

Problem Set 1

Applied Stats/Quant Methods 1

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. 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,  
2       112, 98, 80, 97, 95, 111, 114, 89, 95, 126, 98)  
3 t_score <- qt(0.95, df=length(y)-1)  
4 lower_90_t <- mean(y)-(t_score)*(sd(y)/sqrt(length(y)))  
5 upper_90_t <- mean(y)+(t_score)*(sd(y)/sqrt(length(y)))  
6  
7 lower_90_t  
8 mean(y)  
9 upper_90_t
```

I have calculated that the 90% confidence interval is between 93.95993 and 102.9201.

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")
```

After running this code, I got a p-value of 0.7215, thereby allowing me to conclude that there is insufficient evidence to reject the null hypothesis.

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.

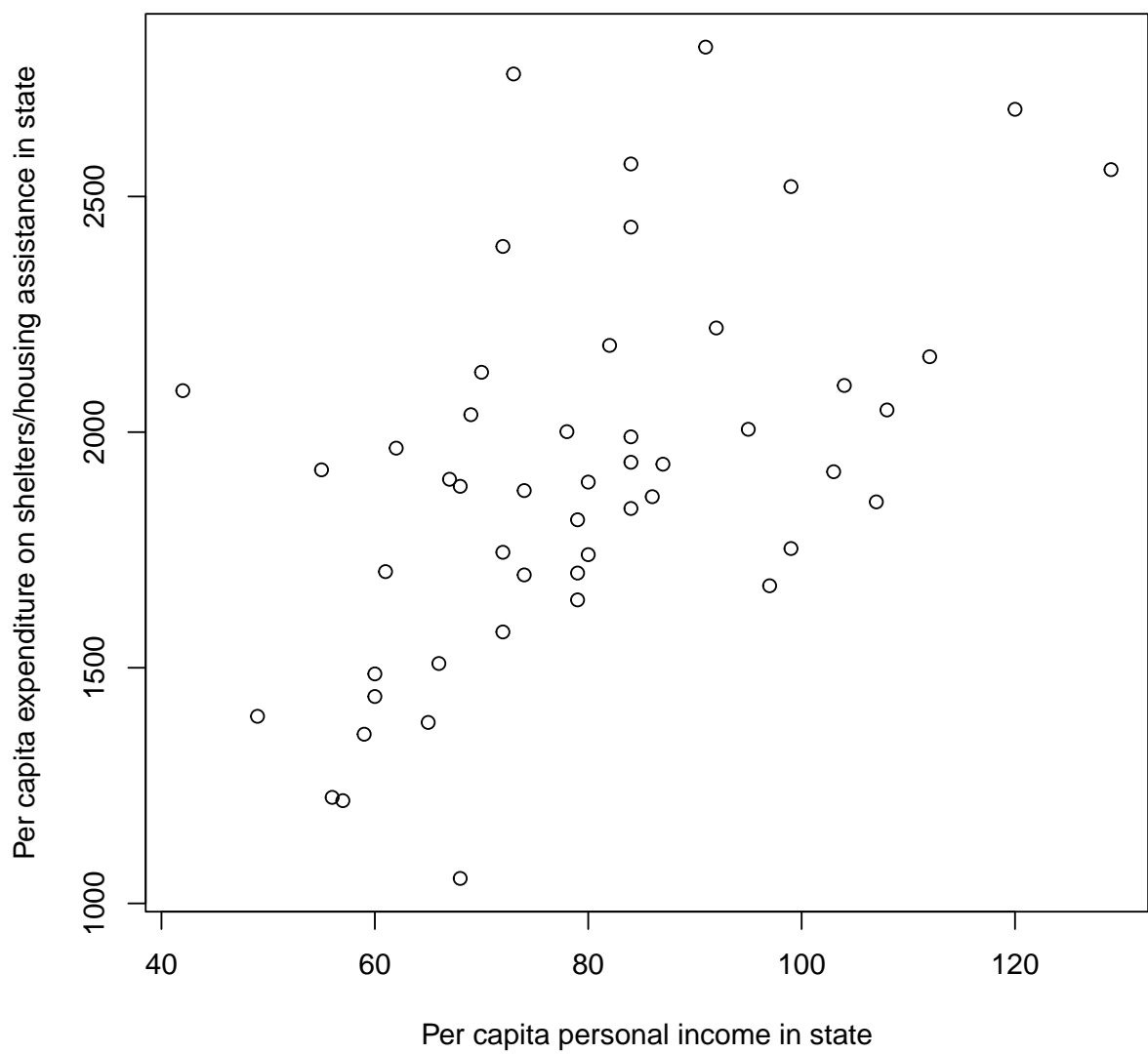
State	50 states in US
Y	per capita expenditure on shelters/housing assistance in state
X1	per capita personal income in state
X2	Number of residents per 100,000 that are "financially insecure" in state
X3	Number of people per thousand residing in urban areas in state
Region	1=Northeast, 2= North Central,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

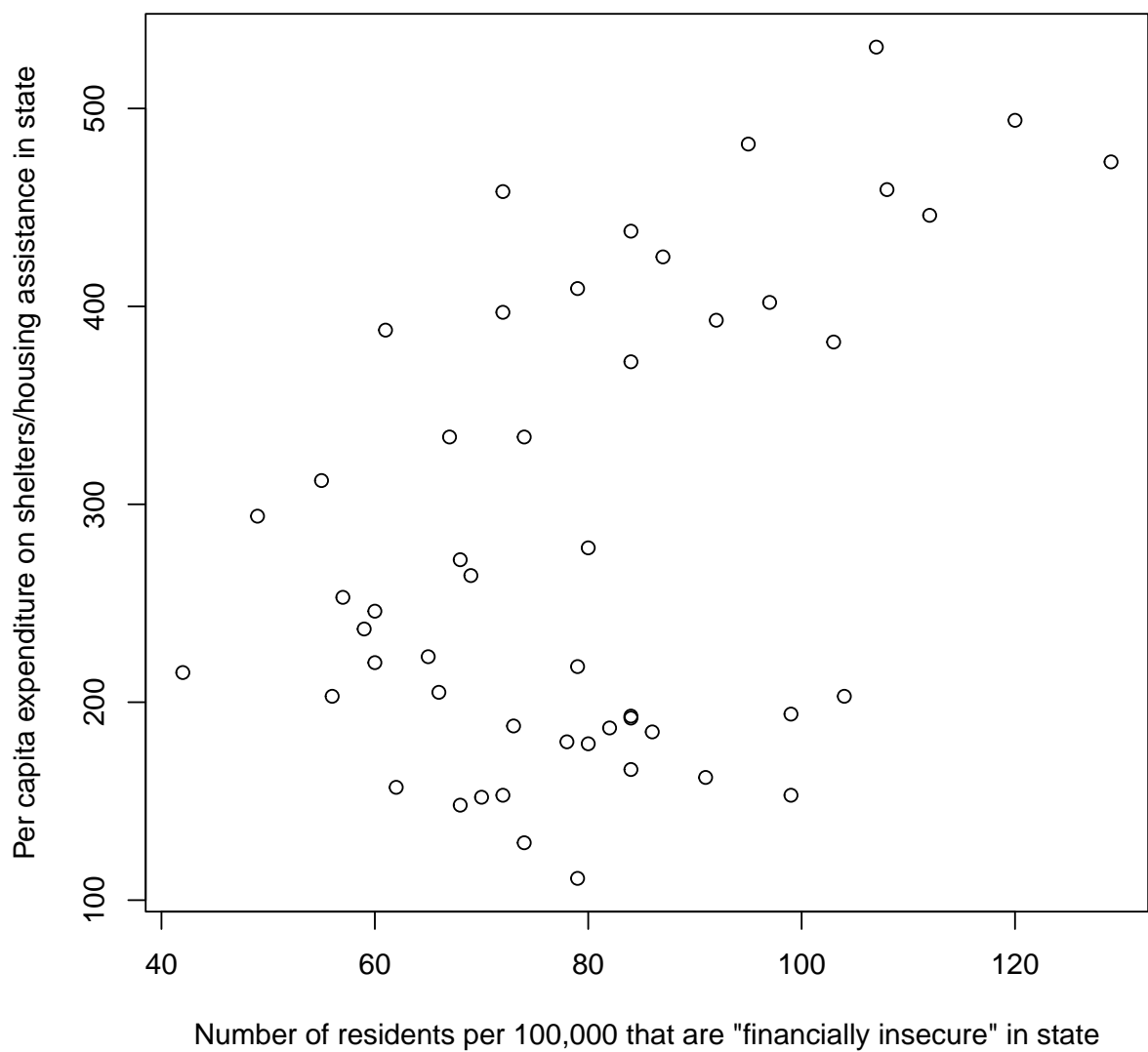
Explore the `expenditure` data set and import data into R.

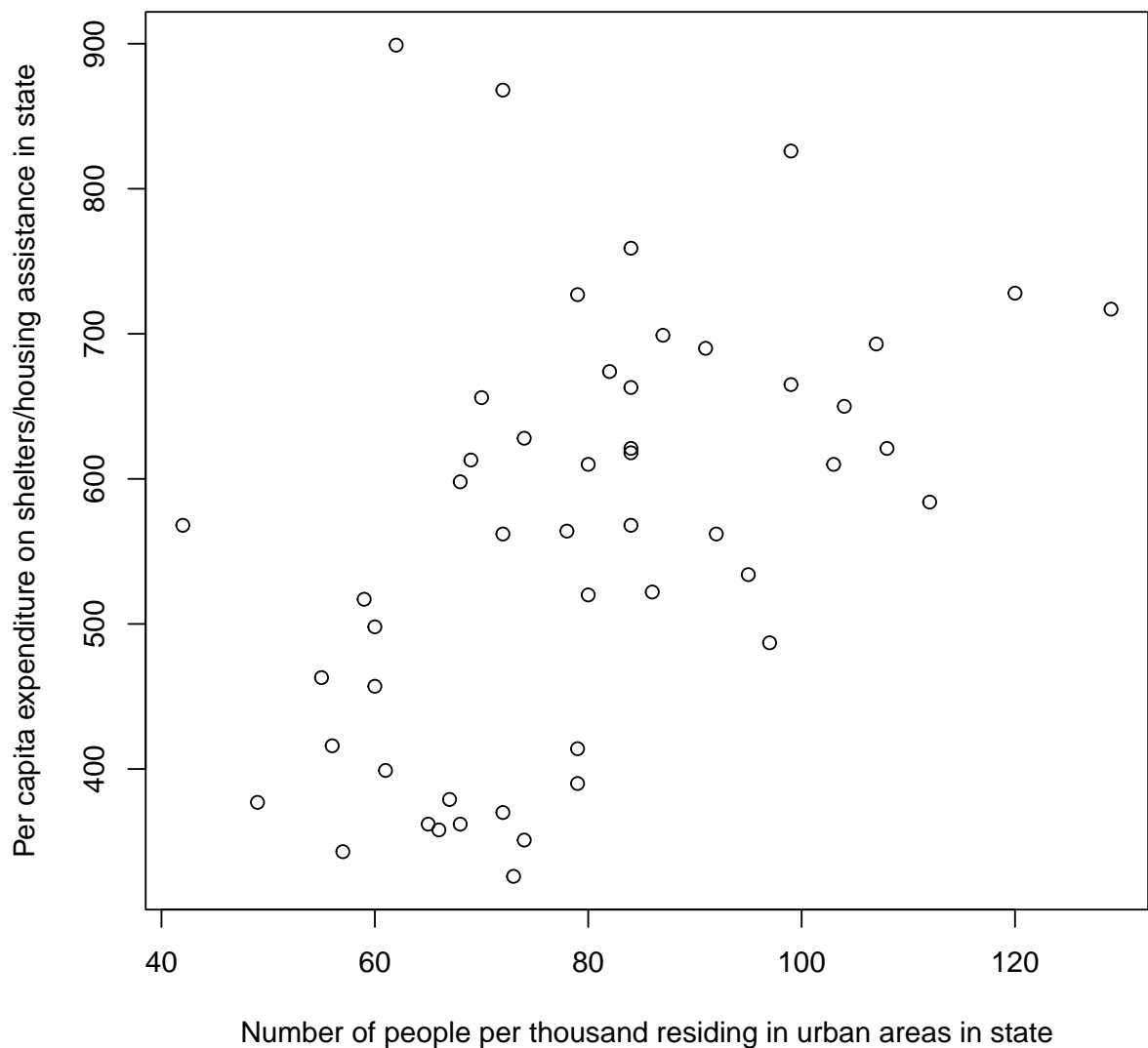
- Please plot the relationships among Y , $X1$, $X2$, and $X3$? What are the correlations among them (you just need to describe the graph and the relationships among them)?

All of these have weak to moderate positive correlations, with Y and $X1$, and Y and $X3$ exhibiting the strongest correlations.

```
1 pdf( file="plot1.pdf" ) +
2 plot( expenditure$Y, expenditure$X1,
3       xlab="Per capita personal income in state",
4       ylab="Per capita expenditure on shelters/housing assistance in state")
5 dev.off()
6
7 pdf( file="plot2.pdf" ) +
8 plot( expenditure$Y, expenditure$X2,
9       xlab="Number of residents per 100,000 that are 'financially insecure
10        ' in state",
11       ylab="Per capita expenditure on shelters/housing assistance in state
12        ") +
13 dev.off()
14
15 pdf( file="plot3.pdf" ) +
16 plot( expenditure$Y, expenditure$X3,
17       xlab="Number of people per thousand residing in urban areas in state
18        ",
19       ylab="Per capita expenditure on shelters/housing assistance in state
20        ") +
21 dev.off()
```







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

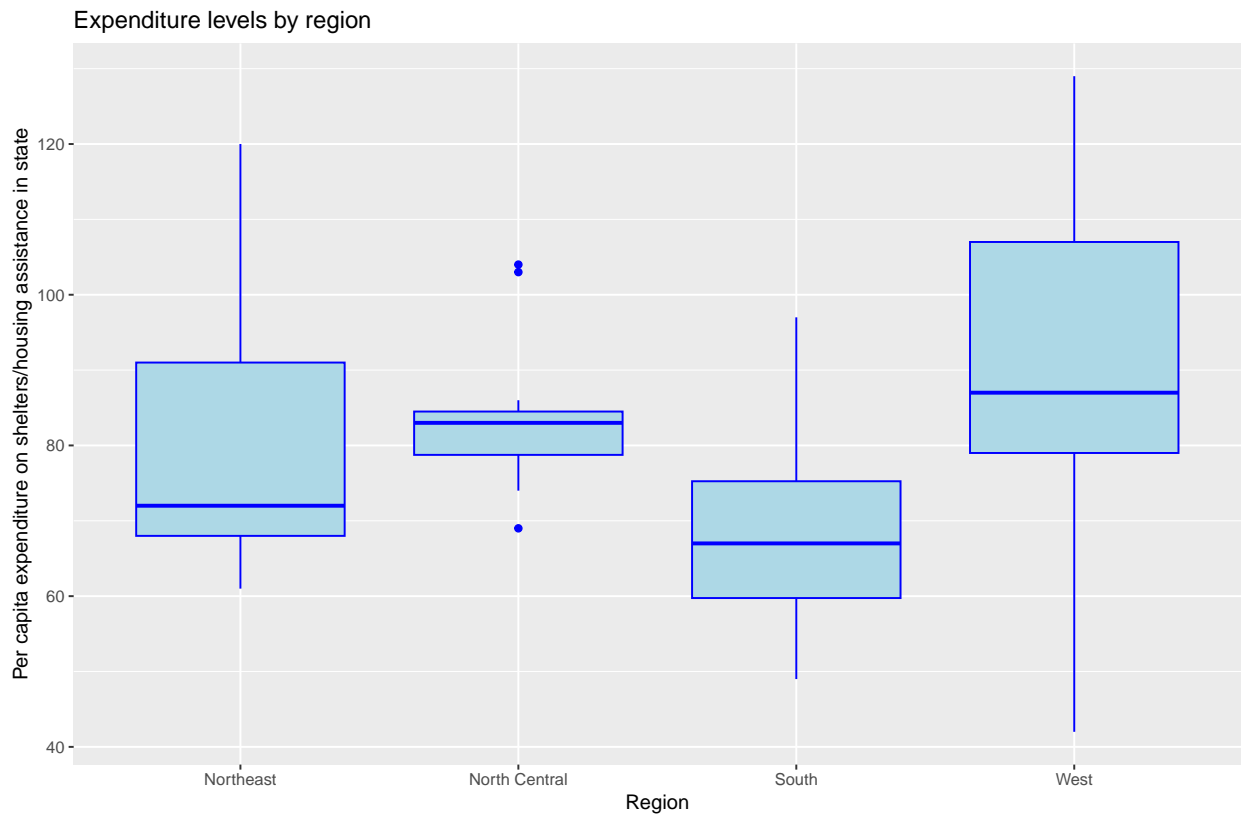
```

1 expenditure$Region <- factor(expenditure$Region, levels = c(1, 2, 3, 4),
2   labels = c("Northeast", "North Central", "South", "West"))
3
4 ggplot(data = expenditure, aes(x = Region, y = Y)) +
5   geom_boxplot(fill = "lightblue", color = "blue") +
6   labs(title = "Expenditure levels by region", x = "Region",
7     y = "Per capita expenditure on shelters/housing assistance in state")

```

```
7 ggsave(filename = "expenditure_levels_by_region.pdf")
```

It is clear that the West has the highest per capita expenditure on housing assistance.

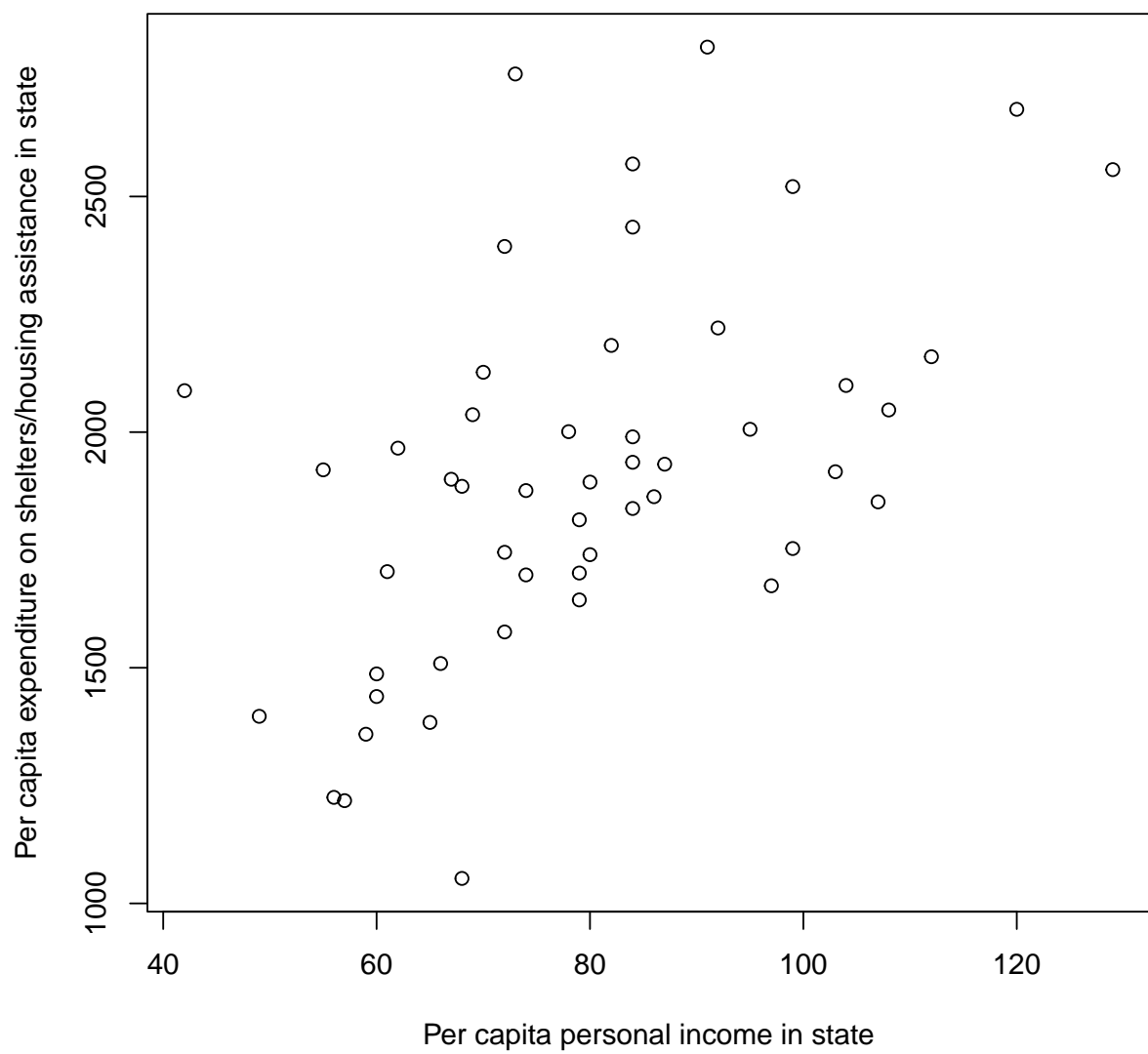


- Please plot the relationship between Y and $X1$? Describe this graph and the relationship. Reproduce the above graph including one more variable *Region* and display different regions with different types of symbols and colors.

There is a weak to moderate positive correlation.

```
1
2 ggplot(data = expenditure, aes(x = Region, y = Y)) +
3   geom_boxplot(fill = "lightblue", color = "blue") +
4   labs(title = "Expenditure levels by region", x = "Region",
5     y = "Per capita expenditure on shelters/housing assistance in state")
6 ggsave(filename = "expenditure_levels_by_region.pdf")
7
8 pdf(file="plot1_2.pdf") +
9   plot(expenditure$Y, expenditure$X1,
10     xlab="Per capita personal income in state",
11     ylab="Per capita expenditure on shelters/housing assistance in state")
+

```



Expenditure levels by region

