Is there a relationship between ocean health and happiness?

Contents

1 Introduction (150 words)

In a recent paper published in the Journal of Environmental Research and Public Health, Wheaton *et al.* [1] describes the need for a better understanding of our relationship with the ocean and human wellbeing. The wellbeing benefits of being near or in the ocean have been qualitatively described on an individual or community level [2, 3, 4] and Wuebben and Gonzalez-Trueba [5] have identified coastal tourism as having a positive impact on human wellbeing. In addition, 2021 - 2030 has been declared by the United Nations as the 'Decade of Ocean Science for Sustainable Development' [6, 1] which highlights how important our relationship to this ecosystem is. However, research also describes the inequity present with regards to access to the ocean, with certain socioeconomic and disadvantaged groups having less availability to the ocean than others [7, 8]. Identifying the relationship between the ocean and human wellbeing on a regional or global scale could help to reinforce the importance of protecting coastal ecosystems and help to identify where more work is needed in bringing a sense of connection to the ocean in certain areas. Therefore, this analysis takes a more quantitative approach and aims to investigate the relationship on a more macro level by exploring how our connection with, and reliance on, the ocean may influence our happiness on a country by country basis.

2 Aims and Objectives (150 words)

The aims of this analysis are to determine whether there is a significant correlation between a countries happiness and the health of the ocean that the country borders. Furthermore, it aims to identify what it is about our connection with the ocean that has the most influence on human wellbeing. To achieve these aims the following objectives need to be achieved: The first step will be to obtain data which include the required information. Since the data is likely to come from different sources, the data need to be cleaned, wrangled and merged together to form one dataset to be used for analysis. This will then allow any missing data to be identified and the appropriate method for handling these data can be applied. From here, descriptive statistics can be carried out to provide more insight into the data. With a better understanding of the features of the dataset a statistical model can then be created to identify any relationship between the features, testing different methodologies to create the most appropraite model for the data. Finally, the results of the analysis will be presented using both visual assets and model outputs followed by discussion around the significance of the results and any limitations of the methods used for analysis. For the data cleaning and wrangling, Python 3.10.2 [9] will be used, for the statistical analysis, R version 4.1.2 [10] will be used, and for the geographical presentation of the data, QGIS 3.22.3-Białowieża [11] will be used.

3 Data (500 words)

For the analysis, multiple datasets were identified as appropriate sources: the Ocean Health Index (OHI) dataset and the World Happiness Report (WHR) dataset.

3.1 Ocean Health Index

The OHI is a global assessment of 220 countries and assess the health of the ocean within each countries exclusive economic zone (EEZ). The assessment has been conducted from 2012 - 2021 and is still an ongoing project. The researchers use 10 goals and subgoals that they consider to be an import attributing factor to the health of the ocean for a country. These goals are weighted and used to calculated an overall index score for a country. A description of each goal involved in the calculation of the index score can be seen in Figure 1. In the context of the blue economy and understanding the current status of our oceans, as well as the future threats, the OHI has been beneficial is exploring the link between the indicators of ocean health and the gaps in the research where more data or analysis is need to help protect and improve the state of the global oceans [12].

The data set comes in the format of a CSV file and contains an observation for each country for each goal,

year and an overall index score. The goals and index are values between 0 - 100 where the index is the weighted calculations of the goals. Since the index is calculated based on the goals, the index was not used in the analysis and instead each goal was used to determine its influence over the response variable. The data was filtered to only include 2015 - 2019 since this was the data range available in both the WHR and OHI datasets. The data also needed pivoted so that each goal was a column and to tidy the data.

Goal	Abbreviation	Description
Food Provision	FP	The sustainable harvest of seafood from wild-caught fisheries and mariculture
Artisanal Fishing Opportunity	AO	The opportunity for small-scale fishers to supply catch for their families, members of their local communities, or sell in local markets
Natural Products	NP	The natural resources that are sustainably extracted from living marine resources
Carbon Storage	CS	The condition of coastal habitats that store and sequester atmospheric carbon
Coastal Livelihoods and Economies	LE	Coastal and ocean-dependent livelihoods (job quantity and quality) and economies (revenues) produced by marine sectors
Tourism and Recreation	TR	The value people have for experiencing and enjoying coastal areas through activities such as sailing, recreational fishing, beach-going, and bird watching
Sense of Place	SP	The conservation status of iconic species (e.g., salmon, whales) and geographic locations that contribute to cultural identity
Clean Waters	CW	The degree to which ocean regions are free of contaminants such as chemicals, eutrophication, harmful algal blooms, disease pathogens, and trash
Biodiversity	BD	The conservation status of native marine species and key habitats that serve as a proxy for the suite of species that depend upon them
Coastal Protection	CP	The amount of protection provided by marine and coastal habitats serving as natural buffers against incoming waves

Figure 1: The goals and their definitions used to calcualted the index score for each country

3.2 World Happiness Report

The WHR is a publication by the Sustainable Development Solutions Network and has been published from 2012 - 2021 (excl. 2014). The report synthesis information from the Gallup World Poll and Lloyd's Register Foundation which provides data for the World Risk Poll. The data is collected using qualitative interviews from across 142 countries with a specific focus on ensuring that data is collected in countries where there has traditionally been a lack or of or limited data. The authors of the WHR then synthesize these data add calculate an overall happiness score for each country. There are six factors used to calculate the overall happiness score: GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity and perceptions of corruption. The format of the data was such that each country had a value for each factor and an overall happiness score. Since the happiness score is a composite index and the values for each factors was the extent to which it influenced the happiness score, the happiness score was used for analysis and considered as the response variable. Each year was its own data set so for each file, a 'year' column was added and all the files were merged into one dataset.

Once a tidy version of the OHI and the WHR had been made containing data for 2015 - 2019, the two

datasets needed to be joined together. This was carried out using Pandas in Python and involved some pre-processing steps. The datasets were to be joined using matching values for country and year, however some countries were not consistently named between datasets. These were identified and made consistent then the two datasets were merged. The columns were renamed and ordered in a sensible order to make analysis easier.

To make the world map in Figure 2 the shape files of the OHI dataset were used. To this, the cleaned merged dataset was joined to the shape files to allow the ocean health index scores and the world happiness scores to be visualized. QGIS was used for the joining and visualization of this image.

3.3 Missing Data

Missing data was imputed using the `sklearn` package in Python. Two algorithms were used to impute the missing data: K-nearest-neighbour (KNN) and IterativeImputer. Both of the algorithms used are multivariate imputation methods as this gives a more precise estimation than linear imputation [13]. The KNN algorithm was the chosen method used to impute the missing data since it impute the missing values based on related observations [14]. A hyper-parameter of 2 was chosen for the number of neighbors as this estimated the missing values while preserving the distribution of the data.

Insert world maps here and maybe some graphs of imputed data or DESCRIPTIVE STATS OF DATA

4 Analysis and Results (1500 words)

The world map in Figure 2 was used to initially compare the two datasets adn to identify any possible relationship. This initial comparison shows that there is some overlap between countries having high happiness and a high ocean health index. In areas where the happiness score is high (dark green), such as Australia and New Zealand, North America, and Scandinavia, the respective EEZ also have a high ocean health index score (dark blue). Similarly in places with lower happiness scores, such as Western and Sub-Sarahan Africa, and South Asia, there seems to be a relatively low ocean index score as well. To investigate this further a basic linear model was fitted to the data to see the relationship between the index and the happinessScore. Figure 3 shows that there seems to be a positive correlation between the variables. To gain a better understanding of which regions have the higher happiness and ocean health index scores the data was grouped by region and the mean of the index and happiness scores were taken. Table 1 are the results of this and shows that Australia and New Zealand, North America, and Western Europe have both the highest average ocean health
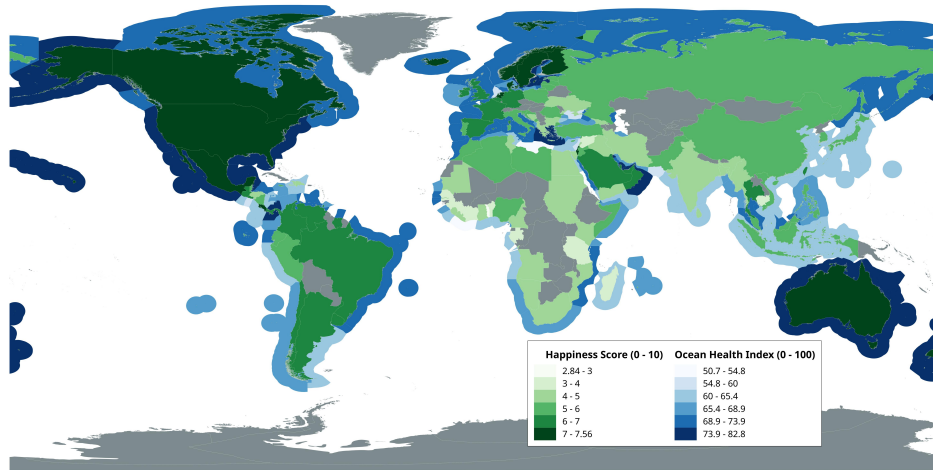
Figure 2: The happiness score for each country and the ocean health index for the connected exclusive economic zone (EEZ) for that country. Darker colours represent higher scores and lighter colours represent lower scores. While the ocean health index score was not used in the analysis this gives an initial overview of the health of each countries ocean

index score and average happiness score.

Include some basic descriptive stats like average OHI for each region and WHR for each region

Start with the relationship between sense of place (OR JUST OHI INDEX) and world happiness as this is what I was initially interested in

- Scatter plot of graph with model fitted

- Talk about R value and why it is low

- Maybe include plot split by region and state how this changes the relationship. Say that I will first create a model by including other varibles and then perform the heirarchical modelling to investigate the effect of region

- Initally used OHI index and happiness score but since OHI index is a composite index it was not included in the models which used the other features

Then describe the final basic model used and present some daignostic charts

- Compare it to the initial model

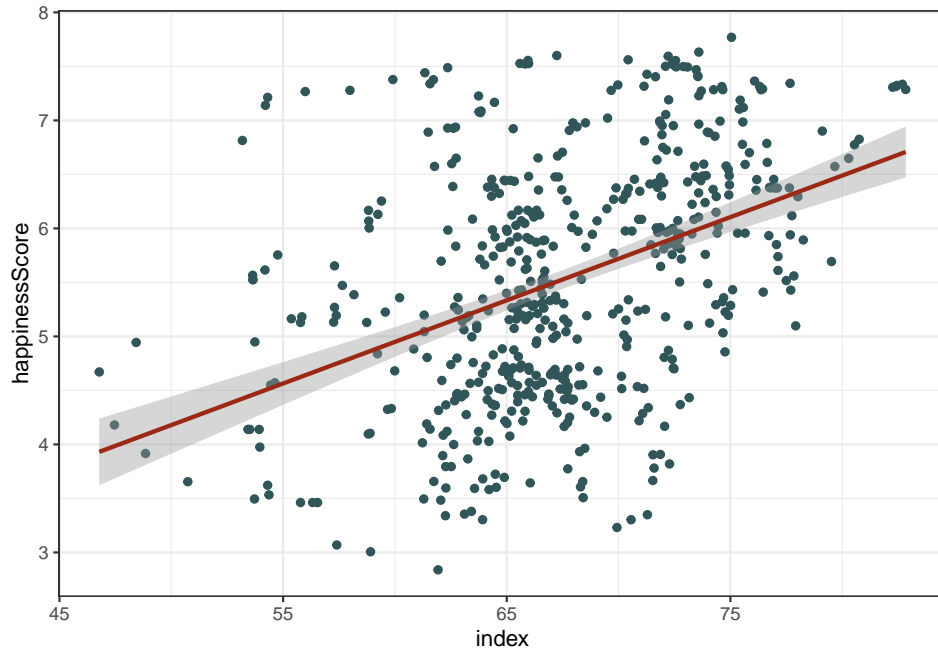*Go into hierarchical modelling and fixed effects modelling*

Figure 3: An initial model comparing Happiness Index using the built stats method in ggplot2. The index feature was used here to gain an initial understanding of the relationship between ocean health and happiness however further analysis will use the goal features from the OHI as a proxy for the ocean health index.

Table 1: Mean ocean health index and happiness score for each region

region	index	happinessScore
Australia and New Zealand	78.35	7.29
North America	72.72	7.17
Western Europe	71.06	6.73
Central and Eastern Europe	69.12	5.32
Southeastern Asia	67.96	5.22
Latin America and Caribbean	67.89	6.04
Middle East and Northern Africa	66.14	5.33
Sub-Saharan Africa	64.26	4.29
Eastern Asia	64.22	5.83
Southern Asia	62.57	4.65

- Fixed effects graph
- random effects graph
- Regression table of mixed effect model

5 Limitations (300 words)

- Didn't compare the happiness difference between coastal and land locked countries (if I don't end up doing it)
- Some of the variables in the OHI data are heavy assumptions (sense of place in particular but check others)
- Talk about how there may be crossed random effects of the years due to repeated observations of each country

6 Conclusion (180 words)

- Restate aims of the analysis
- Data used
- Steps taken and what they showed
- Why a simple linear model/fixed effects model overstates relationship
- How accounting for region changed this relationship#
- ***Link back to the first paper cited and how more qual research is needed***

IF I HAVE TIME I COULD LOOK AT THE DIFFERENCE BETWEEN HAPPINESS OF COUNTRIES CONNECTED TO THE OCEAN AND NOT CONNECTED TO THE OCEAN

References

[1] B. Wheaton, J. T. A. Waiti, R. Olive, and R. Kearns, "Coastal Communities, Leisure and Wellbeing: Advancing a Trans-Disciplinary Agenda for Understanding Ocean-Human Relationships in Aotearoa New Zealand," *International Journal of Environmental Research and Public Health*, vol. 18, no. 2, p. 450, Jan. 2021, number: 2 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/1660-4601/18/2/450

[2] R. Foley, "Swimming in Ireland: Immersions in therapeutic blue space," *Health & Place*, vol. 35, pp. 218–225, Sep. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1353829214001452

[3] S. L. Bell, C. Phoenix, R. Lovell, and B. W. Wheeler, "Seeking everyday wellbeing: The coast as a therapeutic landscape," *Social Science & Medicine*, vol. 142, pp. 56–67, Oct. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0277953615300630

[4] D. Burdsey, *Race, place and the seaside: Postcards from the edge.* Springer, 2016.

[5] D. Wuebben and J. J. González-Trueba, "Surfing between Blue Humanities and Blue Economies in Cantabria, Spain," *symplokē*, vol. 27, no. 1-2, pp. 65–78, 2019, publisher: JSTOR.

[6] https://plus.google.com/+UNESCO, "United Nations Decade of Ocean Science for Sustainable Development (2021-2030)," Feb. 2017. [Online]. Available: https://en.unesco.org/ocean-decade

[7] R. Olive and B. Wheaton, "Understanding Blue Spaces: Sport, Bodies, Wellbeing, and the Sea," *Journal of Sport and Social Issues*, vol. 45, no. 1, pp. 3–19, Feb. 2021, publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/0193723520950549

[8] C. Phoenix, S. L. Bell, and J. Hollenbeck, "Segregation and the Sea: Toward a Critical Understanding of Race and Coastal Blue Space in Greater Miami," *Journal of Sport and Social Issues*, vol. 45, no. 2, pp. 115–137, Apr. 2021, publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/0193723520950536

[9] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual.* Scotts Valley, CA: CreateSpace, 2009.

[10] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[11] QGIS Development Team, *QGIS Geographic Information System*, QGIS Association, 2022. [Online]. Available: https://www.qgis.org

[12] B. S. Halpern, "Building on a Decade of the Ocean Health Index," *One Earth*, vol. 2, no. 1, pp. 30–33, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590332219302702

[13] K. S. Htoon, "A Guide To KNN Imputation," Jul. 2020. [Online]. Available: https://medium.com/@kyawsawhtoon/a-guide-to-knn-imputation-95e2dc496e

[14] L. Beretta and A. Santaniello, "Nearest neighbor imputation algorithms: a critical evaluation," *BMC Medical Informatics and Decision Making*, vol. 16, no. 3, p. 74, Jul. 2016. [Online]. Available: https://doi.org/10.1186/s12911-016-0318-z