### Always Remember "google is your friend!"

1. Create a variable "x" that has value 5. And another variable "y" that has value 3. Add them and store in z. Print z in the console

2. Create a vector called "tamim" with runs in five matches. The runs are 40, 50, 100, 20, 10,35, 40. Get the average run of tamin in the tournament

3. Create a boxplot of variable tamim. It should like like the following. Which one is the outlier here?
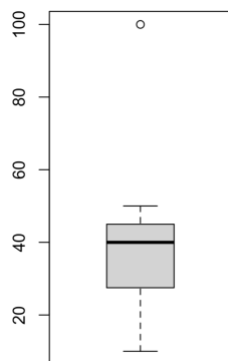
4. Create matrix with number 1 to 100. The matrix should contain 4 columns and 25 rows.

5. Give the 4 column names. "one, two, three, four". It should look like following

```
      one two three four
 [1,]   1   2     3    4
 [2,]   5   6     7    8
 [3,]   9  10    11   12
 [4,]  13  14    15   16
 [5,]  17  18    19   20
 [6,]  21  22    23   24
 [7,]  25  26    27   28
 [8,]  29  30    31   32
 [9,]  33  34    35   36
[10,]  37  38    39   40
[11,]  41  42    43   44
[12,]  45  46    47   48
[13,]  49  50    51   52
[14,]  53  54    55   56
[15,]  57  58    59   60
[16,]  61  62    63   64
[17,]  65  66    67   68
[18,]  69  70    71   72
[19,]  73  74    75   76
[20,]  77  78    79   80
[21,]  81  82    83   84
[22,]  85  86    87   88
[23,]  89  90    91   92
[24,]  93  94    95   96
[25,]  97  98    99  100
```

6. Create a data frame that looks like following.

name	region	sales	expenses
William	East	50000	42000
Emma	North	52000	43000
Sofia	East	90000	50000
Markus	South	34000	44000
Edward	West	42000	38000
Thomas	West	72000	39000

7. Create the following list:

```
$mother
[1] "Veronique"
$father
[1] "Michel"
$sisters
[1] "Alicia" "Monica"
$sister_age
[1] 12 22
```

8. Write an if else condition where it says, if x is greater than 0 then print positive, if less than 0 print negative and if x is 0 than print zero and if anything else print please type a new number. Check what the value shows if x <- 5 and x <- 0

9. Create the following data frame:

```
DF1 = data.frame(c1= c(1, 5, 14, 23, 54), c2= c(9, 15, 85, 3, 42), c3= c(9, 7, 42, 87, 16))
DF1
```

Convert this data frame into matrix. Such that you get the following output. Check "google" to know how to convert a data frame into matrix. You should get the following output

```
      c1 c2 c3
[1,]   1  9  9
[2,]   5 15  7
[3,]  14 85 42
[4,]  23  3 87
[5,]  54 42 16
```

10. Use for loop to get the following output

```
[1] "the year is,1"
[1] "the year is,2"
[1] "the year is,3"
[1] "the year is,4"
[1] "the year is,5"
[1] "the year is,6"
[1] "the year is,7"
[1] "the year is,8"
[1] "the year is,9"
[1] "the year is,10"
```

11. Install Bioconductor in R. install Deseq2 package in R. Check the following
https://www.bioconductor.org/packages/release/bioc/html/DESeq2.html

12. Load the inbuild "mtcars" data frame and store it as "cars". If you View the cars you can see that the first column is "mpg" and fourth column is "hp". Swap this column such that 1st column is "hp" and the fourth column is "mpg".

13. Make a scatter plot "mpg" vs "hp" and color it based on "gear". What is the difference when you use factor and when you don't

14. Load the file "house.csv" in R. you can see that the "aircond" column has 0 or 1. Where 0 means no facility of air-condition and 1 means containing facility of air-condition. Create a new column with "yes" and "no" for each columns.

15. Make a new column titled "New_created_column" which contains "yes" or "no" value which is consistent with 0 and 1 in "aircond" column.

16. Make a boxplot with x axis as room number and y axis as price and compare how difference in price it makes if you have air condition facility or not. Use new_created_column

17. Create a frequency polygon with facet_grid based on "New_created_column"
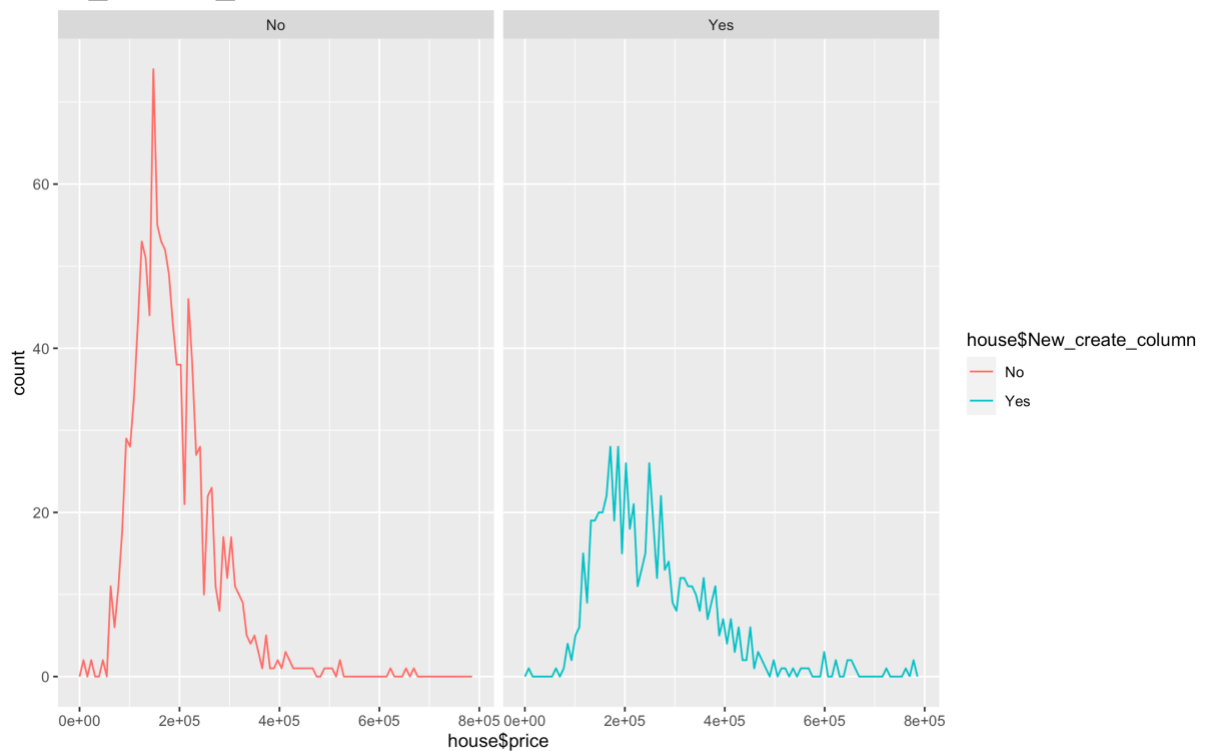
18. Say you have three genes EGFR, FOX1, TF1. You have calculated that in three cell type "cell1, cell2, cell3". You want to create a matrix where each gene will have a value for each cell type

```
      cell_1 cell_2 cell_3
EGFR      10     12     15
TF_1     100     12     20
FOX_1     12      2      3
```

19. Calculate the mean of each row and add it as the fourth column and calculate the row sum and add it as the fifth column

20. Say you have collected some samples from 5 person. You have asked them if they believe in aliens. The response was as follows :

```
responses <- factor(c("Agree", "Agree", "Strongly Agree", "Disagree",
"Agree"))
```

```
[1] Agree           Agree           Strongly Agree Disagree       Agree
Levels: Agree Disagree Strongly Agree
```

You want to change is such the the level shows it in following order:

```
[1] Agree           Agree           Strongly Agree Disagree       Agree
Levels: Strongly Agree Agree Disagree
```

21. Create the following data frame,

	Age	Height	Weight	Sex
Alex	25	177	57	F
Lilly	31	163	69	F
Mark	23	190	83	M
Oliver	52	179	75	M
Martha	76	163	70	F
Lucas	49	183	83	M
Caroline	26	164	53	F

afterwards invert Sex for all individuals. Sothat it shows M in place of F and F in place of M.

22. Create this data frame (make sure you import the variable Working as character and not factor).

	Working
Alex	Yes
Lilly	No
Mark	No
Oliver	Yes
Martha	Yes
Lucas	No
Caroline	Yes

Add this data frame column-wise to the previous one. a) How many rows and columns does the new data frame have? b) What class of data is in each column?

23. Write two string "hello" and "why am I doing this". Add this two string together and separate by ","

24. If name_list <- list(a = 1:200, b = "this is a string", c = "hello"). You will write a code that will add 1 to each element of the first vector of the new list. Also, add a new item z = "newItem" to the list name_list

25. Download the small_counts.txt from the following location

https://figshare.com/s/1d788fd384d33e913a2a

26. Read the file in R and save it as small_counts. View the file.

27. Get the following output from the file

```
          Sample_1 Sample_2
Xkr4           438      300
Sox17          106      182
Mrpl15         309      234
Lypla1         652      515
Tcea1         1604     1495
Rgs20            4        2
Atp6v1h        769      752
Rb1cc1        1494     1412
Pcmtd1        1344     1242
Rrs1          1691     1808
```

28. Get log of the small_count so that it looks like the following

```
         Sample_1   Sample_2  Sample_3  Sample_4
Xkr4     6.082219  5.7037825  4.174387  5.468060
Sox17    4.663439  5.2040067  4.406719  4.653960
Mrpl15   5.733341  5.4553211  5.820083  5.703782
Lypla1   6.480045  6.2441669  6.854355  6.840547
Tcea1    7.380256  7.3098815  7.450661  7.183112
Rgs20    1.386294  0.6931472  2.639057  1.386294
Atp6v1h  6.645091  6.6227363  6.967909  6.894670
Rb1cc1   7.309212  7.2527624  7.053586  6.874198
Pcmtd1   7.203406  7.1244783  7.225481  7.373374
Rrs1     7.433075  7.4999765  7.662468  7.410347
```

29. Download the "ResultsTable_small.txt" from the following location
https://figshare.com/s/1d788fd384d33e913a2a

30. This is a file which contains the gene expression data. The Entrez id is the gene name. You can search Entrez id in google to get more information.Symbol is the gene name. And "logFC" value which means how much more a gene is expressed in treatment condition compared to control condition. Read the file in R and store it as
"results"

	ENTREZID	SYMBOL	logFC	AveExpr	t	P.Value	adj.P.Val
1	24117	Wif1	1.819943	2.975545	20.10780	1.063770e-10	1.01624e-06
2	381290	Atp2b4	-2.143885	3.944066	-19.07495	1.982934e-10	1.01624e-06
3	78896	1500015010Rik	2.807548	3.036519	18.54773	2.758828e-10	1.01624e-06
4	226101	Myof	-2.329744	6.223525	-18.26861	3.297667e-10	1.01624e-06
5	16012	Igfbp6	-2.896115	1.978449	-18.21525	3.413066e-10	1.01624e-06
6	231830	Micall2	2.253400	4.760597	18.02627	3.858161e-10	1.01624e-06

31. Sort the file such thath the genes are orderd in highest to lowest value of "logFC".

	ENTREZID	SYMBOL	logFC	AveExpr	t	P.Value	adj.P.Val
22	16878	Lif	3.738933	6.68203417	13.73344	9.105708e-09	6.541210e-06
10	14620	Gjb3	3.600094	3.52528051	16.46627	1.113755e-09	1.718703e-06
25	12977	Csf1	2.835624	7.47759094	13.41902	1.187300e-08	7.505634e-06
3	78896	1500015010Rik	2.807548	3.03651950	18.54773	2.758828e-10	1.016240e-06
15	11636	Ak1	2.766745	4.30347462	15.27694	2.664640e-09	2.807465e-06
26	12654	Chil1	2.342914	5.57645724	13.21976	1.408760e-08	8.306595e-06
29	217166	Nr1d1	2.278879	6.26087761	13.12885	1.524242e-08	8.306595e-06
6	231830	Micall2	2.253400	4.76059697	18.02627	3.858161e-10	1.016240e-06
13	74747	Ddit4	2.180370	6.86479110	15.70145	1.938279e-09	2.356351e-06
20	17131	Smad7	1.972771	6.71751902	14.14348	6.493642e-09	5.131276e-06

32. See the following figure

```
counts_matrix <- data.frame(cell_1 = rpois(10, 10),
                cell_2 = rpois(10, 10),
                cell_3 = rpois(10, 30))
rownames(counts_matrix) <- paste0("gene_", 1:10)
counts_matrix <- as.matrix(counts_matrix) # must be a matrix object!
```

Type the above code in your console and check counts_matrix to see what does it create. What do you think "rpois" comment did here? And try to find what is the difference between paste and paste0. Always remember "google" is your friend

33. Create a heatmap using the following file: "basketball.csv". Explain what was done in each line. Use "##" to comment on your code file in R.
Load the file "basketball.csv". Make sure you change the read.csv location from the following code.

```
nba <- read.csv("Desktop/PBT_LAb/cBLAST/Curriculum for workshop/basketball.csv", header =TRUE, sep = ",")

nba <- nba[order(nba$PTS),]

row.names(nba) <- nba$Name

nba <- nba[,2:20]

nba_matrix <- data.matrix(nba)


nba_heatmap <- heatmap(nba_matrix, Rowv=NA, Colv=NA, col = cm.colors(256), scale="column", margins=c(5,10))
```

The code will/might show you error!.copy the error and put it in google and see what is the suggestion from the internet. Try to understand and

solve the problem. The end of the code will show something like the following: