

# JJohnson\_hw1

*Jen*

9/18/2017

## Q1

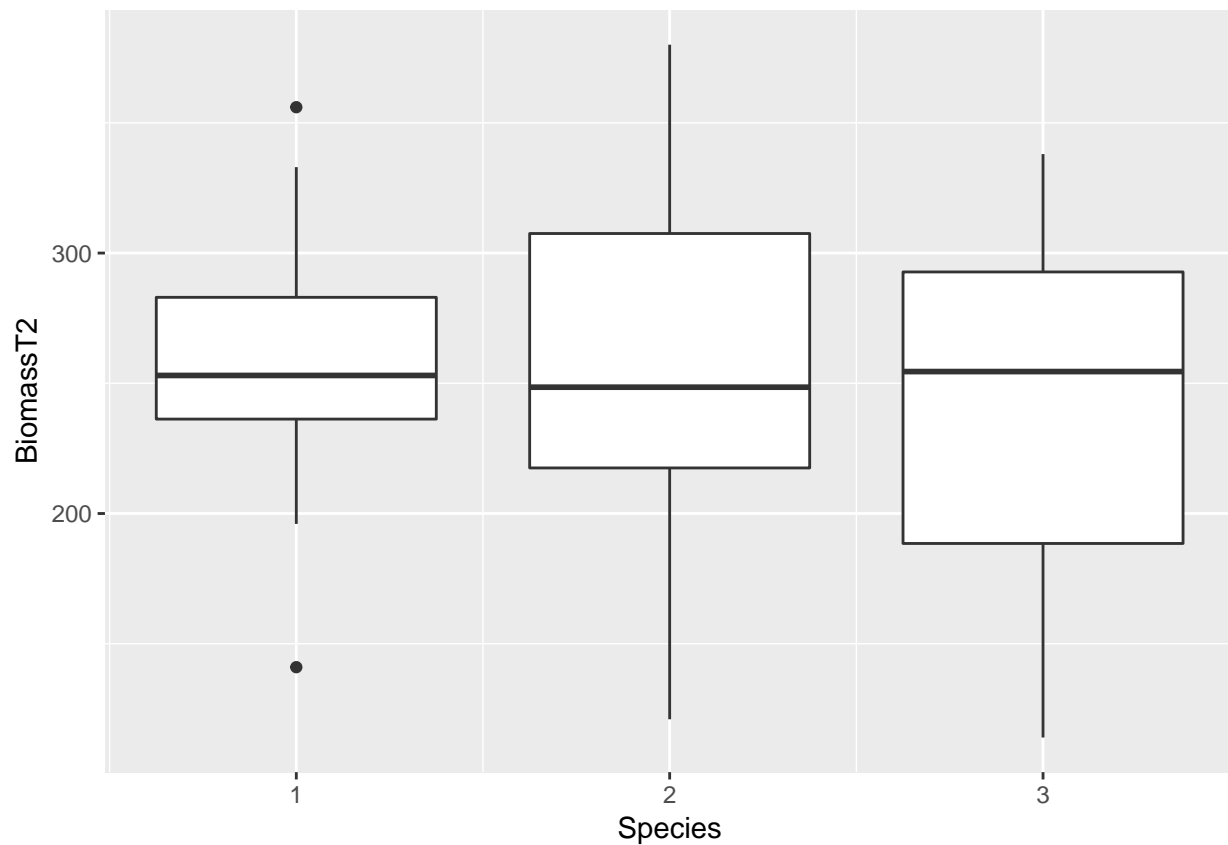
```
table(data$Species, data$Fert)
```

```
##  
##      A  B  C  
##    1  8 10 12  
##    2  9 10 11  
##    3  9 10 11
```

All 3 species got treatments of the 3 different fertilizers. Each species/fertilizer had ~11 samples, so the distribution was even. Therefore, the experimental design was acceptable. The design could be improved by having a control group for each species, where no fertilizer was added. This would have improved the validity of the comparison of fertilizers.

## Q2

```
ggplot(data, aes(x = Species, y = BiomassT2)) + geom_boxplot(aes(group = Species))
```



The distribution of final masses in the middle 50% is slightly higher for Species 2 than for Species 3. The distribution of masses for Species 1 is more precise and the middle 50% rests in a smaller box. Species 2 seems to be generally larger than the other 2 species. Species 3 has a larger distribution of sizes than Species 1, but they have similar medians.

### Q3

```
mean.species.1 <- mean(data$BiomassT2[data$Species=="1"])
mean.species.2 <- mean(data$BiomassT2[data$Species=="2"])
mean.species.3 <- mean(data$BiomassT2[data$Species=="3"])

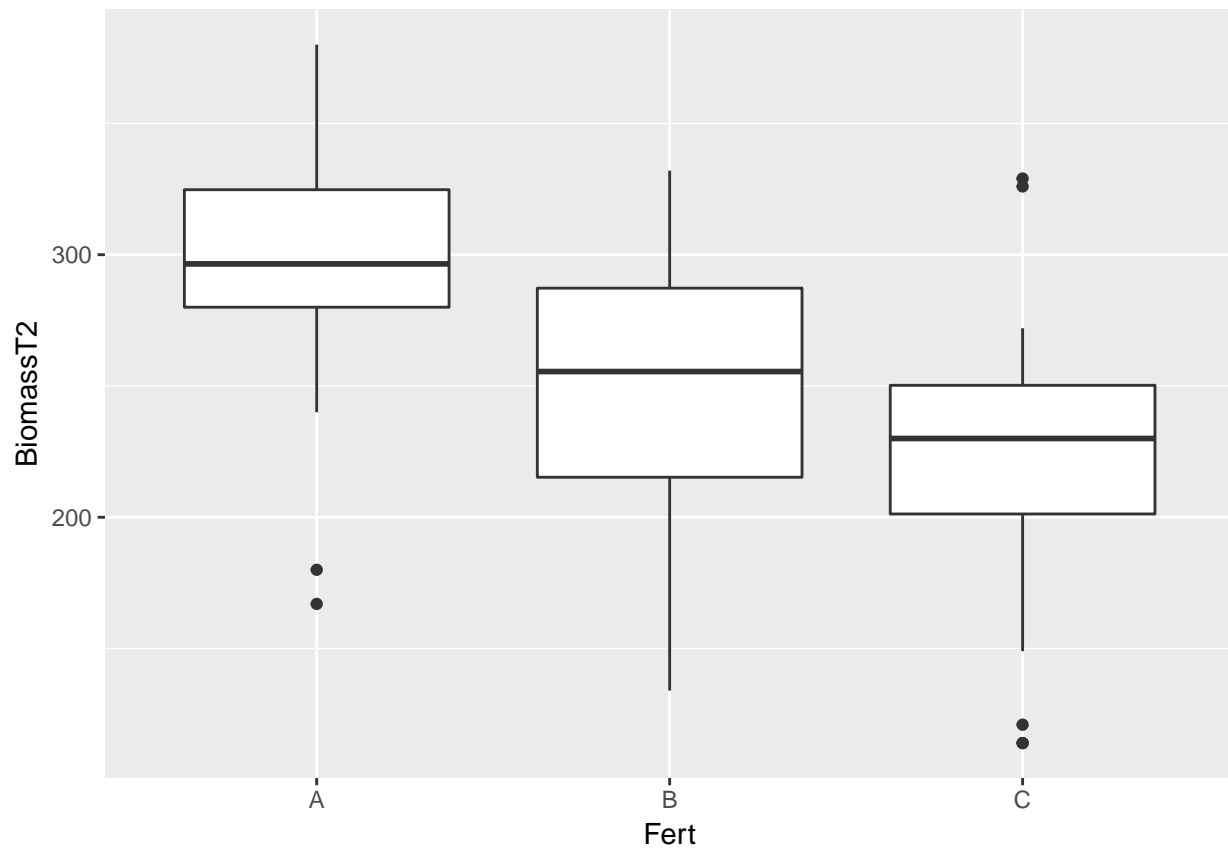
paste(mean.species.1, mean.species.2, mean.species.3, sep = " ")
```

```
## [1] "256.4 254.3 239.733333333333"
```

The mean final biomass of Species 1 is the smallest, followed by Species 3 and then Species 2. The difference between Species 1 and 2 is small.

### Q4

```
ggplot(data, aes(x = Fert, y = BiomassT2)) + geom_boxplot(aes(group = Fert))
```



```
mean_fert_A <- mean(data$BiomassT2[data$Fert=="A"])
mean_fert_B <- mean(data$BiomassT2[data$Fert=="B"])
```

```
mean_fert_C <- mean(data$BiomassT2[data$Fert=="C"])

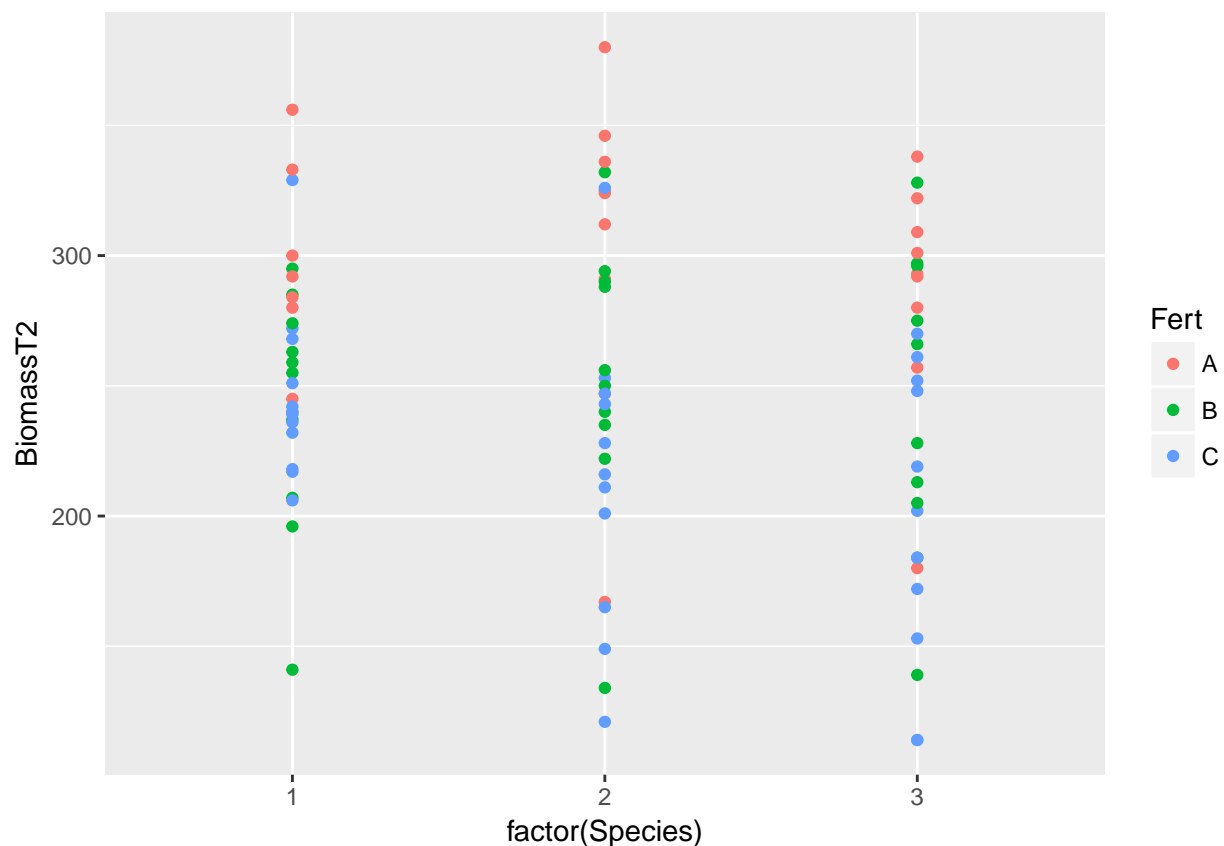
paste(mean_fert_A, mean_fert_B, mean_fert_C, sep = " ")
```

```
## [1] "293.461538461538 246.133333333333 220.558823529412"
```

Fertilizer A is the most effective because the plants with this treatment has the largest mean biomass. Furthermore, the middle 50% of Fertilizer A plants has a small range. Fertilizer B is the next effective, and Fertilizer C is the least effective.

## Q5

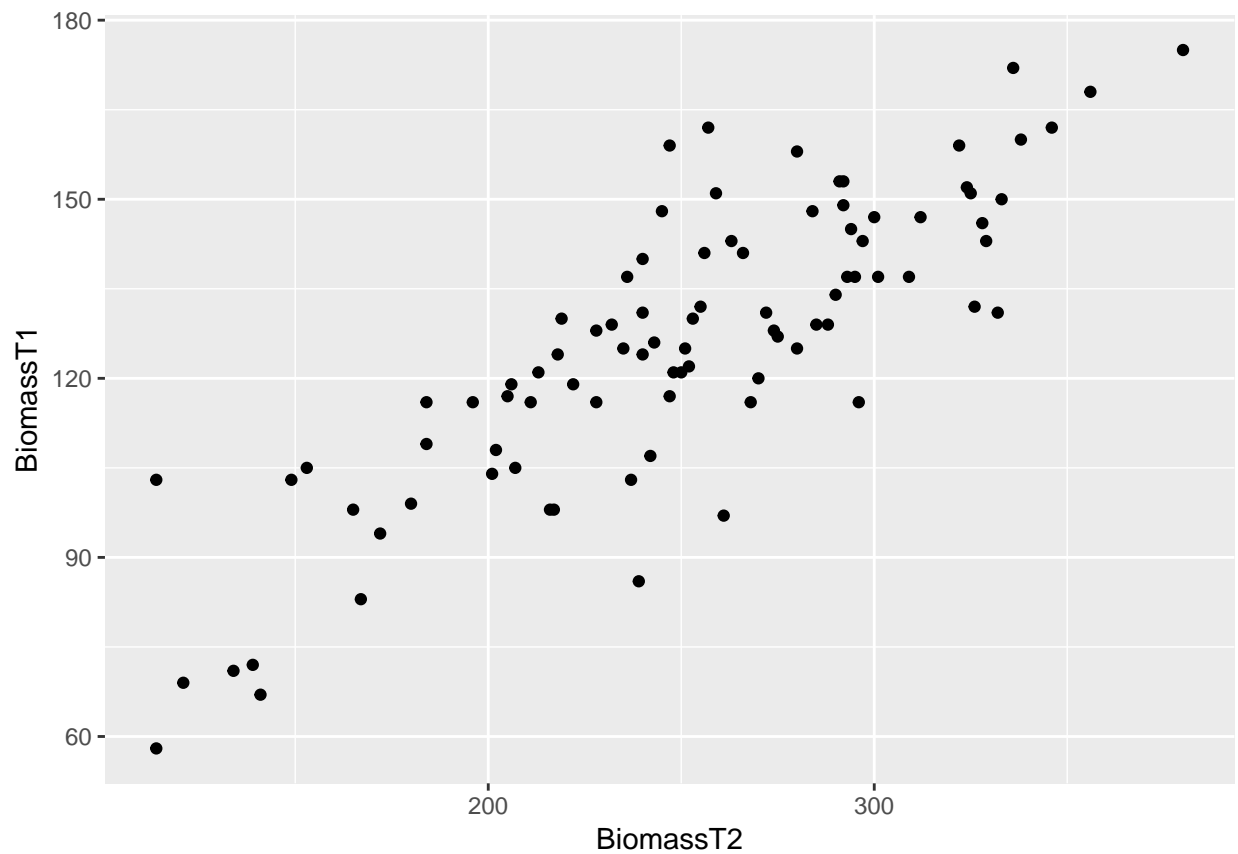
```
ggplot(data, mapping = aes(x = factor(Species), y = BiomassT2)) + geom_point(aes(color = Fert))
```



I chose biomass for the Y axis because it is the dependent variable. I added factor() to the x axis so that the Species variable would be seen as categorical instead of continuous and therefore clean up the axis labels. The pattern from Q2 is present, since there is a wide range of biomasses for Species 2 and 3, while there are fewer extreme biomasses for Species 1. The pattern from Q4 is also present, since the pink dots representing fertilizer A are at the higher biomasses for all of the species. You can also see how the pink dots for Fertilizer A are at the top of the plot for all of the species. Overall, the fertilizer has a larger impact on the final biomass than the species.

Q6

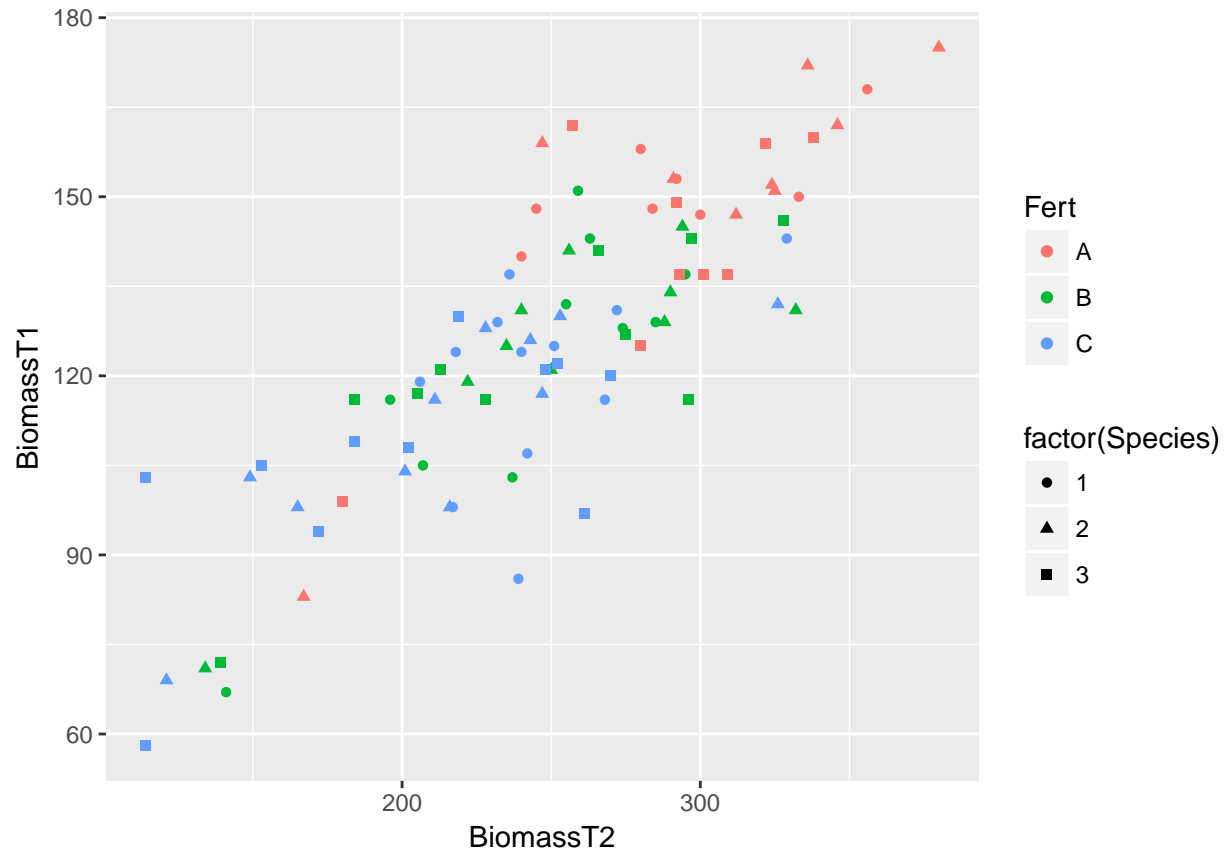
```
ggplot(data, mapping = aes(x = BiomassT2, y = BiomassT1)) + geom_point()
```



This plot shows a linear relationship between initial and final biomasses. This suggests that either measure could be used, and that the observed impact of fertilizer is still valid.

Q7

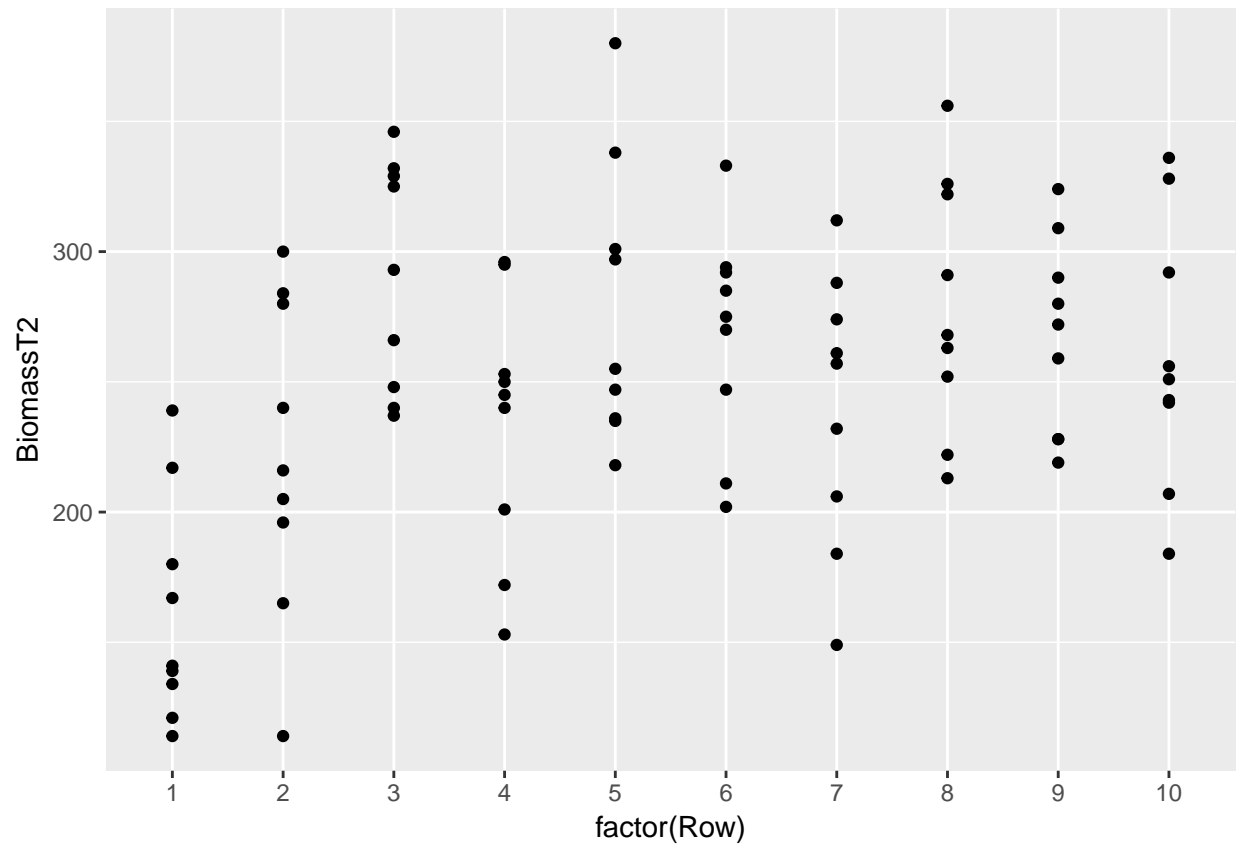
```
ggplot(data, mapping = aes(x = BiomassT2, y = BiomassT1)) + geom_point(aes(color = Fert, shape = factor
```



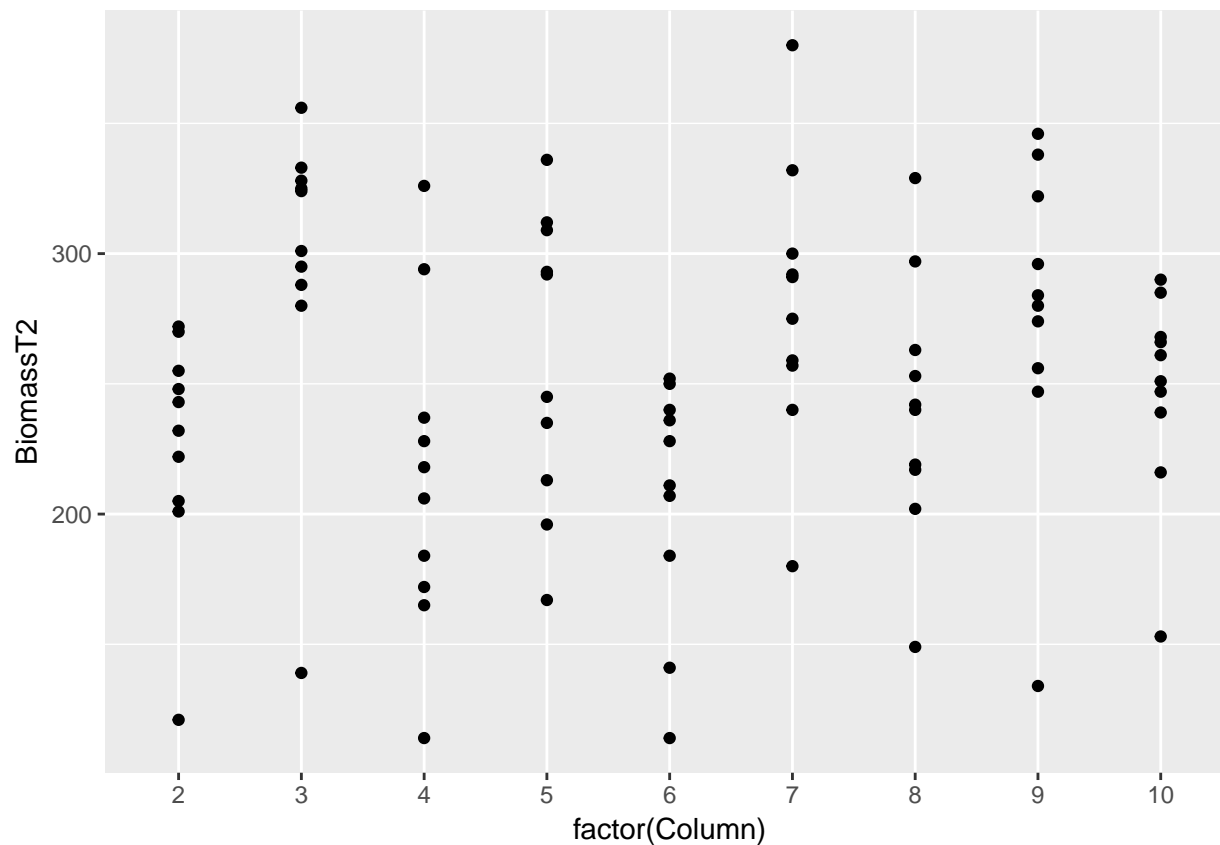
This graph is consistent with Q5. It is strange that there are a few Fertilizer A-treated plants in the lower left hand corner. There points were also visible in Q5; the 3 pink dots at the bottom of each species column. One possible explanation could be that these plants were small to begin with, so their final biomasses will also be smaller.

## Q8

```
ggplot(data, mapping = aes(x = factor(Row), y = BiomassT2)) + geom_point()
```



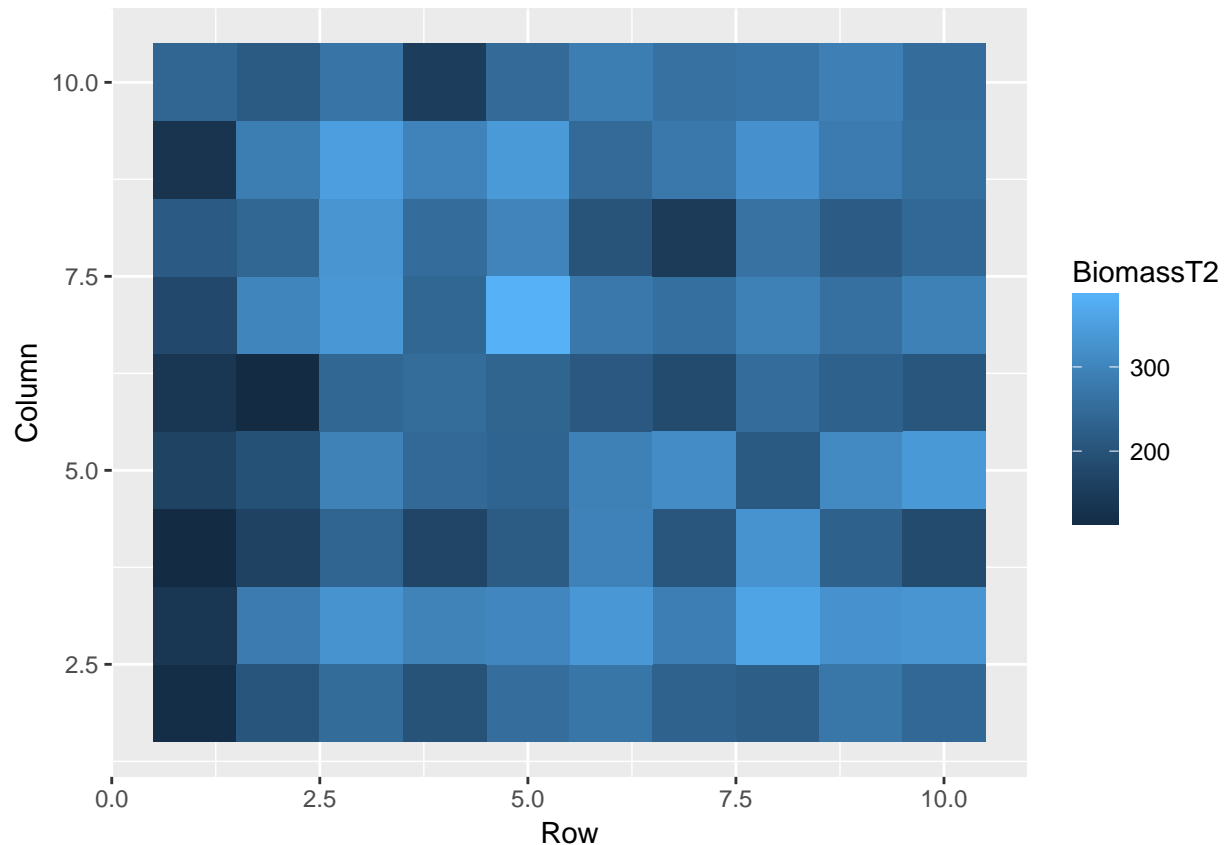
```
ggplot(data, mapping = aes(x = factor(Column), y = BiomassT2)) + geom_point()
```



I used `factor()` again so that the units for the x axis were reasonable and not representing row number 2.5. Row 1 had a lower final biomass, while row 5 had a higher biomass. Columns 2 and 6 had the lowest biomass. All of the columns had at least one low biomass. These points could correspond to the row 1 plant for each column.

Q9

```
ggplot(data, mapping = aes(x = Row, y = Column)) + geom_tile(aes(fill = BiomassT2))
```



In this graph, the darkest squares are mostly found on the left edge, in row 1. This suggests that another factor impacted the plants only in column 1. Other factors that impact plant growth that should have been controlled are light or water. One source of the systematic error could have been that the sprinklers for that column were faulty.

## Q10

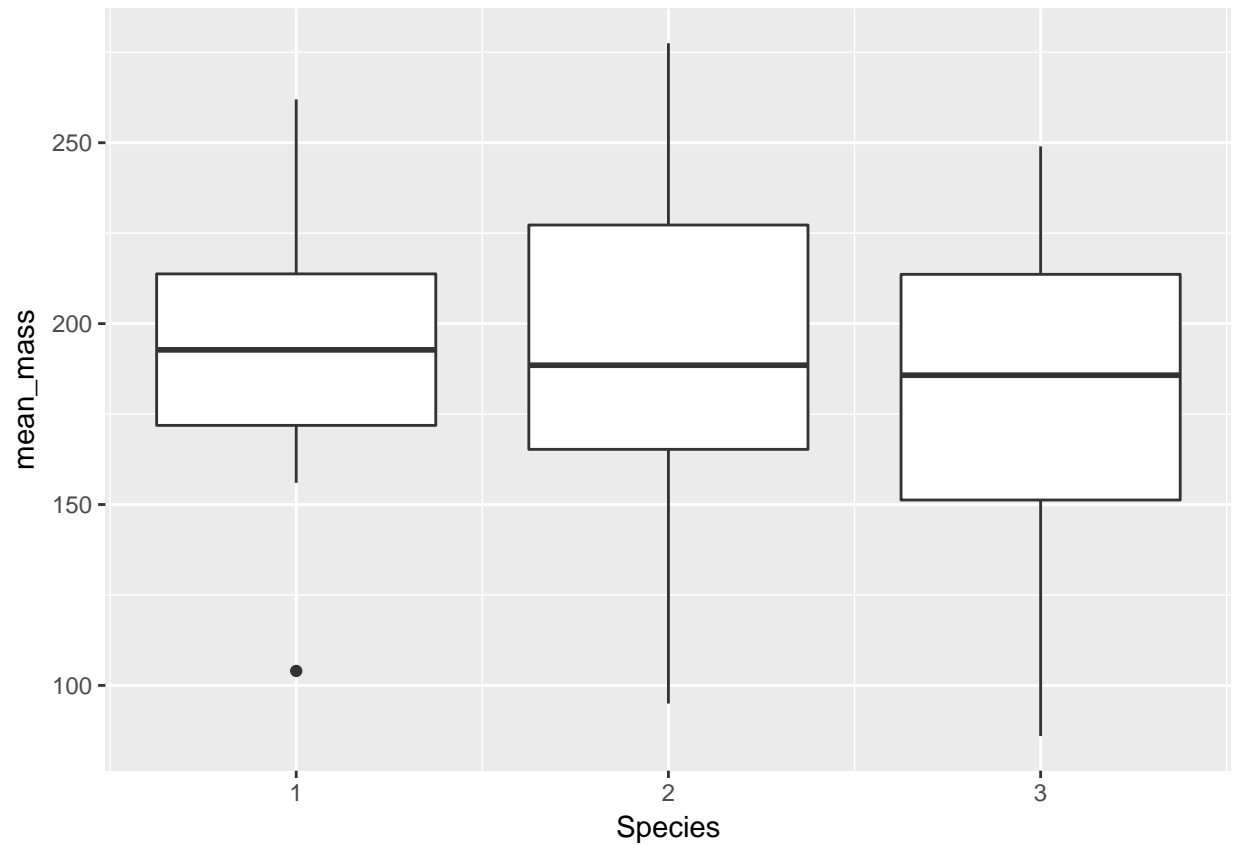
I made a subset of the data in line 13. I know that it updated because the numbers in the call to `table()` in Q1 have decreased. In Q2, the means of Species 1 and 2 are now more similar, and the mean of Species 3 is smaller. In Q4, there is now a larger difference between Fertilizer B and C. The distribution of the middle 50% is much smaller for Fertilizer C, and therefore this is definitely the worst one.

## Q11

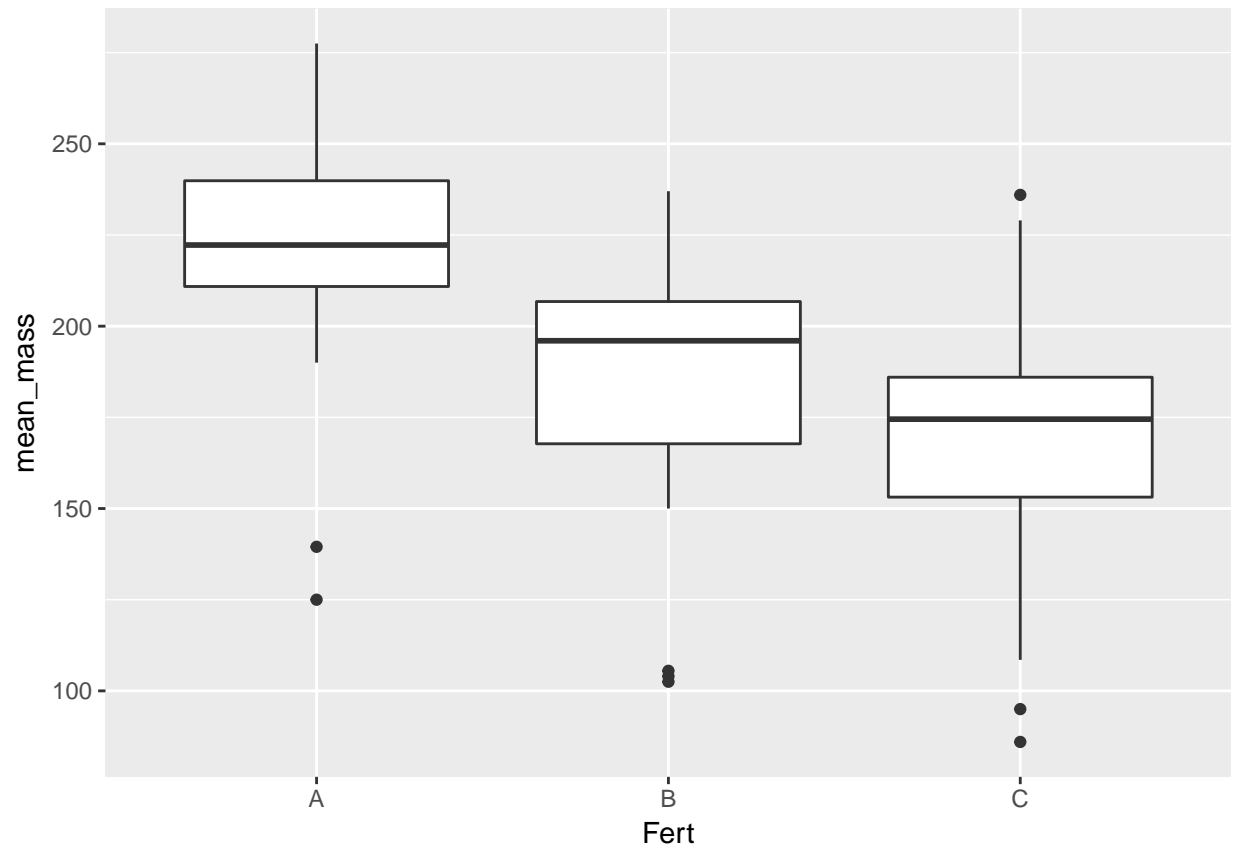
```
data$mean_mass <- rowMeans(data[c("BiomassT1", "BiomassT2")])

ggplot(data, aes(x = Species, y = mean_mass)) + geom_boxplot(aes(group = Species))
```

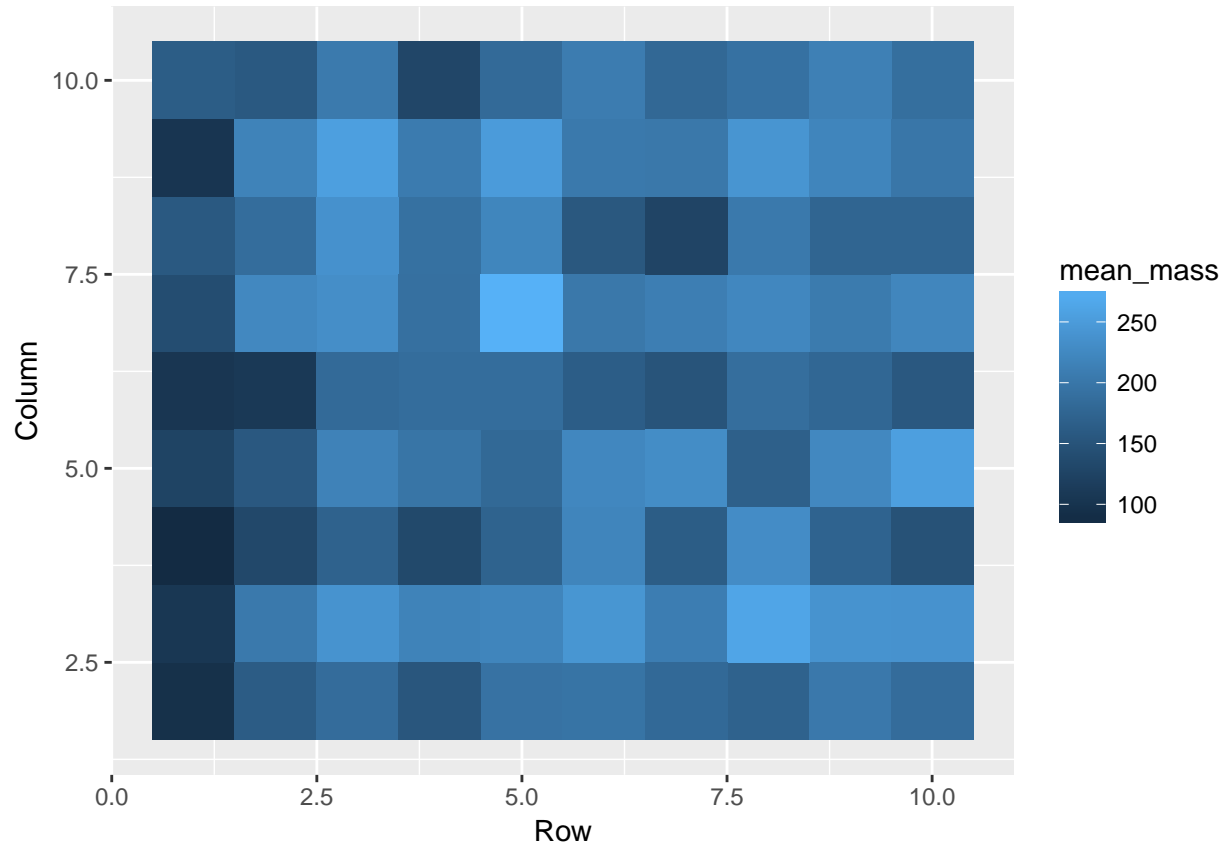




```
ggplot(data, aes(x = Fert, y = mean_mass)) + geom_boxplot(aes(group = Fert))
```



```
ggplot(data, mapping = aes(x = Row, y = Column)) + geom_tile(aes(fill = mean_mass))
```



The relationship between species remained the same. The differences between fertilizer became even more pronounced. The 3 middle 50% boxes overlap each other on the y axis even less.

## Q12

To grow the largest hydrangeas, the researchers should use fertilizer C on plants that already have a large biomass. The difference between Species 1 and 2 is not very large, so it does not matter which species is used. However, Species 3 should be avoided. The researchers should make sure all of the external factors are optimal, and fix the problem that impacted column 1 by changing lightbulbs and/or getting a new sprinkler.