To begin the process of exploratory data analysis and data cleaning, I first identified the months in which we have observations and the year of the first observation for each location-age class combination we're interested in. I also labeled the observations by "season" because the observations from November and December should be grouped with the observations from January, February, and March of the following year as this constitutes one birthing/breeding season.

Looking only at the cow counts, I have observations for every year from 1981 to 2020 for Point Reyes Headlands. Observations from Drakes Beach and South Beach start in 1995. Using Condit et al.'s (2020) correction factors for Point Reyes found in table A3, I estimated the total female population for each year and site individually using the maximum count between January 26 and February 2 when available and the overall maximum count when there are no observations during this period. I manually corrected the estimate for two year-location combinations. Point Reyes Headlands did not have any observations between January 15 and February 8 in 1994, so I used the count from February 9 and the correction factor for February 8 to get the estimate for that year[1]. In 2019, Drakes Beach had bad visibility between January 26 and February 2, so Sarah Codde advised me to use the count on January 23. Condit et al. provide two additional correction factors that allowed me to create a 95% confidence interval and I have included these upper and lower estimates in my final table. Other than Point Reyes Headlands 1994, if the maximum count fell outside of date range Condit et al. included correction factors for I wrote in the point estimate, upper estimate, and lower estimate as the maximum count in order to have these measures for each year-location combination.

As for pups, PR Headlands has observations starting in 1981, Drakes Beach from 1995, and South Beach has one observation in 1983 and then none until 1995. The data I was given had two categories for pups: "PUP" and "EPUP". "EPUP"s only appeared in 2018, 2019, and 2020. After visually inspecting the data I did not find any overlap in date and location between "PUP" and "EPUP", so I relabeled all the "EPUP"s to "PUPs". I did not use any correction factors or estimates, so for each year I found the maximum pup count between January 26 and February 2. If there were no observations in this date range, I used the maximum count regardless of date. I also looked at pup and weaner counts together. Unlike cows and pups, weaners move between the beaches so in order to avoid double counting, I combined the counts across all locations

[1] Condit et al. only provide correction factors for dates between January 11 and February 8.

before finding the maximum count from all locations on a single day, again between January 26 and February 2 when possible[2].

For cows, pups, and pups/weaners I created basic R plots. Each plot is for a specific year and age class combination across time where the maximum count or estimate is plotted. While I worked primarily with the cow, pup, and weaner age classes, I also individually plotted the maximum daily count for each year of the other age classes (excluding dead pups) and aggregated all age classes across all locations to determine and plot the maximum daily count for each year. The code for these basic plots can be found in the "Other Age Classes.R" file.

I began fitting linear models to the data for each age class-location combination where the log count is a function of the year. I used `lm()` for basic linear models, `glm()`for generalized linear models with the errors poisson distributed, `gam()`for generalized additive mixed models, and `gamm4()` for generalized additive mixed models where the model divides itself based on the location. For all models except `gamm4()`, I passed in each unique age class-location combination for pups/weaners and cows to train a model for each age class-location combination.

[2] From here I can find the count for each location on the day determined to have the maximum count