Analysing the Fish Hunting Numbers in North Pacific Ocean*

Using Baysian Modeling, to find that number of fish caught has increased

Shreya Sakura Noskor

November 26, 2024

Abstract

We analyzed data on fish catches in the North Pacific Ocean using Bayesian modeling. Our analysis shows that the number of fish caught has increased over time. This suggests that fishing activities in the region have intensified. Understanding this trend is important for managing fish populations and ensuring the sustainability of the ocean's resources. This is important as the Pacific is full of different kind of fish like, trout and salmon that are essential in the food chain, but are being overhunted for food and recreational purposes.

Table of contents

1 Introduction

Estimand is the number of fishes that were caught by each country, for every year. Or more specifically what the rate was and if they are more likely to be fished for commercial purposes.

*Code and data are available at: https://github.com/NotSakura/FisheriesData.git.

2 Data

2.1 Overview

The data was downloaded from (NPAFC) (2024) and was cleaned using R (R Core Team 2023). The data was read using Schauberger and Walker (2024) and Wickham and Bryan (2023), while the data was cleaned using Wickham et al. (2019), Wickham et al. (2019), Firke (2023), Wickham et al. (2023), Xie (2023). The data was modeled using Arel-Bundock (2022), Robinson, Hayes, and Couch (2023), Goodrich et al. (2022), and Bürkner (2017).

To download the data go to NPFAC's official data portal and look for "NPAFC Catch Statistics (updated 28 June 2024)". Click on that link to get the csv file containing all the data.

2.2 Methodology and Measurement

NPAFC has this data to download from their website. The way they gathered this data was that they are an inter-government organisation so they have access to government data based on how much fish were hunted in the respective countries. The countries in the data include Canada, Russia, Korea, Japan and Unites States of America. The way each of these countries measured this data was that when fish are being caught on international waters and report it to each other. This is strictly enforced, especially after the fall of salmon and trout population in the Pacific, majorly due to environment purposes.

This paper look at multiple variables. First, variable we look at is Country which are either "Canada", "Russia", "Korea", "Japan", and "United States", all representing the countries that are members of this organization. This data was left unchanged, Next variables we look at is "Whole Country/Province/State", where the instances of these variables may either be Whole country, or the different states or provinces that was fishing and gathered that data. So for example, if the value was British Columbia then the corresponding number of fishes caught reported is the number of fish caught by the province. Throughout our data we filtered for the "Whole country" value, assuming that the numbers in each province and/or state would add up to the number in "Whole country" (which it did). This was because we were more interested in comparing the fishing trends between countries rather then within a country. The next variable is "Reporting Area" which accounts for where the fishes were caught. This was also filtered by "Whole country" due to the previous reasoning. The next variable that we filtered was "Species" which contained "Cherry", "Chinook", "Chum", "Coho", "Pink", "Sockeye", "Steelhead" and "Total". These are all types of salmon except for Steelhead which is a trout and "Total" which represents all the fishes that were hunted. Although there is a lot of interesting information to uncover if we did a deeper analysis on each fish, but, we decided that the best way to compare the fishing trends between countries would be to just look at the total fishes caught. The next variable we used and actually analyse is the "catch type". This tells us whether is the fishes were caught for commercial purposes (caught for profit purposes like selling), sporting purposes (which means they were caught recreationaly) or subsistence purposes (which means they were caught to provide food, not as a profit). The last variable we filtered was "Data Type" which was the unit that these numbers were reported in. There was Numbers in 1000s or Round weight in metric tonne. We chose to filter with numbers in 1000s as the other option was done only by the US, who provided both units.

Thoese were the variables that we filtered or focused on but there is one more variable that is important for this paper; the number of fishes caught. That number in the raw data was provided as the year as a column and the corresponding value as the number of fishes caught. This format meant that when we are creating models or graphs it is very difficult to work with it. And so we actually shifted, rather pivotted the table so that analysis is easier. To pivot the table we created 2 new rows to the dataset, "Year" and "Catch". The year corresponds to the column's title which is the year this data is for and the "catch" refers to the number of fishes that were caught in that year, in that country. This helped significantly with making the models and such.

2.3 Data Visualization

we make the assumption that the coloumns that say whole country it also include the provinces and different areas number as well.

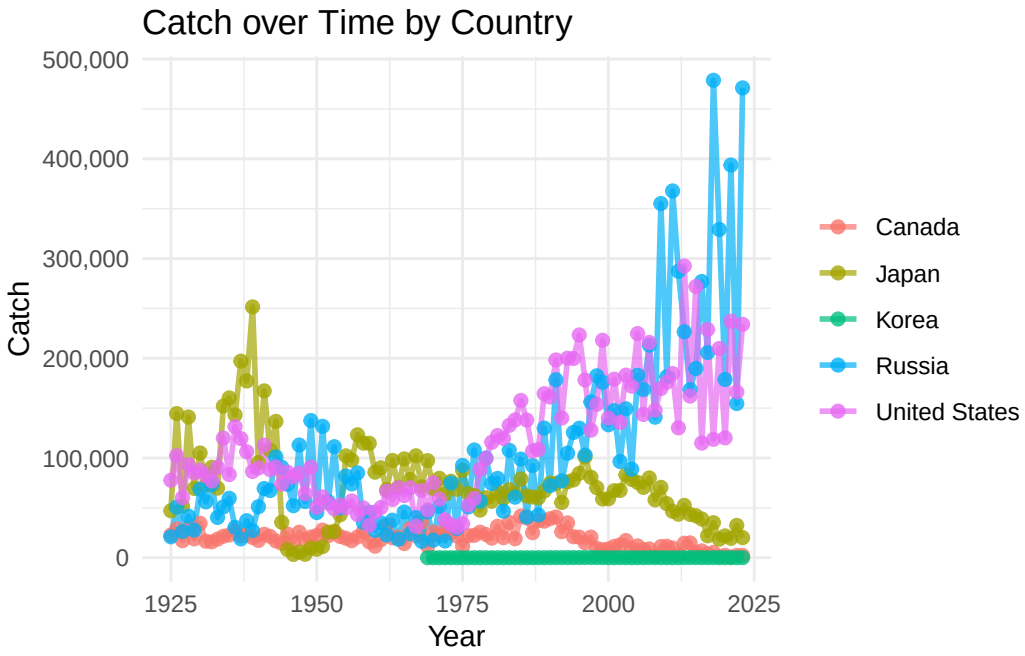


Figure 1: **?(caption)**

2.4 Outcome and Predictor variables

3 Model

Here we model the data in 2 ways. The first way is to model the rate at which the fishes are being caught for any country. And the second model looks at what is the probability that a country is fishing for commercial purposes.

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`. We also use `brm` package from Bürkner (2017)

3.1 Model Step- Up

Model 2

the base line probability of having a country that is fishing for commercial purpose is log of 0.838 which is 69.8%. which means that there is 69.8% probability that the country is fishing for commercial purposes, without even knowing the country.

```
[1] "Estimate.Intercept"  "Est.Error.Intercept" "Q2.5.Intercept"
[4] "Q97.5.Intercept"
```

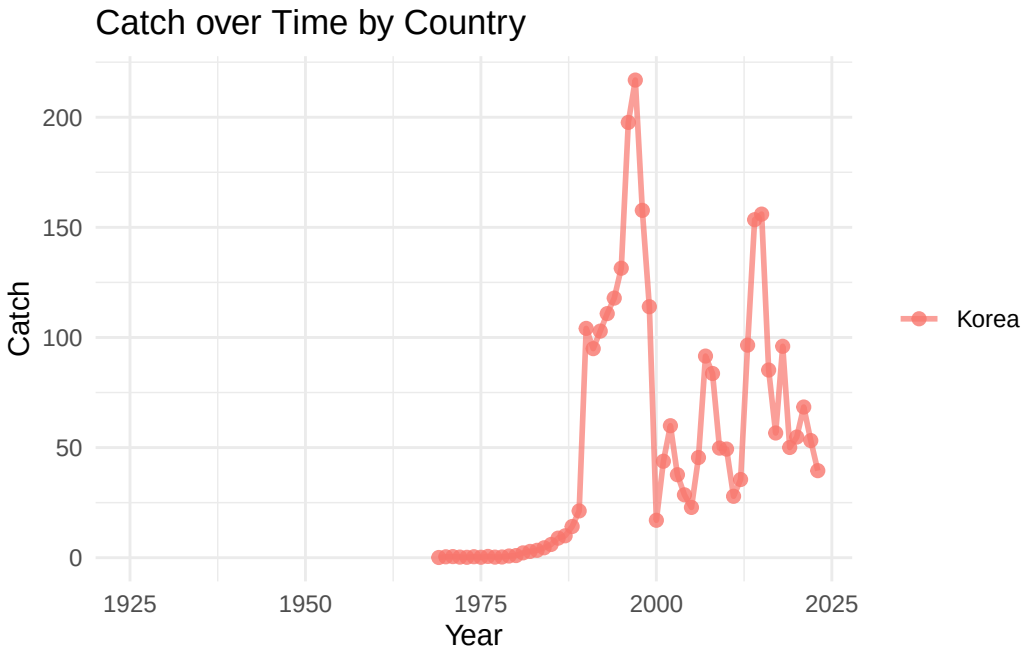


Figure 2: **?(caption)**

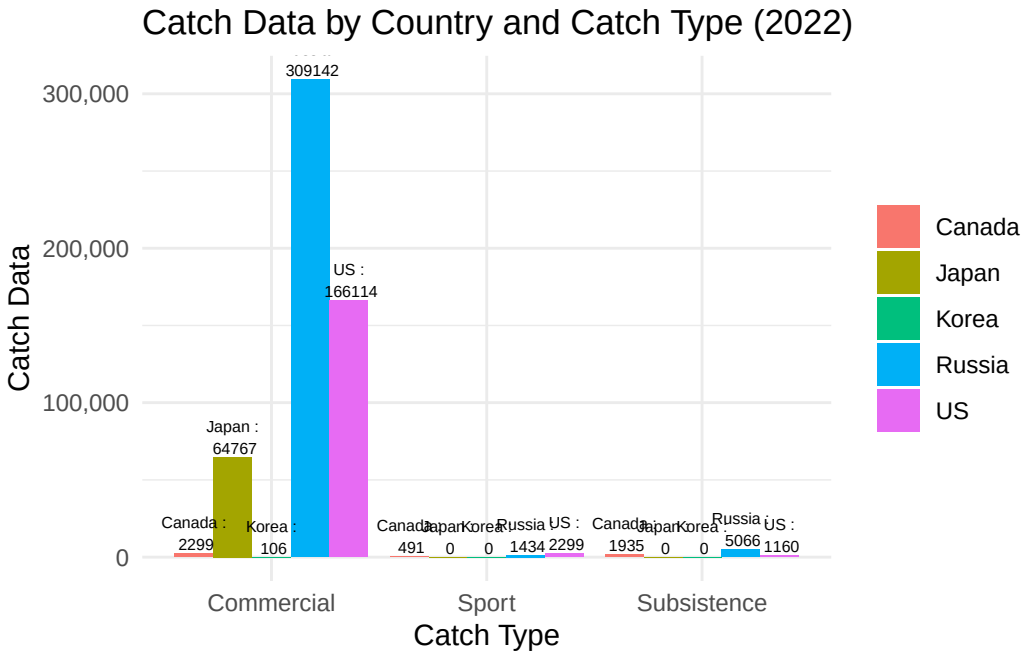
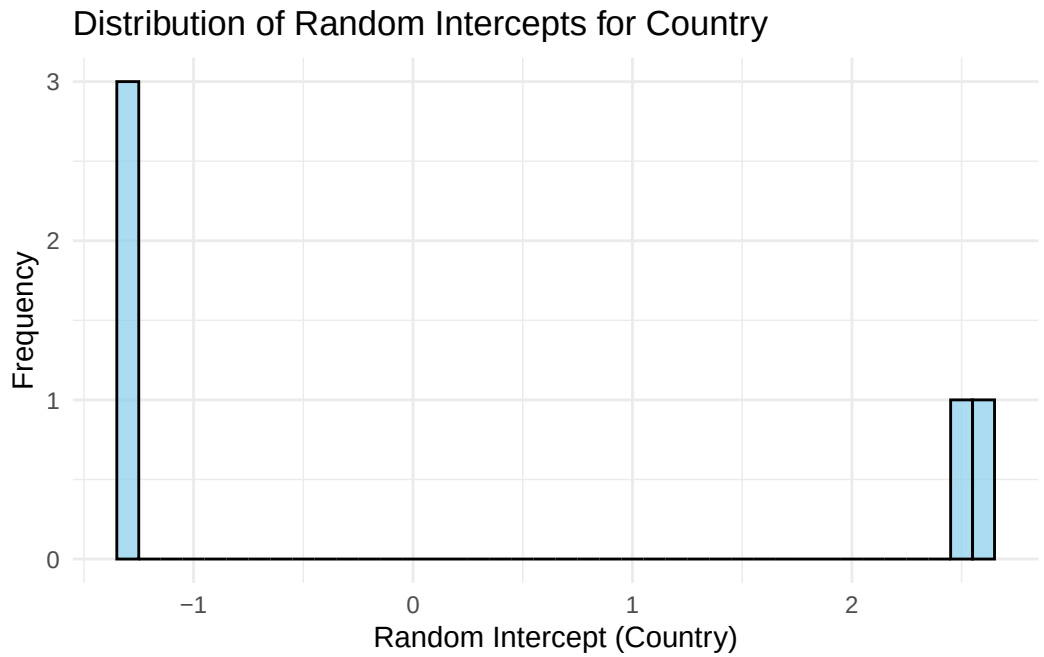
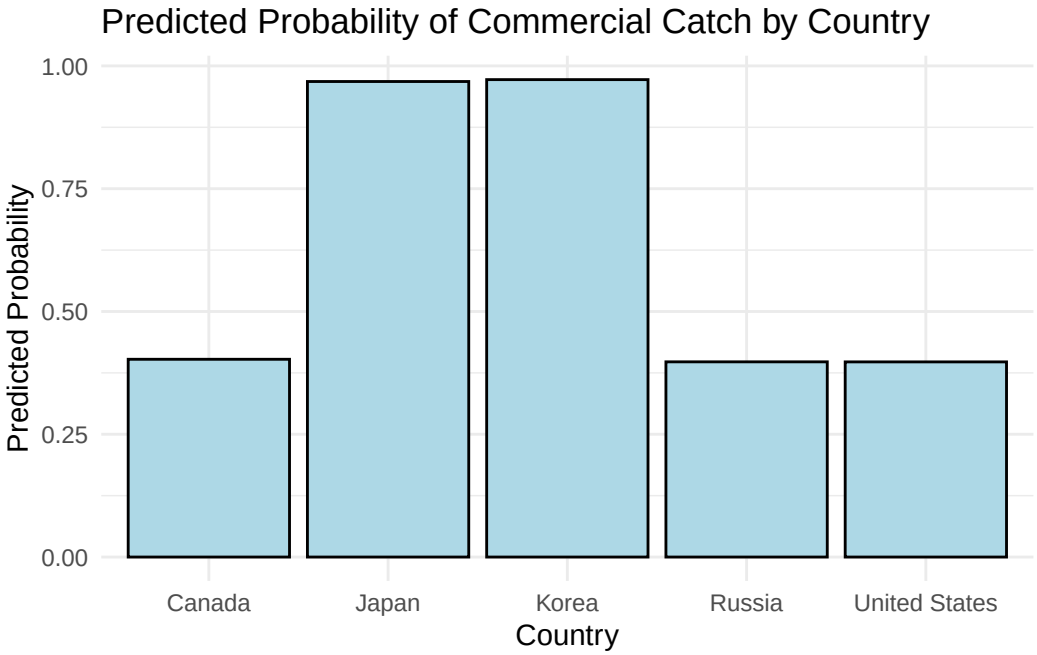


Figure 3: **?(caption)**



	Estimate.Intercept	Est.Error.Intercept	Q2.5.Intercept	Q97.5.Intercept	Country
Canada	-1.300951	1.544680	-4.6789422	1.5447404	Canada
Japan	2.511131	2.130351	-0.5104108	7.7166799	Japan
Korea	2.645726	2.093863	-0.3000692	7.8361163	Korea
Russia	-1.322698	1.171082	-3.7730737	0.9469157	Russia
United States	-1.323180	1.480182	-4.5734170	1.3398343	United States


Predicted Probability of Commercial Catch by Country
1.00
0.75
0.50
0.25
0.00
Predicted Probability
Canada
Japan
Korea
Russia
United States
Country

3.1.1 Model justification

4 Results

4.1 Model Results

Table 1: baysian model summary for predicting the rate of fishes being caught

	First model
(Intercept)	−1 367 335.76
	$(1.902\,962 \times 10^5)$
Year	721.74
	$(9.452\,000 \times 10^1)$
Sigma[Country × (Intercept),(Intercept)]	3 487 096 712.07
	$(2.258\,065 \times 10^9)$
Num.Obs.	451
R2	0.448
R2 Adj.	0.439
R2 Marg.	0.077
Log.Lik.	−5547.420
ELPD	−5555.7
ELPD s.e.	36.7
LOOIC	11 111.3
LOOIC s.e.	73.4
WAIC	11 111.1
RMSE	53 056.21
r2.adjusted.marginal	0.438511212758486

Table 2: baysian model summary for predicting the probability of country fishing for commercial purposes.

	Second model
b_Intercept	0.84
	(1.03)
sd_Country___Intercept	2.32
	(1.07)
Num.Obs.	45
R2	0.267
R2 Marg.	0.000
ICC	0.7
ELPD	−26.1
ELPD s.e.	2.6
LOOIC	52.2
LOOIC s.e.	5.2
WAIC	51.6
RMSE	0.41

5 Discussion

5.1 Weaknesses and next steps

Korea didn't report data until 1969 so they are left out of the first model in general.

A Appendix 1

B Appendix 2 (datasheet)

1. For what purpose was the dataset created? The dataset was originally created by North Pacific Anadormous Fish Commission ((NPAFC) 2024) created this dataset to see how much fishing of trout and salmon was done in international waters. They strictly prohibits mass fishing of these highly demanded fishes and hense the dataset. We use the dataset to analyse the number of these fishes caught and see if there is a reason for NPAFC to be worried, and if there is a solution to this.

2.What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? The instances of these data is either numeric or catagorical. Thee first couple of columns tells you the country the data is from, where the fishes were hunted as well as getting to specific regions. Why they were hunted and what the unit were, was also in the data set. The majority of the dataset is numbers expressing how much fish was hunted, either in thousands or tonnes.

3.Is any information missing from individual instances? There are several instances of data missing. Most are from the number of fishes column as some of the data goes back to 1925 but, not all of them so, there are some rows of data where the total number of fishes collected in 1928, for example, is not present.

4. How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/ derived from other data (for example, part-of-speech tags, model-based guesses for age or language)?

This dataset was reported by the subjects of each country. Meaning this organisation asked the government of Canada, US, Korea, Russia and Japan, send in the numbers and they compiled this data.

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bürkner, Paul-Christian. 2017. "Brms: An r Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80 (1): 1–28. https://doi.org/10.18637/jss.v080.i01.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

(NPAFC), North Pacific Anadromous Fish Commission. 2024. "Statistics – North Pacific Anadromous Fish Commission." https://www.npafc.org/statistics/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Robinson, David, Alex Hayes, and Simon Couch. 2023. *Broom: Convert Statistical Objects into Tidy Tibbles.* https://CRAN.R-project.org/package=broom.

Schauberger, Philipp, and Alexander Walker. 2024. *Openxlsx: Read, Write and Edit Xlsx Files.* https://CRAN.R-project.org/package=openxlsx.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Jennifer Bryan. 2023. *Readxl: Read Excel Files.* https://CRAN.R-project.org/package=readxl.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.