

Midterm EDA: Group 3

2022-10-26

From 2015- 2022, in response to a deep lack of reporting within government sources, The Washington Post compiled a database of every fatal police shooting in the United States. We are interested in exploring this data, specifically as it shows the differences between US States.

Setting the Data Up

First we call our packages: dplyr and ggplot2 as well as reading our data:

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

Then we remove the null values from our dataset

After Accounting for Null Values: The dataset we are working with has 6574 observations.

```
## [1] 6574
```

```
##   id          name      date manner_of_death armed age gender race
## 1  3      Tim Elliot 10/4/2022      shot      gun  53      M    A
## 2  4  Lewis Lee Lembke 10/4/2022      shot      gun  47      M    W
## 3  5 John Paul Quintero 10/3/2022 shot and Tasered  unarmed  23      M    H
## 4  8  Matthew Hoffman 10/2/2022      shot toy weapon  32      M    W
## 5  9  Michael Rodriguez 10/2/2022      shot  nail gun  39      M    H
## 6 11 Kenneth Joe Brown 10/2/2022      shot      gun  18      M    W
##   city state signs_of_mental_illness threat_level flee
## 1  Shelton  WA              TRUE      attack Not fleeing
## 2   Aloha  OR              FALSE      attack Not fleeing
## 3   Wichita KS              FALSE      other Not fleeing
## 4 San Francisco CA              TRUE      attack Not fleeing
## 5    Evans  CO              FALSE      attack Not fleeing
## 6   Guthrie OK              FALSE      attack Not fleeing
## body_camera longitude latitude is_geocoding_exact
## 1      FALSE   -123.1      47.2              TRUE
## 2      FALSE   -122.9      45.5              TRUE
```

```

## 3      FALSE      -97.3      37.7      TRUE
## 4      FALSE      -122.4     37.8      TRUE
## 5      FALSE      -104.7     40.4      TRUE
## 6      FALSE      -97.4      35.9      TRUE

## 'data.frame':    6574 obs. of  17 variables:
## $ id          : int  3 4 5 8 9 11 13 15 16 17 ...
## $ name         : chr  "Tim Elliot" "Lewis Lee Lembke" "John Paul Quintero" "Matthew Hoffma
## $ date         : chr  "10/4/2022" "10/4/2022" "10/3/2022" "10/2/2022" ...
## $ manner_of_death : chr  "shot" "shot" "shot and Tasered" "shot" ...
## $ armed        : chr  "gun" "gun" "unarmed" "toy weapon" ...
## $ age          : int  53 47 23 32 39 18 22 35 34 47 ...
## $ gender       : chr  "M" "M" "M" "M" ...
## $ race         : chr  "A" "W" "H" "W" ...
## $ city         : chr  "Shelton" "Aloha" "Wichita" "San Francisco" ...
## $ state        : chr  "WA" "OR" "KS" "CA" ...
## $ signs_of_mental_illness: logi  TRUE FALSE FALSE TRUE FALSE FALSE ...
## $ threat_level  : chr  "attack" "attack" "other" "attack" ...
## $ flee         : chr  "Not fleeing" "Not fleeing" "Not fleeing" "Not fleeing" ...
## $ body_camera   : logi  FALSE FALSE FALSE FALSE FALSE ...
## $ longitude     : num  -123.1 -122.9 -97.3 -122.4 -104.7 ...
## $ latitude      : num  47.2 45.5 37.7 37.8 40.4 ...
## $ is_geocoding_exact : logi  TRUE TRUE TRUE TRUE TRUE TRUE ...
## - attr(*, "na.action")= 'omit' Named int [1:1229] 128 770 810 820 933 941 966 991 1338 1353 ...
## ..- attr(*, "names")= chr [1:1229] "128" "770" "810" "820" ...

## [1] 37.2

## [1] 35

## [1] "Tim Elliot"
## [1] 1
## [1] "Lewis Lee Lembke"
## [1] 2
## [1] "John Paul Quintero"
## [1] 3
## [1] "Matthew Hoffman"
## [1] 4
## [1] "Michael Rodriguez"
## [1] 5
## [1] "Kenneth Joe Brown"
## [1] 6
## [1] "Kenneth Arnold Buck"
## [1] 7

## [1] 1810

## [1] 1226

## [1] 1080

## [1] 1890

```

```
## [1] 568
```

```
## [1] 6574
```

```
## [1] 6574
```

```
##
```

```
##      A      B      H      N      O      W  
## 853 102 1525 1045   89   45 2915
```

```
##      state      regions      stbcp      gen.p      smi.p  
## Length:6574      MW:1080      Min.      :0.000      Min.      :0.818      Min.      :0.000  
## Class :character      NE: 568      1st Qu.:0.101      1st Qu.:0.938      1st Qu.:0.200  
## Mode  :character      NW:1810      Median :0.133      Median :0.952      Median :0.219  
##      SE:1890      Mean  :0.144      Mean  :0.952      Mean  :0.223  
##      SW:1226      3rd Qu.:0.183      3rd Qu.:0.966      3rd Qu.:0.265  
##      Max.      :0.409      Max.      :1.000      Max.      :0.556  
##      flee.p      att.p      armed.p      MoD.p      age.avg  
## Min.      :0      Min.      :0.350      Min.      :0.778      Min.      :0.810      Min.      :33.1  
## 1st Qu.:0      1st Qu.:0.564      1st Qu.:0.918      1st Qu.:0.938      1st Qu.:35.7  
## Median :0      Median :0.644      Median :0.934      Median :0.948      Median :36.9  
## Mean  :0      Mean  :0.635      Mean  :0.937      Mean  :0.951      Mean  :37.2  
## 3rd Qu.:0      3rd Qu.:0.679      3rd Qu.:0.958      3rd Qu.:0.969      3rd Qu.:38.6  
## Max.      :0      Max.      :1.000      Max.      :1.000      Max.      :1.000      Max.      :44.4  
## Non_White_prop  
## Min.      :0.250  
## 1st Qu.:0.455  
## Median :0.563  
## Mean  :0.557  
## 3rd Qu.:0.635  
## Max.      :0.939
```

```
##      state      stbcp      gen.p      smi.p      flee.p  
## Length:6574      Min.      :0.000      Min.      :0.818      Min.      :0.000      Min.      :0  
## Class :character      1st Qu.:0.101      1st Qu.:0.938      1st Qu.:0.200      1st Qu.:0  
## Mode  :character      Median :0.133      Median :0.952      Median :0.219      Median :0  
##      Mean  :0.144      Mean  :0.952      Mean  :0.223      Mean  :0  
##      3rd Qu.:0.183      3rd Qu.:0.966      3rd Qu.:0.265      3rd Qu.:0  
##      Max.      :0.409      Max.      :1.000      Max.      :0.556      Max.      :0  
##      att.p      armed.p      MoD.p      age.avg      Non_White_prop  
## Min.      :0.350      Min.      :0.778      Min.      :0.810      Min.      :33.1      Min.      :0.250  
## 1st Qu.:0.564      1st Qu.:0.918      1st Qu.:0.938      1st Qu.:35.7      1st Qu.:0.455  
## Median :0.644      Median :0.934      Median :0.948      Median :36.9      Median :0.563  
## Mean  :0.635      Mean  :0.937      Mean  :0.951      Mean  :37.2      Mean  :0.557  
## 3rd Qu.:0.679      3rd Qu.:0.958      3rd Qu.:0.969      3rd Qu.:38.6      3rd Qu.:0.635  
## Max.      :1.000      Max.      :1.000      Max.      :1.000      Max.      :44.4      Max.      :0.939
```

```
## [1] 0.144
```

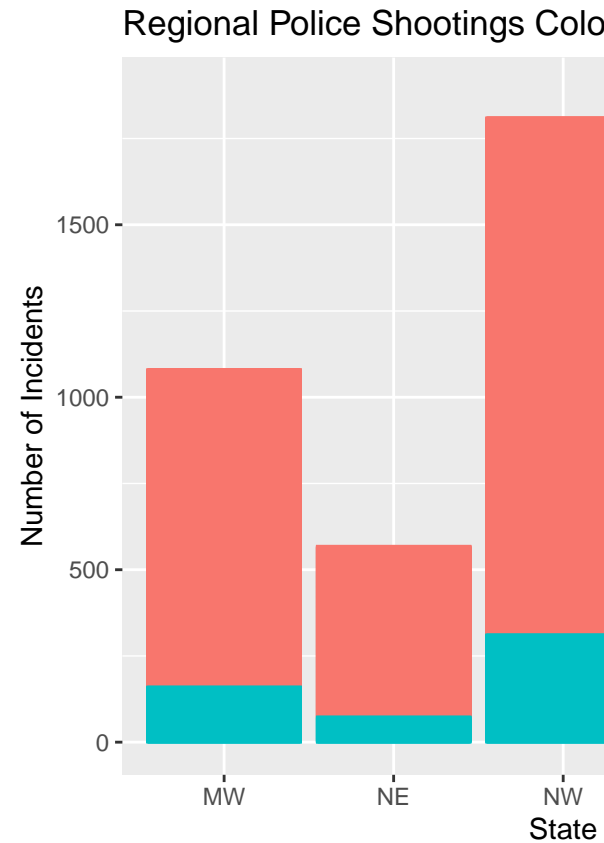
```
## # A tibble: 6 x 11
```

```
## # Groups:   state [6]
```

```
## state regions stbcp gen.p smi.p flee.p att.p armed.p MoD.p age.avg Non_Whit~1
```

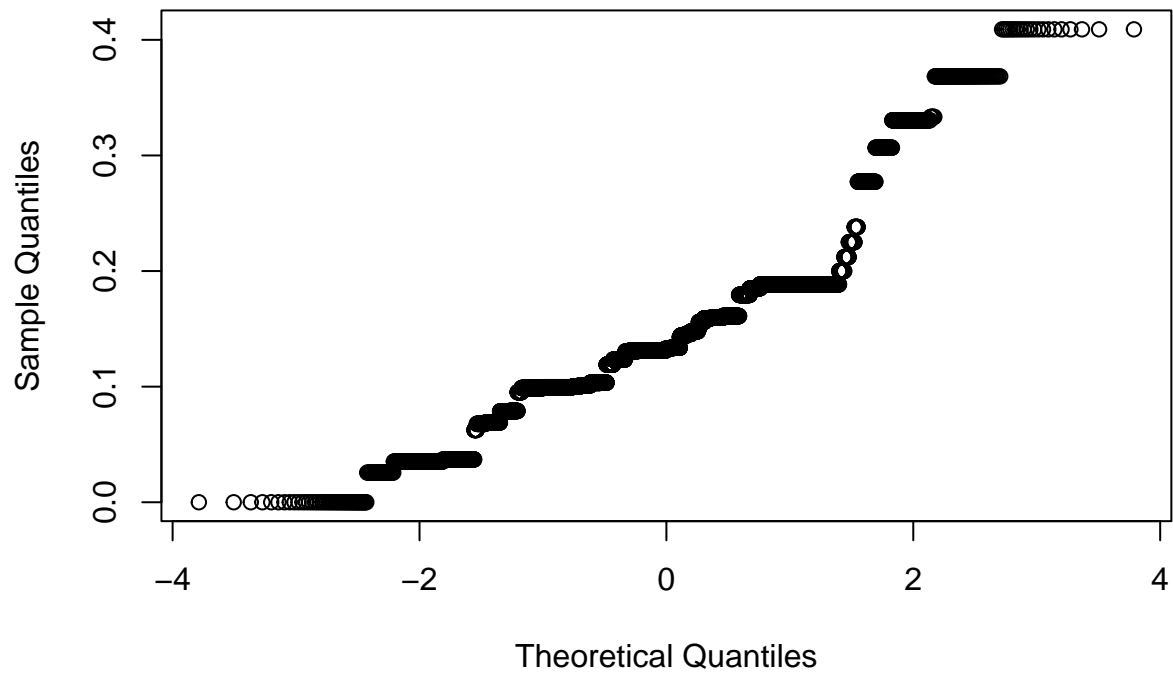
```
##   <chr> <fct>      <dbl> <dbl> <dbl>  <dbl> <dbl>    <dbl> <dbl>    <dbl>    <dbl>
## 1 WA    NW        0.103 0.967 0.337    0 0.549    0.935 0.946    36.9     0.576
## 2 OR    NW        0.0690 0.974 0.302    0 0.517    0.957 0.957    39.2     0.328
## 3 KS    MW        0.130 0.913 0.217    0 0.696    0.928 0.942    36.7     0.406
## 4 CA    NW        0.188 0.952 0.219    0 0.564    0.918 0.938    35.5     0.736
## 5 CO    NW        0.123 0.952 0.137    0 0.634    0.952 0.969    35.7     0.507
## 6 OK    SW        0.179 0.978 0.212    0 0.707    0.908 0.924    37.5     0.413
## # ... with abbreviated variable name 1: Non_White_prop
```

Explanation:

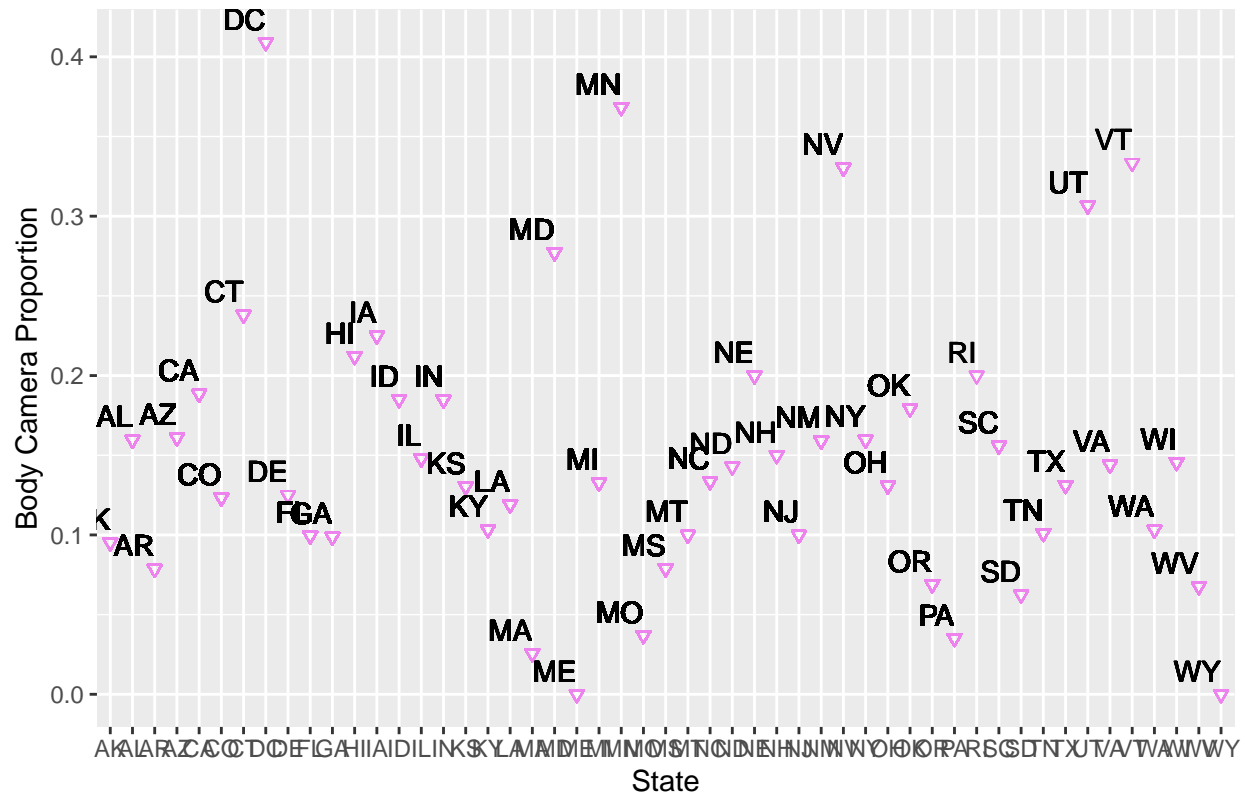


I made two new datasets. One grouped by State, one grouped by Region.4

Normal Q-Q Plot



Body Camera On Proportion By US State



```
##      state      regions      stbcp      gen.p      smi.p
## Length:6574      MW:1080      Min.   :0.000      Min.   :0.818      Min.   :0.000
## Class :character      NE: 568      1st Qu.:0.101      1st Qu.:0.938      1st Qu.:0.200
## Mode  :character      NW:1810      Median :0.133      Median :0.952      Median :0.219
##                                     SE:1890      Mean   :0.144      Mean   :0.952      Mean   :0.223
##                                     SW:1226      3rd Qu.:0.183      3rd Qu.:0.966      3rd Qu.:0.265
##                                     Max.   :0.409      Max.   :1.000      Max.   :0.556
##      flee.p      att.p      armed.p      MoD.p      age.avg
## Min.   :0      Min.   :0.350      Min.   :0.778      Min.   :0.810      Min.   :33.1
## 1st Qu.:0      1st Qu.:0.564      1st Qu.:0.918      1st Qu.:0.938      1st Qu.:35.7
## Median :0      Median :0.644      Median :0.934      Median :0.948      Median :36.9
## Mean   :0      Mean   :0.635      Mean   :0.937      Mean   :0.951      Mean   :37.2
## 3rd Qu.:0      3rd Qu.:0.679      3rd Qu.:0.958      3rd Qu.:0.969      3rd Qu.:38.6
## Max.   :0      Max.   :1.000      Max.   :1.000      Max.   :1.000      Max.   :44.4
## Non_White_prop
## Min.   :0.250
## 1st Qu.:0.455
## Median :0.563
## Mean   :0.557
## 3rd Qu.:0.635
## Max.   :0.939

##      Df Sum Sq Mean Sq F value Pr(>F)
## age.avg      1   1.48    1.48  918.29 < 2e-16 ***
## Non_White_prop      1   0.41    0.41  252.63 < 2e-16 ***
```

```
## gen.p          1  0.58    0.58  362.43 < 2e-16 ***
## smi.p          1  0.71    0.71  442.48 < 2e-16 ***
## MoD.p          1  0.00    0.00    0.16 0.69246
## att.p          1  0.00    0.00    0.00 0.97645
## armed.p        1  0.13    0.13   78.44 < 2e-16 ***
## age.avg:Non_White_prop 1  0.31    0.31  194.58 < 2e-16 ***
## age.avg:gen.p  1  0.02    0.02   13.01 0.00031 ***
## Non_White_prop:gen.p 1  2.29    2.29 1419.39 < 2e-16 ***
## smi.p:MoD.p    1  0.08    0.08   49.23 2.5e-12 ***
## smi.p:att.p    1  0.01    0.01    4.96 0.02596 *
## MoD.p:att.p    1  0.16    0.16   97.20 < 2e-16 ***
## smi.p:armed.p  1  0.05    0.05   31.64 1.9e-08 ***
## MoD.p:armed.p  1  0.00    0.00    0.00 0.99539
## att.p:armed.p  1  4.96    4.96 3070.78 < 2e-16 ***
## age.avg:Non_White_prop:gen.p 1  3.78    3.78 2340.27 < 2e-16 ***
## smi.p:MoD.p:att.p 1  0.31    0.31  194.88 < 2e-16 ***
## smi.p:MoD.p:armed.p 1  0.08    0.08   50.99 1.0e-12 ***
## smi.p:att.p:armed.p 1  0.08    0.08   50.19 1.5e-12 ***
## MoD.p:att.p:armed.p 1  0.00    0.00    0.00 0.95418
## smi.p:MoD.p:att.p:armed.p 1  0.53    0.53  331.42 < 2e-16 ***
## Residuals      6551 10.57    0.00
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Warning: Unknown or uninitialised column: 'rgbcp'.
```

```
## Warning in mean.default(rfp$rgbcp): argument is not numeric or logical:
## returning NA
```

```
## [1] NA
```

```
## # A tibble: 6 x 11
## # Groups:   state [6]
##   state regions stbcp gen.p smi.p flee.p att.p armed.p MoD.p age.avg Non_Whit~1
##   <chr> <fct>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 WA     NW      0.103 0.967 0.337    0 0.549 0.935 0.946 36.9 0.576
## 2 OR     NW      0.0690 0.974 0.302    0 0.517 0.957 0.957 39.2 0.328
## 3 KS     MW      0.130 0.913 0.217    0 0.696 0.928 0.942 36.7 0.406
## 4 CA     NW      0.188 0.952 0.219    0 0.564 0.918 0.938 35.5 0.736
## 5 CO     NW      0.123 0.952 0.137    0 0.634 0.952 0.969 35.7 0.507
## 6 OK     SW      0.179 0.978 0.212    0 0.707 0.908 0.924 37.5 0.413
## # ... with abbreviated variable name 1: Non_White_prop
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.000  0.101   0.133   0.144  0.183   0.409
```

```
## Warning in chisq.test(contable): Chi-squared approximation may be incorrect
```

```
##
## Pearson's Chi-squared test
##
## data:  contable
## X-squared = 3e+05, df = 2300, p-value <2e-16
```