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: 9/1/2020
Chapter Number: #4-6
Title of Chapter: Module Objectives

Lecture 2.1 - Module Objectives

The main focus of this chapter is to introduce us to the basic data modeling and basic data frame, and use it on R.(Rstudio)

Lecture 2.2 Data Modeling

This section has introduced the Data Flow Diagram(DFD) to us. This focused on data flow more. But Entity Relationship Diagram(ERD) focus more on relationships between entities. The hierarchy of Information System Types are really like the hierarchy of the value of data, most unprocessed data refer to the direct transacted data, wisdom requires an analytics

Roundtable 2.2a - Data Modeling

Reveal the other point why we are doing data modeling: effectively converting unstructive data to structed data.

Roundtable 2.2b - Data Modeling Roundtable (Part 2)

Different models gives different aspect of data, DFD is easy to start with and ERD is good while you actually start to analyze the data.

Lecture 2.3 Data Frames

Introducing us the function of <- data.frame()
str() reveal the input data frame and also shows what type of data it have
summary() gives statistic of the input dataframe
dataframe(1,1)(1,)(,1)(-1,)(,-1)negative here means exclude

Lecture 2.4 Data Frames

It is important to under the characteristics of the dataset you are looking at.
Each row or column referring to what?
Questions:
I will create two columns, studentID and studentGPA and combine them into a dataframe student.
StudentID StudentGPA
N1 3.8

N2 4.0
N3 3.3
N4 3.5
N5 3.9

Lecture 2.5 Data Frames

Professor gave his example of the grid.

R Coding 2.1 - Vector Review

Introducing us the function of <- data.frame()
str() reveal the input data frame and also shows what type of data it have
summary() gives statistic of the input dataframe
dataframe(1,1)(1,)(,1)(-1,)(,-1)negative here means exclude
introducing data.frame(x,y,stringAsFactors=false) to allow us add new strings to the dataframe using <- rbind(dataframe, c(x,y))
<- cbind(dataframe, newcolumn)
use dataframe$column if you want to change or refer to certain column
if you need to exclude multiple row or column use -x:-y:-z.
use dataframe$column <- as.numeric(dataframe$column) to convert string into number
use which.max() to get the largest number stored row of that column.

R Coding 2.2 - Data Frames

Dataframe works the same as vector math
If you run code like
X <- y[y$z>29] it returns you a dataframe with the elements you want
order() return the index in order of your preference(bet it is very useful) you can use negative to give the order upside down

Lecture 2 - Data Science Interview

Give us an idea of what people actually do in the industry and what the distribution of jobs are. Also, explain data science's difference to traditional report: future trend and past trend analyze.

Exercise Review

Chapter 4 challenge:
The pro on using the ERD is that it can reviews the relationships between entities better, the counter on using the ERD is that it doesn't show the data flow directly so you wouldn't know where your data is coming origin and how does it do cycles.

Chapter 5 challenge:
myFamilyIQ<- c(120,130,125,125,999)

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.

myFamily$myFamilyIQ<- cbind(myFamily,myFamilyIQ)

Chapter 6 challenge:
It requires a membership to download the csv… I can't finish this challenge, sorry.

R Code Fragment and Explanation
I have take notes on the Rcode in previous section, thank you!!

Question for Class
I am still a bit confuse of the exercise part but I will finish any exercise I see from the video or the book.