

Lina Clifford, Olivia Dinkelacker, Laura Haynes, Jess Martinez

Eco Girllies

Ginkgo Data Exploration

11/15/2022

**Q1 (1 pt.): How many trees are in the dataset? Hint: Try using `subset()` with the `select` argument to create a `data.frame` that contains only the `seeds_present` and `site_id` columns. Next check out the `unique()` function.**

22

```
(nrow(ginkgo))/10
```

```
trees=(data.frame(ginkgo$site_id, ginkgo$seeds_present))
```

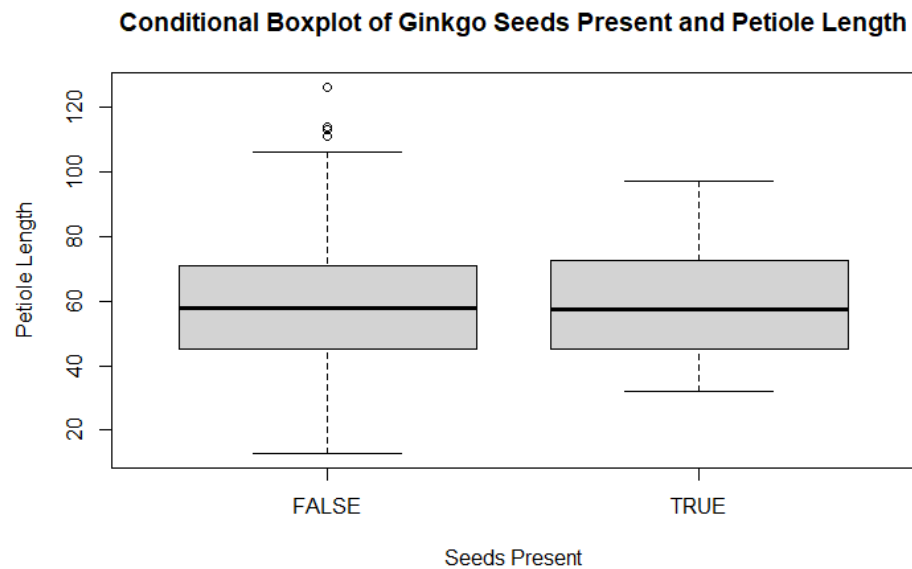
```
nrow(subset(unique(trees, select=site_id)))
```

**Q2 (1 pt.): How many trees had seeds?**

Four trees had seeds.

```
sum(ginkgo$seeds_present=="TRUE")/10
```

**Q3 (1 pt.): Include a *conditional boxplot* of one of the continuous variables conditioned on the `seeds_present` column in your report.**

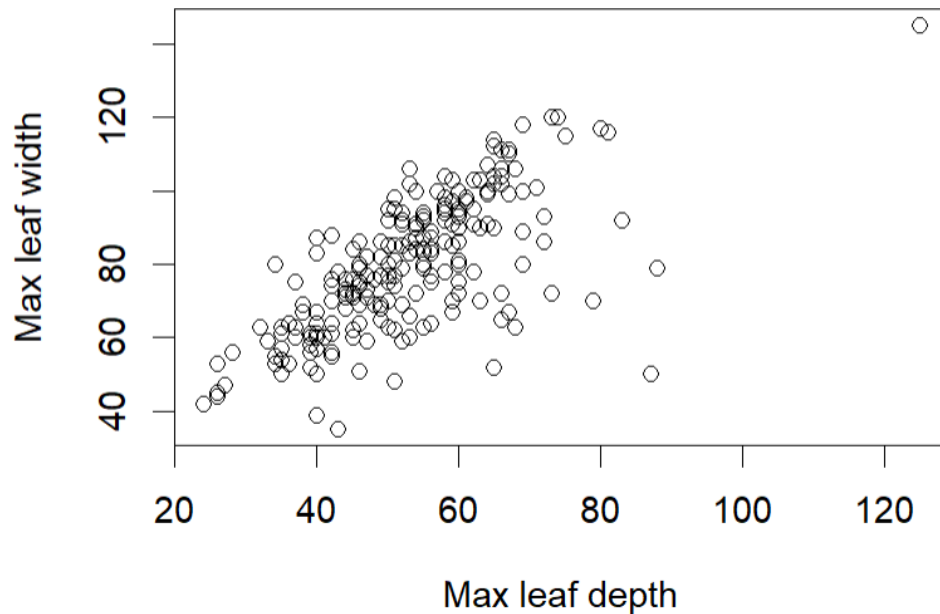


**Q4 (1 pt.):** Based on your boxplot, do you think there is any difference between seed bearing and non seed bearing trees? Note: this is just a preliminary data exploration, you may change your mind based on further analysis!

Based on preliminary exploration of the boxplot above, there does not appear to be any significant difference between seed bearing and non-seed-bearing trees when petiole length is conditioned on seed presence. The shape and size of both boxplots are quite similar, indicating that there is homogeneity of variance between the two groups.

**Q5 (1 pt.):** Create a *scatterplot* of max leaf depth (x) and max leaf width (y).

## Ginkgo Leaves



**Q6 (1 pt.): Qualitatively describe the patterns you see in the scatterplot.**

As leaf width increases, leaf depth also increases. There is a positive relationship between the two variables.

**Q7 (1 pt.): Explain how our data collection procedure might have violated the *fixed x* assumption.**

The fixed x assumption is that there are perfect measurements of the explanatory or predictor variables. Since we used paper rulers and had not pinned down or perfectly smoothed the leaves, there's likely to be many errors in measurements. Also, many different people contributed to measuring the leaves and this likely added to decreased precision.

**Q8 (1 pt.): Name 1 or more concepts you'd like me to review or discuss in more detail.**

More on ANOVA

Loops (were these part of the lecture or just lab?)

Model selection, more specifically penalizing the number of parameters.