Exploring and Visualizing Data in R

**Overview**

In this exercise, you will look at how to sample and simulate data within R. You will also look at data distributions and how visualization can aid in this process.

**1: Generating samples and reviewing ranges**

1. Take a random sample of size 5 from the weather data set (weather 2.csv), and detach the weather data set once completed:

**Mysample <-weather[sample(1:nrow(weather),5,replace=FALSE),]**

**rm(weather)**

1. Using a sequence of numbers 1 to 100, use the sample() function to generate a

3x3 data frame of test data, similar to the following:



x <- 1:100

y <- data.frame("a"=sample(x,3), "b"=sample(x,3), "c"=sample(x,3))

1. Generate 100 random numbers with a normal distribution. Set a seed of 1 for reproducible results:

**set.seed(1)**

**rnorm(n=100,mean=0,sd=1)**

1. Use the rpois() function to simulate 15 independent Poisson random variables with parameter lambda= 4:

**rpois(15, 4)**

* 1. **3 3 4 7 2 7 7 5 5 1 2 2 5 3 5**

1. Generate 10 random numbers between 1 and 3:

**runif(10, min=1, max=3)**

6. Calculate the range of the MinTemp in the weather data set:

**range(weather$MinTemp)**

7. Calculate the mean and median of car's mpg:

**mean(mtcars$mpg) median(mtcars$mpg)**

8. Explore how much the minimum temperature has changed every hour:

**diff(weather$MinTemp)**

9. Calculate the interquartile range:

**IQR(weather$MinTemp)**

1. Calculate the observation that cuts off the first 25 percent of the data values when it is sorted in ascending order:

**quantile(weather$MinTemp,0.25)**

11. Display a six-number summary (including the mean of the MinTemp):

**summary(weather$MinTemp)**

**2: Visualizing data**

12. Generate a histogram of mtcar's mpg:

**hist(mtcars$mpg)**

13. Modify the bin width to increments of 5:

**hist(mtcars$mpg, c(5,10,15,20,25,30,35))**

14. Return the density of the data and plot the results:

**d <- density(mtcars$mpg) plot(d)**

15. Display a box plot of the mpg values, with a red shaded box:

**boxplot(mtcars$mpg, col="red")**

1. Cars can have 3, 4, or 5 gears. We need to find how many cars of each type are in the data set. Use a box plot with red bars to visualize your answer, and title the

plot **Number of Gears**

**counts<- table(mtcars$gear) barplot(counts, main="Cars", col="red", xlab="Number of Gears")**