

Statistical Analysis Assignment

Laura McPhillips

September 2021

1 Problem 1: Education

```
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.

```
t.test(y, con.level = 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 $\alpha = 0.05$

```
t.test(y, mu = 100, alternative = "greater")  
p-value = 0.7215  
We fail to reject the null hypothesis because p-value > 0.05
```

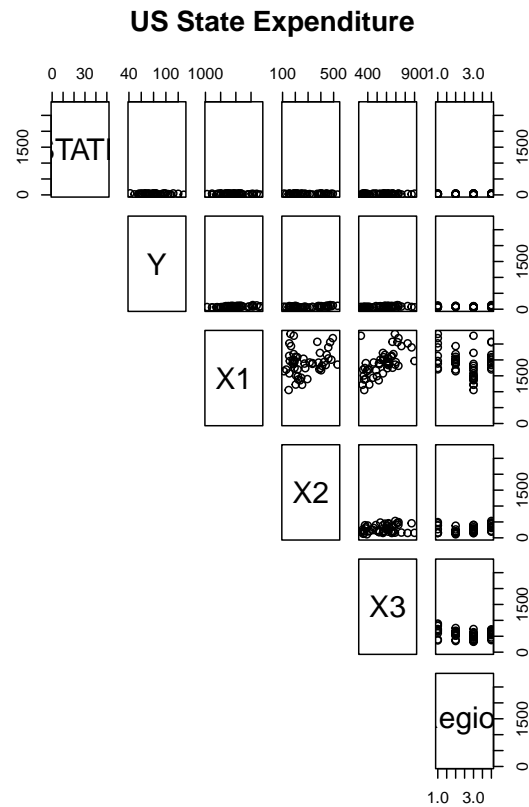
2 Problem 2: Political Economy

1. 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.

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)?

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



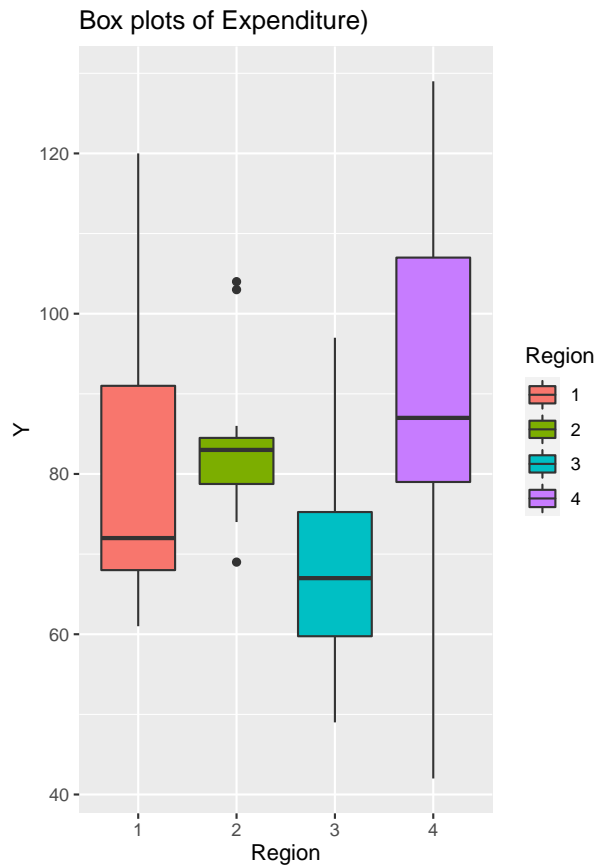
```
expenditure <-
  read.table("https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2021/main/datasets/expenditure.csv",
    header=T)
View(expenditure)
```

There is no relationship between State and any of the other variables.

There is no relationship between housing assistance per capita and any of the other variables.

There is a strong positive relationship between X1 and X2: personal income and financial insecurity and X2 and X3: financial insecurity urban residents.

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

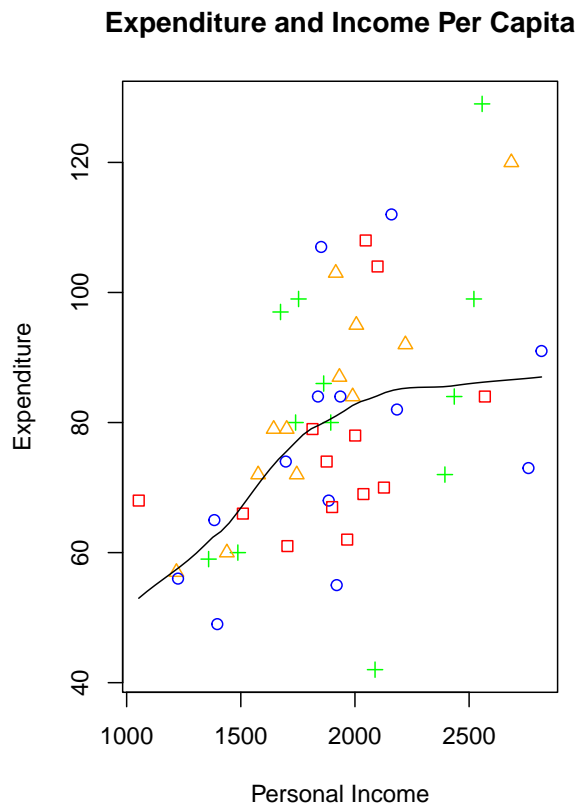


Region 4 - The West has the highest mean expenditure on housing assistance per capita

```
install.packages("ggplot2")
library(ggplot2)
expenditure$Region = as.factor(expenditure$Region)
barplot(table(expenditure$Region))
data=as.data.frame(expenditure[,c(2,6)])
data$Region = as.factor(data$Region)
mode(data$Region)

ggplot(aes(y=Y, x = Region, fill = Region), data = data) +
  geom_boxplot()+ggtitle("Box plots of Expenditure")
```

- 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.



```
scatter.smooth(expenditure$X1, expenditure$Y, xlab = "Personal Income",
  ylab = "Expenditure", main = "Expenditure and Income Per Capita",
  col = c("red", "blue", "orange", "green"), pch = c(0,1,2,3))
```
