MATH 3070 Lab Project 8

Pranav Rajan

October 21, 2020

Contents

Remember: I expect to see commentary either in the text, in the code with comments created using `#`, or (preferably) both! **Failing to do so may result in lost points!**

Problem 1 (Verzani problem 3.17)

The `state.x77` data set contains various information for each of the fifty United States. We wish to explore possible relationships among the variables. First, we make the data set easier to work with by turning it into a data frame.

```
x77 <- data.frame(state.x77)
```

Now, make scatter plots of `Population` and `Frost`; `Population` and `Murder`; `Population` and `Area`; and `Income` and `HS.Grad`. Do any relationships appear linear? Are there any surprising correlations? # Income versus HS.Grad seems linear

```
# Your code here
plot(x77$Population, x77$Frost, main = "Population vs. Frost", xlab = "Population", ylab = "Frost")
```
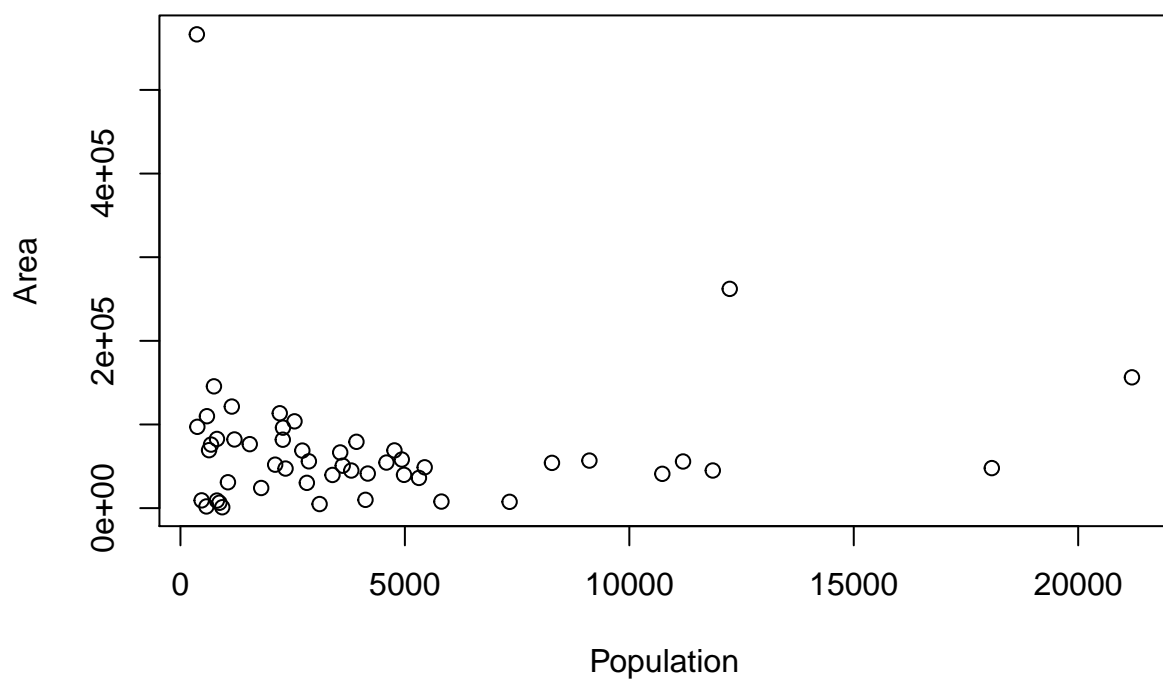
Population vs. Frost
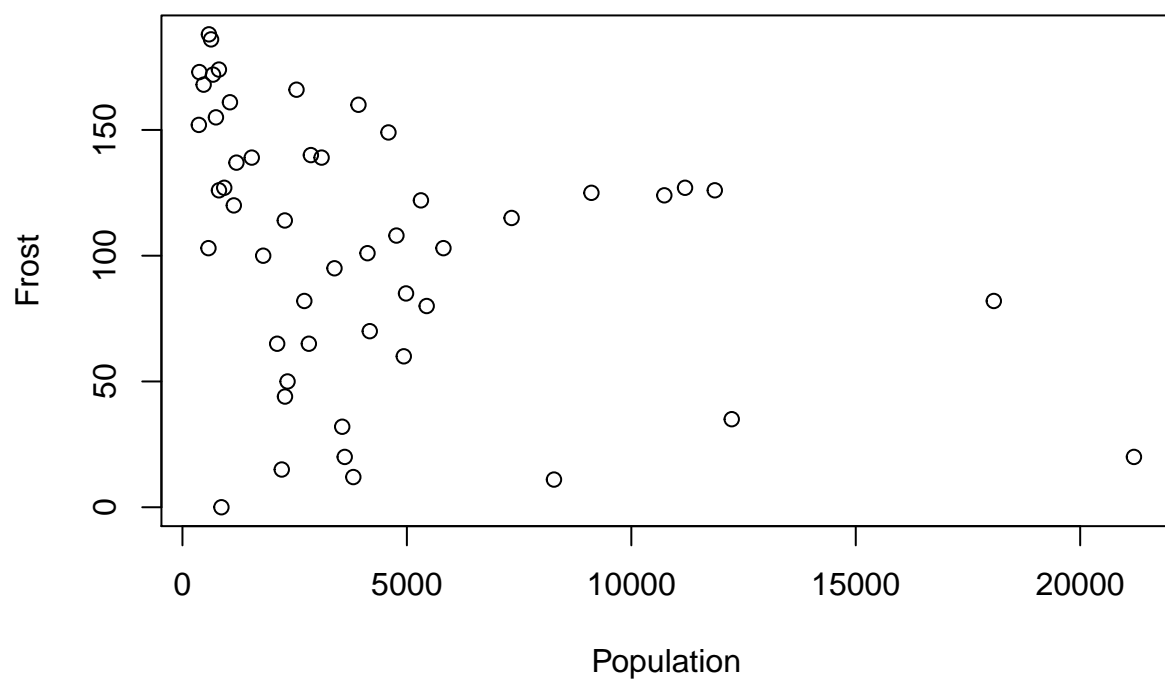

```
plot(x77$Population, x77$Murder, main = "Population vs. Murder", xlab = "Population", ylab = "Murder")
```
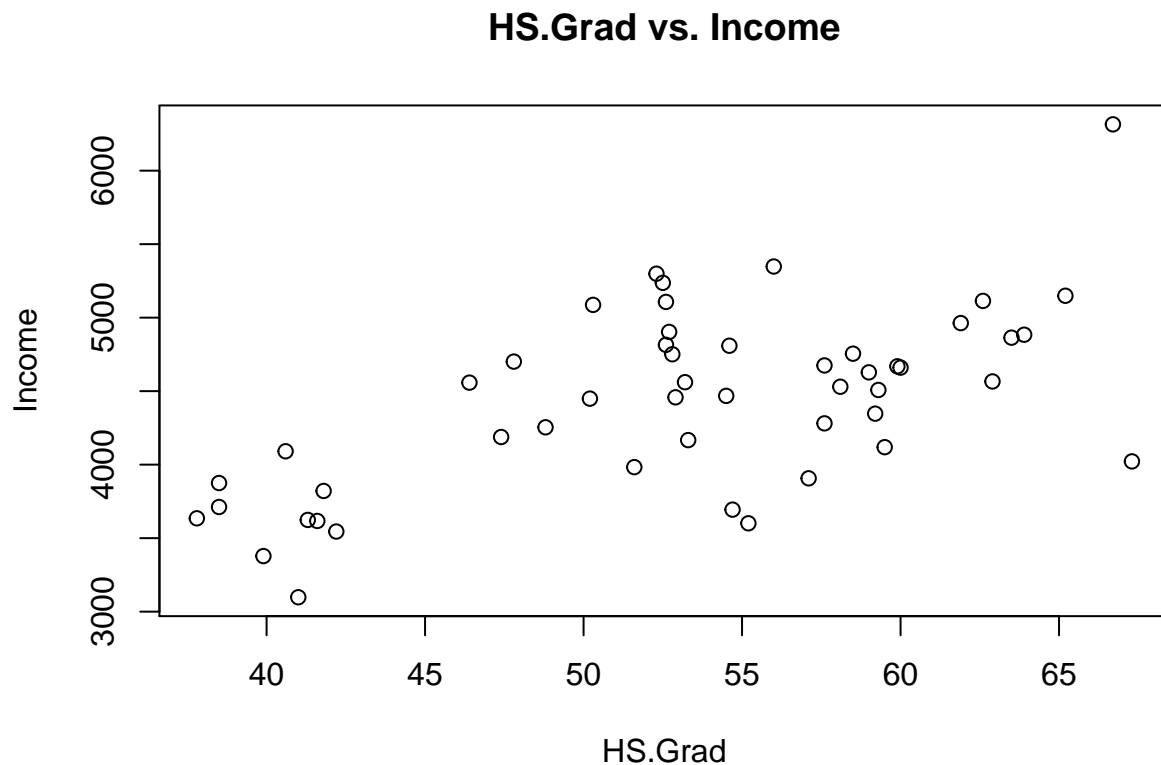
Population vs. Murder
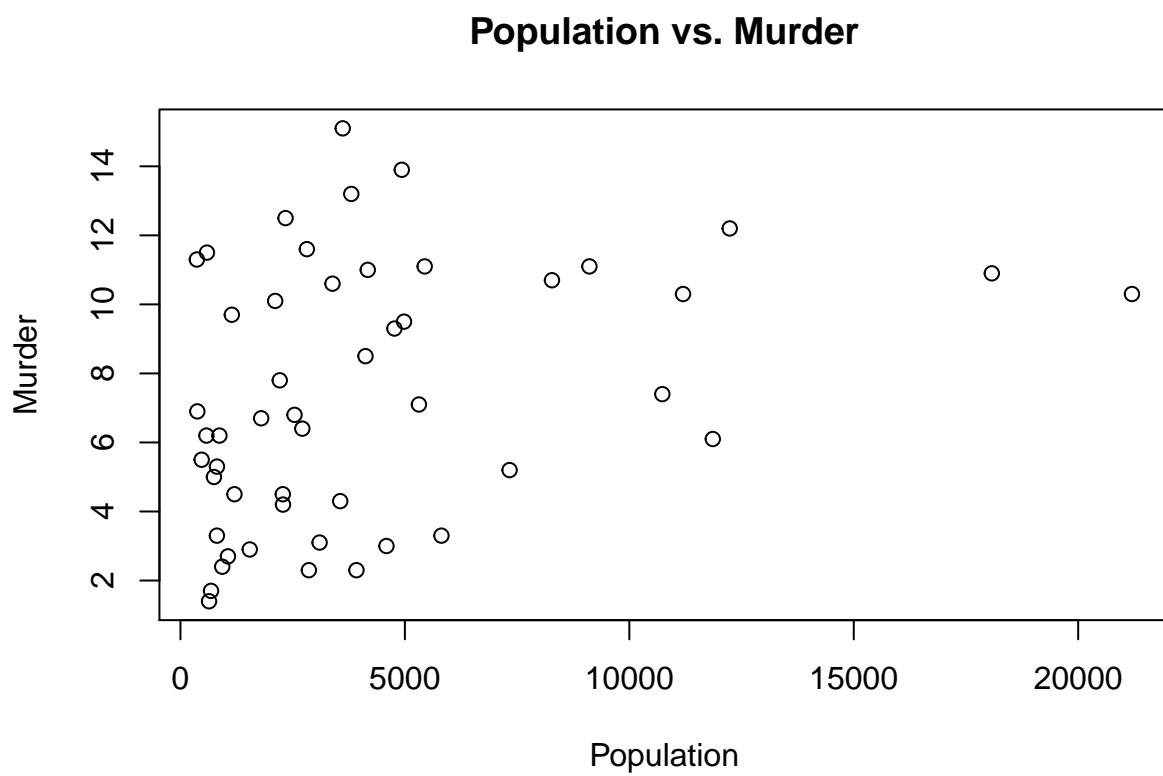

```
plot(x77$Population, x77$Area, main = "Population vs. Area", xlab = "Population", ylab = "Area")
```

Population vs. Area


```
plot(x77$HS.Grad, x77$Income, main = "HS.Grad vs. Income", xlab = "HS.Grad", ylab = "Income")
```

HS.Grad vs. Income

Problem 2 (Verzani problem 3.20)

*The `batting` (**UsingR**) data set contains baseball statistics for the 2002 Major League Baseball season. What is the correlation between the number of strikeouts (`SO`) and the number of home runs (`HR`)? Make a scatter plot to see whether there is any trend. Does the data suggest that in order to hit a lot of home runs one should strike out a lot?*

```
# Your code here
require(UsingR)
```

```
## Loading required package: UsingR
```

```
## Loading required package: MASS
```

```
## Loading required package: HistData
```

```
## Loading required package: Hmisc
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units

##
## Attaching package: 'UsingR'

## The following object is masked from 'package:survival':
##
##     cancer
```

```r
plot(batting$SO, batting$HR, main = "Home Runs by Strike Outs")
abline(v=mean(batting$SO), lty=2)
abline(h=mean(batting$HR), lty=2)
points(mean(batting$SO), mean(batting$HR), pch=16, cex=4, col=rgb(0, 0, 0, .25))
```

Home Runs by Strike Outs
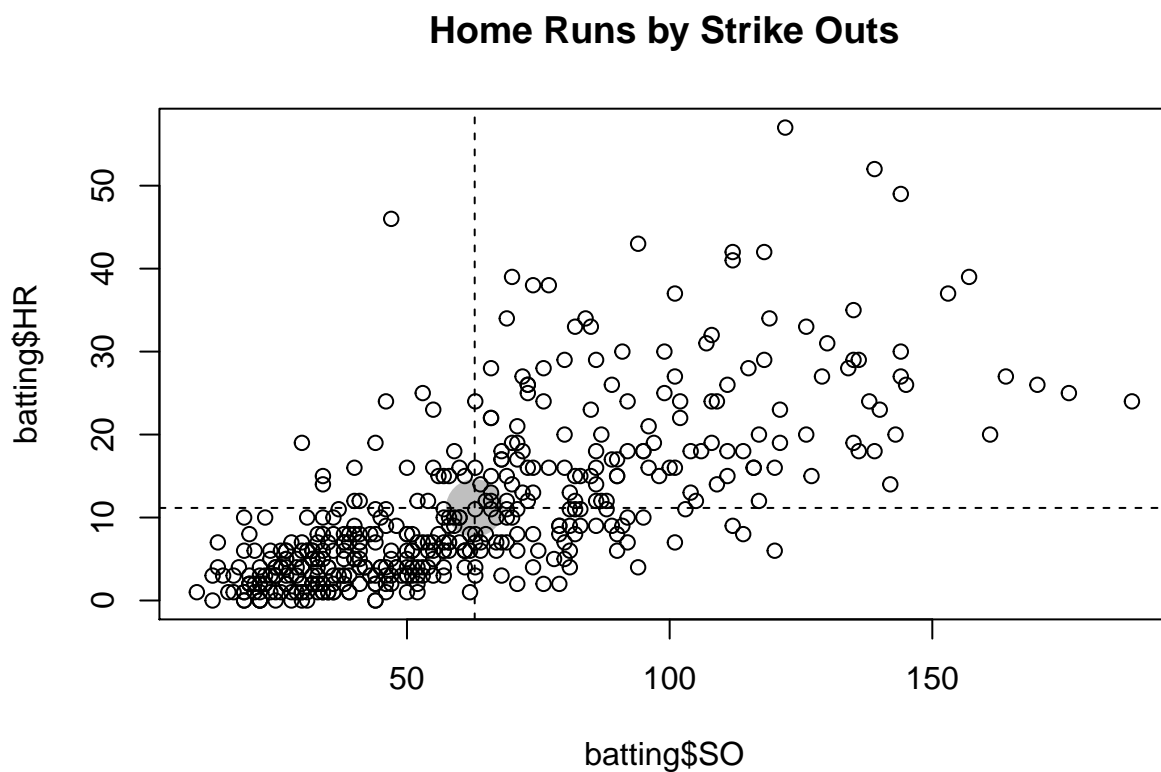

Problem 3 (Verzani problem 3.24)

*The data frame **x77** contains data from each of the fifty United States. First coerce the **state.x77** variable into a data frame with:*

```
x77 <- data.frame(state.x77)
```

*For the following, make a scatter plot with the regression line:*

1. *The model of illiteracy rate (Illiteracy) modeled by high school graduation rate (HS.Grad).*

```
# Your code here
res <- lm(x77$Illiteracy ~ x77$HS.Grad)
plot(x77$Illiteracy ~ x77$HS.Grad, main="Illiteracy vs HS.Grad", xlab="HS.Grad", ylab = "Illiteracy")
abline(res)
```
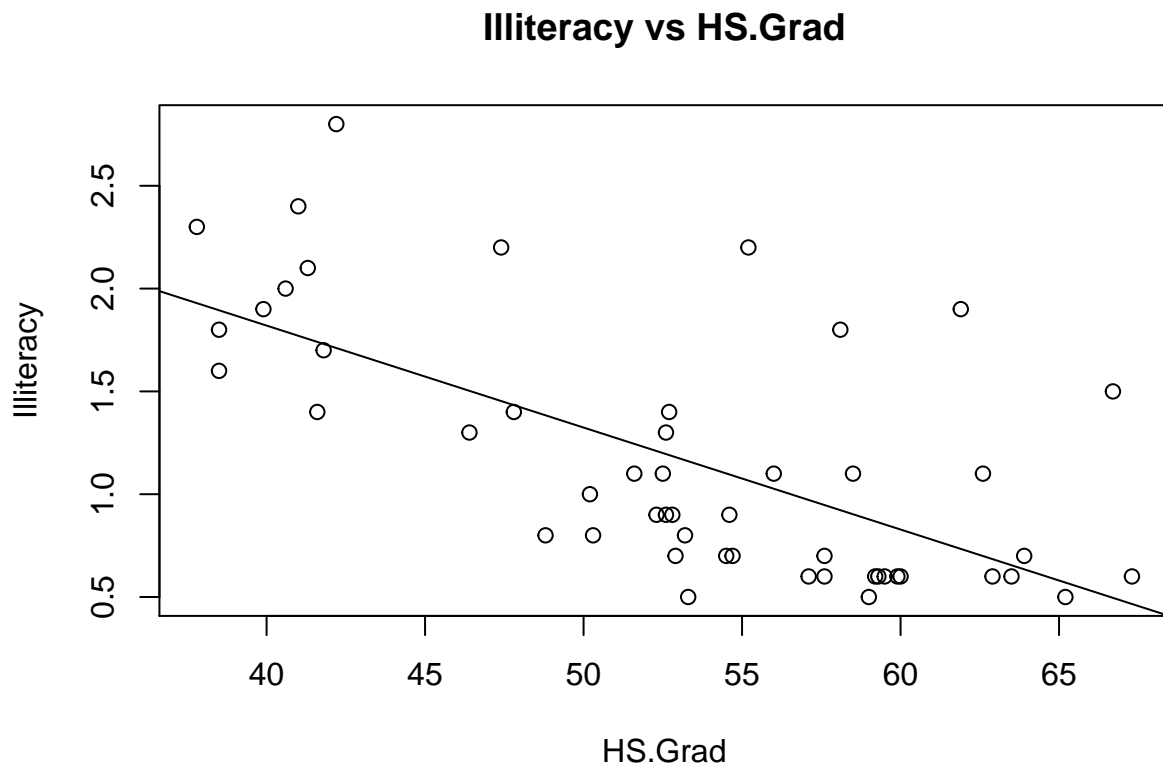

2. *The model of life expectancy (Life.Exp) modeled by the murder rate (Murder).*

```
# Your code here
res <- lm(x77$Life.Exp ~ x77$Murder)
plot(x77$Life.Exp ~ x77$Murder, main="Life Expectancy vs Murder", xlab="Murder", ylab = "Life Expectancy
abline(res)
```

Life Expectancy vs Murder
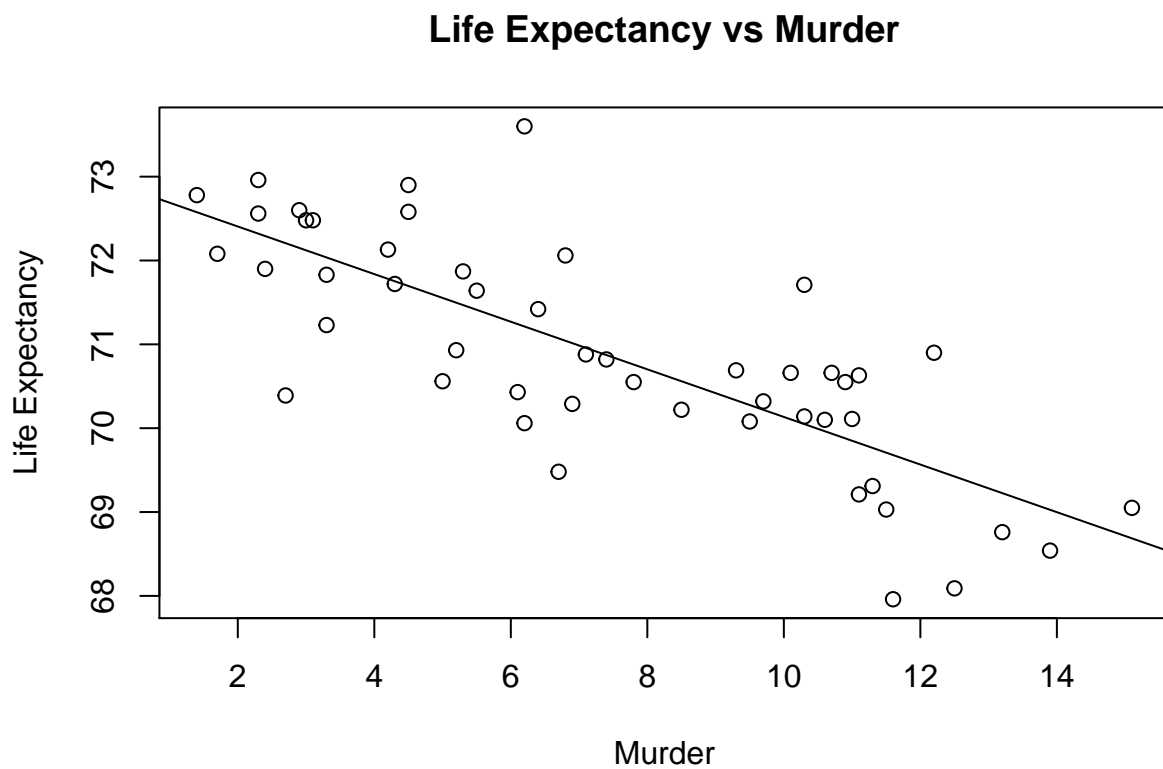
3. *The model of income (`Income`) modeled by the illiteracy rate (`Illiteracy`).*

```
# Your code here
res <- lm(x77$Income ~ x77$Illiteracy)
plot(x77$Income ~ x77$Illiteracy, main="Income vs Illiteracy", xlab="Illiteracy", ylab = "Income")
abline(res)
```

Income vs Illiteracy
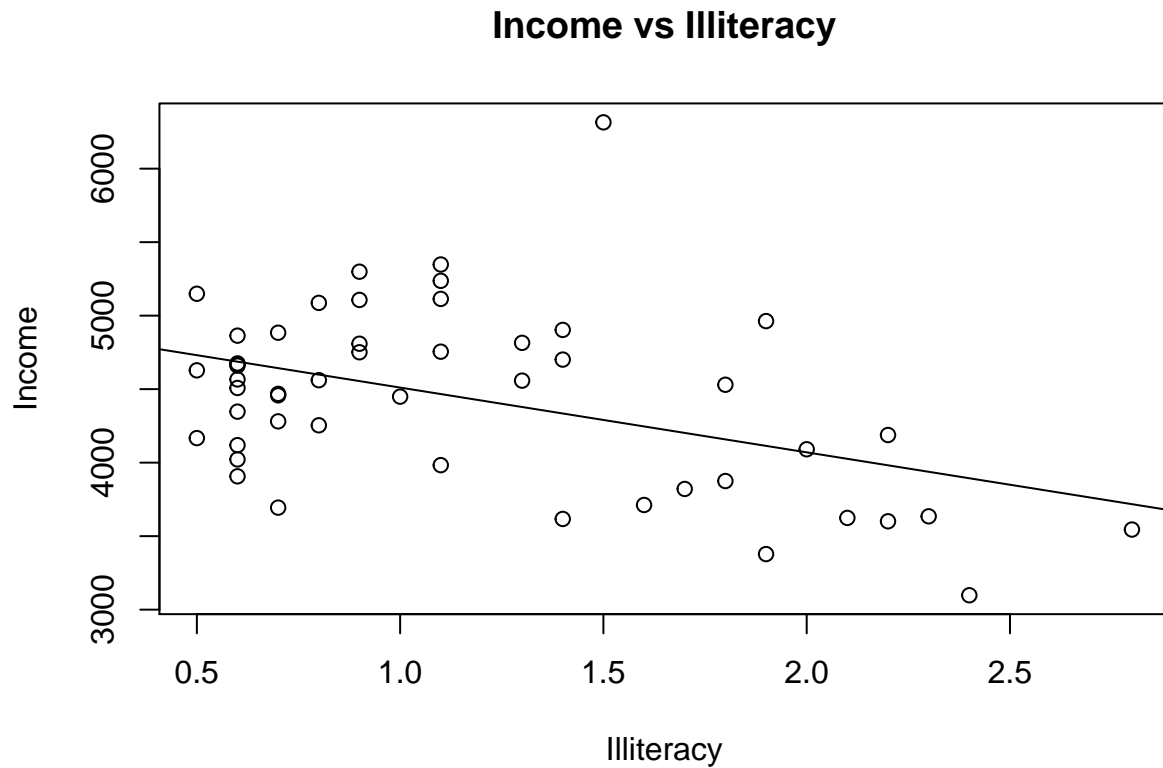
Write a sentence or two describing any relationship for each fo the relationships examined above. In particular, do you find it as expected or surprising? # The relationship is linear using least squares regression

Problem 4 (Verzani problem 3.32)

The data set **UScereal** *(**MASS**) contains information about cereals on a shelf of a United States grocery store. Make a table showing the relationship between manufacturer,* **mfr**, *and shelf placement,* **shelf**. *Are there any obvious differences between manufacturers?*

```
# Your code here
require(UsingR)
table(UScereal$mfr, UScereal$shelf)
```

```
##
##      1  2  3
##   G  6  7  9
##   K  4  7 10
##   N  2  0  1
##   P  2  1  6
##   Q  0  3  2
##   R  4  0  1
```

Problem 5 (Verzani problem 3.33)

The help page for `mosaicplot()` demonstrates the data set `HairEyeColor`, which records `sex`, `Hair` color, and `Eye` color for 592 statistics students. The data set comes as a flattened table, so simply passing the object to `mosaicplot()` will create the plot. (Or, as demonstrated, passing `shade = TRUE`, as in `mosaicplot(HairEyeColor, shade = TRUE)`, will produce a colored version.) Make the plot. Why does the help page note, "there are more blue-eyed, blonde females than expected?" # There are more blue-eyed blonde females than expected because of # the model used according to my understanding of the documentation

```
# Your code here
require(stats)
mosaicplot(HairEyeColor, shade = TRUE)
```