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:

```
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.

```
\usepackage[svgnames]{xcolor}
\usepackage{listings}

\lstset{language=R,
```

```
  basicstyle=\small\ttfamily,
  stringstyle=\color{DarkGreen},
  otherkeywords={0,1,2,3,4,5,6,7,8,9},
  morekeywords={TRUE,FALSE},
  deletekeywords={data,frame,length,as,character},
  keywordstyle=\color{blue},
  commentstyle=\color{DarkGreen},
\lstset{language=R,
  basicstyle=\small\ttfamily,
  stringstyle=\color{DarkGreen},
  otherkeywords={0,1,2,3,4,5,6,7,8,9},
  morekeywords={TRUE,FALSE},
  deletekeywords={data,frame,length,as,character},
  keywordstyle=\color{blue},
  commentstyle=\color{DarkGreen},
  \definecolor{codegreen}{rgb}{0,0.6,0}
  \definecolor{codegray}{rgb}{0.5,0.5,0.5}
```

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 $\alpha = 0.05$.

```

  \lstdefinestyle{mystyle}{
    backgroundcolor=\color{backcolour},
    commentstyle=\color{codegreen},
    keywordstyle=\color{magenta},
    numberstyle=\tiny\color{codegray},
    stringstyle=\color{codepurple},
    basicstyle=\footnotesize,
```

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.

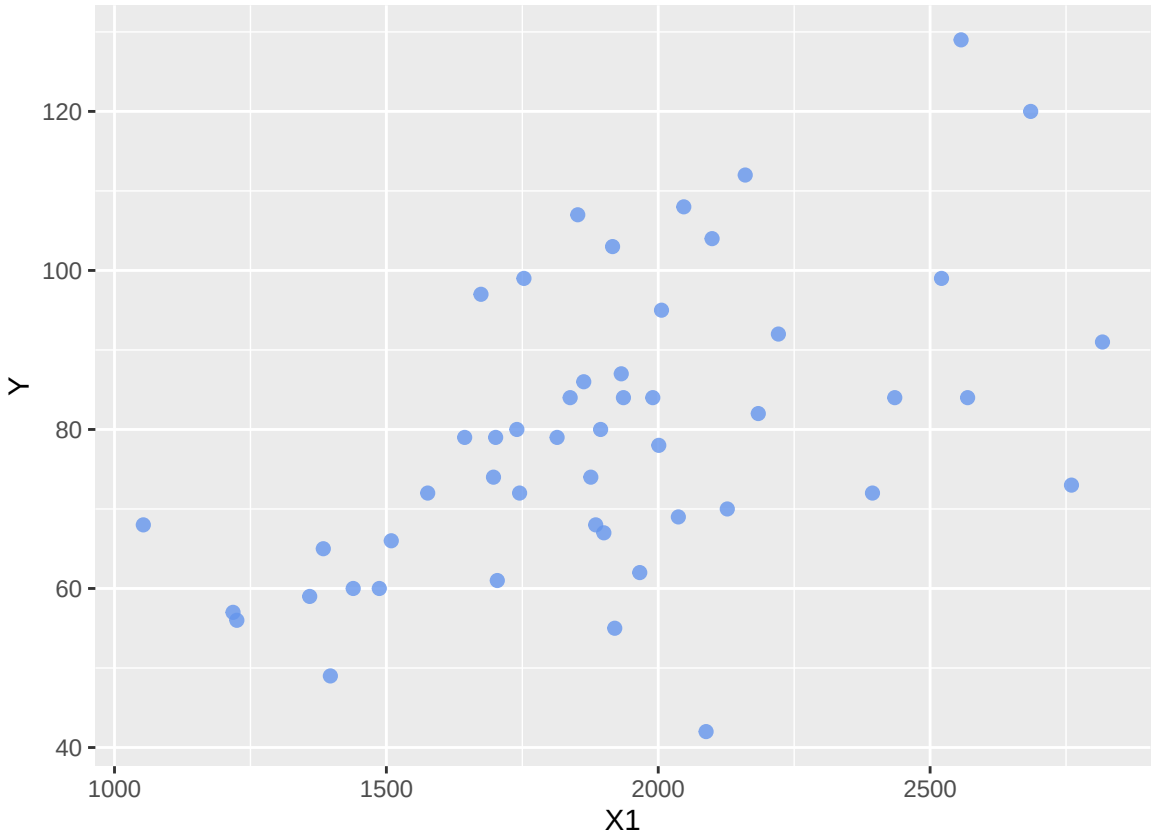
`State`	*50 states in US*
`Y`	*per capita expenditure on shelters/housing assistance in state*
`X1`	*per capita personal income in state*
`X2`	*Number of residents per 100,000 that are "financially insecure" in state*
`X3`	*Number of people per thousand residing in urban areas in state*
`Region`	*1=Northeast,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

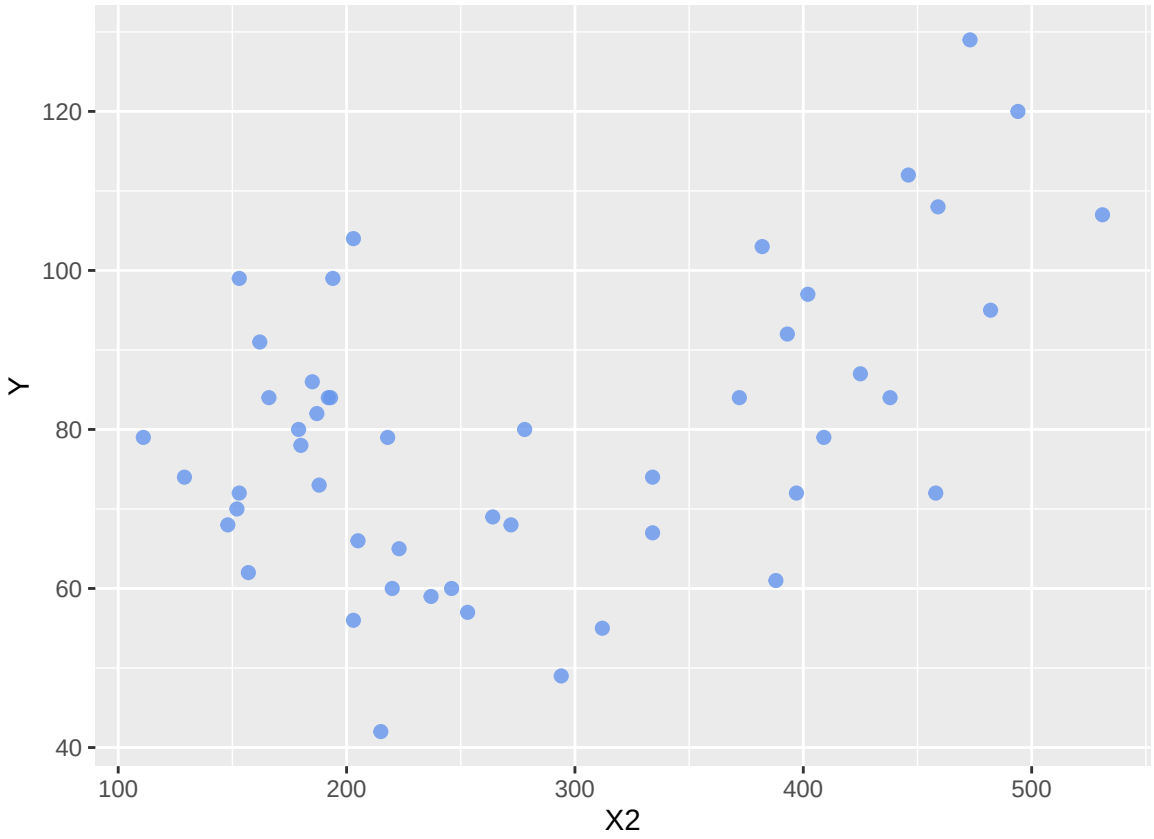
Explore the `expenditure` data set and import data into `R`.

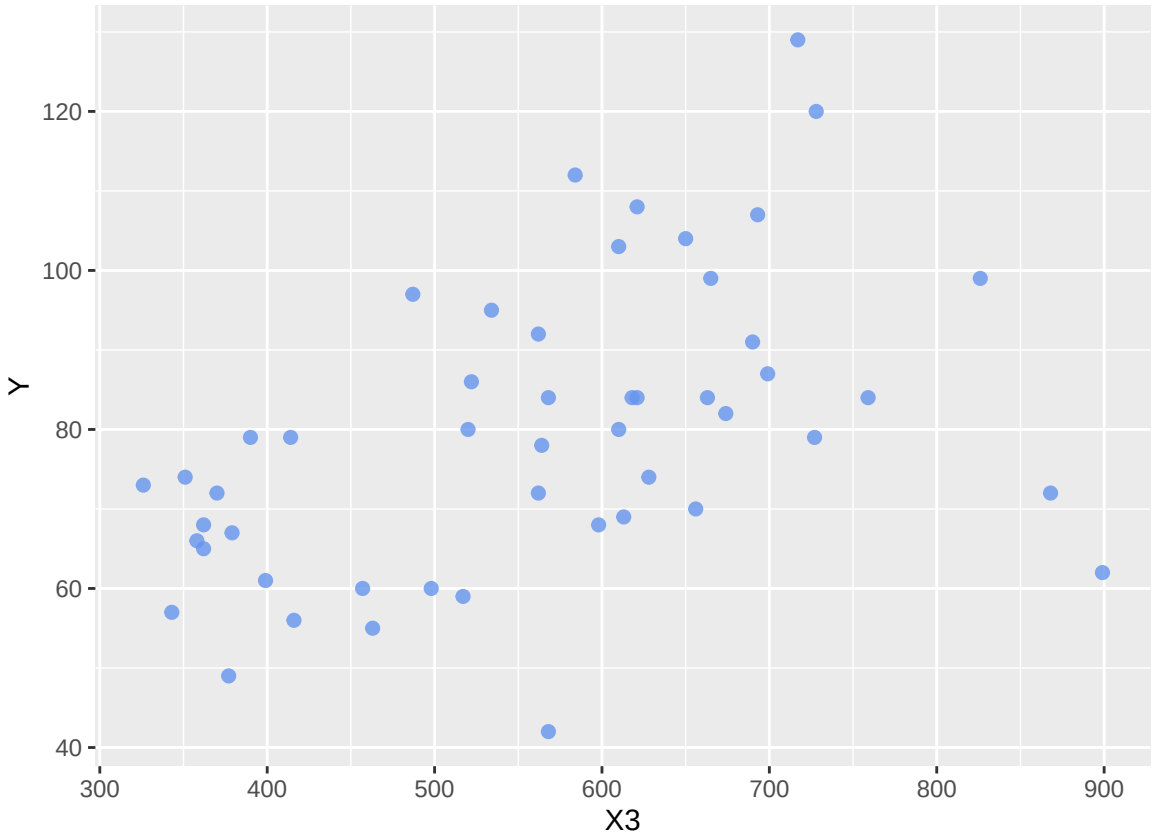
- Please plot the relationships among *Y*, *X1*, *X2*, and *X3*? What are the correlations among them (you just need to describe the graph and the relationships among them)?

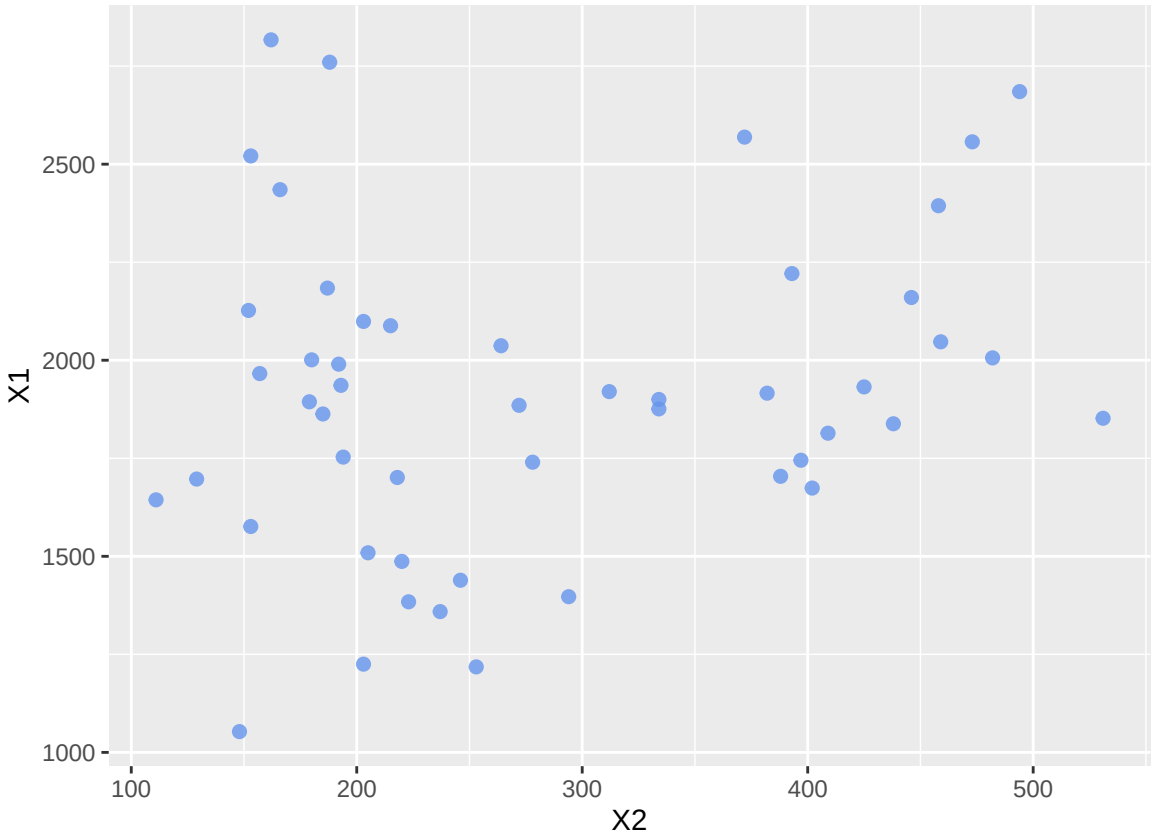
```
    tabsize=2
}
\lstset{style=mystyle}
\newcommand{\Sref}[1]{Section~\ref{#1}}
\newtheorem{hyp}{Hypothesis}

\title{Problem Set 1}
\date{Due: October 1, 2021}
\author{Applied Stats/Quant Methods 1}

\begin{document}
  \maketitle

  \section*{Instructions}
  \begin{itemize}
    \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.
    \item Your homework should be submitted electronically on GitHub in \texttt{.pdf} form.
    \item This problem set is due before 8:00 on Friday October 1, 2021. No late assignments will be accepted.
```

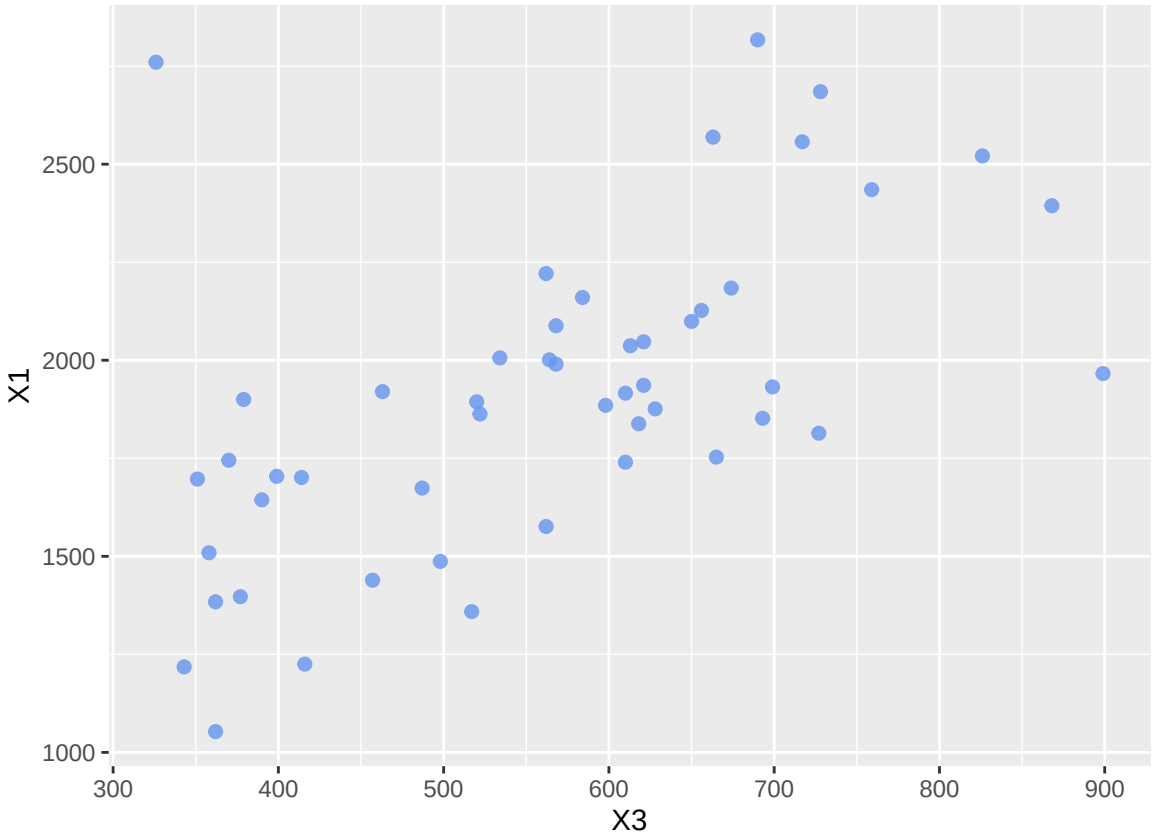
```
    \item Total available points for this homework is 100.
  \end{itemize}

  \vspace{1cm}
  \section*{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:\\

  \vspace{1cm}

  \begin{enumerate}
    \item Find a 90\% confidence interval for the average student IQ in
   the school.\\
    \vspace{.5cm}
    \begin{lstlisting}
      #Create Object Y
      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)

      #Rename Y to student_IQ
      student_IQ <- y

      #Finf mean of Student IQs
      mean_student_IQ <- mean(student_IQ)
      print(mean_student_IQ)

      #Set up Confidence Interval
      z90 <- qnorm((1 - .90)/2, lower.tail = FALSE)
      n <- length(na.omit(student_IQ))
      mean_student_IQ <- mean(student_IQ, na.rm = TRUE)
      SD_student_IQ <- sd(student_IQ, na.rm = TRUE)
      lower_90 <- mean_student_IQ - (z90 * (SD_student_IQ/sqrt(n)))
```

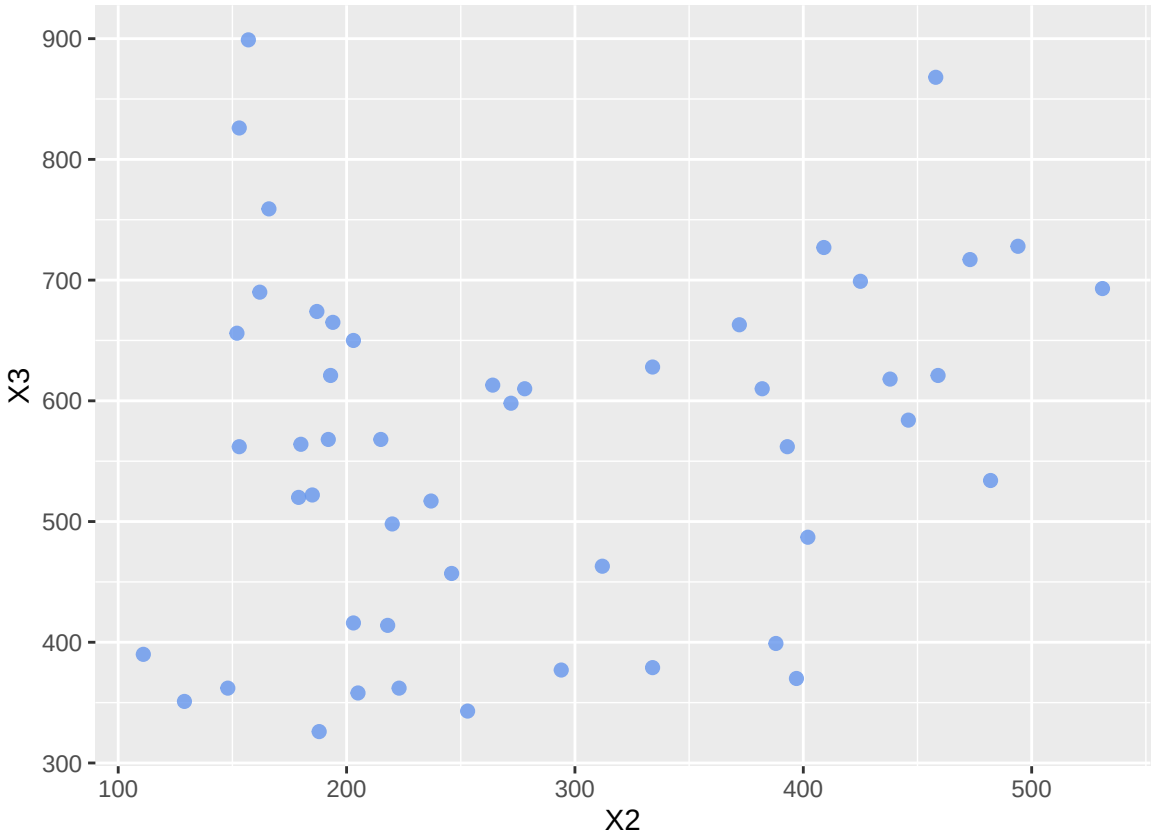

120
100
80
60
40
Y
1000
1500
2000
2500
X1


120
100
80
60
40
Y
100
200
300
400
500
X2


Y
120
100
80
60
40
300
400
500
600
700
800
900
X3


2500
2000
1500
1000
X1
100
200
300
400
500
X2

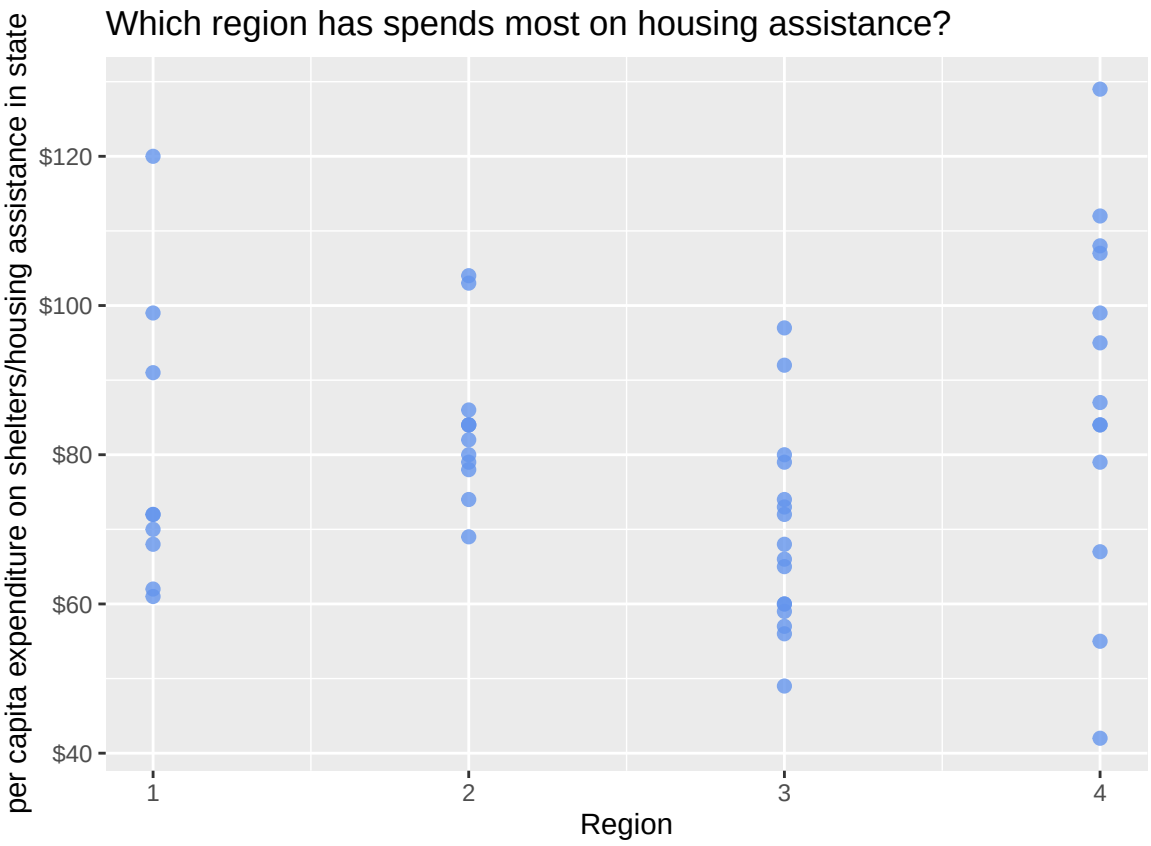

2500
2000
1500
1000
X1
300
400
500
600
700
800
900
X3



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

```
        confint90 <- c(lower_90, upper_90)
        print(confint90)content...

        # 94.13283       102.74717
    \end{lstlisting}
    \vspace{1cm}

    \item 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.\\

    \noindent Using the same sample, conduct the appropriate hypothesis test with $\alpha=0.05$.
    \vspace{.5cm}
    \begin{lstlisting}
      #Find our variables
      str(student_IQ)
```

```

```



- 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.

```
#or equal to 100

IQ_null <- t.test(student_IQ, mu = 100)
IQ_null

print(IQ_null)
One Sample t-test

data:  student_IQ
t = -0.59574, df = 24, p-value = 0.5569
alternative hypothesis: true mean is not equal to 100
95 percent confidence interval:
```

```
        93.03553 103.84447
        sample estimates:
        mean of x
        98.44

    \end{lstlisting}
  \end{enumerate}
```