Problem Set 1

Valeriia Babaian

Due: October 1, 2023

Question 1 (40 points): Education

1. Find a 90% confidence interval for the average student IQ in the school.
   Answer: [93.95993; 102.92007].

   ```
   t_score <- qt(0.95, df=length(y)-1)
   lower_90_t <- mean(y)-(t_score)*(sd(y)/sqrt(length(y)))
   upper_90_t <- mean(y)+(t_score)*(sd(y)/sqrt(length(y)))

   # Confidence interval boundaries: 93.95993; 102.9201
   lower_90_t
   mean(y) #98.44
   upper_90_t

   #double checking:
   t.test(y, conf.level = 0.9, alternative = "two.sided")
   # 90 percent confidence interval: 93.95993; 102.92007
   ```

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.
   Answer: It is not higher:

   ```
   t.test(y, mu = 100, alternative = "greater")
   # We cannot reject the null hypothesis that the average student IQ in the school is not higher than the average IQ score across all schools in the country (p-value = 0.7215)
   ```

   Using the same sample, conduct the appropriate hypothesis test with $\alpha = 0.05$.
   Answer: Expectedly, the answer is the same as the average student IQ score in the given school was not greater than average IQ across the country on any of conventional levels of confidence.

```
##Using the same sample, conduct the appropriate hypothesis test with \
    alpha = 0.05.
t.test(y, mu = 100, conf.level = 0.95) # we cannot first reject the H0
    that the averages are equal on 0.05 significance (p-value = 0.5569)
```

Question 2 (40 points): Political Economy

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

```
expenditure <- read.table(”https://raw.githubusercontent.com/ASDS-TCD/StatsI_
    Fall2023/main/datasets/expenditure.txt”, header=T)
str(expenditure)
```

- 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)?

```
library(GGally)
library(ggplot2)
ggpairs(expenditure,
        columns = 2:5,
        columnLabels=c(”Housing assistance in state”,
       ”Personal income in state”,
       ”Financially insecure share”,
       ”Urban population share”))
ggsave(”Babaian-plot1.pdf”, width = 8,  height = 6)
```

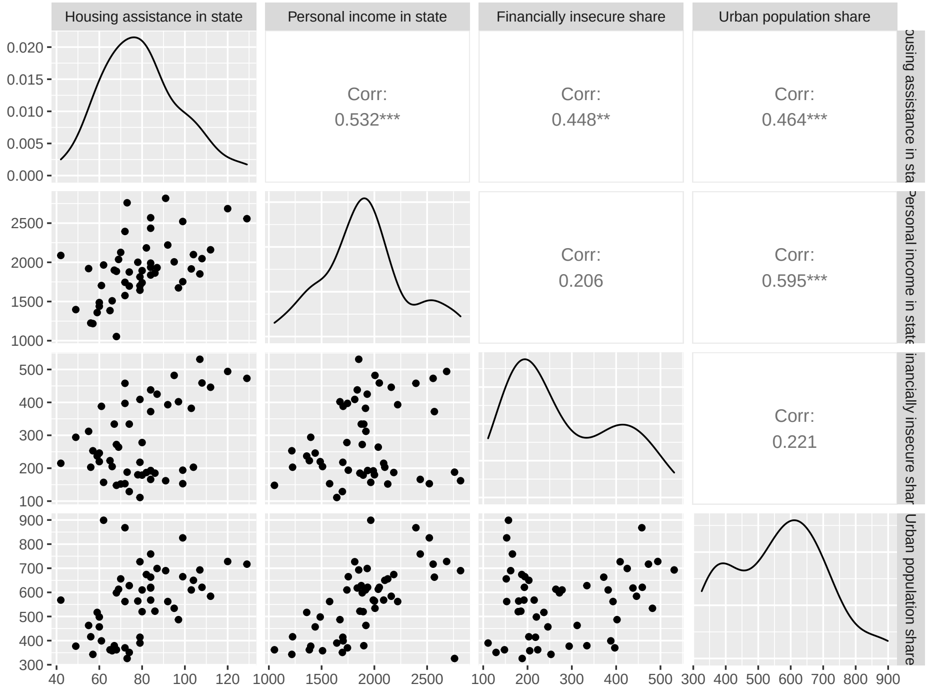

Figure 1: Fig. 1

According to the visualized relationships between all meaningful quantities in the data, the variables are positively related, but there are no strong correlations between any of them. The upper-right part of the graphics proves this point.

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

```
expenditure$Region <- factor(expenditure$Region, labels = c("Northeast", "North Central", "South", "West"))

ggplot(data=expenditure, mapping=aes(x=Region, y=Y)) +
  stat_summary(fun.data=mean_sdl, geom="bar") +
  labs(title="Per capita expenditure on shelters/housing assistance in state by region",
       x="", y="")
ggsave("Babaian-plot2.pdf", width = 8,   height = 6)
```

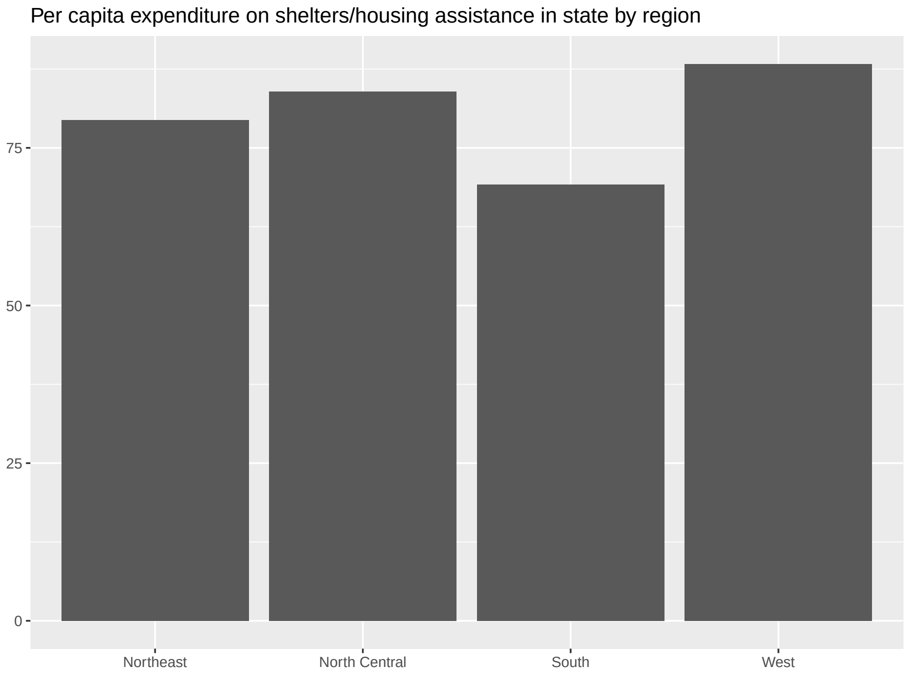

Figure 2: Fig. 2

The highest per capita expenditure on housing assistance is in the Western states.

- 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.

```r
plot(expenditure$X1, expenditure$Y,
     xlab=”Per capita personal income in state”,
     ylab=”Housing assistance in state”,
     main=”The Relationship between state-level personal income and housing assistance”)
ggsave(”Babaian-plot3.pdf”, width = 8,  height = 6)
# There is a positive relationship between per capita personal income in state
```

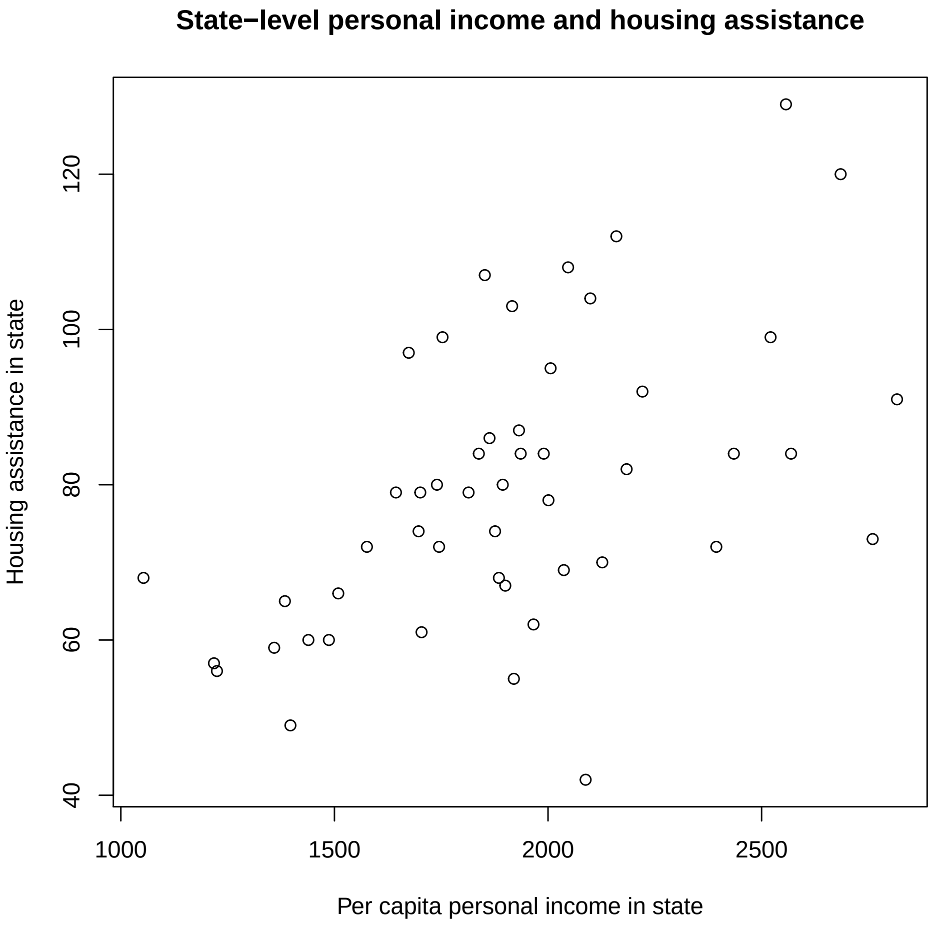

Figure 3: Fig. 3

There is a positive relationship between per capita personal income in state and its housing assistance expenditures, however, there is no strong correlation as there is a significant amount on deviations on both sides of the distributions.

```
plot(expenditure$X1, expenditure$Y,
     col= expenditure$Region,
     pch=c(21,22,23,24)[as.numeric(expenditure$Region)],
     xlab=”Per capita personal income in state”,
     ylab=”Housing assistance in state”,
     main=”State-level personal income and housing assistance by regions”)
legend(”topleft”, cex=1, legend=levels(expenditure$Region),
          pch=unique(expenditure$Region), col=unique(expenditure$Region))
dev.off()
```

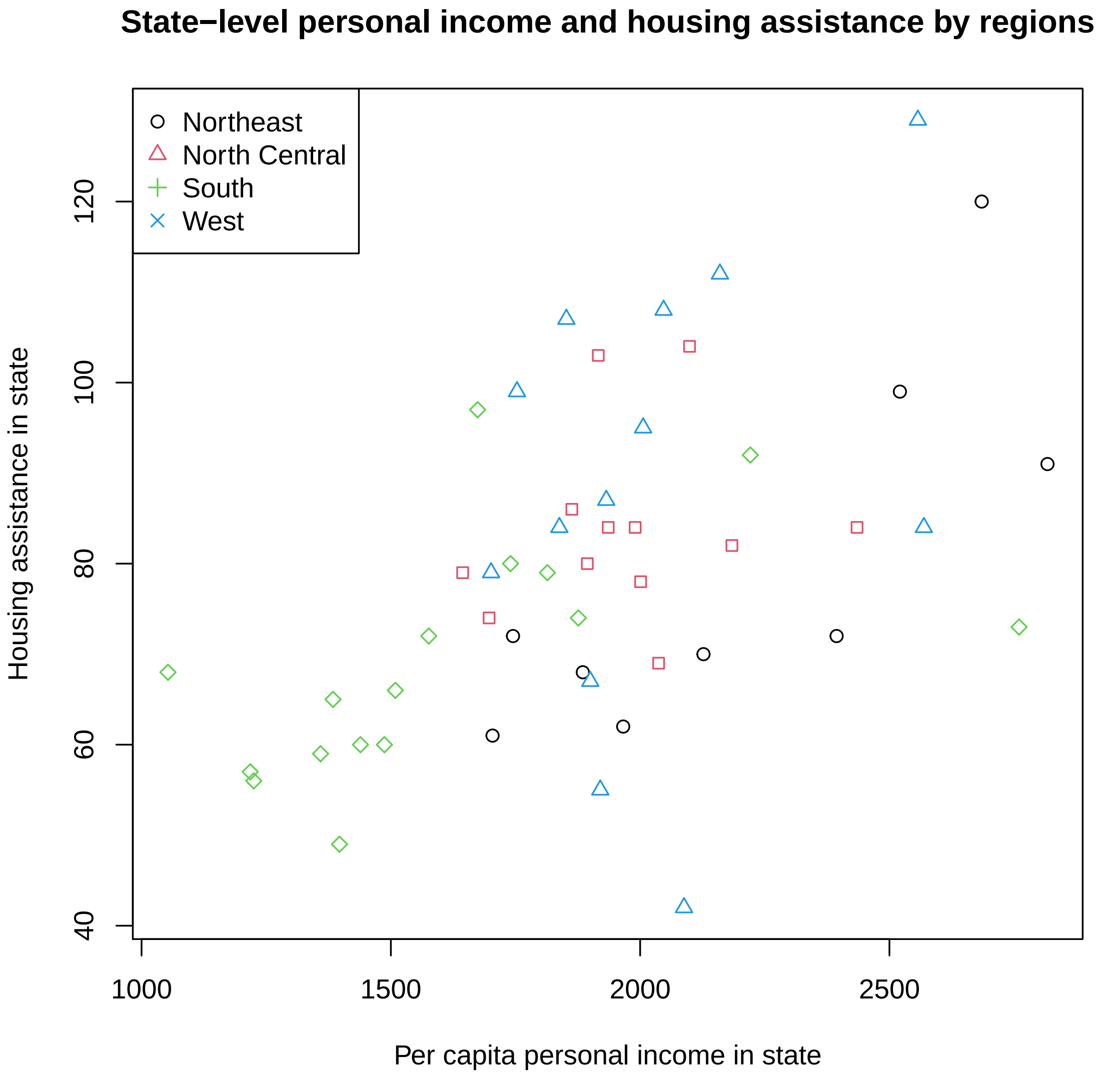

Figure 4: Fig. 4