**Homework 3 – Some Programming Practice**

A picture containing map

Description automatically generated

The goal of this assignment is to have you practice some basic R programming / data wrangling. The attached file, “CRCP\_Reef\_Fish\_Surveys\_Hawaii\_expanded.csv”, includes survey data of reef fish on Hawaiʻi island collected by NOAA (yellow circles in the map above = survey location, [see here](https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:NCRMP-Fish-HI)). I have slightly modified the data (added zeros for when a species it not observed at a location, removed some columns we will not be using).

(1) Let’s look at some basic summary statistics. In total there are 209 fish species encountered and counted in these surveys. Identify the top twelve species in terms of these three statistics: (1) mean abundance, (2) maximum abundance, (3) coefficient of variation in abundance. In other words, which species are most abundant on average, which reach the highest abundances, and which have the most variable abundances?

To make these calculations, you’ll want to focus on the ‘count’ column (number of fish observed at a location), as well as the ‘commonname’ or ‘taxonname’ columns. The latter two are nearly identical, you can choose which you prefer to use. To do these calculations in base R, tapply() or aggregate() will be useful, or to use the ‘tidyverse’ you can look into the summarize() function in the package dplyr.

(2) Now let’s visualize how the abundances of the most common species vary with depth, to get a sense for whether species have different depth niches. Using the top twelve species based on mean abundance, plot a scatterplot of count vs. depth (the column is named ‘depth’). Add a smoother to help visualize mean count vs. depth, put all twelve species on a single plot (with twelve panels), and make sure each panel is titled using the species name. The patterns in abundance will be easier to see if you transform the counts, for example by taking the square root of the counts, because the counts are very skewed.

I would like you to make this twelve-paneled figure in two ways. First, write a ‘for’ loop in which you iterate your plotting code for the twelve species. The function scatter.smooth() is a simple one for making a scatterplot with a smoother. Second, use ggplot() to do the same thing. The functions geom\_point(), geom\_smooth(), and facet\_wrap() will be helpful.

The ggplot way of doing this may be slightly easier (once you’ve figured out the ggplot syntax), but I would like you to do it both ways so that I know that you know how to write a loop, which is a very general and important programming technique.

Finally, what are your conclusions from this visual inspection of the data?