IST-687 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

Name: Yicun Deng
Class: M003
Date: 9/8/10
Chapter Number: #789
Title of Chapter: **Descriptive Statistics**

Descriptive Statistics

We are learning how to write functions in R this week and learning more into statistic.No matter how much of the data we have, the sample we have would be a sample portion of the data, but we can use statistic to get an tendency of the real world running. We have two main ways to do statistics, one is measure of central tendency, which focus on the mean, mode, median, while measure of dispersion would be focusing on how the data point are sitting on the dataset by range variance and standard deviation.
hist() give you the histogram of selected data.
hist(,breaks=)give you an option of getting more bars hence more representative of the data.
Dispersion talks more about continuity of the dataset
Central tendency talks more about the centered number of the dataset
rnorm(samples,mean, standard deviation) to generate vector or normal distribution dataset

Lecture 3.2- Writing Functions

A function is a bucket of word that you can reuse and you input something, run something, then returning something.
unique()returns the unique objects in a vector.
tabulate() counts for each instance how many there are.
using code like unique(input)[which.max(tabulate(input))] get us the mode of the input vector
use match()to pair the output vectore or unique() and tabulate()

Lecture 3.3 - Data Science in the Real World

Data science has been used in many industry, research and NPO.
Take myself as an example, Data science is super useful for Linguistics research. Take the coke machine for an example, it even get data from temperature and vibration for maintenance control. Often time, there are certain pattern we commonly use to produce strategy from data.
Data science can be used in analyzing the customer’s preference and when and how the supply chain works. Just like the first weeks of classes.

IST-687 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

Exercise Review

Chapter 7

Data variables are strings most of the time, value variables are numbers most of the time. But numbers can be stored in string form too”123”.

Chapter 8

```
MyVectorInfo<- function(mv){
  return(summary(mv))
}
```

Chapter 9

```
rpareto(51,1)
```

Question for Class

I had learnt that rpareto() can be used to generate pareto distribution from http://ugrad.stat.ubc.ca/R/library/rmutil/html/Pareto.html, but Rstudio didn’t allow me to run it. I am wondering why