Lab 5: Plant Differentiation and Growth

Pre-Setup

Before we begin to use R today, calculate the average height of your light and dark seedlings at each interval (2 days, 4 days, and 6 days) using Google Sheets.

By calculating the average of each data point you are getting a pretty good estimate of your sample. This minimizes the error in your estimate of the height of your seedlings (i.e. measurement error). However, it means you only have one height estimate per treatment/day.

Once you have an estimate for your sample, enter this into the class data sheet (your TA will provide this). This data sheet is what you'll use during this lab.

Now you have good estimates of multiple samples to work with. There will be some variation among samples due to the soil, the location the plants were grown, and how different people measured height. Having multiple samples allows you to estimate a mean height despite all this variation.

1. Load Data Into R

We will use a similar procedure as the last lab to load data into R so that we can work with it.

- Make a new project in a new folder and open a new script.
- Load libraries you will need: `ggplot2`, `gsheet`, and `dplyr`

For this lab we will provide you with the directions of what to do by showing you the comments you should enter in you R script.

- Copy the comment lines into your script.
- Enter the necessary code based on the previous lab. In any case where you have not seen the necessary code before, it is provided.
- Answer questions in this assignment using additional comments.

```
# Load the libraries ggplot2, gsheet, and dplyr.
```

Next, load in the data. Today's data is found at:

https://docs.google.com/spreadsheets/d/1MOkh3SNsuuTixjRFi4ptYmjimUJ9rGXIta1iJDjre9k

```
# Assign website address for the data to a variable
```

```
# Load the data from google sheets
```

- Note: the follwoing instructions show assigned the data to a variable named plant_data. You can use this name and follow these intructions or change to a new variable name.

Remember, once we have our data in, it is always good to check to make sure the data was imported correctly. Use `head` to check your imported data.

```
# Check your data looks right (first lines only)
head(plant_data)
```

- Note: here the data are called `plant_data` - your may have a different name

2. Graphing the Seedling Data

Just like last time, we are going to use `ggplot` to graph the data. Because we want to compare the seedlings grown under light and dark conditions, we are going to graph both sets of data together. We are first going to graph the data as points. This way you can see how much variation is in your data.

First create the base layer of your plot

- What is the independent variable (x axis)?
- What is the dependent variable (y axis)?
- How will you plot light and dark conditions separately?
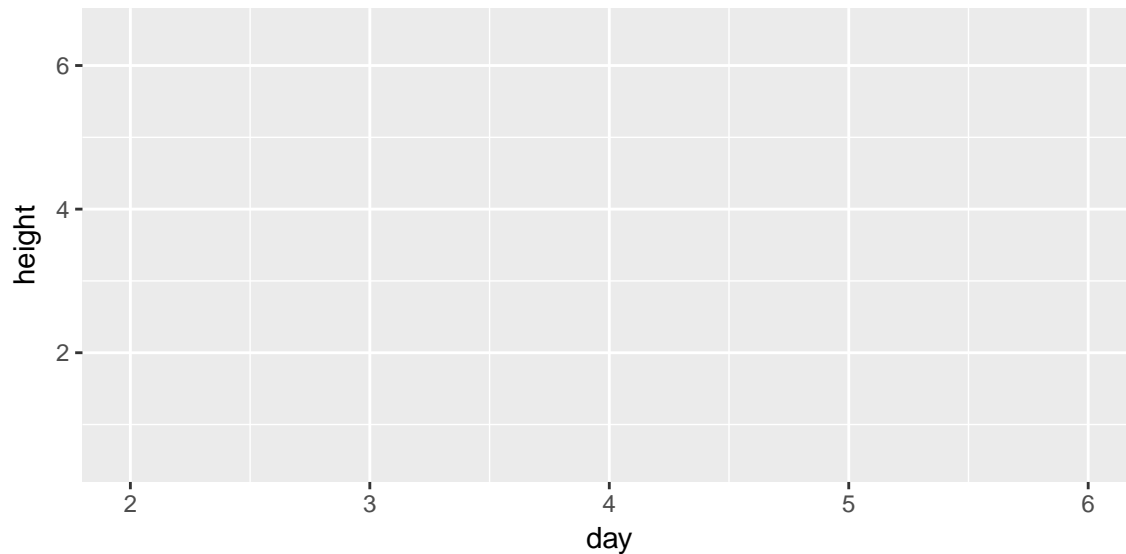
Second, add the data to the plot (using `geom_point`)
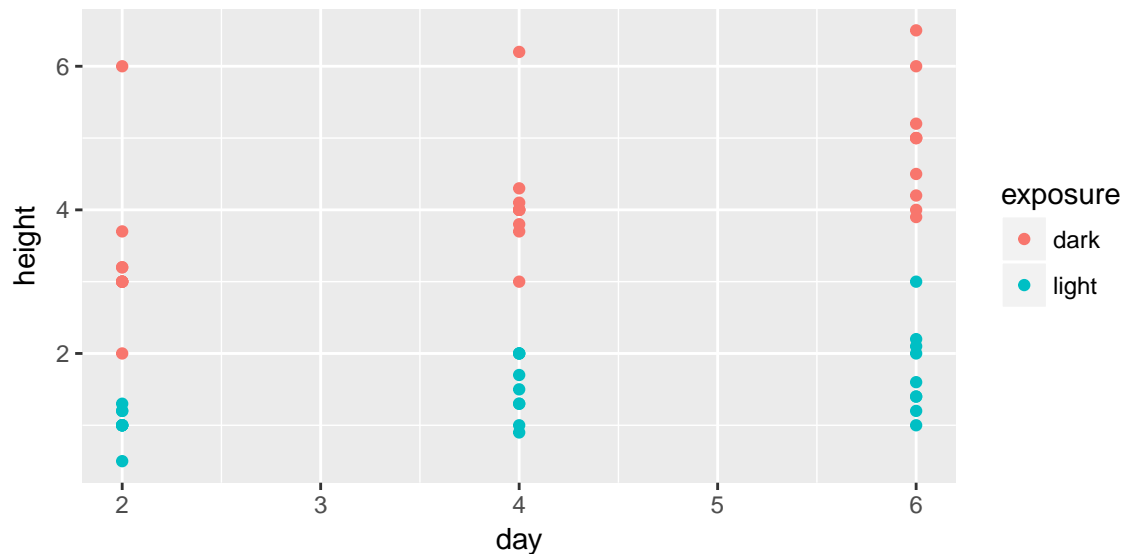
Does it look like there might be a difference between the two groups?

What can you say at this point about the difference? How confident do you feel?

Use your knowledge from the previous lab to visualize the variation in the data without plotting every point.

2. Compare the Height of Seedlings in Different Conditions

We are interested in comparing the growth rate of seedlings in light and dark conditions. To make these comparisons we need to know the mean of each group of data and how much variation there is in that group.

- Use the function `group_by` in the package `dplyr` to group the data by exposure and day.

```
# Group data by exposure and day
```

- Calculate the mean of each group

```
# Calculate the mean height of the seedlings at each interval
plant_data_means <- summarise(grouped_plant_data,
                              mean = mean(height, na.rm=TRUE))
```

This code is just like you used last time, with the addition of `na.rm=TRUE`. This additional code tells R to ignore areas in the data set that had no values entered. For example, if no seedlings grew under your dark condition, you would not record any data under height for that experiment. By using this code, R can now ignore these blank spaces and calculate the mean correctly.

Check the first few lines of your data using `head`

4. Graphing the seedling Data

Just like last time, we are going to use `ggplot` to graph the data. Because we want to compare the seedlings grown under light and dark conditions, we are going to graph both sets of data together.

First create the base layer of your plot

- What is the independent variable (x axis)?
- What is the dependent variable (y axis)?
- How will you plot light and dark conditions separately?
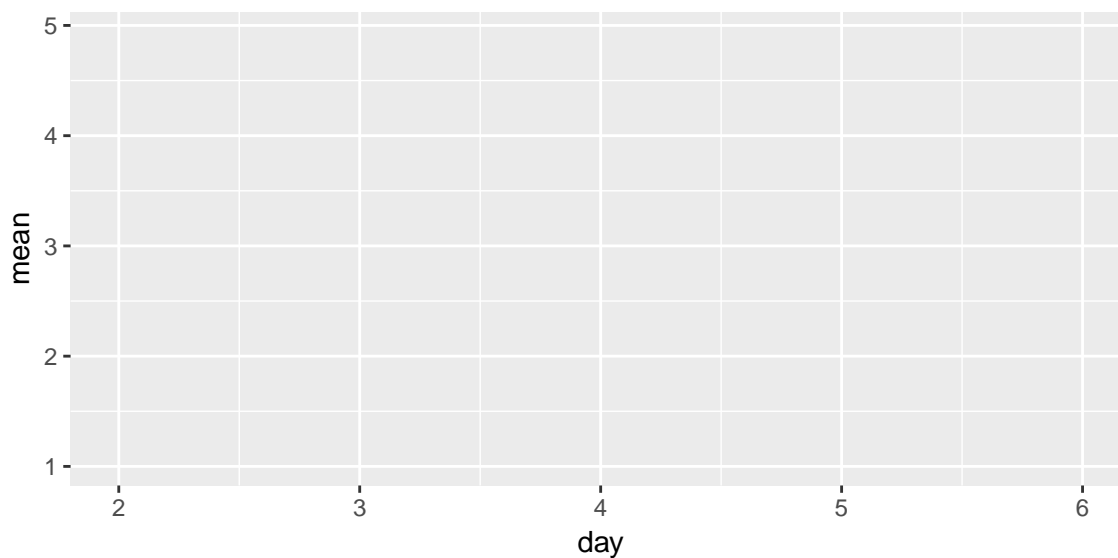
Second, add the data to the plot (using `geom_point`)
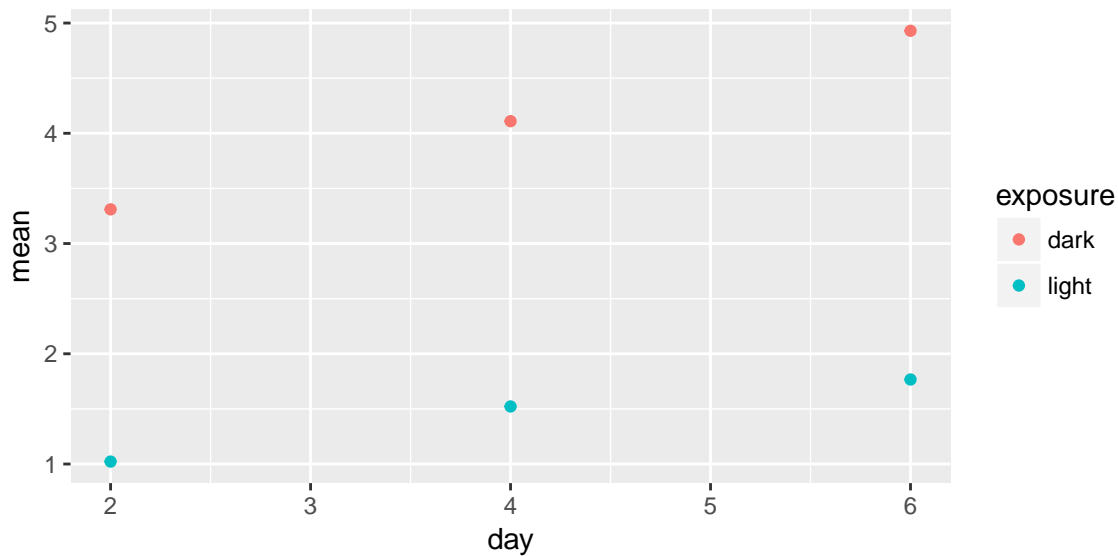
**Third add standard deviation bars to your plot**

- It appears that there are differences in seedling heights between the two different growing conditions. However, if there is a lot of variation in our data those differences may not be significant (remember the t-test from last lab).

- The *standard deviation* is how much the individual data points differ from the overall mean. This puts a single value on how much variation there is in the data. For example, if we saw seedling heights of 3, 2.9, 3, and 3.1, we would have a small standard deviation. However, if we had seedling heights of 0.5, 5.5, 2, and 4, the standard deviation would be high.

- Standard deviation is calculated in R using the function `sd`.

```
# Remake your table of means so it includes std deviation
plant_data_means <- summarise(grouped_plant_data,
                              mean = mean(height, na.rm=TRUE),
                              stdev = sd(height,na.rm=TRUE))
```

Next add the standard deviation bars to your graphs by adding a layer using `geom_errorbar`

- `geom_errorbar` draws an errorbar that has an upper and lower value. In this case, the upper value is the mean + the standard deviation and the lower value is the mean - the standard deviation.

```
# Add the following layer to your plot
  geom_errorbar(aes(ymin=mean+stdev, ymax=mean-stdev))
```
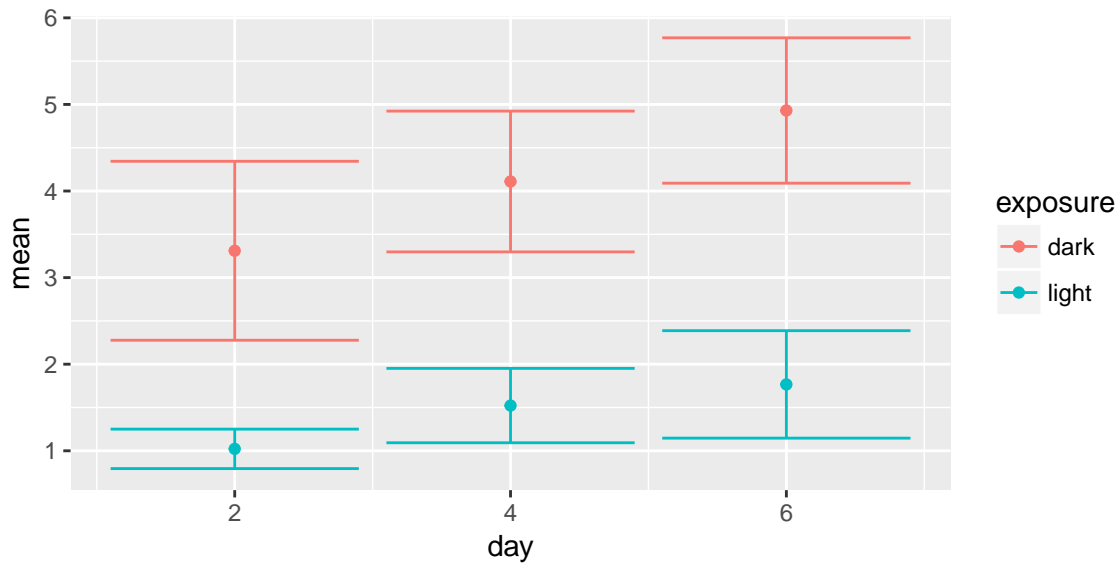
Fourth, add labels to your graph

Just like last class, you'll want to clean up your graph and make it look professional.

- To label the axes make a new layer using `labs(x="", y="", color="")`. Enter labels for the x axis, y axis, and legend in the quotes.

- Last, its good to clean up your graph. As you can see, those error bars are a little messy. Try changing their width to make them more professional looking by adding `width = 0.2` after `ymax` in your error bar layer.

```
#Label your plot and change axes titles
```
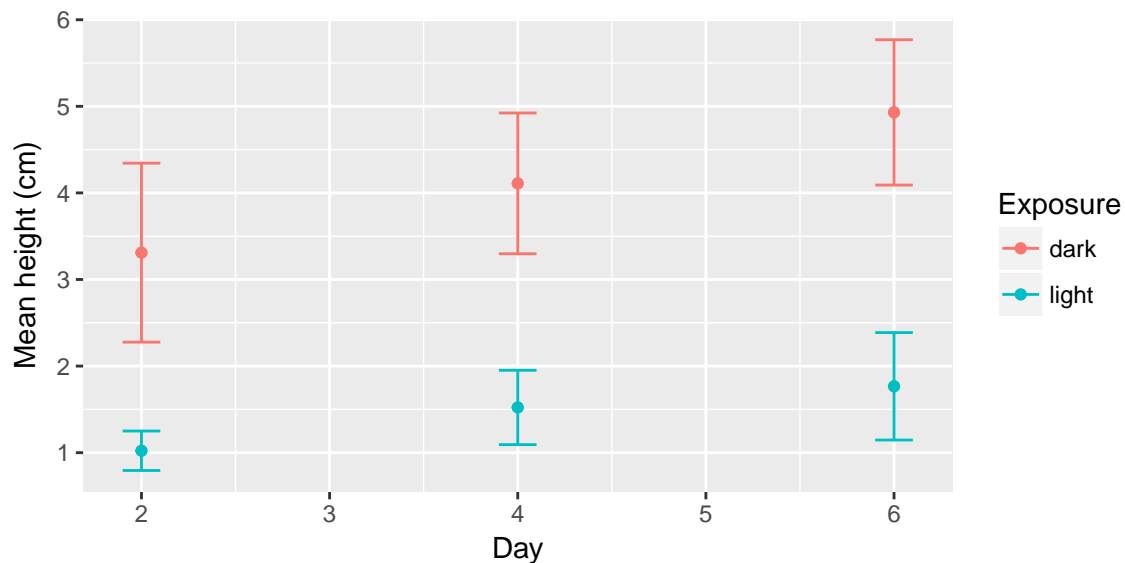

Answer the following questions

- Under which condition did the seedlings grow better? Light or dark?
- How confident do you feel? More or less than when you plotted all the data?

4. Compare height under different conditions using statistics (a t-test)

Just like in the Mechanisms of Evolution lab, we now have data for which we want to see if there is a statistical difference between two groups. In this case, we're interested in whether there is a difference between seedling heights in the 2 day old, 4 day old, and 6 day old seedlings. Let's walk through how to do a t-test for the 2 day old seedlings.

- Filter light and dark grown seedlings for the day 2 time period.

- What is your null hypothesis for the 2 day old seedling data?

Compare the data at 4 and 6 days on your own.

- Under which condition did the seedlings grow better? Light or dark?

More material to help you understand mean and standard deviation

Check out this interactive website for a better understanding of mean and standard deviation!

http://www.zoology.ubc.ca/~whitlock/Kingfisher/SamplingNormal.htm

Click on the tutorial button to work though the example.

- When you sample multiple individuals do you see variation?
- When you calculate the mean of your samples what are you estimating?
- Why do you need to calculate the mean of many samples (each of which is the mean of multiple measurements)?

More material to help you work better in R

Try swirl, an interactive tutorial to learn R in R and learn some statistics and data analysis.

- Make a new project in a new folder called swirl
- In the console window (lower left) of R Studio type the following then follow the instructions.

```
library(swirl)
swirl_options(swirl_data_dir='~')
swirl()
```

- You can continue your tutorial where you left off at any time. Your progress is automatically saved.
- When you open a different or new project after using swirl R will ask if you want to save workspace image. Answer yes to record the commands you have typed in swirl so your TA can see your progress.
- Make sure to use the same name to continue your work.