CIND 123 Winter 2018 - Assignment #3

Write your name here

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

Use RStudio for this assignment. Edit the file `assignment-3.Rmd` and insert your R code where wherever you see the string "INSERT YOUR ANSWER HERE"

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

Sample Question and Solution

Use `seq()` to create the vector $(2, 4, 6, \ldots, 20)$.

```
#Insert your code here.
seq(2,20,by = 2)
```

```
##  [1]  2  4  6  8 10 12 14 16 18 20
```

This assignment questions 1 - 4 make use of data that is provided by the `mosaic` package. (install mosaic package and load KidsFeet using data(KidsFeet)).

```
#install.packages('mosaic')
#library(mosaic)
#data(KidsFeet)
```

Question 1 - 30%

In this question you will explore the KidsFeet dataset available in the datasets library in the mosaic package.

a) Display the first 6 rows of the KidsFeet dataset.

```
#Insert your code here.
```

b) Display the type of each column of the airquality dataset, use only one function in R to do so.

```
#Insert your code here.
```

c) Use a histogram to assess the normality of the `width` variable.(In order to get the output diagram inserted in your answer use attach(dataframe name))

```
#Insert your code here.
```

d) Does it appear normally distributed? Why?

```
#Insert your answer here. (Do not remove the #)
# The distribution does not appear normally distributed.
#
```

e) Create a boxplot which shows the distribution of `width` in each `birth month`. Use different colors for each `birth month`.

```
#Insert your code here.
```

f) Create one scatter plot matrix of the numeric variables (length, width) within the KidsFeet dataset. (Hint investigate pairs())

```
#Insert your code here
```

Question 2 - 20%

a)Use simulation to estimate the mean and variance of a binomial random variable with n = 45 and p = 0.32. One time use 100 samples and the other time use 10000 samples.

```
#Insert your code here
```

b) Calculate the values using the theroy (state the value and the equation in your answer),compare the values you get with the values you got in (a), wirte one sentence sumurizing the comparision. Explain the difference between 100 samples and 10000 samples and which one seems to be more accurate and why?

```
#Insert your answer here (Do not remove the #)
# Theoretical values:
```

Question 3 - 30%

a)If there are twelve customers entering a mall per minute on average, find the probability of having seventeen or more customers entering the mall in a particular minute.

```
#Insert your code here
```

b) Estimate the mean and variance of a Poisson random variable in the previuos question by simulating 100 and 10,000 Poisson random numbers.

```
#Insert your code here
```

c) Compare the mean value you got in (b),with the one stated in the question. wirte one sentence summarizing the comparision. Explain the difference between 100 samples and 10000 samples and which one seems to be more accurate and why?

```
#Insert your answer here (Do not remove the #)
#
```

Question 4 - 20%

a) Generate normally distributed random numbers for three categories: A (n = 200, mean = 100, sd = 20), B (n = 200, mean = 120, sd = 20), and C (n = 200, mean = 80, sd = 20)

```
a1 = rnorm(n = 200, mean = 100, sd = 20)
a2 = rnorm(n = 200, mean = 120, sd = 20)
a3 = rnorm(n = 200, mean = 80, sd = 20)
```

b) Combine all the three categories in one dataframe then generate a density plot colored by category.

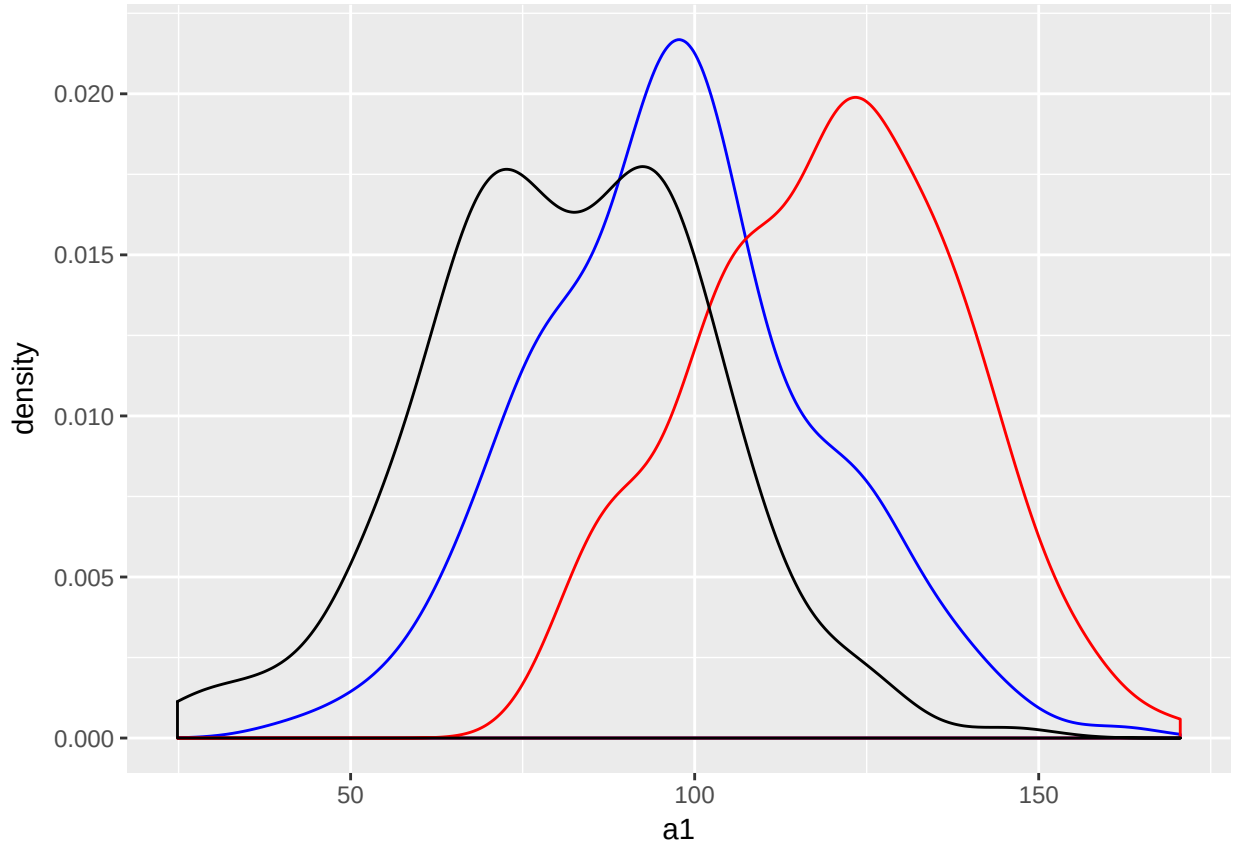
```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
data = data.frame(a1, a2, a3)

ggplot(data) +
```

```
  geom_density(aes(x = a1), color="blue")+
  geom_density(aes(x = a2), color="red")+
  geom_density(aes(x = a3))
```



END of Assignment #3.