Full Solution - Skilltest on R for Data Science

R (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/R/)

Introduction

R is the most commonly used tool in analytics industry today. No doubt, python is catching up quickly. Many companies which were heavily reliant on SAS, have now started R in their day to day analysis. Since R is easy to learn, your proficiency in R can be a massive advantage to your

candidature.

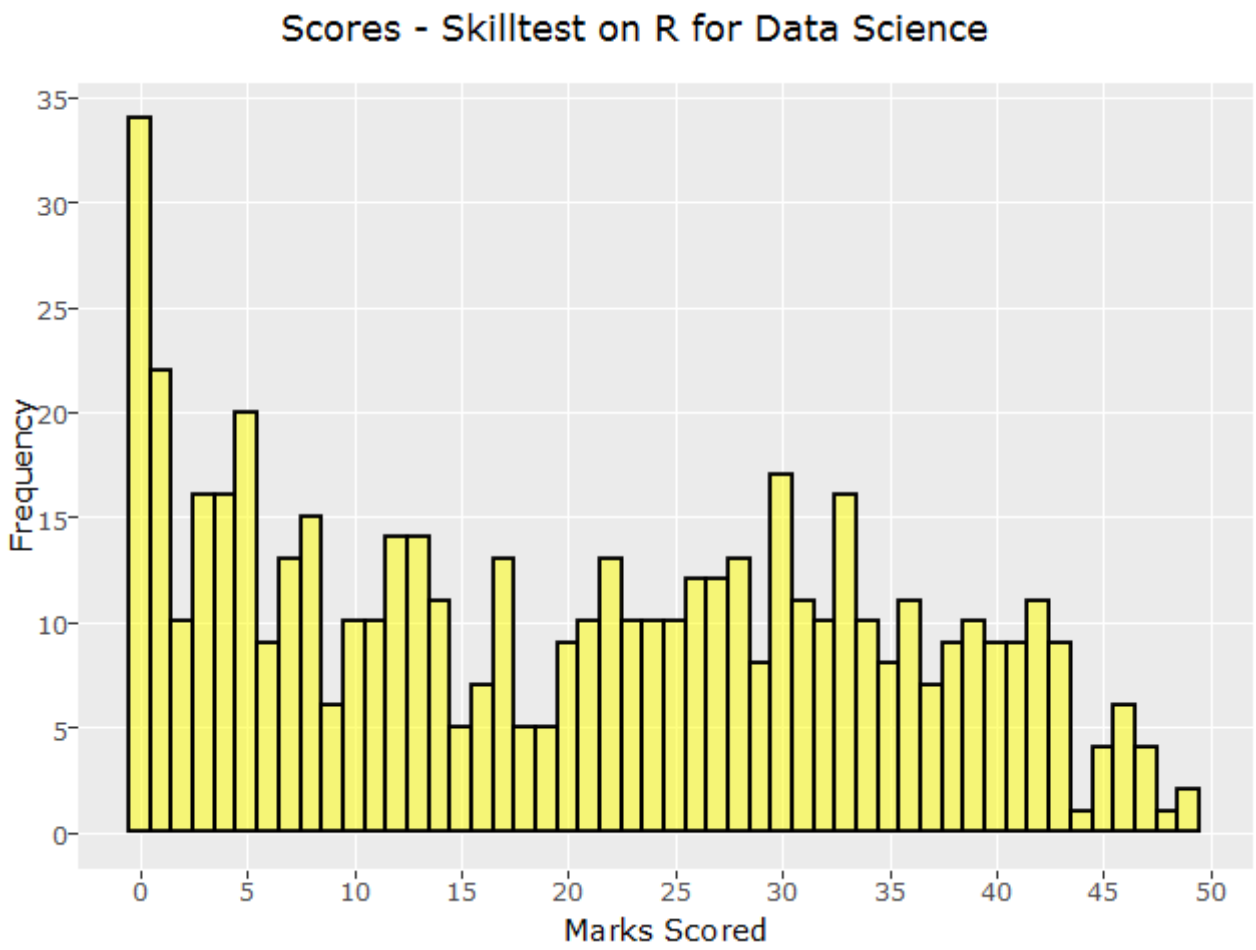
This test wasn't designed for freshers. But, for people having some knowledge of R. If you've taken this test thoroughly, you might be either disappointed or happy with your performance and keen to know the solutions. As expected, we've complied the list of Q&A so that you can learn and improve.

A best way to learn is to solve these questions at your end. You'll learn multiple ways to perform a task in R. In other words, you'll be able to add more weapons to your R armory.

If you don't understand anything, drop your question in comments!

Overall Results

Below are the distribution of the scores. This will help you to evaluate your performance.



Some of the interesting statistics from this competition:

Mean – 20.16

Median – 20

Mode – 0

Range – 49

Standard Deviation – 14.09

95% Confidence Interval – [-7.45,47.77]

Heartiest Congratulations to participants who have scored 32 & above, they are in the top 25 percentile. And, people scoring more than 40 are in top 10 percentile, score 47 & above makes you in top 1 percentile.

Due to wide range, the confidence interval doesn't seem so practical mathematically. Looks like many participants didn't take the complete test and left in between.

Since majority of the questions were fairly easy, if you have scored less than 20, you are in an alarming situation. You need to spend more time practicing on R.

Helpful Resources on R

- Complete Tutorial on R (https://www.analyticsvidhya.com/blog/2016/02/complete-tutorial-learn-data-science-scratch/)
- Summarizing Data in R (https://www.analyticsvidhya.com/blog/2015/12/7-important-ways-summarise-data/)
- Data Manipulation in R (https://www.analyticsvidhya.com/blog/2015/12/faster-data-manipulation-7-packages/)
- Missing Values Treatment in R (https://www.analyticsvidhya.com/blog/2016/03/tutorial-powerful-packages-imputing-missing-values/)

Skill Test Questions & Answers

1). Two vectors X and Y are defined as follows – X <- c(3, 2, 4) and Y <- c(1, 2). What will be output of vector Z that is defined as Z <- X*Y

A – 3,4,0

B – 3,4,4

C – error

D – 3,4,8

Solution: B

Vector recycling takes place when 2 vectors of unequal lengths are multiplied.

2). If you want to know all the values in c (1, 3, 5, 7, 10) that are not in c (1, 5, 10, 12, 14). Which code in R can be used to do this?

A – setdiff(c(1,3,5,7),c(1,5,10,12,14))

B – diff(c(1,3,5,7),c(1,5,10,12,14))

C – unique(c(1,3,5,7),c(1,5,10,12,14))

D – None of the Above.

Solution: A

setdiff() function finds the values which are different in any given two vectors.

3). What is the output of f(2) ?

```
b <- 4
f <- function (a) {
b <- 3
b^3 + g (a)
}

g <- function (a)  {
a*b
}
```

A – 33

B – 35

C – 37

D – 31

Solution: B

g(a) uses `b <- 4` because it is globally available. Globally means to every variable in the environment. f(a) uses b <- 3 because it is locally available for the function. Therefore, for a function locally available information takes precedence over global information.

4) The data shown below is from a csv file. Which of the following commands can read this csv file as a dataframe into R?

Male	25.5	0
Female	35.6	1
Female	12.03	0
Female	11.30	0
Male	65.46	1

Table1.csv

A – read.csv("Table1.csv")

B – read.csv("Table1.csv",header=FALSE)

C – read.table("Table1.csv")

D – read.csv2("Table1.csv",header=FALSE)

Solution: B

Since the table has no headers, it is imperative to specify it in the read.csv command.

5). The missing values in the data shown from a csv file have been represented by '?'. Which of the below code will read this csv file correctly into R?

A	10	Sam
B	?	Peter
C	30	Harry
D	40	?
E	50	Mark

Table2.csv

A – read.csv("Table2.csv")

B – read.csv("Table2.csv",header=FALSE,strings.na="?")

C – read.csv2("Table2.csv",header=FALSE,sep=",",na.strings="?")

D – read.table("Table2.csv")

Solution: C

Since missing values comes in many forms and not just standard NA, it is essential to define by what character the NA values are represented. `na.strings` will tell `read.csv` to treat every question mark `?` as a missing value.

6). The table shown below from a csv file has row names as well as column names. This table will be used in the following questions:

Which of the following code can read this csv file properly into R?

	Column 1	Column 2	Column 3
Row 1	15.5	14.12	69.5
Row 2	18.6	56.23	52.4
Row 3	21.4	47.02	63.21
Row 4	36.1	56.63	36.12

Table3.csv

A – read.delim("Train3.csv",header=T,sep=",",row.names=1)

B – read.csv2("Train3.csv",header=TRUE,row.names=TRUE)

C – read.table("Train3.csv",header=TRUE,sep=",")

D – read.csv("Train3.csv",row.names=TRUE,header=TRUE,sep=",")

Solution: A

Since the first column has row names, it is important to specify it using `row.names` while loading data. `row.names = 1` says that row names are available in the first column of the table.

7). Which of the following code will fail to read the first two rows of the csv file?

	Column 1	Column 2	Column 3
Row 1	15.5	14.12	69.5
Row 2	18.6	56.23	52.4
Row 3	21.4	47.02	63.21
Row 4	36.1	56.63	36.12

Table3.csv

A – read.csv("Table3.csv",header=TRUE,row.names=1,sep=",",nrows=2)

B – read.csv("Table3.csv",row.names=1,nrows=2)

C – read.delim2("Table3.csv",header=T,row.names=1,sep=",",nrows=2)

D – read.table("Table3.csv",header=TRUE,row.names=1,sep=",",skip.last=2)

Solution- D

Except D, rest all the options will successfully read the first 2 lines of this table. `nrows` parameter helps to determine how many rows from a data set should be read.

8). Which of the following code will read only the second and the third column into R?

	Column 1	Column 2	Column 3
Row 1	15.5	14.12	69.5
Row 2	18.6	56.23	52.4
Row 3	21.4	47.02	63.21
Row 4	36.1	56.63	36.12

Table3.csv

A – read.table("Table3.csv",header=T,row.names=1,sep=",",colClasses=c("NULL",NA,NA))

B – read.csv("Table3.csv",header=TRUE,row.names=1,sep=",",colClasses=c("NULL","NA","NA"))

C – read.csv("Table3.csv",row.names=1,colClasses=c("Null",na,na))

D – read.csv("Table3.csv",row.names=T, colClasses=TRUE)

Solution: A

You can skip reading columns using NULL in `colclasses` parameter while reading data.

9). Below is a data frame which has already been read into R and stored in a variable named dataframe1.

Which of the below code will produce a summary (mean, mode, median etc if applicable) of the entire data set in a single line of code?

	V1	V2	V3
1	Male	12.5	46
2	Female	56	135
3	Male	45	698
4	Female	63	12
5	Male	12.36	230
6	Male	25.23	456
7	Female	12	457

Dataframe 1

A – summary(dataframe1)

B – stats(dataframe1)

C – summarize(dataframe1)

D – summarise(dataframe1)

Solution:A

10) dataframe2 has been read into R properly with missing values labelled as NA. This dataframe2 will be used for the following questions:

Which of the following code will return the total number of missing values in the dataframe?

A	10	Sam
B	NA	Peter
C	30	Harry
D	40	NA
E	50	Mark

dataframe2

A – table(dataframe2==NA)

B – table(is.na(dataframe2))

C – table(hasNA(dataframe2))

D – which(is.na(dataframe2)

Solution: B

11). Which of the following code will not return the number of missing values in each column?

A	10	Sam
B	NA	Peter
C	30	Harry
D	40	NA
E	50	Mark

dataframe2

A – colSums(is.na(dataframe2))

B – apply(is.na(dataframe2),2,sum)

C – sapply(dataframe2,function(x) sum(is.na(x))

D – table(is.na(dataframe2))

Solution: D

Rest of the options will traverse through every column to calculate and return the number of missing values per variable.

12). The data shown below has been loaded into R in a variable named dataframe3. The first row of data represent column names. The powerful data manipulation package 'dplyr' has been loaded. This data set will be used in following questions:

Which of the following code can select only the rows for which Gender is Male?

Gender	Marital Status	Age	Dependents
Male	Married	50	2
Female	Married	45	5
Female	Unmarried	25	0
Male	Unmarried	21	0
Male	Unmarried	26	1
Female	Married	30	2
Female	Unmarried	18	0

dataframe3

A – subset(dataframe3, Gender="Male")

B – subset(dataframe3, Gender=="Male")

C – filter(dataframe3,Gender=="Male")

D – option 2 and 3

Solution: D

filter function comes from dplyr package. subset is the base function. Both does the same job.

13). Which of the following code can select the data with married females only?

Gender	Marital Status	Age	Dependents
Male	Married	50	2
Female	Married	45	5
Female	Unmarried	25	0
Male	Unmarried	21	0
Male	Unmarried	26	1
Female	Married	30	2
Female	Unmarried	18	0

dataframe 3

A – subset(dataframe3,Gender==”Female” & Marital Status==”Married”)

Marital Status==”Married”)

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

14). Which of the following code can select all the rows from Age and Dependents?

Gender	Marital Status	Age	Dependents
Male	Married	50	2
Female	Married	45	5
Female	Unmarried	25	0
Male	Unmarried	21	0
Male	Unmarried	26	1
Female	Married	30	2

Female	Unmarried	18	0

dataframe3

A – subset(dataframe3, select=c("Age","Dependents"))

B – select(dataframe3, Age,Dependants)

C – dataframe3[,c("Age","Dependants")]

D – All of the above

Solution: D

If you got this wrong, refer to the basics of sub-setting a data frame.

15). Which of the following codes will convert the class of the Dependents variable to a factor

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

		Age	Dependents
		50	2
		45	5
		25	0
		21	0
		26	1
Female	Married	30	2
Female	Unmarried	18	0

Dataframe 3

A – dataframe3$Dependents=as.factor(dataframe3$Dependents)

B – dataframe3['Dependents']=as.factor(dataframe3['Dependents'])

C – transform(dataframe3,Dependents=as.factor(Dependents))

D – All of the Above

Solution: D

`as.factor()` is used to coerce class type to factor.

16). Which of the following code can calculate the mean age of Female?

Gender	Marital Status	Age	Dependents
Male	Married	50	2
Female	Married	45	5
Female	Unmarried	25	0
Male	Unmarried	21	0
Male	Unmarried	26	1
Female	Married	30	2
		18	0

Dataframe3

e")%>%summarise(mean(Age))

me3$Gender=="Female")])

Female)

Solution: D

Option A describes the method using dplyr package. Option B uses the base functions to accomplish this task.

17). The data shown below has been read into R and stored in a dataframe named dataframe4. It is given that Has_Dependents column is read as a factor variable. We wish to convert this variable to numeric class. Which code will help us achieve this?

Gender	Marital Status	Age	Has_Dependents
Male	Married	50	0
Female	Married	45	1
Female	Unmarried	25	0
Male	Unmarried	21	0
Male	Unmarried	26	1
Female	Married	30	1
Female	Unmarried	18	0

dataframe4

A – dataframe4$Has Dependents=as.numeric(dataframe4$Has_Dependents)

B – dataframe4[,"Has Dependents"]=as.numeric(as.character(dataframe4$Has_ Dependents))

C – transform(dataframe4,Has Dependents=as.numeric(Has_Dependents))

l in two respective variables named Dataframe1 and

Dataframe1 Dataframe2
(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

Feature1	Feature2	Feature3
A	1000	25.5
B	2000	35.5
C	3000	45.5
D	4000	55.5

Feature1	Feature2	Feature3
E	5000	65.5
F	6000	75.5
G	7000	85.5
H	8000	95.5

Which of the following codes will produce the output as shown below?

Feature1	Feature2	Feature3

A	1000	25.5
B	2000	35.5
C	3000	45.5
D	4000	55.5
E	5000	65.5
F	6000	75.5
G	7000	85.5
H	8000	95.5

A – merge(dataframe1,dataframe2,all=TRUE)

B – merge(dataframe1,dataframe2)

C – merge(dataframe1,dataframe2,by=intersect(names(x),names(y))

D – None of the above

both the data sets, and even if there is no match found for

create a new column named Size(MB) from the existing
ed in a variable named dataframe5. Given 1MB = 1024KB

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

Package Name	Creator	Size(kB)
Swirl	Sean Kross	2568
Ggplot	Hadley Wickham	5463
Dplyr	Hadley Wickham	8961
Lattice	Deepayan Sarkar	3785

dataframe5

A – dataframe5$Size(MB)=dataframe$Size(KB)/1024

B – dataframe5$Size(KB)=dataframe$Size(KB)/1024

C – dataframe5%>%mutate(Size(MB)=Size(KB)/1024)

D – Both 1 and 3

Solution: D

20). Following question will use the dataframe shown below:

Gender	Marital Status	Age	Has Dependents
Male	Married	50	0
Female	Married	45	1
Female	Unmarried	25	0
Male	Unmarried	21	0
Male	Unmarried	26	1
		30	1
		18	0

Dataframe6

nly with numerical data. In that case, categorical variables
MY variables which represent the presence or absence of
aset. From Dataframe6, after creating the dummy variable

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

Gender_Male	Gender_Female	Marital Status	Age	Has Dependents
1	0	Married	50	0
0	1	Married	45	1
0	1	Unmarried	25	0
1	0	Unmarried	21	0
1	0	Unmarried	26	1
0	1	Married	30	1
0	1	Unmarried	18	0

Which of the following commands would have helped us to achieve this?

A – dummies:: dummy.data.frame(dataframe6,names=c("Gender"))

B – dataframe6[,"Gender"] <- split(dataframe6$Gender, ifelse(dataframe6$Gender == "Male",0,1))

C – contrasts(dataframe6$Gender) <- contr.treatment(2)

D – None of the above

Solution: A

For Option A, install and load dummies package. With its fairly easy code syntax, one hot encoding in R was never easy before.

21). We wish to calculate the correlation between column 2 and column 3. Which of the below

				Column4	Column5	Column6
			olumn3			
				54	0	Alpha
				51	1	Beta
			4	32	0	Gamma
				84	1	Delta
				32	0	Phi
Name6	Male	12	24	12	0	Zeta
Name7	Female	32	64	64	1	Sigma
Name8	Male	42	84	54	0	Mu
Name9	Male	56	112	31	1	Eta

Dataframe 7

A – cor(dataframe7$column2,dataframe7$column3)

B – (cov(dataframe7$column2,dataframe7$column3))/(sd(dataframe7$column4)*sd(dataframe7$column

C –
(cov(dataframe7$column2,dataframe7$column3))/(var(dataframe7$column4)*var(dataframe7$colum

D – All of the above

Solution: A

`cor` is the base function used to calculate correlation between two numerical variables.

22). Column 3 has 2 missing values represented as NA in the dataframe below stored in the variable named dataframe8. We wish to impute the missing values using the mean of the column 3. Which code will help us do that?

	Column1	Column2	Column3	Column4	Column5	Column6
Name1	Male	12	24	54	0	Alpha
				51	1	Beta
			4	32	0	Gamma
				84	1	Delta
			A	32	0	Phi
				12	0	Zeta
			A	64	1	Sigma
				54	0	Mu
			2	31	1	Eta

Dataframe 8
(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

A – dataframe8$Column3[which(dataframe8$Column3==NA)]=mean(dataframe8$Column3)

B – dataframe8$Column3[which(is.na(dataframe8$Column3))]=mean(dataframe8$Column3)

C –
dataframe8$Column3[which(is.na(dataframe8$Column3))]=mean(dataframe8$Column3,na.rm=TRUE)

D –
dataframe8$Column3[which(is.na(dataframe8$Column3))]=mean(dataframe8$Column3,rm.na=TRUE)

Solution: C

Option `na.rm=TRUE` says that impute the missing values by calculating the mean of all available observations.

23). Column7 contains some names with the salutations. In such cases, it is always advisable to extract salutations in a new column since they can provide more information to our predictive model. Your work is to choose the code that cannot extract the salutations out of names in Column7 and store the salutations in Column8.

	Column1	Column2	Column3	Column4	Column5	Column6	Column7
Name1	Male	12	24	54	0	Alpha	Mr.Sam
Name2	Female	16	32	51	1	Beta	Ms.Lilly
Name3	Male	52	104	32	0	Gamma	Mr.Mark
				84	1	Delta	Ms.Shae
				32	0	Phi	Ms.Ria
				12	0	Zeta	Mr.Patrick
				64	1	Sigma	Ms.Rose
				54	0	Mu	Mr.Peter
				31	1	Eta	Mr.Roose

Dataframe 9

A – dataframe9$Column8<-sapply(strsplit(as.character(dataframe9$Column7),split = "[.]"),function(x)[x[1]])

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

B – dataframe9$Column8<-sapply(strsplit(as.character(dataframe9$Column7),split = "."),function(x)[x[1]])

C – dataframe9$Column8<-sapply(strsplit(as.character(dataframe9$Column7),split = ".",fixed=TRUE),function(x)[x[1]])

D – dataframe9$Column8<-unlist(strsplit(as.character(dataframe9$Column7),split = ".",fixed=TRUE))[seq(1,18,2)]

Solution: B

`strsplit` is used to split a text variable based on some splitting criteria. Try running these codes at your end, you'll understand the difference.

24). Column 3 in the data frame shown below is supposed to contain dates in ddmmyyyy format but as you can see, there is some problem with its format. Which of the following code can convert the values present in Column 3 into date format?

	Column1	Column2	Column3	Column4	Column5	Column6	Column7
Name1	Male	12	24081997	54	0	Alpha	Mr.Sam
Name2	Female	16	30062001	51	1	Beta	Ms.Lilly
Name3	Male	52	10041998	32	0	Gamma	Mr.Mark
Name4	Female	36	17021947	84	1	Delta	Ms.Shae
				32	0	Phi	Ms.Ria
				12	0	Zeta	Mr.Patrick
				64	1	Sigma	Ms.Rose
				54	0	Mu	Mr.Peter
				31	1	Eta	Mr.Roose

Dataframe 10

lumn3),format="%d%m%Y")

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

B –as.Date(dataframe10$Column3,format="%d%m%Y")

C -as.Date(as.character(dataframe10$Column3),format="%d%m%y")

D -as.Date(as.character(dataframe10$column3),format="%d%B%Y")

Solution: A

25). Some algorithms work very well with normalized data. Your task is to convert the Column2 in the dataframe shown below into a normalised one. Which of the following code would not achieve that? The normalised column should be stored in a column named column8.

	Column1	Column2	Column3	Column4	Column5	Column6	Column7
Name1	Male	12	24081997	54	0	Alpha	Mr.Sam
Name2	Female	16	30062001	51	1	Beta	Ms.Lilly
Name3	Male	52	10041998	32	0	Gamma	Mr.Mark
Name4	Female	36	17021947	84	1	Delta	Ms.Shae
Name5	Female	45	15031965	32	0	Phi	Ms.Ria
Name6	Male	12	24111989	12	0	Zeta	Mr.Patrick
Name7	Female	32	26052015	64	1	Sigma	Ms.Rose
Name8	Male	42	18041999	54	0	Mu	Mr.Peter
Name9	Male	56	11021994	31	1	Eta	Mr.Roose

dataframe 11

1$Column2-
me11$Column2)

ame11$Column2)

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

Option A describes simply the mathematical formula for standarization i.e x – μ/σ

26). dataframe12 is the output of a certain task. We wish to save this dataframe into a csv file named "result.csv". Which of the following commands would help us accomplish this task?

	Column1	Column2	Column3	Column4	Column5	Column6	Column7
Name1	Male	12	24081997	54	0	Alpha	Mr.Sam
Name2	Female	16	30062001	51	1	Beta	Ms.Lilly

Name3	Male	52	10041998	32	0	Gamma	Mr.Mark
Name4	Female	36	17021947	84	1	Delta	Ms.Shae
Name5	Female	45	15031965	32	0	Phi	Ms.Ria
Name6	Male	12	24111989	12	0	Zeta	Mr.Patrick
Name7	Female	32	26052015	64	1	Sigma	Ms.Rose
Name8	Male	42	18041999	54	0	Mu	Mr.Peter
Name9	Male	56	11021994	31	1	Eta	Mr.Roose

dataframe 12

A – write.csv("result.csv", dataframe12)

B – write.csv(dataframe12,"result.csv", row.names = FALSE)

C – write.csv(file="result.csv",x=dataframe12,row.names = FALSE)

D – Both a and c

B – 1999

C – 1999

D – 1998

Solution: C

28). The dataset has been stored in a variable named dataframe13. We wish to see the location of all those persons who have "Ms" in their names stored in Column7. Which of the following code will not help us achieve that?

	Column1	Column2	Column3	Column4	Column5	Column6	Column7
Name1	Male	12	24081997	54	0	Alpha	Mr.Sam
Name2	Female	16	30062001	51	1	Beta	Ms.Lilly
Name3	Male	52	10041998	32	0	Gamma	Mr.Mark
Name4	Female	36	17021947	84	1	Delta	Ms.Shae
Name5	Female	45	15031965	32	0	Phi	Ms.Ria
Name6	Male	12	24111989	12	0	Zeta	Mr.Patrick
Name7	Female	32	26052015	64	1	Sigma	Ms.Rose
Name8	Male	42	18041999	54	0	Mu	Mr.Peter
Name9	Male	56	11021994	31	1	Eta	Mr.Roose

dataframe13

umn7)

lumn7, ignore.case=T)

lumn7,fixed=T)

lumn7,ignore.case=T,fixed=T)

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

In option D, we tell the function to find the match irrespective of lower or upper case i.e. it just matches the spelling the and return the output.

29). The data below has been stored in a variable named dataframe14. We wish to find and replace all the instances of Male in Column1 with Man. Which of the following code will not help us do that?

	Column1	Column2	Column3	Column4	Column5	Column6	Column7

Name1	Male	12	24081997	54	0	Alpha	Mr.Sam
Name2	Female	16	30062001	51	1	Beta	Ms.Lilly
Name3	Male	52	10041998	32	0	Gamma	Mr.Mark
Name4	Female	36	17021947	84	1	Delta	Ms.Shae
Name5	Female	45	15031965	32	0	Phi	Ms.Ria
Name6	Male	12	24111989	12	0	Zeta	Mr.Patrick
Name7	Female	32	26052015	64	1	Sigma	Ms.Rose
Name8	Male	42	18041999	54	0	Mu	Mr.Peter
Name9	Male	56	11021994	31	1	Eta	Mr.Roose

dataframe 14

A – sub("Male","Man",dataframe14$Column1)

B – gsub("Male","Man",dataframe14$Column1)

me14$Column1=="Male")]="Man"

ry option will do this task gracefully.

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

30) Which of the following command will display the classes of each column for the following dataframe ?

	Column1	Column2	Column3	Column4	Column5	Column6	Column7
Name1	Male	12	24081997	54	0	Alpha	Mr.Sam
Name2	Female	16	30062001	51	1	Beta	Ms.Lilly
Name3	Male	52	10041998	32	0	Gamma	Mr.Mark
Name4	Female	36	17021947	84	1	Delta	Ms.Shae
Name5	Female	45	15031965	32	0	Phi	Ms.Ria
Name6	Male	12	24111989	12	0	Zeta	Mr.Patrick

Name7	Female	32	26052015	64	1	Sigma	Ms.Rose
Name8	Male	42	18041999	54	0	Mu	Mr.Peter
Name9	Male	56	11021994	31	1	Eta	Mr.Roose

A – lapply(dataframe,class)

B – sapply(dataframe,class)

C – Both 2 and 3

D – None of the above

Solution: C

The only difference in the answer of lapply and sapply is that lapply will return a list and sapply will return a vector/matrix.

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

idyr package which forms an important part of the data

combine Male and Female column into a single column ble named Count as the count of male or female per

		Female
		6
B	5	9

Initial dataframe

Name	Sex	Count
A	Male	1
B	Male	5
A	Female	6
B	Female	9

Final dataframe

A – collect(dataframe,Male:Female,Sex,Count)

B – gather(dataframe,Sex,Count,-Name)

C – gather(dataframe,Sex,Count)

D – collect(dataframe,Male:Female,Sex,Count,-Name)

Solution: B

32). The dataframe below contains one category of messy data where multiple columns are stacked into one column which is highly undesirable.

Sex_Class	Count
Male_1	1
	2
	3
	4

ert the above dataframe to the dataframe below ? The l dataframe.

		Count
Female	1	1
		2
Female	1	3
Female	2	4

A – separate(dataframe,Sex_Class,c("Sex","Class"))

B – split(dataframe,Sex_Class,c("Sex","Class"))

C – disjoint(dataframe,Sex_Class,c("Sex","Class"))

D – None of the above

Solution: A

33). The dataset below suffers from a problem where variables "Term" and "Grade" are stored in separate columns which can be displayed more effectively. We wish to convert the structure of these variables into each separate variable named Mid and Final.

Name	Class	Term	Grade
Alaska	1	Mid	A
Alaska	1	Final	B
Banta	2	Mid	A
Banta	2	Final	A

Which of the following code will convert the above dataset into the one showed below? The dataframe is stored in a variable named dataframe.

Name	Class	Mid	Final
		A	B
		A	A

e)

ade)

Solution: C
(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

34). The _____ function takes an arbitrary number of arguments and concatenates them one by one into character strings.

A – copy()

B – paste()

C – bind()

D – None of the above.

Solution: B

35). Point out the correct statement :

A – Character strings are entered using either matching double (") or single (') quote.

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 above

Solution:D

wing code ?

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

2- [,1] [,2] [,3] x 1 2 3
y 10 11

3- [,1] [,2] [,3] x 1 2 3
y 4 5 6

4 – All of the above

Solution: A

37). Which of the following method make vector of repeated values?

A – rep()

B – data()

C – view()

D – None of the above

Solution: A

38). Which of the following finds the position of quantile in a dataset ?

A – quantile()

B – barplot()

ss-tabulate tables using formulas ?

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

A – table()

B – stem()

C – xtabs()

D – All of the above

Solution: D

40) What is the output of the following function?

```
>     f <- function(num = 1) {
                hello <- "Hello, world!\n"
        for(i in seq_len(num)) {
                  cat(hello)
         }
         chars <- nchar(hello) * num
         chars
 }

> f()
```

A – Hello, world!

14

13
(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)
Solution: A

41- Which is the missing value from running the quantile function on a numeric vector in comparison to running the summary function on the same vector ?

A – Median

B – Mean

C – Maximum

D – Minimum

Solution: B

42- Which of the following command will plot a blue boxplot of a numeric vector named vec?

A – boxplot(vec,col="blue")

B – boxplot(vec,color="blue")

C – boxplot(vec,color="BLUE")

D – None of the above

ill create a histogram with 100 buckets of data ?

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

C – hist(vec,breaks=100)

D – None of the above

Solution: C

44- What does the "main" parameter in barplot command does ?

A – x axis label

B – Title of the graph

C – I can't tell

D – y axis label

Solution: B

45- The below dataframe is stored in a variable named sam:

A	B
12	East
15	West
13	East
15	East
15	West

ine of code per B i.e a total of two boxplots (one for East
g command will achieve the purpose ?

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

D – None of the above

Solution: A

46- Which of the following command will split the plotting window into 3 X 4 windows and where the plots enter the window row wise.

A – par(split=c(3,4))

B – par(mfcol=c(3,4))

C – par(mfrow=c(3,4))

D – par(col=c(3,4))

Solution – C

47- A dataframe named frame contains two numerical columns named A and B. Which of the following commands will draw a scatter plot between the two columns of the dataframe?

A – with(frame,plot(A,B))

B – plot(frame$A,frame$B)

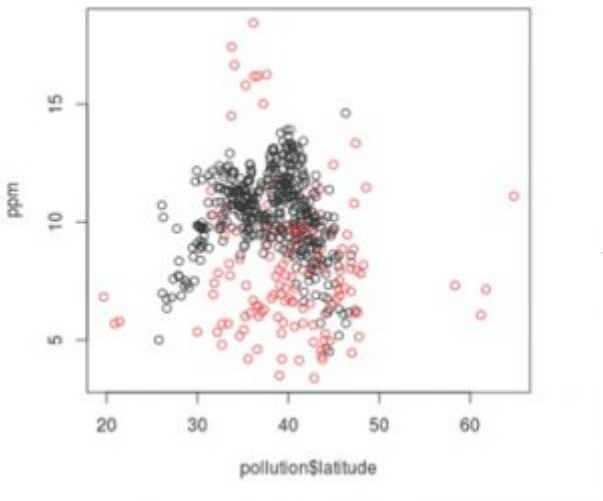
C – ggplot(data = frame, aes(A,B))+geom_point()
D – All of the above

Solution: D

variable named frame.

		C
		East
		West
		East
45	63	East
12	26	West

Which of the following command will draw a scatter plot between A and B differentiated by different color of C like the one below.



(https://www.analyticsvidhya.com/wp-content/uploads/2016/08/Q.8-Image-1-e1471953083964.jpg)

A – plot(frame$A,frame$B,col=frame$C)

B – with(frame,plot(A,B,col=C)

ly to R's base plotting system ?

started.(eg: to change margins etc)
(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

B – It is convenient and mirrors how we think of building plots and analysing data

C – starts with plot(or similar) function

D – Use annotation functions to add/modify (text, lines etc)

Solution: A

The following questions revolve around the ggplot2 package, which is the most widely used plotting package used in the R community and provides great customisation and flexibility over plotting.

50- Which of the following function is used to create plots in ggplot2 ?

A – qplot

B – gplot

C – plot

D – xyplot

Solution: A

51- What is true regarding the relation between the number of plots drawn by facet_wrap and

Solution: C

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

52- Which function in ggplot2 allows the coordinates to be flipped? (i.e x bexomes y and vice-versa) ?

A – coordinate_flip

B – coord_flip

C – coordinate_rotate

D – coord_rotate

Solution: B

53- The below dataset is stored in a variable called frame.

A	B
alpha	100
beta	120
gamma	80
delta	110

Which of the following commands will create a bar plot for the above dataset with the values in column B being the height of the bar?

t="identity")

t="bin")

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-data-probability-1)

54- The following dataframe is stored in a variable named frame and is a subset of a very popular dataset named mtcars.

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

We wish to create a stacked bar chart for cyl variable with stacking criteria being vs variable .which of the following commands will help us do this ?

A – qplot(factor(cyl),data=frame,geom="bar",fill=factor(vs))

B – ggplot(mtcars,aes(factor(cyl),fill=factor(vs)))+geom_bar()

C – All of the above

D – None of the above

Solution: C

55 – The question is same as above . The only difference is that you have to create a dodged bar chart instead of a stacked one. Which of the following command will help us do that ?

A – qplot(factor(cyl),data=frame,geom="bar",fill=factor(vs),position="dodge")

B – ggplot(mtcars,aes(factor(cyl),fill=factor(vs)))+geom_bar(position="dodge")

End Notes

I hope you had fun participating in the assessment challenge and reading this article. We tried to answer all your queries but if we still haven't cleared all your doubts , then feel free to post your questions in the comments below.

And, since it was a new thing which we tried to enrich your experience we would like to know your thoughts / suggestions / feedback on the same. This will help us serve you better and help us understand where should we improve.

Also, make sure you register in Statistics Skill Test – 2 (https://datahack.analyticsvidhya.com/contest/skilltest-statistics-ii/).

You want to apply your analytical skills and test your potential? Then participate in our Hackathons (http://datahack.analyticsvidhya.com/contest/all) and compete with Top Data Scientists from all over the world.

Share this:

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

(https://www.analyticsvidhya.com/blog/2014/09/test-level-expertise-sasrpython/)
Test your level of expertise with SAS/R/Python (https://www.analyticsvidhya.com/blog/2014/09/test-level-expertise-sasrpython/)
In "Big data"

(https://www.analyticsvidhya.com/blog/2014/05/tricky-interview-questions/)
4 Tricky R interview questions (https://www.analyticsvidhya.com/blog/2014/05/tricky-interview-questions/)
In "Business Analytics"

(https://www.analyticsvidhya.com/blog/2016/05/data-table-data-frame-work-large-data-sets/)
data.table() vs data.frame() - Learn to work on large data sets in R (https://www.analyticsvidhya.com/blog/2016/05/data-table-data-frame-work-large-data-sets/)
In "Machine Learning"

TAGS: R FOR DATA SCIENCE (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/R-FOR-DATA-SCIENCE/), SKILLTEST IN R (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/SKILLTEST-IN-R/), SOLUTION FOR SKILLTEST IN R

(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/SOLUTION-FOR-SKILLTEST-IN-R/), TEST FOR R (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/TEST-FOR-R/)

Previous Article

Solutions for Skilltest in Statistics Revealed (https://www.analyticsvidhya.com/blog/2016/08/solutions-for-skilltest-in-statistics-revealed/)

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

(https://www.analyticsvidhya.com/blog/author/kunalj/)

Author

Kunal Jain (https://www.analyticsvidhya.com/blog/author/kunalj/)

Kunal is a post graduate from IIT Bombay in Aerospace Engineering. He has spent more than 8 years in field of Data Science. His work experience ranges from mature markets like UK to a developing market like India. During this period he has lead teams of various sizes and has worked on various tools like SAS, SPSS, Qlikview, R, Python, Adobe Insight (Omniture) and Matlab.

2 COMMENTS

Rahul Suman says: (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/08/FULL-SOLUTION-SKILL-TEST-ON-R-FOR-DATA-SCIENCE/?REPLYTOCOM=115258#RESPOND)
AUGUST 29, 2016 AT 4:22 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/08/FULL-SOLUTION-SKILL-TEST-ON-R-FOR-DATA-SCIENCE/#COMMENT-115258)

It was a awesome experience and explanation is also good.some of the code is even very useful and goona help us in future in our respective fields.Thanks again to AV and Please conduct this type of test again.

Regards
Rahul

Hari Narayan Singh says: (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/08/FULL-SOLUTION-SKILL-TEST-ON-R-FOR-DATA-SCIENCE/?REPLYTOCOM=115263#RESPOND)
AUGUST 29, 2016 AT 6:12 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2016/08/FULL-SOLUTION-SKILL-TEST-ON-R-FOR-DATA-SCIENCE/#COMMENT-115263)

ease continue this.

p-login.php?
(https://datahack.analyticsvidhya.com/contest/av
action=wordpress_social_authenticate&mode=login&provider=Facebook&redirect_to=https%3A%2F%2Fww
casino-introduction-to-probability/)
solution-skill-test-on-r-for-data-
science%2F%3Futm_source%3Dfeedburner%26utm_medium%3Demail%26utm_campaign%3DFeed%253A%

Your email address will not be published.

Comment

Name (required)

Email (required)

Website

SUBMIT COMMENT

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

			Points
		nalyticsvidhya.com/user/profile/SRK)	5378
2		(https://datahack.analyticsvidhya.com/user/profile/aayushmnit)	4818
3		Nalin Pasricha (https://datahack.analyticsvidhya.com/user/profile/Nalin)	4407
4		vopani (https://datahack.analyticsvidhya.com/user/profile/Rohan Rao)	4353
5		binga (https://datahack.analyticsvidhya.com/user/profile/binga)	3371

More Rankings (http://datahack.analyticsvidhya.com/users)

POPULAR POSTS

- A Complete Tutorial to Learn Data Science with Python from Scratch (https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-

ms (with Python and R Codes)
/2015/08/common-machine-learning-algorithms/)
should know!
/2015/08/comprehensive-guide-regression/)
leling from Scratch (in R & Python)
/2016/04/complete-tutorial-tree-based-modeling-scratch-

deling in R
/2015/12/complete-tutorial-time-series-modeling/)
Forecast (with Codes in Python)
(https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/)

- Beginner's guide to Web Scraping in Python (using BeautifulSoup) (https://www.analyticsvidhya.com/blog/2015/10/beginner-guide-web-scraping-beautiful-soup-python/)
- SAS vs. R (vs. Python) – which tool should I learn? (https://www.analyticsvidhya.com/blog/2014/03/sas-vs-vs-python-tool-learn/)

FEATURED VIDEO

Mr. Kunal Jain (Analytics Vidhya) speaks at Marketing Conclave 2014

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

(http://imarticus.org/sas-online)

RECENT POSTS

(https://www.analyticsvidhya.com/blog/2016/08/full-solution-skill-test-on-r-for-data-science/)

Full Solution - Skilltest on R for Data Science (https://www.analyticsvidhya.com/blog/2016/08/full-solution-skill-test-on-r-for-data-science/)

KUNAL JAIN , AUGUST 28, 2016

(https://www.analyticsvidhya.com/blog/2016/08/solutions-for-skilltest-in-statistics-revealed/)

Solutions for Skilltest in Statistics Revealed (https://www.analyticsvidhya.com/blog/2016/08/solutions-for-skilltest-in-statistics-revealed/)

KUNAL JAIN , AUGUST 27, 2016

om/blog/2016/08/10-youtube-videos-explaining-the-

s-iot/)

ernet of Things (IoT) - Explained in Videos

om/blog/2016/08/10-youtube-videos-explaining-the-real-world-

s-iot/)

om/blog/2016/08/beginners-guide-to-topic-modeling-

(https://datahack.analyticsvidhya.com/contest/av-

in-python-intro/) duction-to-probability/)

Beginners Guide to Topic Modeling in Python (https://www.analyticsvidhya.com/blog/2016/08/beginners-guide-to-topic-modeling-in-python/)

SHIVAM BANSAL , AUGUST 24, 2016

GET CONNECTED

6,094
FOLLOWERS

18,874
FOLLOWERS
(http://www.facebook.com/Analyticsvidhya)

Email
SUBSCRIBE
(http://feedburner.google.com/fb/a/mailverify?uri=analyticsvidhya)

ABOUT US

For those of you, who are wondering what is “Analytics Vidhya”, “Analytics” can be defined as the science of extracting insights from raw data. The spectrum of analytics starts from capturing data and evolves into using insights / trends from this data to make informed decisions. Read More (http://www.analyticsvidhya.com/about-me/)

STAY CONNECTED

LATEST POSTS

(https://www.analyticsvidhya.com/blog/2016/08/full-solution-skill-test-on-r-for-data-science/)

Full Solution - Skilltest on R for Data Science (https://www.analyticsvidhya.com/blog/2016/08/full-solution-skill-test-on-r-for-data-science/)

KUNAL JAIN , AUGUST 28, 2016

(https://www.analyticsvidhya.com/blog/2016/08/solutions-for-skilltest-in-statistics-revealed/)

Solutions for Skilltest in Statistics Revealed (https://www.analyticsvidhya.com/blog/2016/08/solutions-for-skilltest-in-statistics-revealed/)

KUNAL JAIN , AUGUST 27, 2016

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)

(https://www.analyticsvidhya.com/blog/2016/08/10-youtube-videos-explaining-the-real-world-applications-of-internet-of-things-iot/)

10Real World Applications of Internet of Things (IoT) - Explained in Videos (https://www.analyticsvidhya.com/blog/2016/08/10-youtube-videos-explaining-the-real-world-applications-of-internet-of-things-iot/)

SWATI KASHYAP , AUGUST 26, 2016

(https://www.analyticsvidhya.com/blog/2016/08/beginners-guide-to-topic-modeling-in-python/)

Beginners Guide to Topic Modeling in Python (https://www.analyticsvidhya.com/blog/2016/08/beginners-guide-to-topic-modeling-in-python/)

SHIVAM BANSAL , AUGUST 24, 2016

QUICK LINKS

Home (https://www.analyticsvidhya.com/)
About Us (https://www.analyticsvidhya.com/about-me/)
Our team (https://www.analyticsvidhya.com/about-me/team/)
Privacy Policy (https://www.analyticsvidhya.com/privacy-policy/)
Refund Policy (https://www.analyticsvidhya.com/refund-policy/)
Terms of Use (https://www.analyticsvidhya.com/terms/)

TOP REVIEWS

(https://datahack.analyticsvidhya.com/contest/av-casino-introduction-to-probability/)