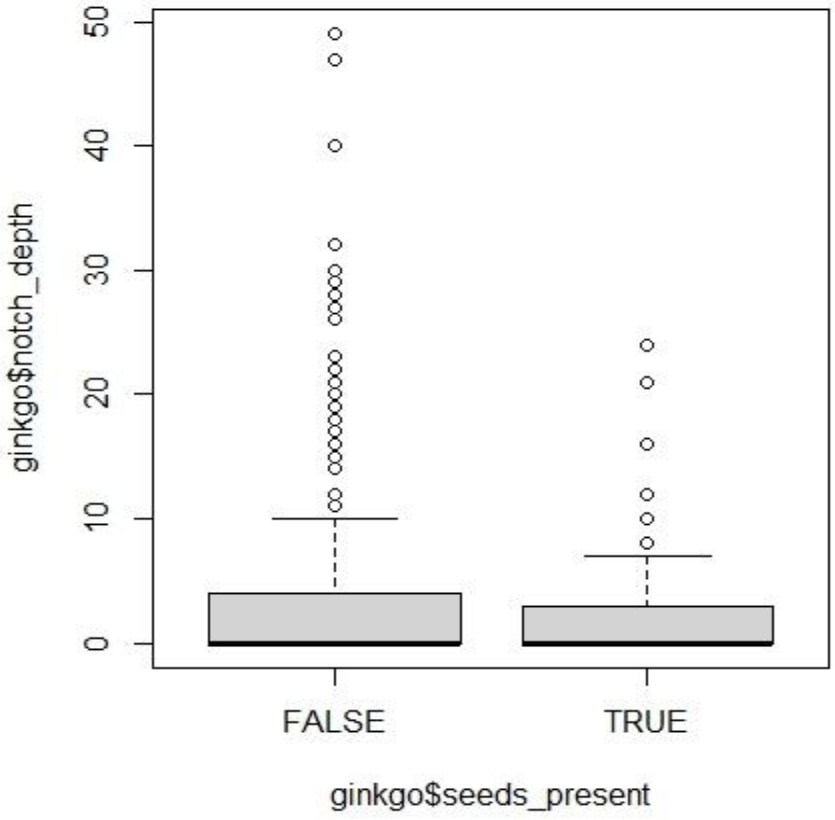
Hannah Korte, Chloe Lang, Abigail Guinan, Mercy Melo, Jeffery Larkin, Hazel Ortiz, Melissa Langley, Devon Parsons, Matthew Larosee



Q1 (1 pt.): Create a conditional boxplot of one of the continuous variables (notch) conditioned on the seeds_present column.

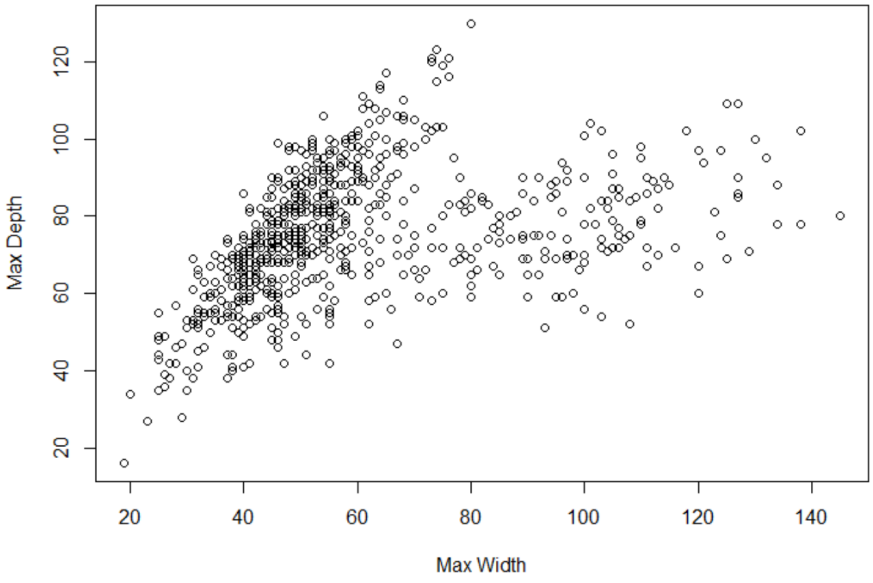
```
boxplot(ginkgo$notch_depth ~ ginkgo$seeds_present)
```

Q2 (1 pt.): Based on your boxplot, do you think there is any difference between seed bearing and non seed bearing trees? Note: this is just a preliminary data exploration, you may change your mind based on further analysis!

The variability appears to be higher on the non-seed bearing trees compared to the seed-bearing trees, but this difference does not appear to be significant considering that most of the variation occurs in the outliers.

Q3 (1 pt.): Create a scatterplot of max leaf depth (x) and max leaf width (y).

plot(ginkgos$max_depth, ginkgos$max_width xlab = "Max Width", ylab = "Max Depth")



Q4 (1 pt.): Qualitatively describe the patterns you see in the scatterplot.

From the scatterplot, we are able to observe a positive correlation between the maximum leaf width and the maximum depth. As we increase our max width (or max depth), the max depth (or vice versa) will also increase, according to the trend.

Q5 (1 pt.): Explain how our data collection procedure might have violated the fixed x assumption.

The fixed x assumption assumes that there are no errors in the predictor variable, which in this case is maximum width. Maximum width was measured by hand, leaving the potential for human error in the x-value measurements, meaning this data has the potential to violate the fixed x assumption.

Q6 (1 pt.): Name 1 or more concepts you'd like me to review or discuss in more detail during our last two class meetings.
Log transformations and examples of them