Problem Set 1

Exercise 1

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

 # Question 1 - Part 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)

str(y) # sample size = 25 (t test needed)
mean(y) # sample mean = 98.44
sd(y) # sd of sample y = 13.09287
median(y) # median = 98
summary(y)

plot(density(y),
     main="Distribution of class IQ",
     xlab="IQ")

# confidence coefficient = .90 add and subtract
# df (degrees of freedom) for a t-test is n-1 = 25-1 = 24

error <- qt(0.95, df=24)*13.09/sqrt(24) # to find the standard error with 90% confidernce interval, df of 24, an sd of 13.09 and an 'n' of 25.
print(error) # error of 4.571451
interval_1 <- 98.44-error # interval to the left side of the mean
interval_2 <- 98.44+error # interval to the right side of the mean

interval_1 # 93.86855
interval_2 # 103.0155

# So the 90% confidence interval for the average student IQ in the school is 93.87 - 103
```

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 $\alpha = 0.05$.

Question 1 - Part 2

Null Hypothesis > Average school IQ is equal or less than 100
Alternative Hypothesis > Average school IQ is greater than 100
α = 0.05
#

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

data: y
#t = 37.593, df = 24, p-value < 2.2e-16
#alternative hypothesis: true mean is greater than 0
#95 percent confidence interval:
93.95993   Inf
#sample estimates:
mean of x
98.44

Results
p-value < 2.2e-16 which is not less than α (0.05), therefore I fail the reject the null hypothesis, so the average school IQ is not higher than the average IQ among all the schools in the country, but less than or equal to it.

Exercise 2

Chart info

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

Chart 1 (below): expenditure on shelters / housing assistance in state &
per capita expenditure on shelters/housing assistance in state
There seems to be a positive correlation between per capita personal
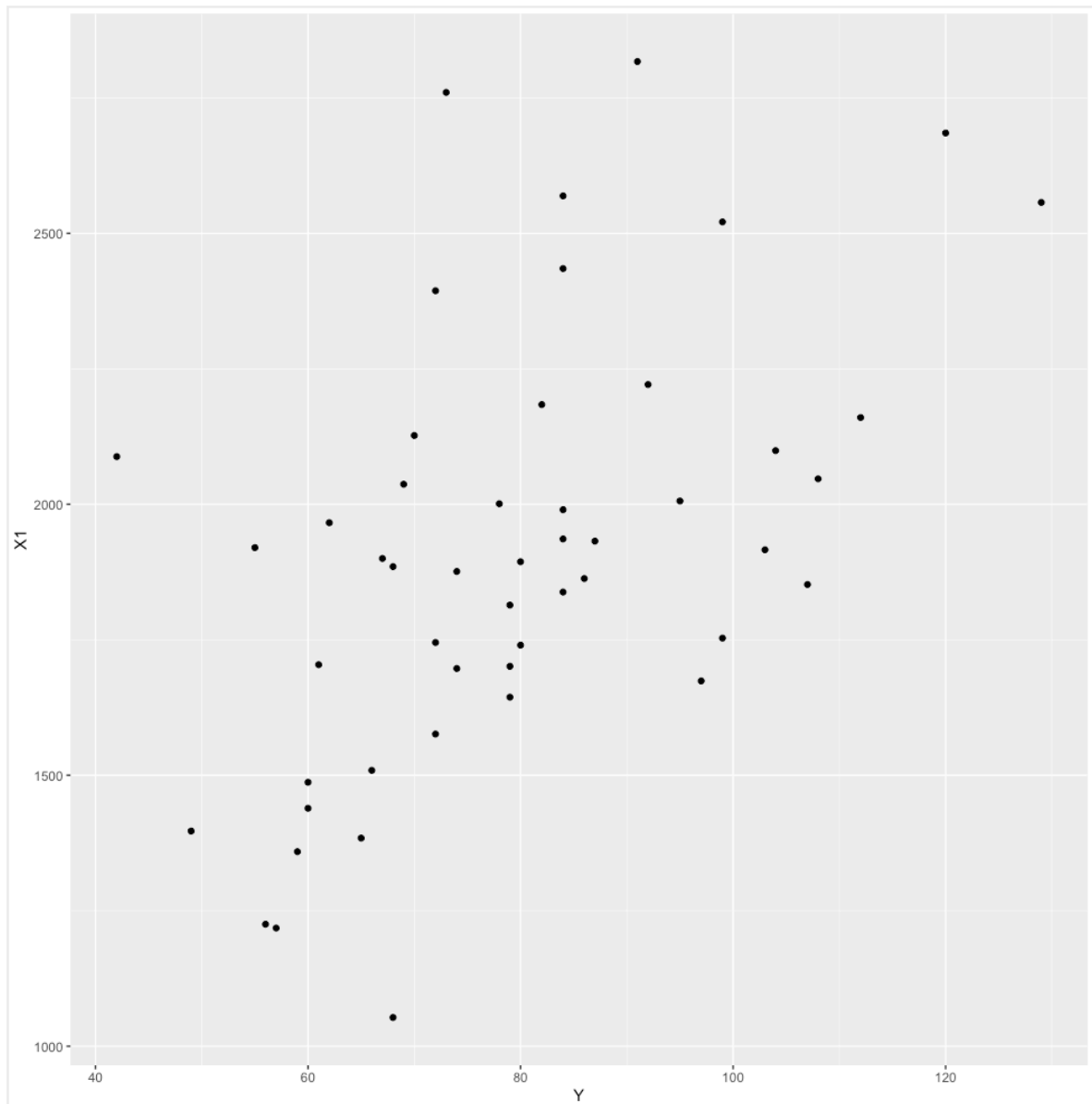income and expenditure on shelters / housing assistance in state

Chart 2 (below): expenditure on shelters / housing assistance in state & number of residents that are 'financially insecure'

There doesn't seem to be much of a relationship between the number of residents that are 'financially insecure' and expenditure on shelters / housing assistance in state.
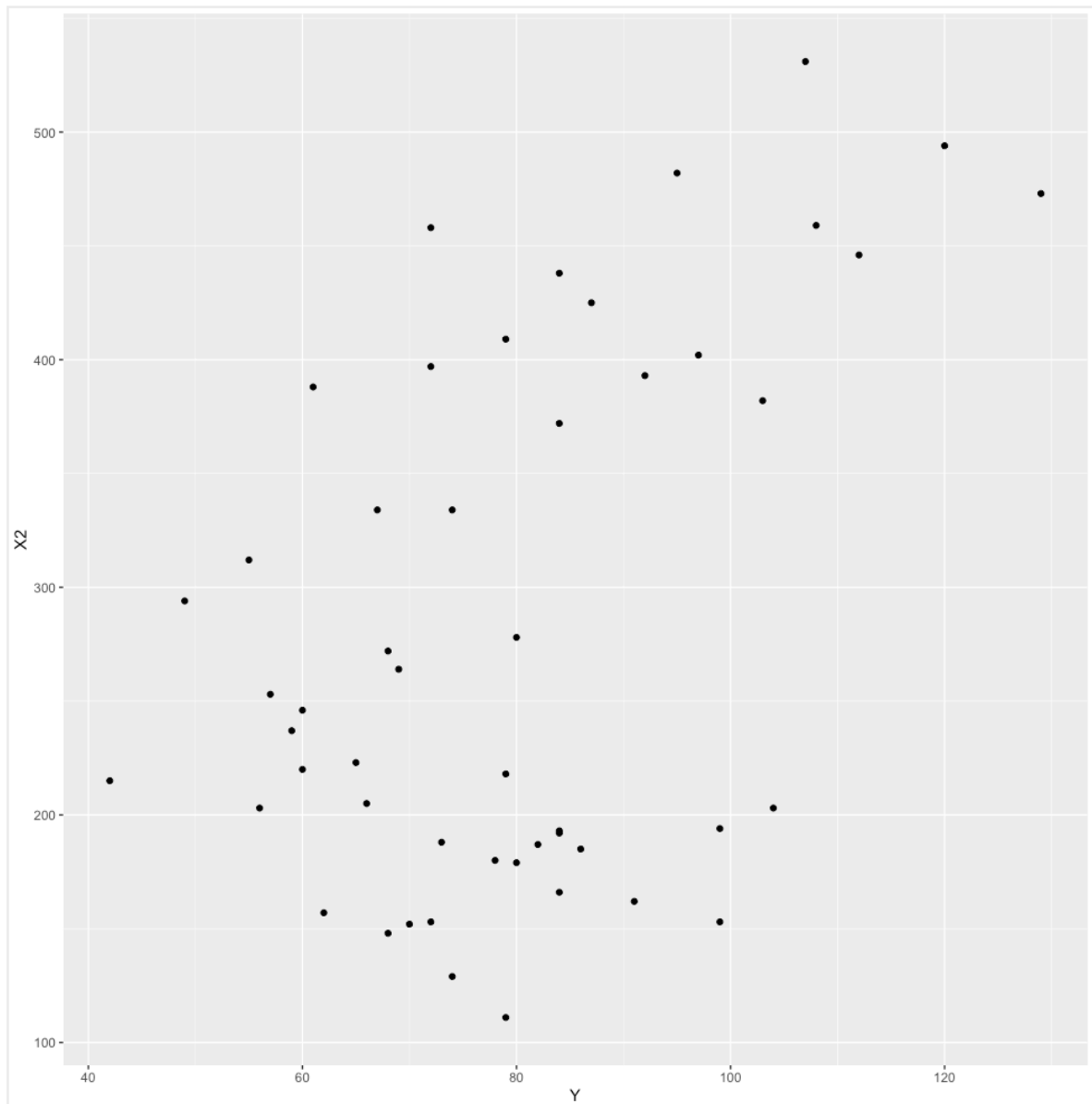
Chart 3 (below): number of people in thousands residing in urban areas in state & expenditure on shelters / housing assistance in state
There seems to be a positive correlation between the number of people residing in urban areas in state and expenditure on shelters / housing assistance in state.
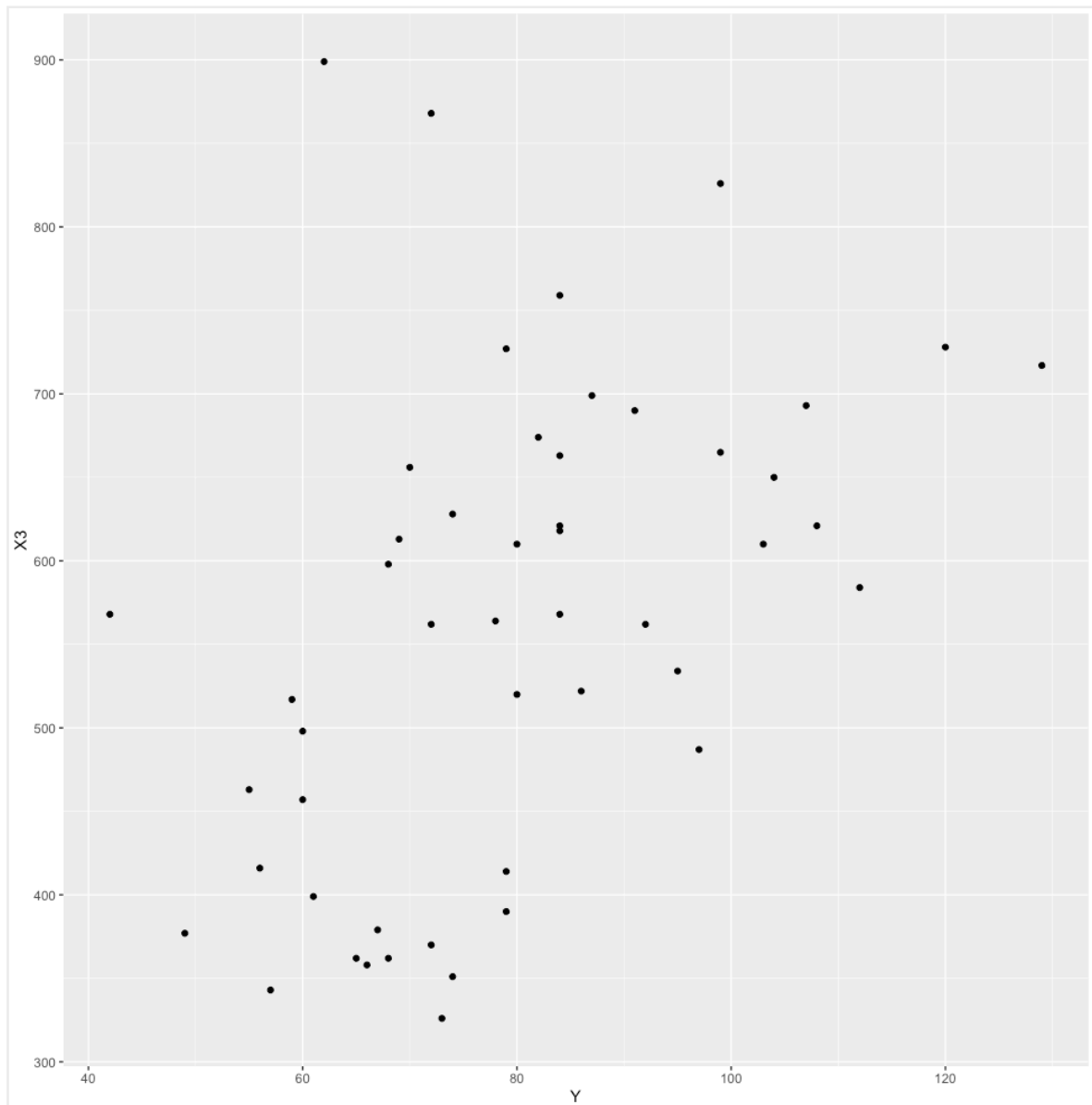
Chart 4 (below): per capita personal income in state & number of residents per 100,000 that are "financially insecure" in state.
There doesn't seem to be a correlation between per capita personal income in state and number of residents per 100,000 that are "financially insecure" in state.
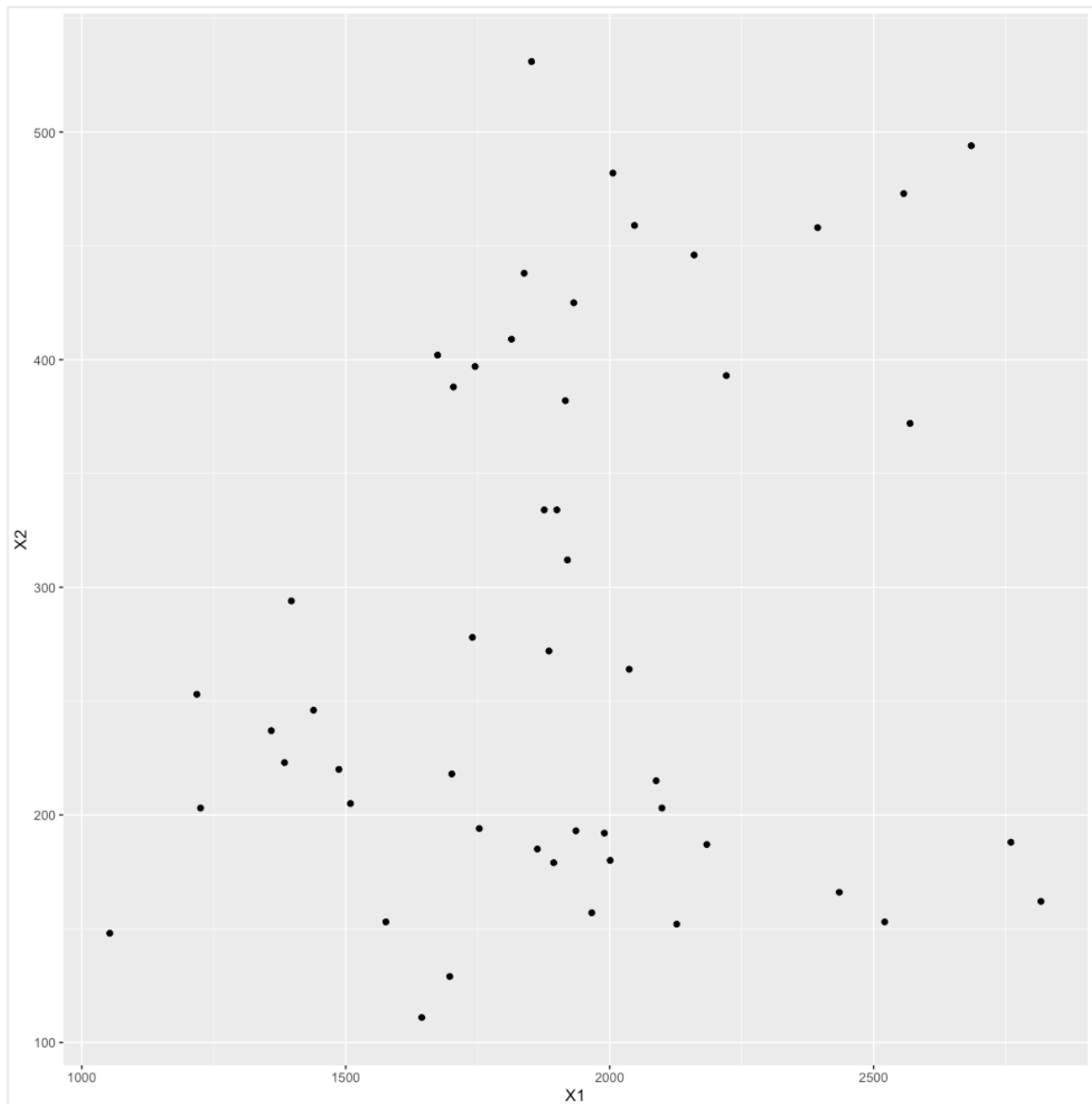
Chart 5 (below): per capita personal income in state & number of people per thousand residing in urban areas in state.
There seems to be a positive correlation between per capita personal income in state and the number of people per thousand residing in urban areas in state.
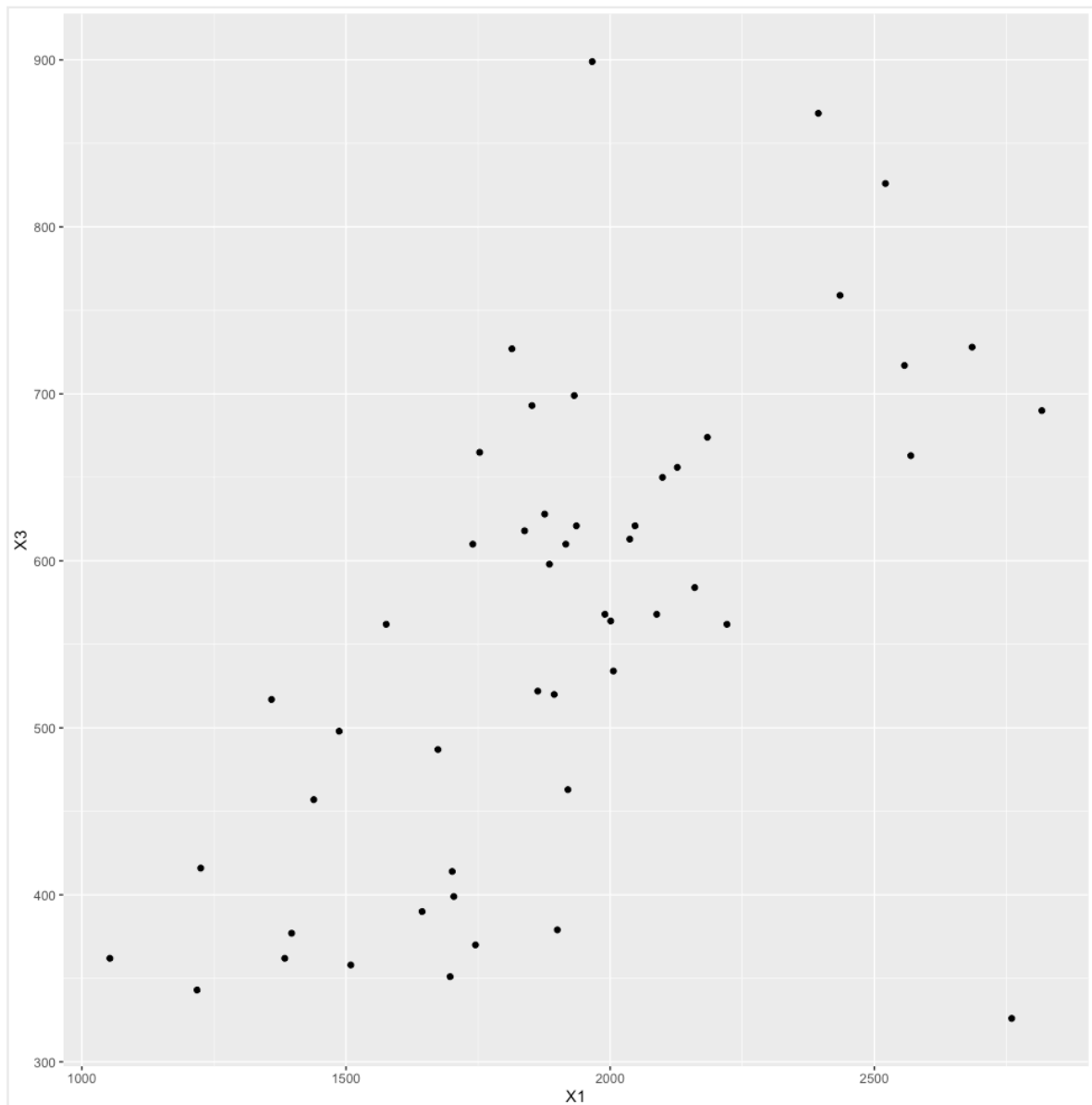
Chart 6 (below): Number of residents per 100,000 that are "financially insecure" in state & Number of people per thousand residing in urban areas in state.

These doesn't seem to be a correlation between the number of residents per 100,000 that are "financially insecure" in state & Number of people per thousand residing in urban areas in state.
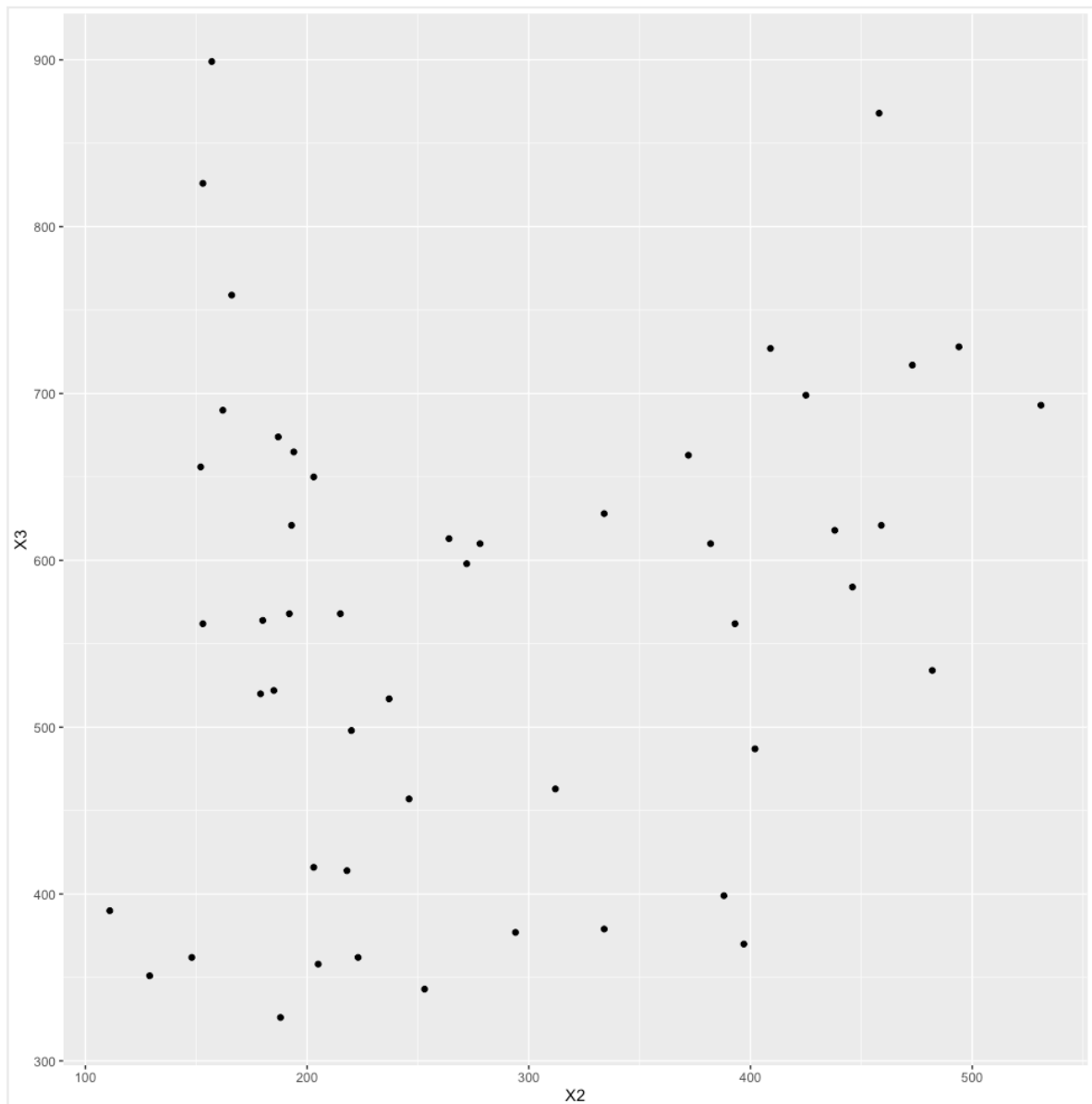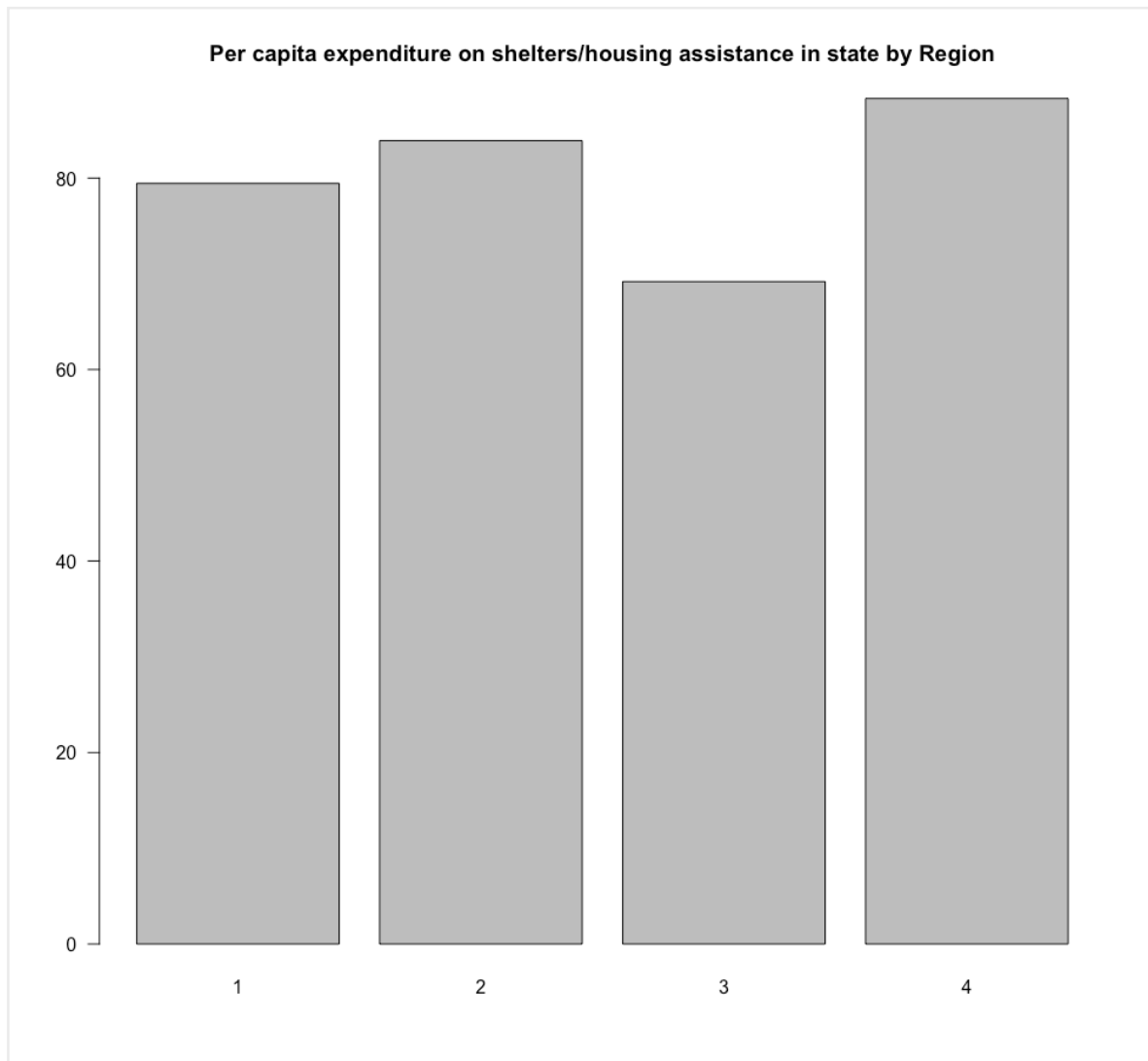
Chart 7 (below): Per capita expenditure on shelters / housing assistance in state by Region
From the bar chart it seems that region 4 spends the most per capita on shelters / housing assistance.

**Per capita expenditure on shelters/housing assistance in state by Region**

Q - 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.

ggplot(data = expenditure) + geom_point(mapping = aes(x = Y, y = X1)) #relationship between Y and X1 (again)

There seems to be a positive correlation between per capita expenditure on shelters/housing assistance in state and per capita personal income in state

ggplot(data = expenditure, mapping = aes(x = Y, y = X1)) + geom_point(aes(color = Region))
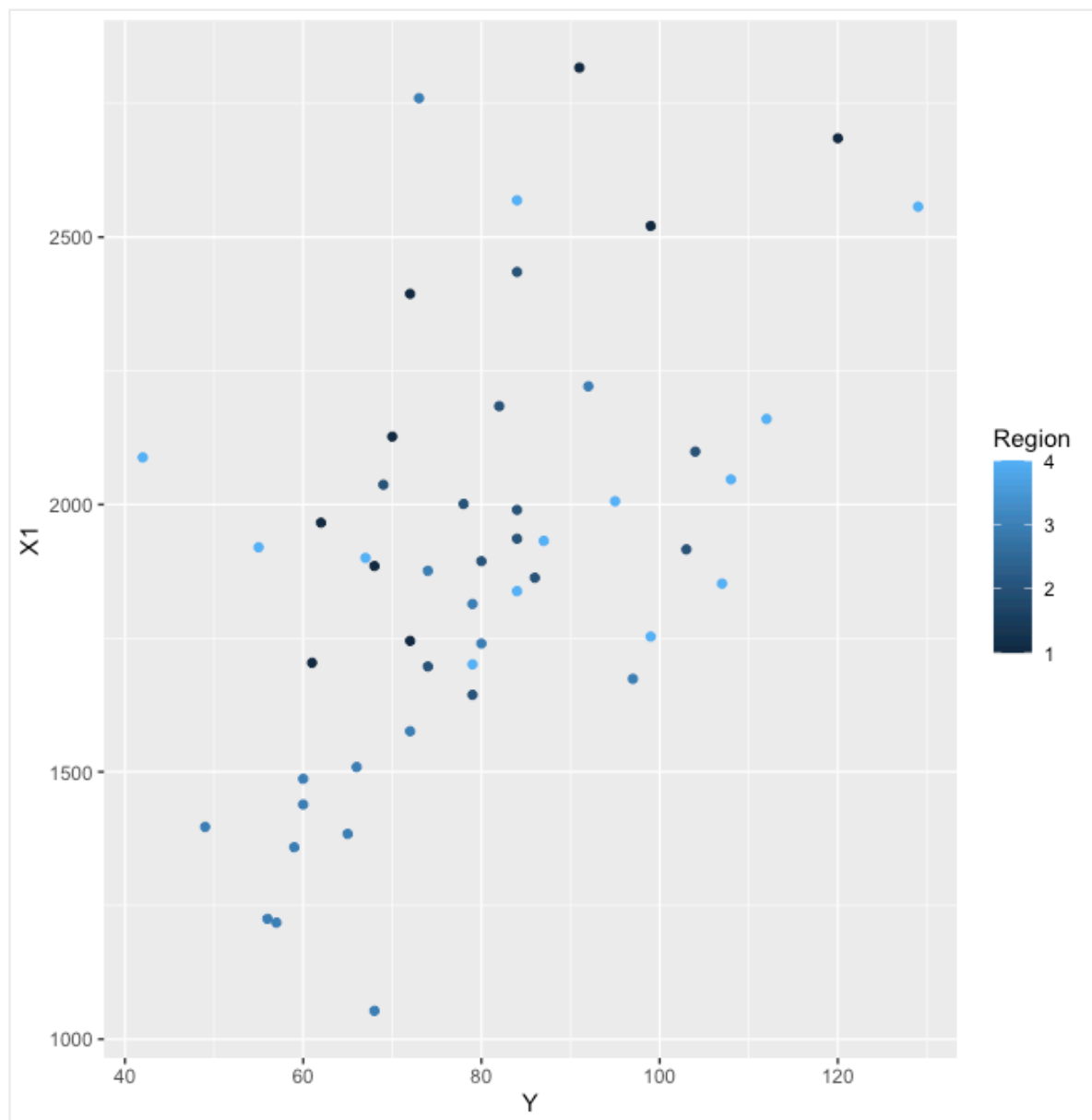
Variable 'region', added with different coloured points on the scatterplot.