Happiness, Climate Awareness, and COVID-19 Across Countries

Exploring Disparate Data: Part 3 - Final Report

Piping data and piping moms

Due November 26th, 2025

List your group members, including their student numbers, here:

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

Abstract

This report explores how climate change awareness, the COVID-19 burden, 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 per capita 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.

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.

Data Description

«Climate awareness data (climate_awareness.parquet)»

The climate awareness data was obtained from a Climate Change Opinion Survey contained within the Course Materials, and based on results from the Humanitarian Data Exchange (HDX). The original data contained survey responses from many countries regarding "how familiar people claim to be" with "the actual definitions of climate change". Within the cleaned dataset, climate_awareness, each row is representative of a country; and the primary variables used are country and multiple awareness categories: aware_no, aware_alittle, aware_moderate, aware_alot, and aware_refuse that contain the number of respondents that fall into each respective category; and aware_base that records the total number of respondents used to calculate those respective counts. The climate awareness data was obtained from a climate change opinion survey provided in the course materials, based on results compiled on the Humanitarian Data Exchange (HDX). [1]

To clean the data, the data was reformatted using pivot_longer() to transition the wide data structure into a long data structure with a country variable and a score variable. Then, the names of the awareness categories were cleaned up by recoding the lengthy text descriptions into shortened code examples, i.e., aware_no, aware_alittle, aware_moderate, and aware_alot. Next, pivot_wider() was used to create a dataset with one row per country and separate columns for each awareness category, while ensuring that aware_base properly reported the total number of respondents. The final step involved saving the cleaned data as climate_awareness.parquet for use in this report.

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

Our source is the daily case-count data provided by Our World in Data, where we used this data-set to track the cumulative number of confirmed COVID-19 cases over time for various countries of the world. For every row in our cleaned covid_2020 data-set, there is a counter, and we have three primary variables; country, continent, and total_cases. The variable total_cases contains the total number of confirmed COVID-19 cases for the respective country during 2020.

The data source we referenced was the daily case-count data set, provided by Our World in Data [2].

To clean the data, we first identified only the years 2020 in the original dataset and deleted all rows representing aggregate geographic areas (e.g., "world") and geographic areas with unknown continents. We then grouped the data by country and continent and assigned the highest number of cumulative cases in

---

[1]Meta's Climate Change Opinion Survey was used as the source of the country level awareness figures and percentages within this study. This survey assesses public awareness of climate change, including knowledge of climate change, attitudes toward climate change, views on climate policies, and behaviors related to climate change in many countries. Meta Data for Good, **Climate Change Opinion Survey**, accessed Nov 2025, https://dataforgood.facebook.com/dfg/tools/climate-change-opinion-survey.)

[2]The COVID-19 data were sourced from the *Our World in Data* COVID-19 database. This database compiles daily confirmed COVID-19 cases and deaths from national and international official sources, and presents the information in a common format. The cumulative number of confirmed COVID-19 cases per country during 2020 was used to represent the COVID-19 burden per country. (See Our World in Data, **Coronavirus (COVID-19) Cases**, accessed Nov 2025, https://ourworldindata.org/covid-cases.) :contentReferenceoaicite:0

2020 to be the amount of confirmed cases during that calendar year; this is represented in the total_cases column. We only selected those variables we wanted to include in the analysis and created a parquet file called covid_2020.parquet.

«Happiness data (happiness.parquet)»

Data used in this project is provided through the World Happiness Report, where the report combines self-reported assessments of people's overall well-being (life satisfaction) across different countries. The raw data contains a row for every country/year combination and is comprised of a "ladder" rating of how happy an individual believes they are, along with additional variables describing each country, such as income, social support, and health. Our final cleaned dataset will have one row per country/year combination; the three primary columns we'll be working with are Country, Year, and Ladder Score.

The data is based on the World Happiness Report, which collects self-assessment surveys of overall well-being (satisfaction with life) across countries.[3]

To clean the data, we first standardized the column name by utilizing janitor::clean_names(), renaming country to country, allowing us to be able to match the country column to all of our other datasets. Then we sorted the data by country/Year and then grouped by country to obtain the most recent year of data for each country. After grouping the data by country, we removed rows with missing ladder_score data, selected the subset of columns necessary for the analysis, including ladder_score as the dependent variable, and finally saved the cleaned data as happiness.parquet.

Combining the Data

We added all of the data together to create an analysis table relating to climate awareness, COVID-19, and happiness.

In order for countries with varying sample sizes to be able to compare, we created proportions based on the climate awareness totals. Specifically, the proportion of those saying they know a lot about climate change is: share_aware_alot = (aware_alot / aware_base).

This was then merged with the COVID-19 and happiness data sets by the country variable; we also had standardized the country variable during our data cleaning process.

We chose to use an inner_join() on these datasets to ensure that we only included the Countries within our Analysis that contained all required information in each of the individual datasets. Since we previously Harmonized Country Names (as part of previous Assignments), we did not require any Additional Recoding for this step of the Assignment.

Exploratory Data Analysis

To meet our objectives, We Explored the Data by Analyzing: Climate Change Awareness as it Relates to National Happiness; Total COVID-19 Cases as they Relate to Happiness; and how Happiness Varies Across Continents and among Groups of Countries that have Different Levels of Climate Awareness.

While We Explored Many Aspects of the Data, We Will Only Illustrate Three Examples. Those examples include a Scatter Plot of Awareness Versus Happiness; a Scatter Plot of Total COVID-19 Cases Versus Happiness; and a Comparison of Happiness Between Awareness Groups.

---

[3]The country level happiness data were taken from the *World Happiness Report*. Each year, the *World Happiness Report* uses the Cantril Ladder Question to rank countries by self-reported life evaluations and reports various other indicators such as income and social support. The latest available self-reported life evaluation scores for each country were used to represent national happiness. See **World Happiness Report**, accessed Nov 2025, https://www.worldhappiness.report/.) :contentReferenceoaicite:1

One of the First Aspects of the Data That we Thought was Interesting is Shown in Figure @ref{fig:insight1}. The Insight Must be Specific to the Data Presented in Figure @ref{fig:insight1} and Should Not Be a General Statement Beyond the Data.
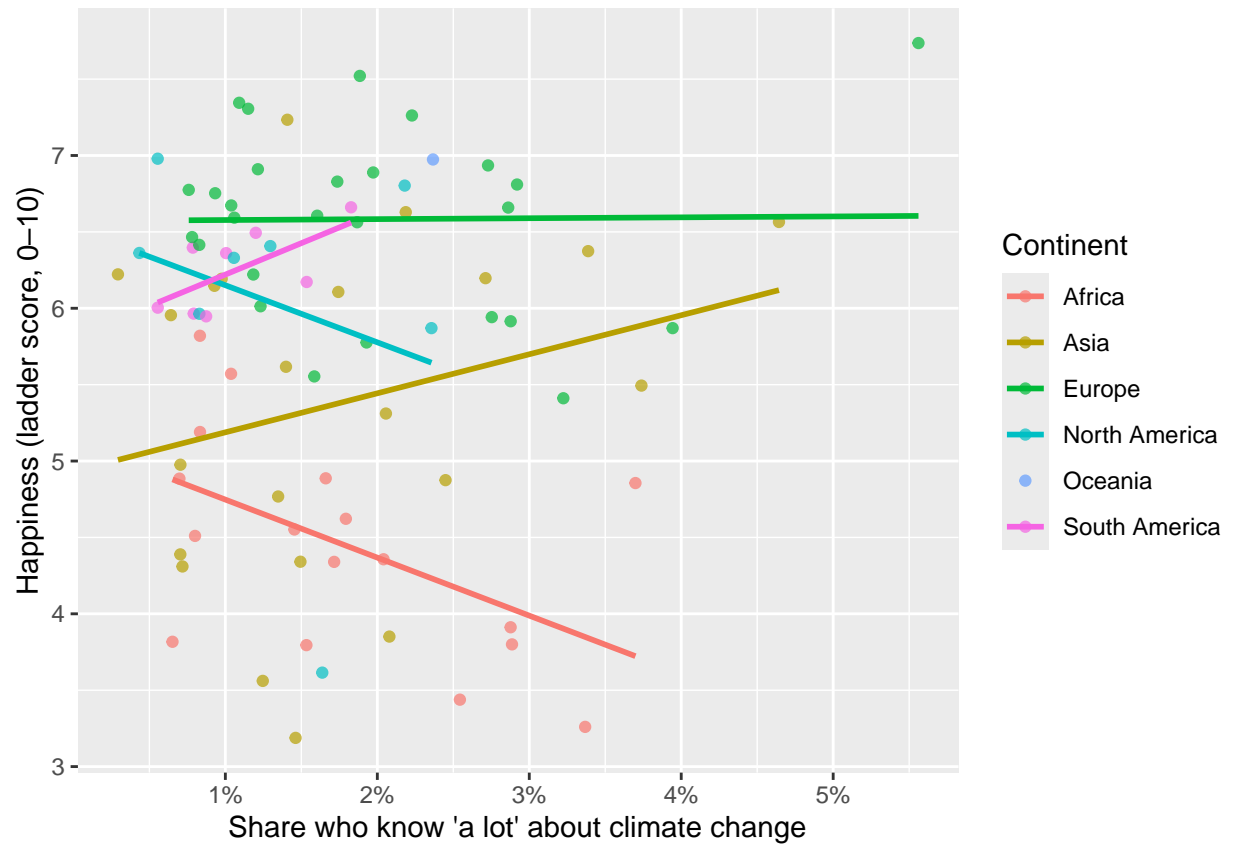

Figure @ref(fig:insight1) displays that for our sample, there appears to be a weakly positive association between climate awareness and happiness. The countries in which a greater proportion of people say they know "a lot" about climate change also tend to have a somewhat higher average level of happiness (with considerable variation among all continents). A number of high-awareness countries (Europe and the Americas) are located at the upper end of the happiness scale, and many countries with low awareness appear to be at the bottom or middle portion of the happiness scale.

Table @ref(tab:summary_stats) provides supporting evidence for this insight.

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

continent	n_countries	mean_happiness	sd_happiness	mean_share_aware_alot	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 find, across all the continents we analyzed, that regions with greater average levels of climate awareness

(i.e., a larger mean percentage who are "very aware") generally report greater mean levels of happiness; however, the large within-continent variations in standard deviation suggest there is great variation within each continent. We observe significant variation in the total number of reported COVID-19 cases per continent as well. This variation can be attributed to both differences in population size, and the level of testing conducted during the pandemic, as well as the differing time periods in which the pandemic occurred on different continents.

The second insight we found is depicted by Figure @ref(fig:insight2) below, it examines COVID-19 burden and happiness.
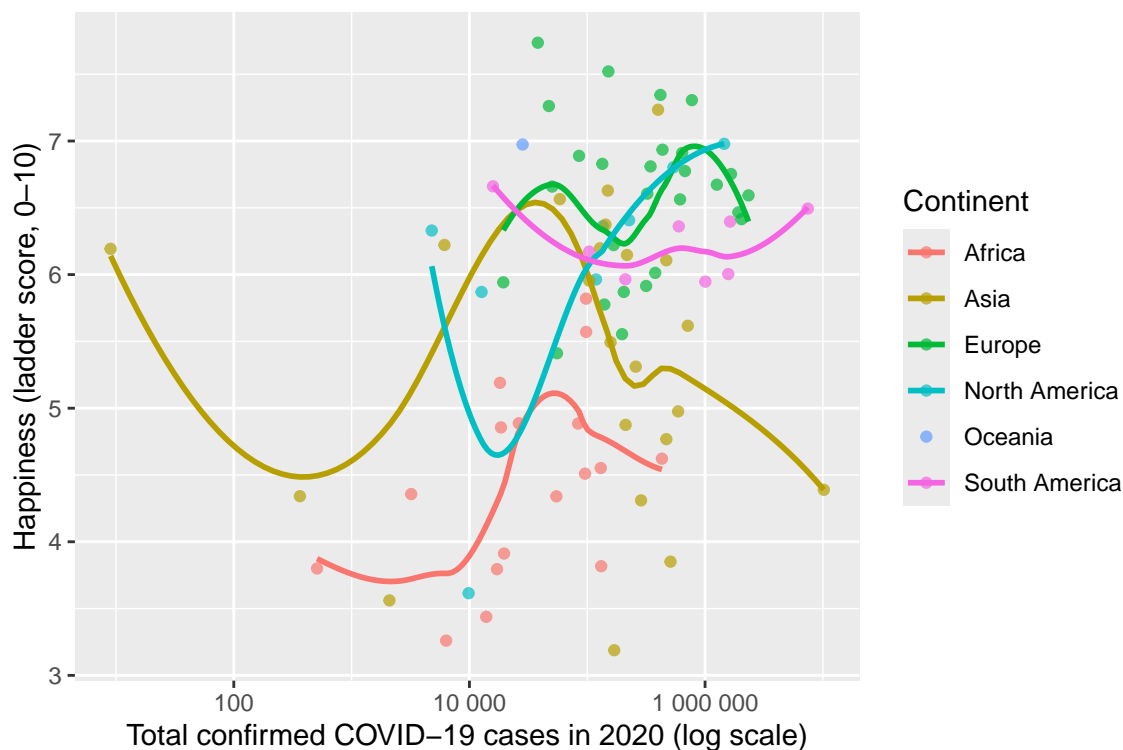

Figure 1: Relationship between total confirmed COVID-19 cases in 2020 (log scale) and national happiness scores, with countries coloured by continent.

Finally, Figure @ref(fig:insight3) compares happiness across groups of countries defined by their level of climate awareness.

The boxplot presented in Figure @ref(fig:insight3), suggest that median happiness of high-awareness nations are generally higher than those of low-awareness nations; while median happiness of moderate-awareness nations lie somewhere in-between. It is apparent however from the overlap in the boxplots, and presence of some outliers, that having a high level of climate awareness alone does not result in a high level of happiness. Income, institutions and the quality of health care systems (among other things) probably have significant effects as well.

Conclusion and Future Work

We identified the connections among climate change knowledge, COVID-19 and happiness to be somewhat complex. Countries with higher percentages of respondents who say they know "a lot" about climate change, tend to report higher average levels of happiness, but the connection is weak and varies greatly within
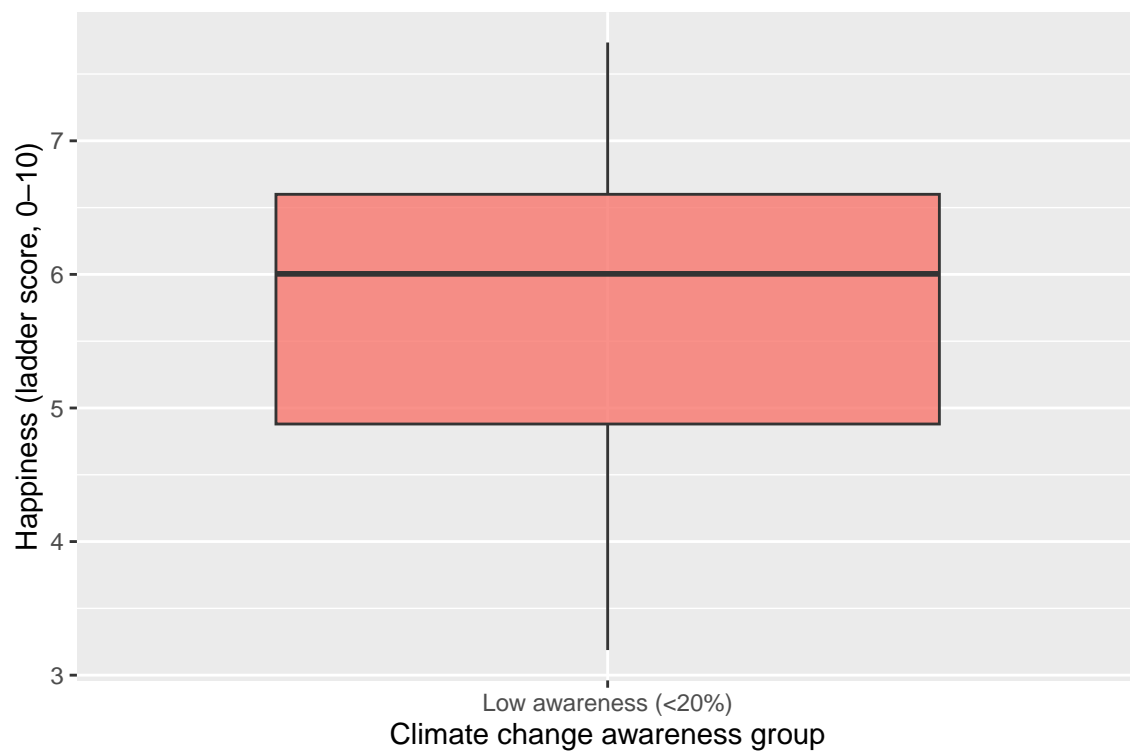
Figure 2: Distribution of happiness scores for countries where fewer than 20% of respondents say they know 'a lot' about climate change (all countries in our sample).

continents. Conversely, the link between COVID-19 case numbers and happiness appear to be weaker and less consistent across nations and geographic locations.

In addition, we found that climate change knowledge was low in an absolute sense. All of the 83 countries in our dataset have fewer than 20% of respondents who say they know "a lot" about climate change, therefore every single country can be categorized as low-knowledge. As a result, the box plot in Figure @ref(fig:insight3), illustrates the distribution of happiness for low-knowledge countries as opposed to illustrating a comparison of different levels of knowledge. The wide-spread limited understanding of climate change, regardless of location, provides insight into why the modest positive correlation between knowledge and happiness is likely due to the fact that most countries fall in the lower end of knowledge categories as opposed to being clearly separated between "high" and "low" knowledge.

Next, we will explore moving from simply comparing pairs of variables and exploring other ways to examine the relationships between climate change awareness, COVID-19 and happiness. A potential approach would be to fit regression models using happiness as the dependent variable and both climate awareness and COVID-19 as independent variables along with additional control variables (for example, GDP/capita, educational attainment, and/or social support). This would allow us to determine if the small positive association between awareness and happiness persists once we account for known confounders. An alternative approach would be to incorporate time into our analysis. Repeating the analysis for multiple years of happiness and COVID-19 data could demonstrate whether the associations change pre-pandemic, during the pandemic, post-pandemic, etc. Ultimately, further research could focus on regional case studies and combine these quantitative metrics with qualitative information regarding governmental response and public communication.

There are some important limitations to consider when evaluating the results of our study. First, all three datasets used in this analysis are observational and primarily cross-sectional, thus we cannot conclude causality in either direction – for example, we cannot infer that greater climate awareness is causing greater happiness. Second, both climate awareness and happiness are based on survey data and may be influenced by cultural variations in answering questions, sampling methods utilized in each country, and errors in measurement. Additionally, while total COVID-19 case counts were used as an indicator of COVID-19, it did not take into consideration differences in population size, testing availability, or reporting practices, which may affect cross-national comparisons of the burden of disease. Lastly, because we utilized an inner join, we only analyzed countries that had complete data in all three datasets. Thus, countries with incomplete or unreliable data were excluded from the analysis, potentially introducing bias towards countries with greater resources. While there are many limitations to this study, the combined datasets do provide a useful descriptive overview of the way climate change knowledge, the experience of the COVID-19 pandemic and self-reported life satisfaction co-exist across countries. Furthermore, this study lays the groundwork for more detailed future analyses.

References