

# Problem Set 1

## 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:

```
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 #Create Object Y  
2 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)  
3
```

```

4 #Rename Y to student_IQ
5 student_IQ <- y
6
7 #Find mean of Student IQs
8 mean_student_IQ <- mean(student_IQ)
9 print(mean_student_IQ)
10
11 #Set up Confidence Interval
12 z90 <- qnorm((1 - .90)/2, lower.tail = FALSE)
13 n <- length(na.omit(student_IQ))
14 mean_student_IQ <- mean(student_IQ, na.rm = TRUE)
15 SD_student_IQ <- sd(student_IQ, na.rm = TRUE)
16 lower_90 <- mean_student_IQ - (z90 * (SD_student_IQ/sqrt(n)))
17 upper_90 <- mean_student_IQ + (z90 * (SD_student_IQ/sqrt(n)))
18
19 confint90 <- c(lower_90, upper_90)
20
21 print(confint90)

```

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  $\alpha = 0.05$ .
3. Set up our null hypothesis H0: Our sample mean is less than or equal to the population mean

```

1 #Find our variables
2 str(student_IQ)
3
4 #Formulate our null hypothesis – NULL HYPOTHESIS: The average student IQ
  is less than
5 #or equal to 100
6
7 IQ_null <- t.test(student_IQ, mu = 100)
8 IQ_null

```

## 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.

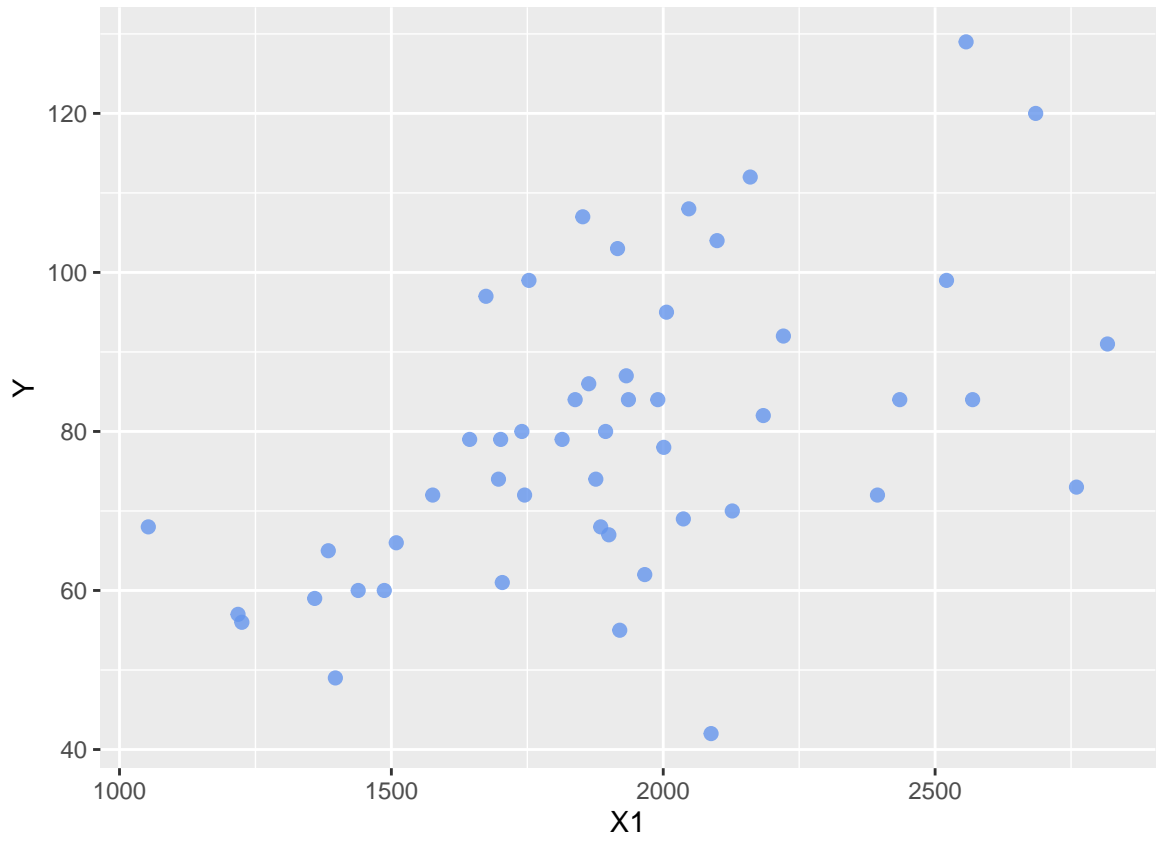
- Please plot the relationships among  $Y$ ,  $X1$ ,  $X2$ , and  $X3$ ? What are the correlations among them (you just need to describe the graph and the relationships among them)?

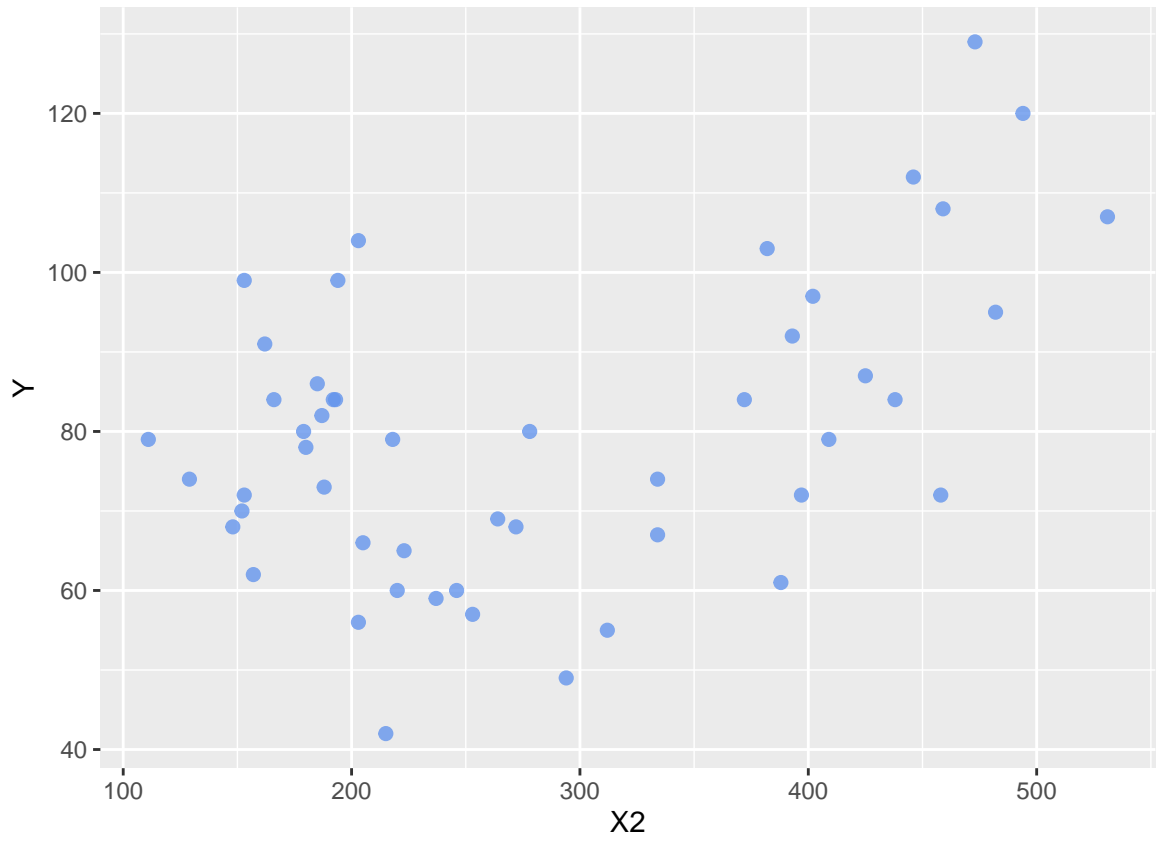
```
1      \tabsize=2
2      }
3      \lstset{style=mystyle}
4      \newcommand{\Sref}[1]{Section~\ref{#1}}
5      \newtheorem{hyp}{Hypothesis}
6
7      \title{Problem Set 1}
8      \date{Due: October 1, 2021}
9      \author{Applied Stats/Quant Methods 1}
10
11     \begin{document}
12       \maketitle
13
14       \section*{Instructions}
15       \begin{itemize}
16         \item Please show your work! You may lose points by simply writing
in the answer. If the problem requires you to execute commands in \
texttt{R}, please include the code you used to get your answers.
Please also include the \texttt{.R} file that contains your code. If
you are not sure if work needs to be shown for a particular problem,
please ask.
17         \item Your homework should be submitted electronically on GitHub in
\texttt{.pdf} form.
18         \item This problem set is due before 8:00 on Friday October 1,
2021. No late assignments will be accepted.
```

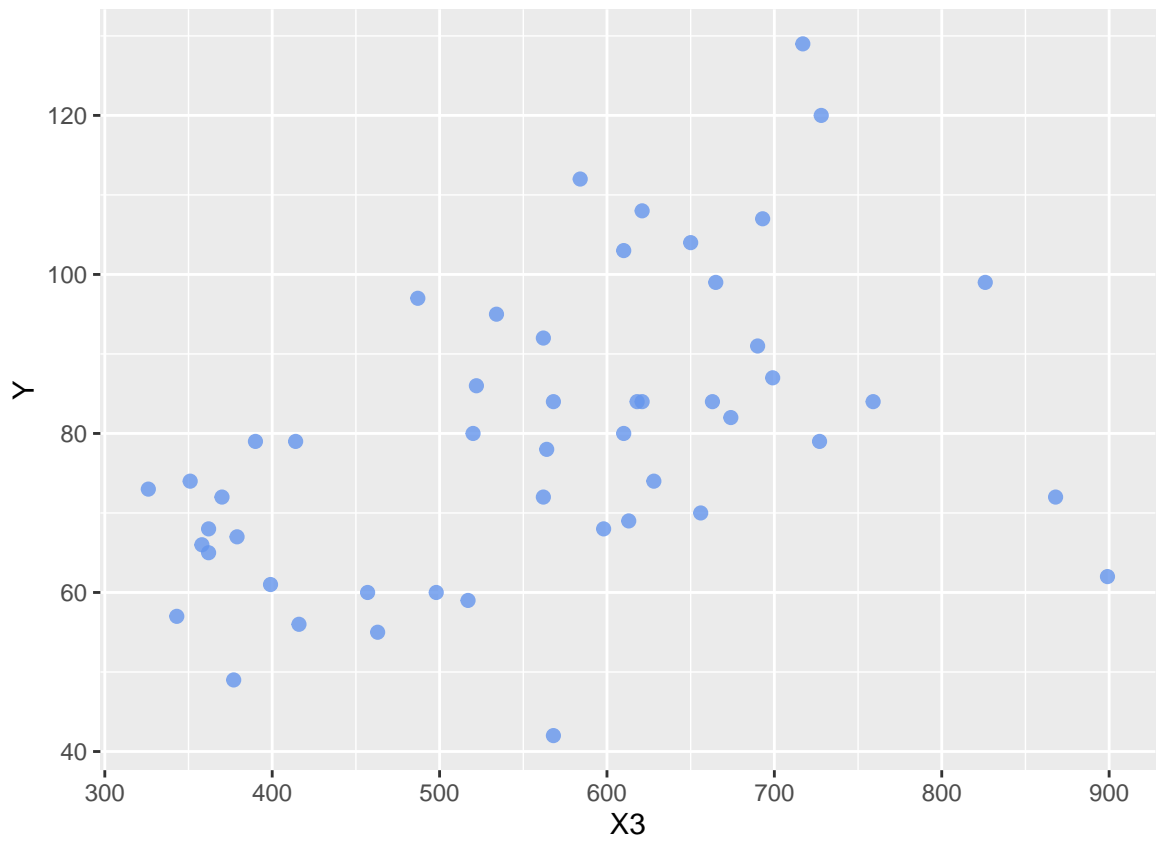
```

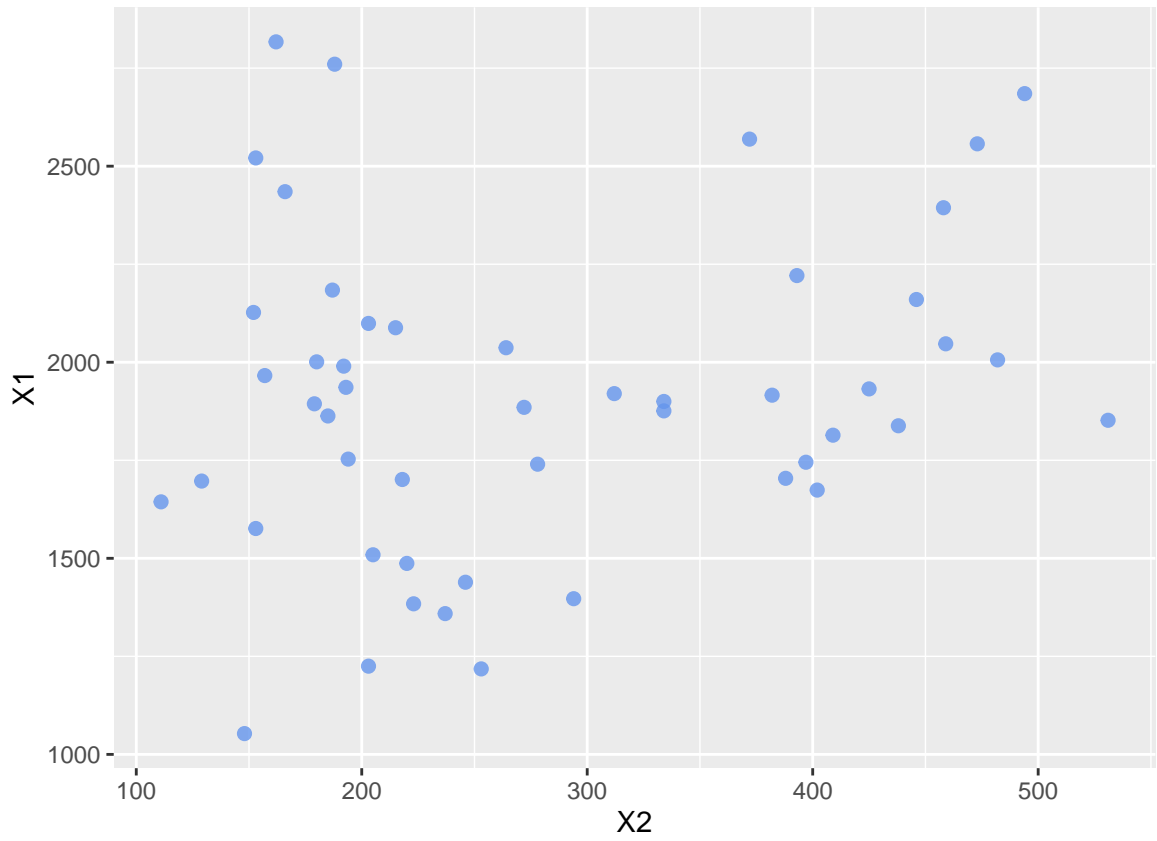
19   \item Total available points for this homework is 100.
20   \end{itemize}
21
22   \vspace{1cm}
23   \section*{Question 1 (50 points): Education}
24
25   A school counselor was curious about the average of IQ of the
26   students in her school and took a random sample of 25 students' IQ
27   scores. The following is the data set:\\
28
29   \vspace{1cm}
30   \begin{enumerate}
31     \item Find a 90\% confidence interval for the average student IQ in
32     the school.\\
33     \vspace{.5cm}
34     \begin{lstlisting}
35       #Create Object Y
36       y <- c(105, 69, 86, 100, 82, 111, 104, 110, 87, 108, 87, 90, 94,
37       113, 112, 98, 80, 97, 95, 111, 114, 89, 95, 126, 98)
38
39       #Rename Y to student_IQ
40       student_IQ <- y
41
42       #Finf mean of Student IQs
43       mean_student_IQ <- mean(student_IQ)
44       print(mean_student_IQ)
45
46       #Set up Confidence Interval
47       z90 <- qnorm((1 - .90)/2, lower.tail = FALSE)
48       n <- length(na.omit(student_IQ))
49       mean_student_IQ <- mean(student_IQ, na.rm = TRUE)
50       SD_student_IQ <- sd(student_IQ, na.rm = TRUE)
51       lower_90 <- mean_student_IQ - (z90 * (SD_student_IQ/sqrt(n)))

```

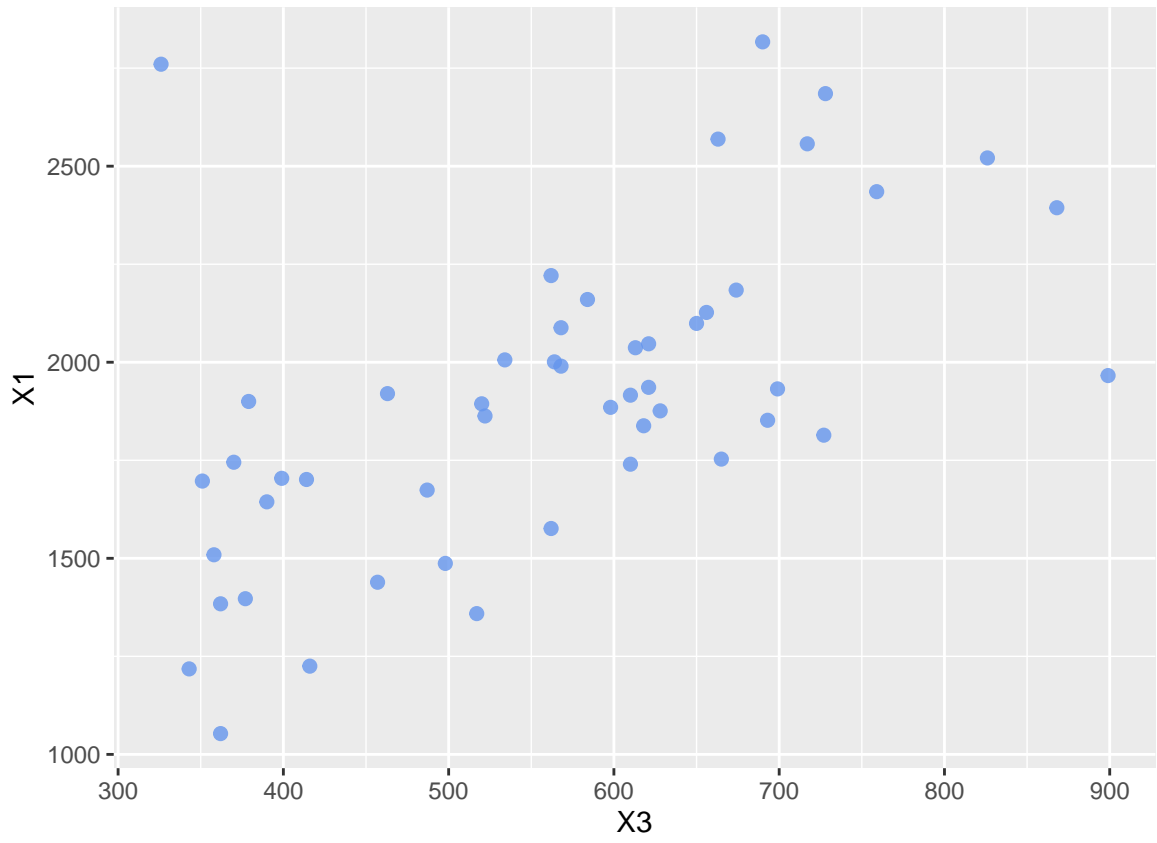


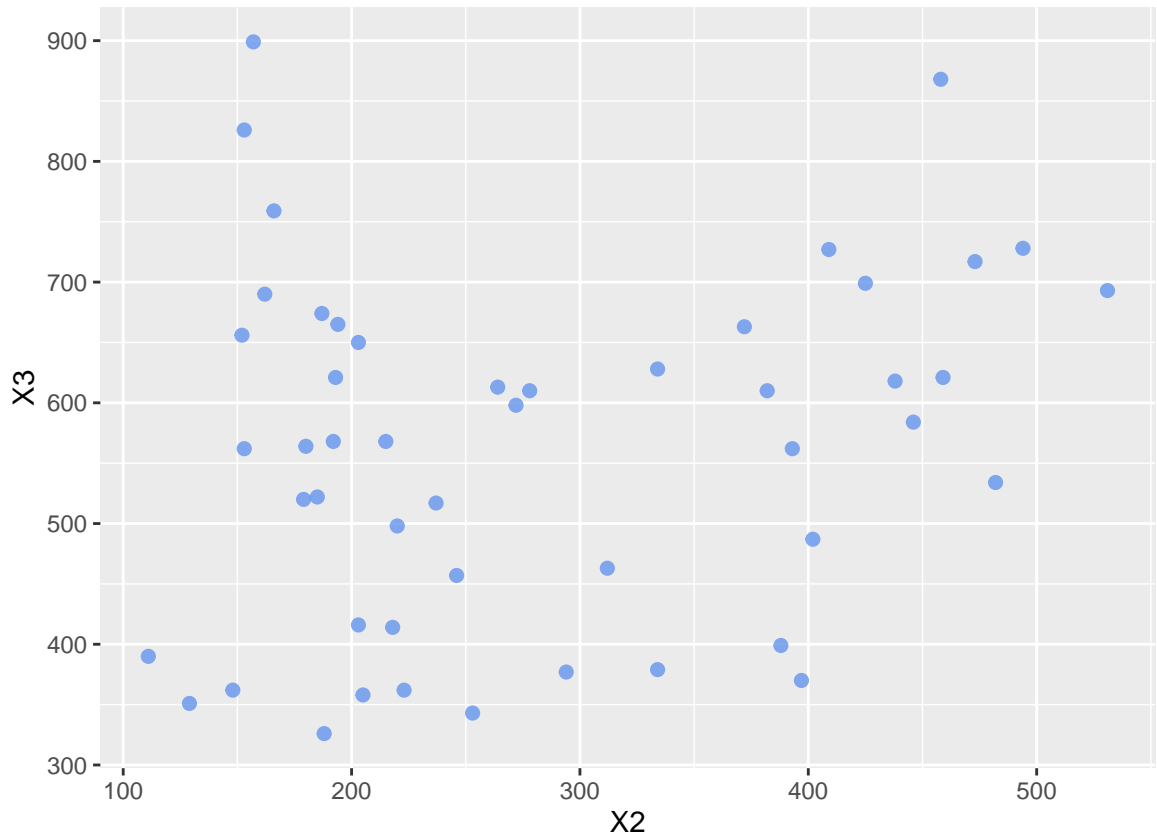












- Please plot the relationship between  $Y$  and *Region*? On average, which region has the highest per capita expenditure on housing assistance?
- Region 4 has the highest per capita expenditure on housing assistance, as illustrated in the below graph.

```

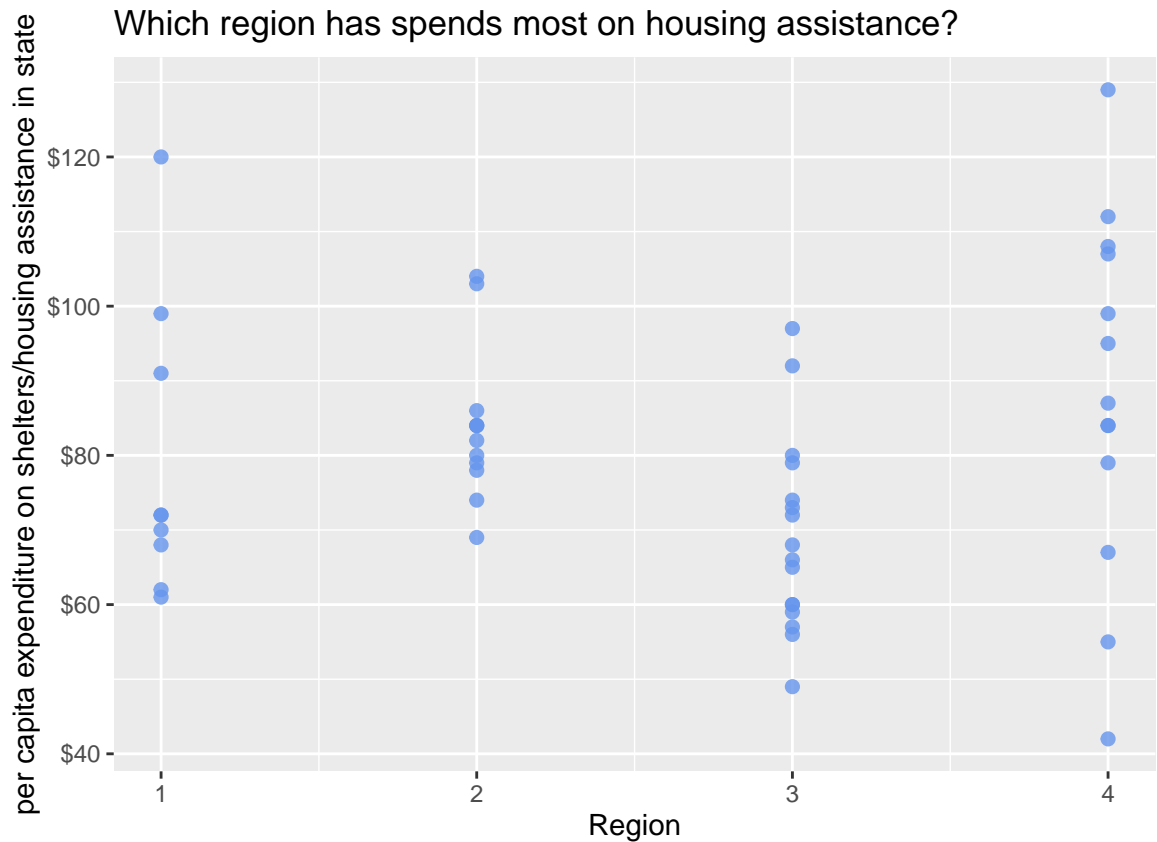
1      confint90 <- c(lower_90, upper_90)
2      print(confint90)content...
3
4      # 94.13283      102.74717
5      \end{lstlisting}
6      \vspace{1cm}
7
8      \item Next, the school counselor was curious whether the average
9      student IQ in her school is higher than the average IQ score (100)
10     among all the schools in the country.\\
11
12     \noindent Using the same sample, conduct the appropriate hypothesis
13     test with  $\alpha=0.05$ .
14     \vspace{.5cm}

```

```

12 \begin{lstlisting}
13   #Find our variables
14   str(student_IQ)
15
16

```



- 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   #or equal to 100
2
3   IQ_null <- t.test(student_IQ, mu = 100)
4   IQ_null
5
6   print(IQ_null)
7   One Sample t-test
8
9   data:  student_IQ

```

```

10      t = -0.59574, df = 24, p-value = 0.5569
11      alternative hypothesis: true mean is not equal to 100
12      95 percent confidence interval:
13      93.03553 103.84447
14      sample estimates:
15      mean of x
16      98.44
17
18      \end{lstlisting}
19      \end{enumerate}

```

