For the final project, I performed EDA on catch statistics for 20 ICES countries in the North Atlantic region for over 200 shellfish and fish species (ICES, n.d.). This paper will summarize the hypothesis, the outcome of the EDA, some things that I thought were missing and challenges or concepts I did not fully grasp.

The hypothesis that I had developed from this data was to see if France had a lower salmon weight reported for 2017 than Norway (Schuitemaker, 2017). I also wanted to see if we could develop a model to determine in what Northern Atlantic area to fish to catch a specific species of fish.

From the data I was able to determine that France had a lower reported weight for salmon than other species France caught and a lower weight than what Norway reported for 2017. This indicates that France exports many of the species caught, the species are shellfish or France imports a lot of salmon to consume. I was also able to show that the distribution for the species reported by France in 2017 is not well represented by a normal distribution even though the CDF looks like a straight line. I would have like to have spent more time trying to find a better distribution for this data. I also created PMF’s for Norway and France’s reported species in 2017 to compare. I was able to perform a correlation between Area\_num and Species\_num and obtained a p-value of 0.0. This means that the null hypothesis, there is not correlation between Area\_num and Species\_num, was rejected. Finally, I was able to create a model, Area\_num = 0.0032\*Species\_num + 69.5928, to determine what Area\_num to fish to catch the desired Species\_num.

I would have liked to have data on the fish consumption information to see how the catch statistics related to the consumption of the species for the different countries. This would have helped to see if France consumed more fish than they caught. I was able to show that salmon was not the highest reported fish and if the article was correct, France should have had lower numbers for salmon if the prefer imported from Norway and Scotland (Schuitemaker, 2017). I would have also obtained the import and export statistics for the 20 ICES countries listed to see what is happening with the species caught.

Challenges that were observed during this analysis involved working with variable names that were numbers; example 2017. This caused issues when trying to reference the columns in the data set for functions with the dot notation. I worked around this by utilizing the bracket notation. I also ran into issues with displaying the data for the weight histograms and feel like this could have been solved with creating bins, but everything I tried did not produce the graph.

There was a lot of information gathered from this analysis, but I do feel like more information is needed to determine if France prefers salmon from Norway and Scotland like the article claims. I was not able to disprove the claim so there is still the possibility that France prefers salmon from outside countries. Scotland was not part of the 20 ICES countries making it impossible with this data to compare Norway, Scotland and France for this article found.

# Works Cited

ICES. (n.d.). *Catch statistics*. Retrieved from ICES.dk: https://www.ices.dk/marine-data/dataset-collections/Pages/Fish-catch-and-stock-assessment.aspx

Schuitemaker, L. (2017, October 23). *Norwegian salmon most consumed fish in France*. Retrieved from Salmon Business: https://salmonbusiness.com/norwegian-salmon-most-consumed-fish-in-france/