Climate & Salmon:
Assessing the Impact of Critical Habitat Designation and Sea Surface Temperature on Smolt-to-Adult Return Rates for Spring-Run Chinook Salmon in Upper Columbia River Basin

Team Nexus:
Paul J. Anderson
Rachel Hughes

CS 512, Winter 2025
Oregon State University

Justification for dataset use:

Since the difference between a good data scientist and a great data scientist is a deep understanding of context, we chose these datasets to combine the expertise of both Nexus team members. Rachel Hughes is a Meteorologist and Paul Anderson holds a Bachelors Degree in Forestry, Fisheries, and Wildlife and has done research using spatial analysis in wetland and aquatic systems. These data provide a framework to use both climatic and legislative information as predictive variables for biological outcomes.

From the predictive side, we have both SST and pre and post listing categories. In 2004, Spring Run Chinook Salmon Evolutionary Significant Unit (ESU) was listed under the Endangered Species Act (NOAA 2024, August 23). In 2006, the Upper Columbia River Basin was designated critical habitat for this ESU. By evaluating datasets pre and post listing of the ESU and receiving Critical Habitat Designation, we hope to understand how management has influenced this important group of fish. SST is listed as a good predictor in the 2024 NOAA report on salmonid Fisheries production in the West Coast (NOAA 2024, December 6).

For the outputs, we have annual SAR, annual smolts to adult return ratio for the ESU described above. We can use these data to determine if the populations are in a healthy range, if SAR is different between the two time periods, and if SAR is strongly correlated with either predictor variable. We also have annual smolts, and PIT returns at the Bonneville Dam, a relative measure of adult returns to the Upper Columbia Basin. While the SAR data is a powerful estimation of health, there may be questions or analyses which require finer data.

Abstract:

Decreases in fish stocks have social, economic, and environmental impacts throughout the western United States. Tracking changes in fish stocks provides vital data for the conservation and protection of this declining resource. As part of fish stock management, various state and federal agencies monitor oceanic and river conditions to determine catch limits. The National Oceanic and Atmospheric Administration (NOAA) cites eight ocean indicators critical for monitoring the health of salmon fish stocks along the western United States (NOAA, 2024). This study focuses on one of these ocean indicators, sea surface temperatures (SST), and examines its influence on the smolt-to-adult return ratio (SAR) of an important group of salmonids: the Upper Columbia spring-run Chinook salmon. Two time periods, encompassing neutral to mild La Niña activity, are studied—before and after the critical habitat designation for the Upper Columbia spring-run Chinook salmon, from 2002-2004 and 2020-2024. The results may

provide insights into the relationship between SST and SAR, and how management since the species and habitat listing has influenced these factors. Preliminary results indicate no evidence that the mean SAR is different before or after the ESU listing and habitat protection (Student's t-test, p-value = 0.1064). Preliminary results do show some evidence that the mean annual SST was different between the two time periods chosen (Student's t-test, p-value = 0.07) with the mean annual SST for 2002-2004 of 10.5C and 11.1C for 2020-2022. Further study is required to determine if the higher mean SST influenced the SAR results.

Obtain:

Climate Data:

The climate data was obtained from the National Centers for Environmental Information, a sub agency of NOAA. The retrieved SST were interpolated to a 5 km grid from measured SST from buoys and satellite data.

Data were retrieved using the Environmental Research Division Data Access Program (ERDDAP) to generate a url to access the specified data. The data was available in a variety of formats to include comma-separated values (.csv) and JavaScript Object Notation (JSON). The .csv format was chosen for initial access.

Access link for climate data: https://coastwatch.pfeg.noaa.gov/erddap/griddap/NOAA_DHW.html

Data from 01/01/2020 to 12/31/2022 from 25N to 50N, -124W to -120W, sampled every 5 days, with SST reported at 12Z:
https://coastwatch.pfeg.noaa.gov/erddap/griddap/NOAA_DHW.csv?CRW_SST%5B(2020-01-01T12:00:00Z):5:(2022-12-30T12:00:00Z)%5D%5B(25):1:(50)%5D%5B(-124):1:(-120)%5D

Data from 01/01/2002 to 12/31/2004 from 25N to 50N, -124W to -120W, sampled every 5 days, with SST reported at 12Z:
https://coastwatch.pfeg.noaa.gov/erddap/griddap/NOAA_DHW.csv?CRW_SST%5B(2002-01-01T12:00:00Z):5:(2004-12-30T12:00:00Z)%5D%5B(25):1:(50)%5D%5B(-124):1:(-120)%5D

Fish Returns Data:

The fish return data was obtained from the Columbia Basin Research program at the University of Washington. The retrieved mean SAR values were sorted by year and by spring-run Chinook evolutionary significant unit (ESU) of the entire Upper Columbia Basin critical habitat unit.

Data were retrieved using the Data Access In Real Time (DART) tool maintained by the Columbia Basin Research program to generate a url to access the specified data. The data was available in a variety of formats to include comma-separated values (.csv) and html with accompanying .csv. The .csv format was chosen.

Access link for SAR data: https://www.cbr.washington.edu/dart/

Data from 01/01/2001 to 12/31/2025 Bonneville (All) to Bonneville Adult, Chinook Upper Columbia R Spring ESU, All, Exclude 0,1-Year Adult Detections, by Release Basin:
https://www.cbr.washington.edu/dart/query/pit_sar_esu

Data from 01/01/2002 to 12/31/2004 PIT Tag Adult Returns by Observation Site, Bonneville Dam Adult Fishways, 1-Chinook, 1-Spring, All Rear Types:
https://www.cbr.washington.edu/dart/query/pitadult_obsyr_detail

Data from 01/01/2020 to 12/31/2022 PIT Tag Adult Returns by Observation Site, Bonneville Dam Adult Fishways, 1-Chinook, 1-Spring, All Rear Types:
https://www.cbr.washington.edu/dart/query/pitadult_obsyr_detail

Data from 01/01/2002 to 12/31/2004 Yearling Smolt Release, BON, Chinook-1 (yearling):
https://www.cbr.washington.edu/dart/query/
https://www.cbr.washington.edu/dart/query/smolt_graph_text

Data from 01/01/2020 to 12/31/2022 Yearling Smolt Release, BON, Chinook-1 (yearling):
https://www.cbr.washington.edu/dart/query/
https://www.cbr.washington.edu/dart/query/smolt_graph_text

Estimate of Points Complexity:
Non-standard dataset: +3
Multiple files to start: +1
>1 type of related data: +1
Accessed beyond database or file download: +1

The non-standard dataset designation was determined because the climate data was merged with the fish data prior to analysis. Multiple files were used for both the climate data and the fish data. More than one type of related data was included via climate and fish data files. The fish data was accessed beyond a database download. The climate data was accessed by file download.

Scrub:

Climate data were analyzed to remove missing gridded observations. Data were filtered to concentrate between latitude 40N and 50N. Resulting grid-point SST were averaged together for a single representative SST for the date. Two scrubbed datasets were developed for further analysis, one by month and one for every 5-day mean and maximum SST. After an initial take at analysis, a third dataset of yearly data was developed. See Appendix II for sample data.

The pandas data manipulation tool was used for the scrubbing step for the climate data. See Appendix I for scripts.

The conversion from .csv to .json and from .json to .csv was conducted on the scrubbed and synthesized SST climate data, as opposed to the original climate data set that included NaN values and extended across a greater range of latitudes. See Appendix I for scripts.

Fish Return data were analyzed to remove fields not necessary for our data analysis. It was further sorted by year of interest. Additional fish data were uploaded to a personal MySQL database and sorted by year of interest. But due to the limitation of time and issues with access, further work on these data was abandoned.

Mean SST and Mean SAR data were joined using the Pandas package in Python (see Appendix I). The year field was an integer for one dataset and a string for the other. This was rectified in Python. The two dataframes were merged by year. To help visualize these data, the .csv data were converted to a JSON file in records orientation. Confirmation was performed to verify there was no dataloss between the two formats.

Explore:

The average and maximum SST per month were extracted for each year of the study. See Appendix I. The same method was applied for every 5 day mean and maximum.

Mean SAR values for the 2002-2004 and 2020-2024 date period were extracted (see Appendix I).

Scrubbed and processed data were uploaded to R for further analysis with the fish SAR return ratio (see Appendix III).

Model:

The model step in the OSEMN framework does not apply to this portion of the study.

Interpret:

Initial analysis of the data was conducted in R. A summary of initial investigations is included in Appendix III.

To further analyze the data, two null hypotheses were considered. First, the mean of the early SAR samples was compared to the mean of the late SAR samples with the hypothesis of zero difference. Results of a Student's t-test conducted in R indicate that there is no statistically significant difference between the means in the selected time periods prior to critical habitat designation or post critical habitat designation (p-value = 0.1064).

Next, the mean of the early annual SST samples was compared to the mean of the late annual SST samples with the hypothesis of zero difference. Results of a Student's t-test conducted in R indicate that there was evidence that there is a difference in mean SST between the two time periods (p-value = 0.071), though outside the given confidence level of 0.05. The mean SST in the early period was 10.5C, in the late period was 11.1C.

These results suggest more investigation is needed to determine if SST plays a part in the lack of change for SAR before and after critical habitat designation.

Obstacles Encountered in Work:

During the retrieval stage of the process, it was difficult to directly import the climate data from the generated urls due to the formatting from ERDDAP. A work around for this project was found by reducing the volume of data to analyze and then downloading the data directly in two .csv files. The volume of data was reduced by selecting every 5 days on the interpolated SST, along with the identifying date and latitude/longitude. Similarly, fish data was found in large unrelated datasets. This was delt with by filtering the data by mean SAR by year, focusing on a single ESU, Spring Run Chinook of the Upper Columbia Basin.

Distribution of Work:

The data retrieval was split between the two team members with Rachel retrieving and scrubbing the climate data while Paul retrieved and scrubbed the fish data. Paul worked on setting up a joint MySQL database through Oregon State University College of Engineering. We ran into permissions issues, and that part of the project is sidelined until later but consumed a lot of time. Both partners worked on the final report from the draft that Rachel wrote. Paul took point on the analysis and merging the final processed data, Rachel finished the hypothesis testing.

References:

NOAA Fisheries. (2024, August 23). *Upper Columbia River Spring-run Chinook Salmon*. Endangered Species Conservation, NOAA Fisheries. Accessed 24 January 2025. https://tinyurl.com/y3rnr7kh

NOAA Fisheries. (2024, October 28). *Salmon and Steelhead Research in the Pacific Northwest.* Science & Data, NOAA Fisheries. Accessed 24 January 2025. https://tinyurl.com/y4jhxrzp

NOAA Fisheries. (2024, December 6). *2024 Summary of Ocean Ecosystem Indicators.* Science & Data, NOAA Fisheries. Accessed 24 January 2025. https://tinyurl.com/29wjp8za

NOAA Fisheries. (2024, March 19). Oceanography of the Northern California Current Study Area. West Coast, NOAA Fisheries. Accessed 24 January 2025. https://tinyurl.com/yw6pj5tx

Coding Sources:

Bobbitt, Z. (2022, March 31). *How to Convert Pandas GroupBy Output to DataFrame*. Statology. Accessed 27 January 2025. https://www.statology.org/pandas-groupby-to-dataframe/.

Datetime – Basic date and time types. (n.d.). Python Standard Library. Accessed 26 January 2025. https://docs.python.org/3/library/datetime.html

Ebahrim, M. (2023, December 11). *Convert CSV to JSON using Python Pandas (Easy Tutorial)*. Like Geeks. Accessed 26 January 2025. https://likegeeks.com/csv-to-json-python-pandas/

Nelamali, N. (2024, October 31). Python Pandas: Convert JSON to CSV. Spark by Examples. Acessed 26 January 2025. https://sparkbyexamples.com/pandas/python-pandas-convert-json-to-csv/

Pandas API Reference. (n.d.). Pandas. Accessed 26 January 2025. https://pandas.pydata.org/docs/reference/index.html

Pandas Read CSV in Python. (2024, November 21). Geeks for Geeks. Accessed 26 January 2025. https://www.geeksforgeeks.org/python-read-csv-using-pandas-read_csv/.

Pykes, K. (2024, December 2). *Pandas read_csv() Tutorial: Importing Data*. Datacamp. Accessed 26 January 2025. https://www.datacamp.com/tutorial/pandas-read-csv

APPENDIX I

Code Samples

CLIM_data_process.py

```
#CS 512
#Data Wrangling Homework #4
#Team NEXUS

import pandas as pd

#read in data in .csv files for 2020-22
c20_22 = pd.read_csv('SST_2020_22.csv', skiprows = [1,1])
#drop the NaN values
c20_22.dropna(inplace = True)
#select only for latitude > 40 (data 25N < lat < 50N)
c20_22 = c20_22[c20_22['latitude'] > 40]
#remove lat/lon columns
c20_22.drop(columns = ['latitude', 'longitude'], inplace = True)
#set date-time to pandas datetime type, extract year & month
c20_22['time'] = pd.to_datetime(c20_22['time'])
c20_22['year'] = c20_22['time'].dt.year
c20_22['month'] = c20_22['time'].dt.month

#repeat operations for .csv file for 2002-04
c02_04 = pd.read_csv('SST_2002_04.csv', skiprows = [1,1])
c02_04.dropna(inplace = True)
c02_04 = c02_04[c02_04['latitude'] > 40]
c02_04.drop(columns = ['latitude', 'longitude'], inplace = True)
c02_04['time'] = pd.to_datetime(c02_04['time'])
c02_04['year'] = c02_04['time'].dt.year
c02_04['month'] = c02_04['time'].dt.month

#calculate mean & max SST by month for each time period
#use reset.index because result of groupby() is not a dataframe
ave_SST_20_22 = c20_22.groupby(['year', 'month'])['CRW_SST'].mean().round(2)
ave_SST_20_22 = ave_SST_20_22.reset_index()
max_SST_20_22 = c20_22.groupby(['year', 'month'])['CRW_SST'].max().round(2)
max_SST_20_22 = max_SST_20_22.reset_index()

ave_SST_02_04 = c02_04.groupby(['year', 'month'])['CRW_SST'].mean().round(2)
ave_SST_02_04 = ave_SST_02_04.reset_index()
max_SST_02_04 = c02_04.groupby(['year', 'month'])['CRW_SST'].max().round(2)
max_SST_02_04 = max_SST_02_04.reset_index()

#merge mean and max SST files for each time period
SST_02_04 = pd.merge(ave_SST_02_04, max_SST_02_04, on = ['year', 'month'])
SST_02_04.rename({'CRW_SST_x': 'Mean_SST', 'CRW_SST_y': 'Max_SST'}, axis = 1, inplace = True)
```

```
SST_20_22 = pd.merge(ave_SST_20_22, max_SST_20_22, on = ['year', 'month'])
SST_20_22.rename({'CRW_SST_x': 'Mean_SST', 'CRW_SST_y': 'Max_SST'}, axis = 1, inplace = True)

#write SST mean & max to file for each time period consecutively
SST_02_04.to_csv('SST_stats.csv', index = False)
SST_20_22.to_csv('SST_stats.csv', mode = 'a', index = False, header = False)
```

Fish + CLIM Data Processing (links to gists below)

scrubbing_1_SAR

```
# this code processes a raw csv file > year and mean SAR csv file

import os
import pandas as pd

# change the working directory
os.chdir('c:/Users/ander/OneDrive - Oregon State University/Classes/2025/Winter/CS512/Nexus/Upper_Columbia_Spring_Chinook_ESU_Data/PIT_SAR_Upper_Columbia_Chinook_ESU/')

# load the CSV file
df = pd.read_csv('raw_SAR.csv')

# columns to remove
columns_to_remove = ['lowerCI', 'upperCI', 'meanJuvArrivalDay', 'juvCount', 'medianJuvArrivalDay', 'SE', 'juvMinArrivalDay', 'juvMaxArrivalDate', 'adultCount', 'juv5%ArrivalDay', 'juv95%ArrivalDay', 'ocean1Count', 'ocean2Count', 'ocean3Count']

# remove the columns
df.drop(columns=columns_to_remove, inplace=True)

# save the file
df.to_csv('year_SAR.csv', index=False)
```

scrubbing_2_SAR

```
# this code filters the rows by year to include only 2002-2004 and 2020-2022

import os
import pandas as pd

# change the working directory
os.chdir('c:/Users/ander/OneDrive - Oregon State University/Classes/2025/Winter/CS512/Nexus/Upper_Columbia_Spring_Chinook_ESU_Data/PIT_SAR_Upper_Columbia_Chinook_ESU/')

# load the csv file
df = pd.read_csv('year_SAR.csv')

# filter the rows for the years 2002-2004 and 2020-2022
filtered_df = df[df['year'].isin(['2002', '2003', '2004', '2020', '2021', '2022'])]

# save the filtered data to a new csv file
filtered_df.to_csv('filtered_year_SAR.csv', index=False)
```

scrubbing_3_SAR_SST

```
import pandas as pd
import os

# change the working directory
os.chdir(r'c:/Users/ander/OneDrive - Oregon State University/Classes/2025/Winter/CS512/Nexus/Upper_Columbia_Spring_Chinook_ESU_Data/PIT_SAR_Upper_Columbia_Chinook_ESU/')

# load the first dataset from CSV
df1 = pd.read_csv('SST_yearly.csv')

# load the second dataset from CSV
df2 = pd.read_csv('year_SAR.csv')

# convert 'year' column to integer type in both dataframes
df1['year'] = pd.to_numeric(df1['year'], errors='coerce')
df2['year'] = pd.to_numeric(df2['year'], errors='coerce')

# merge the dataframes on the 'year' column
merged_df = pd.merge(df1, df2, on='year')

# output the merged dataframe to a new CSV file
merged_df.to_csv('fish_SST.csv', index=False)
```

scrubbing_4_csv_JSON

```
import os
import pandas as pd

# change the working directory
os.chdir('c:/Users/ander/OneDrive - Oregon State University/Classes/2025/Winter/CS512/Nexus/Upper_Columbia_Spring_Chinook_ESU_Data/PIT_SAR_Upper_Columbia_Chinook_ESU/')

# load the CSV file into a DataFrame
df = pd.read_csv('fish_SST.csv')

# convert the DataFrame to a JSON format (with indentation for readability)
df.to_json('fish_SST.json', orient='records', lines=False, indent=4)
```

csvtojson.py

```
#CS 512
#Data Wrangling Homework #4
#Team NEXUS
#Extract processed .csv SST data and convert to json

import pandas as pd

# Read in processed .csv file.
SST_csv = pd.read_csv('SST_stats.csv')
```

```
# Convert .csv file to .json, orienting on records and
# having each record on a separate line.
SST_json = SST_csv.to_json(orient = 'records', lines = True)

# Write converted .json data to output file.
with open('SST.json', 'w') as outfile:
outfile.write(SST_json)
```

jsontocsv.py

```
#CS 512
#Data Wrangling Homework #4
#Team NEXUS
#Extract processed .json SST data and convert to .csv

import pandas as pd

# Read in the .json file, using the orient 'records' option.
# Also using the lines = True option because each record in .json
# file is a separate line.
SST_json = pd.read_json('SST.json', orient = 'records', lines = True)

# Convert to .csv format and write to .csv file at same time.
SST_json.to_csv('SST.csv', index = False)
```

SST_fish_merge.py

```
#CS 512
#Data Wrangling Homework #4
#Team NEXUS

#Merge yearly mean and max sst with yearly SAR

import pandas as pd

#read in data in .csv files for SST and fish
fish = pd.read_csv('filtered_year_SAR.csv')
SST = pd.read_csv('SST_yearly.csv')

#merge the fish and SST
SST_fish = pd.merge(fish, SST, on = ['year'])

#write merged data to .csv file
SST_fish.to_csv('SST_fish.csv', index = False)
```

APPENDIX II

Sample of data from original climate data pull:

time,latitude,longitude,CRW_SST UTC,degrees_north,degrees_east,Celsius

2002-01-01T12:00:00Z,50.025,-123.975,NaN 2002-01-01T12:00:00Z,50.025,-123.925,6.98 2002-01-01T12:00:00Z,50.025,-123.875,6.98 2002-01-01T12:00:00Z,50.025,-123.825,6.99 2002-01-01T12:00:00Z,50.025,-123.775,6.99 2002-01-01T12:00:00Z,50.025,-123.725,NaN 2002-01-01T12:00:00Z,50.025,-123.675,NaN 2002-01-01T12:00:00Z,50.025,-123.625,NaN 2002-01-01T12:00:00Z,50.025,-123.575,NaN 2002-01-01T12:00:00Z,50.025,-123.525,NaN 2002-01-01T12:00:00Z,50.025,-123.475,NaN 2002-01-01T12:00:00Z,50.025,-123.425,NaN 2002-01-01T12:00:00Z,50.025,-123.375,NaN 2002-01-01T12:00:00Z,50.025,-123.325,NaN 2002-01-01T12:00:00Z,50.025,-123.275,NaN 2002-01-01T12:00:00Z,50.025,-123.225,NaN 2002-01-01T12:00:00Z,50.025,-123.175,NaN 2002-01-01T12:00:00Z,50.025,-123.125,NaN 2002-01-01T12:00:00Z,50.025,-123.075,NaN 2002-01-01T12:00:00Z,50.025,-123.025,NaN 2002-01-01T12:00:00Z,50.025,-122.975,NaN 2002-01-01T12:00:00Z,50.025,-122.925,NaN 2002-01-01T12:00:00Z,50.025,-122.875,NaN 2002-01-01T12:00:00Z,50.025,-122.825,NaN 2002-01-01T12:00:00Z,50.025,-122.775,NaN 2002-01-01T12:00:00Z,50.025,-122.725,NaN 2002-01-01T12:00:00Z,50.025,-122.675,NaN 2002-01-01T12:00:00Z,50.025,-122.625,NaN 2002-01-01T12:00:00Z,50.025,-122.575,NaN 2002-01-01T12:00:00Z,50.025,-122.525,NaN 2002-01-01T12:00:00Z,50.025,-122.475,NaN 2002-01-01T12:00:00Z,50.025,-122.425,NaN 2002-01-01T12:00:00Z,50.025,-122.375,NaN 2002-01-01T12:00:00Z,50.025,-122.325,NaN 2002-01-01T12:00:00Z,50.025,-122.275,NaN 2002-01-01T12:00:00Z,50.025,-122.225,NaN 2002-01-01T12:00:00Z,50.025,-122.175,NaN 2002-01-01T12:00:00Z,50.025,-122.125,NaN 2002-01-01T12:00:00Z,50.025,-122.075,NaN 2002-01-01T12:00:00Z,50.025,-122.025,NaN 2002-01-01T12:00:00Z,50.025,-121.975,NaN 2002-01-01T12:00:00Z,50.025,-121.925,NaN 2002-01-01T12:00:00Z,50.025,-121.875,NaN 2002-01-01T12:00:00Z,50.025,-121.825,NaN 2002-01-01T12:00:00Z,50.025,-121.775,NaN 2002-01-01T12:00:00Z,50.025,-121.725,NaN 2002-01-01T12:00:00Z,50.025,-121.675,NaN 2002-01-01T12:00:00Z,50.025,-121.625,NaN 2002-01-01T12:00:00Z,50.025,-121.575,NaN 2002-01-01T12:00:00Z,50.025,-121.525,NaN 2002-01-01T12:00:00Z,50.025,-121.475,NaN 2002-01-01T12:00:00Z,50.025,-121.425,NaN 2002-01-01T12:00:00Z,50.025,-121.375,NaN 2002-01-01T12:00:00Z,50.025,-121.325,NaN 2002-01-01T12:00:00Z,50.025,-121.275,NaN 2002-01-01T12:00:00Z,50.025,-121.225,NaN 2002-01-01T12:00:00Z,50.025,-121.175,NaN 2002-01-01T12:00:00Z,50.025,-121.125,NaN 2002-01-01T12:00:00Z,50.025,-121.075,NaN 2002-01-01T12:00:00Z,50.025,-121.025,NaN 2002-01-01T12:00:00Z,50.025,-120.975,NaN 2002-01-01T12:00:00Z,50.025,-120.925,NaN 2002-01-01T12:00:00Z,50.025,-120.875,NaN 2002-01-01T12:00:00Z,50.025,-120.825,NaN 2002-01-01T12:00:00Z,50.025,-120.775,NaN 2002-01-01T12:00:00Z,50.025,-120.725,NaN 2002-01-01T12:00:00Z,50.025,-120.675,NaN 2002-01-01T12:00:00Z,50.025,-120.625,NaN 2002-01-01T12:00:00Z,50.025,-120.575,NaN 2002-01-01T12:00:00Z,50.025,-120.525,NaN 2002-01-01T12:00:00Z,50.025,-120.475,NaN 2002-01-01T12:00:00Z,50.025,-120.425,NaN 2002-01-01T12:00:00Z,50.025,-120.375,NaN 2002-01-01T12:00:00Z,50.025,-120.325,NaN 2002-01-01T12:00:00Z,50.025,-120.275,NaN 2002-01-01T12:00:00Z,50.025,-120.225,NaN 2002-01-01T12:00:00Z,50.025,-120.175,NaN 2002-01-01T12:00:00Z,50.025,-120.125,NaN 2002-01-01T12:00:00Z,50.025,-120.075,NaN 2002-01-01T12:00:00Z,50.025,-120.025,NaN 2002-01-01T12:00:00Z,50.025,-119.975,NaN 2002-01-01T12:00:00Z,49.975,-123.975,6.96 2002-01-01T12:00:00Z,49.975,-123.925,6.97 2002-01-01T12:00:00Z,49.975,-123.875,6.97 2002-01-01T12:00:00Z,49.975,-123.825,6.97 2002-01-01T12:00:00Z,49.975,-123.775,NaN 2002-01-01T12:00:00Z,49.975,-123.725,NaN 2002-01-01T12:00:00Z,49.975,-123.675,NaN

Sample of data from original fish data pull:

year,meanSAR,lowerCI,upperCI,meanJuvArrivalDay,juvCount,medianJuvArrivalDay,SE,juvMinArrivalDay,juvMaxArrivalDate,adultCount,juv5%ArrivalDay,juv95%ArrivalDay,ocean1Count,ocean2Count,ocean3Count
2001,NA,NA,NA,153,524,153,NA,139,180,2,145,163,0,2,0
2002,1.34486071085495,0.645132768959857,2.04458865275004,140,1041,140,0.357004051987291,119,170,14,129,152,1,13,0

2003,0.689145951267536,0.434763807433569,0.943528095101503,146,4063,146,0.129786808078555,121,184,28,129,165,13,13,2
2004,0.707130229817325,0.308450749476715,1.10580971015793,145,1697,145,0.203407898132964,122,172,12,135,158,6,6,0
2005,NA,NA,NA,138,212,137,NA,128,167,2,130,154,0,2,0
2006,NA,NA,NA,138,626,137,NA,120,336,3,127,149,0,3,0
2007,0.999231360491929,0.458763673087431,1.53969904789643,145,1301,145,0.275748820104336,47,201,13,129,162,3,10,0
2008,2.65060240963855,1.75830357854525,3.54290124073186,150,1245,149,0.455254505659849,56,212,33,132,175,3,22,8
2009,2.46991766941102,1.70436418014349,3.23547115867855,149,1579,147,0.390588514932413,109,347,39,133,176,3,33,3
2010,2.74590163934426,2.28739819291581,3.20440508577272,146,4880,145,0.233930329810437,89,224,134,130,163,69,55,10
2011,NA,NA,NA,139,619,136,NA,100,206,6,129,161,0,5,1
2012,2.02312138728324,1.46240690635338,2.58383586821309,144,2422,144,0.286078816800947,120,187,49,132,157,7,42,0
2013,2.17021276595745,1.5810858683797,2.75933966353519,141,2350,139,0.300574947743749,112,332,51,130,158,15,32,4
2014,1.04468949506674,0.564595140197823,1.52478384993566,141,1723,142,0.244946099422919,110,200,18,128,157,6,11,1
2015,1.23279816513761,0.866596598530811,1.59899973174442,140,3488,140,0.186837533983063,111,183,43,130,152,11,31,1
2016,0.904736562001064,0.601999771504964,1.20747335249716,132,3758,132,0.15445754617148,112,355,34,122,143,8,25,1
2017,1.43668559973271,1.01036017649906,1.86301102296636,139,2993,140,0.217512971037575,112,189,43,126,152,13,28,2
2018,0.718735026353618,0.356314403152454,1.08115564955478,136,2087,135,0.184908481225083,113,183,15,126,151,3,11,1
2019,0.884433962264151,0.569346017161777,1.19952190736653,141,3392,141,0.160759155664477,120,202,30,128,160,4,24,2
2020,1.98159943382873,1.46775490926537,2.4954439583921,137,2826,138,0.26216557375682,115,195,56,126,150,15,41,0
2021,1.57187767933695,1.1597297960136,1.98402556266031,139,3499,138,0.210279532307834,50,345,55,127,155,14,40,1
2022,1.17227319062181,0.695995335292672,1.64855104595096,142,1962,143,0.242998905780175,47,192,23,131,156,8,15,0
2023,0.453367875647668,0.216418219239384,0.690317532055953,139,3088,137,0.120892681840962,121,189,14,131,154,14,0,NA
2024,NA,NA,NA,140,2151,140,NA,53,188,0,129,154,0,NA,NA
2025,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
Notes:
Columbia River DART
Smolt-to-Adult Return (SAR) Survival Bonneville (Juvenile) to Bonneville (Adult)
PIT-Tagged Upper Columbia River Spring Chinook ESU (All)
Adult 0-Year Detections Excluded
As of 2024-03-22; a minimum requirement of 7 adults detected within each pooled period for SAR Survival calculations has been introduced.
Generated 25 Jan 2025 11:29:42 PST. DART PIT Tag Smolt-to-Adult Return (SAR) Survival www.cbr.washington.edu/dart/query/pit_sar_esu.

"DART Data Citation. Columbia River DART, Columbia Basin Research, University of Washington. (2025). PIT Tag Smolt-to-Adult Return (SAR) Survival. Available from https://www.cbr.washington.edu/dart/query/pit_sar_esu."

Processed data from file SST_stats.csv:

year,month,Mean_SST,Max_SST
2002,1,7.52,10.43
2002,2,7.2,9.18
2002,3,7.3,9.09
2002,4,8.16,10.62
2002,5,10.03,14.55
2002,6,12.46,17.18
2002,7,14.41,19.11
2002,8,13.91,18.43
2002,9,12.74,16.76
2002,10,10.44,13.05
2002,11,9.24,12.37
2002,12,8.69,12.02
2003,1,8.01,11.18
2003,2,7.75,10.51
2003,3,8.1,10.29
2003,4,8.82,11.08
2003,5,11.07,14.59
2003,6,13.04,17.69
2003,7,14.64,19.1
2003,8,14.56,19.05
2003,9,13.14,16.42
2003,10,11.32,14.66
2003,11,9.06,12.57
2003,12,8.26,10.81
2004,1,7.42,10.38
2004,2,7.37,10.27
2004,3,7.97,10.81
2004,4,9.74,13.17
2004,5,11.8,16.26
2004,6,13.27,18.28
2004,7,14.65,19.95
2004,8,14.53,18.83
2004,9,13.06,17.49

Processed data from file SST.json:

{"year":2002,"month":1,"Mean_SST":7.52,"Max_SST":10.43} {"year":2002,"month":2,"Mean_SST":7.2,"Max_SST":9.18}
{"year":2002,"month":3,"Mean_SST":7.3,"Max_SST":9.09} {"year":2002,"month":4,"Mean_SST":8.16,"Max_SST":10.62}
{"year":2002,"month":5,"Mean_SST":10.03,"Max_SST":14.55}
{"year":2002,"month":6,"Mean_SST":12.46,"Max_SST":17.18}
{"year":2002,"month":7,"Mean_SST":14.41,"Max_SST":19.11}
{"year":2002,"month":8,"Mean_SST":13.91,"Max_SST":18.43}
{"year":2002,"month":9,"Mean_SST":12.74,"Max_SST":16.76}
{"year":2002,"month":10,"Mean_SST":10.44,"Max_SST":13.05}
{"year":2002,"month":11,"Mean_SST":9.24,"Max_SST":12.37}
{"year":2002,"month":12,"Mean_SST":8.69,"Max_SST":12.02} {"year":2003,"month":1,"Mean_SST":8.01,"Max_SST":11.18}
{"year":2003,"month":2,"Mean_SST":7.75,"Max_SST":10.51} {"year":2003,"month":3,"Mean_SST":8.1,"Max_SST":10.29}
{"year":2003,"month":4,"Mean_SST":8.82,"Max_SST":11.08} {"year":2003,"month":5,"Mean_SST":11.07,"Max_SST":14.59}

{"year":2003,"month":6,"Mean_SST":13.04,"Max_SST":17.69} {"year":2003,"month":7,"Mean_SST":14.64,"Max_SST":19.1}
{"year":2003,"month":8,"Mean_SST":14.56,"Max_SST":19.05}
{"year":2003,"month":9,"Mean_SST":13.14,"Max_SST":16.42}
{"year":2003,"month":10,"Mean_SST":11.32,"Max_SST":14.66}
{"year":2003,"month":11,"Mean_SST":9.06,"Max_SST":12.57}
{"year":2003,"month":12,"Mean_SST":8.26,"Max_SST":10.81} {"year":2004,"month":1,"Mean_SST":7.42,"Max_SST":10.38}
{"year":2004,"month":2,"Mean_SST":7.37,"Max_SST":10.27} {"year":2004,"month":3,"Mean_SST":7.97,"Max_SST":10.81}
{"year":2004,"month":4,"Mean_SST":9.74,"Max_SST":13.17} {"year":2004,"month":5,"Mean_SST":11.8,"Max_SST":16.26}
{"year":2004,"month":6,"Mean_SST":13.27,"Max_SST":18.28}
{"year":2004,"month":7,"Mean_SST":14.65,"Max_SST":19.95}
{"year":2004,"month":8,"Mean_SST":14.53,"Max_SST":18.83}
{"year":2004,"month":9,"Mean_SST":13.06,"Max_SST":17.49}

Processed data from files SST.csv after converting SST.json back to .csv:

```
year,month,Mean_SST,Max_SST
2002,1,7.52,10.43
2002,2,7.2,9.18
2002,3,7.3,9.09
2002,4,8.16,10.62
2002,5,10.03,14.55
2002,6,12.46,17.18
2002,7,14.41,19.11
2002,8,13.91,18.43
2002,9,12.74,16.76
2002,10,10.44,13.05
2002,11,9.24,12.37
2002,12,8.69,12.02
2003,1,8.01,11.18
2003,2,7.75,10.51
2003,3,8.1,10.29
2003,4,8.82,11.08
2003,5,11.07,14.59
2003,6,13.04,17.69
2003,7,14.64,19.1
2003,8,14.56,19.05
2003,9,13.14,16.42
2003,10,11.32,14.66
2003,11,9.06,12.57
2003,12,8.26,10.81
2004,1,7.42,10.38
2004,2,7.37,10.27
2004,3,7.97,10.81
2004,4,9.74,13.17
2004,5,11.8,16.26
2004,6,13.27,18.28
2004,7,14.65,19.95
2004,8,14.53,18.83
2004,9,13.06,17.49
```

Fish + CLIM File Output (links to gists below)

fish_SST.csv

```
year,Mean_SST,Max_SST,meanSAR
2002,10.18,19.11,1.344860711
2003,10.66,19.1,0.689145951
2004,10.73,19.95,0.70713023
2020,11.09,21.28,1.981599434
2021,11.21,23.47,1.571877679
2022,10.87,23.16,1.172273191
```

fish_SST.json


```
[
  {
    "year":2002,
    "Mean_SST":10.18,
    "Max_SST":19.11,
    "meanSAR":1.344860711
  },
  {
    "year":2003,
    "Mean_SST":10.66,
    "Max_SST":19.1,
    "meanSAR":0.689145951
  },
  {
    "year":2004,
    "Mean_SST":10.73,
    "Max_SST":19.95,
    "meanSAR":0.70713023
  },
  {
    "year":2020,
    "Mean_SST":11.09,
    "Max_SST":21.28,
    "meanSAR":1.981599434
  },
  {
    "year":2021,
    "Mean_SST":11.21,
    "Max_SST":23.47,
    "meanSAR":1.571877679
  },
  {
    "year":2022,
    "Mean_SST":10.87,
    "Max_SST":23.16,
    "meanSAR":1.172273191
  }
]
```

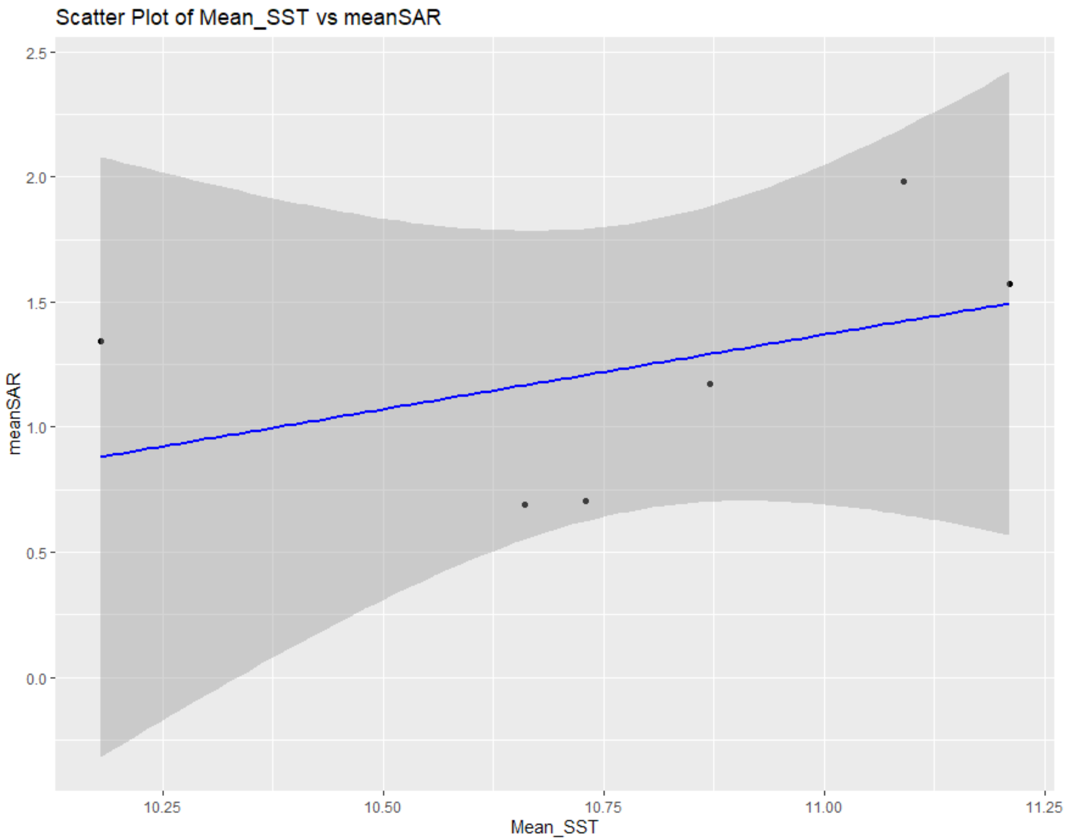
APPENDIX III

Proof of loading data into analysis tool R:

Question 1: Does Mean_SST significantly explain the variability in meanSAR?

R code
Summary of what the code does: The link above takes the reader to a gist, the code loads the data into R, visualizes the data, calculate correlations, fits a linear model to understand the relationship better.
Outputs:

Graph of meanSAR by Mean_SST showing no strong relationship.



Output from the Coefficients, and fit of the linear model:
Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-5.1781	6.7064	-0.772	0.483
Mean_SST	0.5952	0.6212	0.958	0.392

Residual standard error: 0.5068 on 4 degrees of freedom
Multiple R-squared: 0.1867, Adjusted R-squared: -0.01667
F-statistic: 0.918 on 1 and 4 DF, p-value: 0.3923

These data show no evidence (p-value > 0.1) to reject the null hypothesis that meanSAR and

Mean_SST are unrelated, suggesting the Mean_SST does not significantly explain the variability in meanSAR.

Question 2: Is there a significant difference between mean SARs from before listing with the ESA and after listing?

Results from the code below are summarized in the “Interpret” section of the report above.

```
library(tidyverse)

#read in SST & fish data
 SST_fish <- read_csv("SST_fish.csv")

# group by early/late periods and find mean/sd
 SST_fish <- SST_fish |>
   mutate(
     group = case_when(
     year >= 2002 & year <= 2004 ~ "early",
     year >= 2020 & year <= 2022 ~ "late"
     )
   )

SST_fish|>
   group_by(group) |>
   summarise(
     mean_SAR = mean(meanSAR),
     sd_SAR = sd(meanSAR),
     mean_SST = mean(Mean_SST),
     sd_SST = sd(Mean_SST)
     )

# A tibble: 2 × 5
   group mean_SAR sd_SAR mean_SST sd_SST
   <chr>    <dbl>  <dbl>    <dbl>  <dbl>
 1 early    0.914  0.373     10.5  0.299
 2 late     1.58   0.405     11.1  0.172

x = subset(SST_fish, group == "early")
 y = subset(SST_fish, group == "late")

#test hypothesis
 #H_0: mean_SAR_early - mean_SAR_late = 0
 t.test(
   x$meanSAR,
  y$meanSAR,
  var.equal = FALSE
   )
```

```
    Welch Two Sample t-test

data:  x$meanSAR and y$meanSAR
 t = -2.0807, df = 3.9746, p-value = 0.1064
 alternative hypothesis: true difference in means is not equal to 0
 95 percent confidence interval:
  -1.5465150  0.2234394
 sample estimates:
 mean of x mean of y
0.9137123 1.5752501

#test hypothesis
 #H_0: mean_SST_early - mean_SST_late = 0
 t.test(
   x$Mean_SST,
  y$Mean_SST,
  var.equal = FALSE
   )



    Welch Two Sample t-test

data:  x$Mean_SST and y$Mean_SST
 t = -2.6737, df = 3.1954, p-value = 0.07053
 alternative hypothesis: true difference in means is not equal to 0
 95 percent confidence interval:
  -1.14674018  0.08007351
 sample estimates:
 mean of x mean of y
 10.52333  11.05667
```