Happiness, Climate Awareness, and COVID-19 Across Countries
Exploring Disparate Data: Part 3 - Final Report

Piping data and piping moms

Due November 26th, 2025

- Jack St. John (169137076)
- Boden White (169037964)
- Isaac Zeng (169040854)

1 Abstract

This report explores how climate change awareness, the COVID-19, and national happiness relate to each other across countries. Using three cleaned data sets on climate change awareness (climate_awareness), COVID-19 2020 data (covid_2020), and world_happiness levels, we join the data by country to form a single analysis table.

We compute summary statistics and create three visualizations to examine how the share of people who say they know "a lot" about climate change and the burden of COVID-19 cases are associated with national happiness scores. We explore how climate change awareness relates to national happiness, how COVID-19 cases relates to happiness, and how happiness varies across continents and across groups of countries with different levels of climate awareness.

Overall, we find a modest positive association between climate awareness and happiness, while the relationship between COVID-19 case counts and happiness appears weaker and more variable across continents.

2 Introduction

Three of the most and contentious global issues over the last decade are climate change, the COVID-19 pandemic, and the happiness index. In each of these cases, both government and research teams are interested in the mental as much as the physical aspects of these problems, including understanding what the public thinks about them and how they affect overall citizen wellbeing. The public's awareness of climate change can influence support for climate policies and adaptation to climate change, and the COVID-19 pandemic has had far-reaching impacts on healthcare systems, economies, and day-to-day life; however, the data from the happiness index provides an aggregate measure of the extent that people feel positively (or negatively) about their lives across countries.

Moreover, in this report, our three datasets — climate_change_report (awareness by country), covid_2020 (which includes demographic and case count data for every country) and world_happiness (internationally reported life-satisfaction scores) — provide a framework to look at social, health and environmental indicators on an international basis. We use data joins, descriptive statistics and visualizations to identify relationships between these variables.

In this analysis, our objective is to analyze the relationship among awareness of climate change, the burden of COVID-19 and self-reported happiness across countries. More specifically, we seek to determine if countries

that have high awareness of climate change are generally happier than countries that do not, if the impact of COVID-19 has lowered happiness, and how the relationship between climate awareness and happiness differs by continent and/or level of climate awareness. Although we cannot assert causality, the patterns we see could potentially be used as a springboard to discuss global inequities and possible behavioral explanations related to well-being in the context of global health and environmental crises.

3 Data Description

3.1 «Climate awareness data (climate_awareness.parquet)»

In Part I, we imported the climate opinion data from an Excel file and selected the sheet corresponding to climate change awareness. The raw format had countries in columns and awareness categories in rows, so we first used pivot_longer() to convert the wide layout into a long format with a country variable and a score variable. We then cleaned the awareness category names by recoding long text descriptions into shorter codes such as aware_no, aware_alittle, aware_moderate, and aware_alot. After that, we used pivot_wider() to produce a data set with one row per country and separate columns for each awareness category, and ensured that aware_base correctly recorded the total number of respondents. Finally, we saved the cleaned data as climate_awareness.parquet for use in this report.

In order to clean the data, we reformatted it using pivot_longer() to convert the wide layout into a long format with a country variable and a score variable. We then cleaned the awareness category names by recoding long text descriptions into shorter codes such as aware_no, aware_alittle, aware_moderate, and aware_alot. After that, we used pivot_wider() to produce a data set with one row per country and separate columns for each awareness category, and ensured that aware_base correctly recorded the total number of respondents. Finally we saved the cleaned data as climate_awareness.parquet for use in this report.

3.2 «COVID-19 cases data (covid_2020.parquet)»

The data comes from a daily case-count data set, based on data maintained by Our World in Data. The original data track the cumulative number of confirmed COVID-19 cases over time for locations around the world. In our cleaned covid_2020 data set, each row represents a counter, and the main variables we use are country, continent, and total_cases, which records the total number of confirmed COVID-19 cases for that country in 2020.

In order to clean the data, we filtered the original data to keep only observations from the year 2020 and removed rows corresponding to aggregate regions (such as "World") or locations without known continents. We then grouped the data by country and continent and took the maximum value of the cumulative case count over 2020 to represent the total number of confirmed cases that year, stored in the total_cases column, We kept only the variables relevent for this analysis and saved the result as covid_2020.parquet.

3.3 «Happiness data (happiness.parquet)»

The data comes from the World Happiness Report, which compiles survey-based measures of life satisfaction across countries. In the original file, each row corresponds to a country in a particular year and includes a "ladder" score of happiness aong with several explanatory variables related to income, social support, and health. In our cleaned happiness data set, each row represents a country in its most recent available year, and the main variables we use are country, year, and ladder score, the average life satisfaction score from 0 to 10.

In order to clean the data, we first standardized the column names using janitor::clean_names() and renamed the country variable to country so that it would be consistent with our other data sets. We sorted the data by country and year and then grouped by country to keep only the most recent year of data for each country.

Table 1: Summary statistics by continent for happiness, climate awareness, and total COVID-19 cases.

Continent	Number of countries	Mean happiness	SD happiness	Mean share aware a lot	Median total cases
Africa	17	4.45	0.72	0.02	26277.0
Asia	22	5.38	1.10	0.02	190618.0
Europe	27	6.58	0.60	0.02	323367.0
North America	8	6.04	1.05	0.01	126928.5
Oceania	1	6.97	NA	0.02	28296.0
South America	8	6.25	0.27	0.01	801970.0

We removed rows with missing ladder_score values and selected a subset of variables that were relevant for our analysis, including the outcome ladder_score. Finally, we saved the cleaned data as happiness.parquet.

3.4 Combining the Data

To combine the three data sets, we joined them using the common country variable. We started from the happiness data and used a left join to add the climate awareness variables from climate_awareness, so that all countries with happiness scores were retained whenever possible. We then used another left join to merge in the covid_2020 data, which added the total number of confirmed COVID-19 cases in 2020 and a continent label for each country. Finally, we created a new variable, share_aware_alot, which is the proportion of respondents in each country who reported knowing "a lot" about climate change (aware_alot / aware_base). This combined analysis_data table is used for the summary statistics and visualizations in the next section.

4 Exploratory Data Analysis

To achieve our goals, we explored the data by examining: how climate change awareness relates to national happiness, how total COVID-19 cases relate to happiness, and how happiness varies across continents and across groups of countries with different levels of climate awareness.

We explored many aspects of the data, but will demonstrate three. These are a scatterplot of awareness vs. happiness, a scatterplot of total COVID-19 cases vs. happiness, and a comparison of happiness across awareness groups.

The first aspect that we found interesting is shown in 1. The insight should be specific to the data shown, not a general statement beyond the data.

Figure 1 suggests that countries where a larger share of respondents say they know a lot about climate change tend to have higher happiness scores. Although there is considerable scatter, the overall upward trend in the points and the fitted line indicate a positive association between climate awareness and life satisfaction at the country level.

This insight is supported by the summary statistics in Table 1, which summarize the variables by continent.

Table 1 shows that continents with a higher average share of respondents who know a lot about climate change also tend to have higher mean happiness scores. At the same time, the typical number of COVID-19 cases in 2020 varies widely across continents and does not line up as clearly with happiness, which we explore further in the next insight.

The next insight that we found is how the relationship between climate awareness and happiness varies by continent. Figure 2 facets the same scatterplot by continent so that we can compare patterns across regions..

Figure 2 shows that the positive association between climate awareness and happiness is not identical across continents. In some regions, such as [insert the regions that appear strongest/clearest once you knit], the upward trend is more pronounced, while in others the relationship is weaker or based on fewer countries.
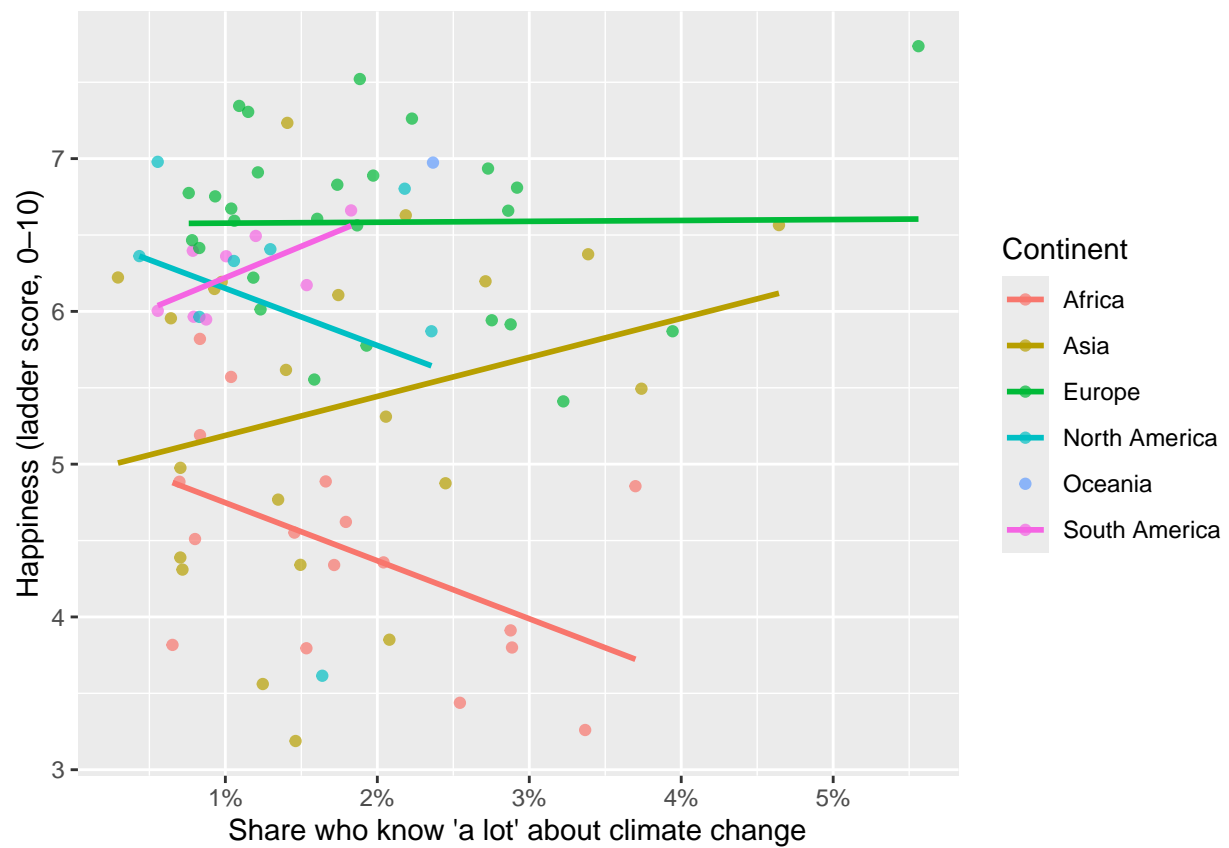
Figure 1: This is a figure caption that you will need to change in order to get good marks in the visualization rubric items.
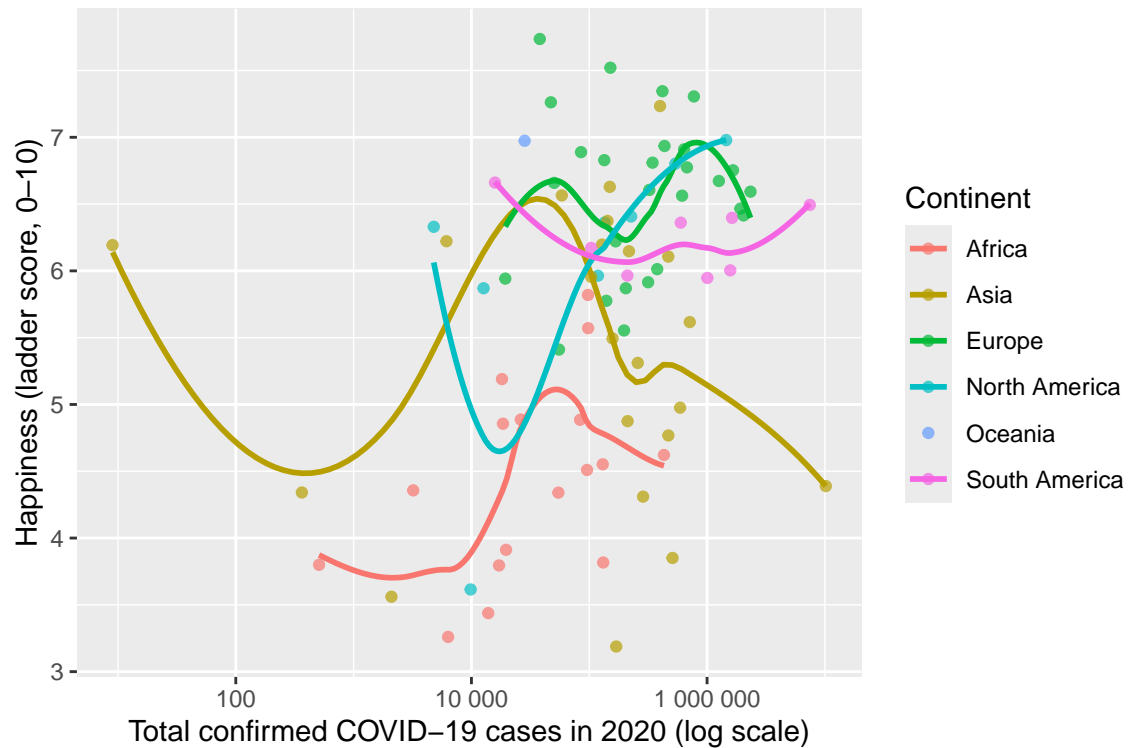
Figure 2: This is a figure caption that you will need to change in order to get good marks in the visualization rubric items.

This suggests that regional context may influence how climate awareness relates to well-being, and that other factors not included in our analysis, such as income or inequality, may also play an important role.

Finally, Figure 3 compares average happiness between countries with higher and lower levels of climate awareness. For this plot we classify countries into two groups based on whether their share_aware_alot is above or below the median value across all countries.
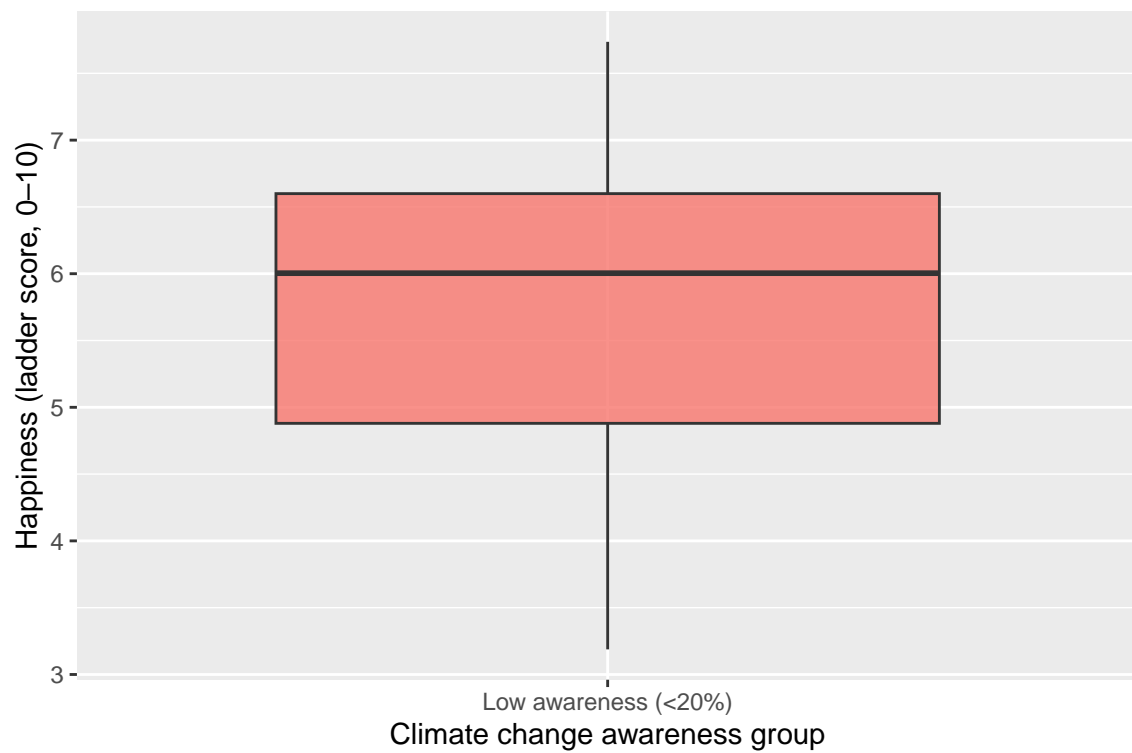
Figure 3: This is a figure caption that you will need to change in order to get good marks in the visualization rubric items.

5 Conclusion and Future Work

Overall, we found that countries with higher climate change awareness tend to report higher national happiness scores. This pattern appears both in the overall scatterplot and in the comparison between higher- and lower-awareness groups. When we looked at the relationship by continent, the positive association remained present in several regions, although the strength and clarity of the pattern varied. In contrast, the link between the total number of COVID-19 cases in 2020 and happiness was less clear from our summary statistics, suggesting that other long-term social and economic factors may be more important for explaining cross-country differences in life satisfaction.

A second conclusion from our analysis is that combining multiple data sources can reveal relationships that are not visible in any single data set alone. By joining climate awareness measures, COVID-19 case data, and happiness scores by country, we were able to examine how attitudes, health events, and well-being intersect across countries. While our findings are descriptive rather than causal, they illustrate how disparate data can be integrated to generate hypotheses about the connections between global challenges and quality of life.

The next steps in this analysis could include incorporating additional variables and more detailed measures. For example, we could add GDP per capita or income inequality measures to see whether the association between climate awareness and happiness persists after accounting for economic factors, or use per-capita or log-transformed measures of COVID-19 burden instead of raw case counts. We could also extend the analysis over time by using multiple years of happiness and COVID-19 data to explore how changes in these variables are related.

The limitations of this analysis are important to keep in mind. First, all three data sets are observational and cross-sectional, so we cannot infer causal relationships from the patterns we observe. Second, happiness and climate awareness are based on self-reported survey responses, which may be affected by sampling differences or response biases across countries. Third, COVID-19 case counts may be under-reported in some regions due to differences in testing and reporting capacity. These limitations mean that our results should be interpreted as descriptive evidence of associations rather than as proof that climate awareness or COVID-19 directly causes changes in happiness.

6 References

We use R Markdown footnotes to document our data sources and external references.
Each footnote explains how the source is relevant for a specific part of our analysis.