EXPLORATORY ANALYSIS USING R

After obtaining single dataframe for our six samples from previous exercise, our data is ready for the actual graphical and statistical analysis.

➢ Herein, exploratory analysis we have several tools including **plot, Bar plot, Box plot, principle component analysis (PCA)** et cetra for analyzing graphically.

The earlier data, we had

geneid <int>	sample1 <int>	sample2 <int>	sample3 <int>	sample4 <int>	sample5 <int>	sample6 <int>
653635	2158	465	1753	3132	3428	2848
100422834	0	0	0	2	0	1
645520	6	0	1	0	0	1
79501	0	0	0	0	0	0
729737	283	18	222	494	381	419
100507658	45	1	23	42	48	46
100132287	491	35	298	616	640	671
100288646	8	4	18	45	36	44
729759	0	1	0	6	9	1
100131754	223447	292009	117502	31219	20557	31935

(DATAFRAME OF SIX SAMPLES (.tsv file format))

NOTE: Above dataset only contains few of the rows.

And phenotypic data,

| SAMPLE	GROUP	SEX	AGE	RIN
R3098_DLPFC_polyA_RNAseq_total_SRR1554535	ADULT	MALE	41.58	8.7
R3098_DLPFC_polyA_RNAseq_total_SRR1554536	ADULT	FEMALE	44.17	5.3
R3467_DLPFC_polyA_RNAseq_total_SRR1554561	ADULT	MALE	43.88	8.7
R3485_DLPFC_polyA_RNAseq_total_SRR1554541	FETAL	MALE	-0.384	5.7
R3452_DLPFC_polyA_RNAseq_total_SRR1554537	FETAL	FEMALE	-0.384	9.6
R4707_DLPFC_polyA_RNAseq_total_SRR1554567	FETAL	MALE	-0.499	8

(PHENOTYPIC INFORMATION OF THE SAMPLE DATASETS)

After performing simple **plot representation** on our data shows: **(Figure**
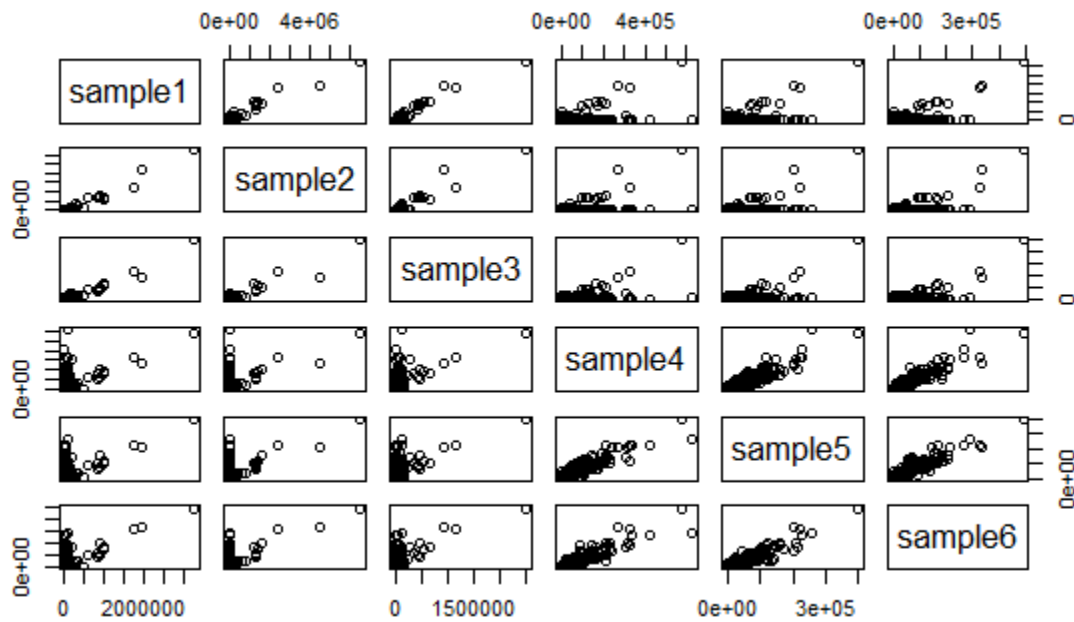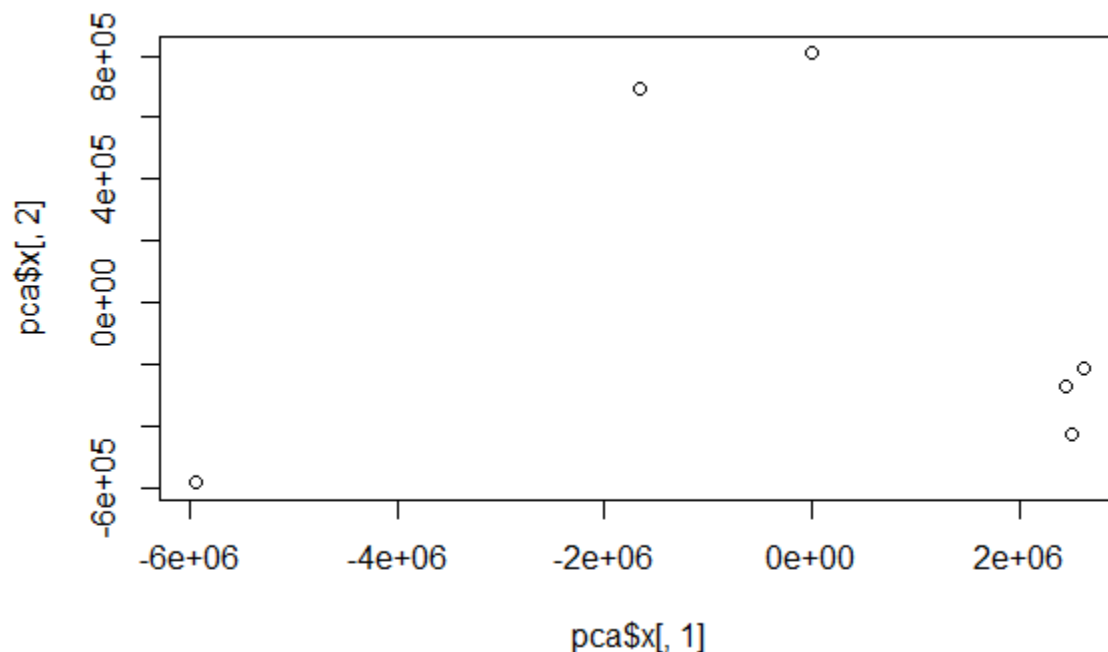

(PLOT REPRESENTATION OF SIX SAMPLES)

After performing simple **BOX PLOT representation** on our sample datasets:
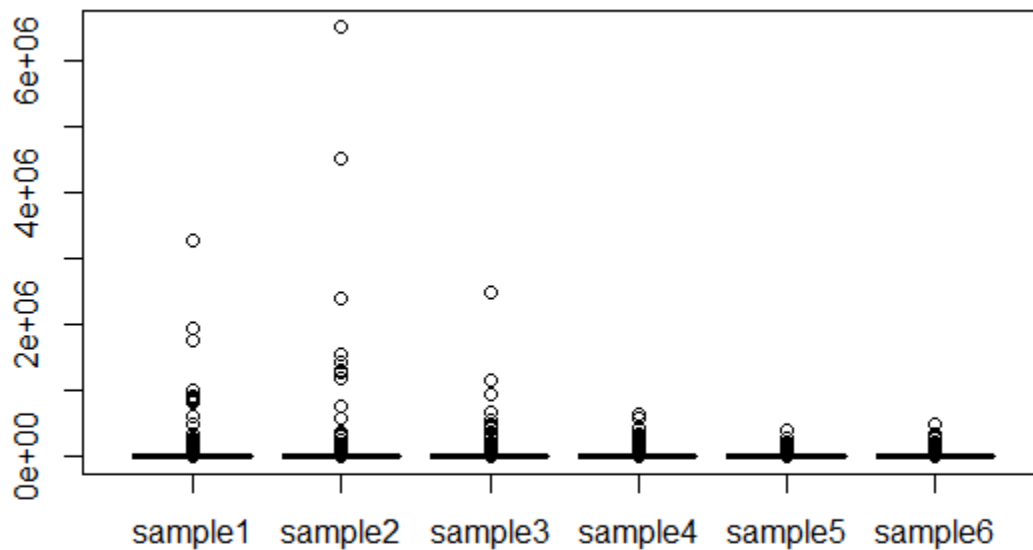
After performing simple **PCA plot representation** on our data shows:

(PCA PLOT REPRESENTATION OF OUR SIX SAMPLES)

INTERPRETATION:

Transcriptome sequences (RNA-seq) of dorsolateral pre frontal cortex brain of fetal were **scattered** ((*Figure 14: Plot representation of six samples* ; *(Figure 15: Barplot representation of our six samples* ; *(Figure 16: PCA plot representation of our six samples*) or could **not** make a compact cluster while on the other hand, adult sample's RNA-seq expression shows **compact or dense clustering.** There how, we can conclude that as human age changes from fetal to an adult the RNA-seq expression gets more compact or it converges towards a certain cluster.

In *(Figure 14: Plot representation of six samples*, sample-1, sample-2 and sample-3 more lies on x-axis. While on the other hand sample-4, sample-5 and sample-6 (adult) having more data points on x=y line. That shows, sample-1,2 and 3 having something similar data and, sample-4,5 and 6 are also having similar data.

In *(Figure 15: Barplot representation of our six samples*, herein it clearly shows that sample-1, sample-2 and sample-3 are scattered while on the other hand, sample-4, sample-5, and sample-6 are dense/compact.

In *(Figure 16: PCA plot representation of our six samples*, we have further supported our finding with PCR (principal component analysis) sample 1, sample 2 and sample 3 can be seen as it scattered on the graph while in adult samples they are dense towards 2e+06.