***Genomic Data Science Capstone: Week 7 (Exploratory Analysis)***

For this assignment I performed an exploratory analysis technique called Principal Component Analysis (PCA) on the results of the previous week's assignment, where we calculated gene expression profile in the different brain samples (fetal tissue vs adult tissue).

First, I calculated the total read for each sample with the reads per million (RPM) method (Graph 1 and Graph 2).

Raw gene counts for each sample

Raw gene counts for each sample
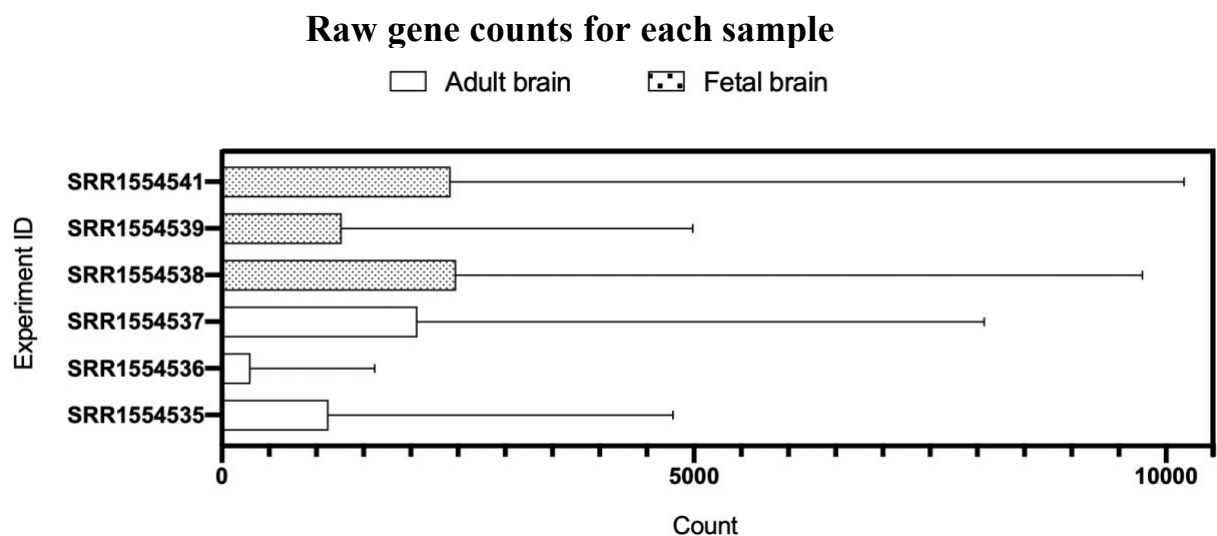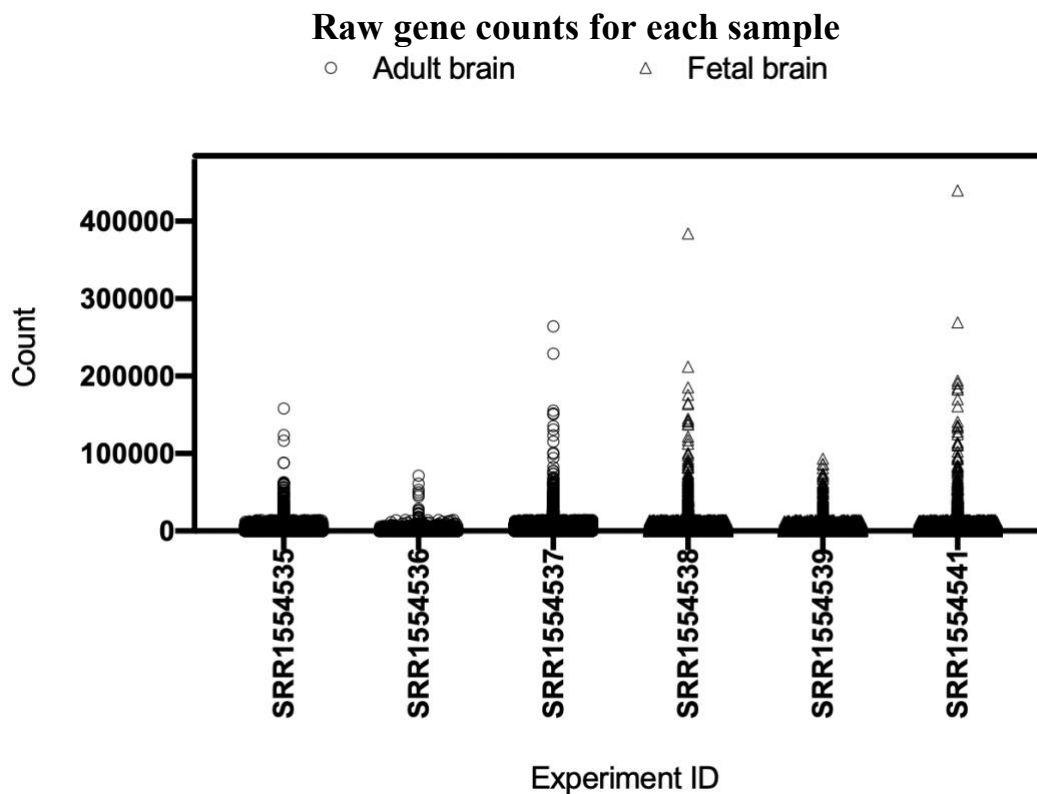
However, the data did not had the best format for statistical analysis, so I applied a log transformation with base 2 (Graph 3 and Graph 4).

Normalized logged base 2 gene counts for each sample

Normalized logged base 2 gene counts for each sample
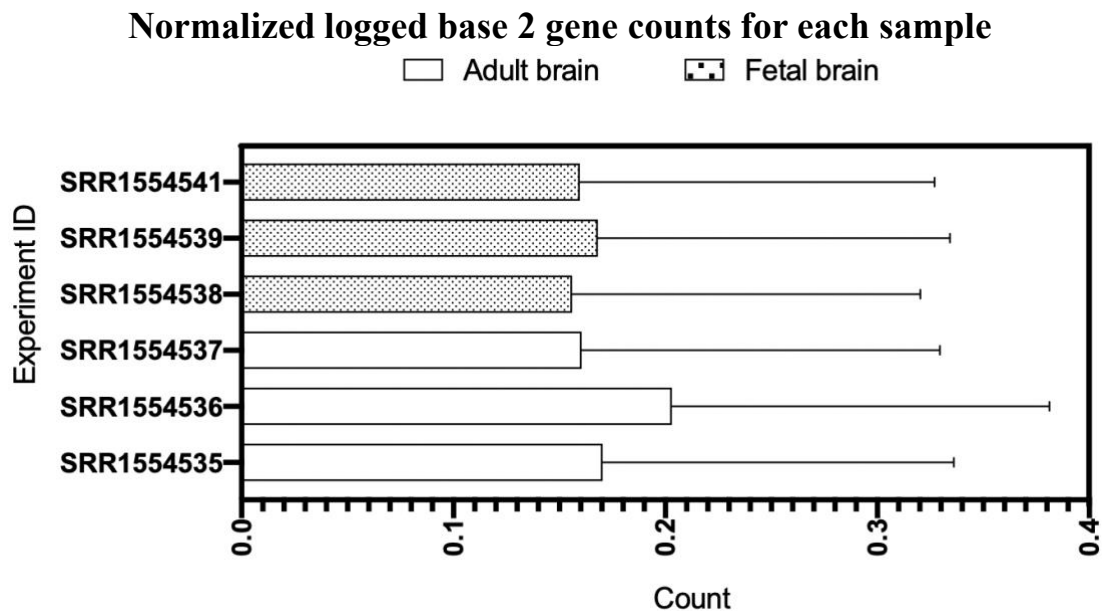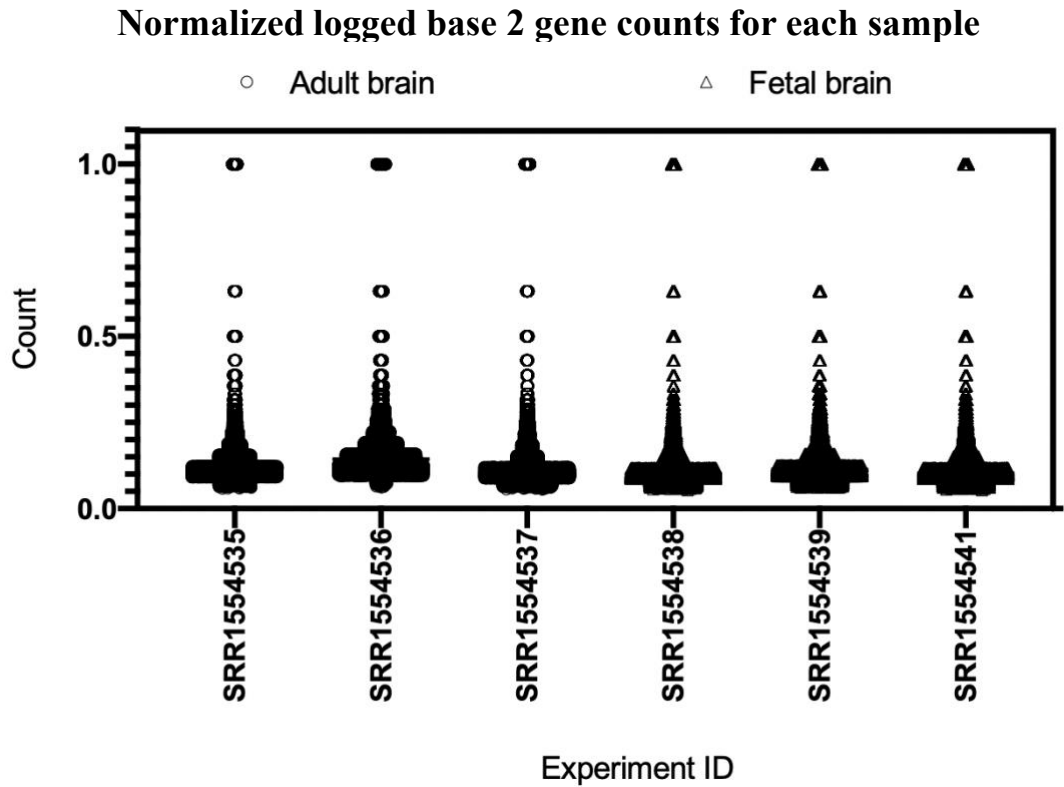
Finally, I applied PCA method with two different categorical variables: (1) Phenoype (fetal vs adult); and (2) Sex (female vs male). The results are shown in Graph 5 and 6, where is possible to observe that the variable Phenotype effectively separate between distinct clusters, but not the variable Sex.

PCA of all genes with Phenotype variable
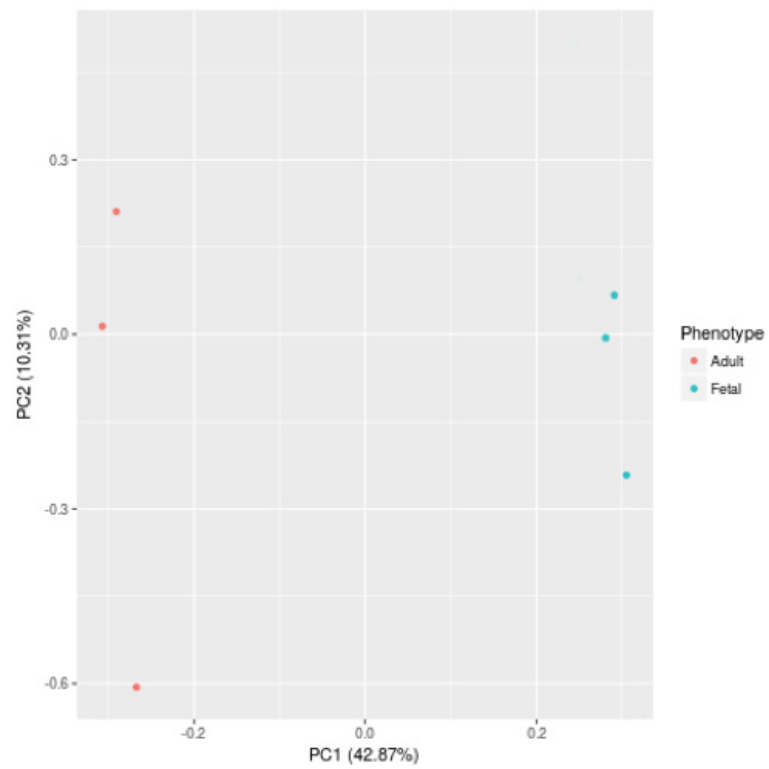
PCA of all genes with Sex variable