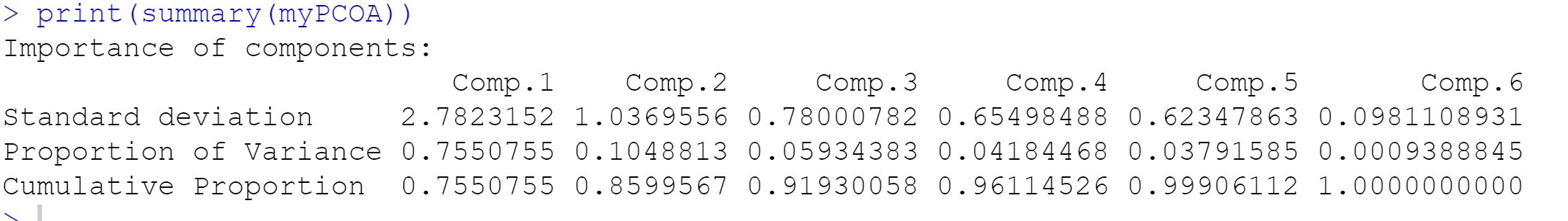
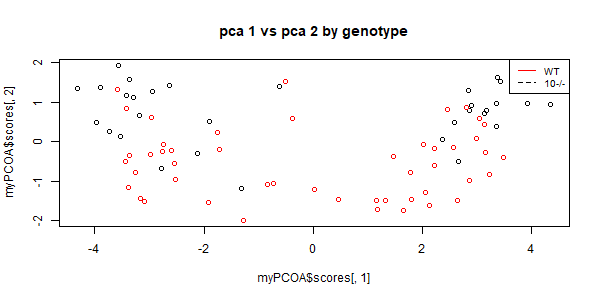
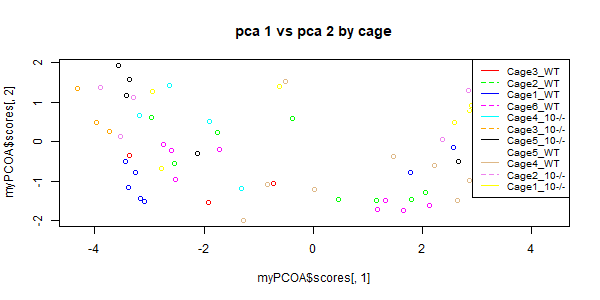
1. **Problem (1): Perform PCA ordination.**

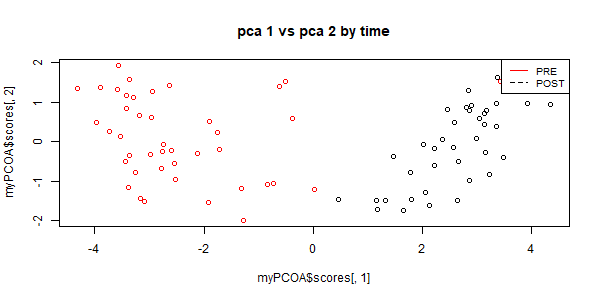
Below is just a summary of the PCA components I got-



1. **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).**







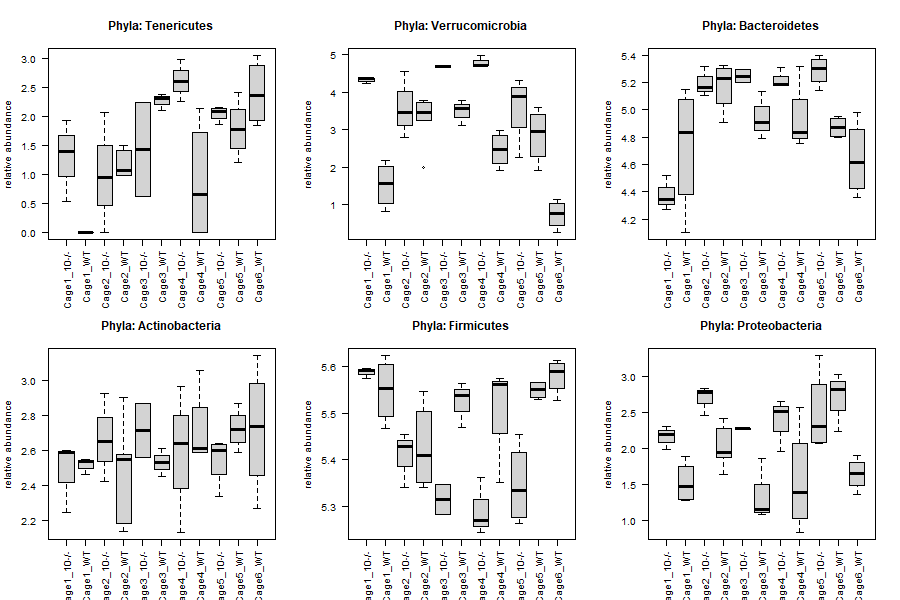
1. **Problem (3): Fill in the following table for p-values testing the null hypothesis for PCA 1 and 2. For cage, use a way one-ANOVA. For genotype and timepoint (“pre” vs “post”) use a t-test.**

|  |  |  |
| --- | --- | --- |
|  | PCA1 | PCA2 |
| Cage | **0.992** | **1.63e-07** |
| Genotype | **0.929** | **1.27e-10** |
| Time (pre vs. post) | **2.52e-29** | **0.427** |

**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.

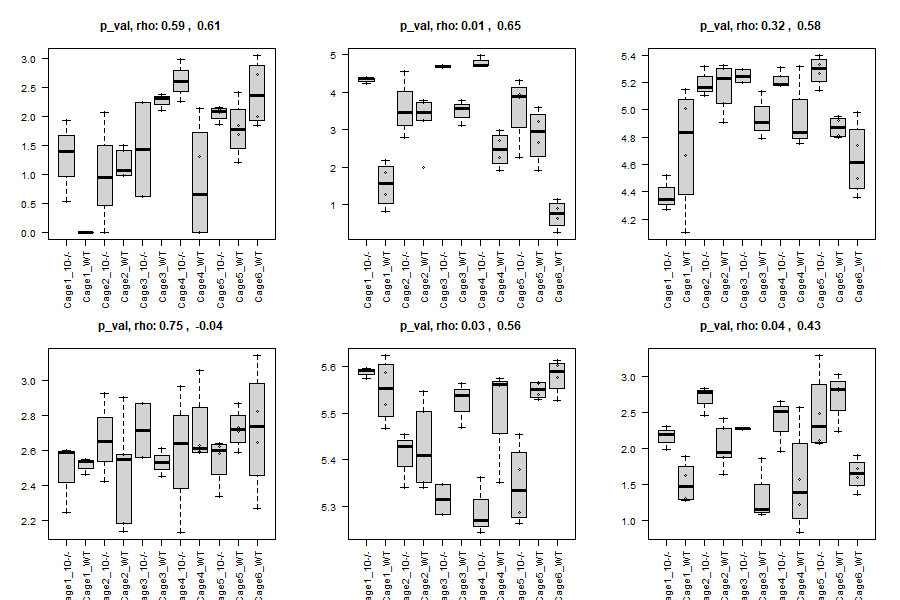
**Which variable is most associated with the second PCA axis?** Genotype, even though for second PCA axis both cage and genotype are significant but it seems like genotype is a bit more associated with the second PCA axis than others.

**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.

1. **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.

1. **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.