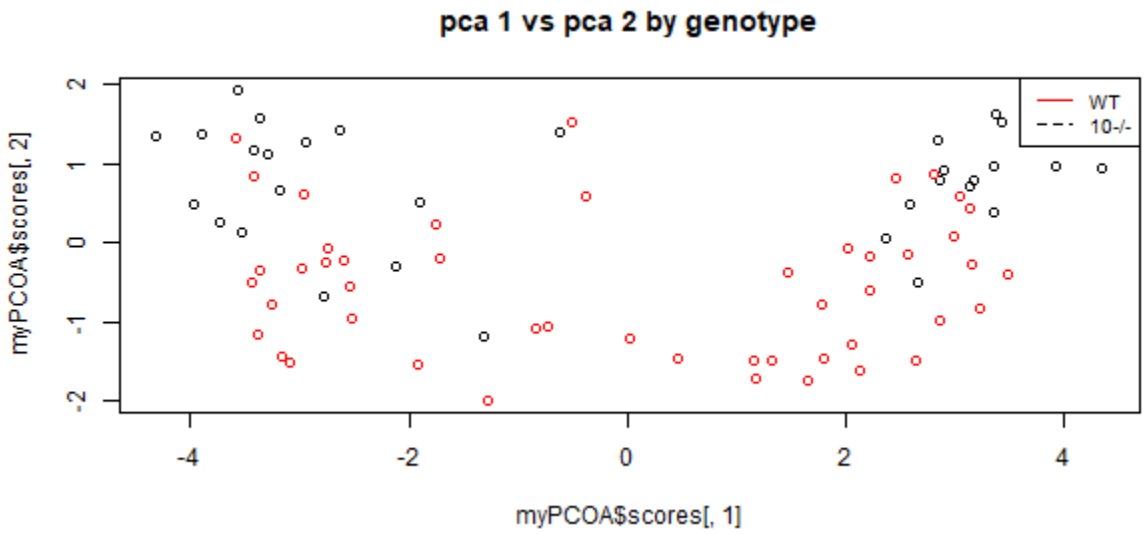
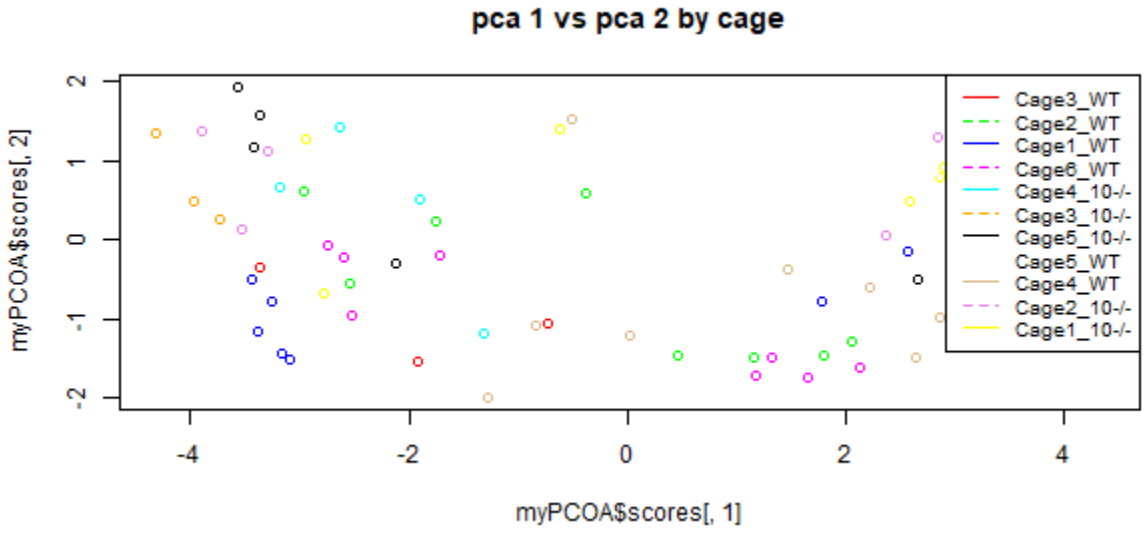
1) **Problem (1): Perform PCA ordination.**

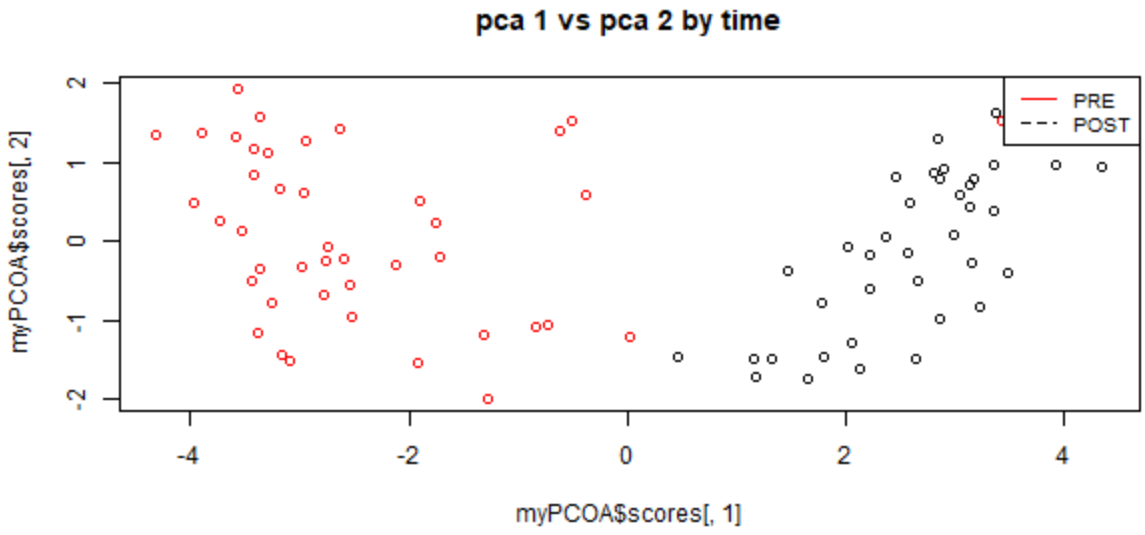
Below is just a summary of the PCA components I got-

```
> print(summary(myPCOA))
Importance of components:
                          Comp.1     Comp.2     Comp.3     Comp.4     Comp.5       Comp.6
Standard deviation     2.7823152 1.0369556 0.78000782 0.65498488 0.62347863 0.0981108931
Proportion of Variance 0.7550755 0.1048813 0.05934383 0.04184468 0.03791585 0.0009388845
Cumulative Proportion  0.7550755 0.8599567 0.91930058 0.96114526 0.99906112 1.0000000000
```

2) **Problem (2): Graph PCA1 vs. PCA2. Make three versions of the graph. One colored by genotype, one colored by cage and one colored by timepoint (pre-vs-post).**







3) **Problem (3): Fill in the following table for p-values testing the null hypothesis for PCA 1 and 2. For cage, use a way one-ANOVA. For genotype and timepoint ("pre" vs "post") use a t-test.**

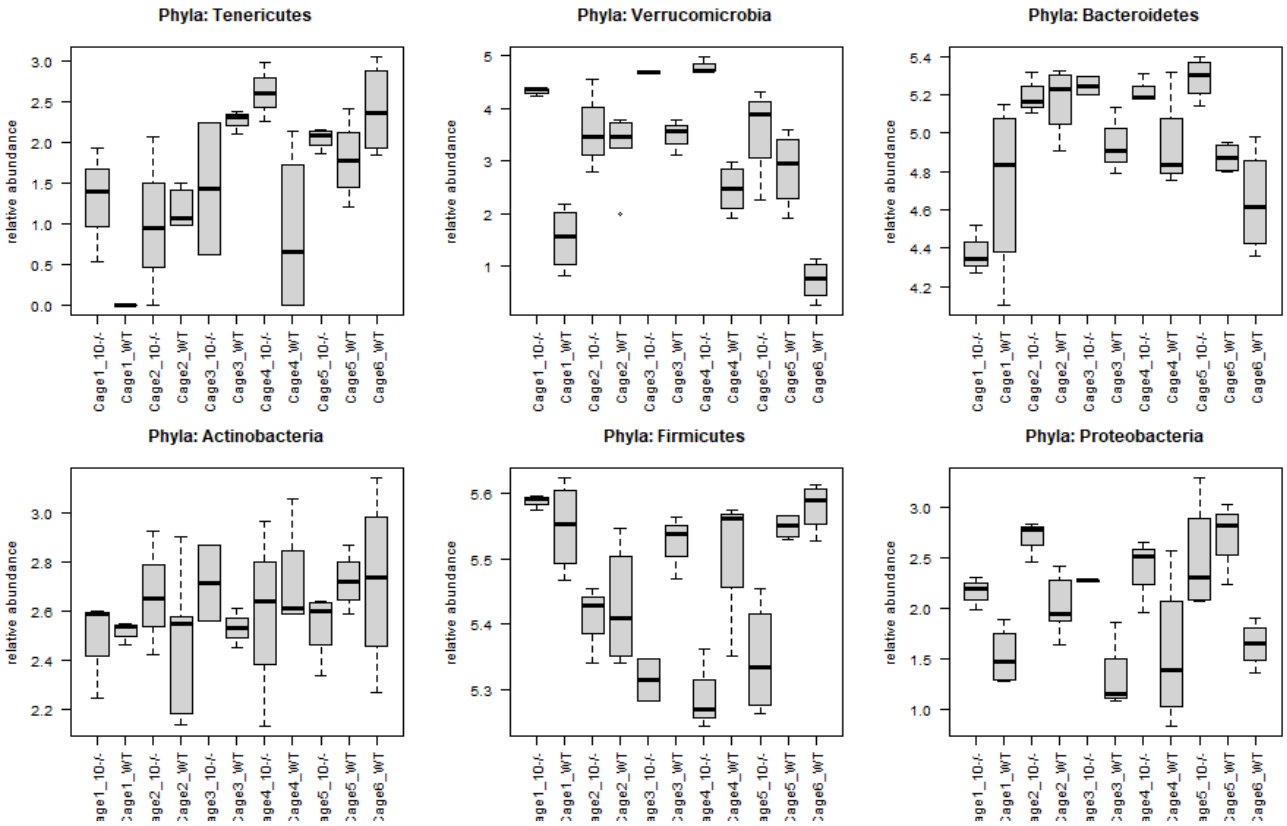
	PCA1	PCA2
Cage	**0.992**	**1.63e-07**
Genotype	**0.926**	**3.54e-10**
Time (pre vs. post)	**2.40e-31**	**0.428**

Which variable seems to be most associated with the first PCA axis? Time (pre vs. post) seems to be the most associated one with first PCA axis, and with cage and genotype the PCA 1 does not seem to vary or does not seem to have a slope which is significant.

Which variable is most associated with the second PCA axis? Genotype, even though for second PCA axis both cage and genotype are significant but it seems like genotype is a bit more associated with the second PCA axis than others.

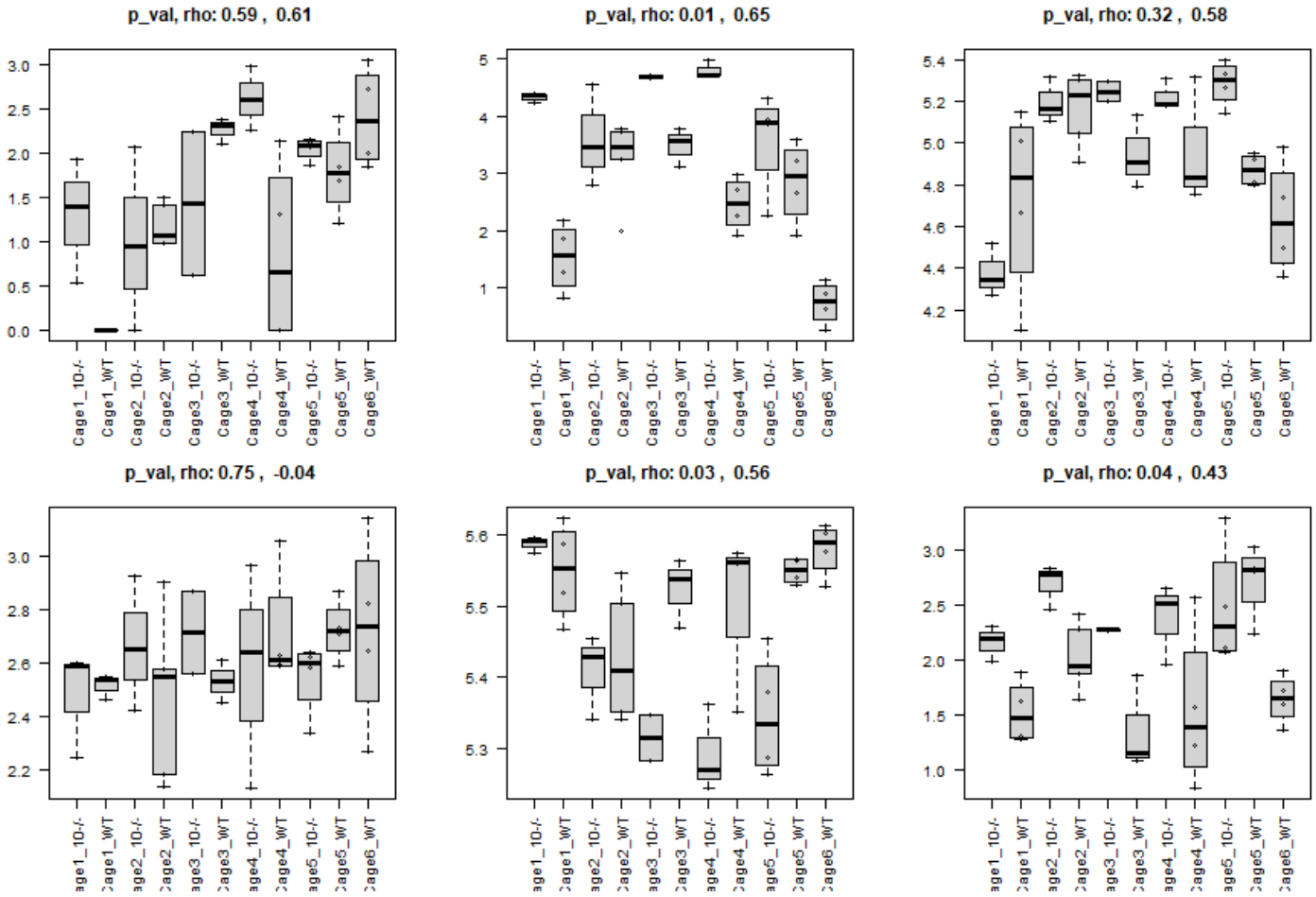
Does cage seem to be having an effect on these data? Not on the first PCA axis but Yes, cage seems to be having an (significant) effect on the second PCA axis.

4) **Problem (4A): For each phyla, graph the relative abundance of that phyla vs. cage.**



Does there appear to be a cage effect across different phyla? Yes, from the above boxplots it seems like there is a strong cage effect across different phyla.

5) **Problem (4B): For each phyla build a mixed linear model with genotype as the fixed variable and cage as a random variable. Report the intraclass correlation coefficient for each phyla.**



Are there any phyla that are significantly different for genotype in the mixed model at a 10% false discovery rate?

```
print(length(p_value_lme)) ## 6
print(sum(p_value_lme < 0.10)) ## unadjusted
p_value_lme_adj <- p.adjust (p_value_lme, method='BH')
print(sum(p_value_lme_adj < 0.10)) ## 3 for adjusted
```

Yes, there seems to be 3 phyla that are significantly different for genotype in the mixed model at a 10% FDR out of 6.