Problem Set 1

Applied Stats/Quant Methods 1

Due: September 30, 2024

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:

```
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.

```
# exploring the data set
mean_y <- mean(y)
cat("Mean:", mean_y, "\n")
sd_y <- sd(y)
cat("Standard Deviation:", sd_y, "\n")
class(y)
length(y)

# 90% confidence interval using t.test
confidence_interval <- t.test(y, conf.level = 0.90)
confidence_interval$conf.int
```

Results:

```
Mean: 98.44
Standard Deviation: 13.09287
> class(y)      [1] "numeric"
> length(y)  [1] 25
> # 90% confidence interval using t.test
[1]  93.95993 102.92007
attr(,"conf.level")
[1] 0.9
```

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 $\alpha = 0.05$.

```r
# one-sample t-test:
t_test_result <- t.test(y, mu = 100, alternative = "greater", conf.level = 0.05)
t_test_result
```

   Results:

```
One Sample t-test

data:  y
t = -0.59574, df = 24, p-value = 0.7215
alternative hypothesis: true mean is greater than 100
5 percent confidence interval:
102.9201      Inf
sample estimates:
mean of x
98.44
```

   Conclusion: Because the p-value is 0.7215 and it's greater than 0.05, I do not have enough evidence to reject the null hypothesis that average student IQ in the school is equal to the national average of 100 IQ score.

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*

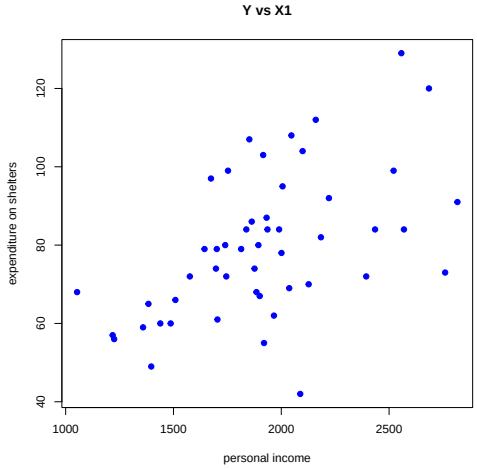
Explore the `expenditure` data set and import data into `R`.

```r
# uploading expenditure data set
getwd()
setwd("/Users/rafaelaalves/Documents/GitHub/StatsI_Fall2024/problemSets/PS01/
    my_answers")


# reading the file
expenditure <- read.table("expenditure.txt", header = TRUE, sep = "\t")
head(expenditure)
str(expenditure)
View(expenditure)


# installing stargazer package to save table view as a .tex
install.packages("stargazer")
library(stargazer)
stargazer(expenditure, type = "text", out = "expenditure_table.tex")
```

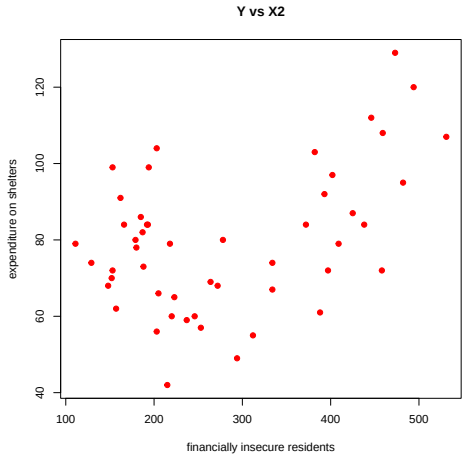
Table 1:

Statistic	N	Mean	St. Dev.	Min	Max
Y	50	79.540	18.507	42	129
X1	50	1,911.900	400.348	1,053	2,817
X2	50	281.780	118.193	111	531
X3	50	561.720	145.037	326	899
Region	50	2.660	1.062	1	4

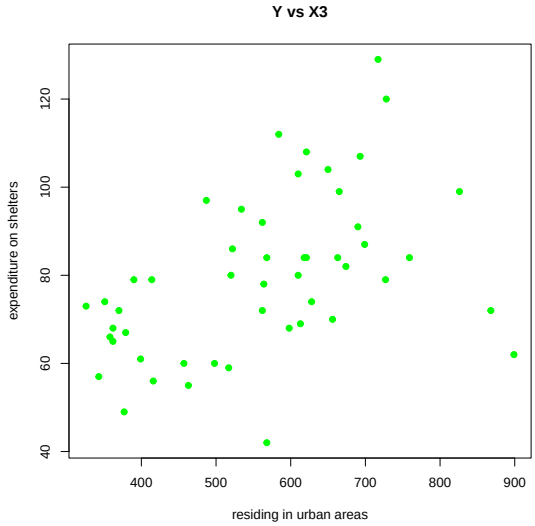
- Please plot the relationships among *Y*, *X1*, *X2*, and *X3*? What are the correlations among them (you just need to describe the graph and the relationships among them)?



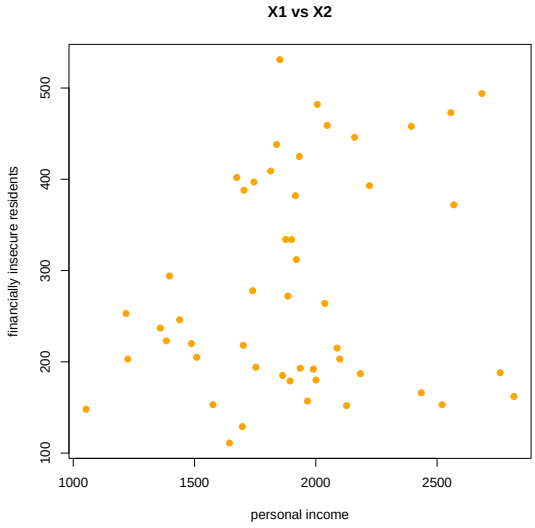
There's a correlation between the per capita expenditure in housing assistance regarding the group that has a personal income between 1,600 to 2,100.



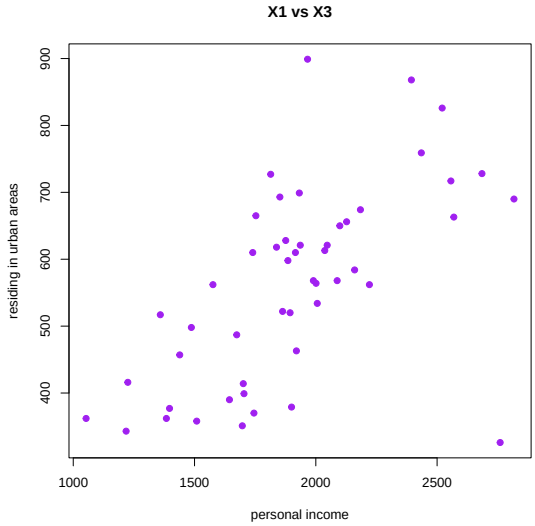
The spent per capita for housing assistance is higher where there are more financially insecure residents.



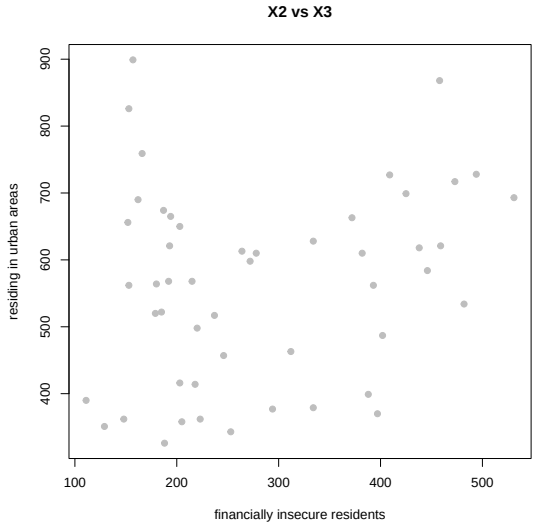
The graph shows that spending in housing assistance is lower in rural areas and hight in urban areas.



The residents with personal income around 1.500 to 2.100 are the ones with a higher financially insecure situation.

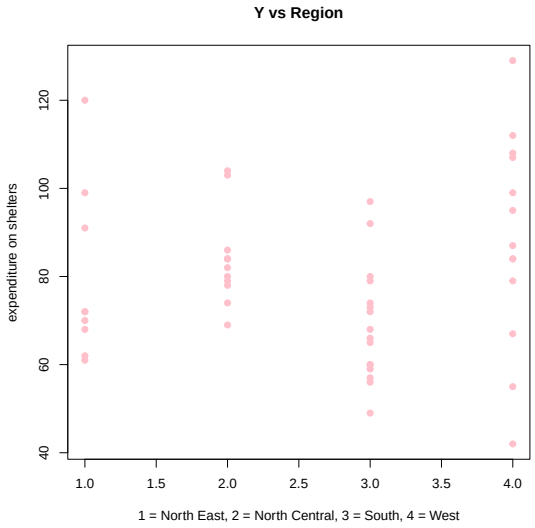


People with lower personal income reside un rural areas, and higher incomes in urban areas.



Even though people with lower personal incomes live in rural areas, there are less people financially insecure living in rural areas. It may be correct to say that personal income is not related with financial security, and/or the income is lower in rural areas but also the cost of life is cheaper.

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



On average, the West region has the highest per capita expenditure on housing assistance.

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