DATA 621: BUSINESS ANALYTICS AND DATA MINING FINA PROJECT

Group 2 - Gabriel Campos, Melissa Bowman, Alexander Khaykin, & Jennifer Abinette

Last edited December 13, 2023

Load the data

hate_crime.csv

```
df_hate_crime <-
  read.csv(paste0(git_url,"hate_crime.csv"))
head(df_hate_crime,n=3)
```

```
##   incident_id data_year      ori   pug_agency_name pub_agency_unit
## 1          43      1991 AR0350100        Pine Bluff
## 2          44      1991 AR0350100        Pine Bluff
## 3          45      1991 AR0600300 North Little Rock
##   agency_type_name state_abbr state_name     division_name region_name
## 1             City         AR   Arkansas West South Central       South
## 2             City         AR   Arkansas West South Central       South
## 3             City         AR   Arkansas West South Central       South
##   population_group_code   population_group_description incident_date
## 1                     3 Cities from 50,000 thru 99,999      7/4/1991
## 2                     3 Cities from 50,000 thru 99,999    12/24/1991
## 3                     3 Cities from 50,000 thru 99,999     7/10/1991
##   adult_victim_count juvenile_victim_count total_offender_count
## 1                 NA                    NA                    1
## 2                 NA                    NA                    1
## 3                 NA                    NA                    1
##   adult_offender_count juvenile_offender_count          offender_race
## 1                   NA                      NA Black or African American
## 2                   NA                      NA Black or African American
## 3                   NA                      NA Black or African American
##   offender_ethnicity victim_count
## 1      Not Specified            1
## 2      Not Specified            2
## 3      Not Specified            2
##                                                 offense_name
## 1                                         Aggravated Assault
## 2 Aggravated Assault;Destruction/Damage/Vandalism of Property
## 3     Aggravated Assault;Murder and Nonnegligent Manslaughter
##   total_individual_victims                     location_name
```

```
## 1                            1                Residence/Home
## 2                            1 Highway/Road/Alley/Street/Sidewalk
## 3                            2                Residence/Home
##                       bias_desc victim_types multiple_offense multiple_bias
## 1 Anti-Black or African American   Individual                S             S
## 2                     Anti-White   Individual                M             S
## 3                     Anti-White   Individual                M             S
```

Summary data

df_hate_crime

Summary

```
##    incident_id        data_year         ori             pug_agency_name
## Min.    :      2   Min.   :1991   Length:241663      Length:241663
## 1st Qu.:  60446   1st Qu.:1999   Class :character   Class :character
## Median :  120873   Median :2006   Mode  :character   Mode  :character
## Mean    : 349025   Mean   :2007
## 3rd Qu.: 181301   3rd Qu.:2016
## Max.    :1494167   Max.   :2022
##
## pub_agency_unit    agency_type_name    state_abbr         state_name
## Length:241663      Length:241663      Length:241663      Length:241663
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
## division_name      region_name        population_group_code
## Length:241663      Length:241663      Length:241663
## Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character
##
##
##
##
## population_group_description incident_date       adult_victim_count
## Length:241663               Length:241663      Min.    :   0.00
## Class :character            Class :character   1st Qu.:   0.00
## Mode  :character            Mode  :character   Median :   1.00
##                                                Mean    :   0.73
##                                                3rd Qu.:   1.00
##                                                Max.    : 146.00
##                                                NA's    :170538
## juvenile_victim_count total_offender_count adult_offender_count
## Min.    : 0.0         Min.    : 0.0000     Min.    : 0.00
## 1st Qu.: 0.0         1st Qu.: 0.0000     1st Qu.: 0.00
## Median : 0.0         Median : 1.0000     Median : 0.00
## Mean    : 0.1         Mean    : 0.9559     Mean    : 0.61
## 3rd Qu.: 0.0         3rd Qu.: 1.0000     3rd Qu.: 1.00
```

```
## Max.    :60.0          Max.    :99.0000      Max.    :60.00
## NA's    :172978                              NA's    :177148
## juvenile_offender_count offender_race        offender_ethnicity
## Min.    : 0.00            Length:241663      Length:241663
## 1st Qu.: 0.00            Class :character    Class :character
## Median : 0.00            Mode  :character    Mode  :character
## Mean    : 0.12
## 3rd Qu.: 0.00
## Max.    :20.00
## NA's    :177155
##   victim_count      offense_name        total_individual_victims
## Min.    :  1.000   Length:241663      Min.    :  0.000
## 1st Qu.:  1.000   Class :character    1st Qu.:  1.000
## Median :  1.000   Mode  :character    Median :  1.000
## Mean    :  1.242                      Mean    :  0.989
## 3rd Qu.:  1.000                       3rd Qu.:  1.000
## Max.    :900.000                      Max.    :147.000
##                                       NA's    :4859
##  location_name       bias_desc         victim_types        multiple_offense
## Length:241663     Length:241663      Length:241663      Length:241663
## Class :character   Class :character    Class :character    Class :character
## Mode  :character   Mode  :character    Mode  :character    Mode  :character
##
##
##
##
##  multiple_bias
## Length:241663
## Class :character
## Mode  :character
##
##
##
##
```

skim

```
## -- Data Summary ------------------------
##                        Values
## Name                   df_hate_crime
## Number of rows         241663
## Number of columns      28
## _____
## Column type frequency:
##   character            19
##   numeric              9
## _____
## Group variables        None
##
## -- Variable type: character ----------------------------------------------------
##   skim_variable            n_missing complete_rate min max  empty n_unique
## 1 ori                              0             1   9   9      0    10409
## 2 pug_agency_name                  0             1   3  67      0     6942
```

3

```
##  3 pub_agency_unit                     0            1  0  37 234474       690
##  4 agency_type_name                    0            1  4  21      0         8
##  5 state_abbr                          0            1  2   2      0        53
##  6 state_name                          0            1  4  20      0        53
##  7 division_name                       0            1  5  18      0        11
##  8 region_name                         0            1  4  16      0         6
##  9 population_group_code               0            1  0   2    555        21
## 10 population_group_description        0            1  0  67    555        21
## 11 incident_date                       0            1  8  10      0     11688
## 12 offender_race                       0            1  5  41      0         8
## 13 offender_ethnicity                  0            1  7  22      0         5
## 14 offense_name                        0            1  4 125      0       399
## 15 location_name                       0            1  9  81      0       149
## 16 bias_desc                           0            1  9 163      0       373
## 17 victim_types                        0            1  5  53      0        55
## 18 multiple_offense                    0            1  1   1      0         2
## 19 multiple_bias                       0            1  1   1      0         2
##    whitespace
## 1           0
## 2           0
## 3           0
## 4           0
## 5           0
## 6           0
## 7           0
## 8           0
## 9           0
## 10          0
## 11          0
## 12          0
## 13          0
## 14          0
## 15          0
## 16          0
## 17          0
## 18          0
## 19          0
##
## -- Variable type: numeric ---------------------------------------------------
##    skim_variable           n_missing complete_rate     mean       sd    p0
## 1 incident_id                      0             1  349025.  517581.     2
## 2 data_year                        0             1    2007.     9.41  1991
## 3 adult_victim_count          170538         0.294    0.729    1.13     0
## 4 juvenile_victim_count       172978         0.284    0.101    0.502     0
## 5 total_offender_count             0             1    0.956    1.32     0
## 6 adult_offender_count        177148         0.267    0.609    0.820     0
## 7 juvenile_offender_count     177155         0.267    0.118    0.516     0
## 8 victim_count                     0             1    1.24     2.18     1
## 9 total_individual_victims      4859         0.980    0.989    1.15     0
##     p25    p50    p75   p100 hist
## 1 60446. 120873 181300. 1494167
## 2  1999   2006   2016   2022
## 3     0      1      1    146
## 4     0      0      0     60
```
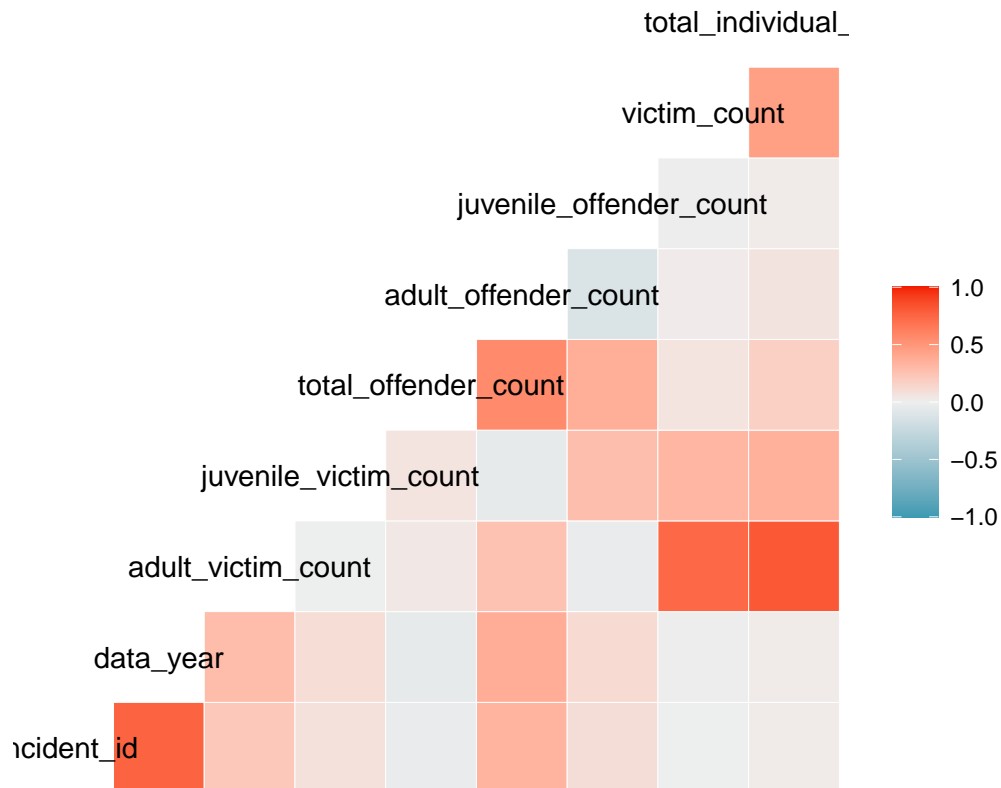
```
## 5       0        1        1       99
## 6       0        0        1       60
## 7       0        0        0       20
## 8       1        1        1      900
## 9       1        1        1      147
```

Correlation plots

df_hate_crime


```
##                               incident_id    data_year          ori*
## incident_id                  1.000000000  0.859283538  0.0782035440
## data_year                    0.859283538  1.000000000 -0.0541109547
## ori*                         0.078203544 -0.054110955  1.0000000000
## pug_agency_name*            -0.033009412 -0.016476665 -0.0175007183
## pub_agency_unit*             0.011678418  0.013319958  0.0209643323
## agency_type_name*            0.015151698  0.001337133  0.0403435447
## state_abbr*                  0.070055311 -0.065496865  0.9902537962
## state_name*                  0.073341933 -0.061295957  0.9869793160
## division_name*               0.081758537  0.118048173 -0.0078020707
## region_name*                 0.044221529  0.105121583 -0.2405972229
## population_group_code*       0.066030872 -0.004435110  0.1434864360
## population_group_description* 0.068212977  0.017623842  0.1184781926
## incident_date*               0.016337309  0.012845945 -0.0023884398
```

```
## adult_victim_count         0.240564646  0.305906299 -0.0360338327
## juvenile_victim_count      0.091156060  0.104788753  0.0110118556
## total_offender_count       0.120630340  0.114078858  0.0088336595
## adult_offender_count       0.358626282  0.420930052 -0.0471162674
## juvenile_offender_count    0.107284492  0.121967394  0.0229087820
## offender_race*            -0.024169655 -0.049473307  0.0161439704
## offender_ethnicity*       -0.231206553 -0.221159707  0.0476549042
## victim_count               0.007030442  0.009421808  0.0095890662
## offense_name*              0.031307583  0.020161170  0.0545426396
## total_individual_victims   0.031261522  0.046810070 -0.0330213790
## location_name*            -0.030816660 -0.036517451  0.0372012294
## bias_desc*                -0.025415287 -0.030729804  0.0398526940
## victim_types*              0.005230162  0.011959708 -0.0239075199
## multiple_offense*         -0.036713099 -0.035619996 -0.0095025495
## multiple_bias*            -0.058309880 -0.068004769  0.0009129238
##                          pug_agency_name* pub_agency_unit*
## incident_id                  -0.0330094118      0.011678418
## data_year                    -0.0164766650      0.013319958
## ori*                         -0.0175007183      0.020964332
## pug_agency_name*              1.0000000000      0.092913856
## pub_agency_unit*              0.0929138559      1.000000000
## agency_type_name*             0.1261319676      0.515040590
## state_abbr*                  -0.0292022314     -0.006638641
## state_name*                  -0.0357618084     -0.010613789
## division_name*                0.0743296331     -0.044157330
## region_name*                  0.1085594222     -0.065310431
## population_group_code*        0.0470853014      0.175193266
## population_group_description*  0.0252659970      0.181896442
## incident_date*               -0.0028696401     -0.002517626
## adult_victim_count           -0.0073968219      0.041460667
## juvenile_victim_count        -0.0153075010      0.012704961
## total_offender_count         -0.0074381662     -0.016622783
## adult_offender_count          0.0052965055      0.011590211
## juvenile_offender_count      -0.0193085442     -0.008951602
## offender_race*               -0.0051278603      0.021488531
## offender_ethnicity*           0.0002357506      0.011905275
## victim_count                 -0.0063421779      0.052067741
## offense_name*                -0.0110151475     -0.005855519
## total_individual_victims     -0.0126020530      0.034328226
## location_name*               -0.0119695054      0.001565919
## bias_desc*                    0.0117210863     -0.011355450
## victim_types*                 0.0212839764      0.019004460
## multiple_offense*            -0.0047833387      0.001321557
## multiple_bias*                0.0077890019     -0.009840455
##                          agency_type_name*    state_abbr*    state_name*
## incident_id                    0.015151698   0.0700553108   0.0733419334
## data_year                      0.001337133  -0.0654968653  -0.0612959574
## ori*                           0.040343545   0.9902537962   0.9869793160
## pug_agency_name*               0.126131968  -0.0292022314  -0.0357618084
## pub_agency_unit*               0.515040590  -0.0066386409  -0.0106137892
## agency_type_name*              1.000000000   0.0387262688   0.0389885469
## state_abbr*                    0.038726269   1.0000000000   0.9967403545
## state_name*                    0.038988547   0.9967403545   1.0000000000
## division_name*                -0.033862794  -0.0396654991  -0.0145030256
```

```
## region_name*                   -0.064330861 -0.2860219832 -0.2904918657
## population_group_code*           0.439173321  0.1589009672  0.1643553908
## population_group_description*    0.433488644  0.1248424247  0.1381141647
## incident_date*                  -0.028462002 -0.0025670445 -0.0022344085
## adult_victim_count              -0.010696918 -0.0548392320 -0.0513885286
## juvenile_victim_count           -0.007050367  0.0111153943  0.0121177211
## total_offender_count            -0.032366445  0.0057312958  0.0068583573
## adult_offender_count            -0.015148654 -0.0566887714 -0.0515005049
## juvenile_offender_count         -0.012074910  0.0243729916  0.0258735513
## offender_race*                   0.019058806  0.0167363453  0.0145378465
## offender_ethnicity*              0.030819477  0.0568864376  0.0573488231
## victim_count                     0.012734823 -0.0019640842 -0.0014331698
## offense_name*                    0.003810604  0.0590152587  0.0594387832
## total_individual_victims        -0.027881773 -0.0478557080 -0.0458888150
## location_name*                   0.047341377  0.0459196797  0.0445136594
## bias_desc*                      -0.000764052  0.0398576810  0.0385598630
## victim_types*                   -0.011114337 -0.0209149590 -0.0245637442
## multiple_offense*                0.008252910 -0.0088763224 -0.0106210792
## multiple_bias*                  -0.004291268  0.0008180858 -0.0004266876
##                                division_name*   region_name*
## incident_id                       0.081758537  0.0442215294
## data_year                         0.118048173  0.1051215825
## ori*                             -0.007802071 -0.2405972229
## pug_agency_name*                  0.074329633  0.1085594222
## pub_agency_unit*                 -0.044157330 -0.0653104313
## agency_type_name*                -0.033862794 -0.0643308606
## state_abbr*                      -0.039665499 -0.2860219832
## state_name*                      -0.014503026 -0.2904918657
## division_name*                    1.000000000  0.3797693397
## region_name*                      0.379769340  1.0000000000
## population_group_code*           -0.006282367 -0.1222555193
## population_group_description*     0.022095684 -0.0890641537
## incident_date*                   -0.002799686 -0.0078113900
## adult_victim_count                0.051073644  0.0600593836
## juvenile_victim_count             0.001027945 -0.0202629208
## total_offender_count              0.035136149  0.0066141899
## adult_offender_count              0.073477162  0.0731449435
## juvenile_offender_count           0.004468837 -0.0245165744
## offender_race*                   -0.015696115  0.0123073273
## offender_ethnicity*              -0.084945718 -0.1215091449
## victim_count                      0.004366312  0.0009202725
## offense_name*                    -0.039422714 -0.0677102983
## total_individual_victims          0.023396908  0.0305618242
## location_name*                   -0.053439816 -0.0712275401
## bias_desc*                       -0.001112582 -0.0032118554
## victim_types*                    -0.022220599 -0.0182386848
## multiple_offense*                -0.005702567 -0.0372944930
## multiple_bias*                   -0.002576454 -0.0057061783
##                                population_group_code*
## incident_id                             0.0660308718
## data_year                              -0.0044351104
## ori*                                    0.1434864360
## pug_agency_name*                        0.0470853014
## pub_agency_unit*                        0.1751932658
```

```
## agency_type_name*                                0.4391733213
## state_abbr*                                       0.1589009672
## state_name*                                       0.1643553908
## division_name*                                   -0.0062823673
## region_name*                                     -0.1222555193
## population_group_code*                            1.0000000000
## population_group_description*                     0.6464337848
## incident_date*                                   -0.0118307283
## adult_victim_count                               -0.0511578838
## juvenile_victim_count                             0.0353400662
## total_offender_count                             -0.0120952230
## adult_offender_count                             -0.0245248640
## juvenile_offender_count                           0.0478315130
## offender_race*                                    0.0970994925
## offender_ethnicity*                               0.0211970831
## victim_count                                     -0.0009439764
## offense_name*                                     0.0157420116
## total_individual_victims                         -0.0436928227
## location_name*                                    0.0536018366
## bias_desc*                                        0.0117638109
## victim_types*                                     0.0162056500
## multiple_offense*                                -0.0114826519
## multiple_bias*                                    0.0100086984
##                               population_group_description* incident_date*
## incident_id                                      6.821298e-02   1.633731e-02
## data_year                                        1.762384e-02   1.284595e-02
## ori*                                             1.184782e-01  -2.388440e-03
## pug_agency_name*                                 2.526600e-02  -2.869640e-03
## pub_agency_unit*                                 1.818964e-01  -2.517626e-03
## agency_type_name*                                4.334886e-01  -2.846200e-02
## state_abbr*                                      1.248424e-01  -2.567044e-03
## state_name*                                      1.381142e-01  -2.234409e-03
## division_name*                                   2.209568e-02  -2.799686e-03
## region_name*                                    -8.906415e-02  -7.811390e-03
## population_group_code*                           6.464338e-01  -1.183073e-02
## population_group_description*                    1.000000e+00   2.520422e-05
## incident_date*                                   2.520422e-05   1.000000e+00
## adult_victim_count                              -3.573302e-03   2.026499e-02
## juvenile_victim_count                            8.042618e-03  -9.924006e-03
## total_offender_count                             1.087793e-02   1.771953e-02
## adult_offender_count                             3.091904e-02   2.280396e-02
## juvenile_offender_count                          1.812902e-02  -1.446913e-02
## offender_race*                                   1.147272e-02  -5.992684e-03
## offender_ethnicity*                              4.700879e-02  -8.686785e-03
## victim_count                                     3.694667e-03   1.102074e-02
## offense_name*                                    1.240057e-02  -1.057336e-02
## total_individual_victims                        -1.952356e-02   1.772207e-02
## location_name*                                   2.152739e-02  -1.268080e-02
## bias_desc*                                       2.243239e-02  -6.268712e-03
## victim_types*                                    2.961477e-02   3.912779e-03
## multiple_offense*                               -2.314861e-02  -9.790237e-03
## multiple_bias*                                   3.192323e-03   3.486868e-03
##                               adult_victim_count juvenile_victim_count
## incident_id                          0.240564646            0.091156060
```

```
## data_year                       0.305906299          0.104788753
## ori*                           -0.036033833          0.011011856
## pug_agency_name*               -0.007396822         -0.015307501
## pub_agency_unit*                0.041460667          0.012704961
## agency_type_name*              -0.010696918         -0.007050367
## state_abbr*                    -0.054839232          0.011115394
## state_name*                    -0.051388529          0.012117721
## division_name*                  0.051073644          0.001027945
## region_name*                    0.060059384         -0.020262921
## population_group_code*         -0.051157884          0.035340066
## population_group_description*  -0.003573302          0.008042618
## incident_date*                  0.020264990         -0.009924006
## adult_victim_count              1.000000000          0.002070753
## juvenile_victim_count           0.002070753          1.000000000
## total_offender_count            0.052539902          0.060780812
## adult_offender_count            0.262432557         -0.027369011
## juvenile_offender_count        -0.034205577          0.262764881
## offender_race*                 -0.030706605          0.003443780
## offender_ethnicity*            -0.054980259         -0.035061703
## victim_count                    0.750578331          0.330350353
## offense_name*                  -0.019710482          0.037979015
## total_individual_victims        0.807340056          0.352900235
## location_name*                 -0.028679019          0.056891130
## bias_desc*                     -0.019967909         -0.022250843
## victim_types*                   0.048472399          0.017901240
## multiple_offense*              -0.092430867         -0.013483774
## multiple_bias*                 -0.054899100         -0.014599145
##                               total_offender_count adult_offender_count
## incident_id                     0.120630340          0.358626282
## data_year                       0.114078858          0.420930052
## ori*                            0.008833660         -0.047116267
## pug_agency_name*               -0.007438166          0.005296505
## pub_agency_unit*               -0.016622783          0.011590211
## agency_type_name*              -0.032366445         -0.015148654
## state_abbr*                     0.005731296         -0.056688771
## state_name*                     0.006858357         -0.051500505
## division_name*                  0.035136149          0.073477162
## region_name*                    0.006614190          0.073144944
## population_group_code*         -0.012095223         -0.024524864
## population_group_description*   0.010877926          0.030919037
## incident_date*                  0.017719525          0.022803959
## adult_victim_count              0.052539902          0.262432557
## juvenile_victim_count           0.060780812         -0.027369011
## total_offender_count            1.000000000          0.543663968
## adult_offender_count            0.543663968          1.000000000
## juvenile_offender_count         0.360546602         -0.121649525
## offender_race*                 -0.126959327         -0.084096188
## offender_ethnicity*            -0.152081901         -0.219812323
## victim_count                    0.057667554          0.041069427
## offense_name*                   0.016229935         -0.005089272
## total_individual_victims        0.083210476          0.067477416
## location_name*                 -0.030292138         -0.072928164
## bias_desc*                      0.003760939         -0.016580056
## victim_types*                   0.027170242          0.034870223
```

```
## multiple_offense*                           -0.044124118        -0.054136556
## multiple_bias*                               -0.023563318        -0.030537219
##                               juvenile_offender_count offender_race*
## incident_id                              0.107284492  -0.024169655
## data_year                                0.121967394  -0.049473307
## ori*                                     0.022908782   0.016143970
## pug_agency_name*                        -0.019308544  -0.005127860
## pub_agency_unit*                        -0.008951602   0.021488531
## agency_type_name*                       -0.012074910   0.019058806
## state_abbr*                              0.024372992   0.016736345
## state_name*                             0.025873551   0.014537847
## division_name*                          0.004468837  -0.015696115
## region_name*                           -0.024516574   0.012307327
## population_group_code*                  0.047831513   0.097099492
## population_group_description*           0.018129024   0.011472717
## incident_date*                         -0.014469130  -0.005992684
## adult_victim_count                     -0.034205577  -0.030706605
## juvenile_victim_count                   0.262764881   0.003443780
## total_offender_count                    0.360546602  -0.126959327
## adult_offender_count                   -0.121649525  -0.084096188
## juvenile_offender_count                 1.000000000  -0.048084094
## offender_race*                         -0.048084094   1.000000000
## offender_ethnicity*                    -0.077877633  -0.010012641
## victim_count                            0.022307498   0.004394189
## offense_name*                           0.035589824  -0.023753753
## total_individual_victims                0.022960776  -0.015863900
## location_name*                          0.055663164   0.037666390
## bias_desc*                              0.014061844  -0.216089760
## victim_types*                          -0.015229376  -0.025007779
## multiple_offense*                      -0.010526738  -0.002578054
## multiple_bias*                         -0.033288983  -0.009442433
##                               offender_ethnicity*   victim_count offense_name*
## incident_id                         -0.2312065526  0.0070304421  0.0313075834
## data_year                           -0.2211597071  0.0094218081  0.0201611702
## ori*                                 0.0476549042  0.0095890662  0.0545426396
## pug_agency_name*                     0.0002357506 -0.0063421779 -0.0110151475
## pub_agency_unit*                     0.0119052750  0.0520677408 -0.0058555194
## agency_type_name*                    0.0308194772  0.0127348234  0.0038106037
## state_abbr*                          0.0568864376 -0.0019640842  0.0590152587
## state_name*                          0.0573488231 -0.0014331698  0.0594387832
## division_name*                      -0.0849457185  0.0043663124 -0.0394227145
## region_name*                        -0.1215091449  0.0009202725 -0.0677102983
## population_group_code*               0.0211970831 -0.0009439764  0.0157420116
## population_group_description*        0.0470087906  0.0036946667  0.0124005690
## incident_date*                      -0.0086867854  0.0110207403 -0.0105733557
## adult_victim_count                  -0.0549802593  0.7505783312 -0.0197104825
## juvenile_victim_count               -0.0350617025  0.3303503526  0.0379790152
## total_offender_count                -0.1520819005  0.0576675543  0.0162299353
## adult_offender_count                -0.2198123230  0.0410694268 -0.0050892723
## juvenile_offender_count             -0.0778776335  0.0223074983  0.0355898237
## offender_race*                      -0.0100126409  0.0043941891 -0.0237537533
## offender_ethnicity*                  1.0000000000 -0.0024189592  0.0013010061
## victim_count                        -0.0024189592  1.0000000000 -0.0429269903
## offense_name*                        0.0013010061 -0.0429269903  1.0000000000
```

```
## total_individual_victims            -0.0235968629  0.9197163629  0.0036030149
## location_name*                        0.0208889352 -0.0061572646 -0.0012823748
## bias_desc*                            0.0192999698 -0.0110481583  0.0117731491
## victim_types*                        -0.0463556272 -0.0025982331  0.0421539893
## multiple_offense*                     0.0090694083 -0.2366278099  0.1239350618
## multiple_bias*                        0.0174483348 -0.0493166840 -0.0008045336
##                              total_individual_victims location_name*
## incident_id                               0.031261522   -0.030816660
## data_year                                 0.046810070   -0.036517451
## ori*                                     -0.033021379    0.037201229
## pug_agency_name*                         -0.012602053   -0.011969505
## pub_agency_unit*                          0.034328226    0.001565919
## agency_type_name*                        -0.027881773    0.047341377
## state_abbr*                              -0.047855708    0.045919680
## state_name*                              -0.045888815    0.044513659
## division_name*                            0.023396908   -0.053439816
## region_name*                              0.030561824   -0.071227540
## population_group_code*                   -0.043692823    0.053601837
## population_group_description*            -0.019523555    0.021527386
## incident_date*                            0.017722067   -0.012680805
## adult_victim_count                        0.807340056   -0.028679019
## juvenile_victim_count                     0.352900235    0.056891130
## total_offender_count                      0.083210476   -0.030292138
## adult_offender_count                      0.067477416   -0.072928164
## juvenile_offender_count                   0.022960776    0.055663164
## offender_race*                           -0.015863900    0.037666390
## offender_ethnicity*                      -0.023596863    0.020888935
## victim_count                              0.919716363   -0.006157265
## offense_name*                             0.003603015   -0.001282375
## total_individual_victims                  1.000000000    0.022398323
## location_name*                            0.022398323    1.000000000
## bias_desc*                               -0.034735239   -0.050704777
## victim_types*                             0.070889690    0.023617072
## multiple_offense*                        -0.109904697    0.011936731
## multiple_bias*                           -0.040382953    0.007595291
##                                 bias_desc* victim_types* multiple_offense*
## incident_id                   -2.541529e-02  5.230162e-03      -0.036713099
## data_year                     -3.072980e-02  1.195971e-02      -0.035619996
## ori*                           3.985269e-02 -2.390752e-02      -0.009502550
## pug_agency_name*               1.172109e-02  2.128398e-02      -0.004783339
## pub_agency_unit*              -1.135545e-02  1.900446e-02       0.001321557
## agency_type_name*             -7.640520e-04 -1.111434e-02       0.008252910
## state_abbr*                    3.985768e-02 -2.091496e-02      -0.008876322
## state_name*                    3.855986e-02 -2.456374e-02      -0.010621079
## division_name*                -1.112582e-03 -2.222060e-02      -0.005702567
## region_name*                  -3.211855e-03 -1.823868e-02      -0.037294493
## population_group_code*         1.176381e-02  1.620565e-02      -0.011482652
## population_group_description*  2.243239e-02  2.961477e-02      -0.023148610
## incident_date*                -6.268712e-03  3.912779e-03      -0.009790237
## adult_victim_count            -1.996791e-02  4.847240e-02      -0.092430867
## juvenile_victim_count         -2.225084e-02  1.790124e-02      -0.013483774
## total_offender_count           3.760939e-03  2.717024e-02      -0.044124118
## adult_offender_count          -1.658006e-02  3.487022e-02      -0.054136556
## juvenile_offender_count        1.406184e-02 -1.522938e-02      -0.010526738
```

```
## offender_race*                -2.160898e-01 -2.500778e-02      -0.002578054
## offender_ethnicity*            1.929997e-02 -4.635563e-02       0.009069408
## victim_count                  -1.104816e-02 -2.598233e-03      -0.236627810
## offense_name*                  1.177315e-02  4.215399e-02       0.123935062
## total_individual_victims      -3.473524e-02  7.088969e-02      -0.109904697
## location_name*                -5.070478e-02  2.361707e-02       0.011936731
## bias_desc*                     1.000000e+00 -9.687026e-06       0.001266085
## victim_types*                 -9.687026e-06  1.000000e+00      -0.002140859
## multiple_offense*              1.266085e-03 -2.140859e-03       1.000000000
## multiple_bias*                 3.718168e-02  1.067044e-02       0.090336564
##                                multiple_bias*
## incident_id                     -0.0583098804
## data_year                       -0.0680047688
## ori*                             0.0009129238
## pug_agency_name*                 0.0077890019
## pub_agency_unit*                -0.0098404551
## agency_type_name*               -0.0042912685
## state_abbr*                      0.0008180858
## state_name*                     -0.0004266876
## division_name*                  -0.0025764540
## region_name*                    -0.0057061783
## population_group_code*           0.0100086984
## population_group_description*    0.0031923229
## incident_date*                   0.0034868682
## adult_victim_count              -0.0548990996
## juvenile_victim_count           -0.0145991450
## total_offender_count            -0.0235633180
## adult_offender_count            -0.0305372188
## juvenile_offender_count         -0.0332889826
## offender_race*                  -0.0094424334
## offender_ethnicity*              0.0174483348
## victim_count                    -0.0493166840
## offense_name*                   -0.0008045336
## total_individual_victims        -0.0403829528
## location_name*                   0.0075952911
## bias_desc*                       0.0371816785
## victim_types*                    0.0106704429
## multiple_offense*                0.0903365639
## multiple_bias*                   1.0000000000
```

Binary Column and Count

```
## Number of Anti_semitic column reports in hatecrime.csv is: 30469  out of  241663
```

Transform to Numeric | Test/Training Dataframes

```r
df_hate_corr <- char2numeric(df_hate_crime)
```

```r
features <- df_hate_corr[, !(names(df_hate_corr) %in% c('Anti_semitic'))]
labels <- df_hate_corr$Anti_semitic_crimes
set.seed(123)
index <- createDataPartition(labels, p = 0.8, list = FALSE)

#training data
df_hate_training <- features[index, ]
labels_hate_training <- labels[index]

# Create testing dataset
df_hate_testing <- features[-index, ]
labels_hate_testing <- labels[-index]
```

Cleanup

```r
df_hate_training <- df_hate_training %>% dplyr::select(-c(adult_victim_count,juvenile_victim_count,adul
df_hate_testing <- df_hate_testing %>% dplyr::select(-c(adult_victim_count,juvenile_victim_count,adult_c
```

Histograms

```r
df_hate_training %>%
  gather() %>%
  ggplot(aes(x = value)) +
  geom_histogram(fill = "cadetblue") +
  facet_wrap(~key, scales = "free")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```