Problem Set 1

Paula Montano/Applied Stats/Quant Methods 1

Due: October 1, 2021

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 in `.pdf` form.
- This problem set is due before 8:00 on Friday October 1, 2021. No late assignments will be accepted.
- Total available points for this homework is 100.

Question 1 (50 points): 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:

```
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.

   Confidence Interval

```r
z90 <-  qnorm((1-.90) / 2, lower.tail = FALSE)
n <- length(na.omit(y))
IQ_sample_mean <- mean(y, na.rm = TRUE)
IQ_sample_sd <- sd(y, na.rm = TRUE)
lower_90 <- IQ_sample_mean - (z90 * (IQ_sample_sd / sqrt(n)))
upper_90 <- IQ_sample_mean + (z90 * (IQ_sample_sd / sqrt(n)))
confint90 <- c(lower_90, upper_90)
```

Answer: We estimate with 90 per cent of certainty that the average student IQ in the school can be between the range of values of 94.1 (lower level) and 102.7 (upper level).

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 $\alpha = 0.05$.

   Hypothesis Testing
   Null Hypothesis: The students IQ in the school is lower than students IQ in the country.
   Alternative Hypothesis: The students IQ in the school is higher than the students IQ in the country.

```r
#### Solution:
IQ_test <- t.test(y, mu = 100, conf.level = 0.95, alternative = "greater")
```

Answer: We cannot reject the null hypothesis as our p.value is greater than alpha = 0.05.

Question 2 (50 points): 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.

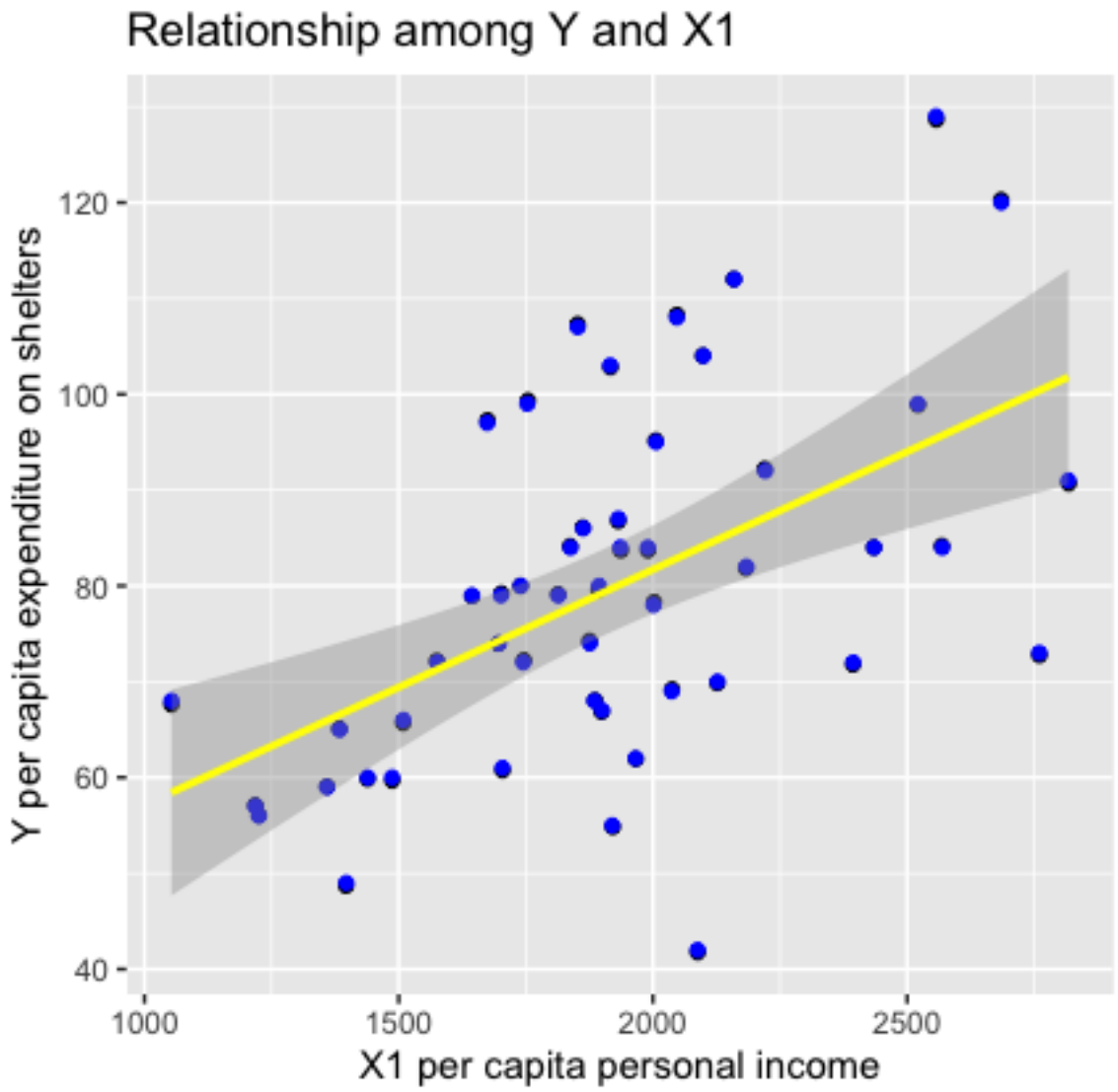
`State`	*50 states in US*
`Y`	*per capita expenditure on shelters/housing assistance in state*
`X1`	*per capita personal income in state*
`X2`	*Number of residents per 100,000 that are "financially insecure" in state*
`X3`	*Number of people per thousand residing in urban areas in state*
`Region`	*1=Northeast, 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)?

- 2.1. Plot 1 Relationship Y, X1

```
##Plot 1 Relationship Y, X1
plot_Y_X1 <-ggplot(expenditure, aes(x = X1, y = Y)) +
  geom_jitter() + geom_point(color = "blue") +
  labs(x = "X1 per capita personal income", y = "Y per capita expenditure
    on shelters",
       title = "Relationship among Y and X1") +
```



Interpretation:
Plot 1 displays an upward trend. The slope also shows this positive linear relationship between X1 and Y.
This indicates a strong and positive correlation between X1 (per capita personal income) and
Y (per capita expenditure on shelters) in 50 States in the US. When X1 increases Y also increases.

Plot 2 Relationship Y, X2

```
##Plot 2  Relationship Y, X2
plot_Y_X2 <-ggplot(expenditure, aes(x = X2, y = Y)) +
  geom_jitter() + geom_point(color = "blue") +
```

```
4   labs(x = "X2 Number of residents per 100,000 that are financially
      insecure", y = "Y per capita expenditure on shelters",
5         title = "Relationship among Y and X2") +
```

Interpretation:
Plot 2 displays an upward trend. This indicates a positive linear correlation between X2 (Number of residents per 100,000 that are financially insecure) and Y (per capita expenditure on shelters) in states.
The slope also shows this positive relationship between X2 and Y.
The data indicates that the higher the number of residents that are "financially insecure" the higher is the value in per capita expenditure on shelters.

Plot 3 Relationship Y, X3

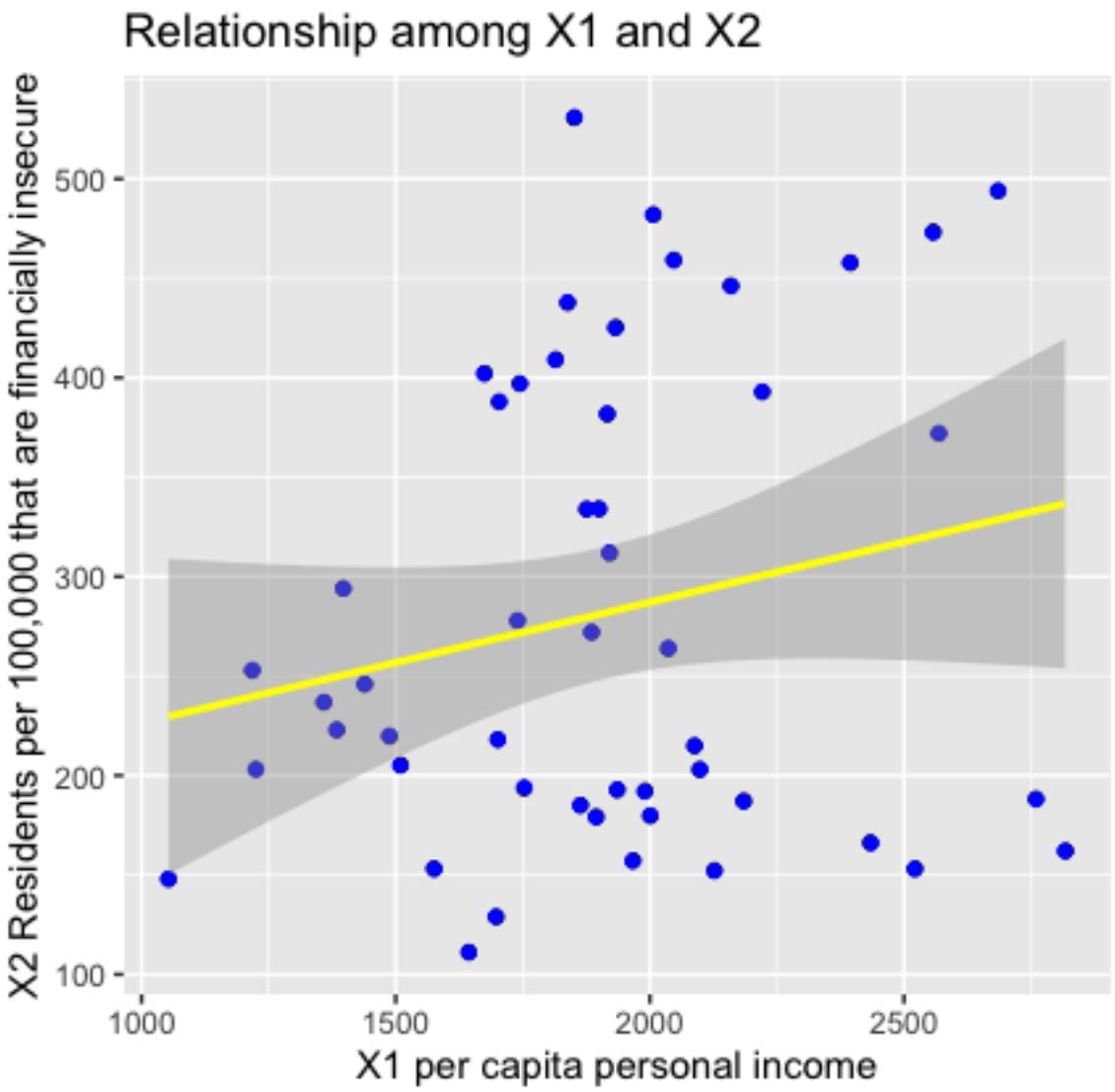
```
##Plot 3 Relationship Y, X3
plot_Y_X3 <-ggplot(expenditure, aes(x = X3, y = Y)) +
  geom_jitter() + geom_point(color = "blue") +
  labs(x = "X3 People per thousand residing in urban areas", y = "Y per
    capita expenditure on shelters",
       title = "Relationship among Y and X3") +
```

Interpretation:
Plot 3 displays a positive upward trend. This indicates a linear correlation between X3 (People per thousand residing in urban areas) and Y (per capita expenditure on shelters) in 50 states in US.
The slope also confirm this positive relationship between X3 and Y. Nonetheless, the

strength appear to be low.
The data indicates that the higher people residing in urban areas the greater is the value per capita in expenditure on shelters.

Plot 4 Relationship X1, X2

```
##Plot 4 Relationship X1, X2
plot_X1_X2 <-ggplot(expenditure, aes(x = X1, y = X2)) +
  geom_jitter() + geom_point(color = "blue") +
  labs(x = "X1 per capita personal income", y = "X2 Residents per 100,000
     that are financially insecure",
        title = "Relationship among X1 and X2") +
```

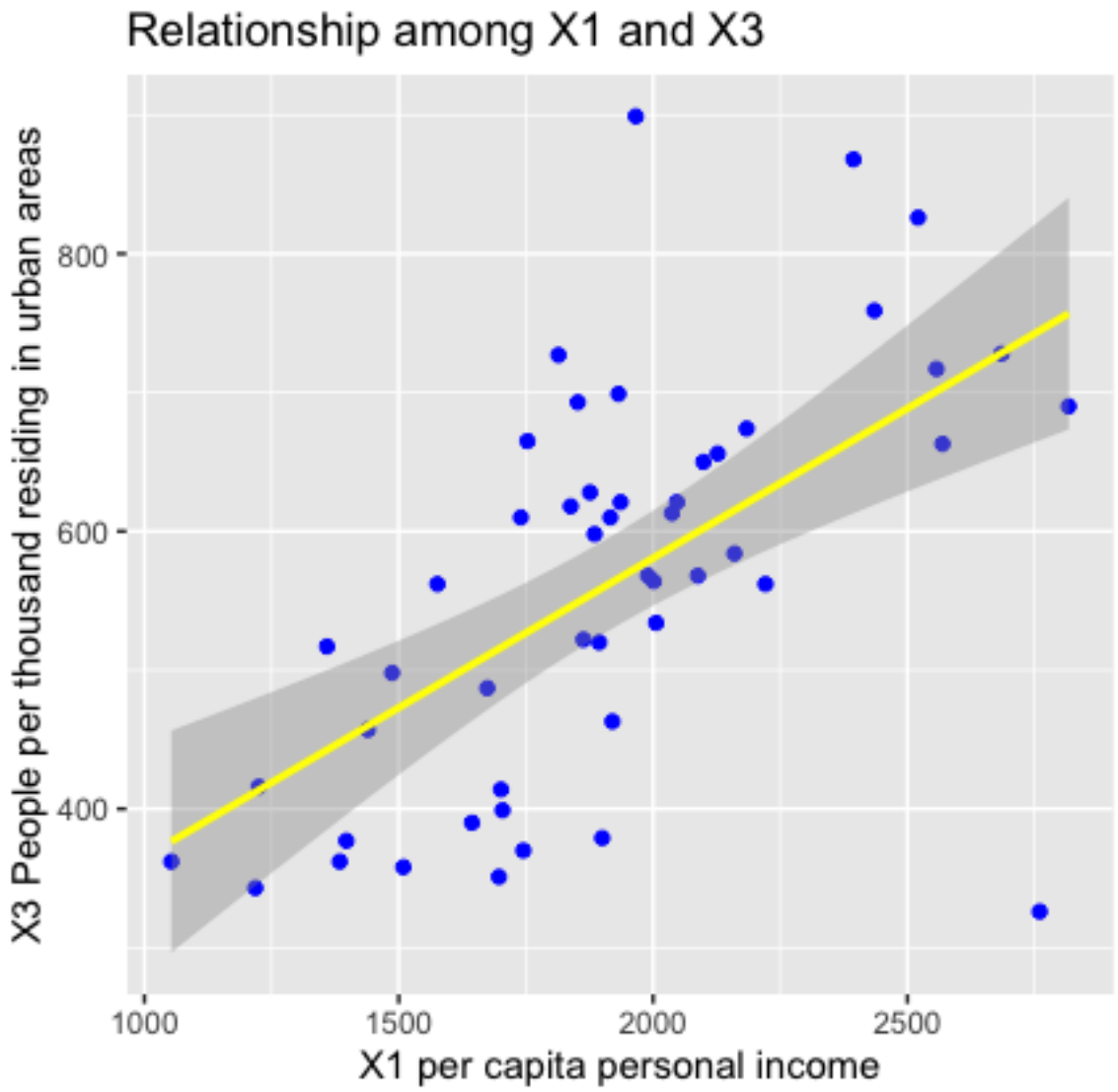


Interpretation:

Plot 4 displays a positive upward trend. This indicates a weak linear correlation between X1 (per capita personal income) and X2 (Residents per 100,000 that are financially insecure) in 50 states in US.
The slope also shows a positive relationship between X1 and X2. Nonetheless, the observations are spread.

Plot 5 Relationship X1, X3:

```
##Plot 5 relationship X1, X3
plot_X1_X3 <-ggplot(expenditure, aes(x = X1, y = X3)) + geom_point(color
    = "blue") +
  labs(x = "X1 per capita personal income", y = "X3 People per thousand
    residing in urban areas",
       title = "Relationship among X1 and X3")  +
```

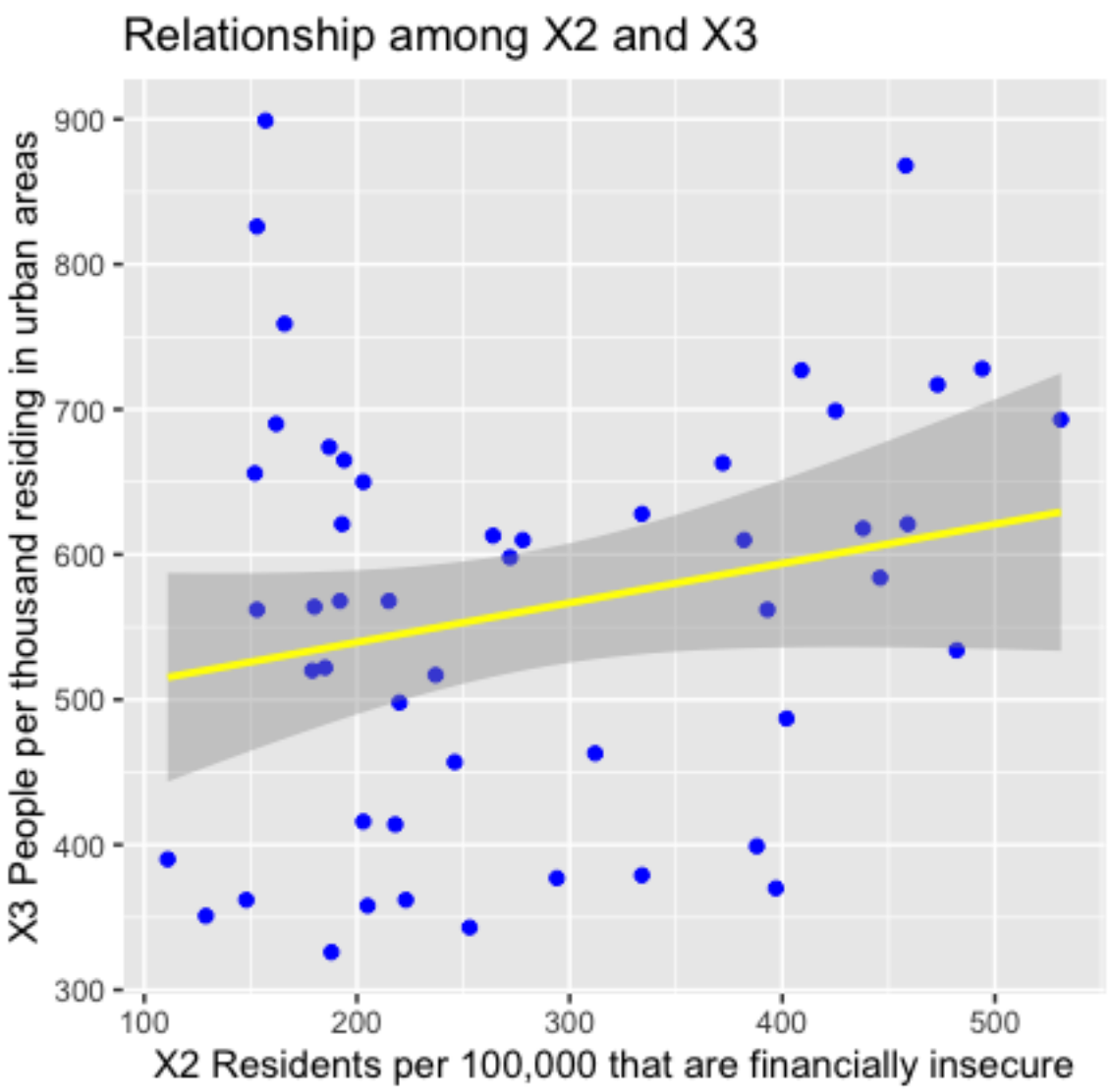


Interpretation::
Plot 5 displays an upward trend. This indicates a strong linear correlation :
between X1 (per capita personal income) and X3 (People per thousand residing in urban areas) in 50 states in US.:
The slope indicates a positive relationship between X1 and X3. The higher the per capita personal income:
the higher is the number of people residing in urban areas.:

Plot 6 relationship X2, X3

```
##Plot 6 relationship X2, X3
plot_X2_X3 <- ggplot(expenditure, aes(x = X2, y = X3)) + geom_point(color
    = "blue") +
  labs(x = "X2 Residents per 100,000 that are financially insecure", y =
```

```
    "X3 People per thousand residing in urban areas",
4       title = "Relationship among X2 and X3") +
```

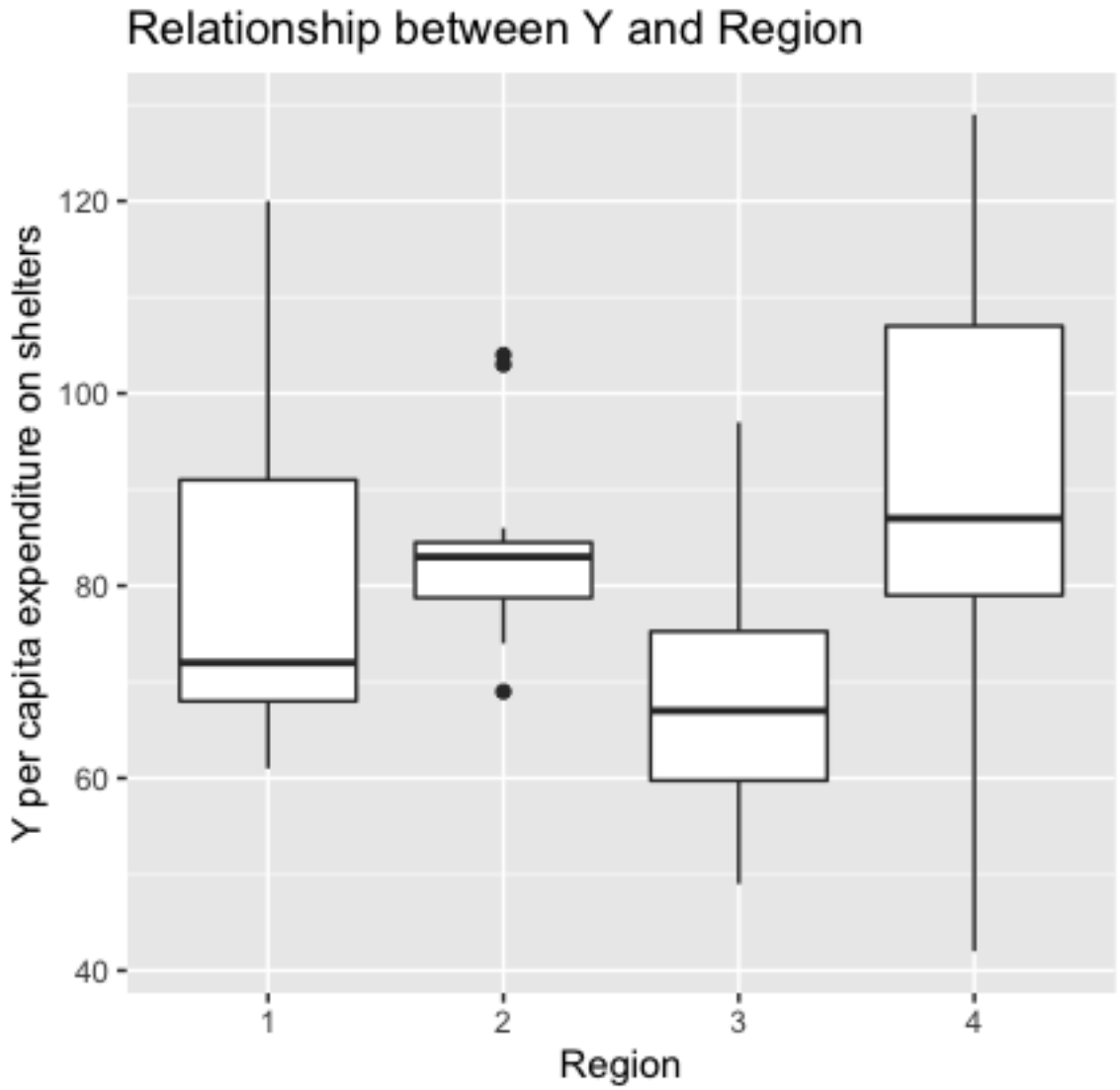


Interpretation:
Plot 6 displays a upward trend. This indicates a positive linear correlation between X2 (Residents per 100,000 that are financially insecure) and X3 (People per thousand residing in urban areas) in 50 states in US. The slope indicates a positive relationship between X2 and X3. However, the relationship is weak and the observations are spread.

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

- 2.2. Plot Relationship between Y and Region

```
plot_Y_Region <- ggplot(expenditure, aes(y = Y, x = as.factor(Region))) +
  geom_boxplot() +
  labs(x = "Region", y = "Y per capita expenditure on shelters",
```



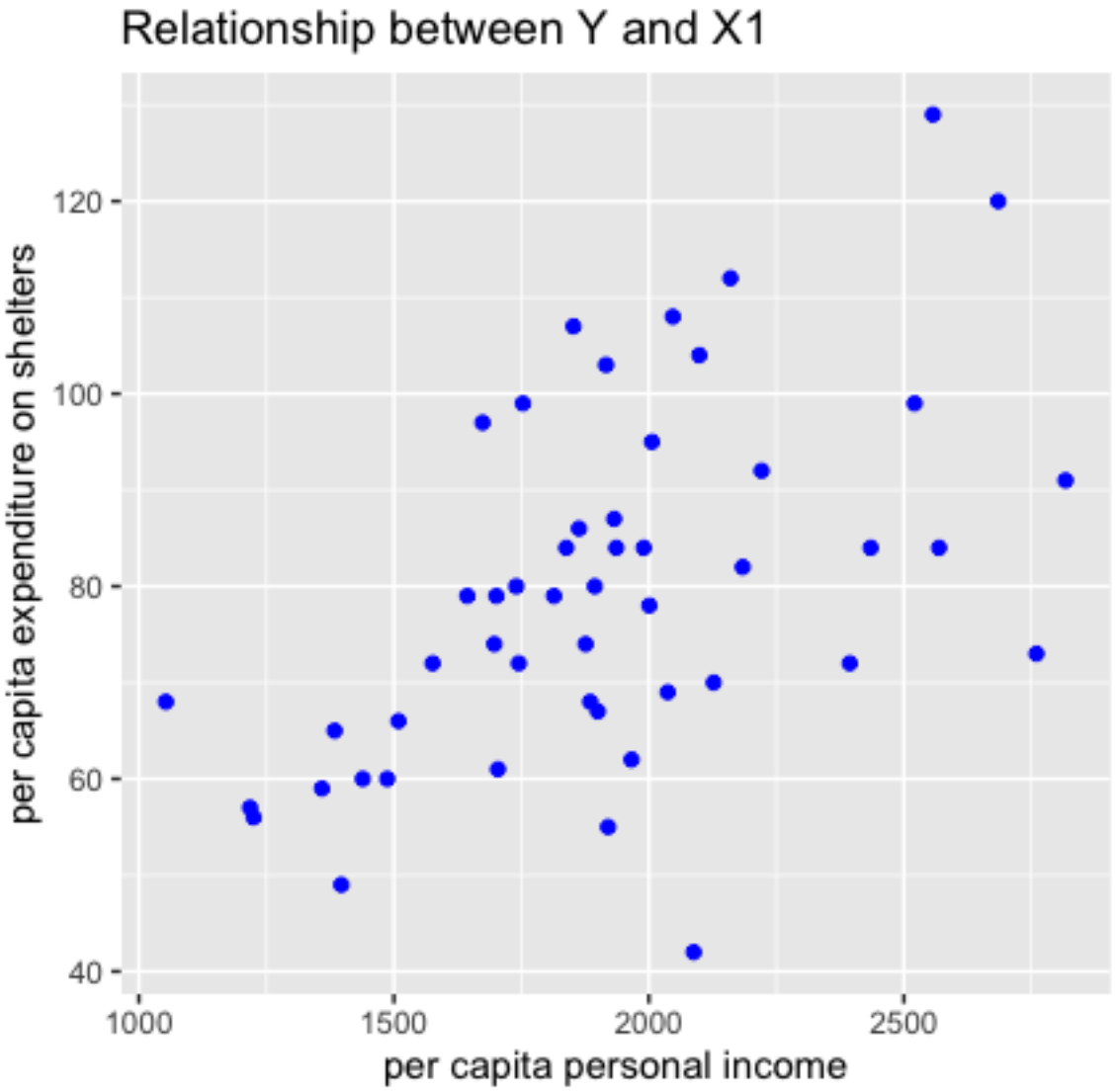
Answer:
On average the West Region has the highest per capita expenditure on housing assistance.

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.

2.3. Plot Relationship between Y and X1

```
1
2 plot_Y_X1 <- ggplot(expenditure, aes(x = X1, y = Y)) + geom_point(color = "
    blue") +
3   labs(x = "per capita personal income", y = "per capita expenditure on
    shelters",
```



Interpretation:
The plot depicts a positive relationship between the variables Y (per capita expenditure on shelters) and X1 (per capita personal income).
When the value on per capita personal income increases, so does the value on per capita

expenditure on housing assistance in 50 States in US.

Plot Relationship between Y and X1 by Region

```
1
2 plot_Y_X1_Region <- ggplot(expenditure) + geom_point(aes(x = X1, y = Y, color=
    as.factor(Region), shape=as.factor(Region))) +
3   labs(x = "X1 per capita personal income", y = "Y per capita expenditure on
    shelters",
4        title = "Relationship between Y and X1 by Region") + scale_shape_manual
    (values=c(3, 16, 17, 18))+
5   scale_color_manual(values=c("#999999","#E69F00", "#56B4E9", "#FC4E07"))+
```