TidyverseMG

10/25/2020

```
knitr::opts_chunk$set(echo = TRUE)
```

Tidyverse CREATE assignment

The purpose of this assignment is to collaborate around a code project with GitHub. We are to build an example on how to use TidyVerse functions.

Our task is to create a "vignette" that demonstrates how to use one or more of the capabilities of a selected TidyVerse package with a dataset from fivethirtyeight.com or Kaggle.

##The first step is to load TidyVerse.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------------------------- tidyve

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------------------------------------------ tidyverse_co
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

##The next step is to read the csv file from fivethirtyeight.com and take a look at the data.

The file is located at: https://raw.githubusercontent.com/fivethirtyeight/data/master/police-locals/police-locals.csv

```
df <- read.csv("https://raw.githubusercontent.com/fivethirtyeight/data/master/police-locals/police-local
head(df, 10)
```

```
##            city police_force_size       all      white non.white       black
## 1      New York             32300 0.6179567 0.44638656 0.7644189 0.770891365
## 2       Chicago             12120 0.8750000 0.87196262 0.8774003  0.89740566
## 3   Los Angeles             10100 0.2282178 0.15277778 0.2638484 0.387387387
## 4    Washington              9340 0.1156317 0.05677419 0.1573651 0.170189099
## 5       Houston              7700 0.2922078 0.17373461 0.3992583  0.36637931
## 6  Philadelphia              6045 0.8354012 0.77689873 0.8994801 0.924657534
## 7       Phoenix              4475 0.3117318 0.27080182 0.4273504  0.52173913
## 8     San Diego              4460 0.3621076 0.37298387 0.3484848 0.538461538
## 9        Dallas              3605 0.1914008 0.17150396 0.2134503 0.214634146
```

```
## 10      Detroit              3265 0.3705972 0.08196721 0.5427873       0.568
##       hispanic       asian
## 1  0.762860728 0.749235474
## 2   0.83982684 0.966666667
## 3  0.217679558 0.305263158
## 4   0.08988764 0.230769231
## 5  0.457142857 0.408163265
## 6  0.817391304          **
## 7  0.427710843          **
## 8  0.297794118    0.515625
## 9  0.256880734          **
## 10 0.333333333          **
```

```r
tail(df, 10)
```

```
##                        city police_force_size       all      white non.white
## 66               St. Louis               950 0.5894737 0.53846154 0.6712329
## 67      Brownsville, Texas               925 0.5135135 0.50000000 0.5141243
## 68            Albany, N.Y.               890 0.1853933 0.16025641 0.3636364
## 69 Colorado Springs, Colo.               860 0.6046512 0.55303030 0.7750000
## 70            Savannah, Ga.              860 0.2151163 0.07692308 0.2990654
## 71     Winston-Salem, N.C.               860 0.5755814 0.42477876 0.8644068
## 72            Toledo, Ohio               805 0.5652174 0.53076923 0.7096774
## 73            Madison, Wis.              790 0.2784810 0.24647887 0.5625000
## 74   Corpus Christi, Texas               770 0.8571429 0.89333333 0.8227848
## 75  San Bernardino, Calif.               755 0.2715232 0.26315789 0.2800000
##          black    hispanic asian
## 66 0.682539683          **    **
## 67          ** 0.520231214    **
## 68          **          **    **
## 69          ** 0.913043478    **
## 70 0.170731707        0.75    **
## 71 0.869565217          **    **
## 72        0.75          **    **
## 73          **          **    **
## 74          ** 0.847222222    **
## 75          ** 0.274509804    **
```

The data from fivethirtyeight.com is the residence of police officers for the 75 largest police forces in the US (except for Honolulu). The data is further broken down by the officer's race.

We can further see that we have some missing data signified by "**" where there are less than 100 police officers of that race in the police force.

Remove unneeded columns

For this data set, I decided to remove the specific races to just focus on white and non-white officers as the data towards the bottom of the list had too many missing data points to be helpful.

In TidyVerse, the select(dataframe, -c(columns)) function will allow us to remove the unwanted columns.

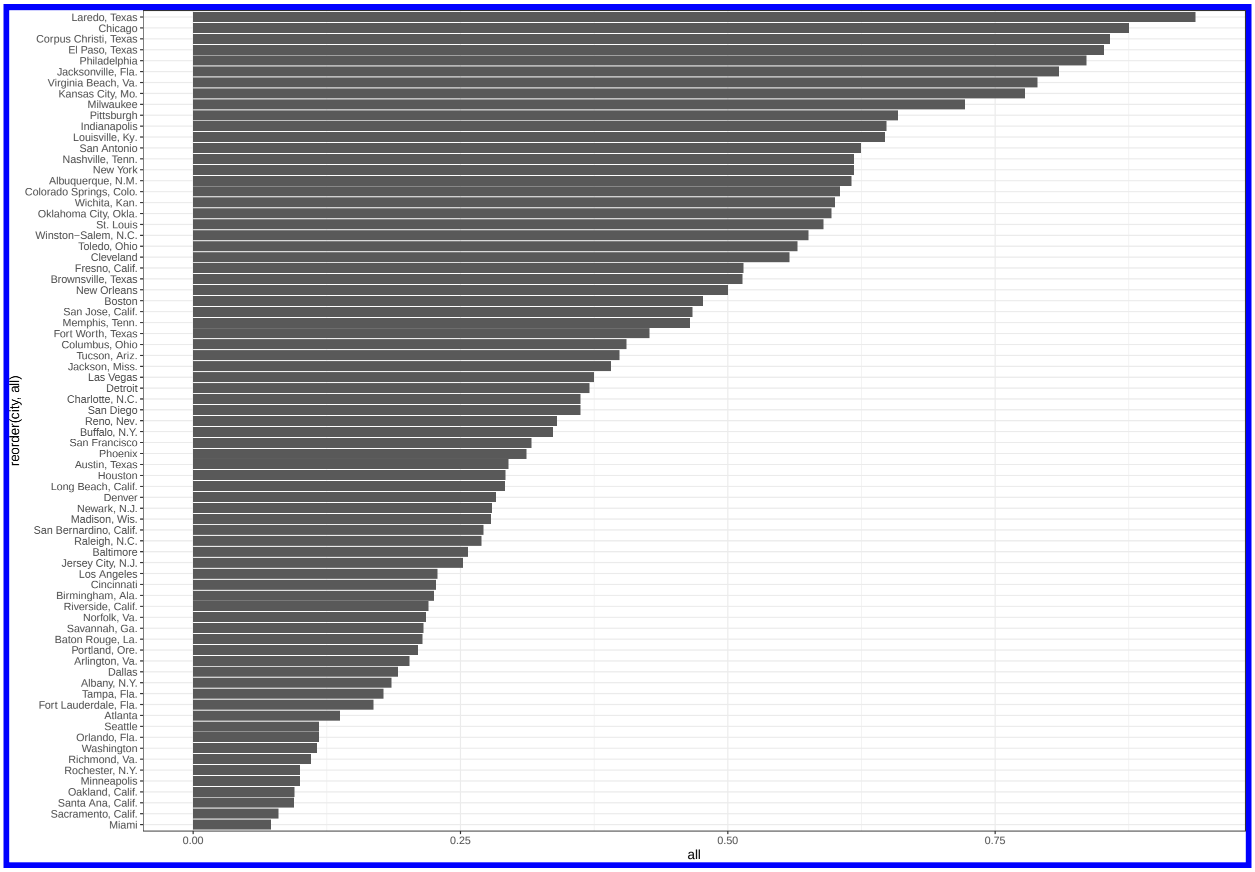
```r
df2 <- select(df, -c(black, hispanic, asian))
df2
```

```
##                      city police_force_size        all      white  non.white
## 1                New York             32300 0.61795666 0.44638656 0.76441894
## 2                 Chicago             12120 0.87500000 0.87196262 0.87740030
## 3             Los Angeles             10100 0.22821782 0.15277778 0.26384840
## 4              Washington              9340 0.11563169 0.05677419 0.15736505
## 5                 Houston              7700 0.29220779 0.17373461 0.39925834
## 6            Philadelphia              6045 0.83540116 0.77689873 0.89948007
## 7                 Phoenix              4475 0.31173184 0.27080182 0.42735043
## 8               San Diego              4460 0.36210762 0.37298387 0.34848485
## 9                  Dallas              3605 0.19140083 0.17150396 0.21345029
## 10                Detroit              3265 0.37059724 0.08196721 0.54278729
## 11          San Francisco              3020 0.31622517 0.25949367 0.37847222
## 12            San Antonio              2955 0.62436548 0.44387755 0.71392405
## 13                Atlanta              2950 0.13728814 0.18627451 0.11139896
## 14              Las Vegas              2830 0.37455830 0.40000000 0.30769231
## 15              Baltimore              2800 0.25714286 0.13281250 0.36184211
## 16                 Boston              2560 0.47656250 0.44155844 0.58267716
## 17      Jacksonville, Fla.             2335 0.80942184 0.71378092 0.95652174
## 18          El Paso, Texas             2260 0.85176991 0.82644628 0.86102719
## 19          Columbus, Ohio             2245 0.40534521 0.37978142 0.51807229
## 20              Cleveland              2045 0.55745721 0.49812734 0.66901409
## 21           Tucson, Ariz.             2020 0.39851485 0.41666667 0.37500000
## 22            Newark, N.J.             2005 0.27930175 0.20796460 0.37142857
## 23           Austin, Texas             1985 0.29471033 0.19469027 0.42690058
## 24          Memphis, Tenn.             1970 0.46446700 0.33913044 0.64024390
## 25              Milwaukee              1960 0.72193878 0.69288390 0.78400000
## 26        San Jose, Calif.             1875 0.46666667 0.47234043 0.45714286
## 27                  Miami              1860 0.07258064 0.03061224 0.08759124
## 28                 Denver              1820 0.28296703 0.14932127 0.48951049
## 29      Sacramento, Calif.             1820 0.07967033 0.06338028 0.13750000
## 30         Charlotte, N.C.             1780 0.36235955 0.29454546 0.59259259
## 31             Tampa, Fla.             1715 0.17784257 0.13191489 0.27777778
## 32           Indianapolis              1620 0.64814815 0.71042471 0.40000000
## 33       Santa Ana, Calif.             1590 0.09433962 0.05882353 0.12087912
## 34            New Orleans              1560 0.50000000 0.32407407 0.59313726
## 35         Oakland, Calif.             1530 0.09477124 0.02666667 0.16025641
## 36           Orlando, Fla.             1530 0.11764706 0.09000000 0.16981132
## 37   Oklahoma City, Okla.              1500 0.59666667 0.54732510 0.80701754
## 38                Seattle              1445 0.11764706 0.11557789 0.12222222
## 39        Kansas City, Mo.             1440 0.77777778 0.76800000 0.84210526
## 40        Nashville, Tenn.             1440 0.61805556 0.43715847 0.93333333
## 41          Laredo, Texas              1435 0.93728223 0.96296296 0.93133047
## 42       Fort Worth, Texas             1430 0.42657343 0.30674847 0.58536585
## 43         Louisville, Ky.             1430 0.64685315 0.62083333 0.78260870
## 44            Norfolk, Va.             1425 0.21754386 0.26708075 0.15322581
## 45          Arlington, Va.             1360 0.20220588 0.22222222 0.17968750
## 46             Pittsburgh              1350 0.65925926 0.67965368 0.53846154
## 47       Albuquerque, N.M.             1340 0.61567164 0.62962963 0.60150376
## 48       Jersey City, N.J.             1170 0.25213675 0.20645161 0.34177215
## 49          Raleigh, N.C.              1150 0.26956522 0.20634921 0.56097561
## 50        Rochester, N.Y.              1150 0.10000000 0.04093567 0.27118644
## 51             Cincinnati              1145 0.22707424 0.14772727 0.49056604
## 52       Long Beach, Calif.            1115 0.29147982 0.27722772 0.30327869
## 53        Birmingham, Ala.             1110 0.22522523 0.08602150 0.32558139
```

```
## 54           Wichita, Kan.                1075 0.60000000 0.51176471 0.93333333
## 55     Virginia Beach, Va.                1070 0.78971963 0.75625000 0.88888889
## 56          Fresno, Calif.                1040 0.51442308 0.50961539 0.51923077
## 57           Buffalo, N.Y.                1010 0.33663366 0.29239766 0.58064516
## 58             Minneapolis                1000 0.10000000 0.05263158 0.37931034
## 59          Portland, Ore.                1000 0.21000000 0.18644068 0.39130435
## 60              Reno, Nev.                1000 0.34000000 0.32386364 0.45833333
## 61           Richmond, Va.                1000 0.11000000 0.10169491 0.12195122
## 62        Baton Rouge, La.                 980 0.21428571 0.14406780 0.32051282
## 63          Jackson, Miss.                 960 0.39062500 0.08219178 0.57983193
## 64       Riverside, Calif.                 955 0.21989529 0.35000000 0.07692308
## 65   Fort Lauderdale, Fla.                 950 0.16842105 0.22018349 0.09876543
## 66               St. Louis                 950 0.58947368 0.53846154 0.67123288
## 67      Brownsville, Texas                 925 0.51351351 0.50000000 0.51412429
## 68            Albany, N.Y.                 890 0.18539326 0.16025641 0.36363636
## 69 Colorado Springs, Colo.                 860 0.60465116 0.55303030 0.77500000
## 70           Savannah, Ga.                 860 0.21511628 0.07692308 0.29906542
## 71     Winston-Salem, N.C.                 860 0.57558140 0.42477876 0.86440678
## 72            Toledo, Ohio                 805 0.56521739 0.53076923 0.70967742
## 73           Madison, Wis.                 790 0.27848101 0.24647887 0.56250000
## 74   Corpus Christi, Texas                 770 0.85714286 0.89333333 0.82278481
## 75  San Bernardino, Calif.                 755 0.27152318 0.26315789 0.28000000
```

Next we'll use a bar chart to show the percentage of all officers who live in the city they protect. We will use the reorder() function to put the highest percentages at the top. And we'll add a blue backgroud around the barchart with the element_rect() function.

```r
ggplot(df2, aes(x = all, y = reorder(city, all))) +
  geom_bar(stat='identity') +
  theme_bw(base_size = 4.5) +
  theme(plot.background = element_rect(color = "blue", size = 2))
```


Laredo, Texas
Chicago
Corpus Christi, Texas
El Paso, Texas
Philadelphia
Jacksonville, Fla.
Virginia Beach, Va.
Kansas City, Mo.
Milwaukee
Pittsburgh
Indianapolis
Louisville, Ky.
San Antonio
Nashville, Tenn.
New York
Albuquerque, N.M.
Colorado Springs, Colo.
Wichita, Kan.
Oklahoma City, Okla.
St. Louis
Winston−Salem, N.C.
Toledo, Ohio
Cleveland
Fresno, Calif.
Brownsville, Texas
New Orleans
Boston
San Jose, Calif.
Memphis, Tenn.
Fort Worth, Texas
Columbus, Ohio
Tucson, Ariz.
Jackson, Miss.
Las Vegas
Detroit
Charlotte, N.C.
San Diego
Reno, Nev.
Buffalo, N.Y.
San Francisco
Phoenix
Austin, Texas
Houston
Long Beach, Calif.
Denver
Newark, N.J.
Madison, Wis.
San Bernardino, Calif.
Raleigh, N.C.
Baltimore
Jersey City, N.J.
Los Angeles
Cincinnati
Birmingham, Ala.
Riverside, Calif.
Norfolk, Va.
Savannah, Ga.
Baton Rouge, La.
Portland, Ore.
Arlington, Va.
Dallas
Albany, N.Y.
Tampa, Fla.
Fort Lauderdale, Fla.
Atlanta
Seattle
Orlando, Fla.
Washington
Richmond, Va.
Rochester, N.Y.
Minneapolis
Oakland, Calif.
Santa Ana, Calif.
Sacramento, Calif.
Miami
reorder(city, all)
0.00
0.25
0.50
0.75
all