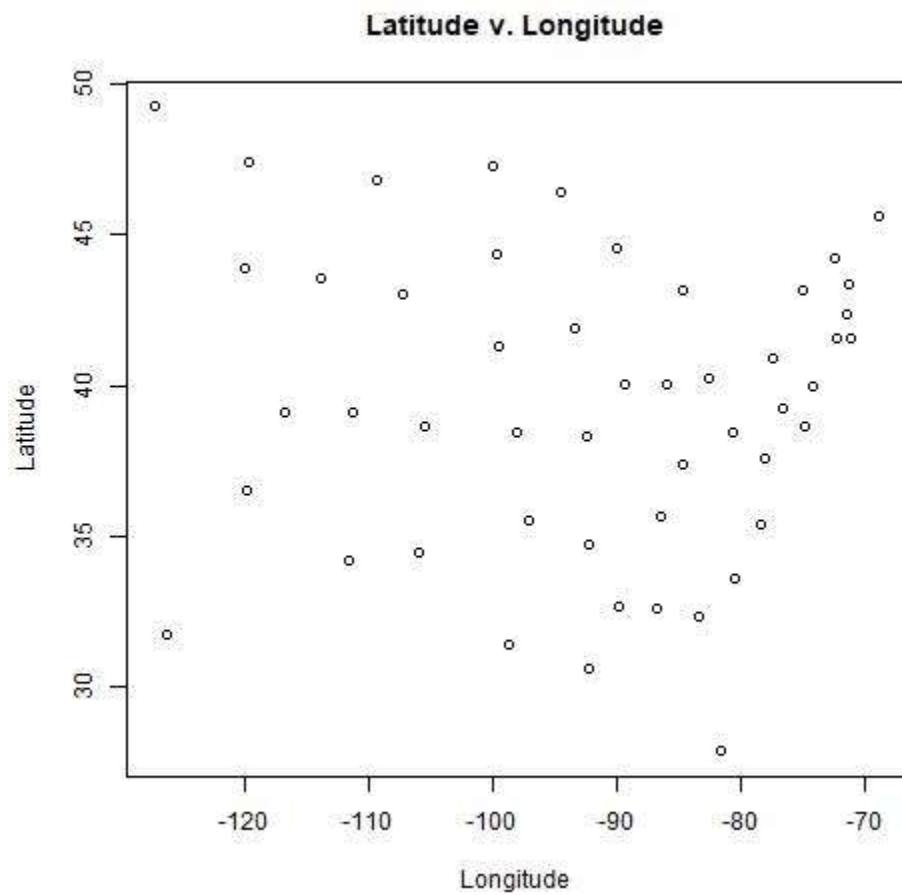Kyle Kolodziej

September 6th, 2021

EMIS 5357 – Analytics in Decision Support

Problem Set #1

A. Scatter plot shown below. The plot does appear to take the shape of an outline of the United States.

Latitude v. Longitude

i.

B. Latitude has to do with the location and, specifically, whether a location is located

farther north or south. It would make sense that the number of Frosts would have a

linear relationship with the Latitude variable. The plot is shown below of Frost and

Latitude. This plot does appear to have a linear relationship as predicted (albeit some

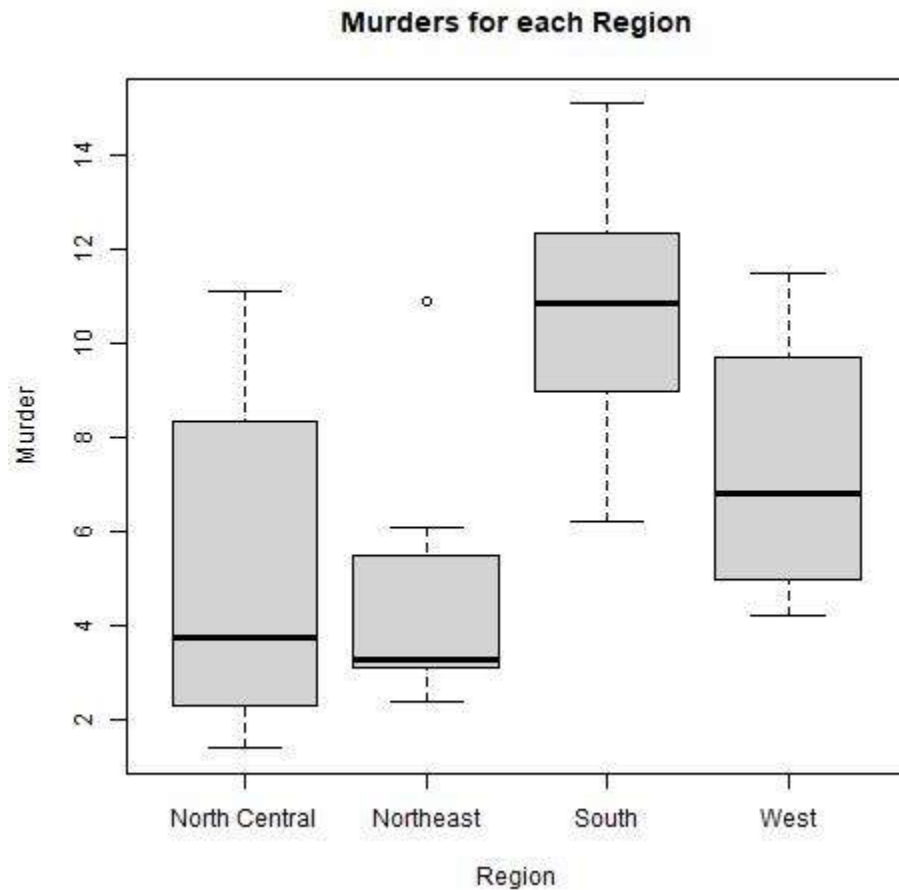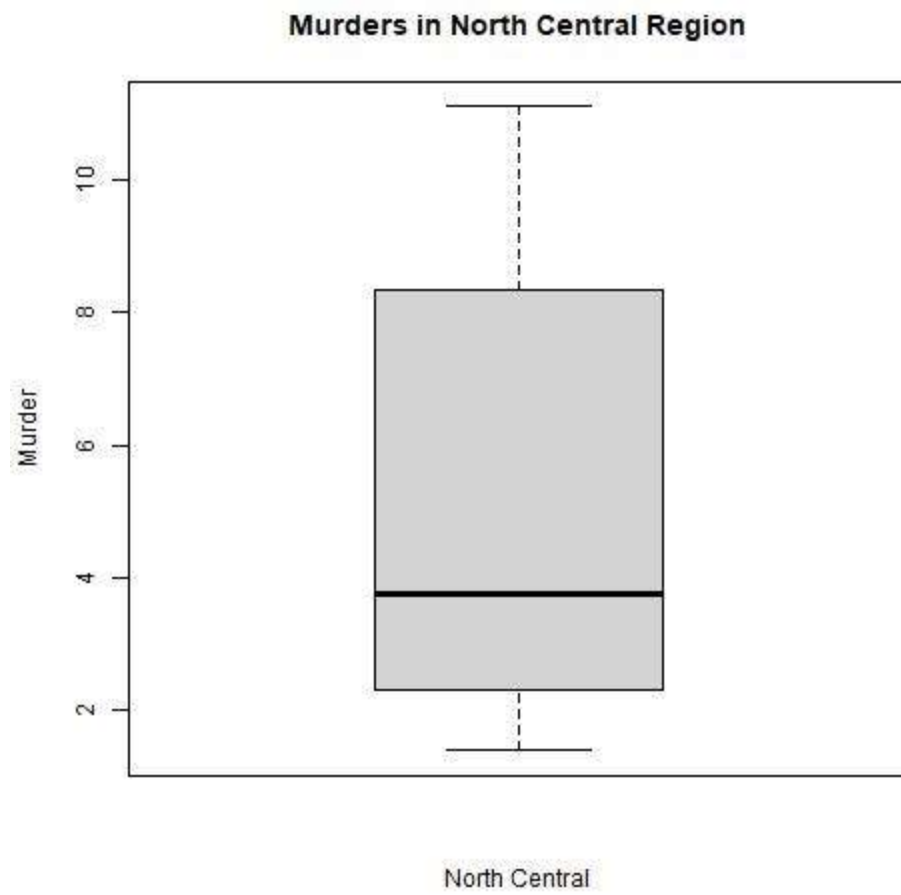outliers that do not follow this linear relationship).

Latitude v. Frost
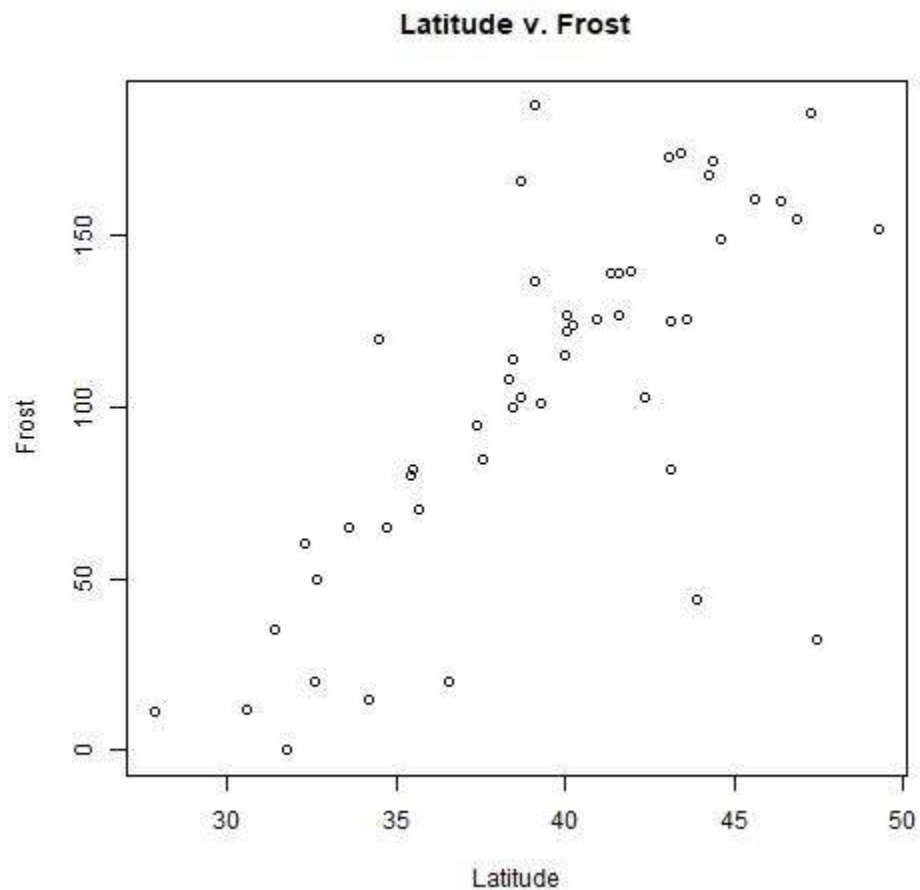
i.

C. Populations for each region

a. North Central = 57,636

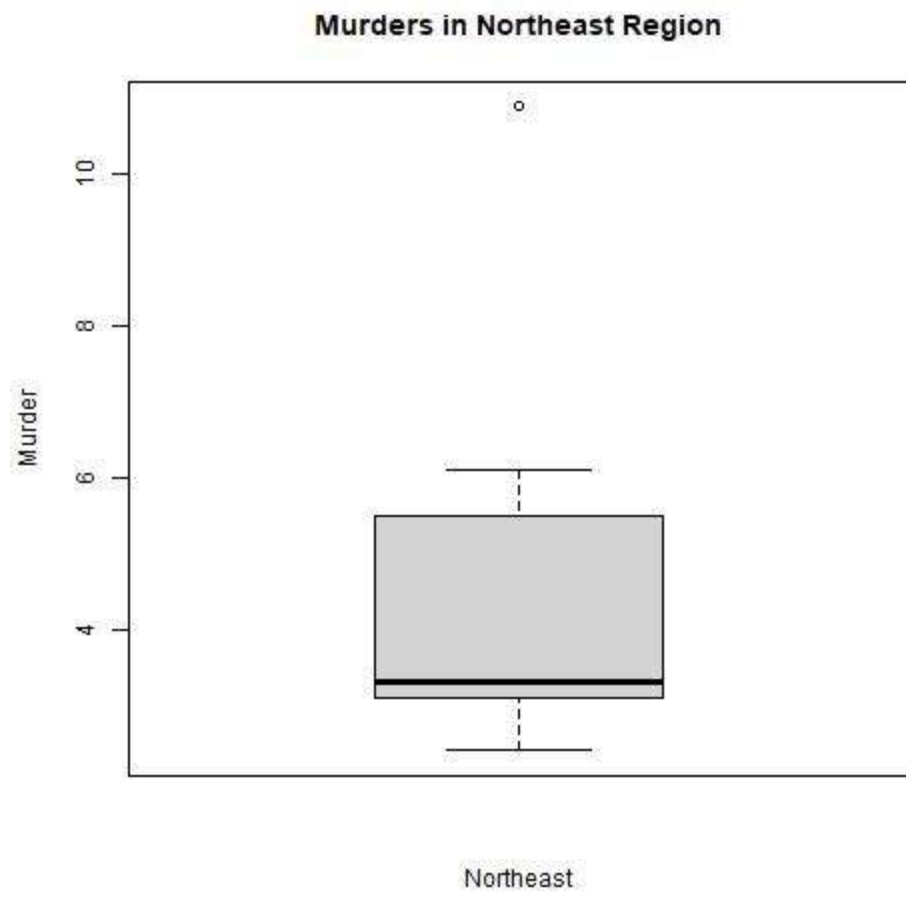b. Northeast = 49,456

c. South = 67,330

d. West = 37,899

D. Box plots comparing the variable Murder for each Region shown below. As seen in the plot below, the Northeast is the safest region with the lowest median murder rate. On the other hand, the South is the most dangerous region with the highest median murder rate.
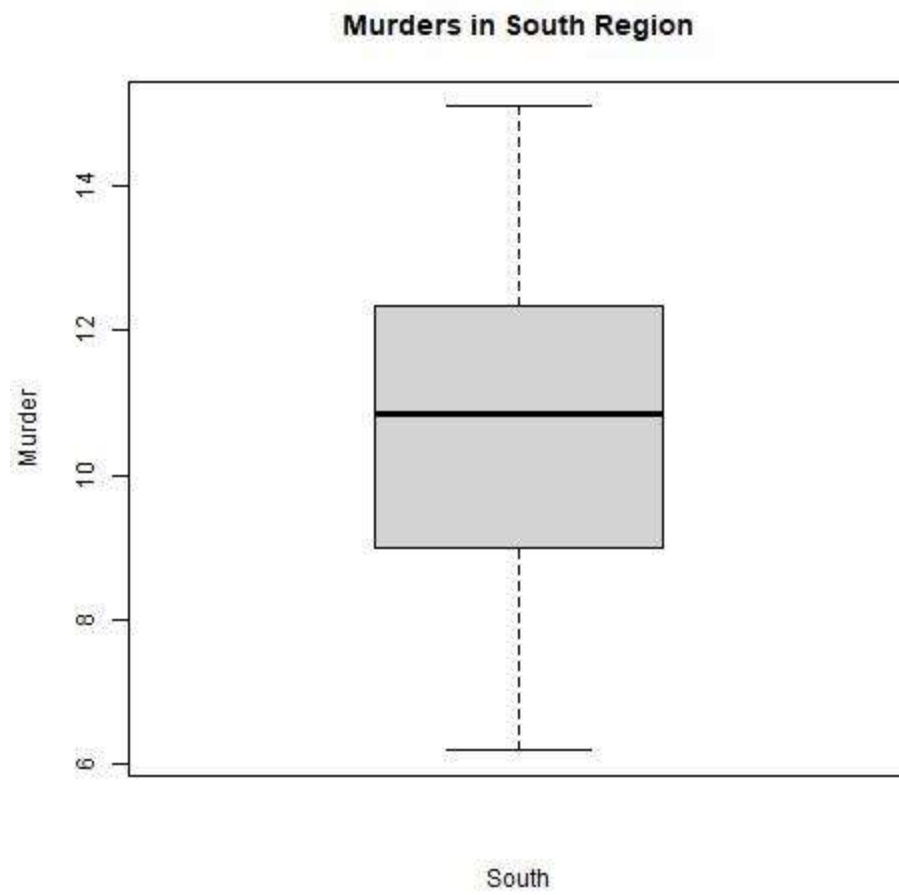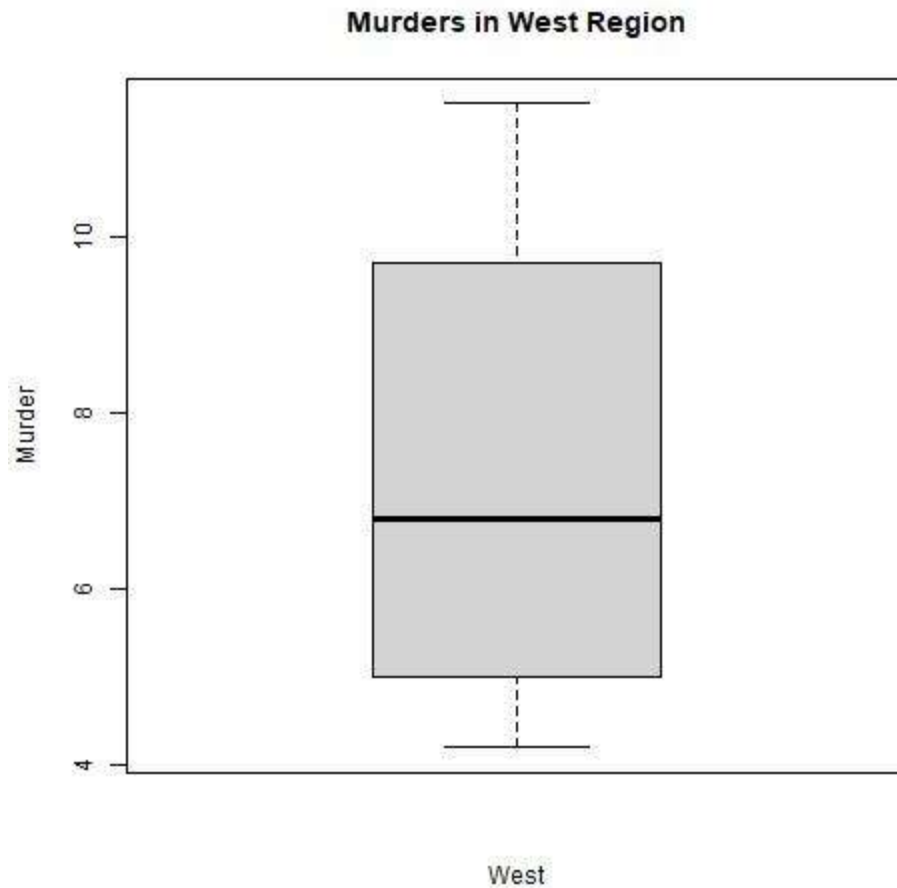
Murders for each Region

a.

Murders in North Central Region

North Central

b.

Murders in Northeast Region

Northeast

C.

Murders in South Region


South

d.

Murders in West Region

West

e.

E. The box plots are shown below. Two things I noticed is that for the income distribution

with high school graduation rates greater than 55%, the South region does not have a

box plot as it does not have a state with a high school graduation rate greater than 55%.

On the other hand, the West does not have a box plot for the second box plot displaying

the income distribution with high school graduation rates below 55%.
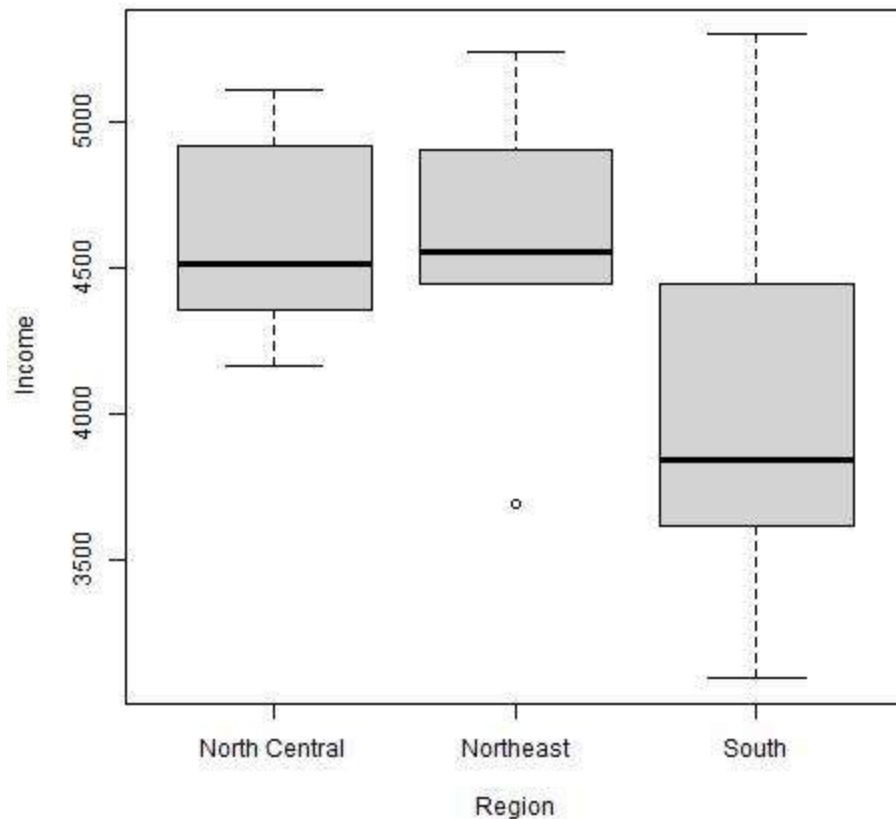
Income Distribution for States with HS Graduation Rate Above 55%
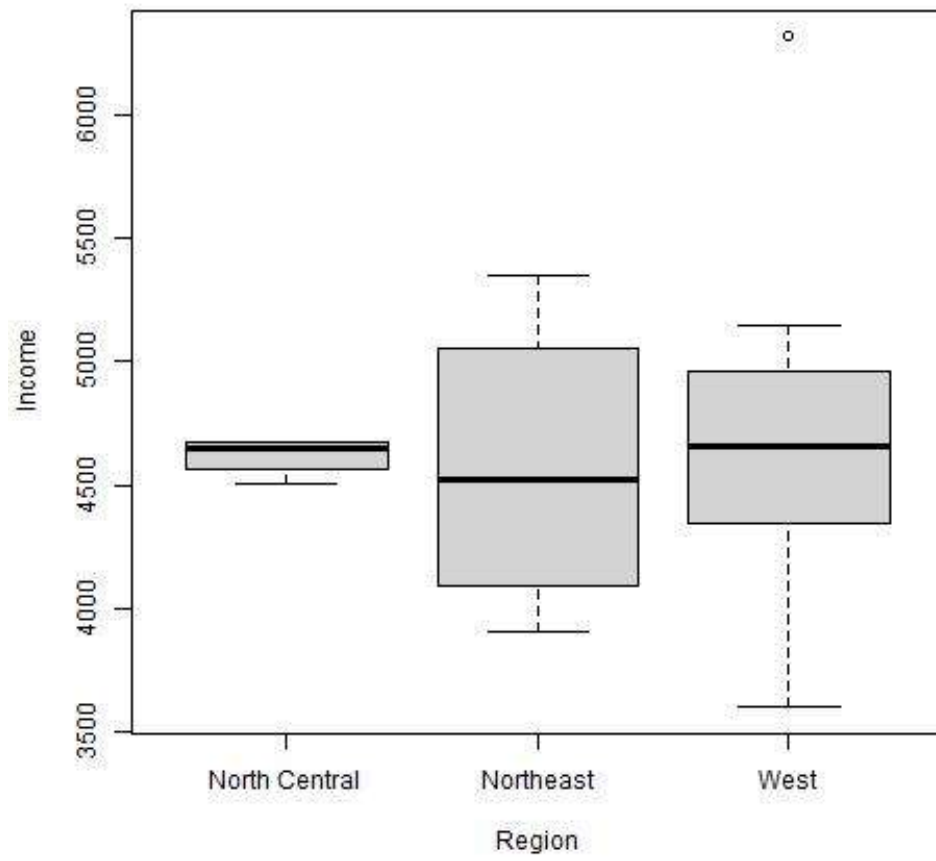
a.

Income Distribution for States with HS Graduation Rate Below 55%

b.

F. I am including a snip-it of the results from this function call below where the column

TRUE is the number of states in that region with an average income greater than $4,500

while FALSE is the number of states in that region that do not have an average income
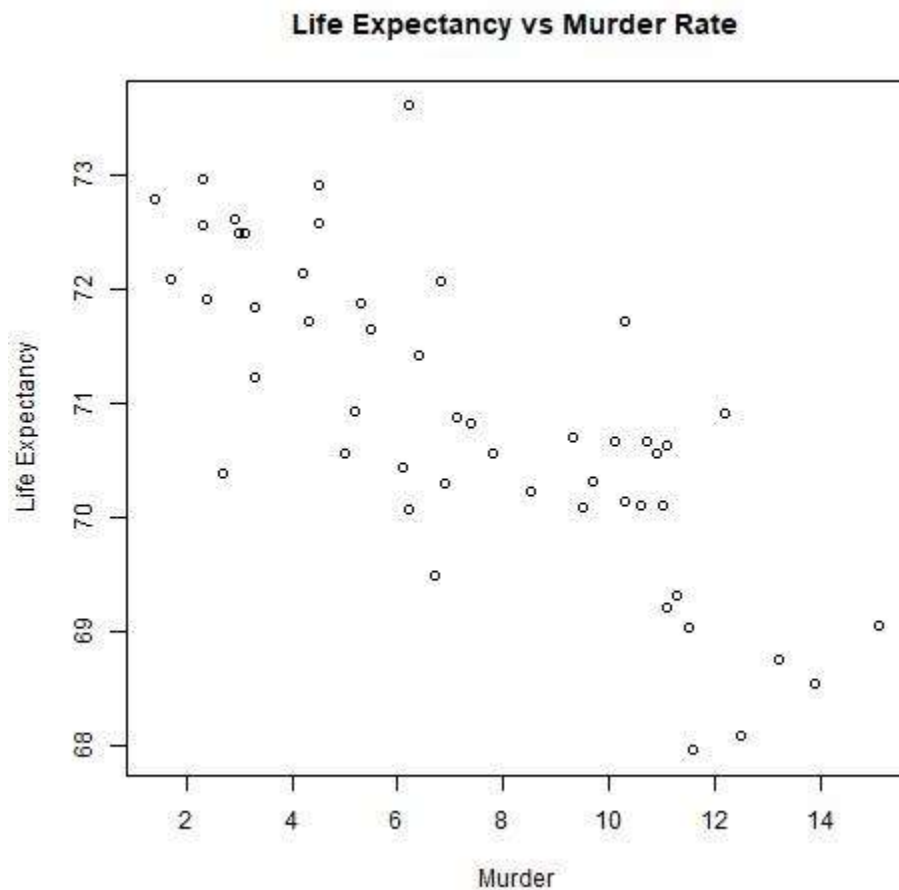
greater than $4,500.

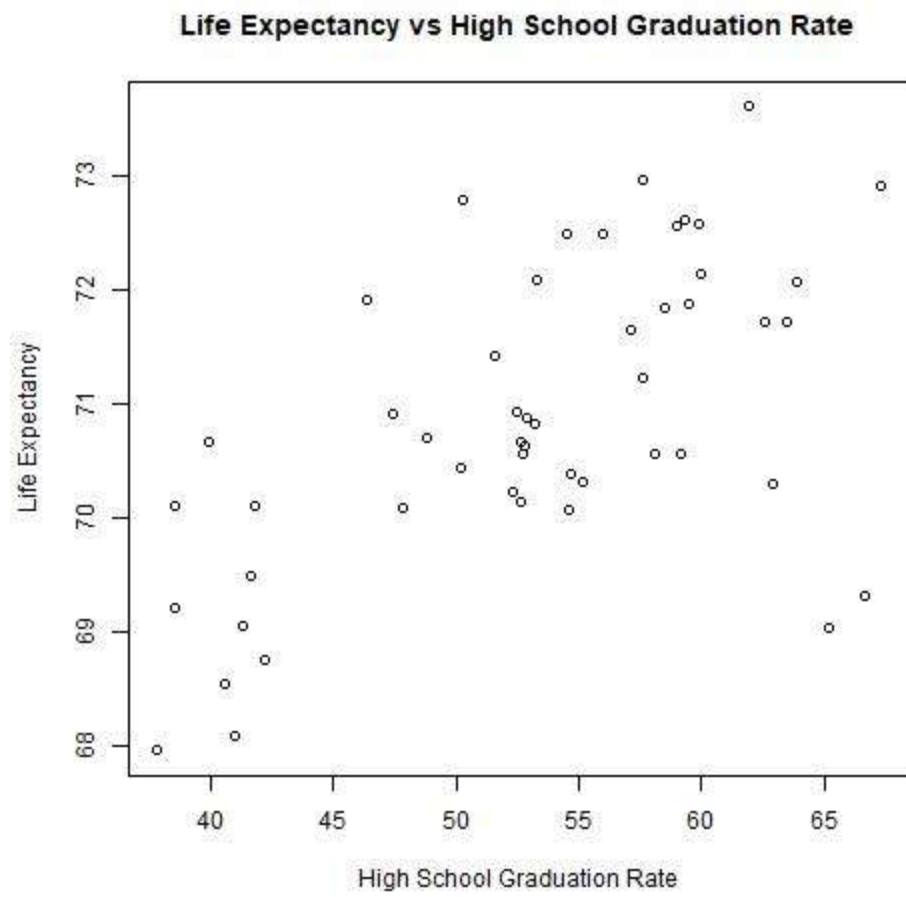|	FALSE	TRUE
North Central	4	8
Northeast	4	5
South	12	4
West	4	9

a.

G. I am including pictures of the scatter plots below. I tried using Murder Rate, High School

Graduation Rate, Population, and Illiteracy Rate. Both the Murder Rate and the Illiteracy

Rate appear to produce a negative relationship in the scatter plot with Life Expectancy.
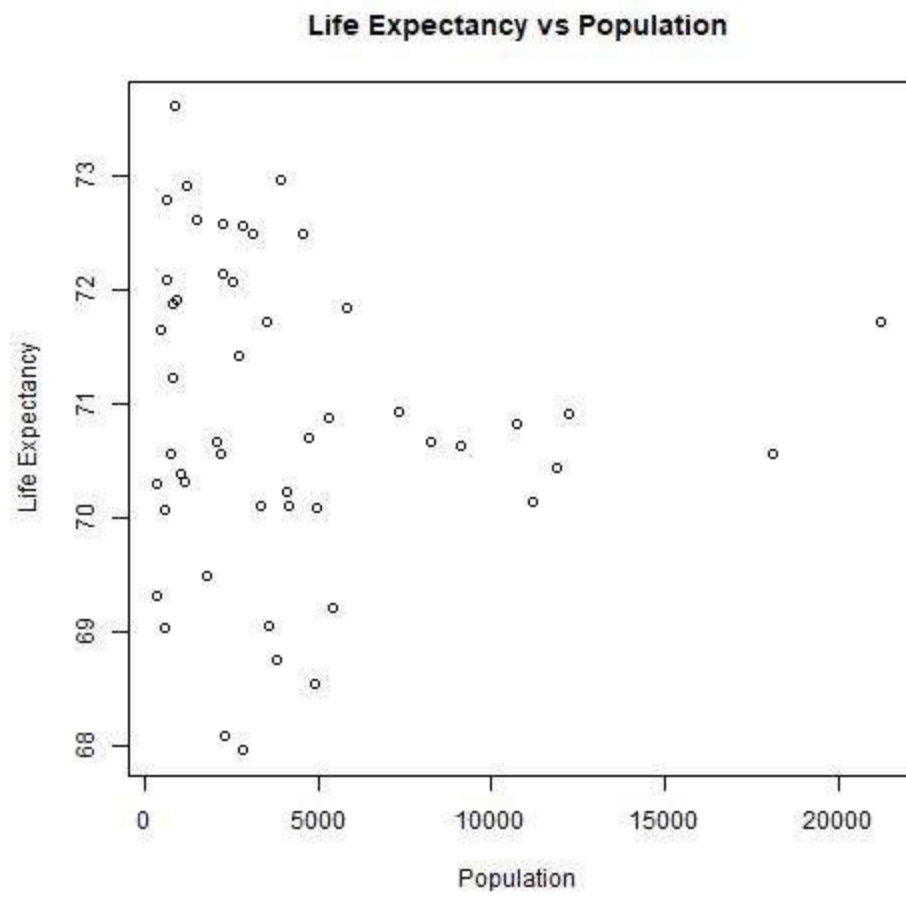
On the other hand, the High School Graduation Rate appears to have a positive

relationship and Population does not appear to have much of a relationship with Life

Expectancy. I expected that Murder Rate would have the strongest negative relationship

with Life Expectancy which it appears it does. However, I did find it surprising that the

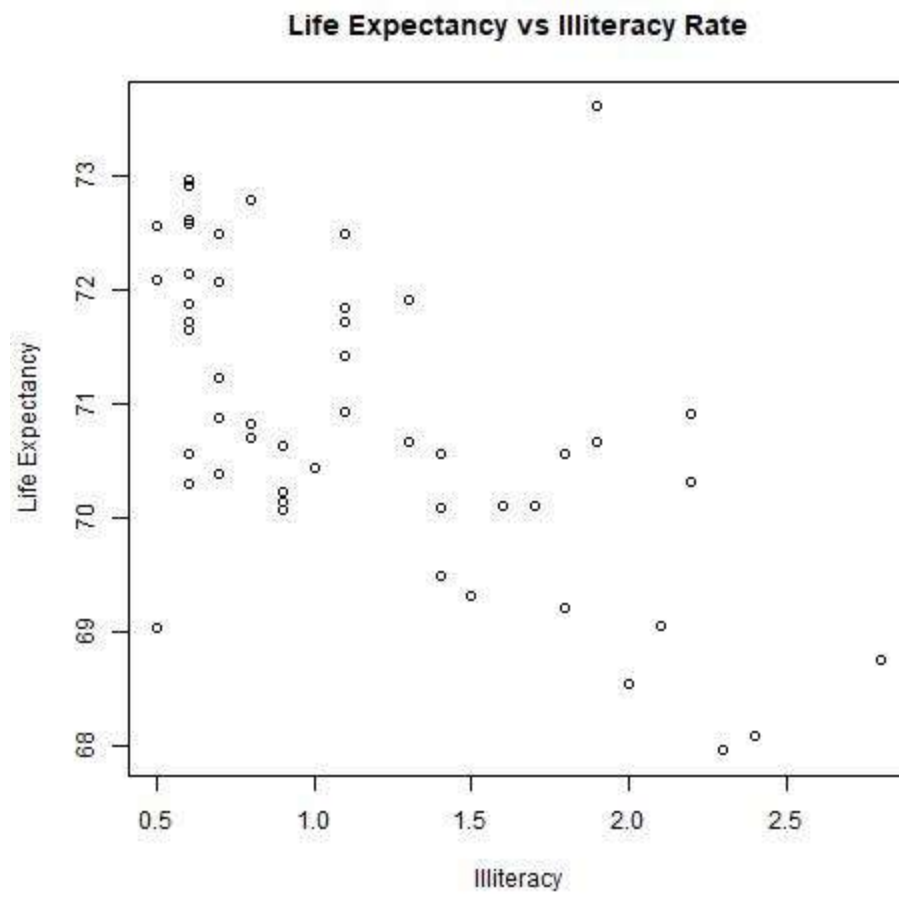Illiteracy Rate scatter plot also appears to show a strong negative relationship with Life

Expectancy.

Life Expectancy vs Murder Rate

a.

Life Expectancy vs High School Graduation Rate

b.

Life Expectancy vs Population

c.

Life Expectancy vs Illiteracy Rate

d.