

Homework5-DataVis-ggplot

Necessary items:

- **libraries:** tidyverse, ggsci, ggforce, patchwork, Hmisc
- **files:** bloom_df.csv, hw_task2_blank.png

Within RStudio you can directly edit these blocks of code and turn in your homework that way with your code and figures printed.

I **HIGHLY** recommend getting your code to work in a normal R script then just copy and pasting the final code over to this document

First: import libraries, set your working directory, and read in bloom_df

```
library(tidyverse)
library(ggforce)
library(ggsci)
library(patchwork)
library(Hmisc)
setwd('~\\Desktop\\PB-HomeworkClub') #change to match your ggplot directory
bloom_df <- read.csv('bloom_df.csv')
```

bloom_df contents

- **taxa:** genus_species
- **genus**
- **species**
- **logbodysize**
- **trophic_position**
- **reg:** diadromous or non-diadromous

Let's take a peak and look at the structure

```
bloom_df[1:5,]

##           taxa genus      species logbodysize trophic_position
## 1  Alosa_alabamae Alosa    alabamae    1.707570      0.4313638
## 2  Alosa_alosa    Alosa      alosa    1.778151      0.5563025
## 3  Alosa_fallax    Alosa    fallax    1.778151      0.5563025
## 4  Alosa_mediocris Alosa   mediocris    1.778151      0.6127839
## 5 Alosa_pseudoharengus Alosa pseudoharengus    1.602060      0.5440680
##           reg
## 1 diadromous
## 2 diadromous
## 3 diadromous
## 4 diadromous
## 5 diadromous

str(bloom_df)
```

```
## 'data.frame':    50 obs. of  6 variables:
## $ taxa          : chr  "Alosa_alabamae" "Alosa_alosa" "Alosa_fallax"
## "Alosa_mediocris" ...
## $ genus         : chr  "Alosa" "Alosa" "Alosa" "Alosa" ...
## $ species       : chr  "alabamae" "alosa" "fallax" "mediocris" ...
## $ logbodysize   : num  1.71 1.78 1.78 1.78 1.6 ...
## $ trophic_position: num  0.431 0.556 0.556 0.613 0.544 ...
## $ reg           : chr  "diadromous" "diadromous" "diadromous"
## "diadromous" ...
```

Graphical tasks

This next tasks are going to ask you to make a series of figures. For each one, I would like you to make all of them *look good*. Change the theme around, change the axis labels, etc. Additionally, each of these will use the **reg** column which has 2 factors. Change the color and/or shape of them both.

Task 1: Continuous vs. continuous plot

Create two scatterplots of logbodysize vs. trophic_position grouped by reg

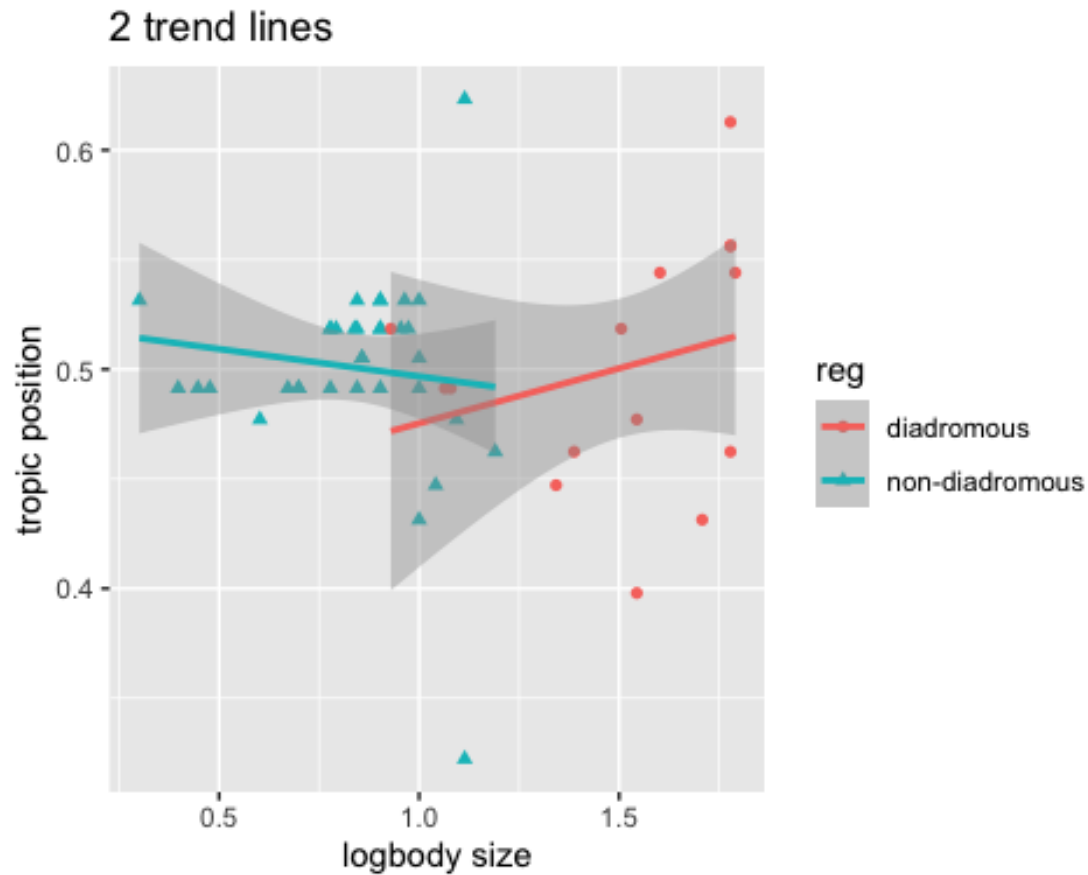
First plot:

- a 2 trend lines (method = 'lm'), one for each reg variable

#plot 1.1

```
ggplot(data = bloom_df, aes(x=logbodysize, y=trophic_position, colour=reg))+
  geom_point(aes(colour=reg, shape=reg))+
  stat_smooth(method = 'lm')+
  ggtitle('2 trend lines')+
  scale_x_continuous(name = 'logbody size')+
  scale_y_continuous(name = 'tropic position')

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



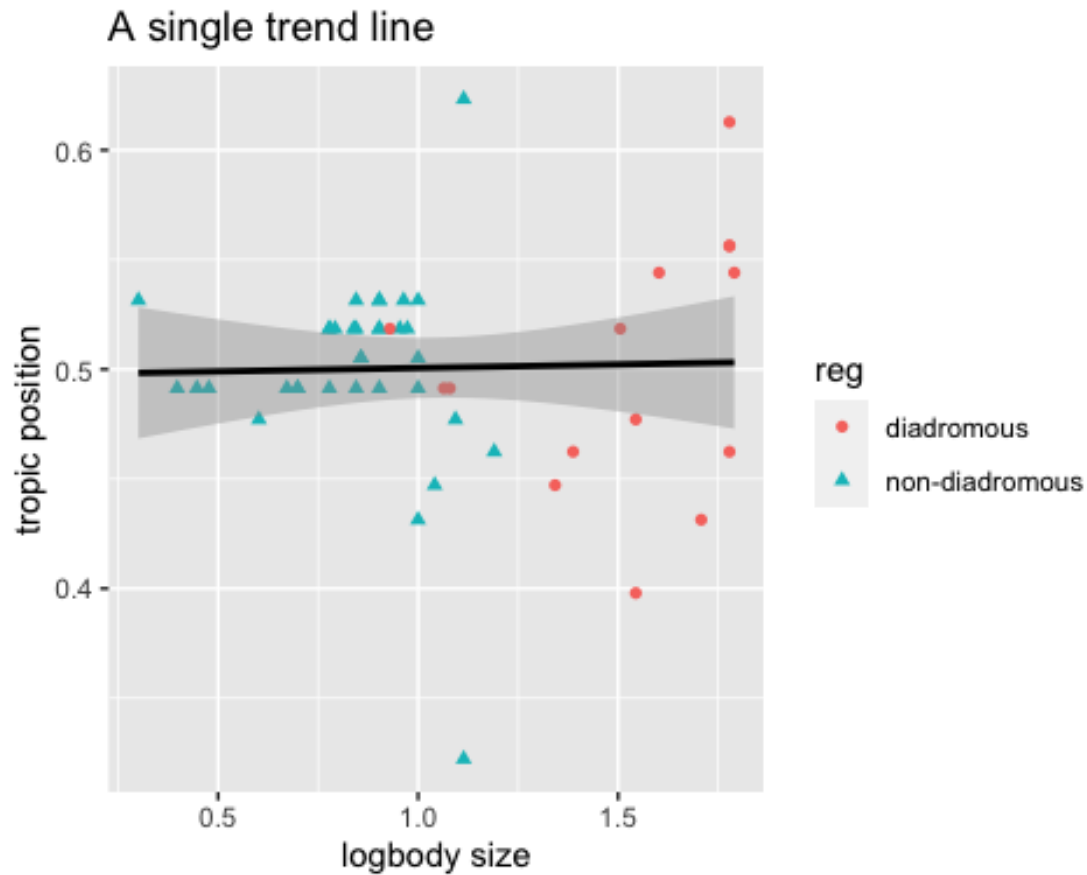
Second plot

- a single trend line for the whole model

#plot 1.2

```
ggplot(data = bloom_df, aes(x=logbodysize, y=trophic_position))+
  geom_point(aes(colour=reg, shape=reg))+
  stat_smooth(method = 'lm', colour = "black")+
  ggtitle('A single trend line')+
  scale_x_continuous(name = 'logbody size')+
  scale_y_continuous(name = 'trophic position')
```

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



Task 2: Continuous vs. categorical

The layout will be:

- 2 panels: trophic_position and logbodysize
- x axis: reg
- y axis: value of trophic_position and logbodysize

Create 4 plots:

- bar and error bars (mean and 95% conf. int.)
- point and error bars (mean and 95% conf. int.)
- box plots
- raw data + point and error bars (mean and 95% conf. int.)

All these plots should have the same general layout but change the type of plot to visualize the data.

HINT: logbodysize and trophic_position need to be turned into long format

Example layout:

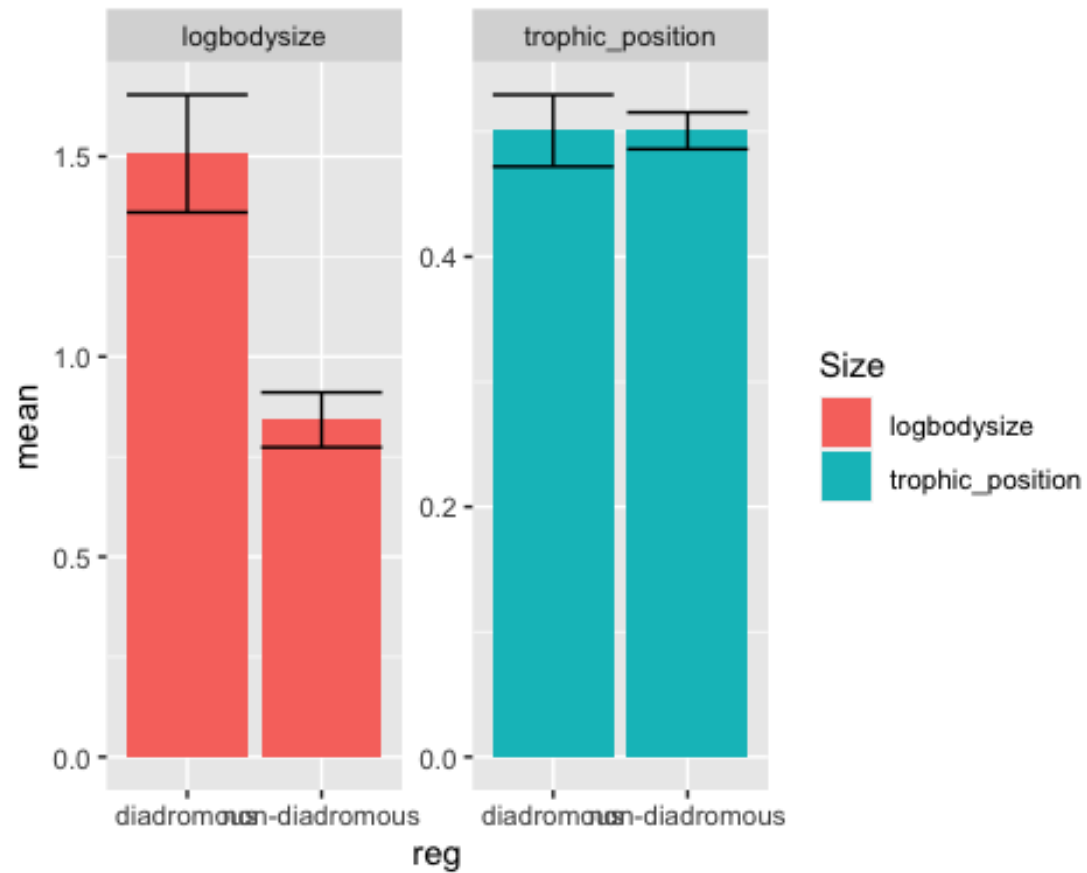
sample layout for task2

sample layout for task2

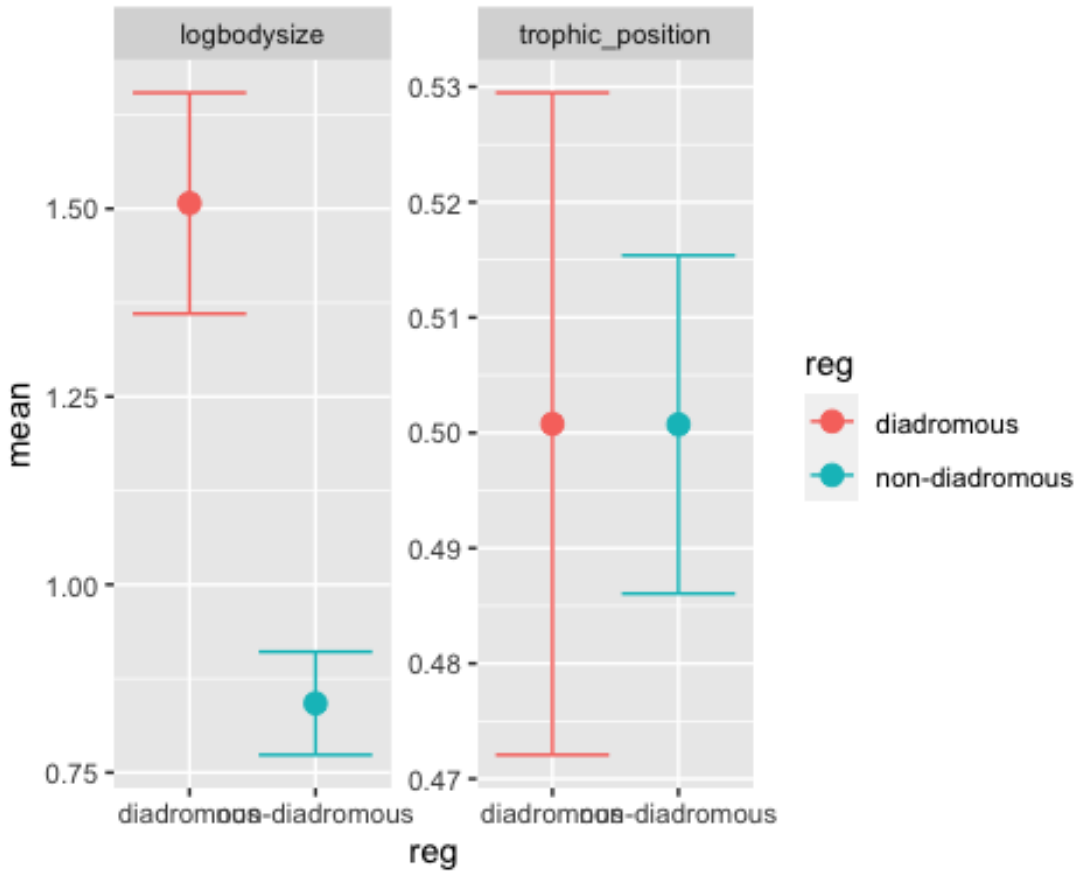
```
#calculation and summarization
logbodysize_sum_df <- bloom_df %>%
  group_by(reg) %>%
  summarise(mean = mean(logbodysize, na.rm= TRUE),
            sd = sd(logbodysize, na.rm= TRUE),
            n = n()) %>%
  mutate(se = sd / sqrt(n),
         ci = 1.96*se)
logbodysize_sum_df$Size <- 'logbodysize'

trophic_position_sum_df <- bloom_df %>%
  group_by(reg) %>%
  summarise(mean = mean(trophic_position, na.rm= TRUE),
            sd = sd(trophic_position, na.rm= TRUE),
            n = n()) %>%
  mutate(se = sd / sqrt(n),
         ci = 1.96*se)
trophic_position_sum_df$Size <- 'trophic_position'
bloom_sum_df <- rbind(logbodysize_sum_df, trophic_position_sum_df)

#plot 2.1
ggplot(data=bloom_sum_df, aes(x=reg, y=mean, fill=Size))+
  facet_wrap(~Size, nrow=1, scales = 'free')+
  geom_bar(stat = 'identity')+
  geom_errorbar(aes(ymin = mean - ci, ymax = mean + ci))
```

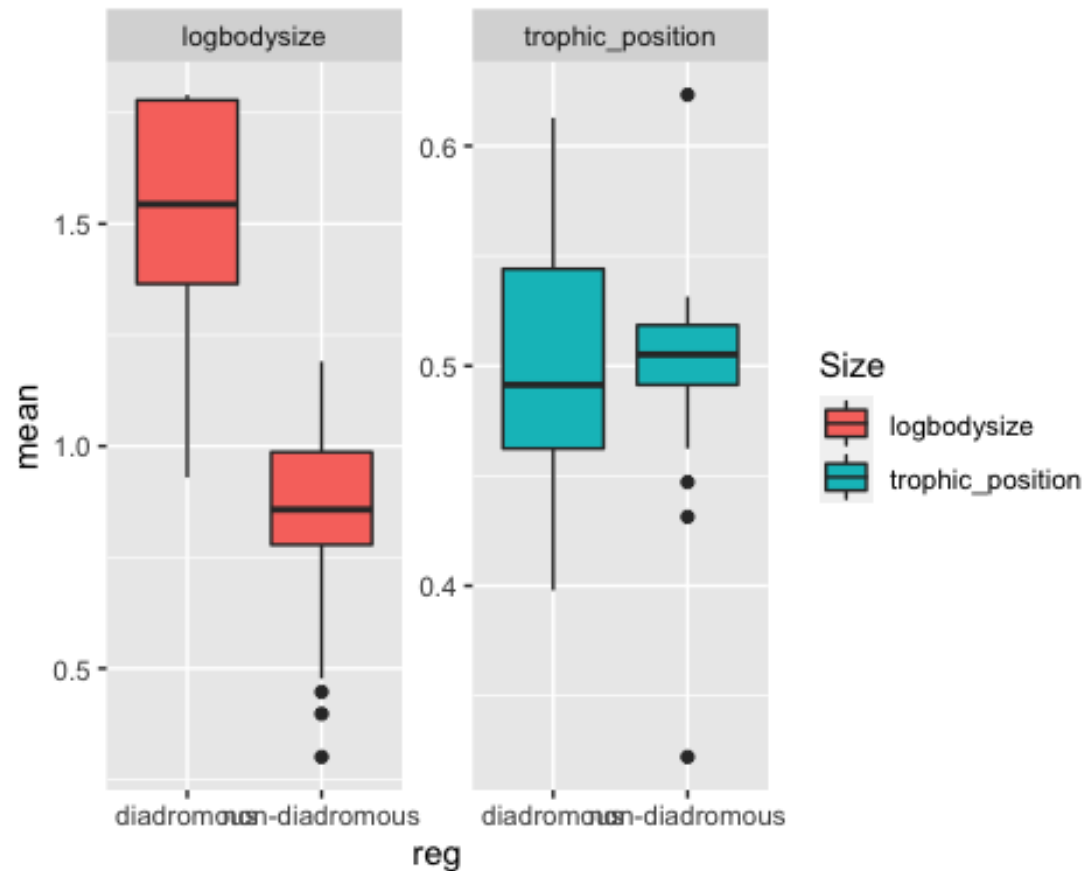


```
#plot 2.2
ggplot(data=bloom_sum_df, aes(x=reg, y=mean, colour=reg))+
  facet_wrap(~Size, nrow=1, scales = 'free')+
  geom_point(size = 3)+
  geom_errorbar(aes(ymin = mean - ci, ymax = mean + ci))
```



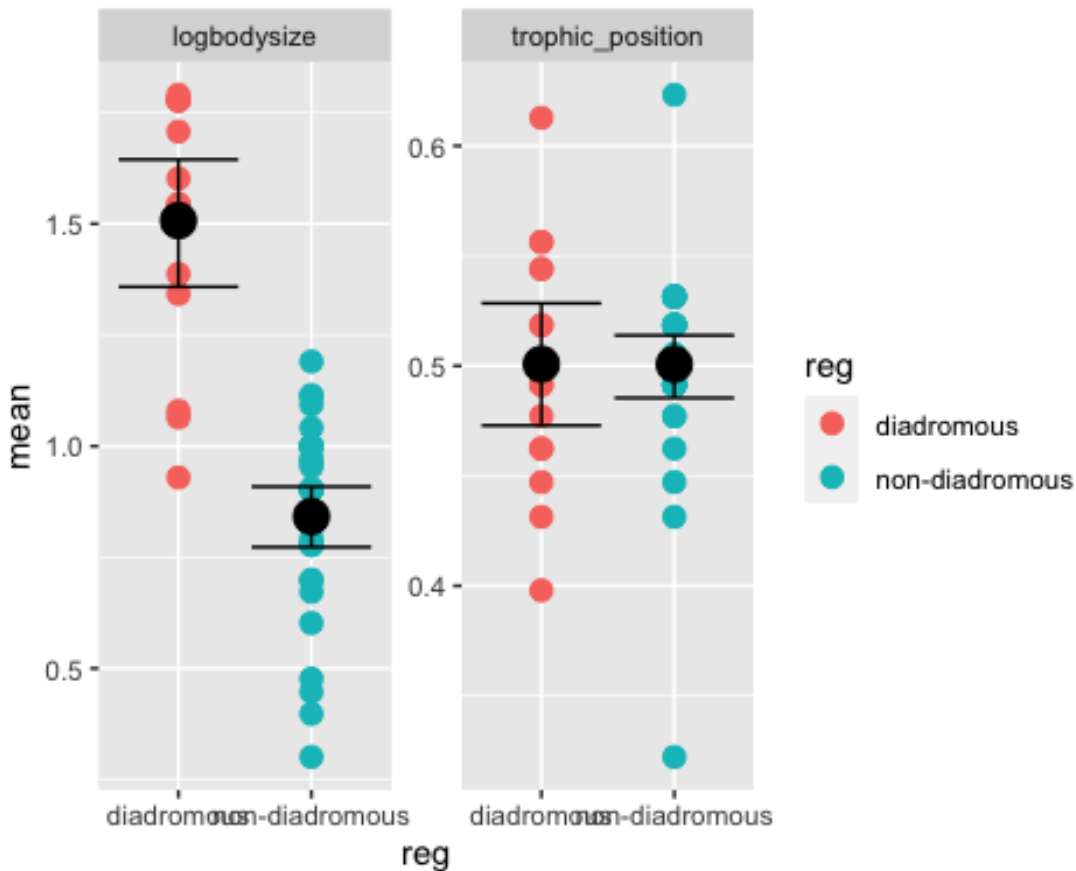
```
#Change data into Long format
bloom_long_df <- bloom_df %>%
  gather(key = Size, value = mean, c(logbodysize, trophic_position))

#plot 2.3
ggplot(data=bloom_long_df, aes(x=reg, y=mean, fill=Size))+
  facet_wrap(~Size, nrow=1, scales='free')+
  geom_boxplot()
```



#plot 2.4

```
ggplot(data=bloom_long_df, aes(x=reg, y=mean, colour=reg))+
  facet_wrap(~Size, nrow=1, scales = 'free')+
  geom_point(size=3)+
  stat_summary(fun.data = mean_cl_boot, geom = "errorbar", color='black')+
  stat_summary(fun = mean, geom = "point", size=5, color='black')
```

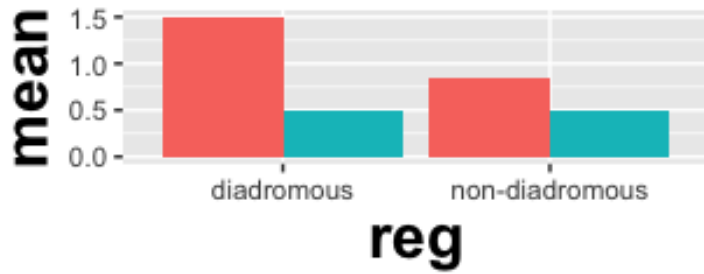
Task 3: layering your plots

Make any two figures with the bloom_df! Can be something we've done or something completely different. Make one TERRIBLE figure and one GREAT figure.

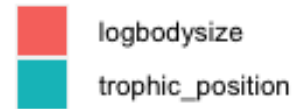
Put them together using *patchwork* or *ggarrange*

```
#plot 3
#This is a worst plot.
worst_plot <- ggplot(data = bloom_sum_df, aes(x=reg, y=mean, fill=Size))+
  geom_bar(stat = 'identity', position = 'dodge')+
  ggtitle('Worst plot')+
  theme(title = element_text(size = 20, colour = "black", face = "bold"))
#This is a good plot.
good_plot <- ggplot(data=bloom_long_df, aes(x=reg, y=mean, colour=reg))+
  facet_wrap(~Size, nrow=1, scales = 'free')+
  ggtitle('Good plot')+
  geom_point(size=2)+
  stat_summary(fun.data = mean_cl_boot, geom = "errorbar", color='black')+
  stat_summary(fun = mean, geom = "point", size=3, color='black')
#We are using patchwork to put worst plot and good plot together.
worst_plot + good_plot + plot_layout(heights = c(3,5))
```

Worst plot



Size



Good plot

