Due Date: October 28, 2021, 7 pm

**Overview**:

For this assignment, you will conduct and present a clean, professional analysis of a dataset on owl limpets in southern California.

You will be graded on your professional presentation using R Markdown, your demonstration of ggplot and dplyr skills, and your thoughtful interpretation of your data analysis. For full credit, your R Markdown document should have:

* A YAML header with your name, the title of your analysis, and the date of the assignment
* Chunks for setup knitr, installing and loading packages, for importing data, with outputs hidden from knitted file using chunk settings.
* An brief summary of the goal of the analysis, after the initial chunks
* Nicely formatted headers to distinguish the parts of the analysis and questions
* Separate, named chunks for each figure and summary, with bulky/uninformative code outputs hidden from output file using chunk settings.
* Text answers to questions in full sentences, explaining methods, results and interpretation of the figure or table, as a full report.

**Set-up & Upload Instructions:**

1. Navigate to the “Problem Set 2” Project in your RStudio Cloud, which should already contain the data files you need for this assignment, as well as these instructions.
2. Create a new .Rmd document, and delete everything under the Setup chunk.
3. Save your .Rmd file with the name convention: Lastname\_PS2\_DDMMYY.Rmd
4. Paste the following questions in your new Markdown document to get you started and structure your analysis.
5. Once you are done, knit your analysis to PDF using the Knit button dropdown, Knit to PDF.
6. To download your project,
   1. go to the Files pane
   2. tick all the boxes next to each file
   3. Click the cogwheel/More button and select “Export…”
   4. Look in your computer’s Downloads folder, where there should be a .zip file containing all the files you need to upload for full credit, including:
      1. Your .Rmd file
      2. The knitted .pdf file (which you will mark your answers on the Gradescope outline)
      3. *(you do not need to upload the data or the .Rproj files this time)*
7. Upload both files to the appropriate Gradescope assignment.

**PART 1 - Comparing Crystal Cove vs. Corona del Mar**

Use the AllOwls\_2020\_CDM\_CCSP.csv data file (the output of the data frame that we created in our class module) to present and interpret the following analyses professionally. Use chunk settings to hide all code and intermediate outputs that are not part of your results!

1. TABLE 1: Present a nicely formatted table with a summary of abundance, mean size, standard deviation of size, and median size at the four sites (CCSP High Impact, CCSP Low Impact, CDM High Impact, CDM Low Impact).
   1. use knitr::kable() to print your final output in the knitted report
   2. use mutate to apply the round() function to columns with lots of digits.
      1. template: mutate(columnname = round(columnname, digits = 2)
2. FIGURE 1: Present a nicely formatted barplot of abundance at the four sites, colored by Impact and Site.
   1. Label your axes, including units on your y-axis.
3. Is there a difference in mean size between the high and low impact sites at CDM? Report your results in a sentence, including the test statistic and p-value.
   1. Test your assumptions
      1. Are the data normally distributed?
      2. Do the two samples have homogeneous variance?
   2. Use a t-test or a Wilcoxon-rank-sum test, depending on the normality of your data.
4. FIGURE 2: Present a nicely formatted boxplot OR a column plot with errorbars for the CDM high/low comparison with Impact on the x axis and size on the y-axis.
   1. Make your decision about which type of plot to present based on whether the data are roughly normally distributed. Explain in your notes.
   2. Include units on your y-axis.
   3. Include an annotation on your graph to show the results of your t-test (\* and p-value)
5. FIGURE 3: Repeat Questions 3 and 4 for the CCSP data.
6. Write a brief paragraph to explain your results, your interpretation, and some hypotheses about their importance in the context of what you learned about owl limpets from Heidi’s lecture.

**PART 2 - Comparing Shaws Cove vs. Dana Point**

The data used in this section is new to you and collected with a different method than the CDM/CCSP data. In this dataset, each site has 5 fixed plots, which are surveyed twice annually. This replication within a site allows us to estimate population density. You will be responsible for tidying and reformatting this dataset using the same methods we used for the previous dataset. There are two files, but all data analysis will be done on the data one. The metadata is just for your reference.

SHCO\_DAPT\_lottia\_count\_sizes.csv : the data

SHCO\_DAPT\_lottia\_metadata.csv : information about each of the columns.

Formatting steps:

1. Reduce this dataset to the columns site\_code, marine\_common\_year, season\_sequence, survey\_date, plot\_code, lottia\_size\_bin, total.
2. Filter to only use season\_sequence == 3, then remove the season\_sequence column.
3. Rename relevant columns to Site, Year, Date, Plot.
4. Uncount the rows so that there is one row per owl limpet.

Part 2 Questions:

1. Create (but do not print) a table that summarizes the count of owl limpets per plot at each site in each year.
   1. hint: This table should have 10 rows.
2. TABLE 2: Create a second table by applying the summarize function to Table 2, to calculate the average plot abundance (i.e., mean of the 5 plots per site) and standard error (of the 5 plot abundances)
   1. Use mutate() to apply the round function to your average plot abundance column.
   2. Use pivot\_wider() again to pivot this table so that Years are on the columns, and site and plot are on the rows. Print using knitr::kable(). This table should have 2 rows.
3. Create a table that summarizes the mean and standard deviation of owl limpet size per plot at each site in each year.
4. TABLE 3: Create a second table by applying the summarize function to Table 2, to calculate the mean mean size (i.e., mean of the 5 plot mean sizes) and the standard error mean size (i.e., the standard error of the 5 plot size means)
   1. Use mutate() to apply the round function to your average plot abundance column.
   2. Use pivot\_wider() again to pivot this table so that Years are on the columns, and site and plot are on the rows. Print using knitr::kable(). This table should have 2 rows.
5. Using Table 2, graph the time series of mean abundance (i.e., mean of the 5 plot counts) of owl limpets over time
   1. Map site to color.
   2. Add error bars for standard error.
   3. Label your axes, including units
6. Using Table 3, graph the time series of mean mean size (i.e., mean of the 5 plot size means) of owl limpets over time.
   1. Map site to color.
   2. Add error bars for standard error.
   3. Label your axes, including units
7. Visually assess these plots, and identify two years **within one site's** time series for either variable (mean size or mean abundance) that you would like to compare using a t-test. Write this question using the example format.
   1. Example: Did average limpet size at SHCO differ significantly between 2013 and 2016?
8. Perform a paired t-test to answer your question using the plots as the replicates. Report your results in a sentence.
9. Write a brief paragraph to explain your results, your interpretation, and some hypotheses about their importance in the context of what you learned about owl limpets from Heidi’s lecture.