Oregon State University
CS512
Paul Anderson
Mid Term Project (Independent)
Rough Outline

Here's the plan based on the data Team Nexus processed for our first project. I've run some initial analyses to spot patterns and found some interesting preliminary results. We're looking at data from 2002-2004 (before the habitat was marked as critical) and 2020-2022 (after the designation), avoiding any El Nino or La Nina impacts. Turns out, there's no significant change in Chinook Salmon SAR between these periods, but meanSST and MaxSST show a noticeable difference.

To dig deeper, I'll process some extra data to figure out these patterns and see if other variables might explain these results. We'll pull in data from stoplight-raw-data-2024.csv, including stuff like Copepod richness, PDO (Dec-March), PDO (May-Sept), and Chinook salmon juvenile catches. I'll use Pandas in Python to wrangle this data—trimming columns and rows, transposing, saving, and joining dataframes. Then, we'll upload everything to a SQL database, visualize it using Matplotlib, and create nested models. We'll evaluate these models with ANOVA, AIC, and BIC to identify any significant relationships behind our findings.

1. Existing Dataset
    a. We were concerned with a period of 2002-2004 before this habitat was designated as critical habitat
    b. And the period 2020-2022 after the habitat was designated as critical habitat
    c. Furthermore we chose these periods to avoid influence by El Nino, La Nina impacts
    d. We found no significant change between our two time periods for Chinook Salmon SAR
    e. But we found a significant change in this time period in meanSST
    f. And a significant change in this period in MaxSST
2. Additional Variables
    a. We wanted to understand more about these variables and if there were additional variables that could further explain these results
    b. We chose to use stoplight-raw-data-2024.csv to incorporate
        i. Copepod richness (number of species; May-Sept)
        ii. PDO (Dec-March)
        iii. PDO (May-Sept) (more important)
        iv. Chinook salmon juvenile catches log(no. km-1; June)
    c. We processed these data using the Pandas package in Python
        i. Limit by colums
        ii. Limit by rows
        iii. Transpose the rows and colums
        iv. Save
    d. We then joined the following dataframes using pandas
        i. meanSST, MaxSST, Chinook SAR – all by year

          ii.   and copepod, pdo dec-march, may-sept, chinook salmon juvenile cachtes

e.   and uploaded them to a SQL database
f.   we then displayed the data
g.   and created graphs of the data in matplot lib to show relationships
h.   we then created the following nested models
i.   and evaluated them using an ANOVA
j.   then using AIC, BIC
k.   showed the results
l.   showed graphs we made in matplot lib
m. Create a graph that compares the raw data