

## Chapter 12 Lab Answer Key

Last Update: November 2018

### Preparation

```
require(knitr)
library(ISLR)
library(ggplot2)
# install.packages('usmap')
library(usmap)
library(data.table)

opts_chunk$set(echo = TRUE)
options(digits = 6)
```

### US Arrests: Exploratory Analysis

```
head(USArrests)
```

```
##           Murder  Assault  UrbanPop  Rape
## Alabama      13.2     236      58 21.2
## Alaska       10.0     263      48 44.5
## Arizona       8.1     294      80 31.0
## Arkansas      8.8     190      50 19.5
## California    9.0     276      91 40.6
## Colorado      7.9     204      78 38.7
```

```
## Summary Stats
```

```
summary(USArrests)
```

```
##           Murder           Assault           UrbanPop           Rape
## Min.      : 0.80    Min.      : 45    Min.      :32.0    Min.      : 7.3
## 1st Qu.: 4.08    1st Qu.:109    1st Qu.:54.5    1st Qu.:15.1
## Median : 7.25    Median :159    Median :66.0    Median :20.1
## Mean     : 7.79    Mean     :171    Mean     :65.5    Mean     :21.2
## 3rd Qu.:11.25    3rd Qu.:249    3rd Qu.:77.8    3rd Qu.:26.2
## Max.     :17.40    Max.     :337    Max.     :91.0    Max.     :46.0
```

```
max(USArrests$Murder)
```

```
## [1] 17.4
```

```
max(USArrests$UrbanPop)
```

```
## [1] 91
```

```
max(USArrests$Rape)
```

```
## [1] 46
```

```

max(USArrests$Assault)

## [1] 337

## Dimensions
dim(USArrests)

## [1] 50  4

## Missingness
sapply(USArrests, function(x) sum(is.na(x)))

##      Murder      Assault UrbanPop      Rape
##           0           0           0           0

# Are variables related to each other
cor(USArrests)

##           Murder      Assault      UrbanPop      Rape
## Murder      1.0000000 0.801873 0.0695726 0.563579
## Assault      0.8018733 1.000000 0.2588717 0.665241
## UrbanPop     0.0695726 0.258872 1.0000000 0.411341
## Rape         0.5635788 0.665241 0.4113412 1.000000

```

Obviously the most recorded incident is the least serious one with the mean for assault at 170.8

Georgia (where I am from) is the most dangerous for murders 91% of people in California are in urban Nevada is most for rape North Carolina is most for assault

There are  $50 \times 4 = 200$  dimensions in the dataset

No missing data

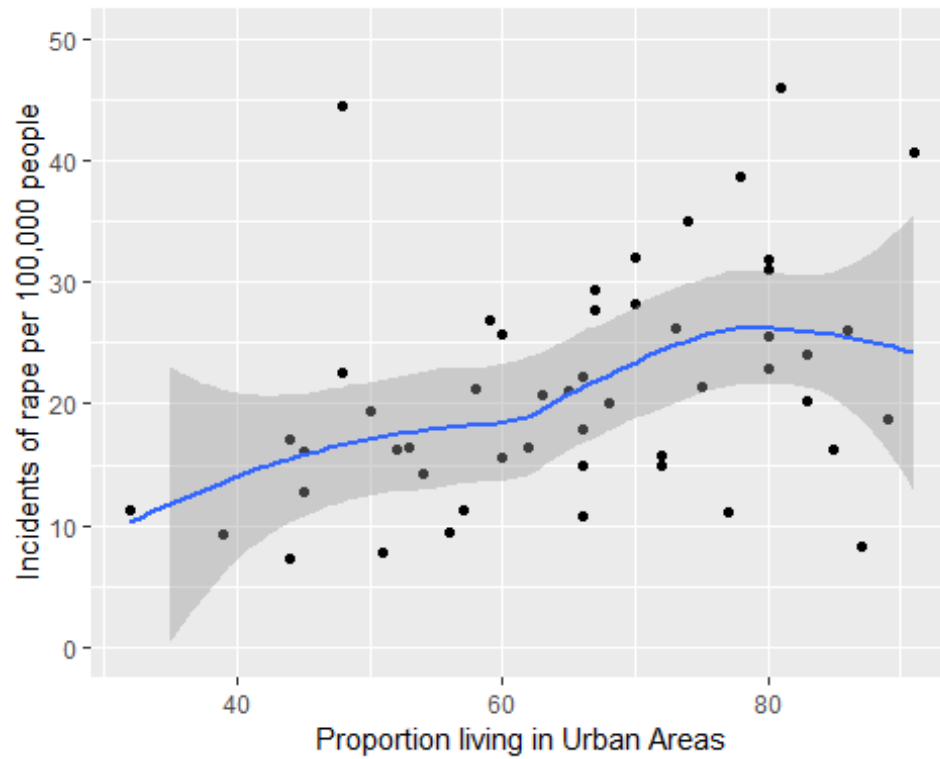
Not much of a correlation between urban and crime reporting as one would expect (Rape is the closest at 0.411)

If a place has a lot of one type of crime, it makes sense that they would have a lot of another type of crime

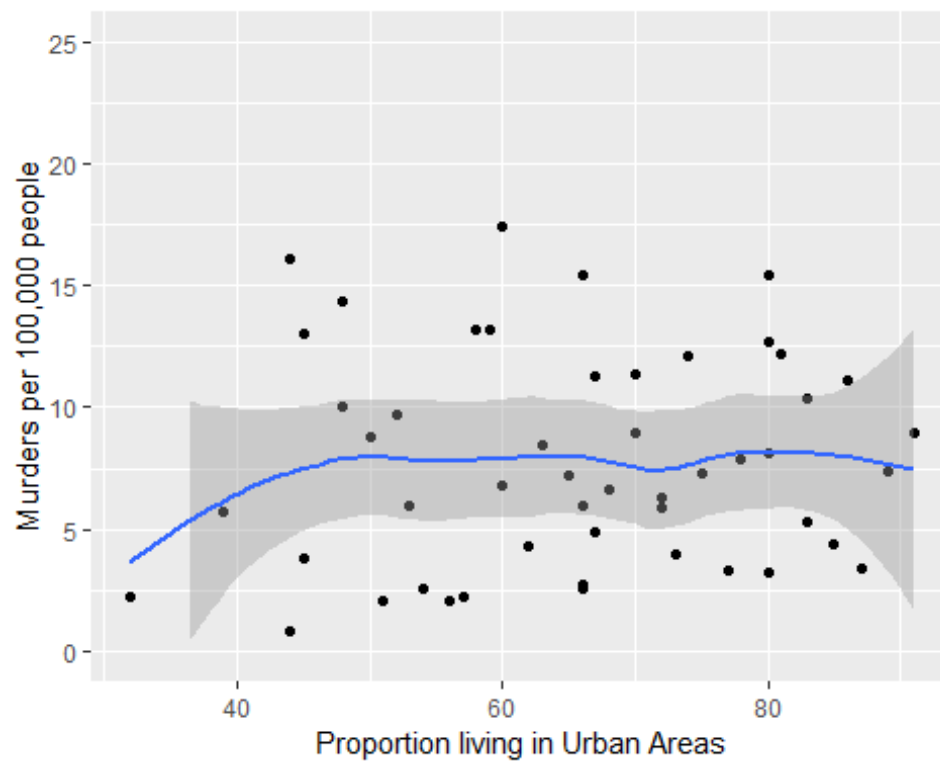
Clearly high between assault and murder and slightly high between rape and those two as well, but not as high as one might expect

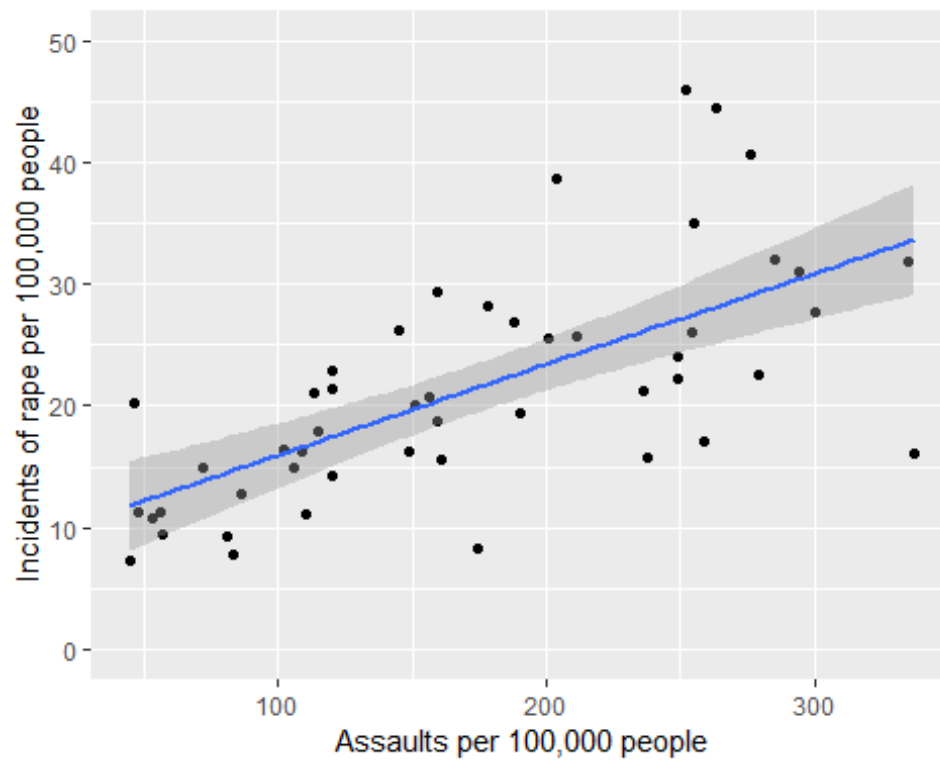
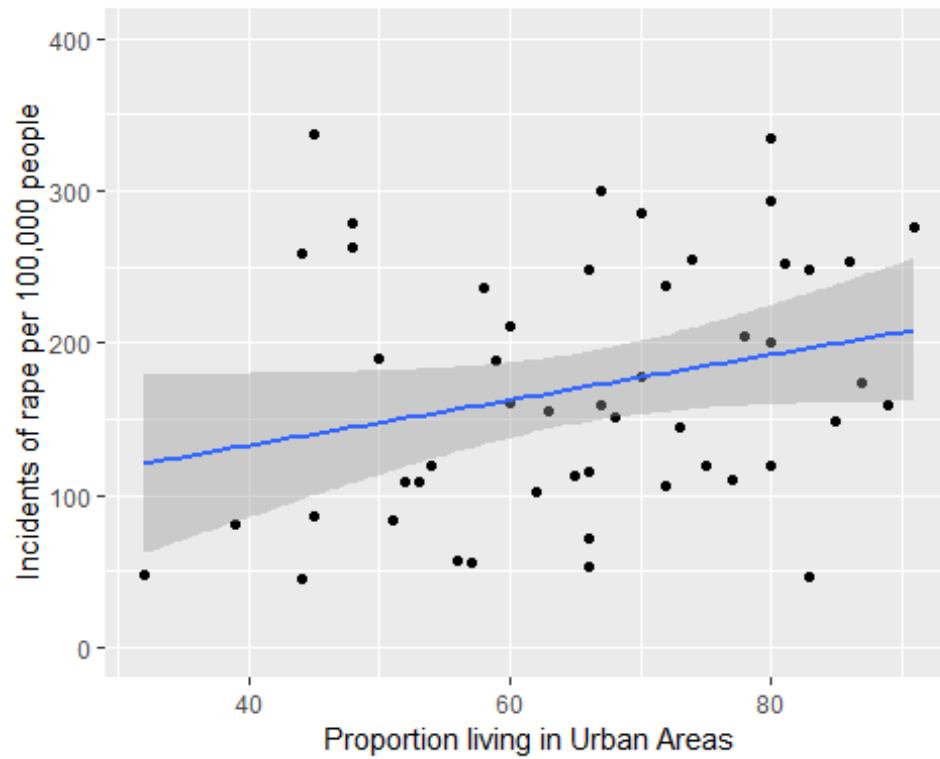
### Graph Analysis

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

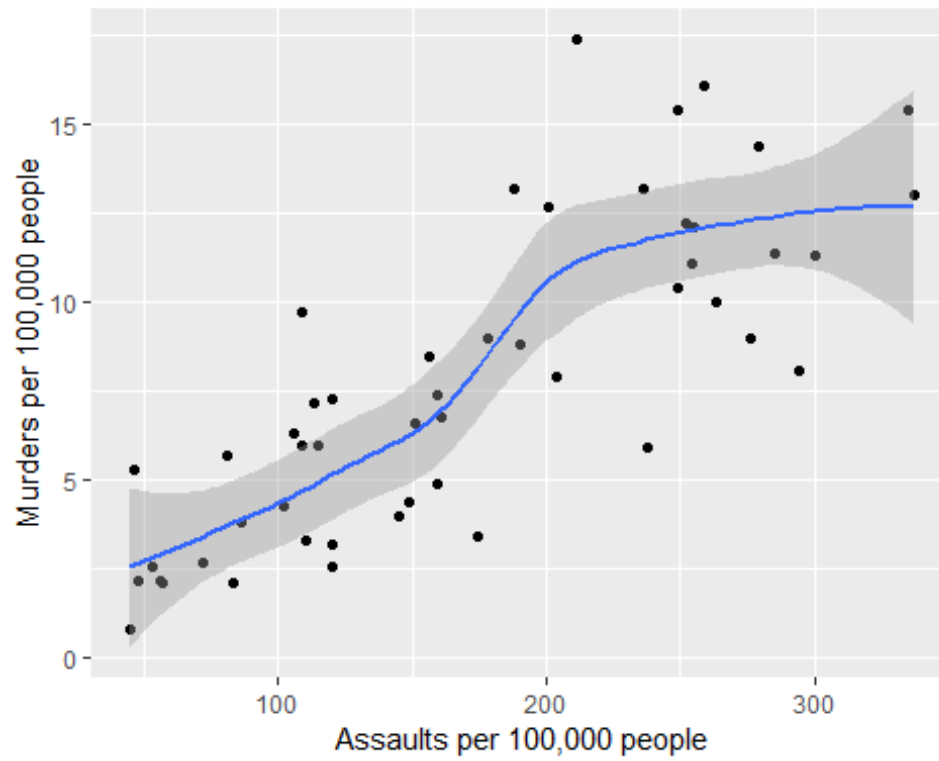


```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

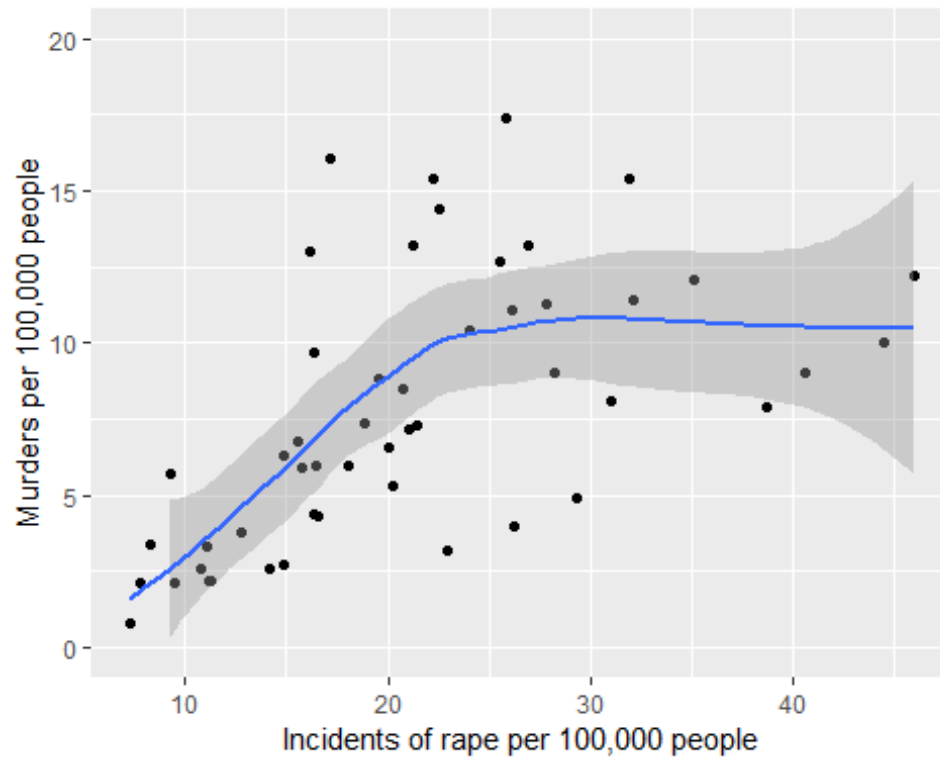




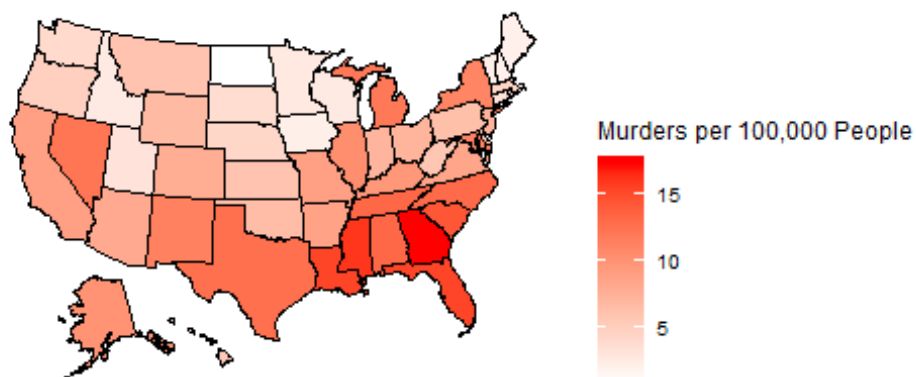
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



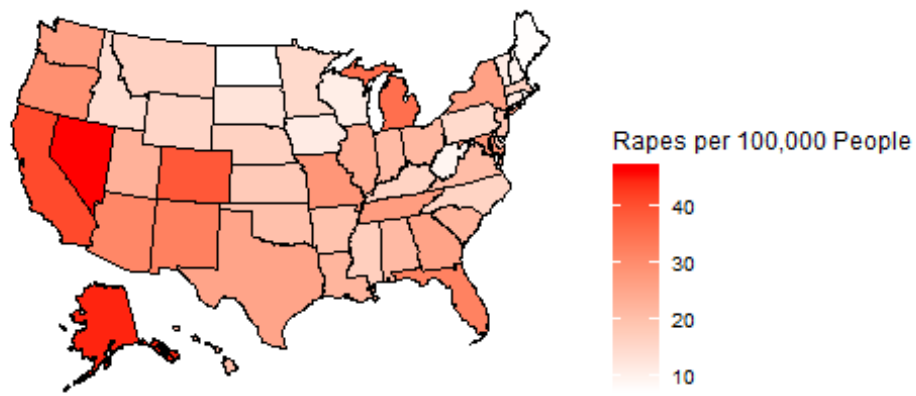
```
## <ScaleContinuousPosition>  
## Range:  
## Limits: 0 -- 1  
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



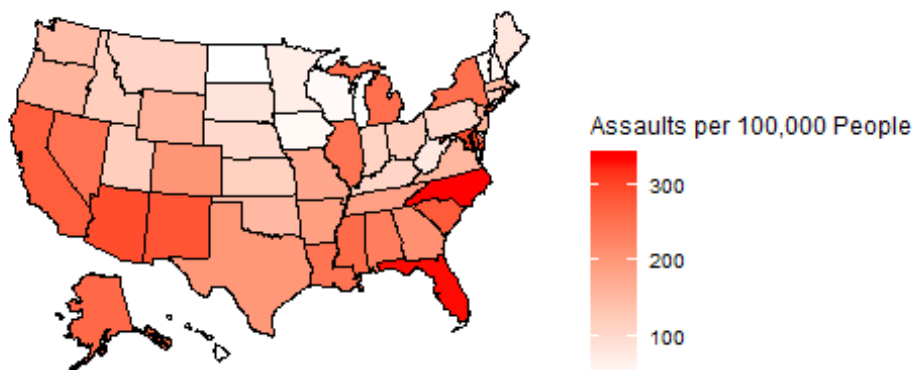
Murder Rates in the US by State



Rape Rates in the US by State



Assault Rates in the US by State



% Urban by US State

