20IT53-R Programming Lab Viva Questions

Week-1

1. R runs on the _____ operating system.

a) Linux b) Windows
c) Ubuntu **d) Any operating system**
Ans: D

2. The primary source code copyright for R is held by the _____

a) R Foundation b) S Foundation
c) R and S foundation d) C Foundation

Explanation: The copyright for the primary source code for R is held by the R Foundation. The S foundation could not have R primary source code. R is a free software environment for statistical computing and graphics.

3. R is published under the _____ General Public License version.

a) A b) B
c) C **d) GNU**

Explanation: R language is published under the GNU General Public License version. It includes compilers for C, C++, Objective-C and a host of other languages. The following licenses are in use for R or associated software such as packages.

4. You can download "base" R system from _____

a) A
b) B
c) CRAN
d) D

Explanation: You can download from CRAN (Comprehensive R Archive Network). The Base R can be found in CRAN. The base R can be found in R studio. The base R system will be found in R studio.

5. There are more than _____ packages on CRAN that have been developed by users and programmers around the world.

a) 40
b) 400
c) 4000
d) 40000

Explanation: There are more than 4000 packages on CRAN. A package bundles together code, data, documentation, and tests, and is easy to share with others. As of January 2015, there were over 4000 packages available on the Comprehensive R Archive Network, or CRAN, the public clearing house for R packages.

6. Which of the following command is used to print an object "x" in R?
a) printf(x)
b) print(x)
c) printx
d) print[x]
Explanation: print(x) command is used to print. Print(x) is the basic syntax for R. We can directly print the variable without print function also. The functions in R are helpful to the user to simplify the problem.

7. Finally, in _____ R version 1.0.0 was released to the public.
a) 2000
b) 2005
c) 2010
d) 2012
Explanation: Source code for the entire R system is accessible to anyone who wanted to tinker with it.
8. R functionality is divided into a number of _____
a) Packages
b) Functions
c) Domains
d) Classes
Explanation: CRAN also hosts many add-on packages that can be used to extend the functionality of R.

9. Which of the following is used for Statistical analysis in R language?
a) RStudio
b) Studio
c) Heck
d) KStudio
Explanation: RStudio is a web application framework for R.

10. What will be the output of the following R program?

```
r<-0:10
r[2]
```

a) 0 **b) 1** c) 2 d) 3
Explanation: 1 is the output of the above code as indexing in R starts from 1. The output can be viewed in the R console. R studio has both R terminal and the R console. Each output format is implemented as a function in R. You can customize the output by passing arguments to the function as sub-values of the output field.

Course Instructor	Module Coordinator	HOD
(Mr.G.Rajendra)	(Dr .K.Lavanya)	(Dr. B, Srinivasa Rao)

Week 2:

1. What will be the output of the following R program?

```
y<-0:5
vector(y)
y[3]
```

a) Error in vector(y): invalid 'mode' argument
b) 1 c) 4 d) 3
Explanation: y is already vector; second line is an invalid argument. The third line will give us the output. When an R vector is printed you will notice that an index for the vector is printed in square brackets [] on the side.

2. In R language, a vector is defined that it can only contain objects of the _____
a) Same class b) Different class c) Similar class d) Any class
Explanation: A vector can only contain objects of the same class. A vector cannot have contain objects of the different class. Same class objects are used mostly. The most basic type of R object is a vector. Empty vectors can be created with the vector() function.

3. A list is represented as a vector but can contain objects of _____
a) Same class **b) Different class** c) Similar class d) Any class
Explanation: A list can contain objects of different class. But a vector can only contain objects of the same class. A vector cannot have contain objects of the different class. Same class objects are used mostly.

4. How can we define 'undefined value' in R language?
a) Inf b) Sup c) Und **d) NaN**
Explanation: NaN is used to define the "undefined" value in the R language. Undefined values also have some value in R. Missing values are denoted by NA or NaN for q undefined mathematical operations. A NaN value is also NA but the converse is not true.

5. What is NaN called?
a) Not a Number b) Not a Numeric
c) Number and Number d) Number a Numeric
Explanation: NaN is called Not a Number. It is the full form of NaN. Full forms can be viewed in R studio by typing help. A NaN value is also NA but the converse is not true. The value NaN represents an undefined value.

6. What will be the output of the following R code?

```
y <- c(TRUE, 2)
```

a) [1] "TRUE" "2" b) [1] "TRUE" 2 c) [1] "0" "2" **d) [1] 1 2**
Explanation: Here TRUE is taken as 1. Then it will give output as 1 and 2. FALSE can be taken as 0. T and F are short-hand ways to specify TRUE and FALSE.

7. Which one of the following is not a basic datatype?

a) Numeric b) Character
c) Data frame d) Integer

Answer: c

Explanation: Data frame is not the basic data type of R. Numeric, character, integer are the basic types of R. The basic data types are used many times. Data frames are used to store tabular data in R. They are an important type of object in R and are used in a variety of statistical modelling applications.

8. How could be the matrix constructed by using the following R code?

```
m <- matrix(1:6, nrow = 2, ncol = 3)
```

a) row-wise **b) column-wise**
c) any manner d) data insufficient

Answer: b

Explanation: If nothing is mentioned, matrix is created column-wise. If we want in row-wise then we have to specify. We have to mention “by row” to create a matrix in row wise. The filter( ) function is used to extract subsets of rows from a data frame. This function is similar to the existing subset( ) function.

9. Which of the following statement is alternative to _____ ?solve

a) help(solve) b) print(solve)
c) bind(solve) d) matrix(solve)

Answer: a

Explanation: help is used to get more information on any specific named function.

10. What is output of getOption(“defaultPackages”) in R studio?

a) Installs a new package **b) Shows default packages in R**
c) Error d) Nothing will print

Answer: b

Explanation: There are base packages (which come with R automatically), and contributed packages. The base packages are maintained by a select group of volunteers, called R Core.

Course Instructor	Module Coordinator	HOD
(Mr.G.Rajendra)	(Dr . K.Lavanya)	(Dr. B, Srinivasa Rao)

Week 3:

1. What will be the output of the following R code?

```
x <- c(3, 7, NA, 4, 7)
y <- c(5, NA, 1, 2, 2)
x + y
```

a) Symbol **b) Missing Data** c) 5 d) 15.5
Answer: b
Explanation: Missing data are a persistent and prevalent problem in many statistical analyses, especially those associated with the social sciences. R reserves the special symbol NA to represent missing data. Ordinary arithmetic with NA value gives NA's (addition, subtraction, etc.) and applying a function to a vector that has a NA in it will usually give a NA.

2. What is the mode of 'a' in the following R code?

```
a <- c(1," a", FALSE)
```

a) Numeric **b) Character** c) Integer d) Logical
Answer: b
Explanation: All three elements can be expressed as a character. Both paste() and cat() will printout text to the console by combining multiple character vectors together. The original data are formatted as character strings so we convert them to R's Date format for easier manipulation.

3. What is the length of b?

```
b <- 2:7
```

a) 4 b) 5 **c) 6** d) 0
Answer: c
Explanation: Length of b [1] 2 3 4 5 6 7 is 6. We can also create an empty list of a prespecified length with the vector() function. Data frames are represented as a special type of list where every element of the list has to have the same length.

4. What are the typeof(x) and mode(x) in the following R syntax?

```
x<-1:3
```

a) Numeric, Integer **b) Integer, Numeric**
c) Integer, Integer d) Numeric, Numeric
Answer: b
Explanation: Here typeof() tells about the data type. They are an important type of object in R and are used in a variety of statistical modelling applications. You can determine an object's type with the typeof function.

5. What is the function to set row names for a data frame?

a) row.names() b) colnames()
c) col.names() d) column name cannot be set for a data frame
Answer: a
Explanation: row.names() is the function to set row names for a data frame. Data frames have a special attribute called row.names, which indicate information about each row of the data frame.

6. A single element of a character vector is referred as _____
a) Character string b) String c) Data strings d) Raw data
Answer: a
Explanation: Single element of a character vector is often referred to as a character string. Dates are represented by the Date class and can be coerced from a character string using the as.Date() function. This is a common way to end up with a Date object in R.

7. R files has an extension _____
a) .R b) .S c) .Rp d) .c
Answer: a
Explanation: All R files have an extension .R. R provides a mechanism for recalling and re-executing previous commands. All S programmed files will have an extension .S. But R has many functions than S.

8. Data frames can be converted into a matrix by calling the following function data _____
a) matr() **b) matrix()** c) matrixf() d) matrixfunc()
Answer: b
Explanation: as.matrix function should be used to coerce a data frame to a matrix. It might seem that the as.matrix() function should be used to coerce a data frame to a matrix.

9. What will be the output of the following R code?

```
x <- (“a”, “b”)
as.logical(x)
```

a) a b b) “a” “b” c) 1 1 **d) Error**
Answer: d
Explanation: It is not possible to coerce, so you will get a Warning. Logical creates a logical vector of the specified length. Each element of the vector is equal to FALSE. as.logical attempts to coerce its argument to be of a logical type.

10. How to install for a package and all of the other packages on which for depends?
a) install.packages (for, depends = TRUE)
b) R.install.packages (“for”, depends = TRUE)
c) install.packages (“for”, depends = TRUE)
d) install (“for”, depends = FALSE)
Answer: c
Explanation: To install a package named for, open up R and type install.packages(“for”). To install foo and additionally install all of the other packages on which for depends, instead type install.packages (“for”, depends = TRUE).

Course Instructor	Module Coordinator	HOD
(Mr.G.Rajendra)	(Dr . K.Lavanya)	(Dr. B, Srinivasa Rao)

Week 4:

1. The longer programs are called _____

a) Files b) Structures
c) Scripts d) Data

Answer: c

Explanation: The longer programs called scripts, there is too much code to write all at once at the command prompt. Furthermore, for longer scripts, it is convenient to be able to only modify a certain piece of the script and run it again in R.

2. Scripts will run on _____

a) Script Editors b) Console
c) Terminal d) GCC Compiler

Answer: a

Explanation: script editors are designed to aid the communication and code writing process. They have all sorts of features including R syntax highlighting, automatic code completion, delimiter matching, and dynamic help on the R functions.

3. Which of the following is a "Recommended" package in R?

a) Util b) Lang c) Stats **d) Spatial**

Answer: d

Explanation: "Recommended" packages also include boot, class, cluster, codetools, foreign, KernSmooth, lattice, mgcv, nlme, rpart, survival, MASS, nnet, Matrix. There are about ten thousand packages in R now.

4. Which programming language is more based on the results?

a) R b) C c) C++ d) Java

Answer: a

Explanation: Compared to other programming languages, the R community tends to be more focussed on results instead of processes. Knowledge of software engineering best practice.

5. R is mostly used in _____

a) Problem solving b) Statistics c) Probability **d) All of the mentioned**

Answer: d

Explanation: Statistics for relatively advanced users. R has thousands of packages, designed, maintained, and widely used by statisticians. We can code ourselves if a command is not present.

6. What is the meaning of "<-"?

a) Functions
b) Loops
c) Addition
d) Assignment

Answer: d

Explanation: The expression a <- 16 creates a variable called a and gives it

the value 16 called assignment. The variable on the left is assigned to the value on the right. The left side should have only a single one.

7. R has many functions regarding _____
a) Statistics, Biotechnology
b) Probability, Microbiology
c) Distributions, Physics
d) Statistics, Probability, Distributions
Answer: d
Explanation: R has many functions for all types of mathematical objects. For example, Statistics, Probability, Distributions like Multivariate, Continuous, Simple, Discrete etc.

8. _____ hosts many add-on packages that can be used to extend the functionality of R.
a) CRAN
b) GNU
c) R studio
d) 450
Answer: a
Explanation: The primary R system is available from the Comprehensive R Archive Network, also known as CRAN. CRAN also hosts many add-on packages that can be used to extend the functionality of R.

9. The entities that R creates and manipulates are known as _____
a) objects b) task
c) container d) packages
Answer: a
Explanation: These may be variables, arrays of numbers, character strings, functions, or more general structures built from such components.

10. Collection of objects currently stored in R is called as _____
a) package **b) workspace**
c) list d) task
Answer: b
Explanation: All objects created during an R session can be stored permanently in a file for use in future R sessions.

Course Instructor	Module Coordinator	HOD
(Mr.G.Rajendra)	(Dr . K.Lavanya)	(Dr. B, Srinivasa Rao)

Week 5:

1. Give any five features of R.

1. Simple and effective programming language.
2. It is a data analysis software.
3. It gives effective storage facility and data handling.
4. It gives high extensible graphical techniques.
5. It is an interpreted language.

2. Differentiate between R and Python in terms of functionality?

For data analysis, R has inbuilt functionality, but in Python, the data analysis functionalities are not inbuilt. They are available by packages like Pandas and Numpy.

3. What are the applications of R?

There are various applications available in real-time. These applications are as follows:
Facebook, Google, Twitter, HRDAG, NDAA

4. Explain what is R?

R is data analysis software which is used by analysts, quants, statisticians, data scientists and others.

5. How can you save your data in R?

To save data in R, there are many ways, but the easiest way of doing this is Go to Data > Active Data Set > Export Active Data Set and a dialogue box will appear, when you click ok the dialogue box let you save your data in the usual way.

6. What are the data structures in R that is used to perform statistical analyses and create graphs?

R has data structures like

- Vectors, Matrices, Arrays, Data frames

7. Explain general format of Matrices in R?

General format is
Mymatrix< - matrix (vector, nrow=r , ncol=c , byrow=FALSE,
dimnames = list (char_vector_ rowname, char_vector_colnames))

8. What is the difference b/w sample() and subset() in R?

The sample() method is used to choose a random sample of size n from a dataset while the subset method is used to choose variables and observations.

9. Which function is used to read csv file? can you give some example?

Ans: read.csv() function to read a CSV file available in your current working directory.

```
data <- read.csv("input.csv")
print(data)
```

10. Which are indexed by either row or column using a specific name or number?

a) Datasets **b) Data frames** c) Data d) Functions
Answer: b
Explanation: Data frames can be indexed by either row or column using a specific name (that corresponds to either the row or column) or a number.

Course Instructor	Module Coordinator	HOD
(Mr.G.Rajendra)	(Dr . K.Lavanya)	(Dr. B, Srinivasa Rao)

Week 6:

1. What is Descriptive Statistics?
Descriptive statistics is the branch of statistics that focuses on **describing** and gaining more insight into the data in its present state. It deals with what the data in its **current state** means. It makes the data easier to understand and also gives us knowledge about the data which is necessary to perform further analysis. Average measures like mean, median, mode, etc. are a good example of descriptive statistics.

2. List various functions used for Summarizing your Data in R
R provides two very simple functions that can instantly summarize our data for us. These are the **str()** and the **summary()** functions.

3. Point out the correct statement?
a) Character strings are entered using either matching double (“) or single (‘) quotes
b) Character vectors may be concatenated into a vector by the c() function
c) Subsets of the elements of a vector may be selected by appending to the name of the vector an index vector in square brackets
d) All of the mentioned
Explanation: Character quantities and character vectors are used frequently in R.

4. What will be the output of the following R code?

```
> x <- c(1, 2, NaN, NA, 4)
> is.na(x)
```

a) FALSE FALSE TRUE TRUE FALSE
b) FALSE TRUE TRUE TRUE FALSE
c) TRUE FALSE TRUE TRUE FALSE
d) TRUE FALSE TRUE FALSE FALSE
Answer: a
Explanation: Missing values are denoted by NA or NaN for q undefined mathematical operations.

5. Lists can be created using the _____ function.
a) Matrix.li b) Matrix.lists
c) Lists.matric **d) List**
Explanation: Lists can be created using the list function. Like data frames, they can incorporate a mixture of modes into the one list and each component can be of a different length or size.

6. What should we use to access elements with a value greater than five?
a) **Subsetting commands** b) Use functions c) Packages d) Interfaces
Explanation: To access elements with a value greater than five we can use some subsetting commands and logical operators to produce the desired result.

7. The length of a list is _____ to the number of components in that list.
a) Double
b) Equal
c) Triple
d) One fourth
Explanation: The length of a list is equal to the number of components in that list. Lists can be created using the list function. Like data frames, they can incorporate a mixture of modes into the one list and each component can be of a different length or size.

8. The four most frequently used types of data objects in R are vectors, matrices, data frames and _____
a) Function
b) Lists
c) Packages
d) Interfaces
Explanation: The four most frequently used types of data objects in R are vectors, matrices, data frames and lists. A list is a generalisation of a vector and represents a collection of data objects.

9. A _____ is a set of elements appearing in rows and columns where the elements are of the same mode whether they are logical, numeric (integer or double), complex or character.
a) Vector
b) Matrix
c) Lists
d) Data frames
Explanation: A matrix is a set of elements appearing in rows and columns where the elements are of the same mode whether they are logical, numeric (integer or double), complex or character.

10. What is the simplest way of creating the vector?
a) C function
b) Create
c) Destroy
d) Invalid
Explanation: The simplest way to create a vector is through the concatenation function, c. This function binds elements together, whether they are of character form, numeric or logical.

Course Instructor	Module Coordinator	HOD
(Mr.G.Rajendra)	(Dr . K.Lavanya)	(Dr. B, Srinivasa Rao)

Week 7:

1. Point out the wrong statement?

a) matrices or more generally arrays are multi-dimensional generalizations of vectors

b) factors provide compact ways to handle categorical data

c) vectors provide a convenient way to return the results of a statistical computation

d) lists can be explicitly created using the list() function

Explanation: The dimension attribute is itself an integer vector of length 2 (number of rows, number of columns).

2. How can you add datasets in R?

rbind () function can be used add datasets in R language provided the columns in the datasets should be same.

3. How will you read a .csv file in R language?

read.csv () function is used to read a .csv file in R language. Below is a simple example –

filcontent <-read.csv (sample.csv)

print (filecontent)

4. How to read data from Excel xls or xlsx file formats into R. This can be done either by:

- ❖ copying data from Excel
- ❖ using readxl package
- ❖ or using xlsx package

5. Importing Excel files into R using readxl package

The **readxl** package, developed by Hadley Wickham, can be used to easily import Excel files (xls|xlsx) into R without any external dependencies.

Installing and loading readxl package

- Install

```
install.packages("readxl")
```

- Load

```
library("readxl")
```

6. Importing Excel files using xlsx package

The **xlsx** package, a java-based solution, is one of the powerful R packages to **read**, **write** and **format Excel files**.

Installing and loading xlsx package

- Install

```
install.packages("xlsx")
```

- Load

```
library("xlsx")
```

7. List two main functions in xlsx **package for reading both xls and xlsx Excel files:**

read.xlsx() and **read.xlsx2**() [faster on big files compared to read.xlsx function].

The simplified formats are:

```
read.xlsx(file, sheetIndex, header=TRUE)
read.xlsx2(file, sheetIndex, header=TRUE)
```

- **file**: file path
- **sheetIndex**: the index of the sheet to be read
- **header**: a logical value. If TRUE, the first row is used as column names.

8. How to read xml data in r?

In R, we can read the xml files by installing "XML" package into the R environment. This package will be installed with the help of the familiar command i.e., install.packages.

9. Point out the correct statement.

a) XLConnect package has more options for manipulating access files
b) XLConnect vignette package can also be used for manipulating excel files
c) write.xlsx write out an excel file with different argument
d) None of the mentioned
Answer: c
Explanation: write.xlsx write out an excel file with similar argument.

10. Which of the following package is used for reading excel data?

a) xlsx
b) xlsc
c) read.sheet
d) all of the mentioned
Answer: a
Explanation: read.xlsx and read.xlsx functions are part of xlsx package.

Course Instructor	Module Coordinator	HOD
(Mr.G.Rajendra)	(Dr . K.Lavanya)	(Dr. B, Srinivasa Rao)

Week 8:

1. How to implement R Pie Charts?

The Pie charts are created with the help of pie () function, which takes positive numbers as vector input. Additional parameters are used to control labels, colors, titles, etc.

There is the following syntax of the pie() function:

pie(X, Labels, Radius, Main, Col, Clockwise)

2. How to implement R Bar Chart?

In R, we can create a bar chart to visualize the data in an efficient manner. For this purpose, R provides the barplot() function, which has the following syntax:

barplot(h,x,y,main, names.arg,col)

S.No	Parameter	Description
1.	H	A vector or matrix which contains numeric values used in the bar chart.
2.	xlab	A label for the x-axis.
3.	ylab	A label for the y-axis.
4.	main	A title of the bar chart.
5.	names.arg	A vector of names that appear under each bar.
6.	col	It is used to give colors to the bars in the graph.

3. What is meant by Bar Chart in R?

A bar chart is a pictorial representation in which numerical values of variables are represented by length or height of lines or rectangles of equal width. A bar chart is used for summarizing a set of categorical data. In bar chart, the data is shown through rectangular bars having the length of the bar proportional to the value of the variable.

4. What does summary () do in R?

A very useful multipurpose function in R is summary(X), where X can be one of any number of objects, including datasets, variables, and linear models, just to name a few. When used, the command provides summary data related to the individual object that was fed into it.

5. What Is Variance?

The variance is the average of the squared differences from the mean. To figure out the variance, first calculate the difference between each point and the mean; then, square and average the results.

6. How do you implement histogram in R?

For creating a histogram, R provides hist() function, which takes a vector as an input and uses more parameters to add more functionality. There is the following syntax of hist() function:

hist(v,main,xlab,ylab,xlim,ylim,breaks,col,border)

7. How do you calculate mean?

t is calculated by taking the sum of the values and dividing with the number of values in a data series.

The function **mean()** is used to calculate this in R.

Syntax

The basic syntax for calculating mean in R is −

```
mean(x, trim = 0, na.rm = FALSE, ...)
```

Following is the description of the parameters used −

- **x** is the input vector.
- **trim** is used to drop some observations from both end of the sorted vector.
- **na.rm** is used to remove the missing values from the input vector.

8. What is meant by Median and how we calculate median in R?

Median

The middle most value in a data series is called the median. The **median()** function is used in R to calculate this value.

Syntax

The basic syntax for calculating median in R is −

```
median(x, na.rm = FALSE)
```

Following is the description of the parameters used −

- **x** is the input vector.
- **na.rm** is used to remove the missing values from the input vector.

9. What is ggplot2 () in R programming?

ggplot2 is an R package used for statistical computing and data representation using data visualization. It follows underlying graphics called Grammar of Graphics which includes certain rules and independent components which can be used to represent data in various formats.

10. What is correct about ggplot2?

ggplot2 is an open-source data visualization package for the statistical programming language R. ... ggplot2 can serve as a replacement for the base graphics in R and contains a number of defaults for web and print display of common scales. Since 2005, ggplot2 has grown in use to become one of the most popular R packages.

Course Instructor	Module Coordinator	HOD
(Mr.G.Rajendra)	(Dr . K.Lavanya)	(Dr. B, Srinivasa Rao)

Week 9:

1. How do you do a normal distribution in R?

In R, there are 4 built-in functions to generate normal distribution:

1. dnorm() dnorm(x, mean, sd)
2. pnorm() pnorm(x, mean, sd)
3. qnorm() qnorm(p, mean, sd)
4. rnorm() rnorm(n, mean, sd)

2. How do you check if the data is normally distributed in R?

Normality Test in R

1. Install required R packages.
2. Load required R packages.
3. Import your data into R.
4. Check your data.
5. Assess the normality of the data in R. Case of large sample sizes. Visual methods. Normality test.
6. Infos.

3. Is binomial distribution a normal distribution?

The normal distribution is a probability distribution for a continuous variable, while binomial distribution is a probability distribution for a discrete variable.

4. What is meant by Binomial Distribution?

The binomial distribution is also known as **discrete probability distribution**, which is used to find the probability of success of an event. The event has only two possible outcomes in a series of experiments. The tossing of the coin is the best example of the binomial distribution. When a coin is tossed, it gives either a head or a tail.

5. What is the use of Linear Regression?

Linear regression is used to predict the value of an outcome variable y on the basis of one or more input predictor variables x. In other words, linear regression is used to establish a linear relationship between the predictor and response variables.

In linear regression, predictor and response variables are related through an equation in which the exponent of both these variables is 1. Mathematically, a linear relationship denotes a straight line, when plotted as a graph.

Course Instructor	Module Coordinator	HOD
(Mr.G.Rajendra)	(Dr . K.Lavanya)	(Dr. B, Srinivasa Rao)

Week 10:

1. How do you analyse Time series data in R?

R language uses many functions to create, manipulate and plot the time series data. The data for the time series is stored in an R object called **time-series object**. It is also a R data object like a vector or data frame.

The time series object is created by using the **ts()** function.

Syntax

The basic syntax for **ts()** function in time series analysis is −

```
timeseries.object.name <-  ts(data, start, end, frequency)
```

Following is the description of the parameters used −

- **data** is a vector or matrix containing the values used in the time series.
- **start** specifies the start time for the first observation in time series.
- **end** specifies the end time for the last observation in time series.
- **frequency** specifies the number of observations per unit time.

2. What are some common *Data Preparation Operations* you would use for Time Series Data?

Answer

Some *data preparation operations* which can be used are:

- *Parsing time series* information from various sources and formats.
- Generating sequences of fixed-frequency dates and time spans.
- Manipulating and converting date times with *time zone* information.
- *Resampling* or *converting* a time series to a particular frequency.
- Performing date and time arithmetic with absolute or relative time increments.

3. What are some examples of *Time-Series Data* which can be *Mined*?

Some examples of such data are as follows:

- **Sensor data:** Sensor data is often collected by a wide variety of hardware and other monitoring devices. Typically, this data contains continuous readings about the underlying data objects. For example,

environmental data is commonly collected with different kinds of sensors that measure temperature, pressure, humidity, and so on. Sensor data is the most common form of time series data.

- **Medical devices:** Many medical devices such as an electrocardiogram (ECG) and electroencephalogram (EEG) produce continuous streams of time series data. These represent measurements of the functioning of the human body, such as the heartbeat, pulse rate, blood pressure, etc. Real-time data is also collected from patients in intensive care units (ICU) to monitor their condition.
- **Financial market data:** Financial data, such as stock prices, is often temporal. Other forms of temporal data include commodity prices, industrial trends, and economic indicators.

4. What are some real-world applications of Time-Series Forecasting?

- **Time-Series in Financial and Business Domain:** Time series analysis and forecasting essential processes for explaining the dynamic and influential behavior of financial markets. Via examining financial data, an expert can predict required forecasts for important financial applications in several areas such as risk evolution, option pricing & trading, portfolio construction, etc.
- **Time-Series in Medical Domain:** Medical instruments like *electrocardiograms*, and *electroencephalograms* are used to diagnose cardiac conditions, and measure electrical activity in the brain, respectively. These inventions made more opportunities for medical practitioners to deploy time series for medical diagnosis.
- **Time-Series in Astronomy:** Being specific in its domain, astronomy hugely relies on plotting objects, trajectories, and accurate measurements, and due to the same, astronomical experts are proficient in time series in calibrating instruments and studying objects of their interest. In the past century, time series analysis was used to discover variable stars that are used to surmise stellar distances, and observe transitory events such as supernovae to understand the mechanism of the changing of the universe with time.

5. How do you represent spatial data in R?

We can represent spatial data as discrete locations (points, lines or polygons) or as a grid of values rendered on a map as pixels. We typically represent the former type of data (discrete locations) as *vector* data, with an associated geometry or shape, and some attributes with information about the locations. Examples are:

- state boundaries with state name and population
- rivers with their flow volume and names
- polygons of watersheds with their names and associated landscape information

Course Instructor	Module Coordinator	HOD
(Mr.G.Rajendra)	(Dr . K.Lavanya)	(Dr. B, Srinivasa Rao)

Week 11:

Introduction Dirty data problems: Missing values, data manipulation, duplicates, forms of data dates, outliers, spelling.

1. What is a dirty dataset?

Dirty data, also known as rogue data, are **inaccurate, incomplete or inconsistent data**, especially in a computer system or database. ... They can be cleaned through a process known as data cleansing.

2. Why do we analyse data?

The basic subject of this book is data analysis, so it is useful to begin by addressing the question of why we might want to do this. There are at least three motivations for analysing data:

- To understand what has happened or what is happening;
- To predict what is likely to happen, either in the future or in other circumstances we haven't seen yet;
- To guide us in making decisions.

3. List Typical Data Quality Problems

Typical solved data quality problems in databases or information systems using data wrangling are:

- Not maintained attributes;
- Abuse of attributes for additional information;
- Incorrect data caused by incorrect input, including, e.g., wrong reading, etc.;
- Typing error;
- Inaccurate data;
- Missing data;
- Redundant and inconsistent data;
- Various incorrect formats;
- Duplicate records;

- Outdated information.

Course Instructor	Module Coordinator	HOD
(Mr.G.Rajendra)	(Dr . K.Lavanya)	(Dr. B, Srinivasa Rao)

Week 12:

1. Is SQLite good for relational database?

SQLite is a self-contained, file-based, and fully open-source RDBMS known for its **portability, reliability**, and strong performance even in low-memory environments. Its transactions are ACID-compliant, even in cases where the system crashes or undergoes a power outage.

2. How tibbles differ from data.frame

There are a number of differences between tibbles and data.frames. To see a full vignette about tibbles and how they differ from data.frame, you'll want to execute `vignette("tibble")` and read through that vignette. However, we'll summarize some of the most important points here:

- **Input type remains unchanged** - data.frame is notorious for treating strings as factors; this will not happen with tibbles
- **Variable names remain unchanged** - In base R, creating data.frames will remove spaces from names, converting them to periods or add "x" before numeric column names. Creating tibbles will not change variable (column) names.
- **There are no `row.names()` for a tibble** - Tidy data requires that variables be stored in a consistent way, removing the need for row names.
- **Tibbles print first ten rows and columns that fit on one screen** - Printing a tibble to screen will never print the entire huge data frame out. By default, it just shows what fits to your screen.

3. What is a SAS file?

A file with the SAS file extension is a **Statistical Analysis Software (SAS) file**. The SAS file is an ASCII (text) file that contains a series of SAS functions that may be run against a data set, or a SAS file may contain the actual data set.

4. How to Read SPSS in R?

We use the read_sav()function to open a SPSS file. The file extension ".sav"

```
PATH_spss <-
'https://github.com/guru99-edu/R-Programming/blob/master/binary.sav?r
aw=true'
df <- read_sav(PATH_spss)
head(df)
```

Course Instructor	Module Coordinator	HOD
(Mr.G.Rajendra)	(Dr . K.Lavanya)	(Dr. B, Srinivasa Rao)