

School of Computer Science, Engineering and Technology

Course Name : R Programming for Data

Course Type: Core

Course Code: C21X0C3

Teaching Scheme			Credits	Examination Marks				Total Marks
L	T	P	C	Theory Marks		Practical Marks		
				External	Internal	External	Internal	
4	1	2	6	40	60	20	30	150

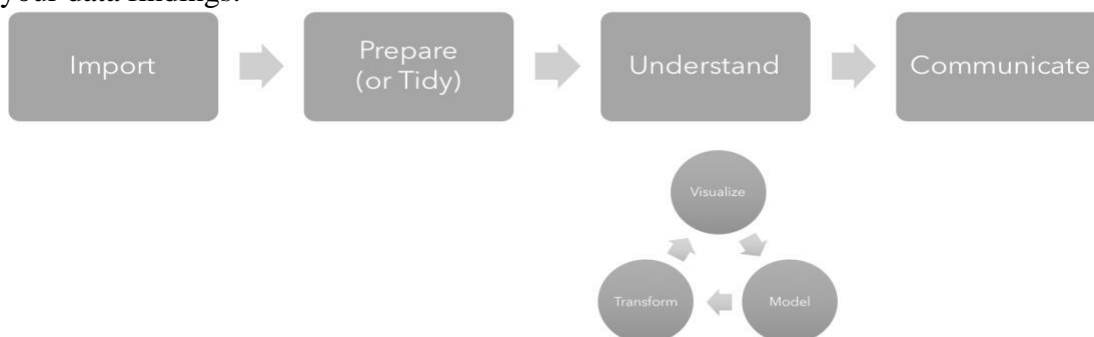
What is R?

R is an open source language and environment for statistical computing, data mining, modeling, and data graphics. It provides a wide variety of statistical and graphical techniques such as linear and non-linear modeling, statistical tests, time series analysis, classification, and clustering.

R is one of the most used business analytics tools. For example, Facebook uses R for behavior analysis related to status updates and profile pictures. Google uses it to analyze advertising effectiveness and economic forecasting. Twitter leverages R for data visualization and semantic clustering. In the 2016 data science salary survey conducted by O'Reilly, R was ranked second in a category of programming languages for data science.

Course overview

In this course, students will learn how to program in R and how to use R for effective data analysis and visualization. "Turn raw data into understanding, insight, and knowledge" (Wickham & Grolemund, 2017, p. ix) using R to import, prepare, understand, and communicate your data findings.



The course begins with developing a basic understanding of the R working environment. Next, students will be introduced the necessary arithmetic and logical operators, salient functions for manipulating data, and getting help using R. Next, the common data structures, variables, and data types used in R will be demonstrated and applied. Students will write R scripts and build

School of Computer Science, Engineering and Technology

R markdown documents to share their code others. They will utilize the various packages available in R for visualization, reporting, data manipulation, and statistical analysis.

Students import data sets, transform and manipulate those datasets for various analytical purposes. Students will learn how to create control structures, such as loops and conditional statements to traverse, sort, merge, and evaluate data. Finally, students create interactive business applications that allow for data querying and data exploration.

Prerequisite

This course is designed for those who have no experience in R or programming. This class give you skills in programming in R and introduce you to 1) A new way of thinking 2) A new language for speaking and reading (vectors, data frames, functions, objects, etc. and 3) a new syntax for writing , e.g. `c()`, `print()`, `cat()`, `sort()`, `require()`, `subset()` for data analysis and presentation.

Course outcomes

By the end of the course students you shall be confident and equipped with all the knowledge required to perform analytical activities in R. Specifically,

	Course Outcome	Bloom's Level
CO1	Understand the fundamental syntax of R through readings, practice exercises, demonstrations, and writing R code.	Understanding
CO2	Import a variety of data formats into R using RStudio	Understanding
CO3	Prepare or tidy datas for in preparation for analysis	Applying
CO4	Query data using SQL and R	Application
CO5	Analyze a data set in R and present findings using the appropriate R packages	Analyze
CO6	Visualize data attributes using ggplot2 and other R packages.	Application

School of Computer Science, Engineering and Technology
CO-PO Mapping

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PO 13	PO 14	PO 15	PO 16
CO1	1	3		3	3	3	2	2		1	1	1	2	3	1	3
CO2	2	3	3	3	1	1	2	2	1				2	3	1	3
CO3	3	3	3	3	3	3	2	3	2	2	2	1	2	3	2	2
CO4	1	3	3	3	3	3	3	2	2		1			3	1	2
CO5	1	3	3	3	3	2	1	2	1			1	2	2		2
CO6		1	1	3	3	1	1	2			2			2		1

Tutorials, video demonstrations, and exercises

Sosulski, K. (2018). R Fundamentals.

Textbook

Sr. No	Book Name
1.	Wickham, H. & Grolemund, G. (2018). for Data Science. O'Reilly: New York
2.	Norman Matloff(2011),The Art of R Programming: A Tour of Statistical Software Design ,No Starch Press,

Reference Book

Sr. No	Book Name
1.	Jared P. Lander(2013), R for Everyone: Advanced Analytics and Graphics, Addison-Wesley Data & Analytics Series
2.	Mark Gardener(2013),Beginning R – The Statistical Programming Language, Wiley
3.	Robert Knell(2013),Introductory R: A Beginner's Guide to Data Visualisation, Statistical Analysis and Programming in R, Amazon Digital South Asia Services Inc,

Other Videos:

School of Computer Science, Engineering and Technology

Sr. No	About Video	Link	Topic
O1.	Stanford Seminar - Expressing yourself in R, Hadley Wickham Rice University	https://www.youtube.com/watch?v=wki0BqlztCo	Data manipulation in R
O2.	MIT 15.071 The Analytics Edge, Spring 2017, MIT penCourseWare	https://www.youtube.com/watch?v=d2CfWJkklvo	Working with Data: History of R
O3.	MIT 15.071 The Analytics Edge, Spring 2017, MIT penCourseWare	https://www.youtube.com/watch?v=E_KUHMuoPLE	Getting Started in R

Related MOOCs courses

Sr.No	MOOC Course
M1	Introduction to R Software by Prof. Prof. Shalabh 8 Weeks on NPTEL.
M2	Data Science: Foundations using R, Johns Hopkins University(Coursera)

Required software

- R: <http://www.r-project.org/> (**FREE**)
- RStudio (additional libraries required): <http://www.rstudio.com/> (**FREE**)

Learning resources

- R Project: <http://www.r-project.org/>
- RStudio (additional libraries required): <http://www.rstudio.com>
- Quick-R <http://www.statmethods.net/>
- Google's R Style Guide: <http://google-styleguide.googlecode.com/svn/trunk/Rguide.xml>

Course Outline

Unit No.	Topics	Lab	Assignment	Hours
1	Introduction to R programming <ul style="list-style-type: none"> • What is R? • Installing R and RStudio • RStudio Overview • Working in the Console • Arithmetic Operators • Logical Operations • Using Functions 	P1	A1	4

School of Computer Science, Engineering and Technology

	<ul style="list-style-type: none"> • Getting Help in R and Quitting RStudio 			
2	Data structures, variables, and data types <ul style="list-style-type: none"> • Creating Variables • Numeric, Character and Logical Data • Vectors • Data Frames • Factors • Sorting Numeric, Character, and Factor Vectors • Special Values 	P2	A2	4
3	R packages and scripts <ul style="list-style-type: none"> • Installing and loading packages • Setting up your working directory • Downloading and importing data • Working with missing data • Extracting a subset of a data frame • Writing R scripts • Adding comments and documentation • Creating reports 	P3	A3	4
4	Descriptive statistics in R <ul style="list-style-type: none"> • Measures of central tendency • Measures of variability • Skewness and kurtosis • Summary functions, describe functions, and descriptive statistics by group • Correlations 	P4	A4	5
5	Statistical graphs <ul style="list-style-type: none"> • Scatter Plots • Box Plots • Scatter Plots and Boxand-Whisker Plots 	P5	A5	2

School of Computer Science, Engineering and Technology

	Together • Histograms			
6	Working with messy data • Messy Data • Renaming Columns (Variable Names) • Attaching / Detaching • Tabulating Data: Constructing Simple Frequency Tables • Ordering Factor Variables	P6	A6	4
7	Iteration • while loops • for loops	P7		4
8	Conditional Statements • If / else • Boolean logical operators	P8		2
9	Writing functions Reporting • Creating functions • Calling functions	P9	A7	5
10	Data Exploration and Visualization • Using the ggplot2 package to visualize data • Applying themes from ggthemes to refine and customize charts and graphs • Building data graphics for dynamic reporting	P10		3
11	Data querying: SQL and R • Writing SQL statements in R • Using the Select, From, Where, Is, Like, Order By, Limit, Max, Min SQL Functions	P 11		4
12	Interactive reporting with	P12	A8	4

School of Computer Science, Engineering and Technology

	Rmarkdown <ul style="list-style-type: none"> • RMarkdown basics • Text formatting • Code chunks • YAML header • Preview of notebooks, presentations, websites, and dashboards 			
Total Hours				45

Lab Experiments:

Sr. No	Practical Name
P1	<p>Compute the following:</p> <ol style="list-style-type: none"> 1. $(123 - 45) / 4 + 4 * (72 / 2.34 - 3)$ 2. $((20*3)-14)^3$ 3. Absolute value of -88 4. Base 10 logarithm of 72 5. Round the square root of 50 to the fourth decimal 6. $e^{1.45} - 2.612$ 7. <ol style="list-style-type: none"> a. Assign a variable year_born to 1984 b. Assign a variable year_current to 2014 c. Assign a variable age and compute it d. Return True / False if person is eligible to vote in US (if age is greater than or equal to 18) 6. Given: formula for area of circle is $\pi * r^2$ Given: Area = 100 <ol style="list-style-type: none"> a. Write statement to find r. (Hint: utilize "sqrt" and "pi" functions) 7. Given: went to lunch and pre-tax bill was \$45.90 <ol style="list-style-type: none"> a. Compute subtotal: add NYC tax of 8.875% b. Compute 15% tip on subtotal c. Compute 20% tip on subtotal 8. <ol style="list-style-type: none"> a. Assign a variable customers to 500

School of Computer Science, Engineering and Technology

	<ul style="list-style-type: none"> b. Assign a variable pizza_price to \$20 c. Assign a variable todays_revenue (customers * pizza_price) and compute today's revenue d. Is today's revenue greater than yesterday's revenue of \$7,000 and less than tomorrow's projected revenue of \$11,000? Show the code that would answer the following question. 																
P2	<p>Dataset: Winter Olympic Medals.</p> <p>Download Link: http://becomingvisual.com/rfundamentals/winter_olympic.csv</p> <p>Data Dictionary: Review the data dictionary for the Winter Olympic Medals data set.</p> <table border="1"> <thead> <tr> <th>Variable</th><th>Description</th></tr> </thead> <tbody> <tr> <td>Rank</td><td>Rank in number of medals</td></tr> <tr> <td>NOC</td><td>Name of country</td></tr> <tr> <td>Gold</td><td>Number of gold medals</td></tr> <tr> <td>Silver</td><td>Number of silver medals</td></tr> <tr> <td>Bronze</td><td>Number of bronze medals</td></tr> <tr> <td>Total</td><td>Total number of medals</td></tr> <tr> <td>Region</td><td>Country Region</td></tr> </tbody> </table> <ul style="list-style-type: none"> 1. Getting to know the data <ul style="list-style-type: none"> a. Import the data b. View the data c. How many variables are in the data frame? d. What are the names of these variables? e. How many countries (rows) are in the data frame? 	Variable	Description	Rank	Rank in number of medals	NOC	Name of country	Gold	Number of gold medals	Silver	Number of silver medals	Bronze	Number of bronze medals	Total	Total number of medals	Region	Country Region
Variable	Description																
Rank	Rank in number of medals																
NOC	Name of country																
Gold	Number of gold medals																
Silver	Number of silver medals																
Bronze	Number of bronze medals																
Total	Total number of medals																
Region	Country Region																

School of Computer Science, Engineering and Technology

	<ol style="list-style-type: none"> 2. Printing data <ol style="list-style-type: none"> a. The first row of data b. The last row of data c. The first 5 rows of data 3. Creating vectors <ol style="list-style-type: none"> a. Create a vector called "country_medals" from data frame b. Create a vector called "gold" from data frame c. What type of variable is "gold"? 4. Create a new data frame that holds data from the region Asia <ol style="list-style-type: none"> a. Call the data frame "asia" b. How many rows and columns are in this data frame? [Hint: use dim()] 5. Create the data frame "total_medals" <ol style="list-style-type: none"> a. Create vector "country" b. Create vector "total_medal_ct" c. Use cbind() to combine the two vectors d. What is the type of object "total_medals"? 6. Vector data counts <ol style="list-style-type: none"> a. What are the different levels of data\$Region? [Hint: use levels()] b. Are any of the other variables factor variables? [Hint: use str()] 7. Subsetting <ol style="list-style-type: none"> a. Create a data frame that holds countries that did not win any gold medals 8. <ol style="list-style-type: none"> a. Create a vector called test_scores with the following values 92, 75, 84, 94, 88, 89, 91 b. Create a vector called students with the following values Jerry, Monica, Felix, James, April, Ruth, Tony c. Create a data frame with these two vectors d. It turns out that Monica's test was regraded and was awarded five extra points - correct this in the data frame. e. Extract the students who got above or equal to 90% f. Sort all the students by their test score in descending order
P3	Dataset: Basketball data from March Madness

School of Computer Science, Engineering and Technology

Download Link: http://becomingvisual.com/rfundamentals/march_madness.csv

Data Dictionary:

Variable	Description
Rank	Team Ranking
Previous	Previous Team Ranking
School	Name of the College or University
Conference	NCAA Conference (30 +)
Record	Overall Record
Neutral	Record with games in a neutral location
Home	Record with games at home
Non Div I	Record with non-divison 1 games

Write a R script to do the following:

1. Set working directory – Hint: setwd()
2. Import the csv file
3. View the file
4. Print number of rows and columns – Hint: dim()
5. Print columns names
6. Change column names to lower case so it is easier to use Hint:
names(df_name) <- tolower(names(df_name))
7. Explore the variable types. – Hint: str()
8. How many different conferences are there?
9. Let's look at the difference in values of first two columns:
 - a. Compute a new vector called "diff" and calculate the difference in rank and previous

School of Computer Science, Engineering and Technology

	<p>b. Print count and list of schools that changed 3 or more places Hint: create subset that satisfies criteria</p> <p>10. Import the GDP dataset and compute the difference in GDP between 2007 and 2017 for each country.</p> <p>Download: http://becomingvisual.com/rfundamentals/gdp.csv</p> <p>a. Create a subset of countries that saw an increase of over one trillion dollars.</p>
P4	<p>Create a RMarkdown (.Rmd) document that answers and addresses the following requirements.</p> <ol style="list-style-type: none"> Getting to know the data: <ol style="list-style-type: none"> Import the data (http://becomingvisual.com/rfundamentals/winter_olympic.csv) View the data Look at column names Look at dimension of data (rows and columns) Data is currently sorted by Rank. Sort data by total medals and country. Assign sorted data to a new data frame. Call it sort_total. Use describe() function to look at data. <ol style="list-style-type: none"> If function does not work, first import library: library(Hmisc) Look at some statistics <ol style="list-style-type: none"> What is median of number of gold, silver, bronze and total medals? Also look at the mean and total number of G, S, B and T medals More statistics <ol style="list-style-type: none"> For Gold, look at summary stats, including: IQR, min, max, mean, var, sd, skew Use summary() and describe(). (May need to install library(psych)) More statistics - subset <ol style="list-style-type: none"> Redo above statistics, this time group by Region Which region won the highest mean total medals? How many countries are in this Geographic Region? How many countries are in the EUROPE group?

School of Computer Science, Engineering and Technology

	<p>e. What is the max number of medals won? What country won the max?</p> <p>7. More statistics – correlations</p> <p>a. Explore correlations between Total medals and number of Gold and Bronze</p> <p>b. What is the correlation between Rank and Total medals? Is this expected or surprising?</p> <p>8. Import the GDP dataset and compute the measures of central tendency for 2017. (Divide by a trillion, and use na.rm = TRUE when computing the measures.)</p> <p>a. Find the mean</p> <p>b. Find the median</p> <p>c. Find the range</p> <p>d. Find the quantile</p>
P5	<p>In an R Markdown document, complete the following with the movies.csv data. Download the data from http://becomingvisual.com/rfundamentals/movies.csv</p> <p>1. Getting to know the data</p> <p>a. Import the data</p> <p>b. View the data</p> <p>c. Look at column names</p> <p>d. Look at dimension of data (rows and columns)</p> <p>2. Scatterplots</p> <p>a. Do scatter plot of Tickets Sold and Gross (Is the trend expected?)</p> <p>b. Redo scatter plot, adjusting scales, divide by 1000</p> <p>c. Redo scatter plot, adjusting scales, divide by 100,000</p> <p>d. Redo scatter plot, adjusting scales, divide by 1,000,000</p> <p>3. What is the correlation between tickets sold and sales? Is this expected?</p> <p>4. Scatterplots with lines</p> <p>a. Do scatter plot with millions scale, add a regression line</p> <p>b. Add label to x and y axis, add plot title label</p> <p>5. Other plots</p> <p>a. Do boxplot</p> <p>b. Do boxplot - horizontal</p>

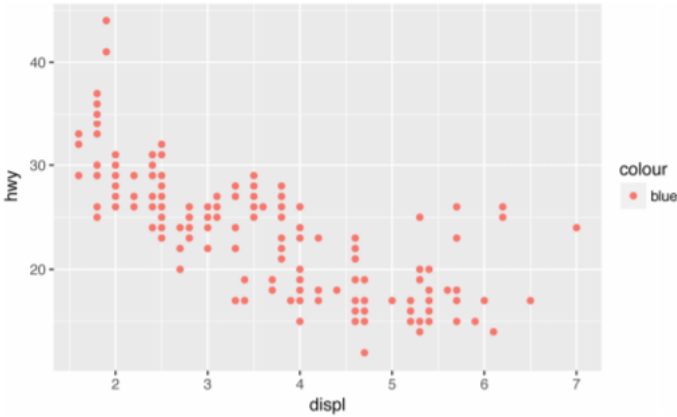
School of Computer Science, Engineering and Technology

	<ul style="list-style-type: none"> c. Do histogram for type of films d. Do histogram of gross sales. How bins are shown by default? e. Do histogram of gross sales with 10 bins. f. Do histogram of ticket sales. Try different bin numbers. g. Do histogram of ticket sales (use millions unit). Add frequency count to top of bars. Add titles. h. Do barplot of genre <p>6. In a R Markdown document, produce plots that describe the GDP (http://becomingvisual.com/rfundamentals/gdp.csv) and Life Expectancy (: http://becomingvisual.com/rfundamentals/life_expectancy.csv) during 2016 You will need to create a new data frame with these columns.</p> <ul style="list-style-type: none"> a. Create a scatter plot of GDP to Life Expectancy b. Create a histogram of GDP c. Create a box and whisker plot of Life Expectancy
P6	<p>Create an RMarkdown document to complete the following:</p> <ul style="list-style-type: none"> 1. Getting to know the data <ul style="list-style-type: none"> a. Import the data (http://becomingvisual.com/rfundamentals/summer_winter_olympics.csv) b. View the data c. Look at column names d. Look at dimension of data (rows and columns) 2. Dealing with Data <ul style="list-style-type: none"> a. Look at the column names and change names to more meaningful names. b. The data represent, in order: <ul style="list-style-type: none"> 1. country 2. number of summer games played, gold, silver, bronze, total, 3. number of winter games played, gold, silver, bronze and total, total 4. total (Winter + Summer) games, gold, silver, bronze, total 3. Summary <ul style="list-style-type: none"> a. Use table() to find frequency of total summer games played


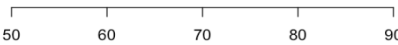
School of Computer Science, Engineering and Technology

	<p>b. Explore the data with other variables</p> <p>4. Graphs</p> <p>a. Do histogram of summer games (total)</p> <p>b. Do histogram of winter games (total)</p> <p>c. Put above two histograms on one page</p> <p>d. Do two histograms on one page: total summer, total winter medals won</p> <p>e. Is there a correlation between number of medals given out in winter and summer? (do plot)</p> <p>f. How about number of games each country competes in. Is there correlation between winter and summer?</p> <p>g. Look at distribution of each of the types of medals, by season (6 histograms on one page)</p> <p>h. Redo g with different number of bins (10 instead of 20)</p> <p>i. Explore data on your own</p> <p>5. Merge the columns for the year 2016 for GDP (http://becomingvisual.com/rfundamentals/gdp.csv), Life Expectancy (http://becomingvisual.com/rfundamentals/life_expectancy.csv), and Employment (http://becomingvisual.com/rfundamentals/employment.csv) into a new data frame and clean-up the new table.</p> <p>a. Rename the appropriate columns to "country", "gdp", "life_expectancy", and "employment".</p> <p>b. Convert the employment number to percentages by dividing by 100</p> <p>c. Then round life expectancy to zero decimals and employment to two decimals</p> <p>d. Create a frequency table for each variable</p> <p>e. Draw histograms for each variable</p>
P7	<p>Import the following AirBnb data set (http://becomingvisual.com/rfundamentals/airbnb.csv)</p> <p>Using a for loop count the number of AirBnbs that are in a particular neighbourhood the NYU area using the Greenwich Village and West Village neighbourhoods.</p>

School of Computer Science, Engineering and Technology

P8	<p>Import the following AirBnb data set http://becomingvisual.com/rfundamentals/airbnb.csv</p> <p>Using if / else statement, count the number of AirBnbs that are in a particular neighbourhood the NYU area using the Greenwich Village and West Village neighbourhoods.</p>
P9	<ol style="list-style-type: none"> 1. Create a function that computes the mean, median, min, and max values. Use this function to compute those values for the attitude data set. 2. Create a new function called checkforna that checks to see if a give value is NA and prints out the row number and column name from the following data set: http://becomingvisual.com/rfundamentals/airbnb.csv
P10	<p>Create a ggplot for following datasets: Data = MPG, Iris, Glass (From UCR Repository)</p> <ol style="list-style-type: none"> 1. How many rows are in each dataset 2. How many columns are in each dataset? 3. Make a scatter plot of hwy vs cyL variables from mpg. 4. <p>What's gone wrong with this code? Why are the points not blue?</p> <pre>ggplot(data = mpg) + geom_point(mapping = aes(x = displ, y = hwy, color = "blue"))</pre> 
P11	<p>SELECT all applicable data</p> <ol style="list-style-type: none"> 1. The players on the San Antonio Spurs in 2014 2. Top 5 blockers in 2010 3. Top 10 combination power-forwards with the most defensive rebounds

School of Computer Science, Engineering and Technology

	<p>4. Top 20 Player-seasons in the NBA 50-40-90 Club (players who have hit over 50% for FG%, 40% for 3P%, 90% for FT%, 300 field goals, 55 3-pointers, and 125 free throws) ordered by their amount of points</p> <p>Top 10 oldest Milwaukee Bucks players with over 1000 points.</p>
P12	<p>Using the http://becomingvisual.com/rfundamentals/nyuclasses.csv file, create a shiny app that displays a box plot of the student grades based on the assignment selection that looks like the image below:</p> <p>Assessment Dashboard</p> <div style="border: 1px solid #ccc; padding: 10px; margin: 10px 0;"> <p>Assessment:</p> <div style="border: 1px solid #ccc; padding: 2px; display: flex; align-items: center;"> Assignment 1 ▼ </div> </div> <div style="text-align: center; margin: 20px 0;">  </div> <div style="text-align: center; margin: 10px 0;">  </div> <ol style="list-style-type: none"> 1. Revise the nyuclasses app to provide a default view of most recent distributions by most recent assignment due date. 2. Revise app to include a selector by one or more students. 3. Revise app to include doughnut charts to show completion, late or incompleted assessments by assessment type.

School of Computer Science, Engineering and Technology

Assignments:

Sr. No.	Assignment Name
A1	<p>a. Which of the following is a logical operator? / - ^</p> <p>b. What value does R return in the statement below? 3 >= 4</p> <p>c. What is the result of this calculation? (45 + 3) * 43 + 3^2</p> <p>d. How would R evaluate the following? carspeed = 70 speedlimit = 65 carspeed > speedlimit</p> <p>e. How would R evaluate the following? (2+2 == 4) (2+2 == 5)</p> <p>f. How would R evaluate the following? !FALSE</p> <p>g. What is the result of this function? round(33.2321435452, 2)</p> <p>h. What is the result of this function? sqrt (64)</p> <p>i. What is the result of this statement? sqrt(64) == 64 ^.5</p> <p>j. What is the result of this statement? abs(-32)</p> <p>k. Which of the following is an arithmetic operator? *, , &, !</p> <p>l. What is wrong with this code? 2 + 3 *4 + sqrt[100]</p>
A2	<p>a. Create a vector called unemploy_rate with 12 values, one for each month in 2013. The values for each month are listed below (beginning with January's rate of 7.9)</p>

	7.9	7.7	7.5	7.5	7.5	7.5	7.3	7.2	7.2	7.2
	<p>b. Create a vector called month and add 12 values, one for the name of each month in a year.</p>									
	Jan	Feb	Mar	Apr	May	Jun	July	Aug	S	
	<p>c. Convert month to a factor variable</p> <p>d. Create a data frame called monthly_rate that is comprised of unemployment_rate and month.</p> <p>e. How would you extract the unemployment rate for March?</p> <p>f. Extract only those months where unemployment was below 7.5%.</p> <p>g. What is a factor variable? When would you want to use a factor variable?</p> <p>h. What is unique about a numeric variable?</p> <p>i. Why would you use a data frame over a vector to store your data?</p>									
A3	<p>a. Create a new R Script in RStudio named fed_stimulus.R</p> <p>b. Add comments to your script that include your name and date</p> <p>c. Go to NYC Open Data and export the Federal Stimulus dataset as a CSV file from https://data.cityofnewyork.us/Business/Federal-Stimulus-Data/ivix-m77e</p> <p>d. Review the details of the variables included in the dataset by selecting the manage button on the NYC Open Data site for the Federal Stimulus data.</p> <p>e. Move the Federal_Stimulus_Data.csv file to your mydata folder on your desktop</p> <p>f. Import the dataset in RStudio. Change the name of the data frame from Federal_Stimulus_Data to fed_stimulus</p> <p>g. Compute the sum and mean for the payment value column</p> <p>h. Create a subset of your data that returns those projects with project status is equal to the completed 50% or more. Do not include fully completed projects.</p> <p>i. Review your R Script and add appropriate explanatory comments</p> <p>j. Try creating a knitr report (you do not need to hand this in).</p>									
A4	For this assignment use a pre-loaded dataset in R named attitude.									

School of Computer Science, Engineering and Technology

	<p>This is from a survey of the clerical employees of a large financial organization, the data are aggregated from the questionnaires of the approximately 35 employees for each of 30 (randomly selected) departments. The numbers give the percent proportion of favorable responses to seven questions in each department. attitude is already pre-loaded in R. To view it, type</p> <pre>>View(attitude).</pre> <p>Create an R script that computes the measures of central tendency and measures of variability and the relationships for each of the seven variables in the attitude dataset. Use the functions below:</p> <p>mean, median, mode, max, min, range, quantile, IQR, var(), sd(), and cor()</p> <p>Check your work by using the summary and/or describe functions.</p>
A5	<p>Return to the attitude dataset. Produce at least one scatter plot, histogram, and box-and-whisker plot for each variable. Complete this as a R Markdown document.</p> <p>Optional: To save time, explore creating a matrix of histograms, a matrix of scatter plots, and a matrix of boxplots.</p>
A6	<p>Use the undergraduate survey data from http://becomingvisual.com/rfundamentals/undergrad.csv to create ordered factor variables for the excel, statistics and programming variables. In a R Markdown draw histograms for your new ordered factor variables.</p>
A7	<p>Import the following AirBnb data set (http://becomingvisual.com/rfundamentals/airbnb.csv)</p> <ol style="list-style-type: none"> Create a set of functions that compute specific metrics by neighborhood: <ol style="list-style-type: none"> average_number_of_reviews average_price Use the functions created in part 1 to compute these metrics by room type and neighborhood.
A8	<ol style="list-style-type: none"> Identify your own data set and create a shiny app that allows the user to explore it (similar to the movie explorer). Host the app on the shinyapps.io website. Submit the URL to your published app.