

Homework5-DataVis-ggplot

YOUR NAME HERE

DUE: Thursday, March 25

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('~/.g/Classes/DataScience/data_vis/') #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          : Factor w/ 50 levels "Alosa_alabamae",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ genus         : Factor w/ 34 levels "Alosa","Amazonsprattus",...: 1 1 1 1 1 1 2 3 3 3 ...
## $ species       : Factor w/ 48 levels "aestuaria","alabamae",...: 2 3 19 31 36 39 41 11 13 14 ...
## $ 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           : Factor w/ 2 levels "diadromous","non-diadromous": 1 1 1 1 1 1 2 2 2 2 ...
```

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
```

Second plot

- a single trend line for the whole model

```
#plot 1.2
```

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

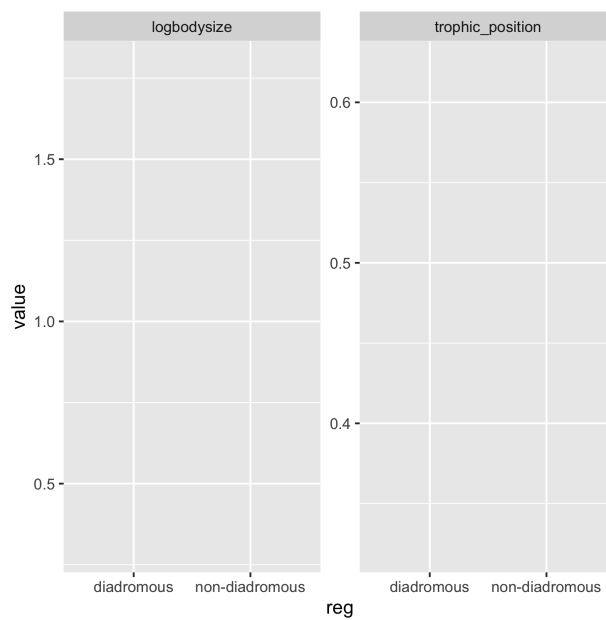


Figure 1: sample layout for task2

#plot 2.1

#plot 2.2

#plot 2.3

#plot 2.4

Example layout:

Task 3: layering your plots

Take your two favorite figures that you made and put them together using *patchwork* or *ggarrange*

Feel free to make any other figure you want with this data and include it here. Have fun with it, play around.

```
#plot 3
```