Analysis of Environmental Data

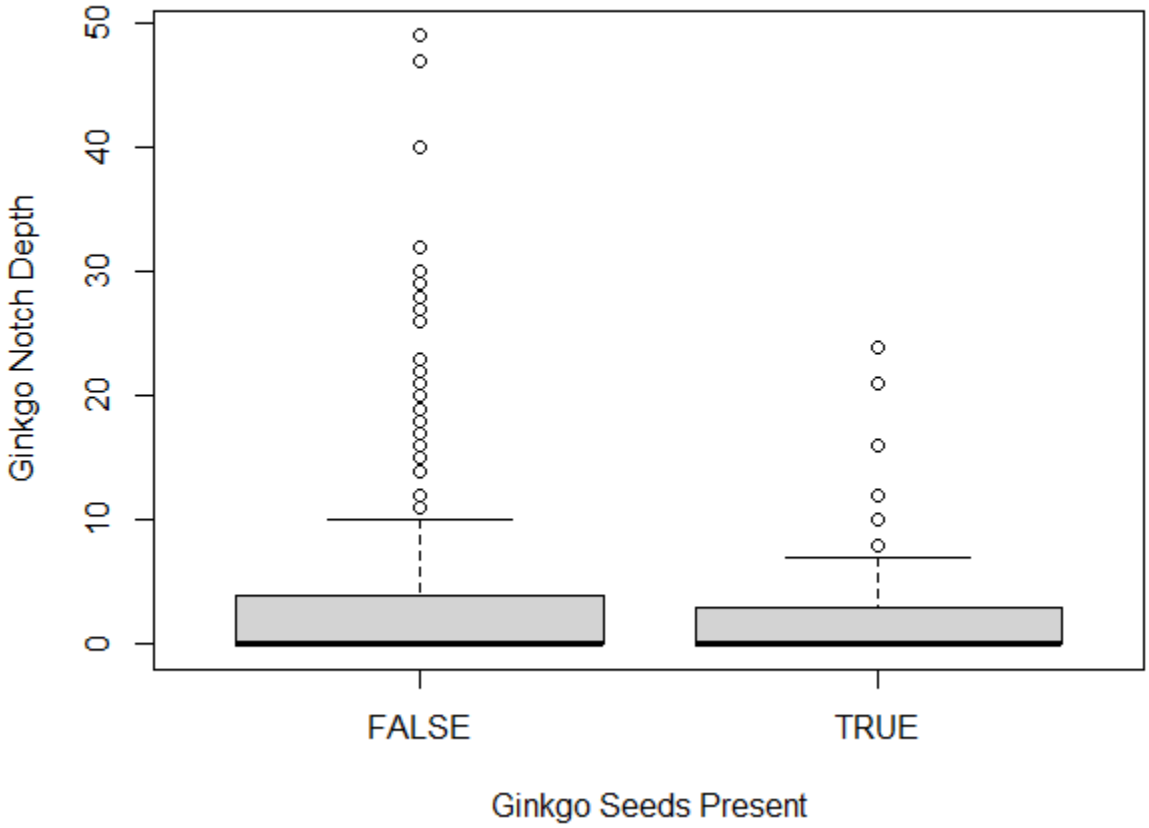
In-Class Ginkgo Data Exploration

(Zoom Group 2)

Amanda Klehr, Jessica Bonin, Amira Spikes, Matthew Fertakos, Megan Mitchell, Min Hein Htike, Pete Crump, Timothy Fistori, Ragib Ahsan

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

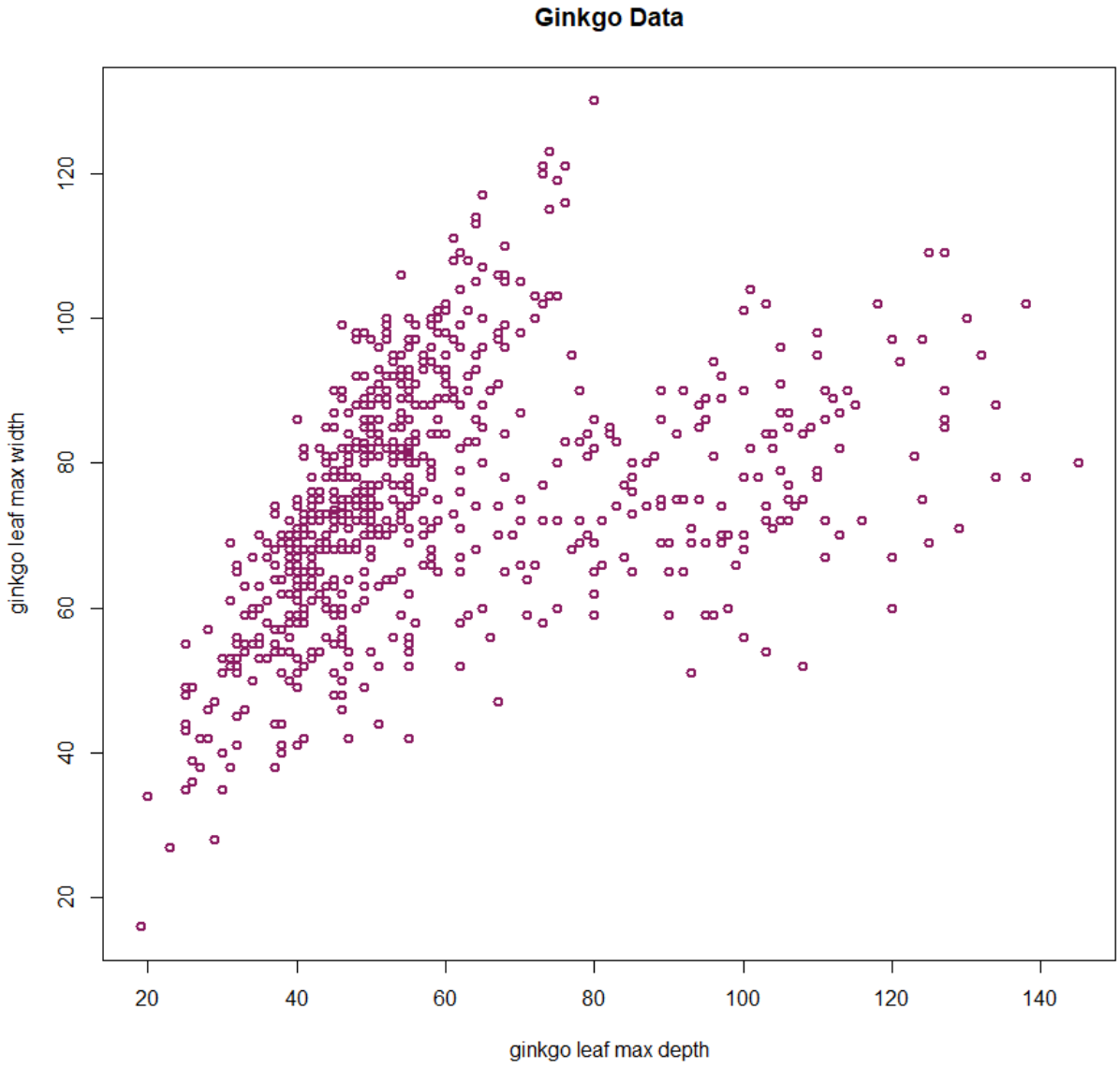
```
boxplot(dat_ginkgo$notch_depth ~ dat_ginkgo$seeds_present, data = dat_ginkgo,
    xlab = "Ginkgo Seeds Present", ylab = "Ginkgo Notch Depth")
```



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!

> Ginkgo trees without seeds present do appear to have a larger and higher range of notch depths, but this might just be because of low sampling of trees that had seeds present, so it's difficult to make an assumption based on the boxplot alone.

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



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

> It appears that as max depth increases, max width increases. There appears to be a positive slope with a strong association between max depth and width around 40 depth and 80 width. There are a few outliers as well.

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

> The “fixed-x” assumption is that the explanatory variable is measured without error. Our data collection procedures might have introduced non-random sampling errors and human errors that would violate the fixed x assumption. For example, there were multiple people collecting data, and there may be discrepancies in the way the leaves were measured.

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.

Resampling methods and frequentist vs bayesian statistics
Conditional Probabilities