# Lab3_YZ_EDA

*Yulia Zamriy*

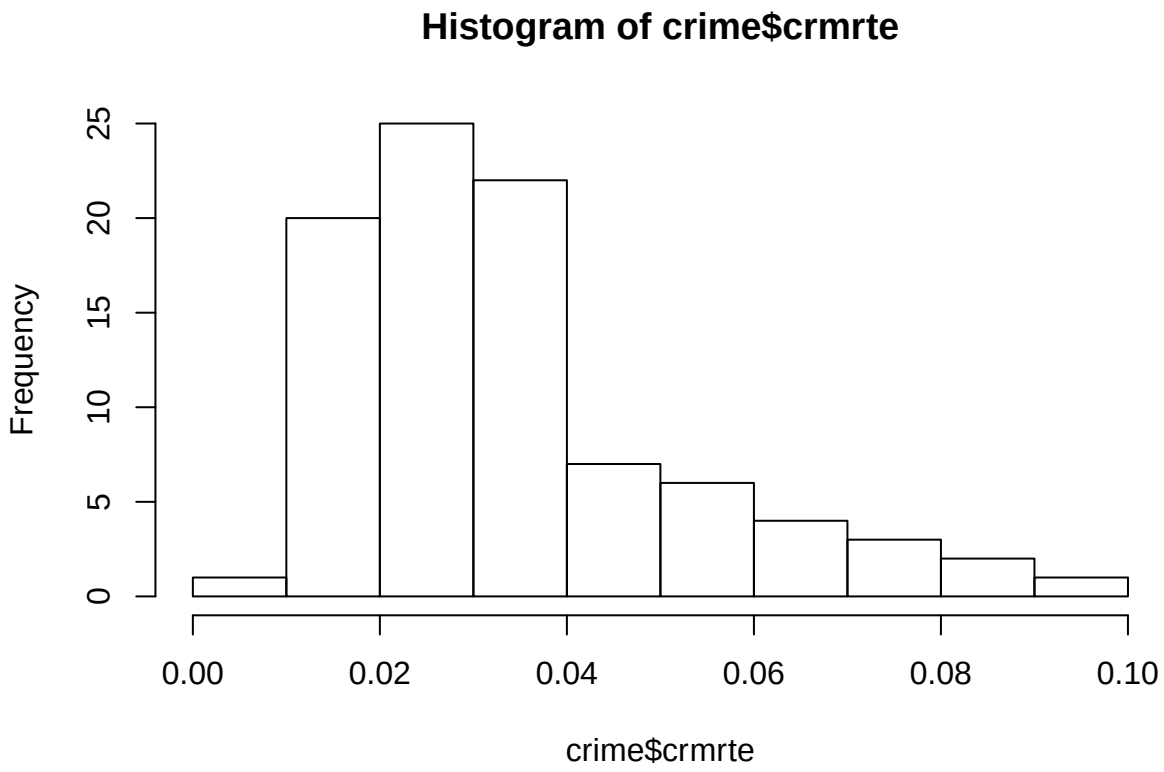
*March 18, 2018*

```
setwd("/home/yulia/Documents/MIDS/W203/Lab_3/")
crime <- read.csv("crime_v2.csv", stringsAsFactors = FALSE)
crime <- na.omit(crime)
```

```
summary(crime$crmrte)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.005533 0.020927 0.029986 0.033400 0.039642 0.098966
```

```
hist(crime$crmrte)
```




```
crime$prbconv <- as.numeric(crime$prbconv)
```

```
summary(crime$prbarr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.09277 0.20568 0.27095 0.29492 0.34438 1.09091
```

```
summary(crime$prbconv)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.06838 0.34541 0.45283 0.55128 0.58886 2.12121
```

```r
summary(crime$prbpris)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1500  0.3648  0.4234  0.4108  0.4568  0.6000
```

```r
nrow(crime[crime$prbarr >= 1,])
```

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

```r
nrow(crime[crime$prbconv >= 1,])
```

```
## [1] 10
```

```r
crime$exclude <- 0
crime[crime$prbarr > 1,]$exclude <- 1
crime[crime$prbconv > 1,]$exclude <- 1
table(crime$exclude)
```

```
##
##  0  1
## 81 10
```

```r
summary(crime$avgsen)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.380   7.340   9.100   9.647  11.420  20.700
```

```r
summary(crime$polpc)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0007459 0.0012308 0.0014853 0.0017022 0.0018768 0.0090543
```

```r
summary(crime$density)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00002 0.54741 0.96226 1.42884 1.56824 8.82765
```

```r
summary(crime$taxpc)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   25.69   30.66   34.87   38.06   40.95  119.76
```

```r
mean(crime$west)
```

```
## [1] 0.2527473
```

```r
mean(crime$central)
```

```
## [1] 0.3736264
```

```r
mean(crime$urban)
```

```
## [1] 0.08791209
```

```r
summary(crime$pctmin80)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.284   9.845  24.312  25.495  38.142  64.348
```

```r
summary(crime$wcon)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   193.6   250.8   281.4   285.4   314.8   436.8
```

```r
summary(crime$wtuc)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   187.6   374.6   406.5   411.7   443.4   613.2
```

```r
summary(crime$wtrd)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   154.2   190.9   203.0   211.6   225.1   354.7
```

```r
summary(crime$wfir)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   170.9   286.5   317.3   322.1   345.4   509.5
```

```r
summary(crime$wser)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   133.0   229.7   253.2   275.6   280.5  2177.1
```

```r
summary(crime$wmfg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   157.4   288.9   320.2   335.6   359.6   646.9
```

```r
summary(crime$wfed)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   326.1   400.2   449.8   442.9   478.0   598.0
```

```r
summary(crime$wsta)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   258.3   329.3   357.7   357.5   382.6   499.6
```

```r
summary(crime$wloc)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   239.2   297.3   308.1   312.7   329.2   388.1
```

```r
summary(crime$mix)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01961 0.08073 0.10186 0.12884 0.15175 0.46512
```

```r
summary(crime$pctymle)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.06216 0.07443 0.07771 0.08396 0.08350 0.24871
```

```r
crime[crime$wser > 2000,]$exclude <- 1
```

```r
crime_sub <- subset(crime, exclude == 0)
crime_sub$exclude <- NULL
crime_cor <- cor(crime_sub)[3,-c(1,2,3)]
```

```
## Warning in cor(crime_sub): the standard deviation is zero
```

```r
crime_cor <- crime_cor[order(crime_cor)]
```

```r
crime_cor_lab <- ifelse(crime_cor < 0, crime_cor-0.15, crime_cor)
```

```r
par(mar = c(2,8,1,0))
b <- barplot(crime_cor,
        horiz = TRUE,
        las = 1,
        xaxt = "n",
        xlim = c(-1,1),
        main = "Correlation of Crime Rate with Other Variables")
text(x = crime_cor_lab,
     y = b,
     label = round(crime_cor,2),
     pos = 4,
     cex = 0.6)
axis(1,
     at = seq(-1,1, by = 0.2),
     labels = seq(-1,1, by = 0.2),
     cex.axis = 0.6)
```