Homework 1: Introduction to R and Statistics Basics

Practice

1. A private 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)
```

   Find a 90% confidence interval for the student IQ in the school assuming the population of IQ from which our random sample has been selected is normally distributed.

2. A private school counselor was curious whether the average of IQ of the students in her school is higher than the average IQ score 100 among all the schools in the country. She 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)
```

   Conduct a test with 0.05 significance level assuming the population of IQ from which our random sample has been selected is normally distributed.

3. Assume y is variable with values 1,2,3,4 standing for "Freshman", "Sophomore", "Junior", and "Senior", convert y from numbers to characters in R:

```
y <- c(1,2,1,3,4,1,1,4,2,1,3,4,3,2,1,3,4,1,2,3,1,1,2,1,1,3,4)
```

   Researchers are curious about what affects the education expenditure on public education. The following is availabe variables in a data set about the education expenditure.

State	*50 states in US*
Y	*per capita expenditure on public education*
X1	*per capita personal income*
X2	*Number of residents per thousand under 18 years of age*
X3	*Number of people per thousand residing in urban areas*
Region	*1=Northeast, 2= North Central, 3= South, 4=West*

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

```
expenditure <- read.table("your directory/expenditure.txt",header=T)
```

4. Which graph is appropriate to visualize the relationships among *Y*, *X1*, *X2*, and *X3*? What are the correlations among them? Produce the graph and describe the relationships among them.

5. Which graph is appropriate to visualize the relationship between *Y* and *Region*? On average, which region does have the highest per capita expenditure on public education?

6. Which graph is appropriate to visualize the relationship between *Y* and *X1*? Produce this graph and describe the relationship. Reproduce the above graph including one more variable *Region* and display different regions with different types of symbols and colors.

7. Write down your homework in *Rmarkdown* and generate it to a HTML document with your title and your name.