

Problem Set 1

Applied Stats/Quant Methods 1

Jianxiong Wu—23354731

Due: September 30, 2024

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in R, please include the code you used to get your answers. Please also include the .R file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.
- Your homework should be submitted electronically on GitHub.
- This problem set is due before 23:59 on Monday September 30, 2024. No late assignments will be accepted.

Question 1: Education

A school counselor was curious about the average of IQ of the students in her school and took a random sample of 25 students' IQ scores. The following is the data set:

```
1 y <- c(105, 69, 86, 100, 82, 111, 104, 110, 87, 108, 87, 90, 94, 113, 112, 98,  
      80, 97, 95, 111, 114, 89, 95, 126, 98)
```

1. Find a 90% confidence interval for the average student IQ in the school.

```
1 y <- c(105, 69, 86, 100, 82, 111, 104, 110, 87, 108, 87, 90, 94, 113,  
      112, 98, 80, 97, 95, 111, 114, 89, 95, 126, 98)  
2 t.test(y, conf.level = 0.90, alternative = "two.sided")
```

Results:

data: y

t = 37.593, df = 24, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

90 percent confidence interval:

93.95993 102.92007

sample estimates:

mean of x

98.44

Answer: 90 percent confidence interval:93.95993 102.92007

2. Next, the school counselor was curious whether the average student IQ in her school is higher than the average IQ score (100) among all the schools in the country.

Using the same sample, conduct the appropriate hypothesis test with $\alpha = 0.05$.

```
1 t.test(y, mu = 100, alternative = "greater")
```

Results:

data: y

t = -0.59574, df = 24, p-value = 0.7215

alternative hypothesis: true mean is greater than 100

95 percent confidence interval:

93.95993 Inf

sample estimates:

mean of x

98.44

Answer: p-value = 0.7215, p-value > 0.05, so there is not sufficient evidence to suggest that the average student IQ in school is higher than the average IQ score (100) among all the schools in the country.

Question 2: Political Economy

Researchers are curious about what affects the amount of money communities spend on addressing homelessness. The following variables constitute our data set about social welfare expenditures in the USA.

State	50 states in US
Y	per capita expenditure on shelters/housing assistance in state
X1	per capita personal income in state
X2	Number of residents per 100,000 that are "financially insecure" in state
X3	Number of people per thousand residing in urban areas in state
Region	1=Northeast, 2= North Central, 3= South, 4=West

Explore the `expenditure` data set and import data into R.

- Please plot the relationships among Y, X1, X2, and X3? What are the correlations among them (you just need to describe the graph and the relationships among them)?

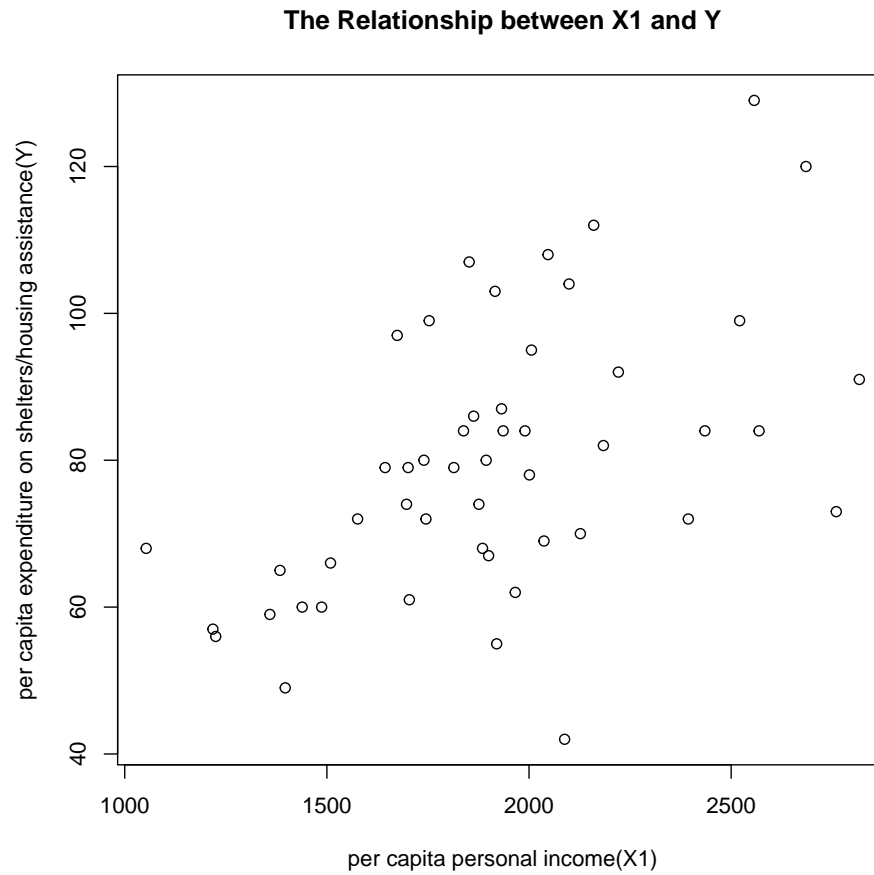
```
1 pdf( file="/Users/wuhadou/Desktop/StatsI_Fall2024-main/problemSets/PS01/my
  _answer/scatter_plot1_X1&Y.pdf")
2 plot(expenditure$X1, expenditure$Y,
3       xlab = "per capita personal income(X1)",
4       ylab = "per capita expenditure on shelters/housing assistance(Y)",
5       main = "The Relationship between X1 and Y")
6 dev.off()
7
8 pdf( file="/Users/wuhadou/Desktop/StatsI_Fall2024-main/problemSets/PS01/my
  _answer/scatter_plot2_X2&Y.pdf")
9 plot(expenditure$X2, expenditure$Y,
10      xlab = "Number of residents per 100,000 that are 'financially
  insecure '(X2)",
11      ylab = "per capita expenditure on shelters/housing assistance(Y)",
12      main = "The Relationship between X2 and Y")
13 dev.off()
14
15 pdf( file="/Users/wuhadou/Desktop/StatsI_Fall2024-main/problemSets/PS01/my
  _answer/scatter_plot3_X3&Y.pdf")
16 plot(expenditure$X3, expenditure$Y,
17      xlab = "Number of people per thousand residing in urban areas(X3)",
18      ylab = "per capita expenditure on shelters/housing assistance(Y)",
19      main = "The Relationship between X3 and Y")
20 dev.off()
21
22 pdf( file="/Users/wuhadou/Desktop/StatsI_Fall2024-main/problemSets/PS01/my
  _answer/scatter_plot4_X1&X2.pdf")
```

```

23 plot(expenditure$X1, expenditure$X2,
24       xlab = "per capita personal income(X1)",
25       ylab = "pNumber of residents per 100,000 that are 'financially
        insecure'(X2)",
26       main = "The Relationship between X1 and X2")
27 dev.off()
28
29 pdf(file="/Users/wuhadou/Desktop/StatsI_Fall2024-main/problemSets/PS01/my
        _answer/scatter_plot5_X1&X3.pdf")
30 plot(expenditure$X1, expenditure$X3,
31       xlab = "per capita personal income(X1)",
32       ylab = "Number of people per thousand residing in urban areas(X3)",
33       main = "The Relationship between X1 and X3")
34 dev.off()
35
36 pdf(file="/Users/wuhadou/Desktop/StatsI_Fall2024-main/problemSets/PS01/my
        _answer/scatter_plot6_X2&X3.pdf")
37 plot(expenditure$X2, expenditure$X3,
38       xlab = "pNumber of residents per 100,000 that are 'financially
        insecure'(X2)",
39       ylab = "Number of people per thousand residing in urban areas(X3)",
40       main = "The Relationship between X2 and X3")
41 dev.off()
42
43 cor(expenditure[,c("Y", "X1", "X2", "X3")])

```

Figure 1: X1 and Y



Results: Figure 1 shows that the correlation coefficient between X1 and Y is 0.5317212, indicating a positive correlation.

Figure 2: X2 and Y

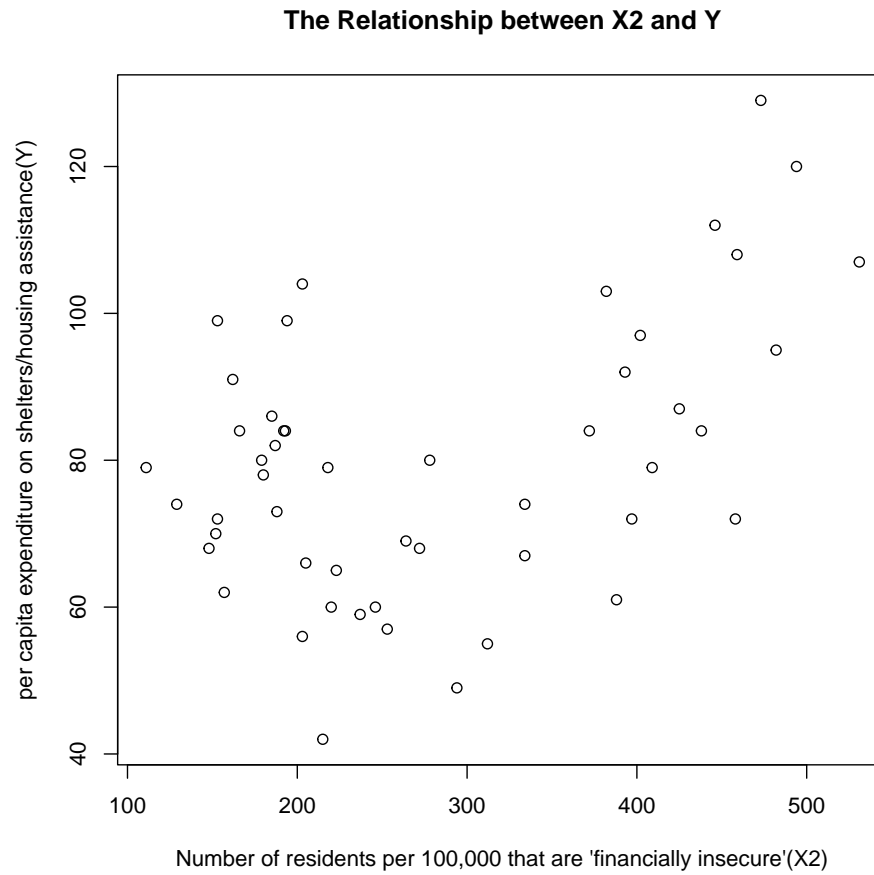
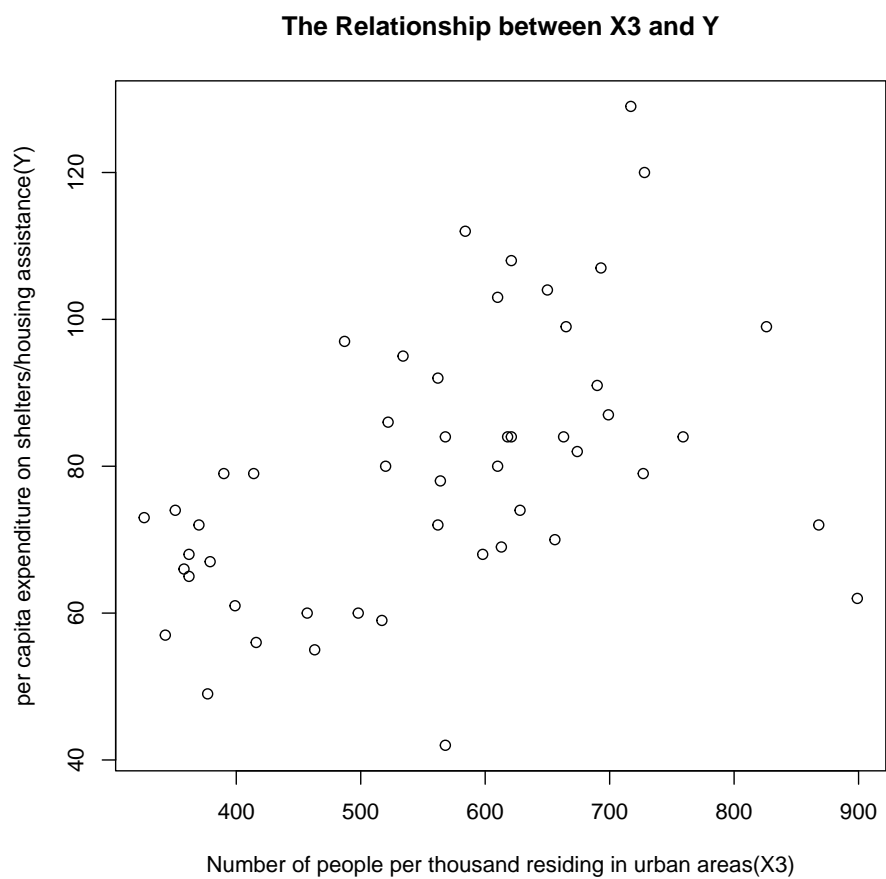
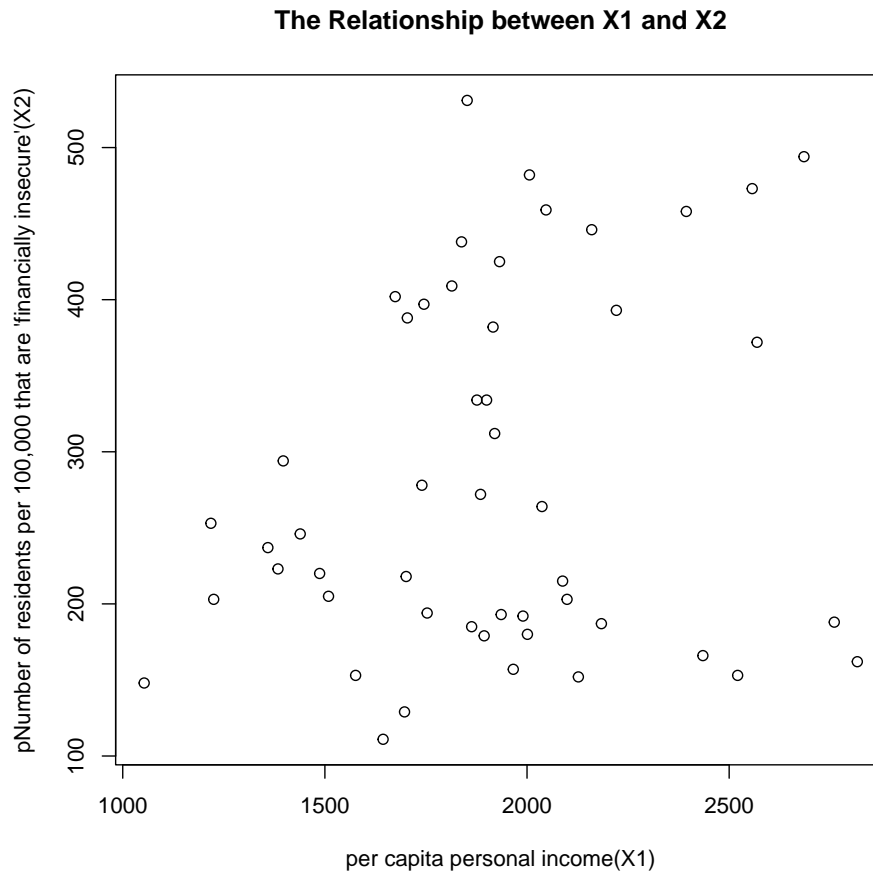


Figure 3: X3 and Y



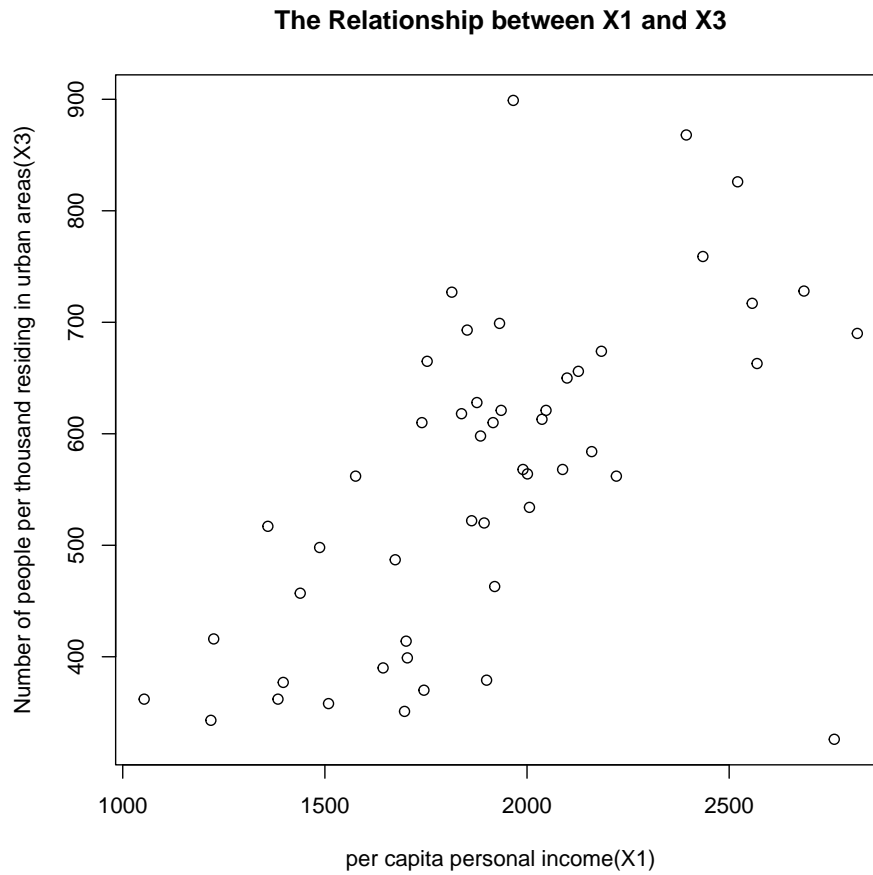
Results: Figure 3 shows that the correlation coefficient between X3 and Y is 0.4636787, indicating a positive correlation.

Figure 4: X1 and X2



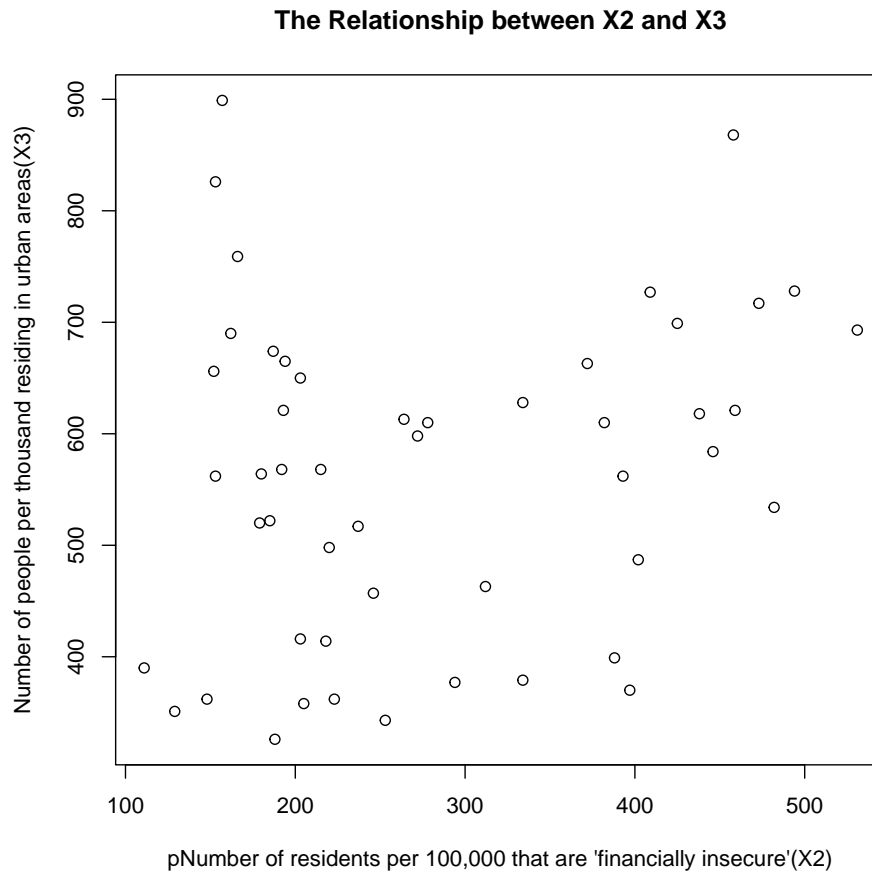
Results: Figure 4 shows that the correlation coefficient between X1 and X2 is 0.2056101, indicating a positive correlation, but this relationship is relatively weak.

Figure 5: X1 and X3



Results: Figure 5 shows that the correlation coefficient between X1 and X3 is 0.5952504, indicating a positive correlation.

Figure 6: X2 and X3



Results: Figure 6 shows that the correlation coefficient between X2 and X3 is 0.2210149, indicating a positive correlation, but this relationship is relatively weak.

Results:

	Y	X1	X2	X3
Y	1.0000000	0.5317212	0.4482876	0.4636787
X1	0.5317212	1.0000000	0.2056101	0.5952504
X2	0.4482876	0.2056101	1.0000000	0.2210149
X3	0.4636787	0.5952504	0.2210149	1.0000000

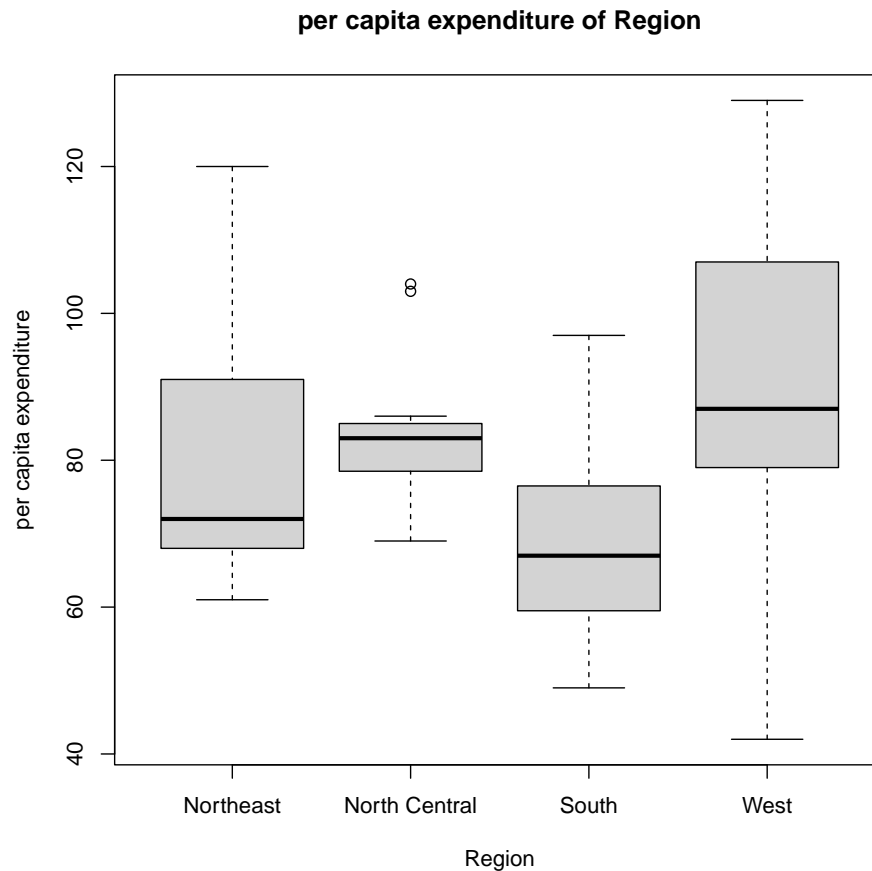
Answer:

The results show that the relationships between Y, X1, X2 and X3 are all positively correlated. The correlation coefficient between X1 and X3 is 0.5952504 and between X1 and Y is 0.5317212, which indicates that the positive relationship between these two groups is relatively strong among all the data. While the correlation coefficient between X1 and X2 is 0.2056101 and between X2 and X3 is 0.2210149, indicating that the positive relationship between these two groups of data is relatively weak.

- Please plot the relationship between *Y* and *Region*? On average, which region has the highest per capita expenditure on housing assistance?

```
1 pdf( file="/Users/wuhadou/Desktop/StatsI_Fall2024-main/problemSets/PS01/my
  _answer/boxplot_plot7_Y&Region.pdf")
2 boxplot(expenditure$Y ~ expenditure$Region,
3         main="per capita expenditure of Region",
4         ylab="per capita expenditure",
5         xlab="Region",
6         names=c("Northeast", "North Central", "South", "West"))
7 dev.off()
```

Figure 7: Y and Region



Answer: Figure 7 shows that, region West has the highest per capita expenditure on housing assistance.

- Please plot the relationship between Y and $X1$? Describe this graph and the relationship. Reproduce the above graph including one more variable *Region* and display different regions with different types of symbols and colors.

```

1 pdf( file="/Users/wuhadou/Desktop/StatsI_Fall2024-main/problemSets/PS01/my_
  answer/scatter_plot8_X1&Y_Region.pdf")
2 plot(expenditure$X1, expenditure$Y,
3       col = expenditure$Region,
4       xlab = "per capita personal income(X1)",
5       ylab = "per capita expenditure on shelters/housing assistance(Y)",
6       main = "The Relationship between per capita personal income and per
  capita expenditure by Region")

```

```

7 legend("topleft",
8       legend = c("Northeast", "North Central", "South", "West"),
9       col = 1:4,
10      pch = 1)
11 dev.off()

```

Figure 8: Y and X1(Region)

relationship between per capita personal income and per capita expenditure

