R Assignment #3 Key

*Marine Biology: Spring 2024*

## **Third R Assignment:**

Using the data we have collected as a class for lab skills (Lab\_Skills\_Data.xlsx), you will complete the following tasks. For each figure you create, make sure you use descriptive X and Y axis labels with units where applicable and write an informative figure caption. All figures, captions, and tables should be saved and placed within a Word document to submit along with a brief description of your results.

1. Make a bar plot of average (+/- standard deviation) value per sample for each of the following parameters:
   * pH
   * temperature
   * salinity
   * weights
   * calipers (L x W x H)
2. Make a scatter plot of the following parameters:
   * salinity versus temperature
   * shell width versus height
3. Find a new data visualization/figure to create in R to present some other relationship in your data that we have not learned yet in another R assignment (i.e., no line, scatter, or bar plots). Be creative and have fun!
4. Perform a statistical test in R (I recommend a T test or ANOVA) on one of the parameters you plotted in part 1 above. Present the results of your statistical test in a table.
5. Finally, write a brief results-style section discussing the statistical test you performed in part 4. In this section, present the p-value and discuss if your results were statistically different or not.

### Question 1: bar plot of average (+/- standard deviation)

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.4.4 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(readxl)

# load temp/salinity data  
temp\_sal\_df <- read\_excel("~/Dropbox/Git/teaching\_Rtutorial/Part\_3/Lab\_Skills\_Data.xlsx", sheet = "Station 1 Salinity and Temp")

## New names:  
## • `` -> `...14`

# load pH data  
pH\_df <- read\_excel("~/Dropbox/Git/teaching\_Rtutorial/Part\_3/Lab\_Skills\_Data.xlsx", sheet = "Station 2 pH")

## New names:  
## • `` -> `...1`  
## • `` -> `...9`

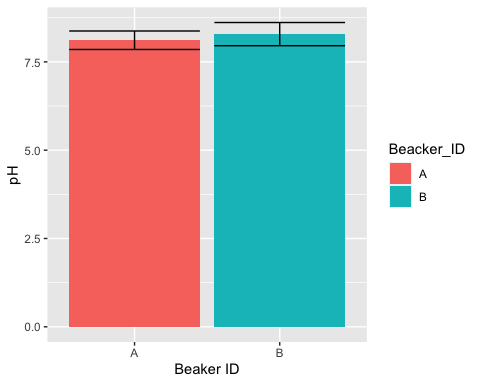
# load weight data  
weight\_df <- read\_excel("~/Dropbox/Git/teaching\_Rtutorial/Part\_3/Lab\_Skills\_Data.xlsx", sheet = "Station 5 Byoyant Weight")  
  
# load length/width data  
calip\_df <- read\_excel("~/Dropbox/Git/teaching\_Rtutorial/Part\_3/Lab\_Skills\_Data.xlsx", sheet = "Station 4 Calipers")

## New names:  
## • `` -> `...8`

#### pH

# remove the first two rows of pH data (one is fake, the other is blank)  
pH\_df <- pH\_df[-1:-2,]  
  
# convert from wide to long to use all pH data  
pH\_long <- pH\_df %>%   
 pivot\_longer(names\_to = "pH\_meas", values\_to = "pH", pH\_1:pH\_3)   
  
# summary of pH data by beaker ID  
pH\_sum <- pH\_long %>%   
 group\_by(Beacker\_ID) %>%   
 summarise(pH\_mean = mean(pH),  
 pH\_sd = sd(pH))

ggplot(data = pH\_sum, aes(x = Beacker\_ID, y = pH\_mean, fill = Beacker\_ID)) +  
 geom\_bar(stat = "identity") +  
 geom\_errorbar(aes(ymin = pH\_mean - pH\_sd, ymax = pH\_mean + pH\_sd)) +  
 labs(x = "Beaker ID", y = "pH")

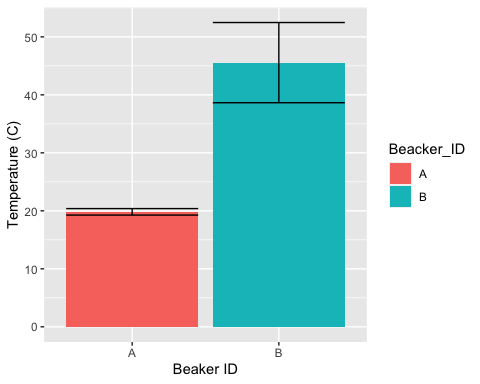


#### temperature

# remove the first two rows of temp data (one is fake, the other is blank)  
temp\_sal\_df <- temp\_sal\_df[-1,] %>%   
 drop\_na(Beacker\_ID)  
  
# convert from wide to long to use all temp data  
temp\_sal\_long <- temp\_sal\_df %>%   
 pivot\_longer(names\_to = "temp\_meas", values\_to = "temp", Temp1:Temp3) %>%   
 pivot\_longer(names\_to = "sal\_meas", values\_to = "sal", Sal\_probe\_1:Sal\_probe\_3)

# summary of temp data by beaker ID  
temp\_sum <- temp\_sal\_long %>%   
 group\_by(Beacker\_ID) %>%   
 summarise(temp\_mean = mean(temp),  
 temp\_sd = sd(temp))

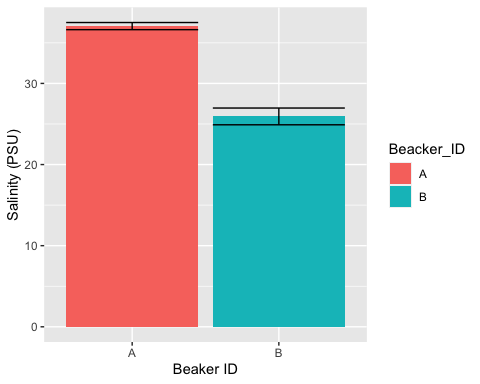
ggplot(data = temp\_sum, aes(x = Beacker\_ID, y = temp\_mean, fill = Beacker\_ID)) +  
 geom\_bar(stat = "identity") +  
 geom\_errorbar(aes(ymin = temp\_mean - temp\_sd, ymax = temp\_mean + temp\_sd)) +  
 labs(x = "Beaker ID", y = "Temperature (C)")



#### salinity

# summary of sal data by beaker ID  
sal\_sum <- temp\_sal\_long %>%   
 group\_by(Beacker\_ID) %>%   
 summarise(sal\_mean = mean(sal),  
 sal\_sd = sd(sal))

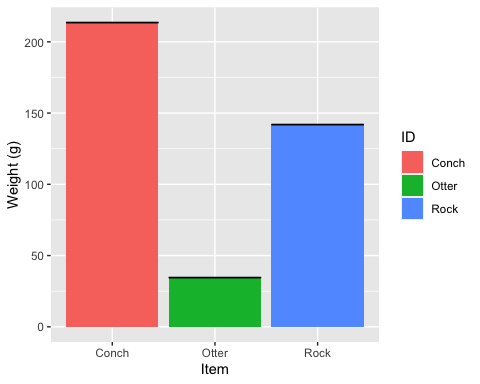
ggplot(data = sal\_sum, aes(x = Beacker\_ID, y = sal\_mean, fill = Beacker\_ID)) +  
 geom\_bar(stat = "identity") +  
 geom\_errorbar(aes(ymin = sal\_mean - sal\_sd, ymax = sal\_mean + sal\_sd)) +  
 labs(x = "Beaker ID", y = "Salinity (PSU)")



#### weights

# convert from wide to long to use all weight data  
weight\_long <- weight\_df %>%   
 drop\_na(weight\_1) %>%   
 pivot\_longer(names\_to = "weight\_meas", values\_to = "weight", weight\_1:weight\_3)  
  
# summary of weight data by beaker ID  
weight\_sum <- weight\_long %>%   
 group\_by(ID) %>%   
 summarise(weight\_mean = mean(weight),  
 weight\_sd = sd(weight))

ggplot(data = weight\_sum, aes(x = ID, y = weight\_mean, fill = ID)) +  
 geom\_bar(stat = "identity") +  
 geom\_errorbar(aes(ymin = weight\_mean - weight\_sd, ymax = weight\_mean + weight\_sd)) +  
 labs(x = "Item", y = "Weight (g)")

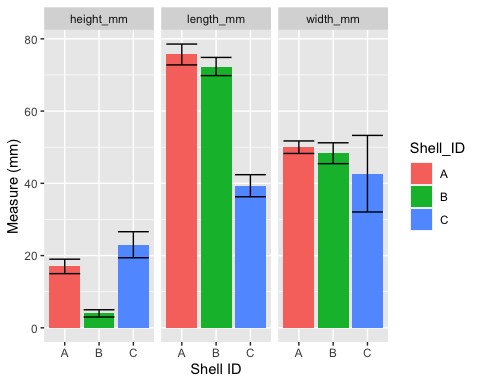


#### calipers (L x W x H)

# convert from wide to long to use all calip data  
calip\_long <- calip\_df[-1,] %>%   
 drop\_na(Shell\_ID) %>%   
 pivot\_longer(names\_to = "calip\_meas", values\_to = "measure", length\_mm:height\_mm)  
  
# summary of calip data by beaker ID  
calip\_sum <- calip\_long %>%   
 group\_by(Shell\_ID, calip\_meas) %>%   
 summarise(calip\_mean = mean(measure),  
 calip\_sd = sd(measure))

## `summarise()` has grouped output by 'Shell\_ID'. You can override using the  
## `.groups` argument.

ggplot(data = calip\_sum, aes(x = Shell\_ID, y = calip\_mean, fill = Shell\_ID)) +  
 geom\_bar(stat = "identity") +  
 geom\_errorbar(aes(ymin = calip\_mean - calip\_sd, ymax = calip\_mean + calip\_sd)) +  
 labs(x = "Shell ID", y = "Measure (mm)") +  
 facet\_wrap(~calip\_meas)



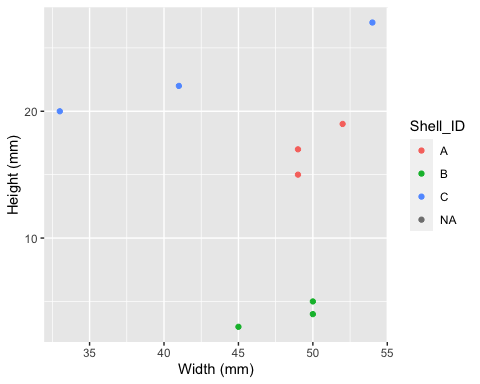
### Question 2:

ggplot(data = temp\_sal\_long, aes(x = sal, y = temp, color = Beacker\_ID)) +  
 geom\_point() +  
 labs(x = "Salinity (PSU)", y = "Temperature (C)")



ggplot(data = calip\_df[-1,], aes(x = width\_mm, y = height\_mm, color = Shell\_ID)) +  
 geom\_point() +  
 labs(x = "Width (mm)", y = "Height (mm)")

## Warning: Removed 4 rows containing missing values (`geom\_point()`).



### Question 3:

Anything goes here! But this website has some great examples of different data visualization options: <https://r-graph-gallery.com/>

### Question 4:

T test option:

library(knitr)  
  
# comparing pH across two beakers  
t.test(pH ~ Beacker\_ID, data = pH\_long)

##   
## Welch Two Sample t-test  
##   
## data: pH by Beacker\_ID  
## t = -1.4305, df = 20.961, p-value = 0.1673  
## alternative hypothesis: true difference in means between group A and group B is not equal to 0  
## 95 percent confidence interval:  
## -0.42534030 0.07867363  
## sample estimates:  
## mean in group A mean in group B   
## 8.112500 8.285833

# OR using the rempsyc package:  
library(rempsyc)  
  
ttest2 <- nice\_t\_test(data = pH\_long, response = "pH", group = "Beacker\_ID")  
kable(ttest2)

| Dependent Variable | t | df | p | d | CI\_lower | CI\_upper |
| --- | --- | --- | --- | --- | --- | --- |
| pH | -1.430545 | 20.96095 | 0.1673001 | -0.5840175 | -1.39604 | 0.2406335 |

ANOVA option:

summary(aov(pH ~ Group, data = pH\_long))

## Df Sum Sq Mean Sq F value Pr(>F)   
## Group 4 1.5992 0.3998 14.64 1.29e-05 \*\*\*  
## Residuals 19 0.5189 0.0273   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Question 5:

Again, this section with vary but key details I am looking for here include:

* the parameter being tested (*e.g.*, pH, temperature, etc.)
* statement of what statistical test was run
* the p value and interpretation of the p value
* broad conclusion of what this may mean

*This answer key was created by Colleen Bove, please feel free to update as needed for your own assignments!*