Is there a relationship between ocean health and happiness?

Contents

1 Introduction

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

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

For the analysis, multiple datasets were identified as appropriate sources: the Ocean Health Index (OHI) dataset (https://oceanhealthindex.org/global-scores/data-download/) and the World Happiness Report (WHR) dataset (https://www.kaggle.com/unsdsn/world-happiness?select=2019.csv).

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 [12]. 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 [13].

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 [14]. 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 [15]. The KNN algorithm

was the chosen method used to impute the missing data since it impute the missing values based on related observations [16]. 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.
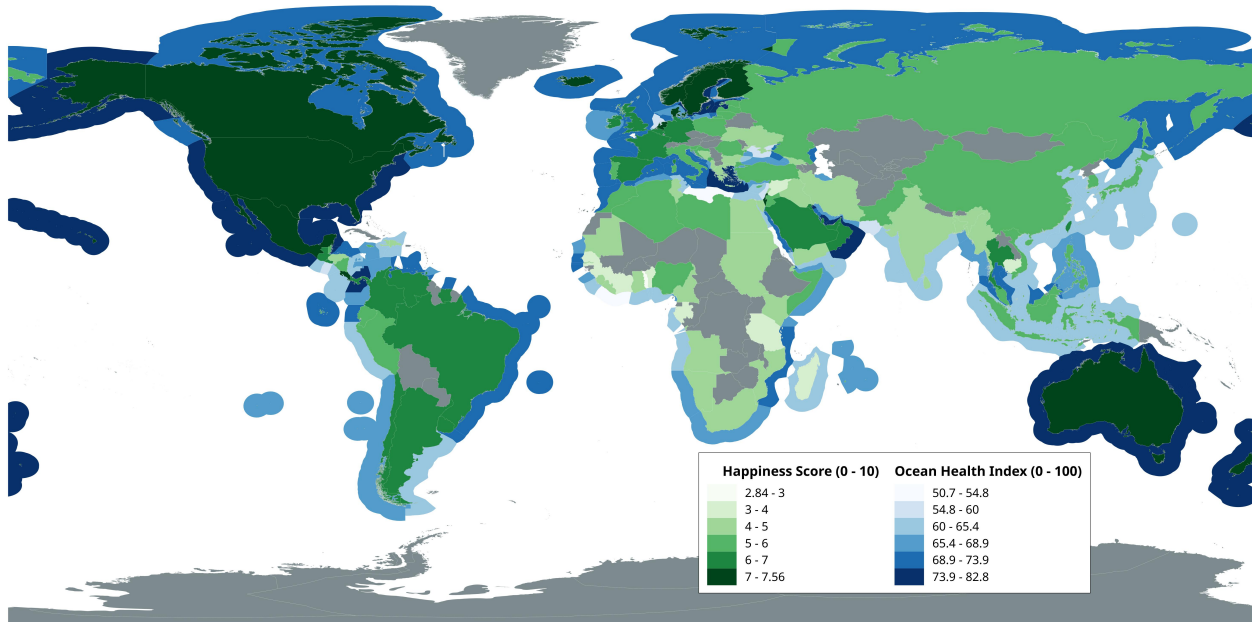

Figure 2: The happiness score for each country and the ocean health index for the connected exclusive economic zone (EEZ) for that country. Darker colours represent higher scores and lighter colours represent lower scores. While the ocean health index score was not used in the analysis this gives an initial overview of the health of each countries ocean

4 Analysis and Results

4.1 Initial visual analysis

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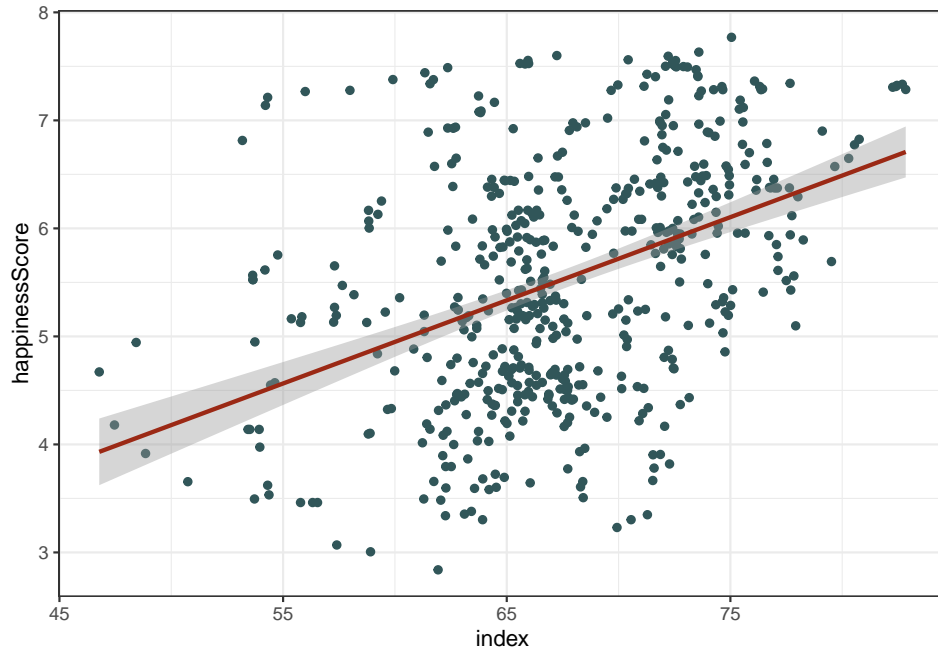
Figure 3: An initial model comparing Happiness Index using the built stats method in ggplot2. The index feature was used here to gain an initial understanding of the relationship between ocean health and happiness however further analysis will use the goal features from the OHI as a proxy for the ocean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

4.2 Statistcal modelling

From the visual analysis there seems to be some relationship between the health score variable and the ocean health index. To dig into this further, statistical modelling was used to investigate whether this relationship was significant and what variables influence this relationship. To identify any correlations between the goals in the OHI data and the happiness score, a correlation matrix was produced (Figure 4). The features that had the highest correlation with the happiness score (`senseOfPlace`, `cleanWater`, `foodProv`, `tourismAndRec`) were then selected for modelling.
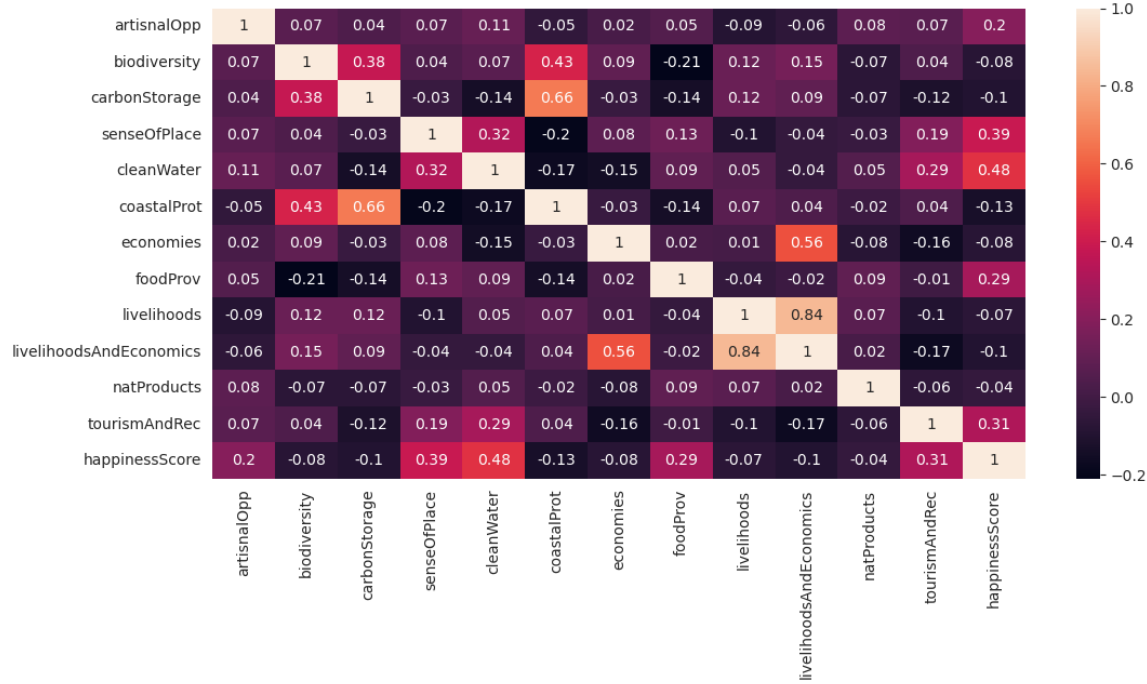

Figure 4: Correlation matrix heat map between variables where light coloured sections have a high positive correlation and dark colors have a negative correlation. Of main interest are the correlation coefficient of the OHI features and the happinessScore and any substantial covariance between the variables.

Initially, a univariate linear regression model using ordinary least squares was created (see code directory rAnalysis/rAnalysis.R -> lm1) using clean water as the explanatory variable since this had the highest correlation. coefficient with the happiness score. While this relationship was significant ($p <= 0.05$) the model had a low `Adjusted R^2` value of 0.229. Clean water had a slight covariation with the sense of place feature so this was added to the model however this was not significant ($p > 0.05$) (see code directory rAnalysis/rAnalysis.R -> lm2). The next model included all the features that had a relatively high correlation with the happiness score (`senseOfPlace`, `cleanWater`, `foodProv`, `tourismAndRec`, `artisinalOpp`, `biodiversity`, `natProducts`) and again a linear regression model was fitted to the data using the formula:`Happiness score ~ Sense of place + Clean Water + Food Provisions + Tourism and Recreation + Artisinal Opportunities + Biodiversity + Natural Products`. This proved to have a higher `Adjusted R^2` value of 0.3919 and all terms included in the model were significant ($p <= 0.05$) so this was one of the models used going forwards (see code directory rAnalysis/rAnalytics.R -> lm3). A possible interpretation of this initial model would be that these features contribute to the happiness of a country more than the other features considered in the OHI database and hence these features have a higher influence on wellbeing. Based on the literature we would expected this relationship since feeling connected to where you live (sense of place) [1], coastal tourism and water sports (tourism and recreation) [5], and being able to provide for your family and community (artisinal opportunities and food provisions) [8] have all been shown to have a beneficial impact on human wellbeing. There are however several reasons why it would be

difficult to draw conclusions from this model. These will be discussed once the goodness of fit of the model has been checked.

4.3 Model diagnostics

Model diagnostics were performed using the 'summary(lm3)' function in R. This produces 4 graphs that help to determine if the model is a good fit for the data and to identify if there are any significant influential outliers that should be investigated. They can also be used to check that the model has not broken any of the assumptions made when using an ordinary least squares regression model. These assumptions are that the there is a linear relationship between the data, the data was collected independently of one another and is homoscedastic, and that the residuals of the fitted model are normally distributed. Figure 5 are the diagnostic plots for the selected model which would suggest the model is a good fit for the data.
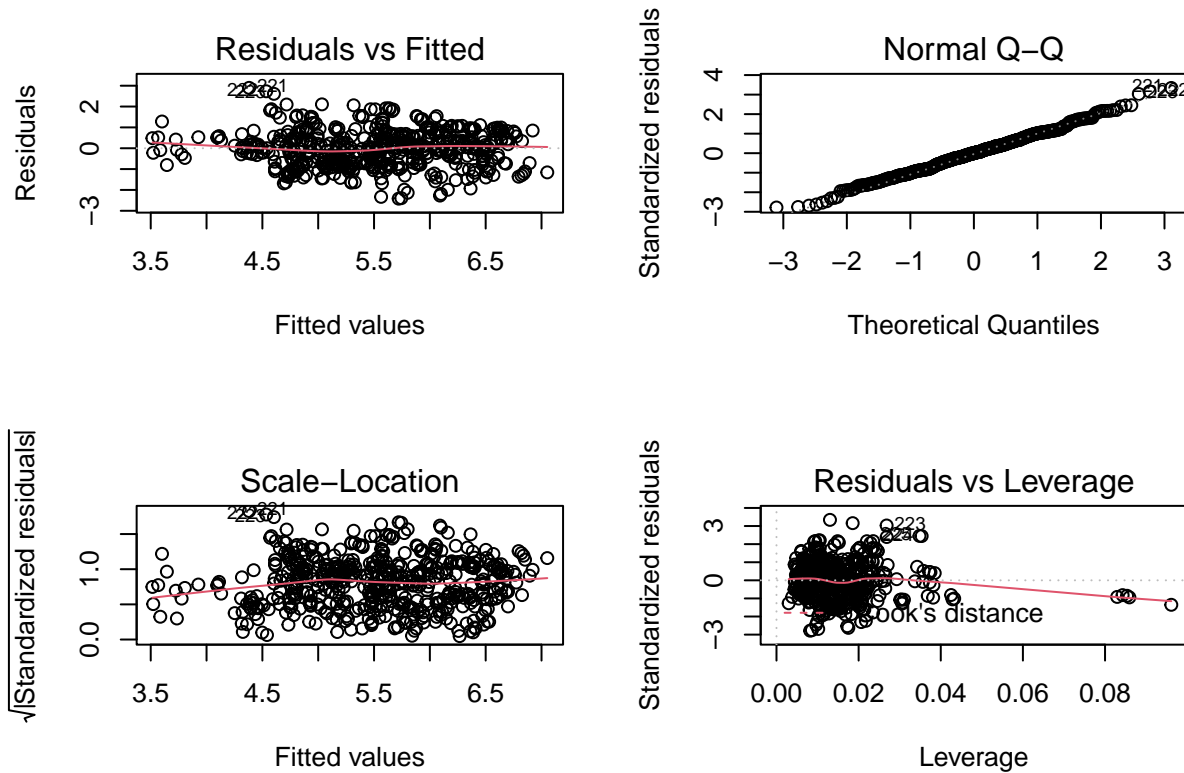

Figure 5: Diagnostic plots for the linear model selected to fit to the data. The Residuals vs Fitted can be used to check that there is a linear relationship between the data, Normal Q-Q can be used to check for normality of residuals, Scale-Location can be used to check that the variance of the residuals is homoscedastic, and Residuals vs Leverage can be used to identify any data points with high influence over the model. Based on these plots, the model seems to not violate any of the assumptions made and is a good fit for the data.

4.4 Further modelling

As mentioned previously, while this model would suggest some relationship between ocean health and happiness, the low `Adjusted R^2` value of the model would suggest that limited assumptions can be made using this model as it only explains a small proportion of the relationship. Furthermore, this model has not accounted for geographical location of the country or the year the data was collected. Since the health of an ocean is not only influenced by the country that borders it directly, but also by the surrounding countries, it is likely that there will be some random effects associated with the grouping of countries within regions. To explore this, fixed effects and hierarchical modelling was explored.

4.4.1 Fixed effects model

To see if there is a difference between regions and the relationship between ocean health and happiness, the intial basic model of `happiness ~ index` was used to plot a graph split by region. This would provide a fixed effect model so that each region would have it's own slope and intercept values and would allow any variation in relationships between regions to be identified. Figure 6 is the plot of happiness against ocean index split by region.
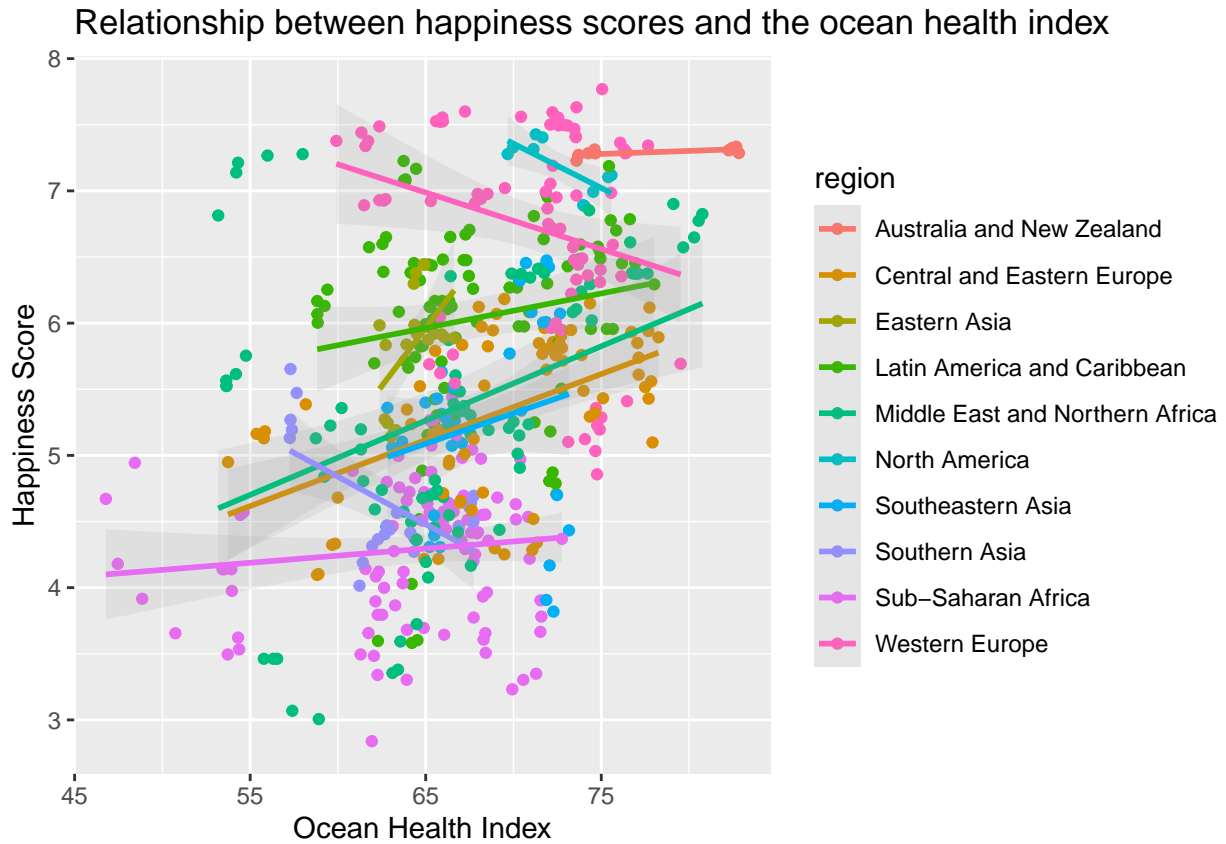

Figure 6: The fixed effect model split by region. This plot shows that there is a substantial difference between regions and the relationship of the variables.

From this it is clear that there is not a consistent relationship across regions and it would not be accurate to assume that there is a positive relationship between ocean health and happiness for all regions. For regions such as Latin America and Caribbean, Middle East and Northern Africa, and Southeastern Asia this assumption does seem to hold true however, for regions such as Southern Asia, Western Europe and North America the opposite relationship occurs.

To see if this can be accounted for in our model, the final model using the formula `Happiness score ~ Sense of place + Clean Water + Food Provisions + Tourism and Recreation + Artisinal Opportunities + Biodiversity + Natural Products` was used and hierarchical modelling applied.

4.4.2 Hierarchical modelling

The final model was adapted to a random intercepts model to allow for any variation between regions to be accounted for. This was achieved by updating the formula to: `Happiness score ~ Sense of place + Clean Water + Food Provisions + Tourism and Recreation + Artisinal Opportunities + Biodiversity + Natural Products + (1 | Region)`. Fitting this model to the data revealed that the variables tourism and recreation, natural products and biodiversity were not significant predictors

(p > 0.05) of the relationship. These were then removed one at a time from the model till on the significant variables were left (`Happiness score ~ Sense of place + Clean Water + Food Provisions + Artisinal Opportunities + (1 | Region)`). With this update model, the variance associated with the region change can then be obtained. This resulted in a variation associated with region differences of 64%. This is high and would suggest that for any prediction model to be accurate, it would need to consider the effect that region has on the happiness variable. Figures 7 and 8 also suggest that the initial linear model overfitted the relationship of the data.
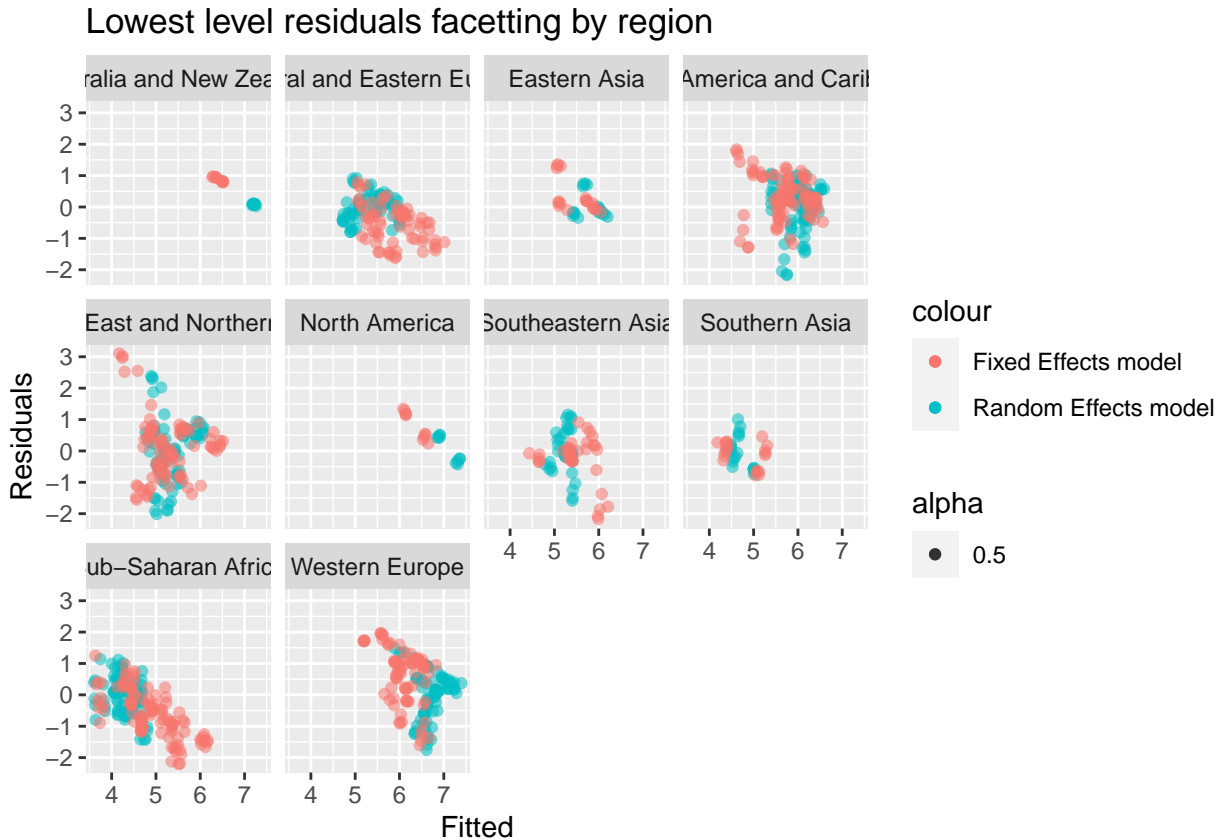

Figure 7: The lowest level residuals plot facetted by region for both the random effects model (green) and the fixed effects model (red). Comparing this to our initial model it would not seem like our either model is a good fit for many of the regions and would not accurately predict the data. It also shows that some regions such as Australia and New Zealand, and North America, have limited amounts of data and may not be enough to perform analysis on if split by region.
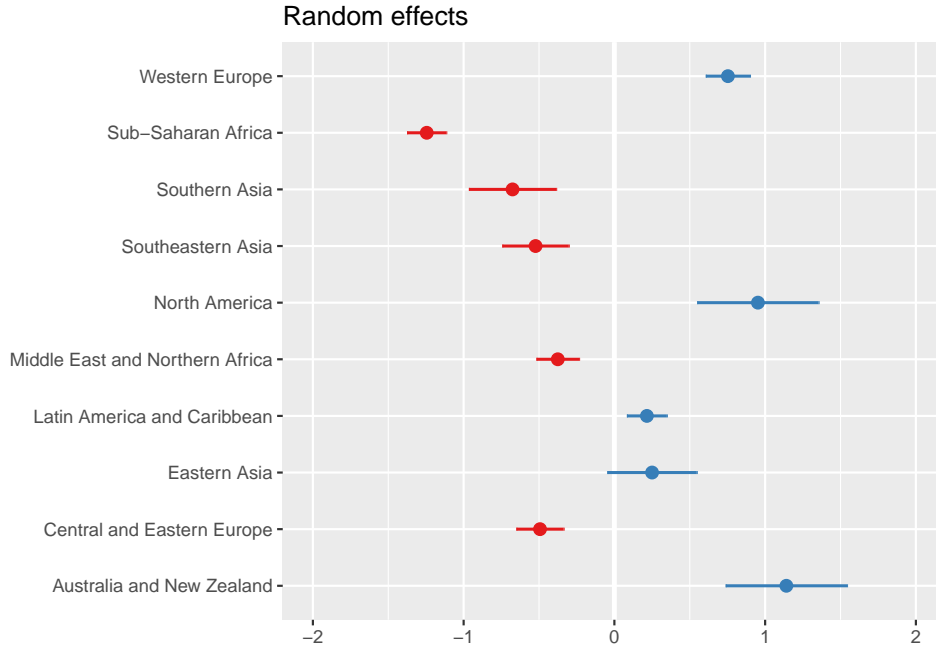
Figure 8: The random effects graph displays the performance of the model. The points should be randomly distributed around the zero mark with no extreme outliers. This plot would suggest that the model does not perform well and that the outliers may have substantial influence over the model

5   Limitations

There were several limitations associated with this analysis. A main limitation is that while processing and cleaning the data, any countries that did not have an associated OHI score were dropped from the dataset. This means that only countries with a border to the ocean were included in the analysis. While this makes sense when looking at the direct relationship between ocean health and happiness, it would have been useful to include the non-coastal countries in the dataset. They could have been used to control for any factors that may impact coastal countries and their happiness just by virtue of location rather than the health of the ocean.

Another limitation is the data used for the analysis. Both the OHI data and the WHR data are estimates made by experts on the status or influence that each factor has on the overall score or index. For example, the OHI goal 'Sense of Place' uses the conservation status of species and geographic locations that contribute towards cultural identity as a proxy for how much the population of that country connects to the ocean. While this is a good estimate, it may not provide an accurate reflection of how connected the citizens of a country feel to the ocean as political or socioeconomic factors may influence how many species or geographical locations are identified as protected. This is to say that any relationship that may have been gleamed from the model may not be an accurate reflection of ocean health and happiness on a country by country basis.

6   Conclusion

This project aimed to investigate the relationship between the health of the ocean connected to a country and the happiness of that country and to see what factors might influence this relationship. Using the Ocean Health Index data base and the World Happiness Report between 2015 - 2019, analysis was carried out to explore these relationships. It was clear from the visual analysis that happiness and ocean health differed between regions and countries. The initial simple linear model fitted to the data suggested that there may be a significant relationship between a sense of place, clean water, tourism and recreation, artisinal opportunities, biodiversity, food provisions, and natural products, and the happiness of a country when looking at the data

on a global scale. However fitting a random effects model controlling for region showed that a significant proportion of the variation in the data was attributable to the effect that the region had on the relationship (over 60%) and that the linear model was over predicting the relationship. Diagnostics of the fixed effects model and the random effects suggest that the model did not perform well for all regions. Further analysis should look further into random effects modelling of the data to account for the variation between regions and also between years. Further models could include nested effects modelling to control for differences between regions and countries within those regions. It should also include non-coastal countries to account for any other factors that may be influencing this relationship.

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/S135382921 4001452

[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, C. Longo, D. Hardy, K. L. McLeod, J. F. Samhouri, S. K. Katona, K. Kleisner, S. E. Lester, J. O'Leary, M. Ranelletti, A. A. Rosenberg, C. Scarborough, E. R. Selig, B. D. Best, D. R. Brumbaugh, F. S. Chapin, L. B. Crowder, K. L. Daly, S. C. Doney, C. Elfes, M. J. Fogarty, S. D. Gaines, K. I. Jacobsen, L. B. Karrer, H. M. Leslie, E. Neeley, D. Pauly, S. Polasky, B. Ris, K. St Martin, G. S. Stone, U. R. Sumaila, and D. Zeller, "An index to assess the health and benefits of the global ocean," *Nature*, vol. 488, no. 7413, pp. 615–620, Aug. 2012. [Online]. Available: http://www.nature.com/articles/nature11397

[13] 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

[14] "World Happiness Report." [Online]. Available: https://worldhappiness.report/

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

[16] 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

7 Appendix

All code can be found on the github repository for this project at: https://github.com/charliegallop/ocean-health-and-happiness or via the shared OneDrive folder with Tim