R Assignment #2 Key

*Marine Biology: Spring 2024*

## **Second R Assignment:**

1. **Create a R markdown document that will knit as a Word document.** You will submit both the R markdown code and the Word document for this assignment so this will be assessed as a part of your submission.
2. **Make a bar plot of average temperature in tank A vs tank B** You will need to calculate the summary statistics of the data before creating your bar plot. Make sure to also include error bars for standard error, appropriate x and y axis labels, and a brief figure caption.
3. **Make a scatter plot to observe the relationship between temperature and salinity for tank B** Create this plot with temperature along the x axis and salinity along the y axis. Your scatter plot should also display the linear trend of the data using geom\_smooth().
4. **Perform a linear regression to assess the relationship between temperature and salinity for tank B** Calculate the equation of the line depicted on your scatter plot for question 4. Show your code for how you performed your regression and then have your markdown file report the equation of your line (), your R2 value, and your p-value from your regression.
5. **Make a line graph showing salinity in BOTH tanks over time** Using what you learned above, create a line graph to demonstrate the change tank salinity over time in both tank **A** and **B**. Make sure you provide informative x and y labels, as well as a brief figure caption.

### Question 1:

For full credit, you should submit both the Rmarkdown (.Rmd) and resulting Word (.doc or .docx) documents demonstrating your code and the resulting output.

### Question 2:

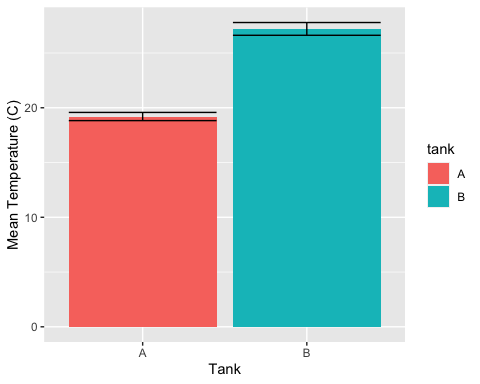
wd <- rprojroot::find\_rstudio\_root\_file() # setting working directory as the Project Directory  
  
  
water\_data <- read.csv(paste0(wd, "/Part\_2/WaterData\_spring24.csv"), header = TRUE)

# loading tidyverse to use for summary  
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

# construct summary dataframe  
temp\_sum <- water\_data %>%   
 group\_by(tank) %>%   
 summarise(mean\_temp = mean(temperature\_C),  
 temp\_sd = sd(temperature\_C),  
 N = n(),  
 temp\_se = temp\_sd / sqrt(N))

# plot of mean +/- SE of temp across tanks  
ggplot(data = temp\_sum, aes(x = tank, y = mean\_temp, fill = tank)) +  
 geom\_bar(stat = "identity") +  
 geom\_errorbar(aes(ymin = mean\_temp - temp\_se, ymax = mean\_temp + temp\_se)) +  
 labs(x = "Tank", y = "Mean Temperature (C)")

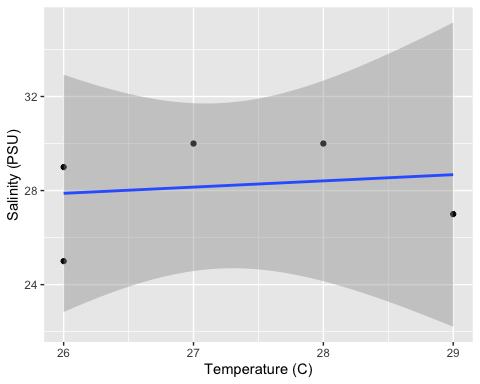
 **Figure 1**: Mean temperature of tank A and tank B with standard error bars.

### Question 3:

# filter the full data for tank B data only  
tankB <- water\_data %>%   
 filter(tank == "B")

ggplot(data = tankB, aes(x = temperature\_C, y = salinity\_psu)) +  
 geom\_point() +  
 geom\_smooth(method = lm) +  
 labs(x = "Temperature (C)", y = "Salinity (PSU)")

## `geom\_smooth()` using formula = 'y ~ x'



**Figure 2**: The relationship between temperature in degrees C against salinity (PSU) measured in tank B. The line represents a linear regression of temperature against salinity.

### Question 4: Perform a linear regression to assess the relationship between temperature and salinity for tank B

Calculate the equation of the line depicted on your scatter plot for question 4. Show your code for how you performed your regression and then have your markdown file report the equation of your line (), your R2 value, and your p-value from your regression.

tankB\_lm <- lm(temperature\_C ~ salinity\_psu, data = tankB)  
tankB\_lm

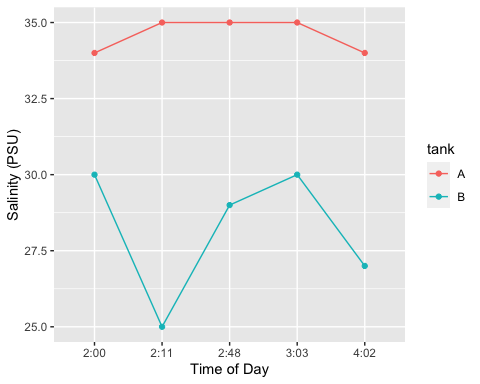
##   
## Call:  
## lm(formula = temperature\_C ~ salinity\_psu, data = tankB)  
##   
## Coefficients:  
## (Intercept) salinity\_psu   
## 24.50000 0.09574

summary(tankB\_lm)

##   
## Call:  
## lm(formula = temperature\_C ~ salinity\_psu, data = tankB)  
##   
## Residuals:  
## 1 2 3 4 5   
## -0.8936 -1.2766 1.9149 0.6277 -0.3723   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 24.50000 9.68978 2.528 0.0855 .  
## salinity\_psu 0.09574 0.34280 0.279 0.7982   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.486 on 3 degrees of freedom  
## Multiple R-squared: 0.02534, Adjusted R-squared: -0.2995   
## F-statistic: 0.07801 on 1 and 3 DF, p-value: 0.7982

### Question 5: Make a line graph showing salinity in BOTH tanks over time

ggplot(data = water\_data, aes(x = time, y = salinity\_psu, color = tank, group = tank)) +  
 geom\_point() +  
 geom\_line() +  
 labs(x = "Time of Day", y = "Salinity (PSU)")

 **Figure 3**: Salinity (PSU) of Tank A (pink) and Tank B (blue) at different times of day (ignoring date).

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