

R Exam Practice Questions

1. Given the cpi data frame used in our lectures the output below was generated using which of the following commands?

Country		y2011	y2010	y2009	y2008
Afghanistan:	1	Min. :1.000	Min. :1.100	Min. :1.000	Min. :1.400
Albania	: 1	1st Qu.:2.500	1st Qu.:2.425	1st Qu.:2.600	1st Qu.:2.500
Algeria	: 1	Median :3.200	Median :3.300	Median :3.300	Median :3.300
Angola	: 1	Mean :4.031	Mean :4.008	Mean :4.012	Mean :4.033
Argentina	: 1	3rd Qu.:5.100	3rd Qu.:5.075	3rd Qu.:5.100	3rd Qu.:5.125
Armenia	: 1	Max. :9.500	Max. :9.300	Max. :9.400	Max. :9.300
(Other)	:178	NA's :1	NA's :6	NA's :7	NA's :8

- a. `str(cpi[,1:5])`
 - b. `summary(cpi[,1:5])`
 - c. `length(cpi[,1:5])`
 - d. `dim(cpi[,1:5])`
 - e. `cpi[1:5,1:5]`
2. Consider a file named STAT604_DATA.csv located in the directory D:\School\TAMU\STAT604. Which command will lead R to import this dataset and save it as the variable "ex" if the return for `getwd()` is the user's C:\ Drive?
 - A. `ex <- read.csv('D:\School\TAMU\STAT604\STAT604_DATA.csv')`
 - B. `ex <- read.csv('D:\\School\\TAMU\\STAT604\\STAT604_DATA')`
 - C. `ex <- read.csv(D:\School\TAMU\STAT604\STAT604_DATA.csv)`
 - D. `ex <- read.csv("D:/School/TAMU/STAT604/STAT604_DATA.csv")`
 - E. `ex <- read.csv("D:\School\TAMU\STAT604\STAT604_DATA.csv')`

3. Given the data from the ToothGrowth data frame shown below,

```
len supp dose
4.2   VC  0.5
11.5  VC  0.5
7.3   VC  0.5
5.8   VC  0.5
6.4   VC  0.5
10.0  VC  0.5
```

which of the following would produce the following output?

```
len supp dose
7.3   VC  0.5
```

- A. ToothGrowth[2,]
- B. ToothGrowth[3,]
- C. ToothGrowth[,2]
- D. ToothGrowth[,3]
- E. ToothGrowth[2]
- F. ToothGrowth[3]
- G. None of the above

4. For the data frame USArrests shown below,

```
      Murder  Assault  UrbanPop  Rape
Alabama   13.2    236      58 21.2
Alaska    10.0    263      48 44.5
Arizona    8.1    294      80 31.0
Arkansas   8.8    190      50 19.5
California 9.0    276      91 40.6
Colorado   7.9    204      78 38.7
```

Which of the following would produce a vector of column names:

- A. col.names(USArrests)
- B. column.names(USArrests)
- C. row.names(USArrests)
- D. names(USArrests)
- E. names(USArrests, type='column')

5. For the data frame USArrests shown below,

	Murder	Assault	UrbanPop	Rape
Alabama	13.2	236	58	21.2
Alaska	10.0	263	48	44.5
Arizona	8.1	294	80	31.0
Arkansas	8.8	190	50	19.5
California	9.0	276	91	40.6
Colorado	7.9	204	78	38.7

Which of the following would produce the value "Arkansas":

- A. `col.names('USArrests')[4]`
- B. `row.names('USArrests')[4]`
- C. `row.names(USArrests)[4]`
- D. `row.names(USArrests)[3]`
- E. `names(USArrests)[3, 1]`
- F. none of the above

6. Which code produced the following output?

```
[1] "1" "2" "3" "4" "6" "8" "-6" "-7" "-8"
```

- A) `c(1:4, c(6, '8'), -(6:8))`
- B) `c(1:4, c(6, 8), -(6:8))`
- C) `c(1:4, c(6, '8'), -6:8)`
- D) `c(1, 2, 3, 4, 6, 7, -6, -7, -8)`

7. Which function(s) can be used to coerce a given data type into another data type? Choose all that apply

- A. `as.numeric`
- B. `as.factor`
- C. `as.mode`
- D. `as.character`

8. What is the output of this command:

```
cat(paste(c(1:5), c('a', 'b', 'c', 'd', 'e'), sep = 'n'), sep = ',')
```

a)

1

a,2

b,3

c,4

d,5

e

b)

1, a

2, b

3, c

4, d

5, e

c)

1na,2nb,3nc,4nd,5ne

d)

1na

2nb

3nc

4nd

5ne

9. From the following code, what will be the output?

```
r <- 3
k <- 1
x <- 0

while(k <= 11) {
  if(k %% 2 == 0) r <- r + (2*k)
  else
    r <- r - (2*k)
  x <- x + r
  k <- k + 1
}
print(r)
print(x)
```

Choices

- A. r is positive, x is positive
- B. r is positive, x is negative
- C. r is negative, x is positive
- D. r is negative, x is negative

10. If object x does not have a class attribute, which of the following functions returns the implicit class of the object (**SELECT ALL THAT APPLY**) :

- A. `implicit.class(x)`
- B. `class(x)`
- C. `str(x)`
- D. `mode(x)`
- E. `summary(x)`

11. Which of the following statements apply to “ifelse”? Select all that apply

- A. vector
- B. scalar
- C. matrix
- D. ‘else’ is optional

12. > fruits

	names	price
1	Apple	2.33
2	Pear	3.00
3	Banana	1.99
4	Orange	2.60
5	Peach	2.30

If you type fruits in your console, it will show the output as above, what's the output of the the below command? names(fruits)[2]

- A) names Price
- B) [1] 2.33 3.00 1.99 2.60 2.30
- C) "price"
- D) "Pear"
- E) 2 Pear 3

13. The modes of three individual vectors of equal length are numeric, character, and logical respectively. When using the cbind function to combine the three vectors into a matrix, the mode of the matrix will be:

- A. Numeric
- B. Character
- C. Logical
- D. List

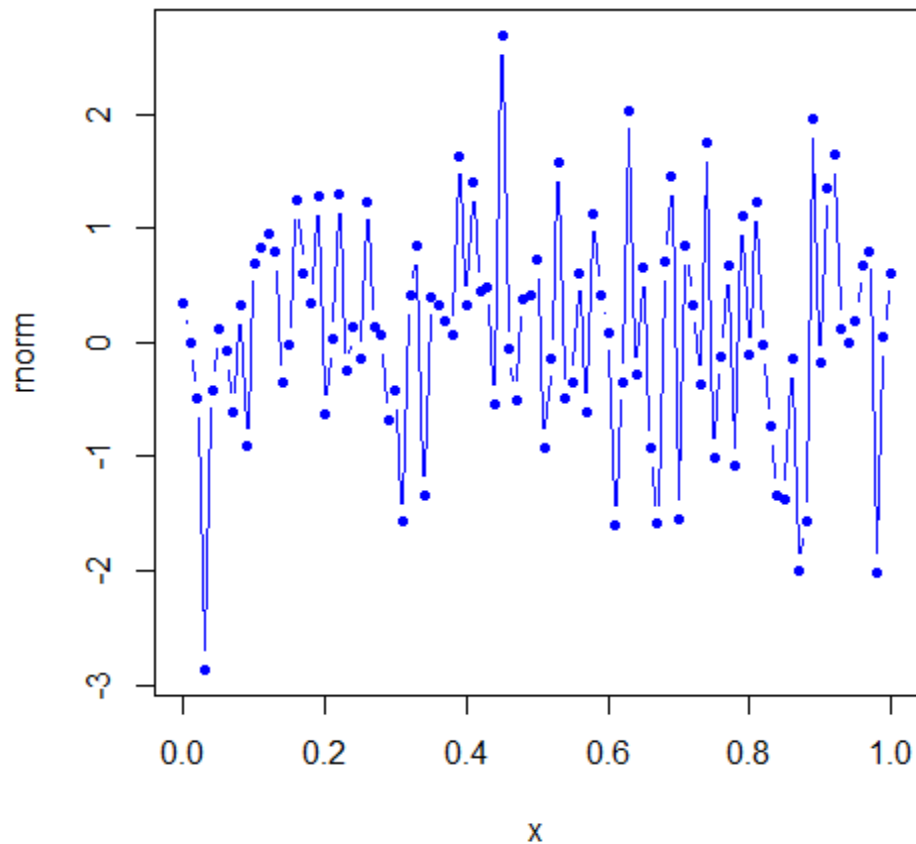
14. Which of the following would give you elements of the second and fourth row and all columns of a matrix?

- A. mat1[(2,4),]
- B. mat1(c(2,4))
- C. mat1[c(2,4),]
- D. mat1[,c(2,4)]
- E. mat1(c(2,4),)

15. You changed your working directory by click "File->Change dir->D:->STAT604". Which of the following function that you can use to verify that you are in the working directory ("D:/STAT604") you wanted?

- A. setwd()
- B. getwd()
- C. ls()
- D. objects()

16. Which code produced the graphic shown below:



- a. `plot(x=rnorm,type="b",pch=20,col="blue")`
- b. `plot(x=rnorm,type="l",pch=20,col="blue")`
- c. `plot(rnorm,type="b",pch=1,col="blue")`
- d. none of these as y-variable is missing in the plot function

17. `X <- seq(-8,10, 0.1)`

What code would you use to plot points on a normal curve with mean = 1 and standard deviation = 3?

- A. `plot(X, c(1,3), type = "p")`
- B. `Y <- dnorm(X, mean = 1, sd = 3)`
`plot(X, Y, type = "p")`
- C. `Y <- dnorm (X)`
`plot(X, Y, type = "b")`
- D. `Y <- dnorm(X, 3, 1)`
`plot(X, Y, type = "p")`

18. Which of the following does not treat a column as a vector?

- A. `FrameName$ColName`
- B. `FrameName[,col#]`
- C. `FrameName.ColName`
- D. `with(FrameName, ColName` or function using `Colname`)
- E. `FrameName[, "ColName"]`

19. Which of the following functions is used to show a list of loaded packages/libraries:

- A. `library()`
- B. `library(foreign)`
- C. `search()`
- D. `package()`
- E. `mode()`

20. Which of the following is a valid input to the `col` parameter in the following example plot command? `plot(x, y, type="p", pch=1, col=_____)`

- A. `"maroon"`
- B. `"#800000"`
- C. `palette()`
- D. All of the above

21. Which of the following parameter of `par()` has logical output by default?

- A. `pch`
- B. `omi`
- C. `ann`
- D. `cex`

22. Which of the following is NOT true about the `barplot` function in R programming?

`barplot (height, width = 1, space= NULL, names.arg = NULL, beside = FALSE, horiz = FALSE)`

- A. If *height* is a matrix, then the values in each column are stacked.
- B. If *names.arg* argument is omitted, then the names are taken from the `names` attribute of `height`.
- C. If *beside* = `FALSE`, the columns of `height` are portrayed as juxtaposed bars.
- D. If *horiz* = `FALSE`, the bars are drawn vertically with the first bar to the left.

23. Your workspace contains only a data frame called **names**. What does the function **sub(a,e, names)** perform for a data frame **names** containing multiple values?

- (a) Replaces all **e** with **a** for each value in names
- (b) Replaces all **a** with **e** for each value in names
- (c) Replaces first occurrence of **a** with **e** for each value in names
- (d) Replaces first occurrence of **a** with **e** for first value in names
- (e) None of the above

24. Which of the following functions is required in order to generate random noise with a standard normal distribution?

- A. dbinom(n)
- B. rnorm(n)
- C. rbinom(n)
- D. qbinom(n)

25. Which of the following functions is required in order to guarantee the same set of random numbers every time?

- A. set.seed(n)
- B. seq(1:n)
- C. y<- 1:n for (i in y)
- D. 1:n

26. The **table()** method in R accepts ...

- a. Exactly one object, including character strings, numeric values and lists which can be interpreted as factors
- b. Exactly one object, excluding character strings, numeric values and lists which can't be interpreted as factors
- c. One or more objects, including character strings, numeric values and lists which can be interpreted as factors
- d. One or more objects, excluding character strings, numeric values and lists which can't be interpreted as factors

27. Which of the following statement about the “grep” and “grepl” function in R is incorrect?

- A. If the “ignore.case” argument is TRUE, then the “grepl” function would be case insensitive.
- B. If you don’t specify the “value” argument, the default setting of “grep” would return indices rather than actual values.
- C. The main difference between “grep” and “grepl” functions is that “grep” is used when looking for values; “grepl” is used when looking for letters and characters.
- D. The reason that “value” argument won’t work for “grepl” function is because “grepl” function returns only TRUE or FALSE.
- E. Adding “^” in front of the pattern means that R will search only the beginning of each element in the vector.

28. How to ensure the same set of 20 random numbers are generated from a normal distribution of mean equal to five and standard deviation equal to eight?

- A. Run `rnorm(20,mean=5,sd=8)` will ensure the same set of 20 random numbers be generated every time
- B. Run `dnorm(20,mean=5,sd=8)` will ensure the same set of 20 random numbers be generated every time
- C. Run `set.seed(20)` once, and then the same set of 20 random numbers will be generated every time `rnorm(20,mean=5,sd=8)` is run
- D. Run `set.seed(20)` first before each time `rnorm(20,mean=5,sd=8)` is run

29. Which code returns the mean for the 4th and 6th columns in a 10 rows x 6 columns matrix named “M1”? Select all that apply:

- a) `apply(M1[,c(4,6)],1,mean)`
- b) `apply(M1[,4:6],2,mean)`
- c) `apply(M1[4,6],1,mean)`
- d) `apply(M1[,c(4,6)],2,mean)`

30. Choose all statements regarding Function that are incorrect.

- a. The purpose of function is to create reusable code.
- b. The components of function consist of name, inputs, code that does something, and output.
- c. You should only use parenthesis when necessary.
- d. It is not necessary to test a function if coded correctly.
- e. You should only include what can be repeated in the function.

31. Which of the following is the right answer according to this command:

```
> Num = c(14,12,24,37)
```

```
> ifelse(Num %% 3 == 0 & Num >= 20 , "3N", ifelse(Num %% 7 == 0, "7N", "NA"))
```

- A) NA 3N 7N NA
- B) 3N NA NA 7N
- C) 7N NA 3N NA
- D) NA NA NA 7N
- E) 3N 7N 3A NA

32. What is the correct syntax to add text into the top margin of a graph plot?

- A. margin(x, y, text, cex=NA, adj=NA)
- B. margintext(text, side=2 , cex=NA, adj=NA)
- C. text(x, y, text, cex=NA, adj=NA)
- D. mtext(text, side=3 , cex=NA, adj=NA)

33.

The two data frames shown below from left to right are named **data1** and **data2** respectively.

	School	Teachers
1	ACHILLE HS	7
2	ADA HS	40
3	ADAIR HS	20

	School	Teachers
1	BRAGGS HS	32
2	BRISTOW HS	23
3	BOWLEGS HS	42

Which of the following shows the result of the following code?

```
rbind(data1,data2)
```

a)

	School	Teachers
1	ACHILLE HS	7
2	ADA HS	40
3	ADAIR HS	20
4	BRAGGS HS	32
5	BRISTOW HS	23
6	BOWLEGS HS	42

c)

	School	Teachers
1	ACHILLE HS	7
2	ADA HS	40
3	BRAGGS HS	32
4	BRISTOW HS	23
5	BOWLEGS HS	42
6	ADAIR HS	20

b)

	School	Teachers
1	BRAGGS HS	32
2	BRISTOW HS	23
3	BOWLEGS HS	42
4	ACHILLE HS	7
5	ADA HS	40
6	ADAIR HS	20

d)

	School	Teachers
1	BRAGGS HS	32
2	BRISTOW HS	23
3	ACHILLE HS	7
4	ADA HS	40
5	ADAIR HS	20
31	BOWLEGS HS	42

34. Read following R script and answer which of the following would be the x in the end?

```
> x <- c("apple","ball","cat","dog","egg")
```

```
> substr(x,1,2) <- "abc"
```

```
> x
```

a) "abple" "ball" "cat" "dog" "egg"

b) "abcle" "abcl" "abc" "abc" "abc"

c) "abple" "abll" "abt" "abg" "abg"

d) "abcle" "ball" "cat" "dog" "egg"

35. If X is defined as `x<-("5,8,9,10,ad,gh")`, which command listed in the options below produces the output `chr "5,8,9,10,ad,gh"`.

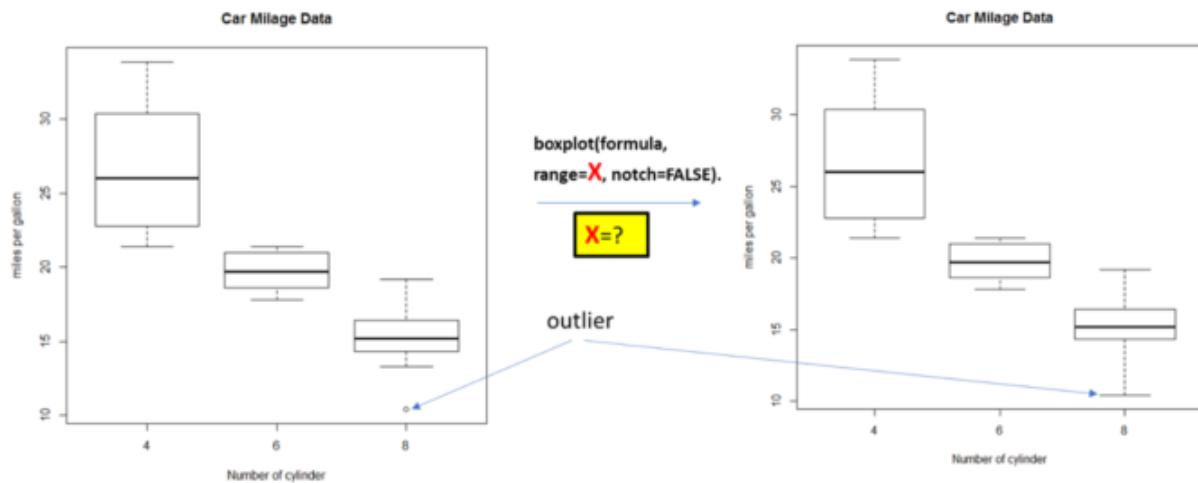
- A. `summary(x)`
- B. `mode(x)`
- C. `class(x)`
- D. `str(x)`
- E. `length(x)`

36.

A typical boxplot shown in figure below can be plotted using the following command in R.

`>boxplot(formula, range=X, notch=FALSE).`

What should be the value of X, so that the whiskers extend upto all the outliers.



- a. 1.5
- b. 2
- c. 0
- d. 1

37. I used "getwd()" to display the default path of my current working directory of the R process.

```
> getwd()
```

```
[1] "C:/Users/xx/Documents"
```

If I would like to send the output to file only in another folder ("C:/Users/xx/STAT604") as "HW07.txt", what should I type in R in order to do so?

[A] sink("C:\\Users\\xx\\STAT604\\HW07.txt", split = TRUE)

[B] Sink("C:\\Users\\xx\\STAT604\\HW07.txt ", split = FALSE)

[C] Sink("C:\\Users\\xx\\STAT604\\HW07.txt ", split = TRUE)

[D] Sink("C:\\Users\\xx\\Documents", "C:\\Users\\xx\\STAT604\\HW07.txt ", split=FALSE)

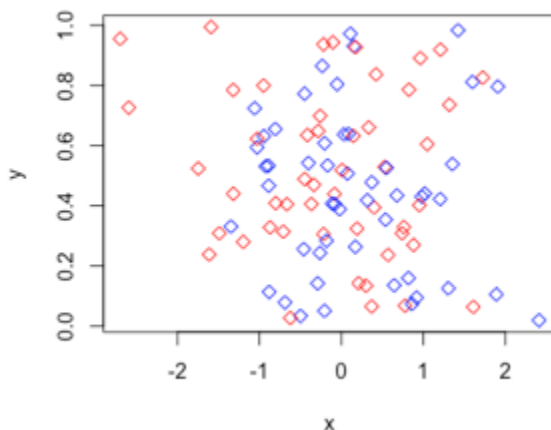
[E] sink("C:\\Users\\xx\\STAT604\\HW07.txt ", split = FALSE)

[F] sink("C:\\Users\\xx\\Documents", "C:\\Users\\xx\\STAT604", "HW07.txt ", split=TRUE)

[G] sink("C:\\Users\\xx\\STAT604", "HW07.txt ", split = TRUE)

[H] sink("C:\\Users\\xx\\STAT604", "HW07.txt ", split = FALSE)

38. Which R code produced the plot below?



```
x <- rnorm(100)
```

```
y <- runif(100)
```

A. plot(x, y, pch = 5, col = rgb(red=1, blue=1, green=0))

B. plot(x, y, pch = 5, col = c("red", "blue", "red"))

C. plot(x, y, pch = 5, col = c("red", "blue", "red", "blue"))

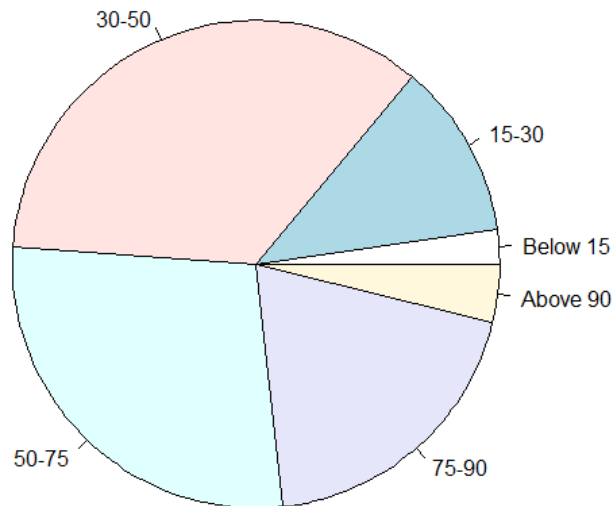
D. plot(x, y, pch = 5, col = rainbow(3))

39. If v1 is a vector given by ('Below 15','15-30','30-50','50-75','75-90','Above 90')

And v2 is another vector given by (3,15,45,36,25,5)

Which codes can produce the graphic shown below?

Pie chart of student scores



- a. `pie(v2,v1,main="Pie chart of student scores")`
- b. `pie(v2,labels=v1,clockwise=TRUE,main="Pie chart of student scores")`
- c. `pie(v2,labels=v1,clockwise=TRUE,init.angle=0,main="Pie chart of student scores")`
- d. `pie(v2,labels=v1,main="Pie chart of student scores")`

40. Complete the code that produced the results shown below.

```
i= 1 and is 6 or less.  
i= 2 and is 6 or less.  
i= 3 and is 6 or less.  
i= 4 and is 6 or less.  
i= 5 and is 6 or less.  
i= 6 and is 6 or less.  
i= 7 and is bigger than 6.  
i= 8 and is bigger than 6.  
i= 9 and is bigger than 6.  
i= 10 and is bigger than 6.
```

```
y<-1:10
```

```
{  
  if(i > 6){  
    (i=i,i,"and is bigger than 6.\n")  
  }  
  (i=i,i,"and is 6 or less.\n")  
}
```

- A. for(i in y)-paste-else-paste
- B. (i in y)-cat-ifelse-cat
- C. (i in y)-paste-ifelse-paste
- D. for(i in y)-cat-else-cat

41. Which of the following is true for object indexing?

- a. Object indexing is a function to access information in an object
- b. In object[x,y], x gives the column indices and y gives the row indices
- c. Negative index values remove specified members
- d. Object indexing only works on data frames
- e. Both c & d

42. x<- c(1:5)

```
y <- c(6:10)
```

```
x&&y
```

- a.TRUE TRUE TRUE TRUE TRUE
- b.TRUE
- c.ERROR
- d.None of the Above

43. Which function would support the argument that there are any TRUE within the command

➤ `X <- 1:20` ?

- A. `any (x > 23)`
- B. `any (x > 15)`
- C. `any (x > 28)`
- D. `any (x > 40)`

44. 100 people were surveyed about their favorite color:

Red: 20

Blue: 35

Green: 10

Purple: 25

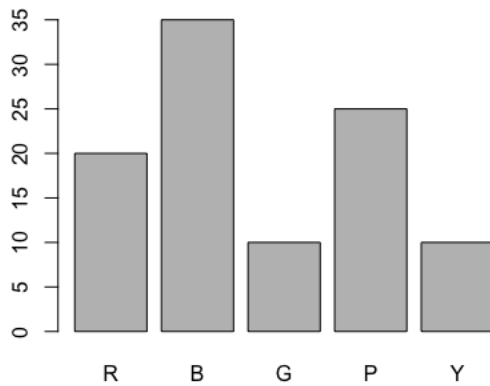
Yellow: 10

The colors and results are stored in the following vectors:

```
colors <- c(R,B,G,P,Y)
```

```
results <- c(20,35,10,25,10)
```

Which of the following commands would produce the graph below:



- a.) `barplot(results, width=1, space=NULL, names.arg=colors, beside=FALSE)`
- b.) `barplot(results, width=1, space=0, names.arg=colors, horiz=FALSE)`
- c.) `barplot(results, names.arg=colors, horiz=TRUE)`
- d.) `barplot(results, space=NULL, beside=FALSE)`

45. What will be the output for the following statements in R:

```
x < -6
```

<Output1>

```
y=8
```

```
!y
```

<Output2>

```
(x<-7 & !y)
```

<Output3>

```
x<-3
```

```
(x==y || y)
```

<Output4>

- A. Error, TRUE, TRUE, FALSE
- B. FALSE, FALSE, FALSE, FALSE
- C. TRUE, FALSE, FALSE, TRUE
- D. Error, FALSE, FALSE, TRUE

46. Choose the correct option for adding legend to a graph.

- A. legend(2000, 2000, c("Bryan", "College Station"), pch=c(1, 2), col=c("blue", "red"))
- B. legend(2000, c("Bryan", "College Station"), pch=c(1, 2), col=c("blue", "red"))
- C. legend(locator(1), c("Bryan", "College Station"), pch=c(1, 2), col=c("blue", "red"))
- D. legend(locator(1, 2000), c("Bryan", "College Station"), pch=c(1, 2), col=c("blue", "red"))

Option 1. B & D Only

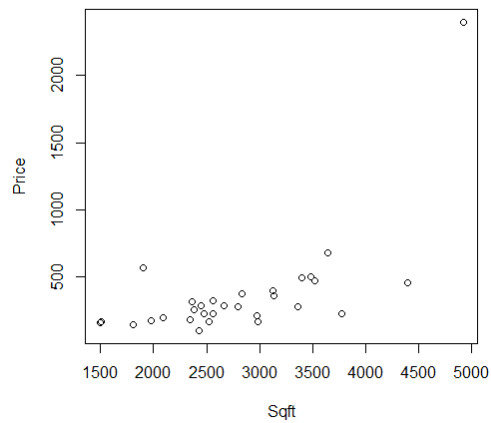
Option 2. A & D Only

Option 3. A & C Only

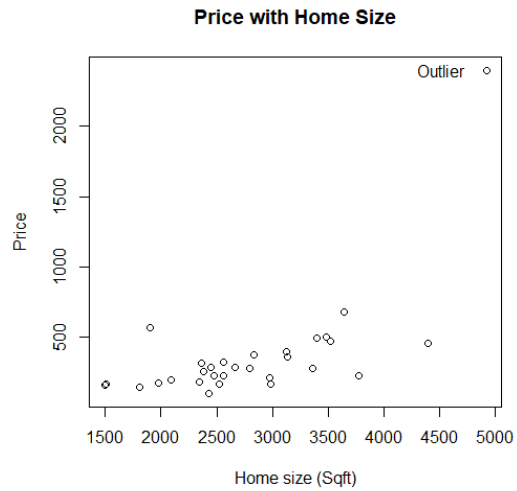
Option 4. B & D Only

47.

Original Plot



New Plot



What code produced the new plot?

- a. `plot(Price~Sqft, title= "Price with Home Size",
 xlab=Home size (Sqft))
 text(4300,2400,'Outlier', adj=0)`
- b. `plot(Price~Sqft, main= "Price with Home Size",
 xlab="Home size (Sqft)")
 text(4300,2400,'Outlier', adj=0)`
- c. `plot(Price~Sqft, main= "Price with Home Size",
 xlab="Home size (Sqft)")
 text(1500,2400, Outlier, adj=0)`
- d. `plot(Price~Sqft, title= "Price with Home Size",
 xlab=Home size (Sqft))
 text(4300,2400,'Outlier', adj=0)`

48. Identify which is **NOT** a method for accessing a row[s] as a vector (all columns) of data frames:

- [a] FrameName[-c(1:4),]
- [b] FrameName[c(1,2,3,4),]
- [c] FrameName[(1,2,3,4),]
- [d] FrameName[1:4,]

49. What is the output of the following code?

```
x<-1
while (x<8){
  print(x);
  if ((x%%2)==0){
    x<-x+2;
  } else{
    x<-x+1
  }
}
```

- a) [1] 1
[1] 2
[1] 4
[1] 6
- b) [1] 2
[1] 4
[1] 6
[1] 8
- c) [1] 1
[1] 3
[1] 5
[1] 7
- d) [1] 2
[1] 3
[1] 5
[1] 7

50. How many lines is the output when executing the code below in R and what is the value of i ?

```
> for(i in y){  
+   if(i>6) {  
+     cat("i is bigger than 6.\n")  
+   } else {  
+     cat("i is 6 or less.\n")  
+   }  
+ }
```

- A. 10 lines; $i = 10$
- B. 1 line; $i = 1$
- C. 1 line; $i = 10$
- D. Error displayed

51. > brands <- c('ford', 'gmc', 'honda', 'jeep', 'kia', 'lexus')

> brands <- toupper(brands[nchar(brands)])

> brands

What is final output of the variable "brands"?

- A. [1] "KIA" "JEEP" "LEXUS" "KIA" "JEEP" "LEXUS"
- B. [1] "JEEP" "HONDA" "KIA" "JEEP" "HONDA" "KIA"
- C. [1] "Ford" "Honda" "Gmc" "Ford" "Honda" "Kia"
- D. [1] "Jeep" "Honda" "Gmc" "Jeep" "Honda" "Gmc"

52. What can be expected as an implicit class when the object does not have a class attribute?

- A. Matrix
- B. Integer
- C. Mode(x)
- D. All of the above

53. Which of the following would give a subset of 10 items from the vector V?

- a. subset(10, V)
- b. sample(10, V)
- c. sample(V, 10)
- d. subset(V, 10)

54. Below is data frame M (on left). The data is from tab-separated file M.txt . Which of the following codes will replace the spaces with NAs and show a new data frame (on right) that excludes the rows containing NAs?

	x	y	z
1		0	160
2	33	1	
3	29	0	122
4	43		152
5	24	1	189
6	33	1	174

M

→

	x	y	z
3	29	0	122
5	24	1	189
6	33	1	174

- A. `is.na(read.delim("M.txt",header=T,na.rm=" "))`
 B. `read.delim("M.txt",header=T,na.strings=" ",na.rm=TRUE)`
 C. `replace(" ",NA),read.delim("M.txt",header=T,), na.rm=TRUE)`
 D. `na.omit(read.delim("M.txt",header=T,na.strings=" "))`
55. Consider two files that you produce in conjunction with R script execution. One file is produced by the `sink()` command: `sink("my_output.txt", split = TRUE)`. The other file shows the console from your session, which you've saved as "console.txt". What would be one major difference between these two files?
- A. "console.txt" prints the commands executed by your R script and "my_output.txt" does not.
 B. "my_output.txt" prints the commands executed by your R script and "console.txt" does not.
 C. "my_output.txt" prints comment lines written in your R script and "console.txt" does not.
 D. These files are exactly the same.
56. Which of the following will help you find out if a number is even or odd?
- a) `1295 %% 2`
 b) `1295 %/% 2`
 c) All of the above
 d) None of the above

57.

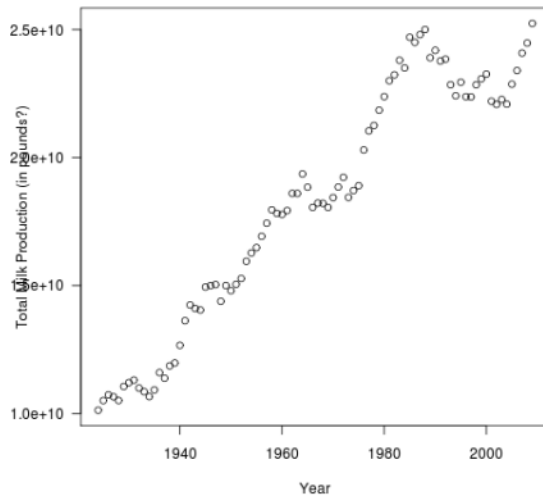


Fig left

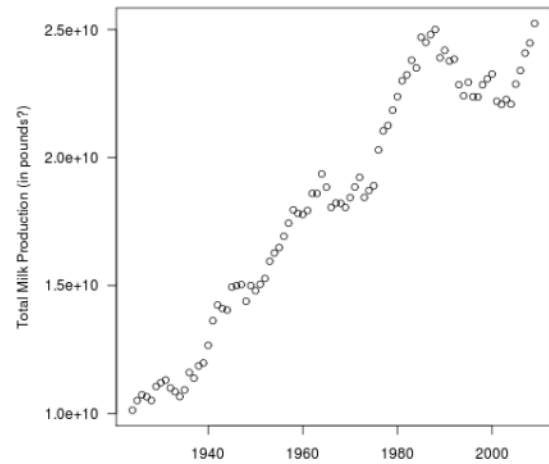


Fig right

In the left graphic, Y axis title overlaps the axis labels(Fig left), Which code produced the right graphic?

- a) `par(mar=c(5,6,4,2)+0.1)`
- b) `par(mar=c(5,4,6,2)+0.1)`
- c) `par(mar=c(5.1,6.1,4.1,2.1))`
- d) `par(mar=c(5,3,4,2)+0.1)`

58. Which of the following will you use to remove members of column 1 of the following data?

C 1	C 2	C 3	C 4	C 5
abc	34	10	11	45
xyz	24	14	12	34
mno	35	53	13	2
pqr	34	77	14	17
efg	56	89	15	16

Options:

- A. `X [remove C1]`
- B. `X [drop C1]`
- C. `X [-C1]`
- D. `X [rm(C1)]`

59.

```
R Console

> class(bcs)
[1] "data.frame"
> class(bcs$Price)
[1] "integer"
> class(bcs$Sqft)
[1] "integer"
> tm<- lm(bcs$Price~bcs$Sqft)
> summary(tm)

Call:
lm(formula = bcs$Price ~ bcs$Sqft)

Residuals:
    Min       1Q   Median       3Q      Max
-463.49 -118.47  -40.54   56.18 1327.62

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -555.09048   217.43518   -2.553  0.016420 *
bcs$Sqft      0.33092     0.07492    4.417  0.000136 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 318.9 on 28 degrees of freedom
Multiple R-squared:  0.4107,    Adjusted R-squared:  0.3896
F-statistic: 19.51 on 1 and 28 DF,  p-value: 0.0001361

> |
```

Question: From the above outputs, what is the object class of tm?

- A) Integer
- B) Data Frame
- C) lm
- D) Function

60.

```
> (x<-1:5)
[1] 1 2 3 4 5
> x[10]<-10
> |
```

Suppose you have entered the above code into R. How many of the following lines of code would return numeric values?

```
mean(x)
mean(x, na.omit=TRUE)
mean(na.rm(x))
mean(x, na.rm=FALSE)
mean(x, na.rm=TRUE)
mean(na.omit(x))
```

- a) 1
- b) 2
- c) 3
- d) 4
- e) 5

61. Which of the following is not used and is not true when data types can be coerced into other types?

- A) as.character
- B) as.numeric
- C) TRUE and FALSE coerce to 1 and 0
- D) as.list

62. Which of the following can be used for multiple graph plots per image?

Select all that apply.

- A. par(mfrow=c(nr, nc))
- B. par(mcol=c(nr, nc))
- C. par(mfcol=c(nr, nc))
- D. par(mrow=c(nr, nc))

63. In R, which method will you use to get the number of characters in a string s?

- A. nChars(s)
- B. nchars(s)
- C. nchar(s)
- D. s.nChars()

64. How can we change the entire margin of the document from figure 1 to figure 2, as shown below?

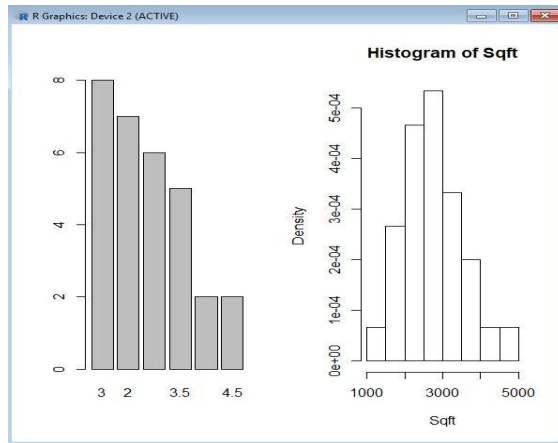


Figure 1

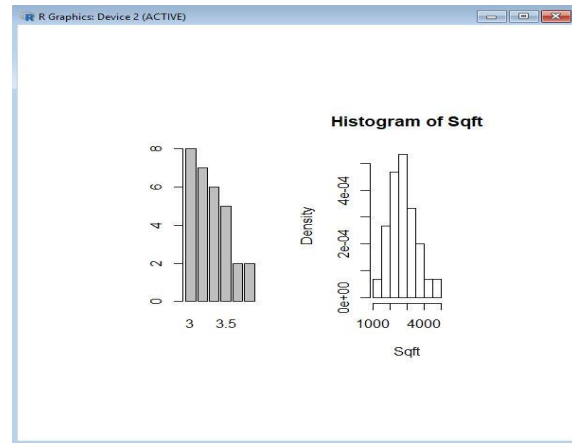


Figure 2

- a. `par(mai = c(1,1,1,1))`
- b. `par(mar = c(1,1,1,1))`
- c. `par(omi = c(1,1,1,1))`
- d. `par(oma = c(1,1,1,1))`

65. How can we set the margin of the plot document as follows?

Top: 0.75 inch, right: 0.5 inch, left: 0.25 inch and bottom: 1 inch

- a. `par(omi = c(1,.25,.75,.5))`
- b. `par(oma = c(1,.25,.75,.5))`
- c. `par(omi = c(.75,.5,1,.25))`
- d. `par(oma = c(.75,.5,1,.25))`

66. Which of the following statements is TRUE for the **cat** function?

Answer Choices

- a. Vectors are concatenated term-by-term.
- b. Linefeeds are automatically output after each element.
- c. It includes a *collapse* value.
- d. It converts its arguments to character vectors.

67. Which of the following is the correct way to display the code of function in R?

- A. `?functionname`
- B. `functionname`
- C. `??functionname`
- D. `demo(functionname)`

68. In the merge function, what does it mean if the parameter “all” is set to “TRUE”?

- a) The merge function creates a new data frame with the rows of data that are identical between data frames.
- b) The merge function creates a new data frame with the rows of data that are identical between data frames as well as every nonidentical row between data frames.
- c) The merge function creates a new data frame with the rows of data that are identical between data frames but removes columns that contain missing values.
- d) The merge function creates a new data frame with the rows of data that are identical between data frames but removes rows that contain missing values.

69. This name of this dataframe is “exam1”

	Name	level
1	Gene_A	123
2	Gene_B	231
3	Gene_C	244
4	Gene_D	121
5	miRNA_A	190
6	miRNA_B	195

The command `grep("Gene",exam1)` will have the following output:

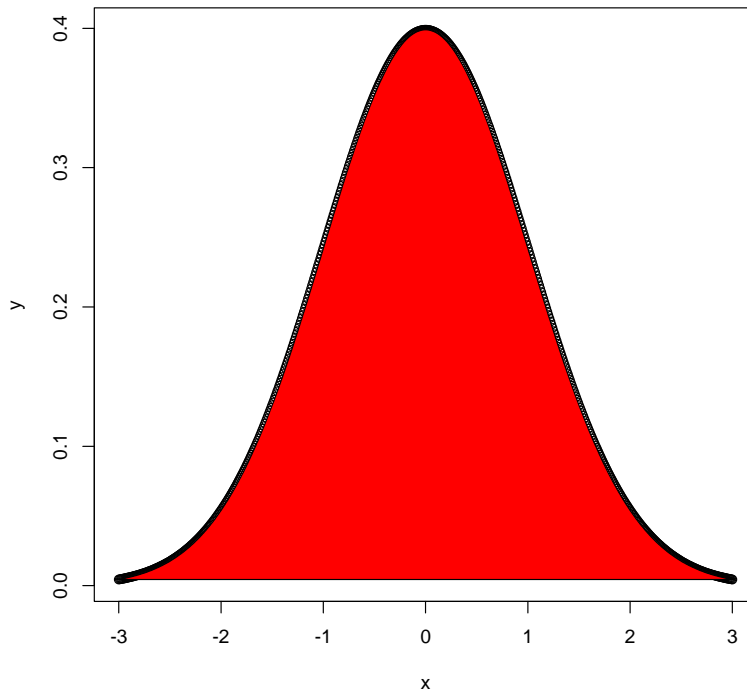
- a- A B C D
- b- 123 231 244 121
- c- 1 2 3 4
- d- Will be the data frame without only row 5 and 6

70. Which code is the proper way to create a box plot encompassing all values in whiskers

- a) `boxplot(Price ~ Location)`
- b) `boxplot(Price ~ Location, range=0)`
- c) `boxplot(Price ~ Location, range=1.5, notch=FALSE)`
- d) `boxplot(Price ~ Location, range=1.5, notch=TRUE)`

71. Which line of code successfully finds the mean of each row of a given matrix, mat1?
- A. `apply(mat1, 2, rowMeans)`
 - B. `apply(mat1, 1, sum()/dim())`
 - C. `mapply(mean, mat1)`
 - D. `apply(mat1, 1, mean)`
 - E. `apply(mat1, c(1, 2), mean)`
 - F. `apply(mat1, 2, mean)`
72. Which assignment of bool cannot be properly evaluated by a traditional "if (bool) {" statement and would require an ifelse? (Assume all data objects are defined in the workspace)
- a. `bool <- mat[1,1] > 1`
 - b. `bool <- 1:20%%2==1`
 - c. `bool <- is.na(nchar("baa"))`
 - d. `if (testscore < 80) { bool <- TRUE } else { bool <- FALSE }`
73. Which of the following is a valid command to load a previously saved workspace called 'SampleWorkspace' in R? The file location is C:\Users\Aggie\STAT604.
- A. `load(C:\\Users\\Aggie\\STAT604\\SampleWorkspace.RData)`
 - B. `load("C://Users//Aggie//STAT604//SampleWorkspace.RData")`
 - C. `load("C:\\Users\\Aggie\\STAT604\\SampleWorkspace.RData")`
 - D. `load("C:/Users/Aggie/STAT604/SampleWorkspace.RData")`
74. What is the output of the following expression?
- `1:10 %/% 2:1`
- A. 1 1 3 3 5 5 7 7 9 9
 - B. 0 0 0 0 0 0 0 0 0 0
 - C. 0 2 1 4 2 6 3 8 4 10
 - D. 1 1 3 2 5 3 7 4 9 5
75. If you want to combine different data types such as vectors and factor into matrix like structure, which one should you choose and what may happen?
- A. `cbind`
 - B. `rbinc`
 - C. `data.frame`
 - D. `merge`
 - E. `recycling`

76.



Question: Which set of R commands would result in the creation of the above graph?

- a.

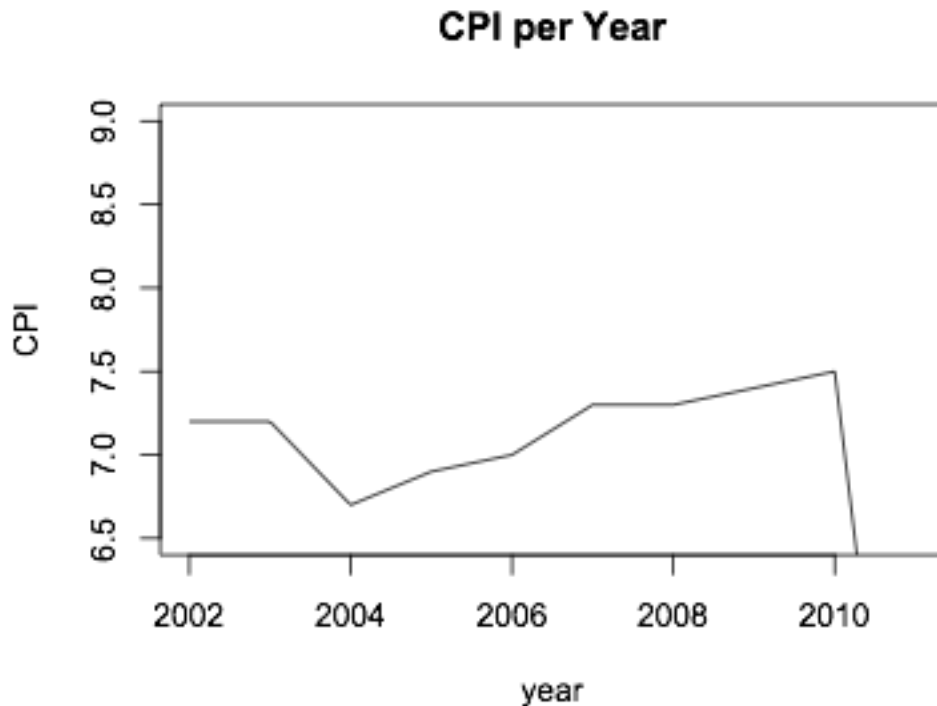
```
x<-seq(-3,3,0.01)
y<-dnorm(x)
polygon(x,y,col="red")
```
- b.

```
x<-seq(-3,3,0.01)
y<-dnorm(x)
plot(x,y,col="red")
```
- c.

```
x<-seq(-3,3,1)
y<-dnorm(x)
polygon(x,y,col="red")
```
- d.

```
x<-seq(-3,3,0.01)
y<-rnorm(x)
polygon(x,y,col="red")
```
- e. None of the above

77.



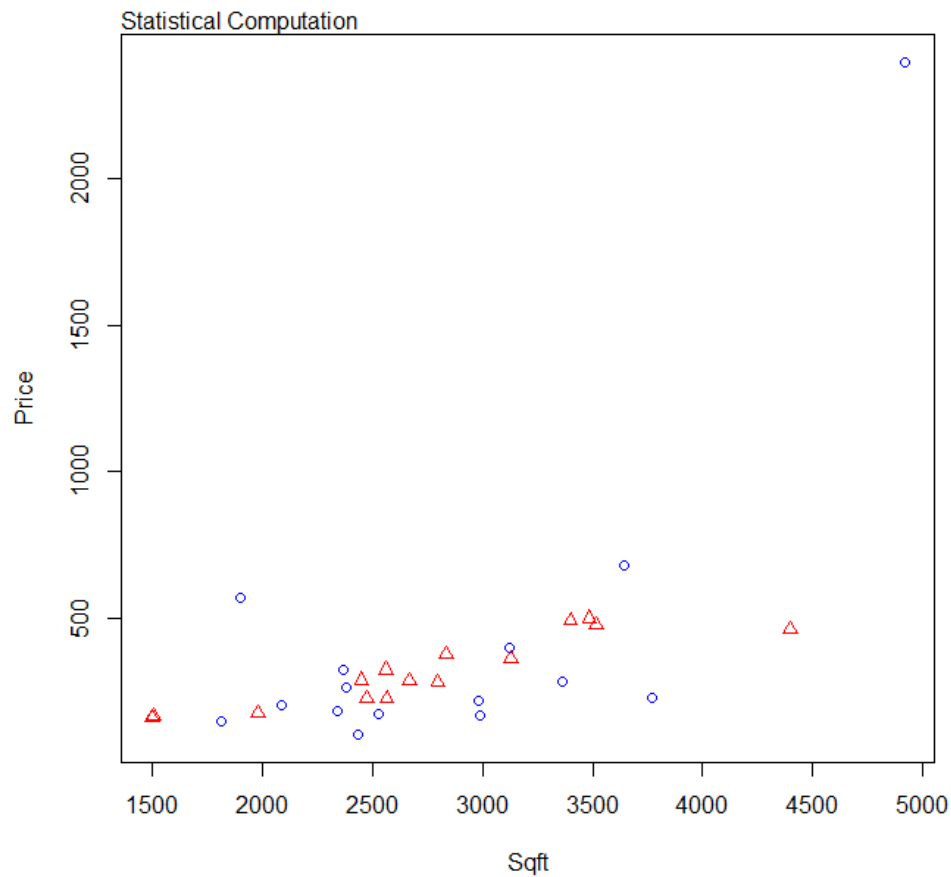
This is a graph of the CPI (Corruption Perception Index value) for Chile per year. We have the same data for Austria in a numeric vector called `cpi_value_Austria`. Which command would add a red line representing the CPI for Austria.

- a) `plot(c(seq(2002, 2011, 1)), cpi_value_Austria, type = "l", col = "red")`
- b) `lines(c(seq(2002, 2011, 1)), cpi_value_Austria, type = "l", col = "red")`
- c) `points(c(seq(2002, 2011, 1)), cpi_value_Austria, col="red")`
- d) `plot(cpi_value_Austria, type="b", col = "red")`
- e) `lines(cpi_value_Austria, type="b", col = "red")`

78. Which of the following is the correct syntax?

- A. `plot(c(2,1,7), c(4,5), c(4,5,3))`
- B. `plot(c(2), c(4,5,8))`
- C. `plot(c(2,1,7), c(4,5,8))`
- D. `plot(c(2,10), 5, 6)`

79. Which code produced the text “Statistical Computation” on graphic shown below?



- A. `mtext('Statistical Computation', side=3)`
- B. `mtext('Statistical Computation', side=1)`
- C. `mtext('Statistical Computation', side=3, adj=0)`
- D. `mtext('Statistical Computation', side=1, adj=0)`
- E. `text(1550,2400,' Statistical Computation')`
- F. `text(1550,2400,' Statistical Computation', adj=0)`

80. Which of the following commands is required in order to remove all variables from R workspace?

- A. `rm(Object)`
- B. `rm(list=ls())`
- C. `rm()`
- D. `rm(list=(ls(), "x"))`

81. Suppose you want to randomly select 6 schools from 20 schools with smallest numbers of teachers in Adair County in Oklahoma data. Which of the following functions you need to use? There might be more than one correct answer. (Notice that the number of teachers may not be integer)

- A. grep, sort, rnorm
- B. grepl, sort, sample
- C. grep, sort, runif
- C. grepl, order, rnorm
- D. grep, order, sample
- E. grepl, order, runif
- F. which, grep, rnorm
- G. which, grep, sample
- H. which, grep, runif

82. What will be the console output when the following script is executed?

```
> a <- c(25,40,10,95)
> order(a)
```

- A. 1 2 3 4
- B. 10 25 40 95
- C. 3 1 2 4
- D. 2 3 1 4

83. The breaks keyword in hist() function:

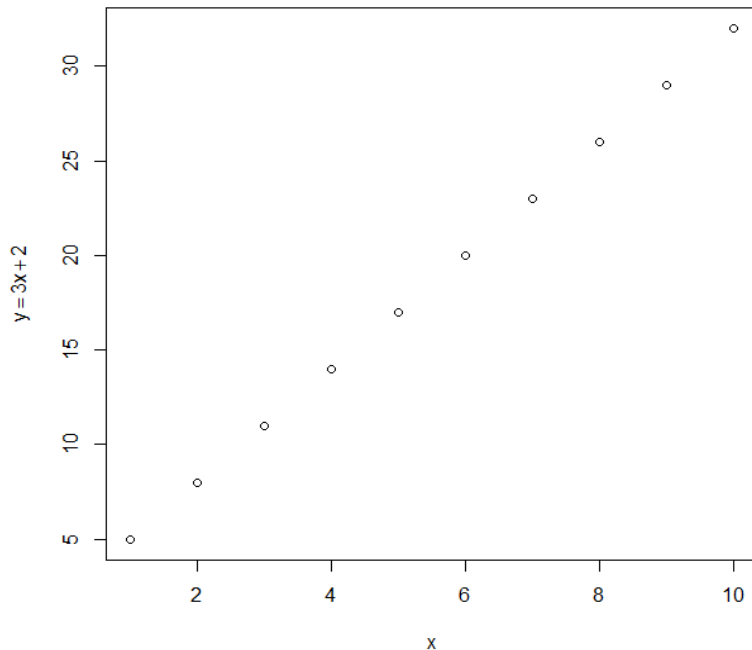
- a) Is a number specifying the length of a cell
- b) Is a vector giving away break points between histogram cells
- c) Can only suggest and not force breaks
- d) Can only take algorithm names as value

84. Which command will produce a linear model of the effect of hours of sleep, "hr_sleep" on test scores, "score", from a data frame "SAT"?

- A) lm(score~hr_sleep, SAT)
- B) lm(hr_sleep,score, SAT)
- C) lm(hr_sleep~score, "SAT")
- D) lm(score=hr_sleep, SAT)

85. Given the following graph with data, which one of the below plot functions will produce the correct result?

```
m<-3;
c<-2;
x<-c(1:10)
y<-m*x+c
```



Options:

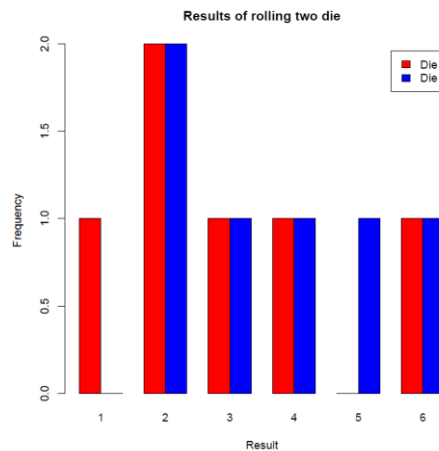
- A. `plot(x,y,xlab = expression(x), ylab = bquote(y==.(m)*x+. (c)))`
- B. `plot(x,y,xlab = expression(x), ylab = expression(y==m*x+c))`
- C. `plot(x,y,xlab = expression(x), ylab = bquote(y==m*x+c))`
- D. `plot(x,y,xlab = bquote(. (m)*x), ylab = bquote(y==.(m)*x+. (c)))`

86. Which of the following function(s) will help change the following string of characters in vector x (`x <-c("Library Book", "Book Store", "Book Club")`) to "Library Magazine", "Magazine Store", "Magazine Club"?

- A. `gsub("Book", "Magazine", x)`
- B. `substitute("Book", "Magazine", x)`
- C. `rm(x, "Book", paste("Magazine"))`
- D. `sub("Book", "Magazine", x)`
- E. `paste("Magazine", x, rm=Book)`

87. You roll two standard 6-sided dice 6 times each with the following results, stored in a data frame named “dice_roll”. The column named “die” refers to which die has been rolled, and the column named “result” shows the result of each roll.

```
> dice_roll
  die result
1    1      3
2    1      4
3    1      2
4    1      2
5    1      1
6    1      6
7    2      2
8    2      3
9    2      6
10   2      5
11   2      2
12   2      4
```



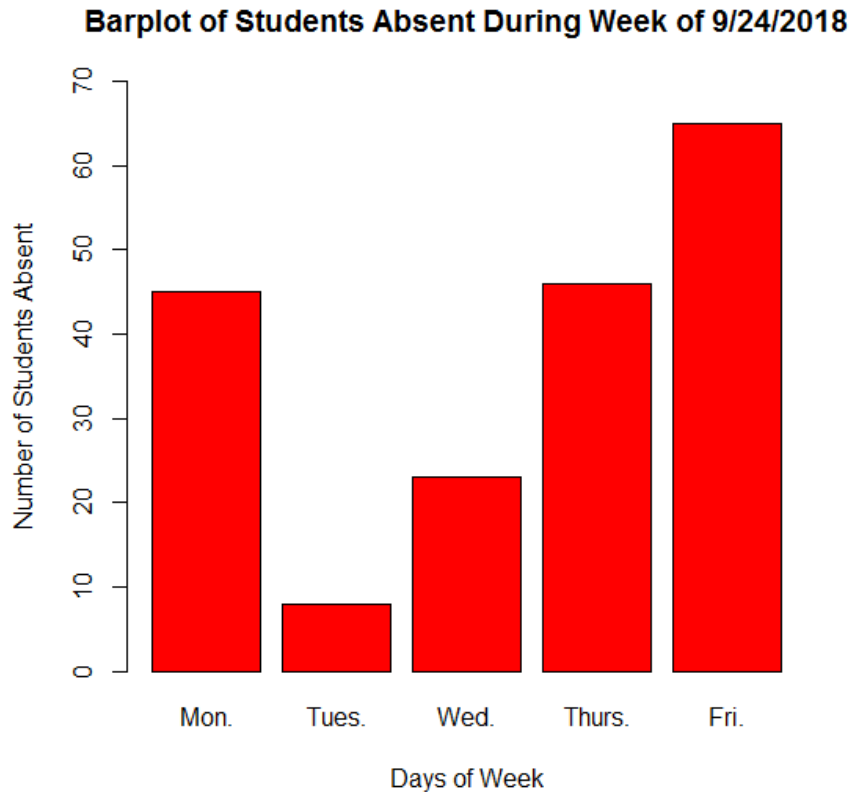
You want to create a barplot showing the frequency of each result for each die. Which of the following lines of script created the following barplot? For your convenience, only the highlighted parts are different among the answers. **SELECT ALL THAT APPLY**

- A) `barplot(table(dice_roll$die, dice_roll$result), beside=TRUE, col=c('red','blue'), legend=c('Die 1','Die 2'), xlab='Result', ylab='Frequency', main='Results of rolling two die')`
- B) `barplot(table(dice_roll), beside=TRUE, col=c('red','blue'), legend=c('Die 1','Die 2'), xlab='Result', ylab='Frequency', main='Results of rolling two die')`
- C) `barplot(table(dice_roll$result, dice_roll$die), beside=TRUE, col=c('red','blue'), legend=c('Die 1','Die 2'), xlab='Result', ylab='Frequency', main='Results of rolling two die')`
- D) `barplot(dice_roll, beside=TRUE, col=c('red','blue'), legend=c('Die 1','Die 2'), xlab='Result', ylab='Frequency', main='Results of rolling two die')`
- E) None of the above scripts could have produced the given barplot

88. What can you type in your console if you want to open help pages about the exponential distribution? **Select all that apply.**

- a) `?exponential`
- b) `??exponential`
- c) `'exponential distribution'`
- d) `?exponential distribution'`
- e) `?exponentialdistribution`
- f) `??'exponential distribution'`

89. Which code produced the graphic shown below?



(A) `barplot(absences1,col="red",names.arg=c("Mon.", "Tues.", "Wed.", "Thurs.", "Fri."),main="Barplot of Students Absent During Week of 9/24/2018", ylab="Days of Week",xlab="Number of Students Absent", xlim=c(70,0))`

(B) `barplot(absences1,col="red",names.arg=c("Mon.", "Tues.", "Wed.", "Thurs.", "Fri."),main="Barplot of Students Absent During Week of 9/24/2018", ylab="Days of Week",xlab="Number of Students Absent", ylim=c(0,70))`

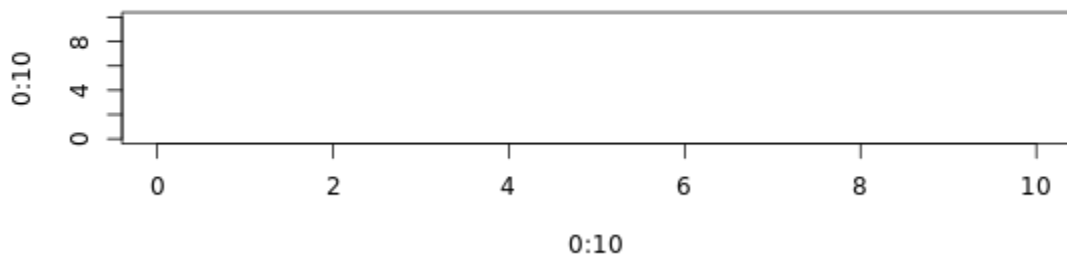
(C) `barplot(absences1,col="red",names.arg=c("Mon.", "Tues.", "Wed.", "Thurs.", "Fri."),main="Barplot of Students Absent During Week of 9/24/2018", xlab="Days of Week",ylab="Number of Students Absent", ylim=c(0,70))`

(D) `barplot(absences1,col="red",names.arg=c("Mon.", "Tues.", "Wed.", "Thurs.", "Fri."),main="Barplot of Students Absent During Week of 9/24/2018", xlab="Days of Week",ylab="Number of Students Absent", ylim=c(70,0))`

90. Which of the following functions outputs a multiple-image file?

- A. `png()`
- B. `tiff()`
- C. `postscript()`
- D. `bmp()`
- E. `PDF()`
- F. None of the above

91. The command
`plot(0:10, 0:10, type = "n")`



What command would you give to add a blue point at coordinates (5,3)?

- a) `points(5,3, col="blue")`
- a) `points(3,5, col="blue")`
- b) `addPoint(5,3,"blue")`
- c) `addPoint(3,5,"blue")`

92. Which is the correct syntax to set up multiple graphs on the same page?

- a. `plot(mfrow=c(nr,nc))`
- b. `par(mfrow=c(nr,nc))`
- c. `legend(mfrow=c(nr,nc))`
- d. `mulplot(mfrow=c(nr,nc))`
- e. `graph(mfrow=c(nr,nc))`

93. Which output is the correct one of the following script?

```
> rxn <- c('A','B','C','D','E','F')  
> H <- c(147,-59,23)  
> (Thermo <- cbind(rxn,H))
```

A. rxn H
[1,] A 147
[2,] B -59
[3,] C 23
[4,] D NA
[5,] E NA
[6,] F NA

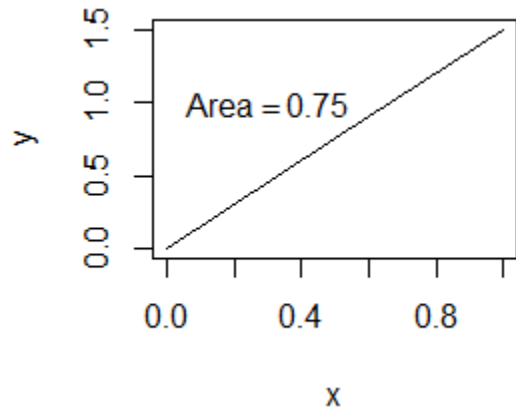
B. rxn H
[1,] "A" "147"
[2,] "B" "-59"
[3,] "C" "23"
[4,] "D" "147"
[5,] "E" "-59"
[6,] "F" "23"

C. rxn H
[1,] A 147
[2,] B -59
[3,] C 23
[4,] D 147
[5,] E -59
[6,] F 23

D. rxn H
[1,] "A" 147
[2,] "B" -59
[3,] "C" 23
[4,] "D" 147
[5,] "E" -59
[6,] "F" 23

E. rxn H
[1,] A 147
[2,] B -59
[3,] C 23
[4,] D 23
[5,] E 23
[6,] F 23

94. Which code produced the text in the graphic shown below? **Select all that apply**



- A. `> a = 1`
`> b = 1.5`
`> text(0.3,0.5,bquote(Area==a*b/2))`
- B. `> a = 1`
`> b = 1.5`
`> text(0.3,0.5,bquote(Area==.(a)*.(b)/2))`
- C. `> a = 1`
`> b = 1.5`
`> text(0.3,0.5,bquote(Area==.(a*b)/2))`
- D. `> a = 1`
`> b = 1.5`
`> text(0.3,0.5,bquote(Area==.(a*b/2)))`

95. Functions can be created in R for the purpose of creating reusable code. In the following function example, which component is considered optional:

```
mysum<-function(a, b, c=0) a+b+c
```

- a. mysum
- b. c=0
- c. a+b+c
- d. function()

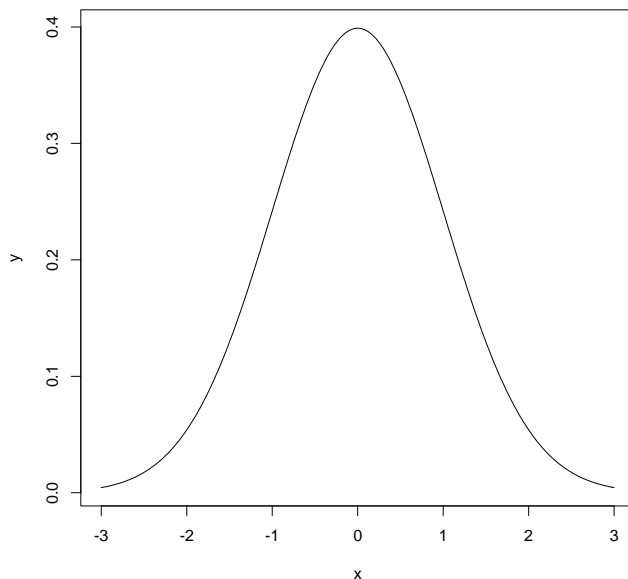
96. Which answer is TRUE?

- A. Normal Distribution functions within R have 4 General forms: `dnorm()`, `rnorm()`, `qnorm()` and `pnorm()`.
- B. `dnorm()` and `rnorm()` are used to determine the densities for the Normal distribution.
- C. `dnorm()` is used to determine the densities for the Normal Distribution while `rnorm()` is used for pseudo-random generation.
- D. Sd stands for string deviation.

97. Which form is the best way to use `rnorm()`?

- A. `Rnorm(x, mean, sd)`
- B. `Rnorm(mean, x=0, sd)`
- C. `Rnorm(cpis, sd,)`
- D. `Rnorm(x)`

98. Which plot matches the following graph?



- A.

```
n<-(-3,3,1,6)
l<-qnorm(x)
plot(n,l)
```
- B.

```
plot(x,y,type=l)
```
- C.

```
paste(x,y, sep= "_", )
```
- D.

```
x<-seq(-3,3,0.03)
y<-dnorm(x)
plot(x,y,type="l")
```

99. Given a data frame called “stationary” with the following (in order) columns: paper, pencils, erasers, scissors. Which of the following are ways to access the data in the pencils column of the “stationary” data frame?

- a) stationary\$pencils
- b) attach(stationary)
- c) stationary[,1]
- d) with(stationary, pencils)

100. Which of these statements regarding conditional code expression are **false**?

- a. The condition expression must evaluate to a logical vector with a length of one
- b. Braces around the command are optional if only executing one expression
i.e., `if(condition)command` is equivalent to `if(condition){command}`
- c. Conditional code execution can only handle one condition per expression set
- d. The else statement expression is optional after the if statement

101. Which of the following commands can clean up ALL objects in a previously saved R workspace?

Select all that apply.

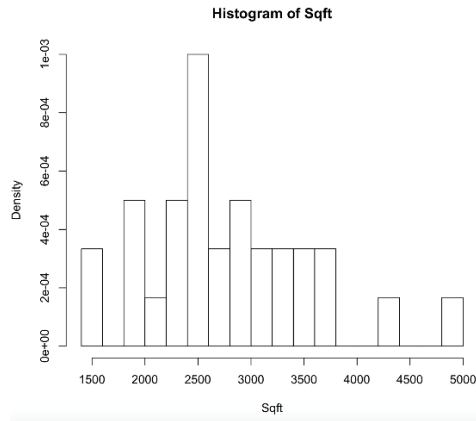
- A. `objects()`
`getwd()`
- B. `rm(list = objects())`
- C. `ls()`
`setwd()`
- D. `rm(list = ls())`
- E. `rm()`

102. Which of the following commands would you use as an Output Device for multiple images?

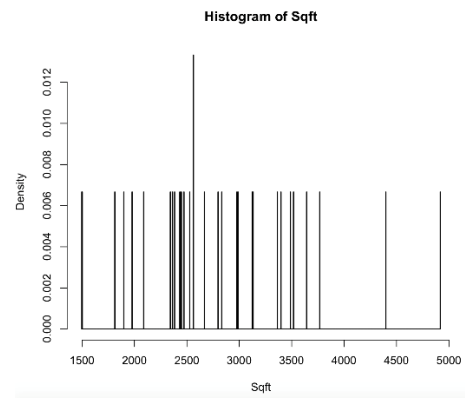
- a) `tiff()`
- b) `postscript()`
- c) `bmp()`
- d) `png()`

103. Breaks is one of a parameter in Histogram. Which of the following graph could have the highest number of breaks in the R console?

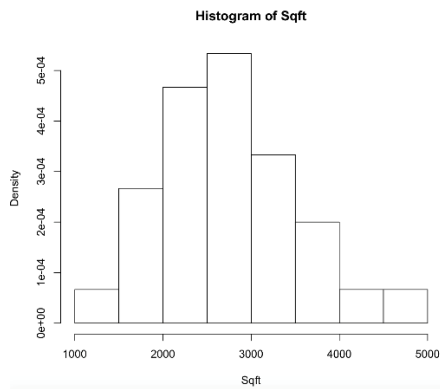
(a)



(c)



(b)



(d)

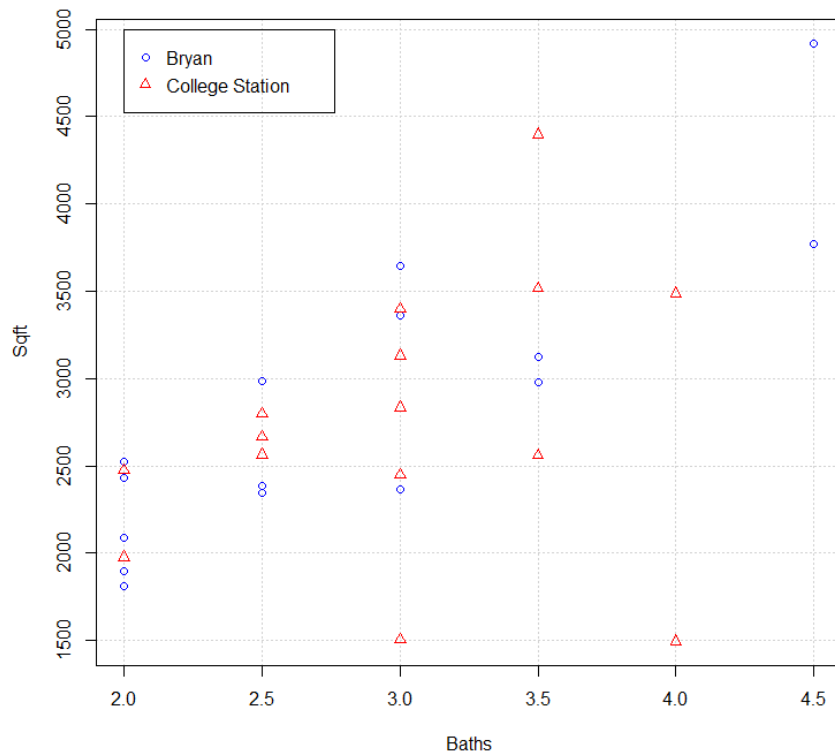


104. Which of these would you use to display names of countries from a data frame (cpi) in reverse alphabetical order?

Note: The data frame is “cpi” and the 3rd column is named “country”.

- a. `order(cpi$country, decreasing = 1)`
- b. `sort(cpi$country, highest = true)`
- c. `order(cpi$country, decreasing = TRUE)`
- d. `sort(cpi$country, increasing = FALSE)`

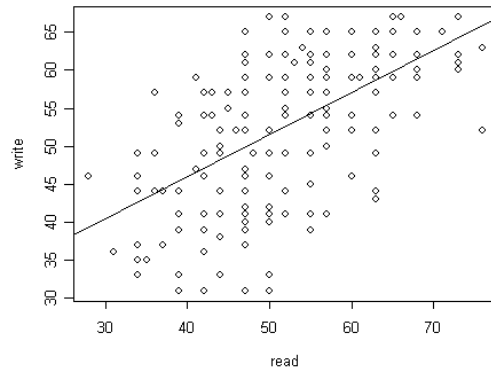
105. Which code created the **Legend** shown in the graph below:



Select all that apply!

- a) `legend(2.0,5000,c("Bryan","College Station"),pch=c(1,2),col=c("blue","red"))`
- b) `legend(2.75,5000,c("Bryan","College Station"),pch=c(1,2),col=c("blue","red"))`
- c) `legend(locator(1),c("Bryan","College Station"),pch=c(1,2),col=c("blue","red"))`
- d) `legend(2.0,4500,c("Bryan","College Station"),pch=c(1,2),col=c("blue","red"))`

106. Which of the following commands will produce an error while execution to plot a chart from dataset HW07.RData that shows a linear relationship of number of people who know to read and write in a city (Assume HW07.RData containing zipdata is already loaded and is ready to use in the workspace)



- a) `abline(data = zipdata, lm(write~read))`
- b) `abline(lm(write~read), zipdata)`
- c) `abline(lm(read~write), zipdata)`
- d) `abline(zipdata,lm(read~write))`

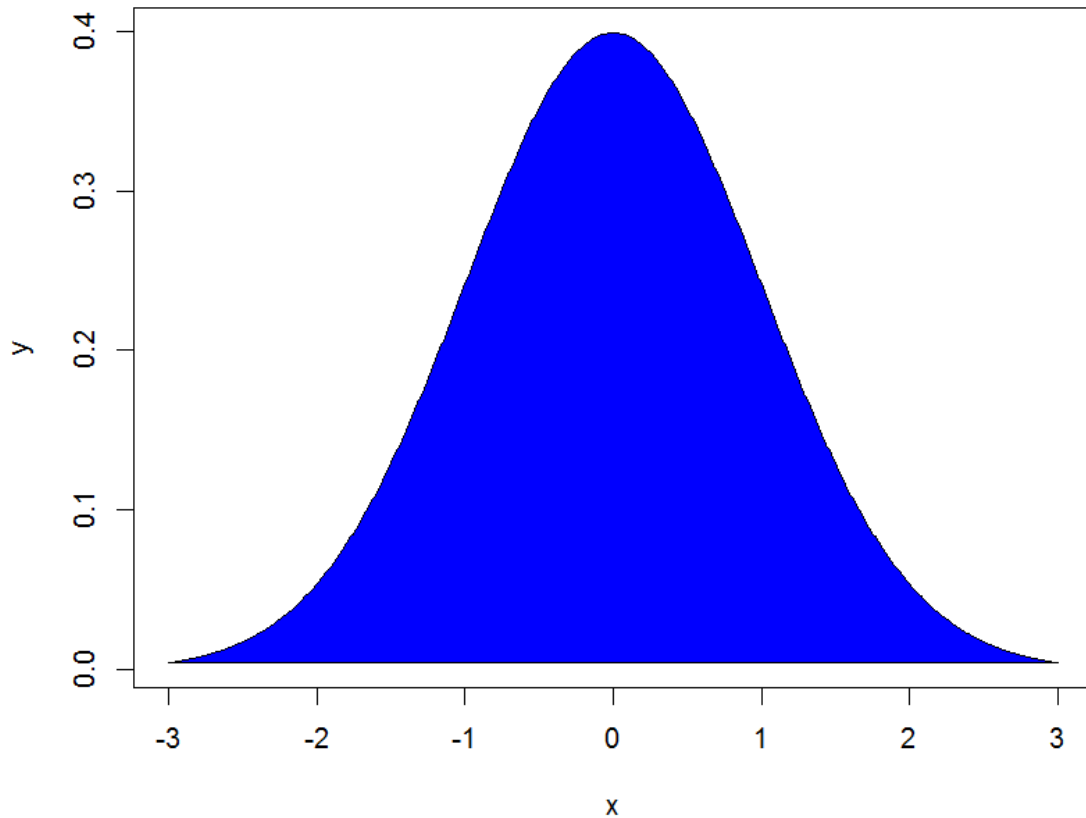
107. Which of the following will be used to extract the first 5 characters in a string?

- A. `substr(1,5,S)`
- B. `substr(S, 1, 5)`
- C. `substr(S, 1:5)`
- D. Both B and C
- E. None of the above

108. To add a straight line with a slope of 5 and a y-intercept of 3 to an existing graphic, you would use which of the following commands:

- A. `abline(3, 5)`
- B. `abline(5, 3)`
- C. `mblne(3, 5)`
- D. `mblne(5, 3)`

109.



Select all the correct coding examples that will generate the graph above:

- A) `x<-seq(-3, 3, 0.01)`
`y<-dnorm(x)`
`plot(x,y,type="l")`
`polygon(x, y, density =1.0, palette(blue))`
- B) `x<-seq(-3, 3, 0.01)`
`plot(x, dnorm(x),type="l")`
`polygon(x, dnorm(x), col="blue")`
- C) `x<-seq(-3, 3, 0.01)`
`y<-dnorm(x)`
`plot(x,y,type="l")`
`polygon(x, y, density =NULL, col="blue")`
- D) `x<-seq(-3, 3, 0.01)`
`y<-dnorm(x)`
`plot(x,y,type="l")`
`polygon(x, y, density =NULL, palette(blue))`

110. Which line of code will correctly read the data from the image below saved in .csv format into a data frame in R? The file path where it is stored is "C:\Users\User1\new 1.csv".

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	A	547	Nicholas Cannady	S	Male															
2	A	582	Emma Olmstead	HR	Female															
3	A	654	Nellie Monahan	CS	Female															
4	B	129	Debbie Ashburn	M	Female															
5	A	266	Shawn Valles	SM	Male															
6	D	746	Jodi Bracey	CS	Female															
7	B	214	Christopher Drummond	S	Male															
8	C	685	Latoya Thornton	SM	Female															
9	A	924	Richard Bickerstaff	HR	Male															
10	B	127	Ida Silverman	M	Female															
11	A	657	Juan Kuehn	SM	Male															
12	A	641	Beverly Teasley	S	Female															
13	C	981	Anne Fraley	HR	Female															
14	B	871	Douglas Mellon	M	Male															
15	C	347	Marilyn Pfeifer	CS	Female															
16	C	169	Bryan Stencil	HR	Male															
17	A	473	Lena Largent	S	Female															
18	A	247	Travis Magana	HR	Male															
19	B	684	Colleen Pray	CS	Female															
20	C	341	Lawrence Nickel	S	Male															

- A. `read.delim("C:\Users\User1\new 1.csv",sep=|)`
- B. `read.csv("C:\Users\User1\new 1.csv",header=TRUE,sep="|")`
- C. `read.csv("C:\Users\User1\new 1.csv",sep="|")`
- D. `data.frame(new 1,nrow=20,ncol=5,header=FALSE)`

111. What will the output of the following code be?

```
>a <- c(1,4,7,10)
>ifelse(a %% 5 == 0 , "Multiple of 5" , "Not a multiple of 5")
```

- a. Error
- b. [1] Multiple of 5 Multiple of 5 Multiple of 5 Not a Multiple of 5
- c. [1] Not a Multiple of 5 Not a Multiple of 5 Not a Multiple of 5 Multiple of 5
- d. [1] Not a Multiple of 5 Multiple of 5 Multiple of 5 Not a Multiple of 5

112. Please select the correct output of the following command:

```
xf <- c(1,2,3,7,8,9)
print(mf <- matrix(xf,nrow=2,byrow = TRUE))
```

- a)

	[,1]	[,2]	[,3]	
[1,]	1	2	3	
[2,]	7	8	9	
- b)

	[,1]	[,2]	[,3]	
[1,]	1	3	8	
[2,]	2	7	9	
- c)

	[,1]	[,2]	
[1,]	1	2	
[2,]	3	7	
[3,]	8	9	
- d)

	[,1]	[,2]	
[1,]	1	7	
[2,]	2	8	
[3,]	3	9	

113. What part of the data frame `mat1` is the following code accessing?

```
mat1[2:4, -c(2,5)]
```

- A. Sub-matrix with the second and fourth columns, and the second and fifth rows removed.
- B. Sub-matrix with the second through fourth rows, and the second and fifth columns removed.
- C. Sub-matrix with the second through fourth rows, and the second and fifth columns.
- D. Sub-matrix with the second and fourth rows, and the second and fifth columns removed.

114. Choose which of the following logical vectors should a Conditional Code Execution evaluate to?

- A. Length – Zero
- B. Length – One
- C. Length – Two
- D. None of the above

115. The following is an R statement to plot values in vector X. A is used to print values of x on the pie chart to label/name the slice. B is used to set the direction of naming slices in the pie-chart. Which of the following arguments represent A and B respectively?

pie(x, A = names(x), B = FALSE, init.angle= if(clockwise) 90 else 0)

- a) Edges, Density
- b) Labels, Clockwise
- c) Labels, Angle
- d) Edges, Clockwise

Example: X<-(1:4)

