Introduction to Data Science (DSE 3(b))

S.No	Topic	Chapters	Reference	# of Lectures
1.	Introduction Data Sciences	Ch 1,2	[1]	2
2.	Data science Installation and Administration	Ch1,3,5,6,9(for windows only) Lab	[2] R-admin	
3.	R Programming Basics: Overview of R, R data types and objects, reading and writing data, Control structures, functions, scoping rules, dates and times, Loop functions, debugging tools, Simulation, code profiling	Ch 1-10	[2] R-Intro	10
4.	Getting and Cleaning Data: Obtaining data from the web , from APIs, from databases and from colleagues in various formats	Ch 1,2,4,9 Web API	[2] R-Data https://tclavelle.github.io/blog/r_and_apis/	14
5.	Exploratory Data Analysis: Essential exploratory techniques for summarizing data, applied before formal modeling commences, eliminating or sharpening potential hypotheses about the world that can be addressed by the data, common multivariate statistical techniques used to visualize highdimensional data	Ch.2(only 2.6), Ch 9 (9.1-9.11) , Ch 10 (10.1 , 10.2, 10.9, 10.11 ,	[3]	12
6.	Reproducible Research: Concepts and tools behind reporting modern data analyses in a reproducible manner, To write a document using R markdown, integrate live R code into a literate statistical program, compile R markdown documents using knitr and related tools, and organize a data analysis .	Ch 3,4,8,9	https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf [4]	12

References:

[1]   Rachel Schutt, Cathy O'Neil, "Doing Data Science: Straight Talk from the Frontiline" by Schroff/O'Reilly, 2013.

[2]   R Software manual(R 3.4 Setup)

[3]"R Cookbook"

http://www.bagualu.net/wordpress/wp-content/uploads/2015/10/R_Cookbook.pdf

[4]   Roger Peng "Report writing for data science in R" http://leanpub.com/reportwriting

List of Practicals for Introduction to Data Sciences

1. Use Git / Github software to create Github account. Also, create a repo using Github.
2. Write a program that prints 'Hello World' to the screen.
3. Write a program that asks the user for a number n and prints the sum of the numbers 1 to n .
4. Write a program that prints a multiplication table for numbers up to 12.
5. Write a function that returns the largest element in a list.
6. Write a function that computes the running total of a list.
7. Write a function that tests whether a string is a palindrome.
8. Implement linear search.
9. Implement binary search.
10. Implement matrices addition , subtraction and Multiplication
11. Fifteen students were enrolled in a course. There ages were:
    20 20 20 20 20 21 21 21 22 22 22 22 23 23 23

    i.   Find the median age of all students under 22 years
    ii.   Find the median age of all students
    iii.   Find the mean age of all students
    iv.   Find the modal age for all students
    v.   Two more students enter the class. The age of both students is 23. What is now mean, mode and median?
12. Following table gives a frequency distribution of systolic blood pressure. Compute all the measures of dispersion.

Midpoint	95.5	105.5	115.5	125.5	135.5	145.5	155.5	165.5	175.5
Number	5	8	22	27	17	9	5	5	2

13. Obtain probability distribution of, where X is number of spots showing when a six-sided symmetric die (i.e. all six faces of the die are equally likely) is rolled. Simulate random samples of sizes 40, 70 and 100 respectively and verify the frequency interpretation of probability.
14. Make visual representations of data using the base, lattice, and ggplot2 plotting systems in R, apply basic principles of data graphics to create rich analytic graphics from available datasets.