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, 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 relate 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 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)»

To clean the data, we were able to convert our wide data into a longer form with both a "country" variable and a "score" variable using the pivot_longer() function, after which we were able to clean the awareness category names by using recode() to turn the lengthy description variables into more condensed variable names (e.g., aware_no, aware_alittle, aware_moderate, and aware_alot). Next, we used pivot_wider() to create a dataset where there was one row per country and separate columns for each awareness category. We also checked to ensure that the base for aware had captured all of the respondents. Lastly, we exported the final cleaned data set as climate_awareness.parquet for later use in this report. [1]

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

This data is from an international dataset of daily case counts (from data provided by "Our World in Data"). This source has tracked the cumulative numbers of confirmed COVID-19 cases by location throughout the world. Within our covid_2020 dataset, every single row corresponds to a different country; and we have three major variables used for this analysis: country, continent, and total_cases, which is the cumulative number of confirmed COVID-19 cases for that country during all of 2020. [2]

In order to get this dataset into an appropriate format for analysis we began filtering the raw dataset down to only data from the year 2020, eliminated all rows that contained aggregate regional data (i.e., "world") and removed all rows which had no available continent information for the location(s) within those regions. Following that, we then combined the remaining data into the appropriate levels of aggregation (by country and by continent). We computed the max cumulative case count per country/continents combination in 2020 and saved the case counts in the total_cases column. Since there are only two major columns being analyzed in this study (country, continent) we then eliminated all the other columns and produced a new parquet file entitled covid_2020.parquet.

3.3 «Happiness data (happiness.parquet)»

The source of this data is The World Happiness Report; it is a report on the happiness of people around the world based upon self-reporting (e.g. surveys) of life satisfaction. The original data had an entry per country per year, and included each country's 'ladder' score for their happiness and also other variables about income, social support and health. The cleaned data has each entry as a country and includes the country, year and ladder score (the average life satisfaction score out of 10). [3]

---

[1] **Relevance:** Provides the country-level climate awareness percentages used to construct `share_aware_alot` and the awareness groups in Figures 1 and 3.
From Meta Data for Good, *Climate Change Opinion Survey*, accessed November 2025.
URL: https://dataforgood.facebook.com/dfg/tools/climate-change-opinion-survey

[2] **Relevance:** Supplies the 2020 cumulative confirmed COVID-19 case counts by country used for `total_cases` in Figure 2 and in the summary statistics table.
From Our World in Data, *Coronavirus (COVID-19) Cases*, accessed November 2025.
URL: https://ourworldindata.org/covid-cases

[3] **Relevance:** Provides the "ladder" life-satisfaction scores used as our outcome variable `ladder_score` throughout the analysis.

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 cleaned this data by first standardizing all the column names using janitor::clean_names() and changing the name of the country variable to country, so that the country variable would match our other data sets. We next sorted the data by country and year, and then by group by country to keep only the most recent year of data for each country. Next we eliminated any entries where there was a missing ladder_score value, and selected a subset of the variables that we thought would be useful for the analysis, including the outcome ladder_score. Last we saved the cleaned data as happiness.parquet.

3.4 Combining the Data

We will use the country variable as a key to join the three data sets together. To start we will take the happiness data and do a left join to add the climate awareness information from the climate_awareness data set, keeping all countries with their respective happiness scores when ever possible. Next we will perform another left join on the merged data set and add the covid_2020 data set. The covid_2020 data set will add a total number of confirmed COVID-19 cases in 2020 and the continent for each country. Finally we will create a new variable, share_aware_alot, and it is defined as the proportion of people responding in each country that said they knew "a lot" about climate change (aware_alot / aware_base) and this combined data table will be the one used for our summary statistics and visualization 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.

The figure shows an overall upward trend (and thus a positive association) in points and the fitted line on Figure 1 indicates a relationship between higher levels of climate knowledge among citizens, as well as their reported happiness scores.

Table 1 provides summary statistics for each of the variables, summarized by continent. The data support this insight.

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
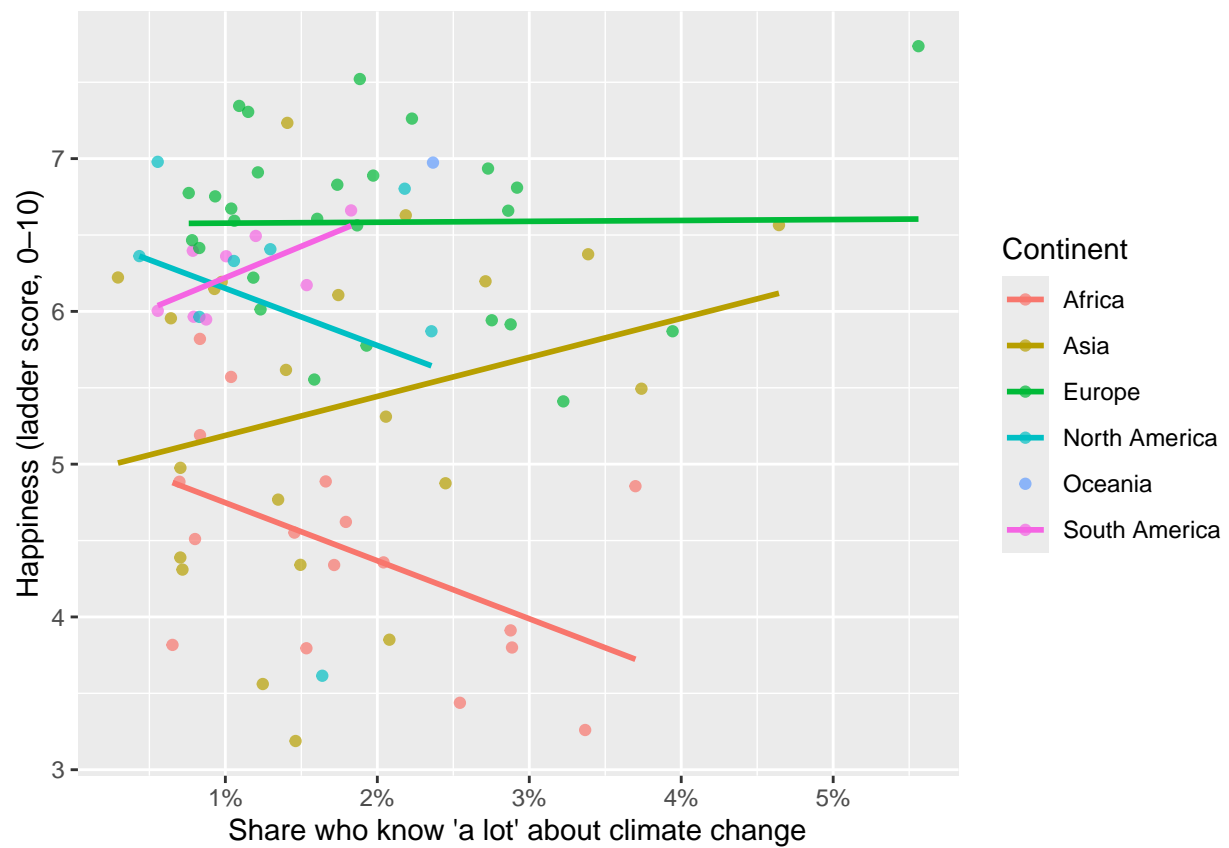
Figure 1: This is a figure caption that you will need to change in order to get good marks in the visualization rubric items.

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.
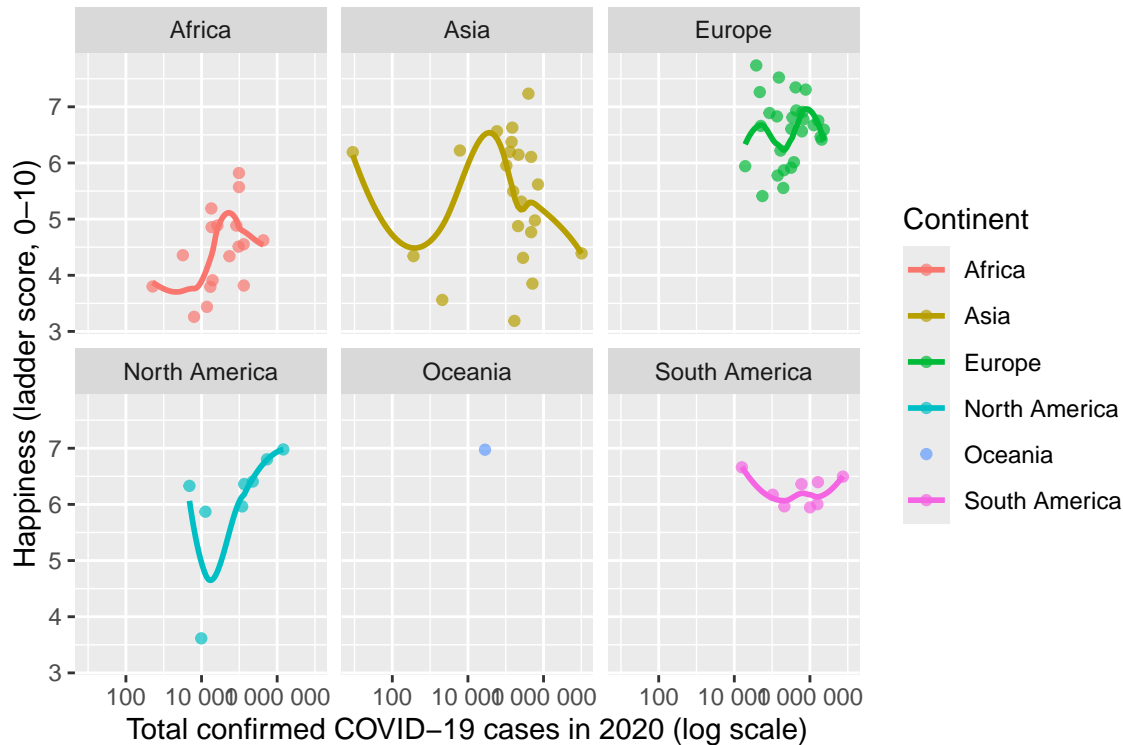

Figure 2: This is a figure caption that you will need to change in order to get good marks in the visualization rubric items.

In contrast to the overall positive relationship between climate awareness and happiness, Figure 2 shows that the strength of this association varies by continent. The upward trend is stronger for some regions, such as Europe and North America, while it is less strong or based on fewer countries in other regions. These results suggest that the relationship between climate awareness and well-being can be influenced by a country's regional context, and that other factors that were not included in our analysis (such as income, and/or inequality), may also contribute to the observed relationships.

Lastly, Figure 3 displays the average happiness among countries classified as having either high or low levels of climate awareness. We have divided all countries into three categories based on their hare_aware_alot variable, namely, Low awareness: <20%, Moderate awareness: 20-40%, High awareness: >=40%.
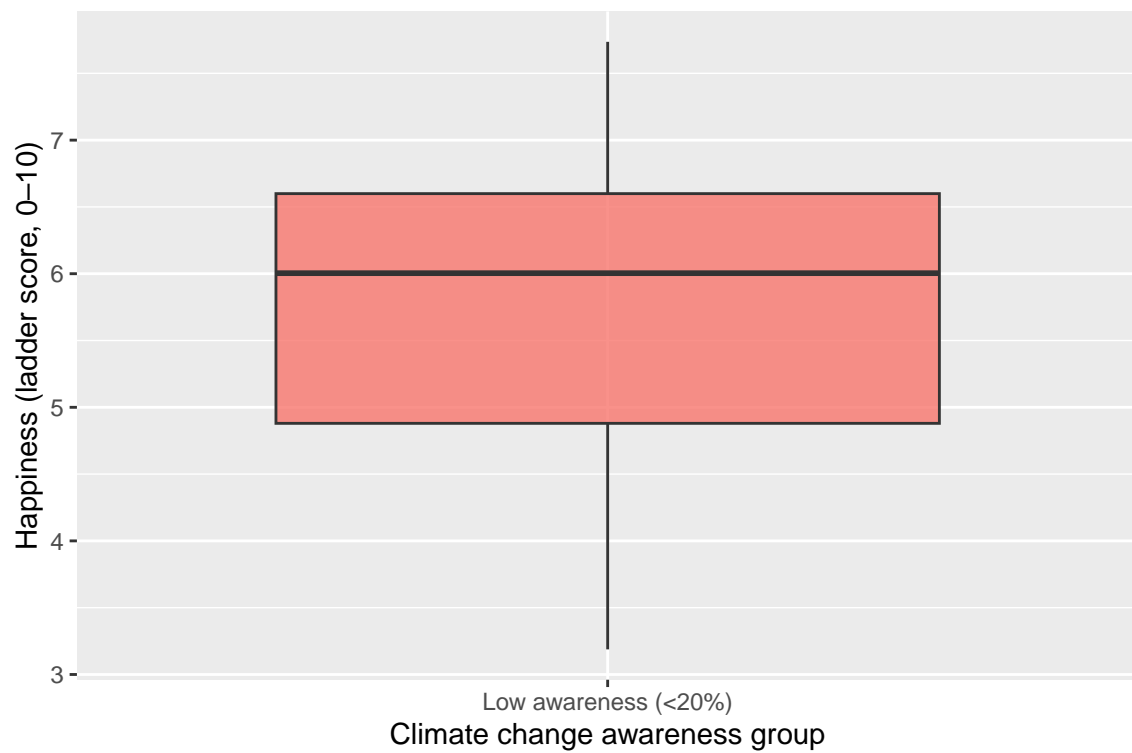
Figure 3: This is a figure caption that you will need to change in order to get good marks in the visualization rubric items.

5 Conclusion and Future Work

In general, we found that countries with a greater level of awareness regarding climate change have reported higher levels of national happiness. The observed relationship exists in both the overall scatter plot and when comparing the two higher and lower awareness groupings. Although the association between the total number of COVID-19 cases during 2020 and national happiness appeared in many of the continents in our study, the consistency and strength of the relationship differed among them. On the other hand, it appeared to be somewhat unclear if there existed an association between the total number of COVID-19 cases in 2020 and national happiness based upon our summary statistics, indicating that there may exist a variety of long-term social and economic factors that are more closely associated with variations in cross-country life satisfaction.

We concluded in addition to the above that the combination of multiple types of data allows for the identification of associations that do not appear in individual datasets. Combining climate awareness, COVID-19 cases, and happiness scores for each country allowed us to consider how attitudes, health crises, and quality of life are interrelated across countries. Our findings are primarily descriptive and therefore do not provide support for causal statements; however, they demonstrate how disparate forms of data can be combined to produce hypotheses regarding potential relationships between global issues and quality of life.

As part of future research in this area, we could incorporate new variables and/or more detailed measures into the analysis. For example, we could consider adding GDP per capita and/or income inequality measures to determine if the association between climate awareness and happiness remains significant even after adjusting for economic factors, or we could replace the raw number of COVID-19 cases with either per-capita or log-transformed measures of COVID-19 burden. Additionally, we could extend the time frame of our analysis by examining the changes in happiness and COVID-19 over time (i.e., using multiple years of happiness and COVID-19 data).

It is also very important to remember the constraints of this analysis. One constraint is that all three data sets used in this analysis are observational, and thus, we cannot draw conclusions regarding causality from the observed associations. A second constraint is that happiness and climate awareness are based upon self-reporting and as such, may be influenced by differences in sample size or response bias across countries. Finally, COVID-19 cases may have been underreported in some areas due to differences in testing and reporting capabilities. As a result, we interpret our results as simply descriptive associations rather than as direct evidence that climate awareness or COVID-19 has caused 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.