

Plant Competition Dataset & Analysis

Katie Wang

4/2/22

Hypotheses

1. At higher nitrogen levels, average plant biomass will be higher because nitrogen is a nutrient and low nitrogen will limit the amount of plant growth possible.
2. At higher competition levels (with competition being defined as more plants growing in a single pot, regardless of species), average biomass of each plant will be lower because each plant must allocate resources toward root growth to access soil nutrients and water rather than toward aboveground growth (which is what we are measuring as biomass).
3. At lower nitrogen levels, the decrease in average plant biomass due to competition will be greater than it is at higher nitrogen levels because the scarcity of nitrogen will compound with competition to cause plants to allocate more resources toward root growth rather than aboveground growth.
4. As competition increases, mustard plants will have a greater decrease in biomass (proportional to the mustard's average biomass) than pea plants will (proportional to the pea's average biomass) because pea plants, as legumes, are able to fix nitrogen and therefore do not have to devote as much root growth toward accessing usable soil nitrogen as mustard does.

Data description

```
plant_data <- read.csv("Plant experiment data.csv") %>%  
  mutate(Avg.biomass.within.species = Mass.of.plant/X..sprouted)  
head(plant_data)
```

##	ID	Species	Nitrogen	X..planted	X..sprouted	Date	Mass.of.plant.weighboat
## 1	A1	Pea	Low	3	2	3/11	5.4655
## 2	A2	Pea	High	3	0	3/11	0.0000
## 3	A2	Mustard	High	3	3	3/11	5.3728
## 4	A3	Pea	High	3	1	3/9	5.4023
## 5	A4	Mustard	Low	3	0	3/9	0.0000
## 6	A5	Pea	Low	6	1	3/9	5.4250
##	Mass.of.weighboat		Mass.of.plant	Avg.biomass.within.species			
## 1	5.2783		0.1872			0.0936	
## 2	0.0000		0.0000			NaN	
## 3	5.3587		0.0141			0.0047	
## 4	5.3741		0.0282			0.0282	
## 5	0.0000		0.0000			NaN	
## 6	5.3127		0.1123			0.1123	

In each pot, we planted pea and mustard seeds in different number combinations (to manipulate competition level, creating both conspecific and heterospecific competition) and in soil of high or low nitrogen levels. Biomass is defined as the dry mass of above-soil portions of plants, and a single measure is taken per species per pot rather than by individual plant. In creating our models, we will assume that the strength of competition from both species is equal and that conspecific and heterospecific competition are also equal in strength. The amount of competition is defined as the number of plants that sprouted in a single pot,

including plants of all species present in that pot. In some pots, none of the seeds of a given species sprouted, but the data is still included to distinguish these instances from instances in which only a single species was planted in a pot.

The column Avg.biomass.within.species will be used for testing hypothesis 4, as the response variable is the average growth of a particular species of plant in a pot rather than the average across the whole pot.

Each pot has a unique ID, so separate rows with the same ID indicate plants of different species from the same pot. The pots were harvested on different dates, so the date of harvest is also included.

In order to test hypotheses 1, 2, and 3, we will need to create a table of the total competition (i.e. number of sprouted plants) and average biomass of a plant in each pot (PerPlant.Biomass), making the assumption that there is about the same number of pea and mustard plants in each competition condition. We are finding the average biomass per pot rather than using absolute biomass in order to normalize by the number of sprouts in each pot. We also need to include nitrogen levels.

```
plant_per_pot <- plant_data %>%
  group_by(ID) %>%
  summarise(Nitrogen = Nitrogen, Total.Sprouted = sum(X..sprouted),
            Total.Biomass = sum(Mass.of.plant),
            PerPlant.Biomass = sum(Mass.of.plant)/sum(X..sprouted)) %>%
  distinct() #summarise() had created identical rows for the pots w/ multiple species

## `summarise()` has grouped output by 'ID'. You can override using the `.groups` argument.
head(plant_per_pot)
```

```
## # A tibble: 6 x 5
## # Groups:   ID [6]
##   ID    Nitrogen Total.Sprouted Total.Biomass PerPlant.Biomass
##   <chr> <chr>          <int>         <dbl>         <dbl>
## 1 A1    Low           2          0.187          0.0936
## 2 A2    High           3          0.0141         0.0047
## 3 A3    High           1          0.0282         0.0282
## 4 A4    Low            0           0           NaN
## 5 A5    Low           3          0.139          0.0465
## 6 A6    High           5          0.0144         0.00288
```

For hypothesis 4, I need to plot biomasses for species separately but as a function of total number of sprouts per pot, so I joined the Total.Sprouted values to the plant_data dataframe (but kept it under a different name for the sake of clarity).

```
plant_data_joined <- left_join(plant_data,
                               plant_per_pot %>% select(ID, Total.Sprouted),
                               by = "ID")
head(plant_data_joined)
```

```
##   ID Species Nitrogen X..planted X..sprouted Date Mass.of.plant.weighboat
## 1 A1    Pea     Low         3           2 3/11          5.4655
## 2 A2    Pea     High        3           0 3/11          0.0000
## 3 A2 Mustard    High        3           3 3/11          5.3728
## 4 A3    Pea     High        3           1 3/9           5.4023
## 5 A4 Mustard    Low         3           0 3/9           0.0000
## 6 A5    Pea     Low         6           1 3/9           5.4250
##   Mass.of.weighboat Mass.of.plant Avg.biomass.within.species Total.Sprouted
## 1          5.2783      0.1872          0.0936              2
## 2          0.0000      0.0000           NaN              3
## 3          5.3587      0.0141          0.0047              3
## 4          5.3741      0.0282          0.0282              1
```

## 5	0.0000	0.0000	NaN	0
## 6	5.3127	0.1123	0.1123	3

Data analysis

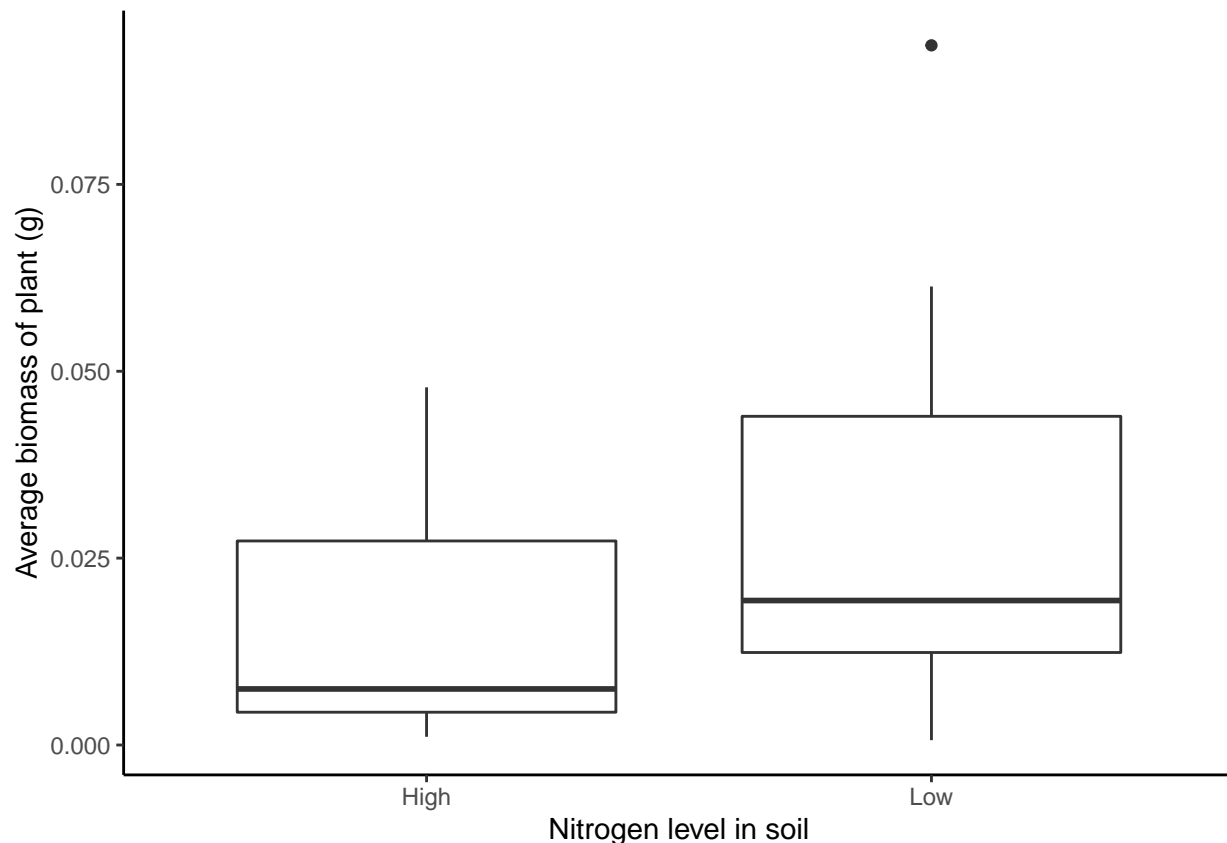
Hypothesis 1

```
summary(lm(PerPlant.Biomass ~ Nitrogen, data = plant_per_pot))
```

```
##
## Call:
## lm(formula = PerPlant.Biomass ~ Nitrogen, data = plant_per_pot)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.028789 -0.013037 -0.009403  0.014000  0.064161
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.015822   0.004248   3.724 0.000634 ***
## NitrogenLow  0.013617   0.006333   2.150 0.037973 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01993 on 38 degrees of freedom
## (8 observations deleted due to missingness)
## Multiple R-squared:  0.1085, Adjusted R-squared:  0.085
## F-statistic: 4.623 on 1 and 38 DF, p-value: 0.03797

ggplot(plant_per_pot, aes(x = Nitrogen, y = PerPlant.Biomass)) +
  geom_boxplot() +
  xlab("Nitrogen level in soil") +
  ylab("Average biomass of plant (g)")
```

```
## Warning: Removed 8 rows containing non-finite values (stat_boxplot).
```



Our data reject the hypothesis. Though linear regression shows that soil nitrogen levels do have a significant effect on biomass, as the slope of the fitted line is significantly different from 0 ($p=0.03797$), the box plot shows that the effect is in the direction opposite of what we had expected. That is, biomass achieves higher values in low nitrogen conditions.

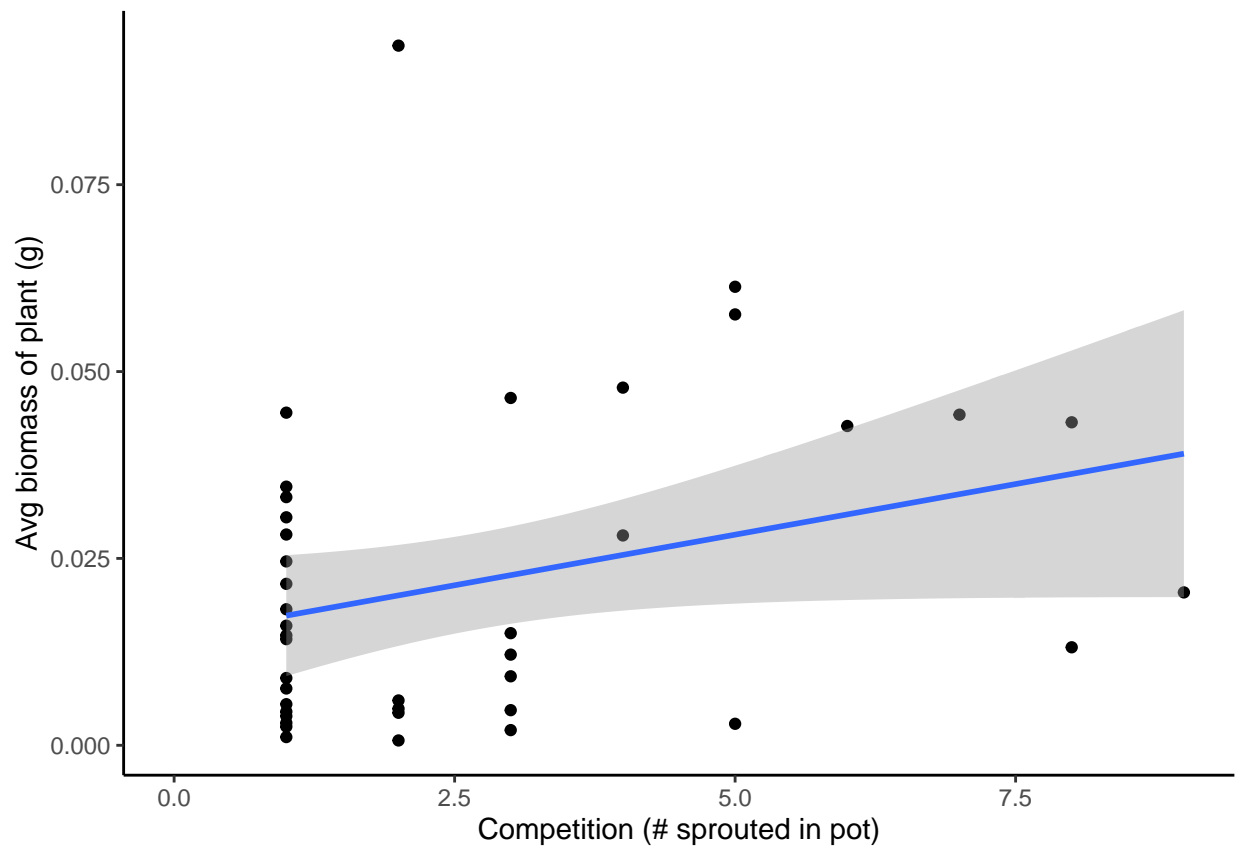
Hypothesis 2

```
summary(lm(PerPlant.Biomass ~ Total.Sprouted, data = plant_per_pot))

##
## Call:
## lm(formula = PerPlant.Biomass ~ Total.Sprouted, data = plant_per_pot)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.025299 -0.014470 -0.005454  0.011098  0.073546
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.014637   0.004979   2.940  0.00556 **
## Total.Sprouted 0.002708   0.001417   1.911  0.06350 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02016 on 38 degrees of freedom
## (8 observations deleted due to missingness)
## Multiple R-squared:  0.08772,    Adjusted R-squared:  0.06371
```

```
## F-statistic: 3.654 on 1 and 38 DF, p-value: 0.0635
ggplot(plant_per_pot, aes(x = Total.Sprouted, y = PerPlant.Biomass)) +
  geom_point() +
  stat_smooth(method = "lm") +
  xlab("Competition (# sprouted in pot)") +
  ylab("Avg biomass of plant (g)")

## `geom_smooth()` using formula 'y ~ x'
## Warning: Removed 8 rows containing non-finite values (stat_smooth).
## Warning: Removed 8 rows containing missing values (geom_point).
```



Our hypothesis was rejected, as linear regression did not find a significant effect of competition on the average biomass of plants. The slope of the fitted line is not significantly different from 0 ($p=0.06371$). Note that this conclusion applies to biomass averaged across entire pots, rather than biomass averaged within plant species in each pot, as is the case in hypothesis 4.

Hypothesis 3

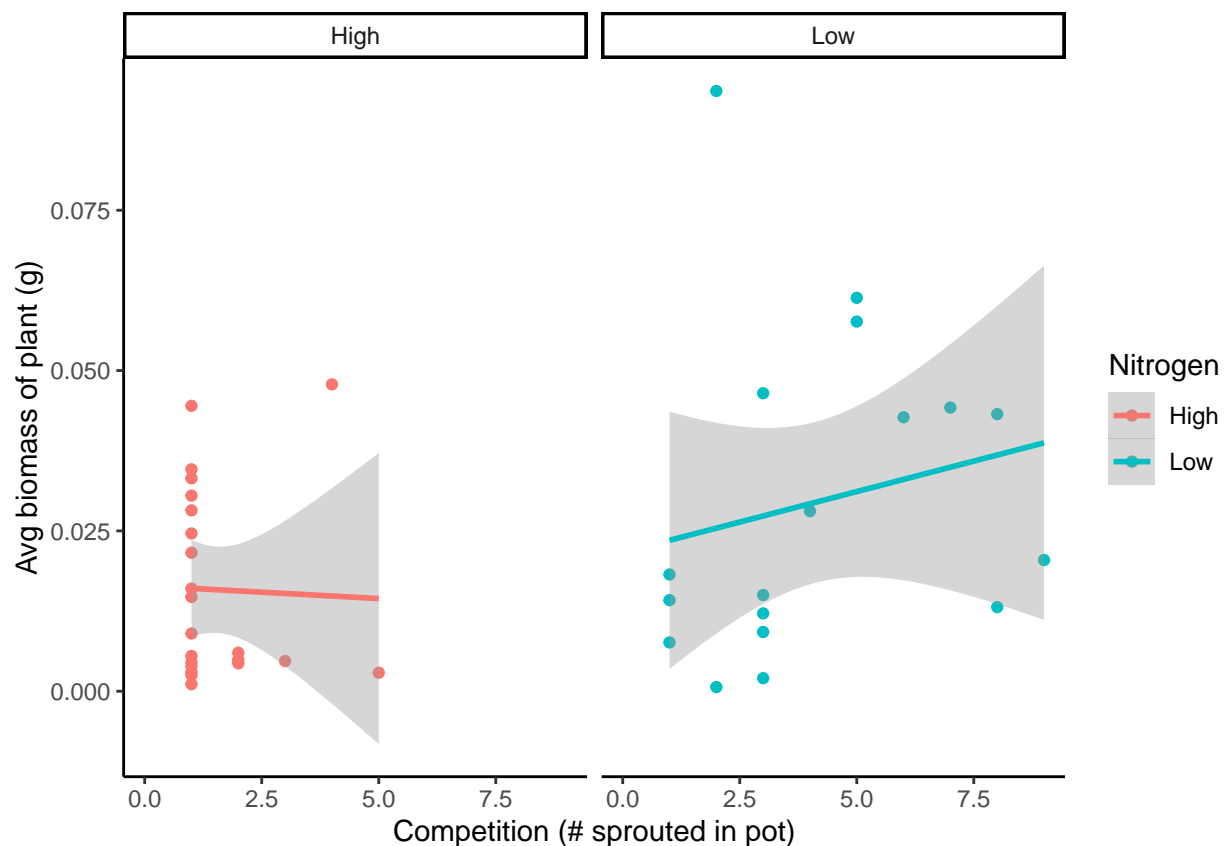
```
# two-way ANOVA interactive model
model1_hyp3 = lm(PerPlant.Biomass ~ Nitrogen + Total.Sprouted, data = plant_per_pot)
model2_hyp3 = lm(PerPlant.Biomass ~ Nitrogen * Total.Sprouted, data = plant_per_pot)
anova(model1_hyp3, model2_hyp3)

## Analysis of Variance Table
##
## Model 1: PerPlant.Biomass ~ Nitrogen + Total.Sprouted
```

```
## Model 2: PerPlant.Biomass ~ Nitrogen * Total.Sprouted
##   Res.Df      RSS Df Sum of Sq    F Pr(>F)
## 1      37 0.014791
## 2      36 0.014681   1 0.00010957 0.2687 0.6074

ggplot(plant_per_pot, aes(x = Total.Sprouted, y = PerPlant.Biomass, color = Nitrogen)) +
  geom_point() +
  geom_smooth(method='lm') +
  xlab("Competition (# sprouted in pot)") +
  ylab("Avg biomass of plant (g)") +
  facet_wrap(~Nitrogen)

## `geom_smooth()` using formula 'y ~ x'
## Warning: Removed 8 rows containing non-finite values (stat_smooth).
## Warning: Removed 8 rows containing missing values (geom_point).
```



According to ANOVA analysis, there is no significant difference between the additive and interactive models for the effect of nitrogen and competition on biomass ($p=0.6074$). This means that there is no interaction between nitrogen and competition in determining biomass, and so we reject our hypothesis. In addition, the plot suggests that average biomass does not decrease as competition increases; this is consistent with our rejection of hypothesis #2.

Hypothesis 4

```
model1_hyp4 = lm(Avg.biomass.within.species ~ Total.Sprouted + Species, data = plant_data_joined)
model2_hyp4 = lm(Avg.biomass.within.species ~ Total.Sprouted * Species, data = plant_data_joined)
anova(model1_hyp4, model2_hyp4)
```

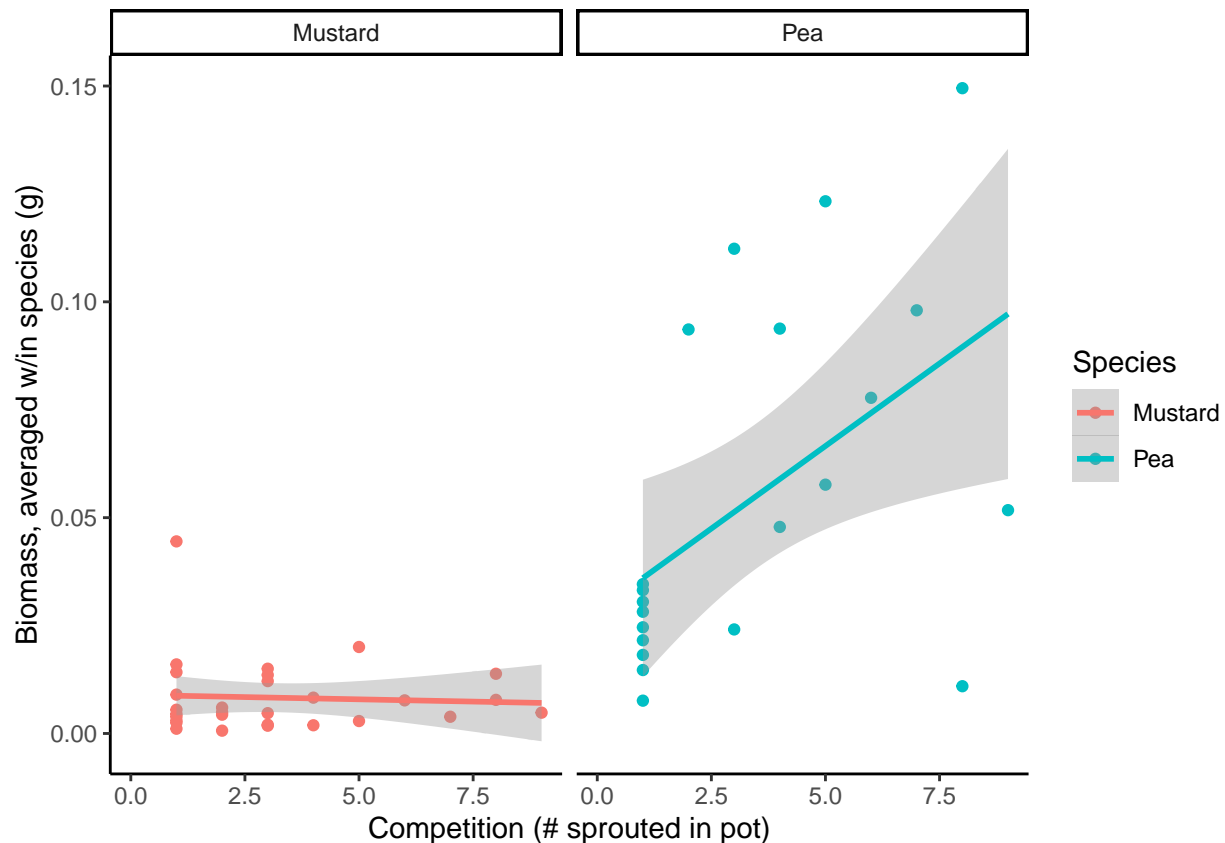
```
## Analysis of Variance Table
##
## Model 1: Avg.biomass.within.species ~ Total.Sprouted + Species
## Model 2: Avg.biomass.within.species ~ Total.Sprouted * Species
##   Res.Df    RSS Df Sum of Sq    F  Pr(>F)
## 1      47 0.032706
## 2      46 0.027813  1 0.0048937 8.0938 0.006609 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ggplot(plant_data_joined, aes(x = Total.Sprouted, y = Avg.biomass.within.species, color = Species)) +
  geom_point() +
  geom_smooth(method='lm') +
  xlab("Competition (# sprouted in pot)") +
  ylab("Biomass, averaged w/in species (g)") +
  facet_wrap(~Species)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 22 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 22 rows containing missing values (geom_point).
```



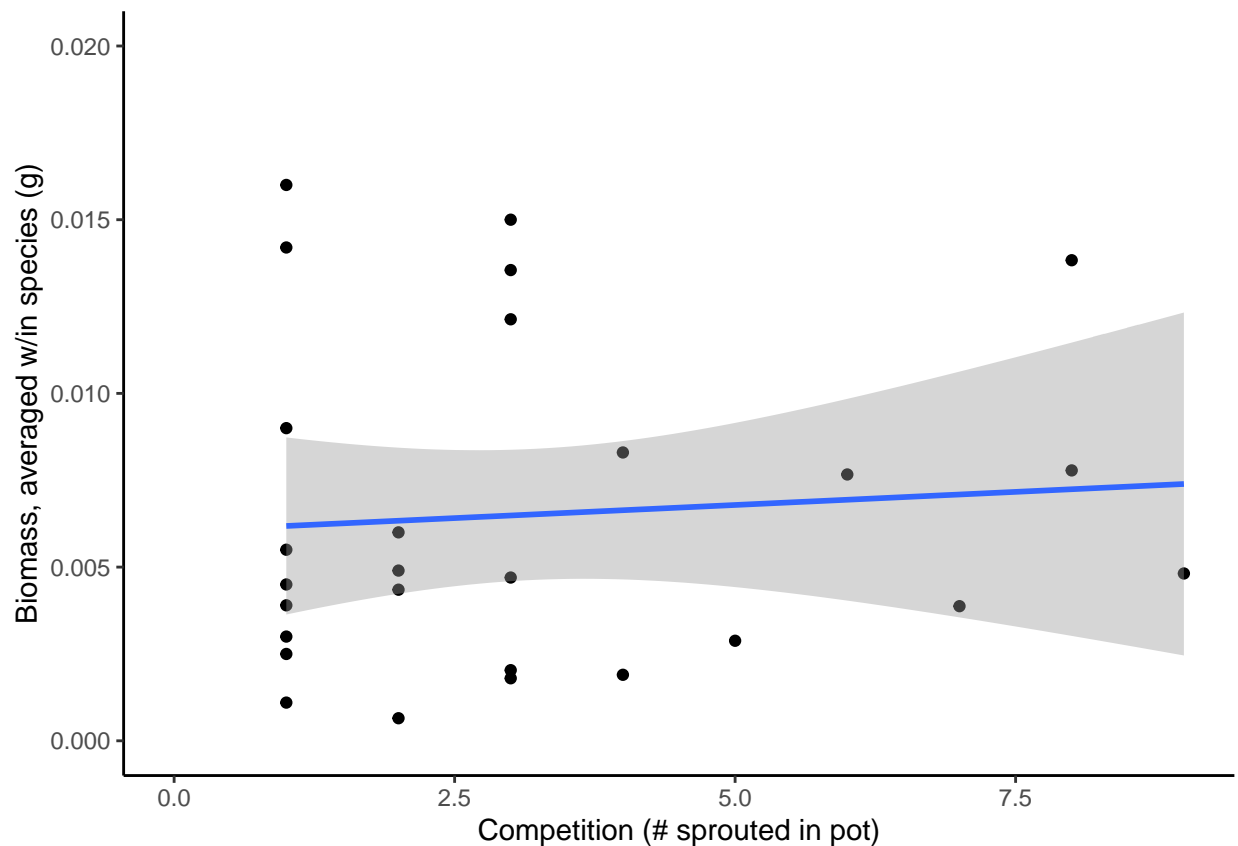
```
ggplot(plant_data_joined %>% filter(Species == "Mustard"),
  aes(x = Total.Sprouted, y = Avg.biomass.within.species)) +
  geom_point() +
  geom_smooth(method='lm') +
```

```
xlab("Competition (# sprouted in pot)") +
ylab("Biomass, averaged w/in species (g)") +
ylim(0, 0.02)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 9 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 9 rows containing missing values (geom_point).
```



In addition to plotting data from both species, I have also plotted the mustard data alone to better show the slope of the regression line.

By comparing the slopes of the two linear regressions in `anova()`, we are doing an assessment of loss (or gain) of biomass growth due to competition that is independent of how large the plant tends to grow. (Or at least that is the intention.)

ANOVA analysis disproves our hypothesis. Although the effect of competition on biomass is very different between the two species (ANOVA comparison of slopes of linear regressions fitted to data grouped by species, $p=0.006609$), we observe that while average mustard plant biomass does not seem to depend on competition, average pea plant biomass seems to increase with competition. There is no decrease in biomass due to competition like we had hypothesized.