :9.000    Max.   :6.000         Max.   :  0.00   Max.   :41.23
##  NA's   :6095     NA's   :6095          NA's   :1580     NA's   :1580
```

One interesting takeaway is that the neighborhood with the highest number of incidents reported is the CentralBusiness District. This makes sense because it is one of the most urban neighborhoods, so there is a high crime rate. This could be a good option to intially rule out because in terms of safety, it is likely not the best.

Histogram of Age Groups

```
hist(policedata$AGE, xlim = c(0,90), breaks = 150, col = "slategrey",
     xlab = "Age", main = "Histogram of Arrests")
```



Tables of Categorical Variables

First, a relative frequency table of gender which shows that the majority of suspects were male.

```
round(table(policedata$GENDER)/length(policedata$GENDER),3)
```

```
##
##     F     M     U
## 0.265 0.734 0.001
```

Next, a relative frequency table of race which shows that the groups with the largest proportion of arrests were Black (60.6%) and White (36.2). The populations in Pittsburgh are 66.85% White, 23.21% Black/African

American, 5.71% Asian, and the rest of the population either falls into two or more race categories or Native American.

```
round(table(policedata$RACE)/length(policedata$GENDER),4)
```

```
##
##      A      B      H      I      O      U      W
## 0.0044 0.6056 0.0095 0.0002 0.0112 0.0071 0.3620
```

I noticed that the dataset with traffic signs that I found also uses police zones and public works divisions, so this could provide an identification method for comparing datasets side by side. I am going to see which police zones and public works divisions have the highest number of traffic signs. This may be an environmental factor that contributes to safety