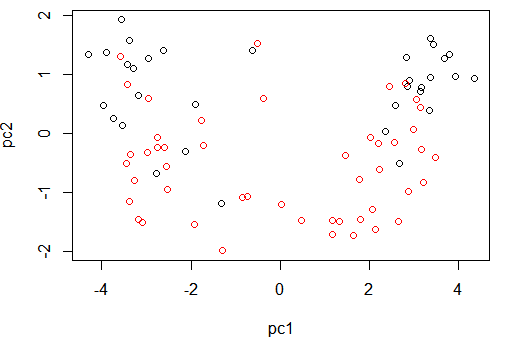
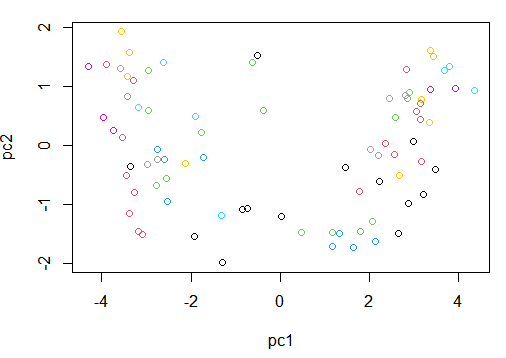
Lab 7 – Advanced Statistics

**Question 2**

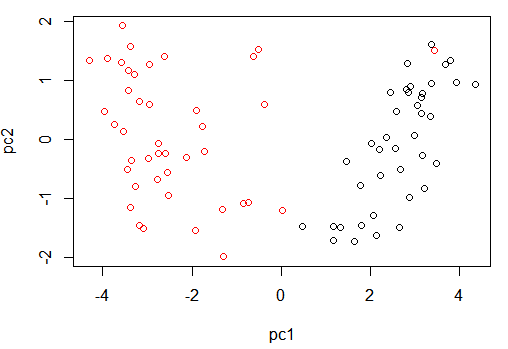
*PCA: Colored by Genotype*



*PCA: Colored by Cage*



*PCA: Colored by Time*

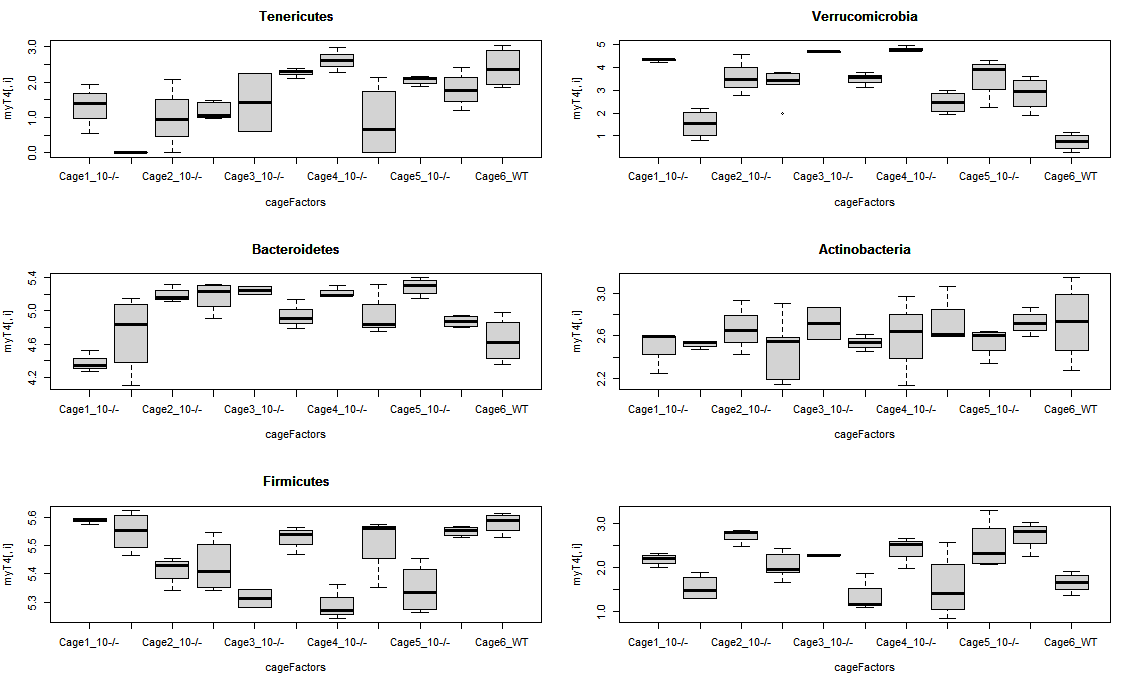


**Question 3**

|  |  |  |
| --- | --- | --- |
|  | PCA1 | PCA2 |
| Cage | 0.99206 | 1.62959e-07 |
| Genotype | 0.92970 | 1.27435e-10 |
| Time (pre vs. post) | 2.51997e-29 | 0.42682 |

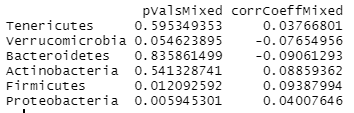
Before we talk about the results I think it’s important to note that percent of variance that is accounted for from each principle component. The principal components account for 75.5% and 10.5% for PC1 and PC2, respectively. Looking at the table above, it is clear that time is most associated with the PCA1 axis. On the other hand, it is clear that cage and genotype are mostly associated with the PCA2 axis. Overall, these results show that cage is having some effect on the data but not by much given that it’s only significant for PC2 which accounts for 10.5% of the variance. Furthermore, looking at the plot (second figure) it is clear that there are no distinct clusters being formed when colored by cage. Finally, it is clear that time is significant in relation to PC1 which accounts for a majority of the data’s variance (75%) and this is illustrated clearly in the plot (third figure) given the clear separation of the two groups.

**Question 4A**

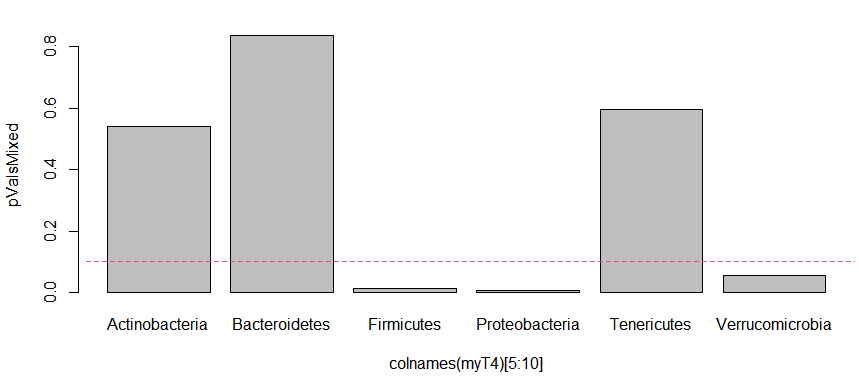


Visually inspecting the figure, there doesn’t appear to be much variation in the means as a result of cage differences. If anything it seems that only Firmicutes and Proteobacteria (figure with no title, R didn’t feel like putting a title on the last plot for some reason) seem to vary quite a bit. On the other hand, the other 4 seem very close to each other in terms of the mean.

**Question 4B**



In the table above we can see the p-vals obtained from the mixed linear model and the correlation coefficients obtained from the GLS model. Since the values of the correlation coefficients are closer to 0, then we can say that cage isn’t having much of an effect on the data.



Looking at the plot of the p-values for each phyla, we notice that Firmicutes, proteobacteria, and verrucomicrobia have p-values less 0.10 (for an FDR of 10%).