

TRGN 527: Applied Data Science and Bioinformatics

UNIT I. Introduction and Basic Data Science

Week 5 – Assignment 5

Enrique I. Velazquez Villarreal, M.D., Ph.D., M.P.H., M.S. | Assistant Professor

Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Leader of the USC Bioinformatics Core – *USC CaRE2 Health Equity Center*

David W. Craig, Ph.D. | Professor and Vice Chair

Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Co-Director, Institute of Translational Genomics

Assignment 5

- Instructions: Please answer the following questions, first 14 questions worth 7 points and last questions worth 2 points (Total: 100 points).
- Load the file “TRGN599_sample_info_A5.csv” and “TRGN599_expression_results_A5.csv” in your R-Studio.
 - For example, using the following code:

```
samples <- read.csv('TRGN599_sample_info_A5.csv', header = TRUE, sep = ",", quote = "", dec = ".", fill = TRUE,  
row.names = 1)

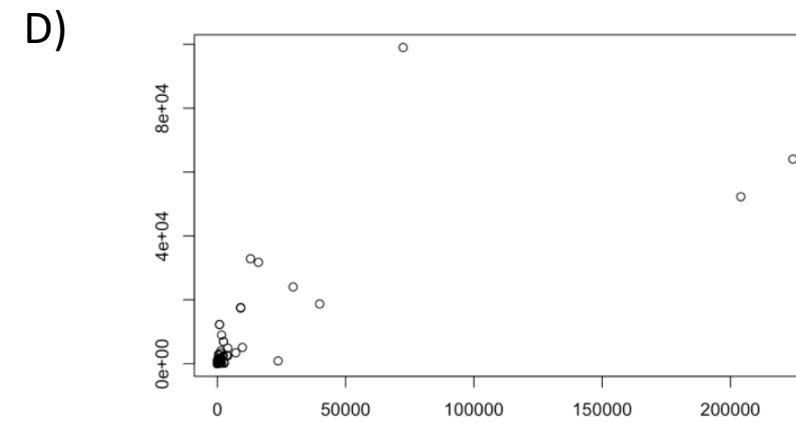
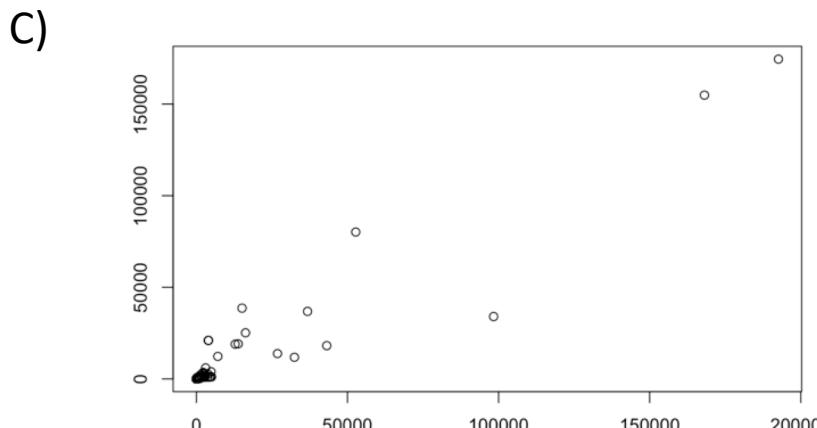
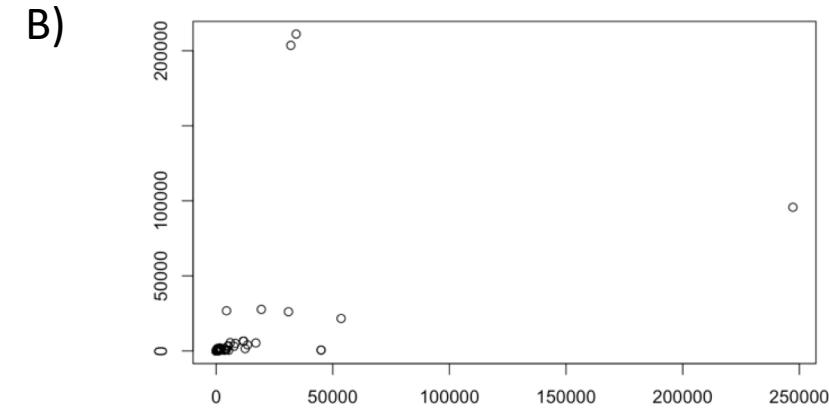
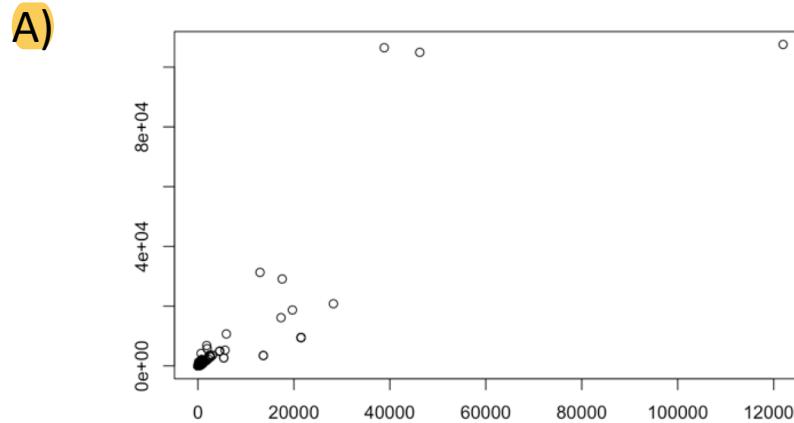
genes <- read.csv('TRGN599_expression_results_A5.csv', header = TRUE, sep = ",", quote = "", dec = ".", fill = TR  
UE, row.names = 1)
```

- In order to answer the following questions you should
- 1) create an R markdown file as explained during the last class,
- 2) copy/paste & run with the new data files (mentioned above) the R markdown of Week_5_Lecture 1 (correlation analysis) that you can find in Blackboard/Lecture 5:
 - File name: Rmarkdown_TRGN_599_Week_5_Lecture_1.Rmd.

Assignment 5

- Question 1

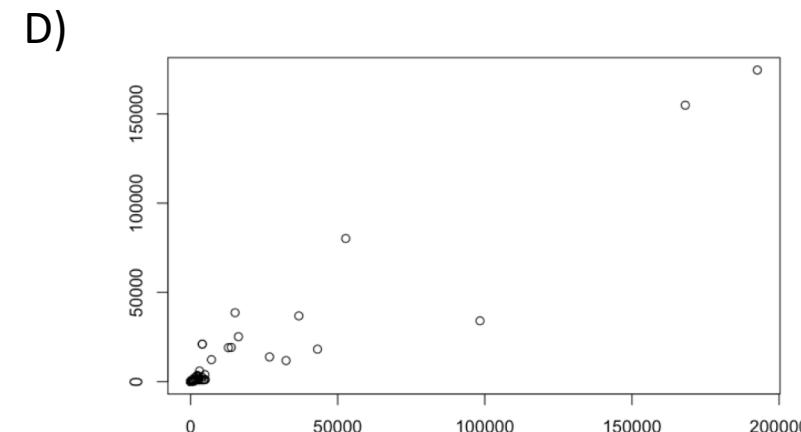
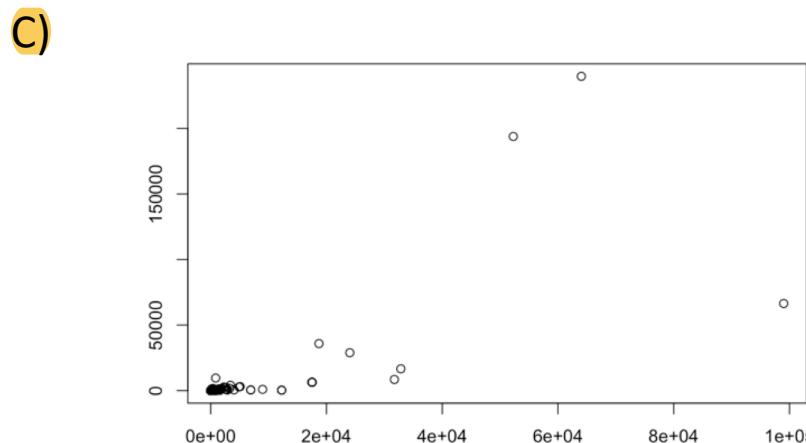
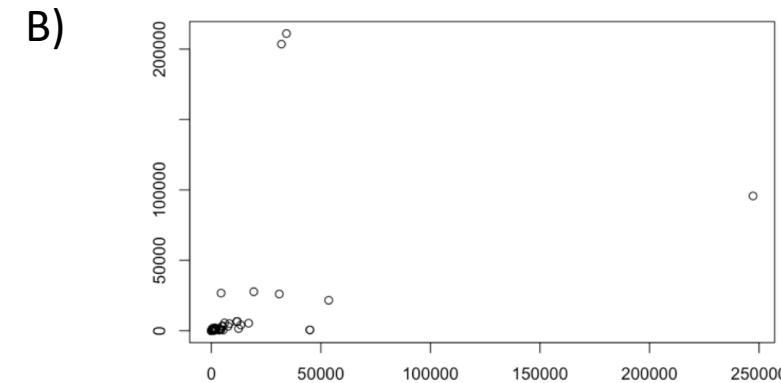
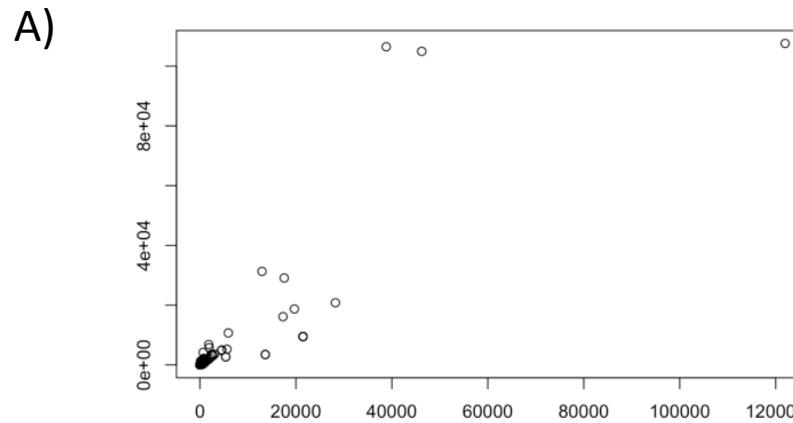
- Plot the genes KITA_23 vs KITA_24. Which of the following graphs is the correct output of the plot function?



Assignment 5

- Question 2

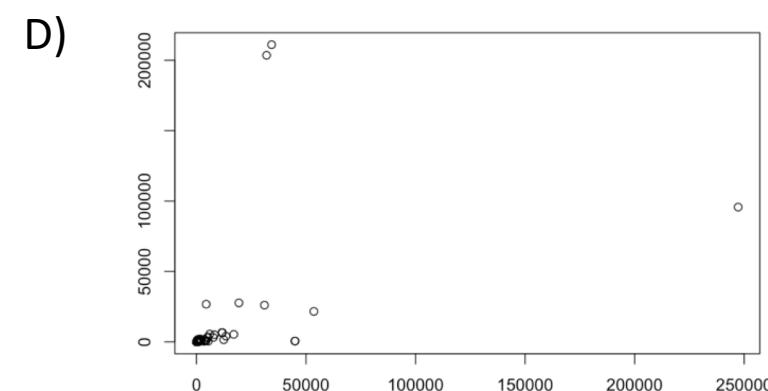
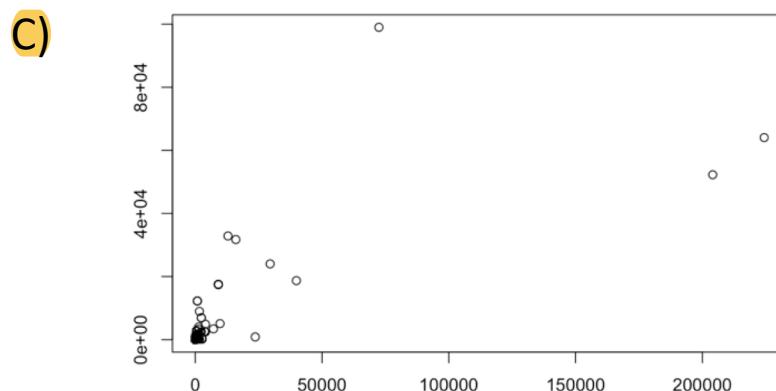
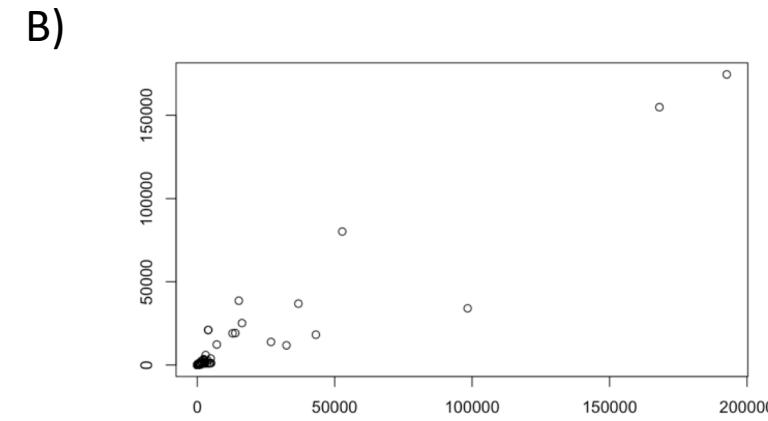
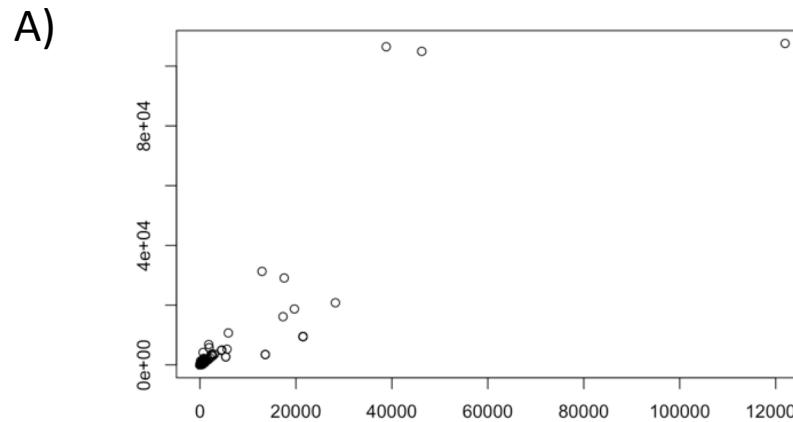
- Plot the genes KITA_11 vs KITA_12. Which of the following graphs is the correct output of the plot function?



Assignment 5

- Question 3

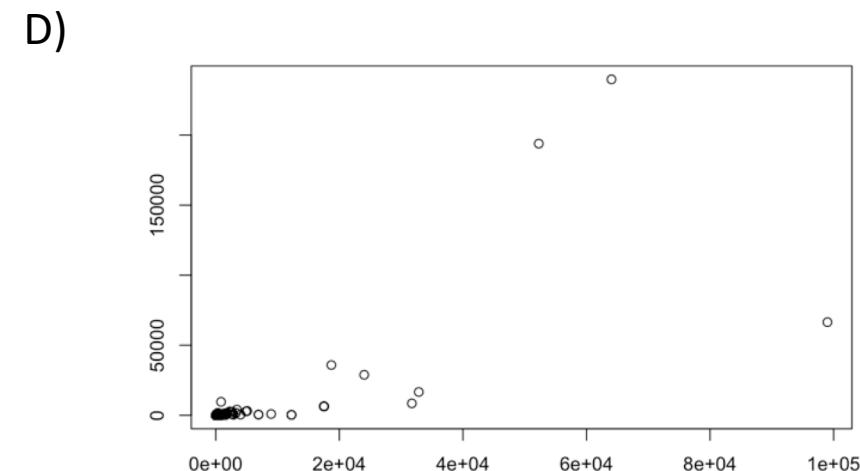
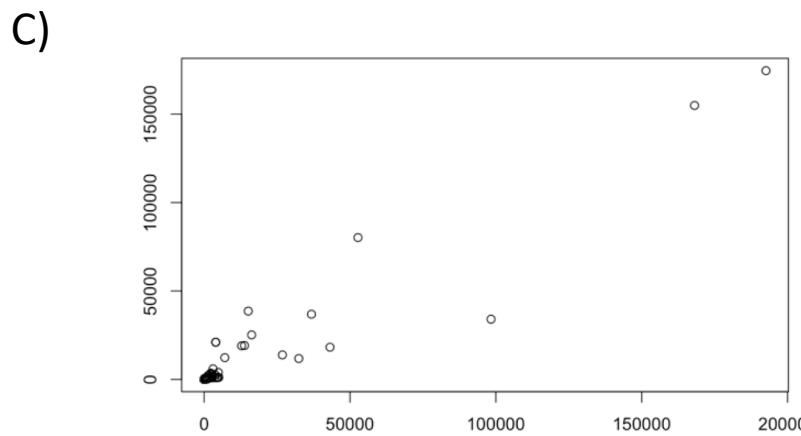
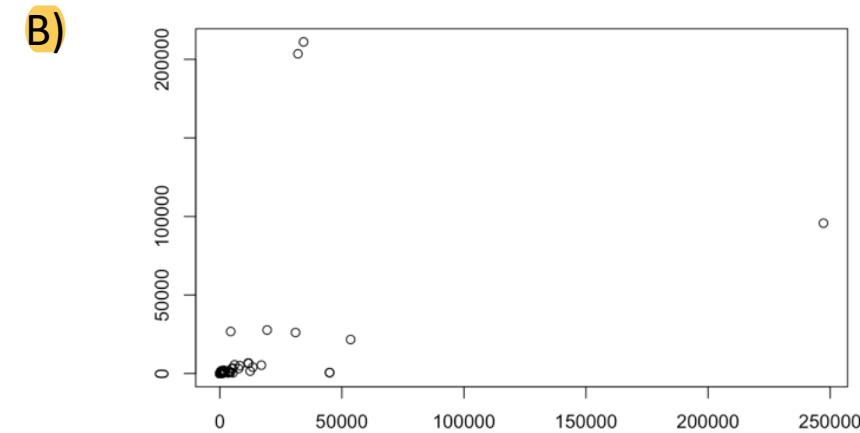
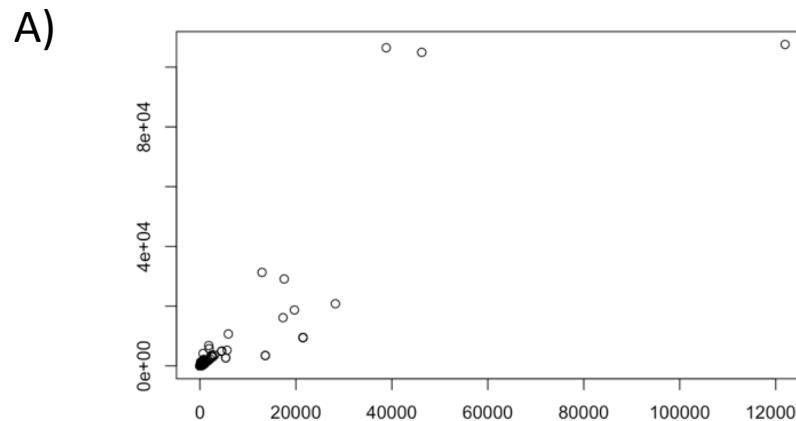
- Plot the genes KITA_10 vs KITA_11. Which of the following graphs is the correct output of the plot function?



Assignment 5

- Question 4

- Plot the genes KITA_15 vs KITA_16. Which of the following graphs is the correct output of the plot function?

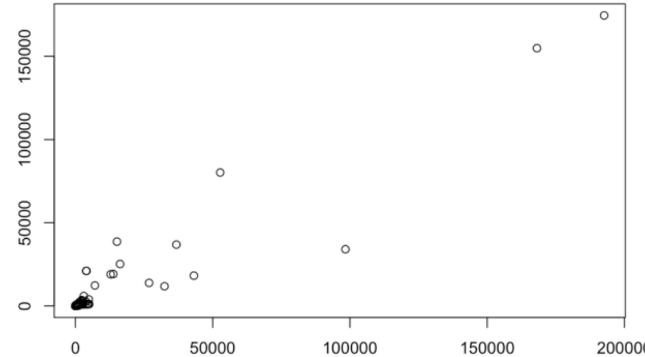


Assignment 5

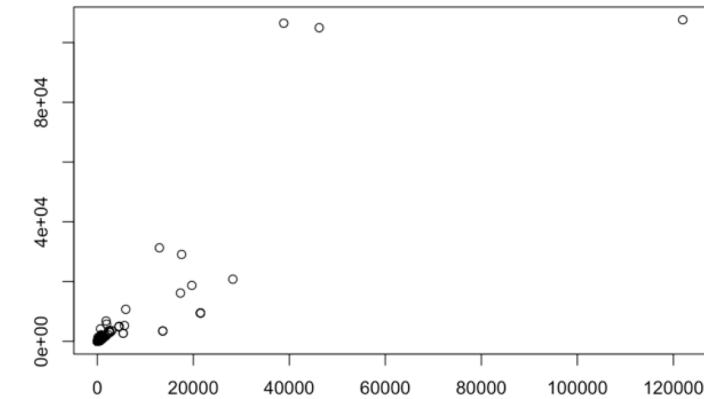
- Question 5

- Plot the genes KITB_15 vs KITB_16. Which of the following graphs is the correct output of the plot function?

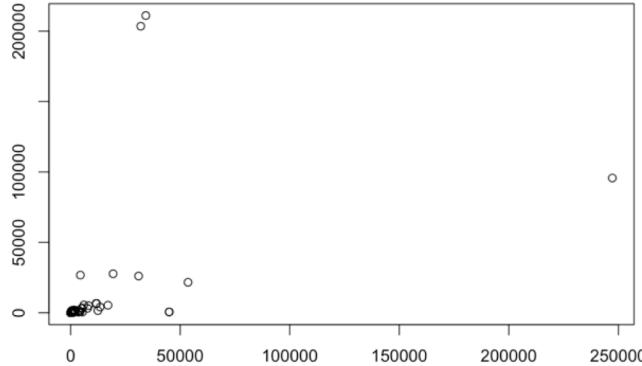
A)



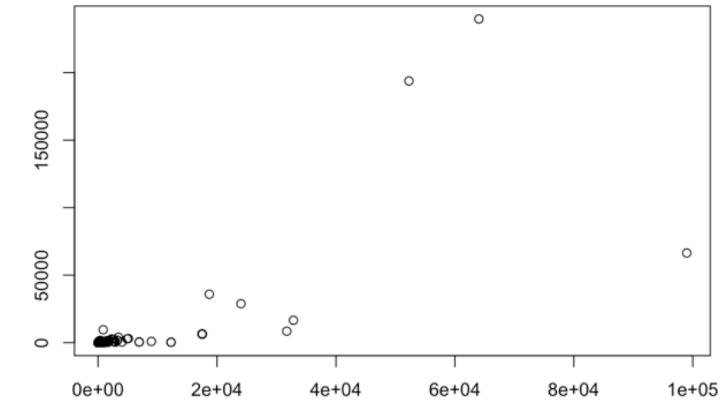
B)



C)



D)

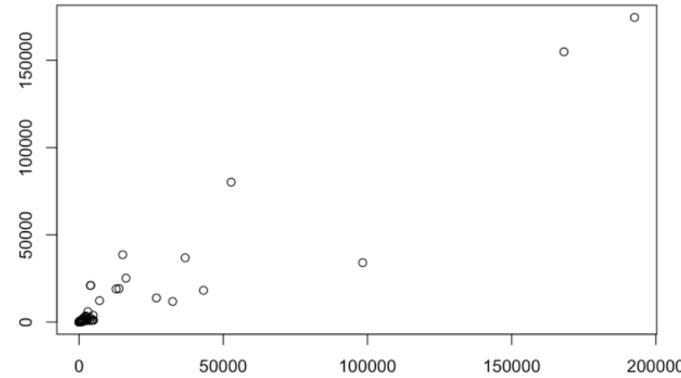


Assignment 5

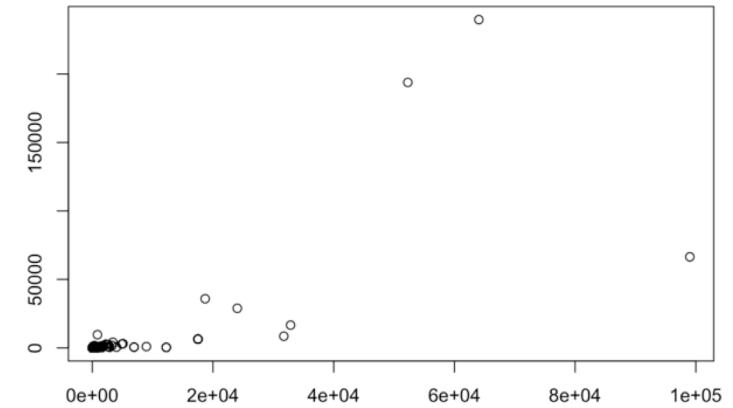
- Question 6

- Make a distribution plot of KITB_11 using the “density()” function. Which of the following graphs is the correct output of the plot function?

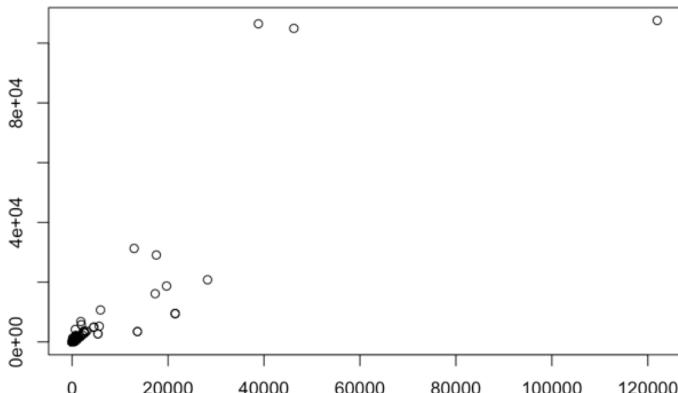
A)



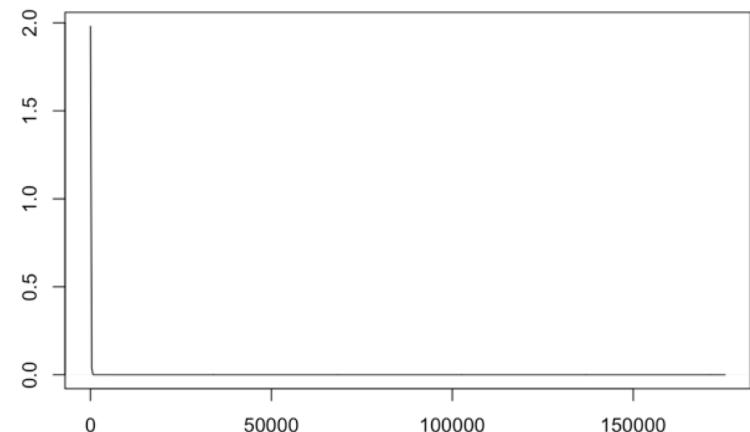
B)



C)



D)



Assignment 5

- Question 7

- What is the N of the KITB_11?
 - A) 61250
 - B) 63677
 - C) 31233
 - D) 63378

Assignment 5

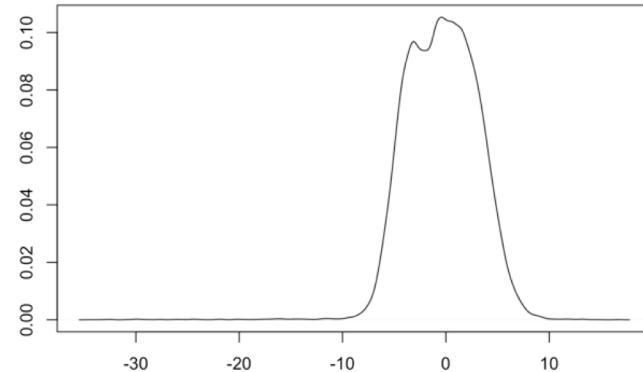
- Question 8
 - What is the Bandwith of the KITB_11?
 - A) 0.250
 - B) 0.197
 - C) 0.233
 - D) 0.378

Assignment 5

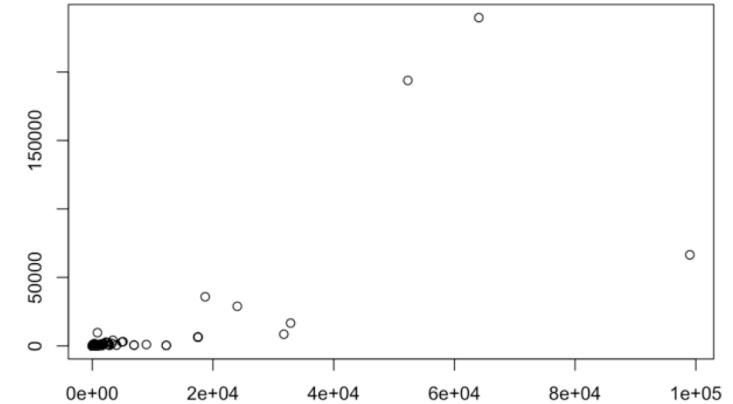
- Question 9

- Make a density plot with the log2 transformation of the values in KITA_24. Which of the following graphs is the correct output of the plot function?

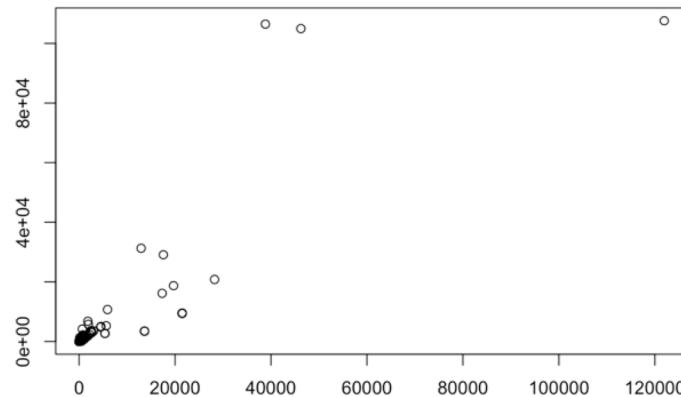
A)



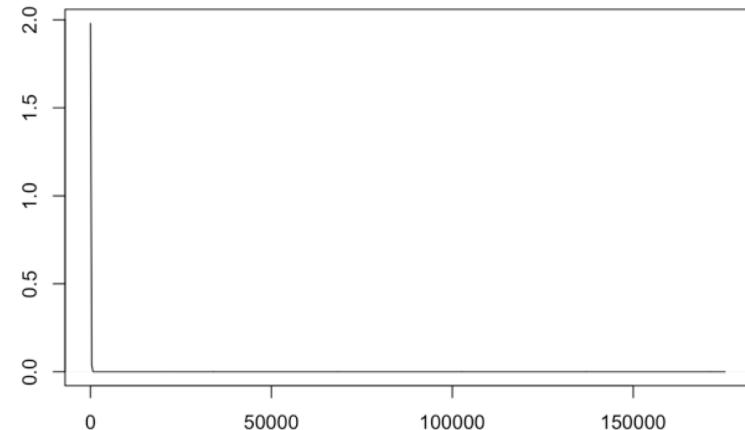
B)



C)



D)



Assignment 5

- Question 10
 - What is the N of the KITA_24?
 - A) 61250
 - B) 63677
 - C) 31233
 - D) 55264

Assignment 5

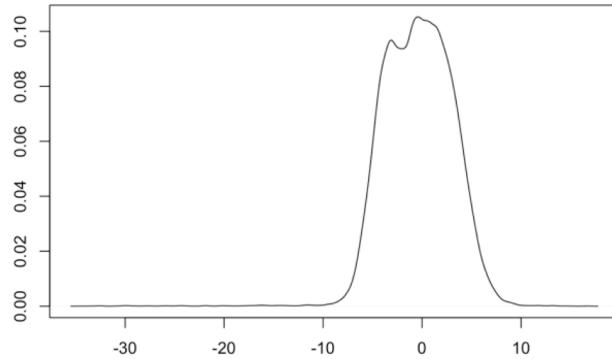
- Question 11
 - What is the Bandwidth of the KITA_24?
 - A) 0.2501
 - B) 0.1978
 - C) 0.3483
 - D) 0.3785

Assignment 5

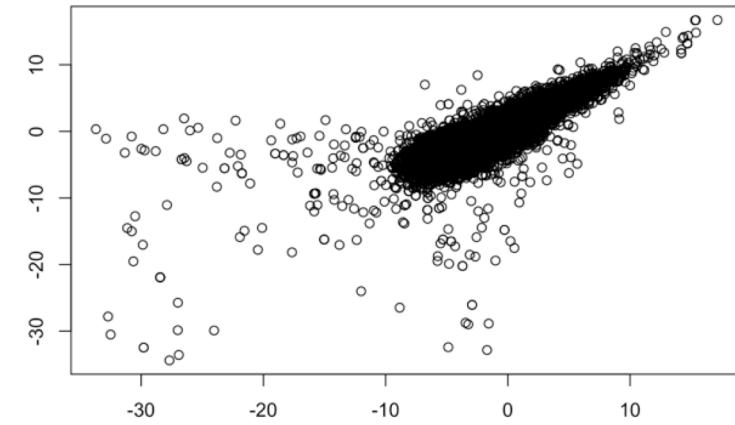
- Question 12

- Plot the log2 values of KITA_23 VS KITA_24. Which of the following graphs is the correct output of the plot function?

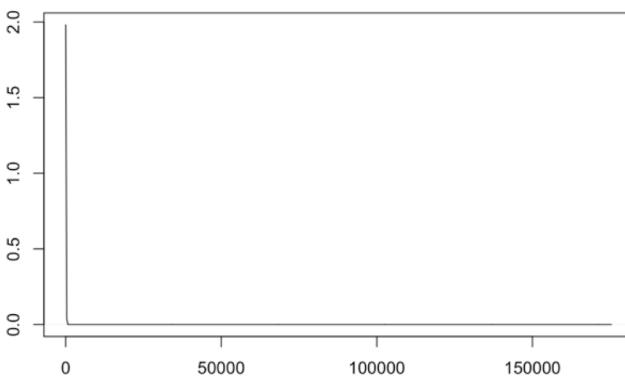
A)



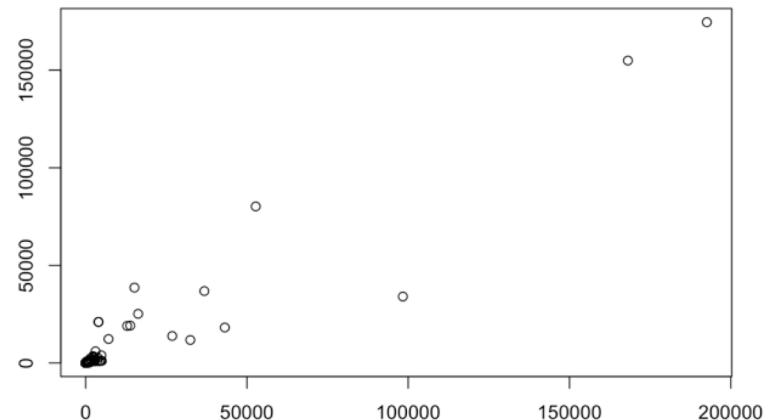
B)



C)



D)



Assignment 5

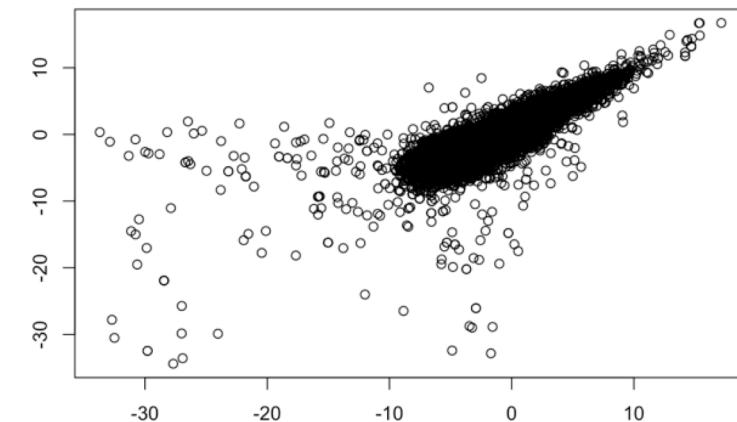
- Question 13

- Correlate all the genes in the expression data file and show the first five rows and the last five columns. Which of the following is the correct output?

A)

```
##          KITA_01   KITA_02   KITA_03   KITA_04   KITA_05  
## KITA_01 1.0000000 0.9172325 0.7059474 0.7890462 0.8727615  
## KITA_02 0.9172325 1.0000000 0.6796419 0.8972362 0.9245204  
## KITA_03 0.7059474 0.6796419 1.0000000 0.7943030 0.7975012  
## KITA_04 0.7890462 0.8972362 0.7943030 1.0000000 0.9774844  
## KITA_05 0.8727615 0.9245204 0.7975012 0.9774844 1.0000000
```

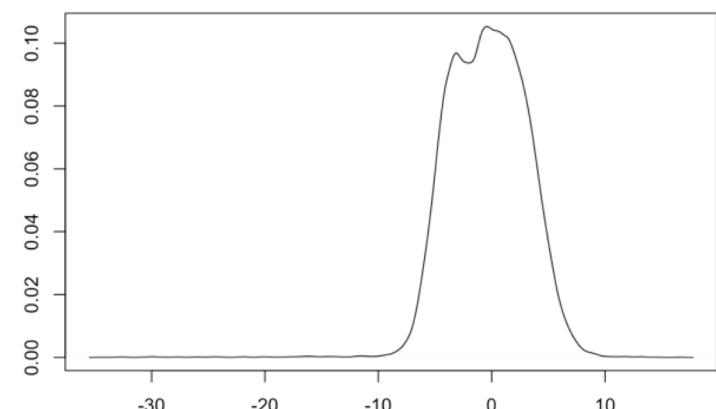
B)



C)

```
##          KITB_20   KITB_21   KITB_22   KITB_23   KITB_24  
## KITA_01 0.6029237 0.6111609 0.5223321 0.7356004 0.5433894  
## KITA_02 0.7448084 0.7477869 0.6528520 0.8656129 0.6747673  
## KITA_03 0.7571740 0.7517584 0.7331815 0.7379933 0.7416544  
## KITA_04 0.9529793 0.9540783 0.9143113 0.9674661 0.9248884  
## KITA_05 0.9020606 0.9048462 0.8526287 0.9441723 0.8655004
```

D)



Assignment 5

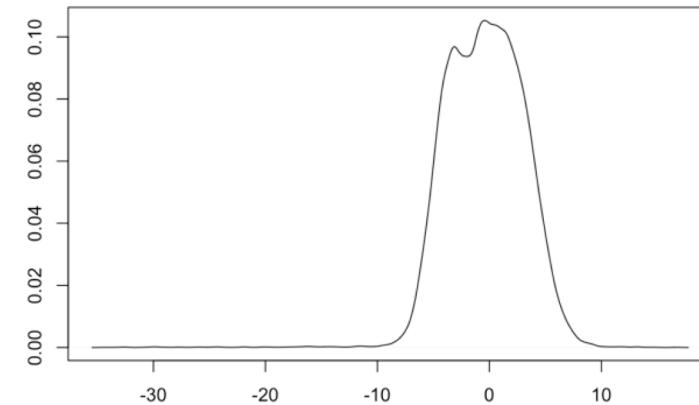
- Question 14

- Correlate all the genes in the expression data file and show the first five rows and the first five columns. Which of the following is the correct output?

A)

```
##          KITA_01   KITA_02   KITA_03   KITA_04   KITA_05
## KITA_01 1.0000000 0.9172325 0.7059474 0.7890462 0.8727615
## KITA_02 0.9172325 1.0000000 0.6796419 0.8972362 0.9245204
## KITA_03 0.7059474 0.6796419 1.0000000 0.7943030 0.7975012
## KITA_04 0.7890462 0.8972362 0.7943030 1.0000000 0.9774844
## KITA_05 0.8727615 0.9245204 0.7975012 0.9774844 1.0000000
```

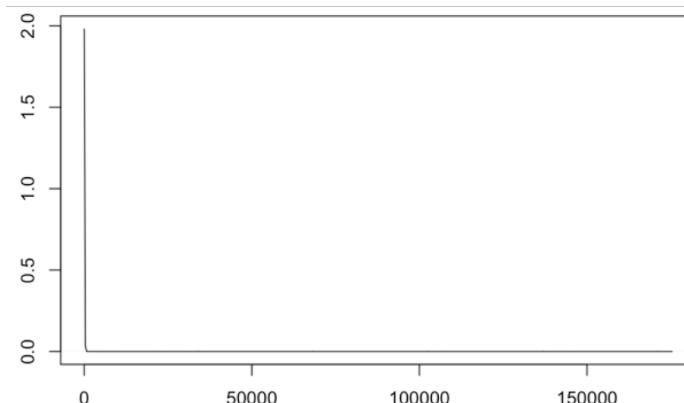
B)



C)

```
##          KITB_20   KITB_21   KITB_22   KITB_23   KITB_24
## KITA_01 0.6029237 0.6111609 0.5223321 0.7356004 0.5433894
## KITA_02 0.7448084 0.7477869 0.6528520 0.8656129 0.6747673
## KITA_03 0.7571740 0.7517584 0.7331815 0.7379933 0.7416544
## KITA_04 0.9529793 0.9540783 0.9143113 0.9674661 0.9248884
## KITA_05 0.9020606 0.9048462 0.8526287 0.9441723 0.8655004
```

D)



Assignment 5

- Question 15

- Using the `melt()` and `ggplotly()` functions, create an interactive heatmap of the expression data. Which of the following graphs is the correct output?

