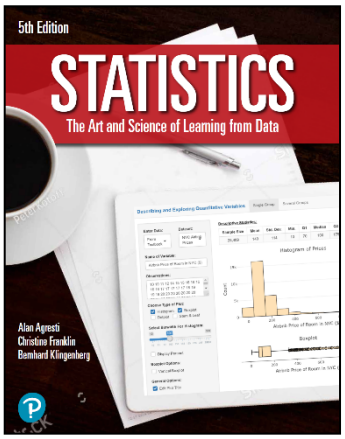


R Code for Examples in the book
"Statistics: The Art and Science of Learning from Data"
by Agresti, Franklin and Klingenberg, 5th edition

Chapter 3

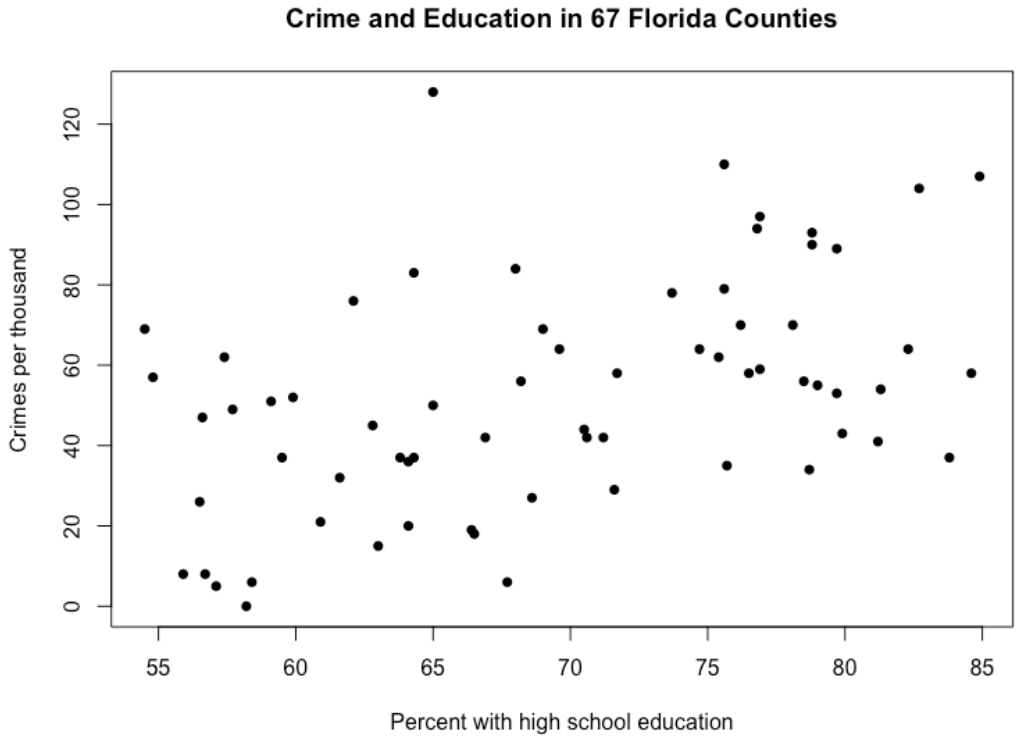
Example 15: Education and Murder – Correlation and Causation

Reading in the data

```
crime <-
read.csv(file='https://raw.githubusercontent.com/artofstat/data/master/Chapter3/fl_crime.csv')
attach(crime) # so we can refer to variable names
```

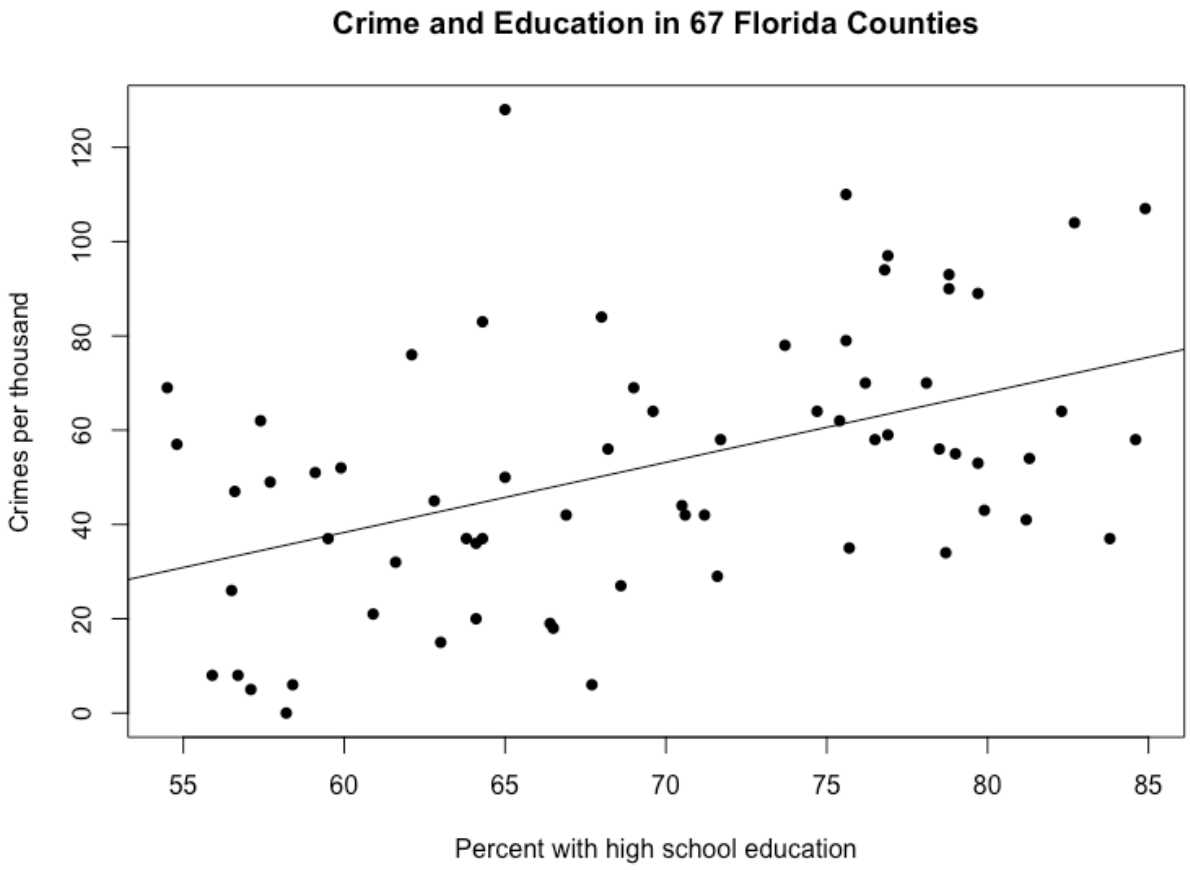
Basic scatterplot of crime rate and percentage with at least a high school education

```
plot(x = education...., y = crime.rate..per.1000., pch = 16,
     main = 'Crime and Education in 67 Florida Counties',
     xlab = 'Percent with high school education',
     ylab = 'Crimes per thousand')
```



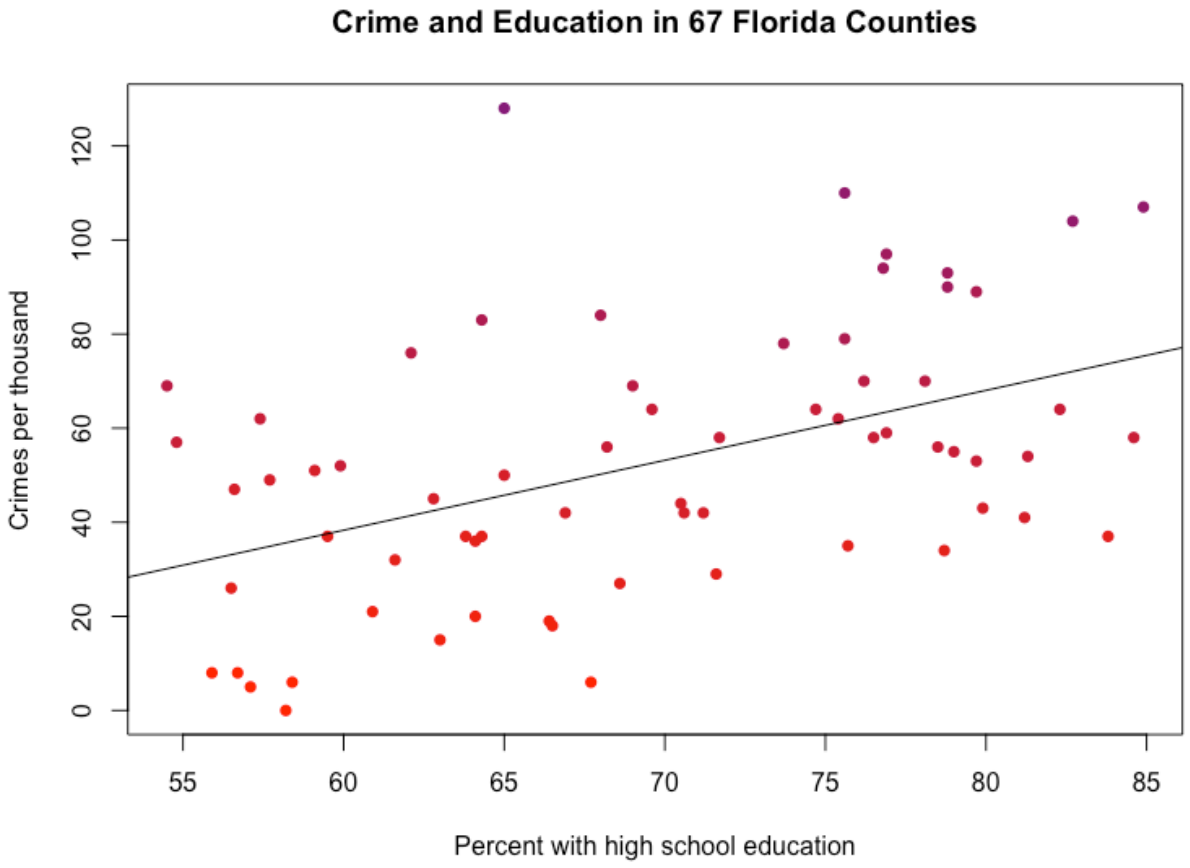
Fitting in regression model and adding to plot

```
linReg <- lm(crime.rate..per.1000. ~ education....)
plot(x = education...., y = crime.rate..per.1000., pch = 16,
     main = 'Crime and Education in 67 Florida Counties',
     xlab = 'Percent with high school education',
     ylab = 'Crimes per thousand')
abline(linReg)
```



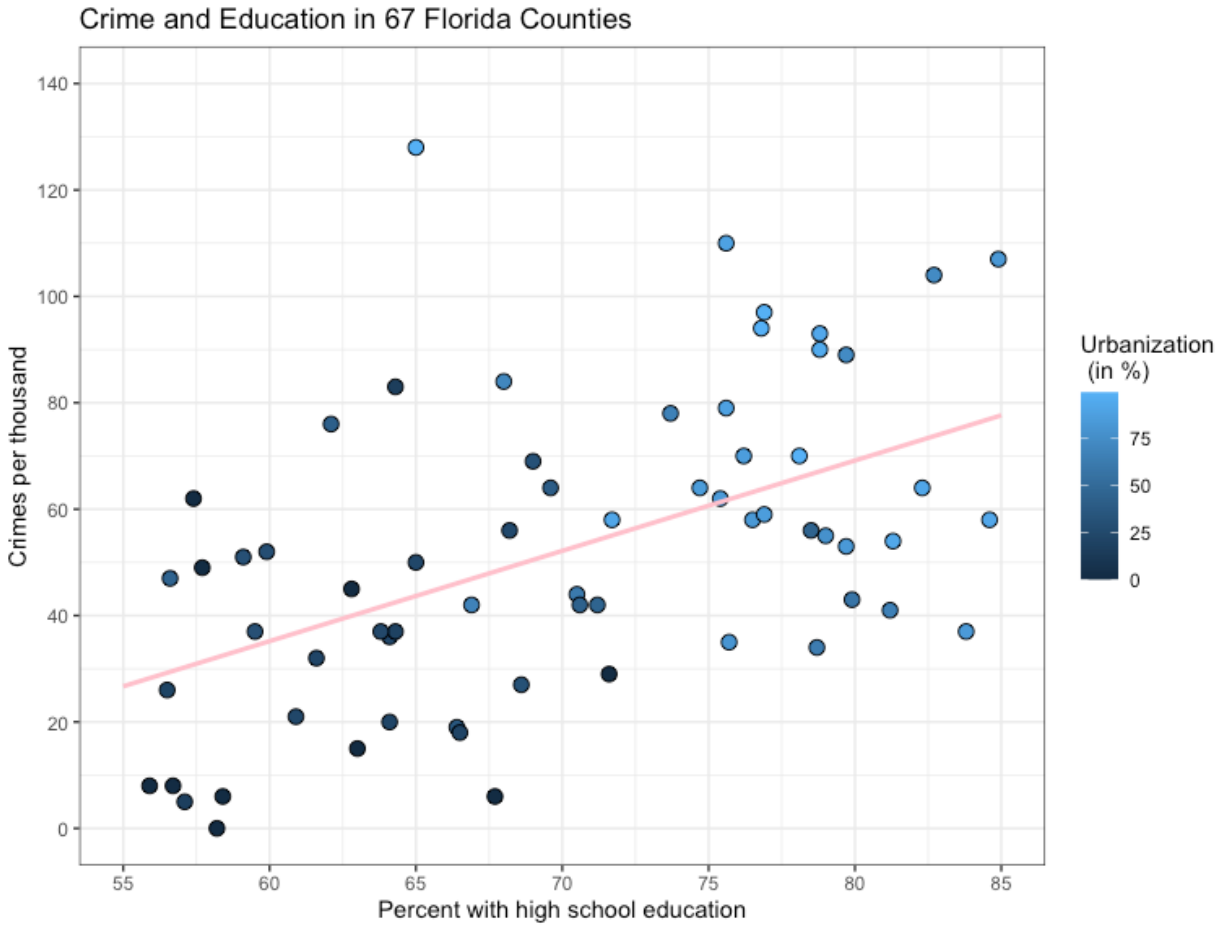
This adds a column of color values based on the y values

```
rbPal <- colorRampPalette(c('red','blue'))
crime$Col <- rbPal(20)[as.numeric(cut(crime$crime.rate..per.1000.,
                                      breaks = 10))]
plot(x = education...., y = crime.rate..per.1000., pch = 16, col = crime$Col,
     main = 'Crime and Education in 67 Florida Counties',
     xlab = 'Percent with high school education',
     ylab = 'Crimes per thousand')
abline(linReg)
```



Scatterplot of crime rate and percentage with at least a high school education with dots colored according to the percentage of urbanization of a county

```
library(ggplot2)
ggplot(crime, aes(x = education...., y = crime.rate..per.1000.)) +
  geom_point(aes(color = urbanization...., fill = urbanization....),
             pch = 21, colour = 'black', size = 3) +
  geom_smooth(method = lm, se=FALSE, fullrange= TRUE, col = 'pink') +
  labs(x = 'Percent with high school education',
       y = 'Crimes per thousand',
       title = 'Crime and Education in 67 Florida Counties',
       fill = 'Urbanization \n (in %)') +
  theme_bw() +
  scale_x_continuous(lim = c(55, 85), breaks = seq(55, 85, 5)) +
  scale_y_continuous(lim = c(0, 140), breaks = seq(0, 140, 20))
```

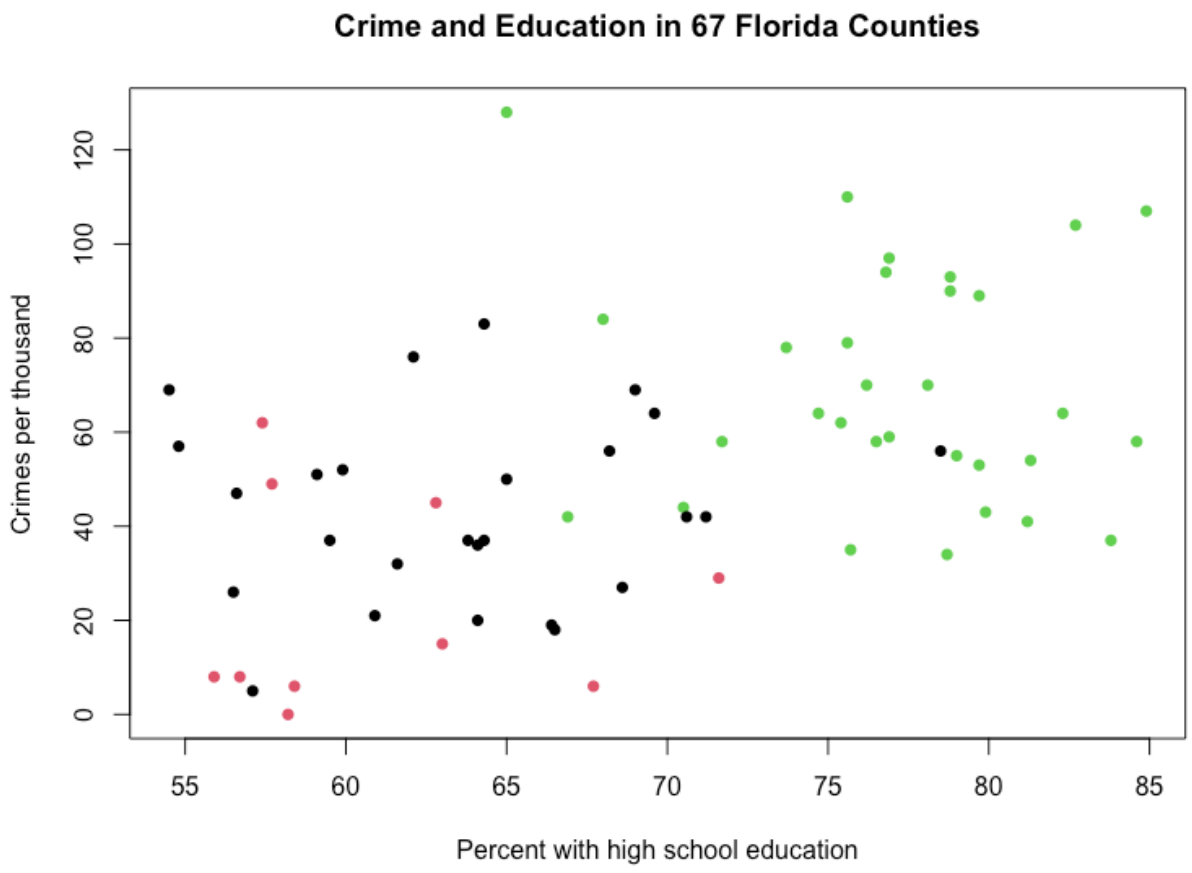


Adding Urbanization variable depending on urbanization percent using `mutate()` function from the dplyr package

```
library(dplyr)
crimeNew <- crime %>%
  mutate(Urbanization = case_when(urbanization.... <= 15 ~ 'rural',
                                  urbanization.... <= 50 ~ 'mixed',
                                  urbanization.... > 50 ~ 'urban'))
```

Basic scatterplot crime rate and percentage with at least a high school education with dots colored according to whether the county is rural, mixed, or urban

```
attach(crimeNew)
plot(x = education...., y = crime.rate..per.1000., pch = 16,
     col = factor(Urbanization),
     main = 'Crime and Education in 67 Florida Counties',
     xlab = 'Percent with high school education',
     ylab = 'Crimes per thousand')
```



Separating observations for rural, mixed, and urban counties

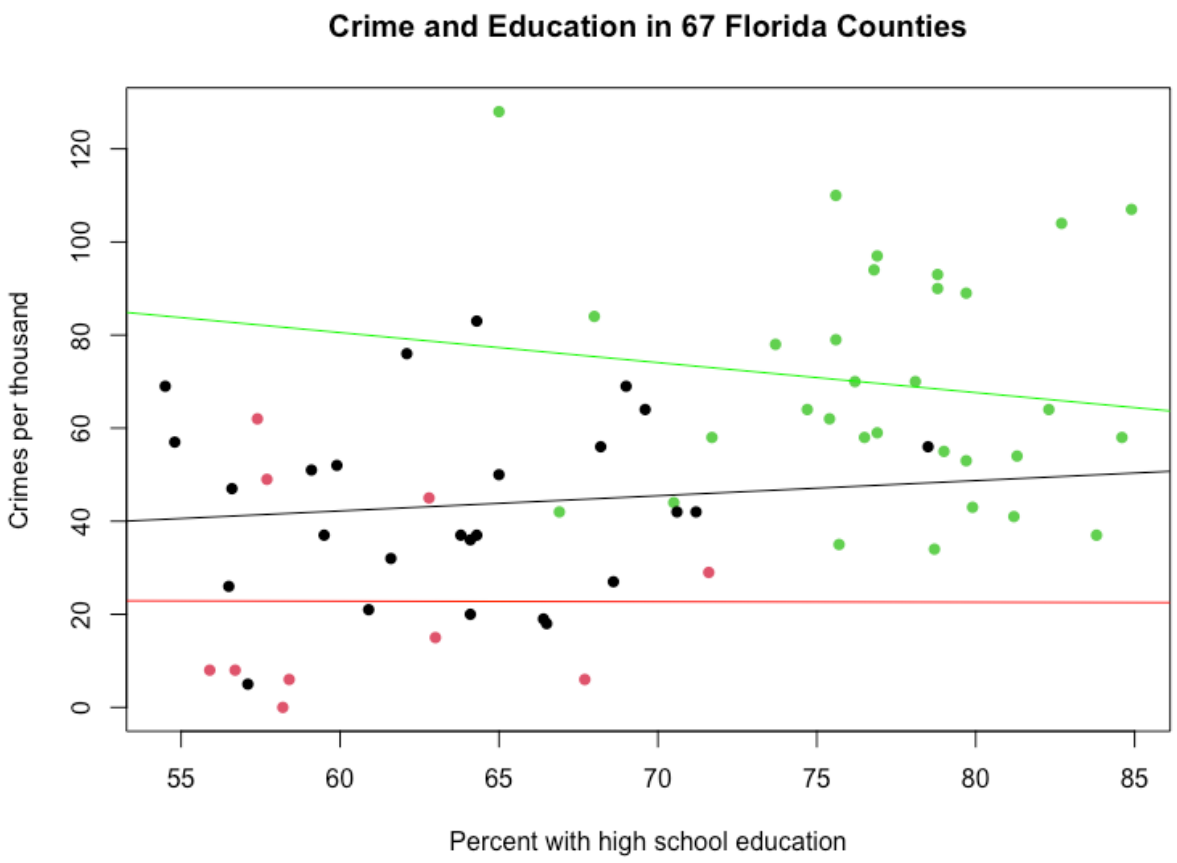
```
ruralObservations <- subset(crimeNew, Urbanization == 'rural')
mixedObservations <- subset(crimeNew, Urbanization == 'mixed')
urbanObservations <- subset(crimeNew, Urbanization == 'urban')
```

Fitting in corresponding regression models for rural, mixed, and urban counties

```
lmRural <- lm(crime.rate..per.1000. ~ education....,
              data = ruralObservations)
lmMixed <- lm(crime.rate..per.1000. ~ education....,
              data = mixedObservations)
lmUrban <- lm(crime.rate..per.1000. ~ education....,
              data = urbanObservations)
```

Adding the regression equations to the plot

```
plot(x = education...., y = crime.rate..per.1000., pch = 16,
     col = factor(Urbanization),
     main = 'Crime and Education in 67 Florida Counties',
     xlab = 'Percent with high school education',
     ylab = 'Crimes per thousand')
abline(lmRural, col = 'red')
abline(lmMixed, col = 'black')
abline(lmUrban, col = 'green')
```



Using the ggplot2 package to make the same scatterplot

```
ggplot(crime, aes(x = education...., y = crime.rate..per.1000.)) +
  geom_point(aes(shape = Urbanization, color = Urbanization,
                 fill = Urbanization), size = 3) +
  geom_smooth(method = lm, se = FALSE, fullrange = TRUE,
              aes(color=Urbanization)) +
  theme_bw() +
  labs(title = 'Crime and Education in 67 Florida Counties',
       x = 'Percent with high school education', y = 'Crimes per thousand')
```