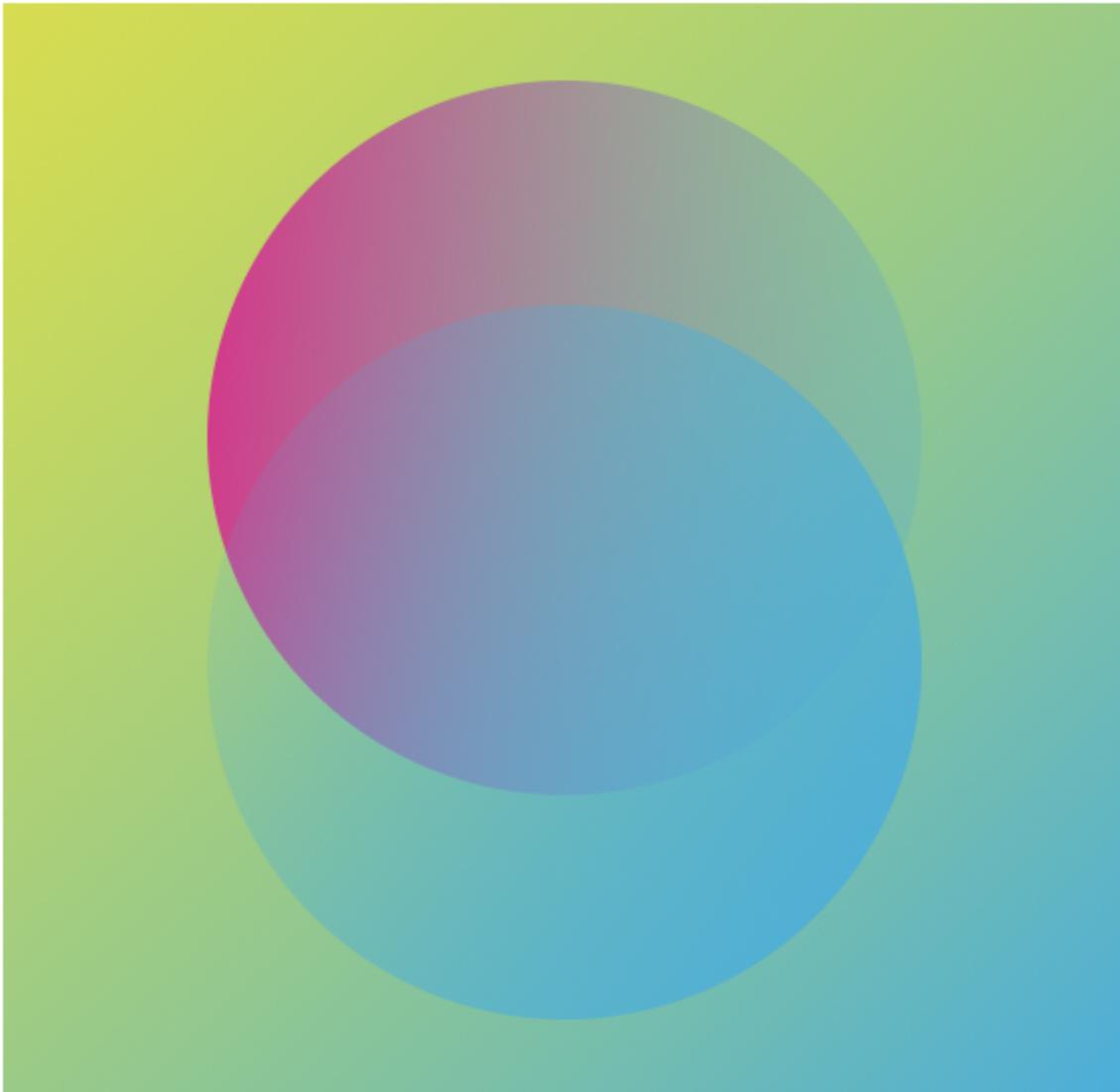


# **World Happiness Report 2021**



**DSC 465 Group 8 Final Project**

Azalea Mendoza, Shivangi Singh, Beeta Sabbithi

# Table of Contents

[Table of Contents](#)

[Introduction](#)

[Exploratory Analysis](#)

[Visualizations](#)

[Analysis and Discussion](#)

[Appendices](#)

[Datasets](#)

[Code](#)

[Individual Components](#)

[Azalea](#)

[Shivangi](#)

[Beeta](#)

# Introduction

For a year that was unlike any other and one which conjures up numerous definitions, putting together the emotional and physical experiences of the year 2020 is a challenge not just now, but most likely will be for several years to come. Nevertheless, our group wants to give readers a mere glimpse into this year using the World Happiness Report datasets.

Our group's overall goal for the project was to visually present findings from the World Happiness Report. Each group member chose to focus on a specific aspect of the data:

- How happiness is affected based on the social factors that were used to measure the happiness score (Ladder Score).
- Deaths and overall life satisfaction measurements related to Covid-19.
- Exploring reasons for happiness in some countries and unhappiness in others.
- Comparing 2020 with previous years.

The dataset we obtained from the World Happiness Report is based on the Gallup World Poll. The polling uses a concept called the Cantril ladder<sup>1</sup>. Respondents are asked to imagine a ladder and asked to rate their life between 0 and 10: a rating of 10 would be considered the best life possible, while a 0 rating would be considered the worst life possible. Each country's responses are then consolidated and translated into a Ladder Score.

The World Happiness Report compares the Ladder Score against six key social indicators, which were determined to be the most likely factors to influence a country's happiness. Their definitions as explained by the Report are included.

- Logged GDP per capita :
- Perception of corruption: *Average from binary (0 or 1) questionnaire questions “Is corruption widespread throughout the government?” and/or “Is corruption widespread throughout business or not”*
- Social support: *Average from binary (0 or 1) questionnaire question, “Do you have relatives or friends you can count on to help you whenever you need them, or not?”*
- Generosity: *Average from binary (0 or 1) questionnaire question, “Have you donated money to a charity in the past month?”*
- Healthy life expectancy: *Based on data provided by the World Health Organization*
- Freedom to make life choices: *Average from binary (0 or 1) questionnaire question, “Are you satisfied or dissatisfied with your freedom to choose what you do with your life?”*

---

<sup>1</sup> <https://happiness-report.s3.amazonaws.com/2021/WHR+21.pdf>

The World Happiness Report included three datasets:

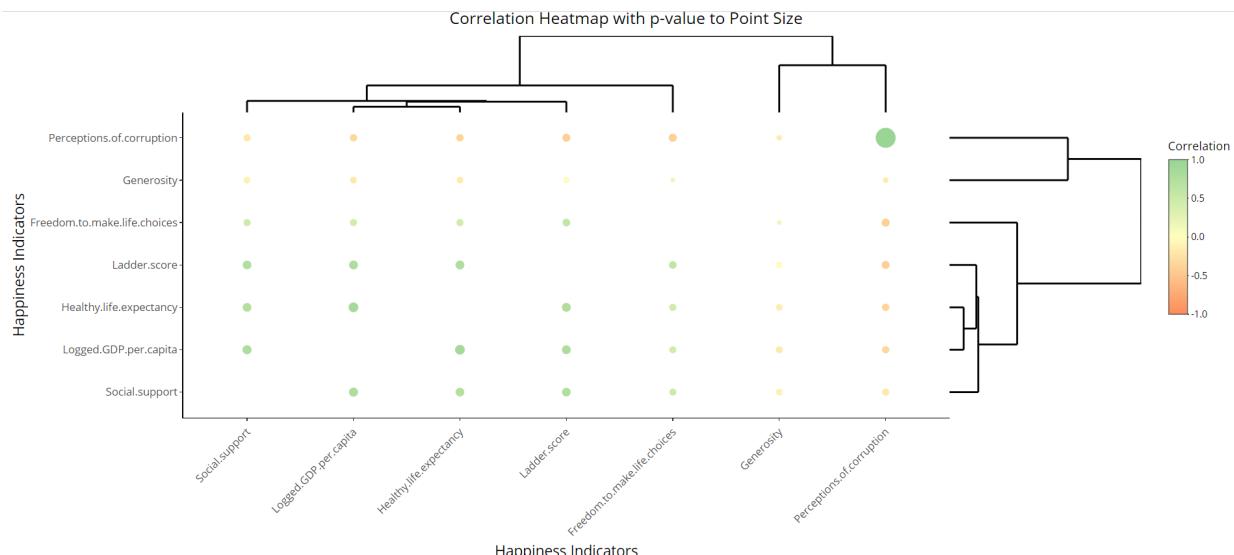
1. Data2020 - Data specific to the year 2020
2. DataPanel - Data from previous years, dating back to 20 years ago
3. Mortality Data - Data about COVID-19 throughout the first year of the pandemic, collected by OurWorldInData and utilized by The World Happiness Report

Although the rankings are based on polls taken from a representative population from each country, one likely limitation of the dataset can be that it does not capture diverse economic samples. A shortcoming of the dataset DataPanel that our group noticed, were gaps in the yearly rankings for some of the countries. Not all countries had rankings every year, with the majority of the countries within low-economic standing regions.

## Exploratory Analysis

Our exploratory analyses of the datasets consisted of scatter plots with trend lines, heatmaps, and box plots. Some of the directions we wanted to explore were, correlation between measured factors, significant influencers of happiness, excess deaths due to Covid-19, and how measurements changed from previous years due to the effects of Covid-19.

### ***Correlation Matrix between measured factors***

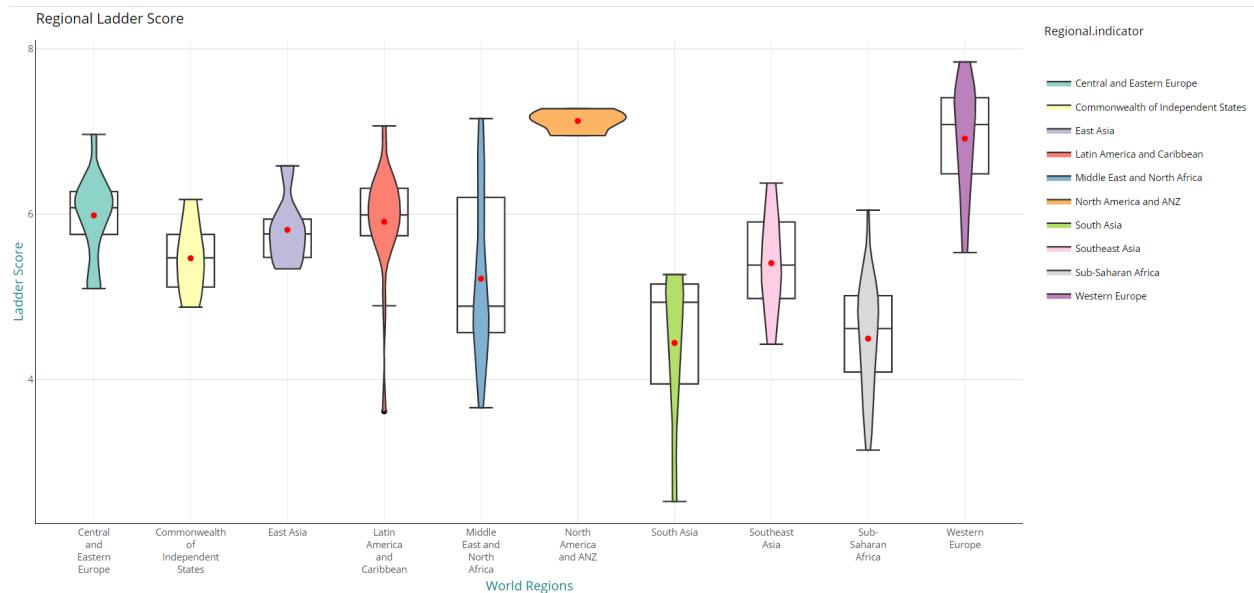


(Beeta)

The above correlation matrix measures dependence between the six social indicators and

Ladder score. In addition, the measures are sized to the p-value<sup>2</sup>. The matrix showed some indicators had higher correlation to “happiness” ratings. One surprising observation from the above matrix was a slight correlation between “Generosity” and “Freedom to make life choices.”

### **Regional Ladder Score rankings (violin plot with median scores)**

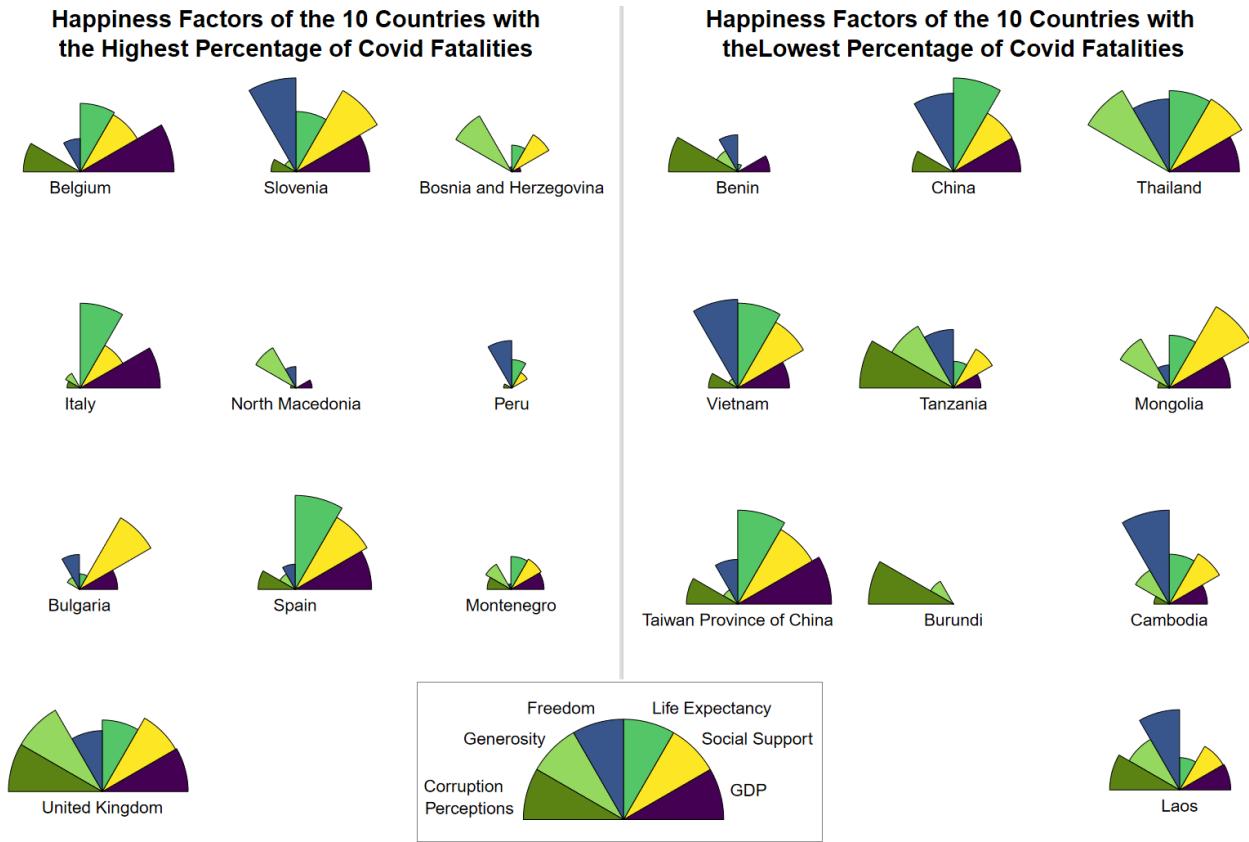


(Beeta)

A violin plot was used to show median Ladder Score in the 10 identified regions<sup>3</sup>. Our question was: do specific regions that are considered “developed,” have a higher happiness ranking than its underdeveloped counterparts? Although there were some variations, specifically in Southeast Asia with very low ranking countries mixed with higher ranking countries, the majority of the developed regions had a stable median Ladder Score.

### **How big an effect does a country’s happiness factors have on COVID-19 mortality rates? (Azalea)**

<sup>2</sup> Interactive graph “Correlation Matrix” included in zip file folder “Beeta”  
<sup>3</sup> Interactive graph “Regional violin plot” included in zip file folder “Beeta”

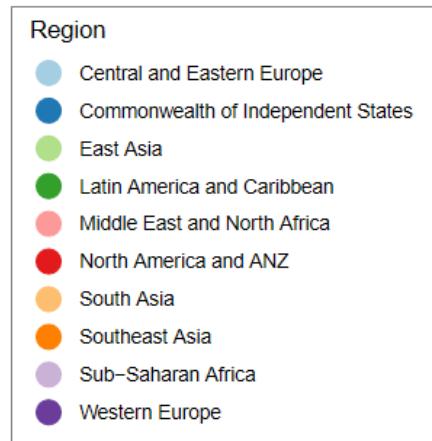
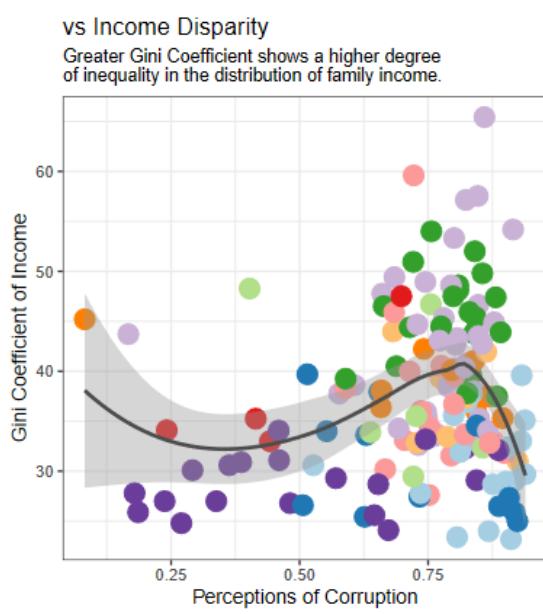
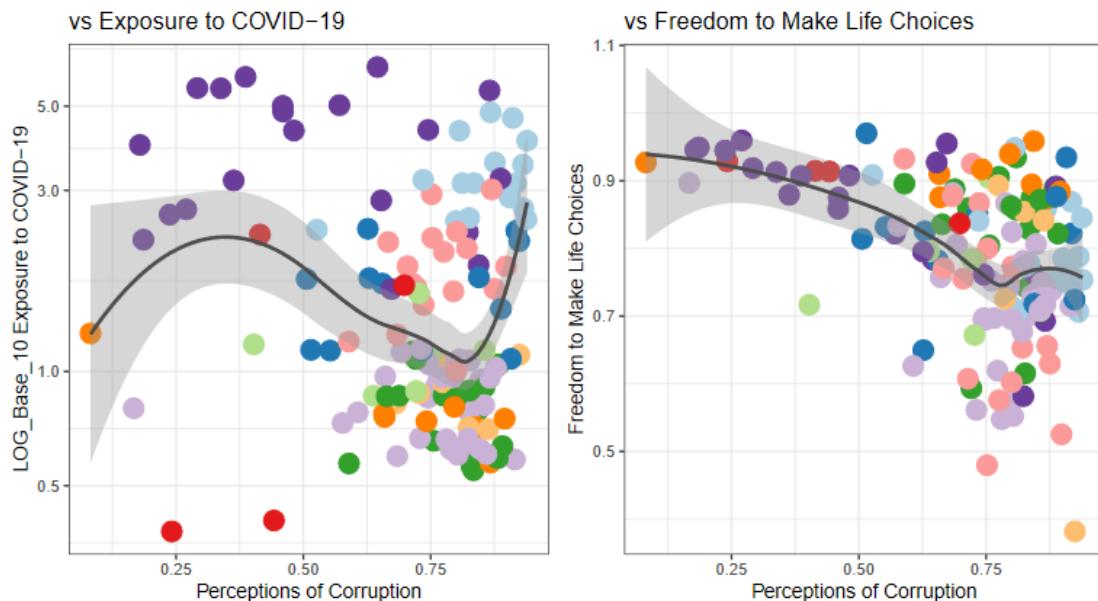
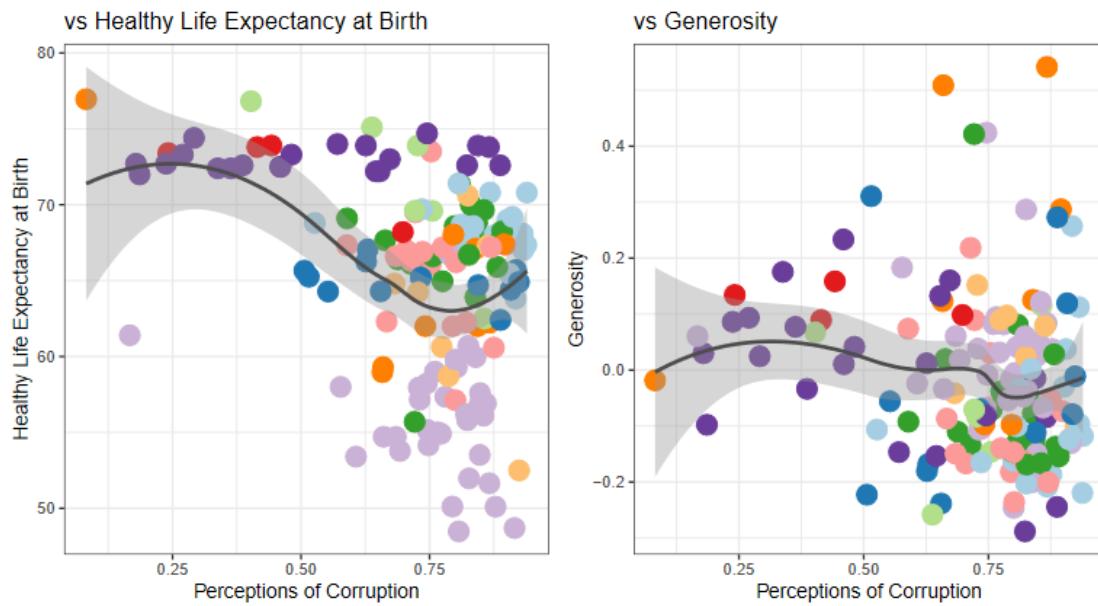


The first graphs to be distributed were country-specific, with star graphs showing the countries with the lowest and highest ratios of COVID fatalities per population and comparing the happiness factors. However, these graphs showed no correlation between countries in either the top or bottom, as it does not include factors that affected COVID-19 rates, such as exposure indices. This led to the creation of a star plot that showed the same variables, but grouped by region, with COVID-19 factored in.

### **Perceptions of Corruption versus other happiness factors, and COVID-19 exposure levels. (Azalea)**

These next visualizations are from a set of hypotheses that a country's perceptions of governmental and business corruption will affect that country's other happiness scores, such as their generosity and their perception of freedom to make life choices. It was also hypothesized that countries with low healthy life expectancy rates would have more perceived corruption, while high levels of income disparity would coincide with high corruption perception. As an additional exploration, the country's exposure indices to COVID-19 was also tested against perception of corruption. The issues that immediately arose with these hypotheses is that perception of corruption clustered globally in the 0.80 range, meaning that correlations would be difficult to find and visualize. For this reason, `geom_smooth` in R was utilized to assist the eye. However these graphs as a whole were decided to not be useful in the visualitative storytelling portion of the report.

## Perceptions of Corruption vs ...



Healthy Life Expectancy at Birth has a marked pull down as Perception of Corruption increases, and the addition of the regional indicators helps us see that this was caused by a markedly lower life expectancy in Sub-Saharan African countries as a whole.

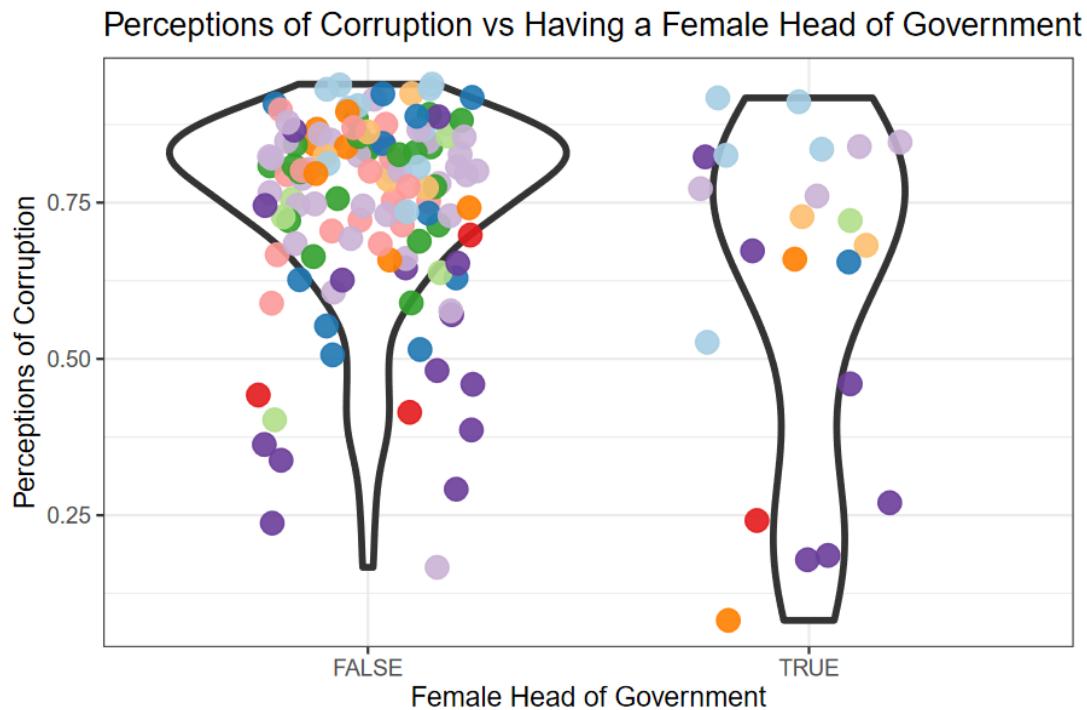
There was no clear relationship between corruption and generosity.

The relationship between exposure to COVID-19 and corruption is none, as there are too many geographical outside factors at play.

Freedom to Make Life Choices and Corruption Perception had the best correlation of the group, with a marked decrease in a country's feeling of life freedom as their perception of corruption increased. This is also the closest one of these graphs is to the hypothesis.

The Gini Coefficient of Income, which I hypothesized to have the clearest relationship to perception of corruption, did not live up to the hypothesis, but still came out to be a very interesting graph. We can see that regionally, Latin America and Sub-Saharan Africa have extremely high levels of Income Disparity, and high Perception of Corruption, as the hypothesis expected. Meanwhile, Central and Eastern Europeans, and those in the Commonwealth of Independent States, like Russia, have the most aggressive ratings towards Perceptions of Corruption. Western Europeans, meanwhile, are living the good life.

As a special interest, one last relationship was explored here, Perceptions of Corruption and having a Female Head of Government.

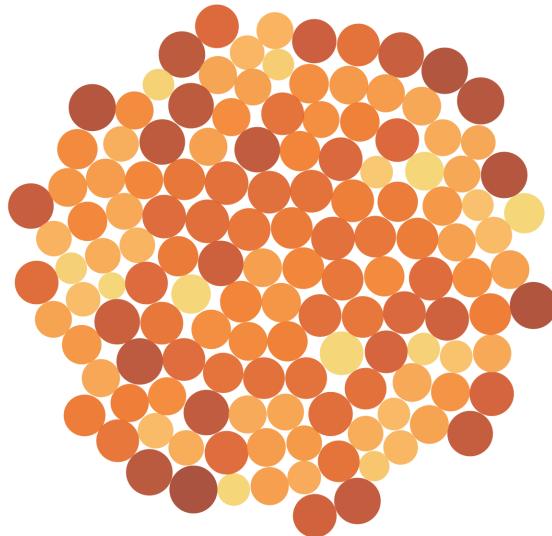


While it's no surprise that there are very few countries with women in charge, it's nice to see that the lowest corruption ratings occur in countries with female leaders. However there is still a cluster of high corruption perceptions in female-leading countries, and there is not enough data from the TRUE section to show any real significance between the two...

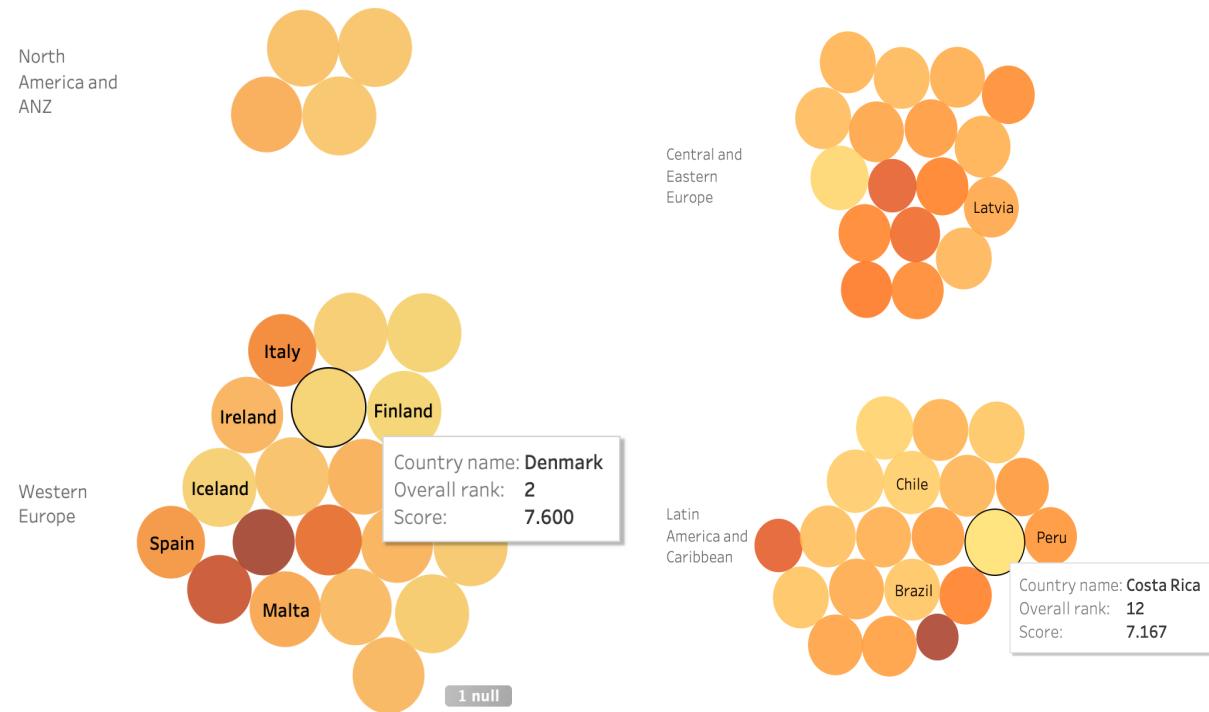
These difficult to describe relationships between Perceptions of Corruption and other factors were largely confusing due to the clustering of corruption perception by the entire global

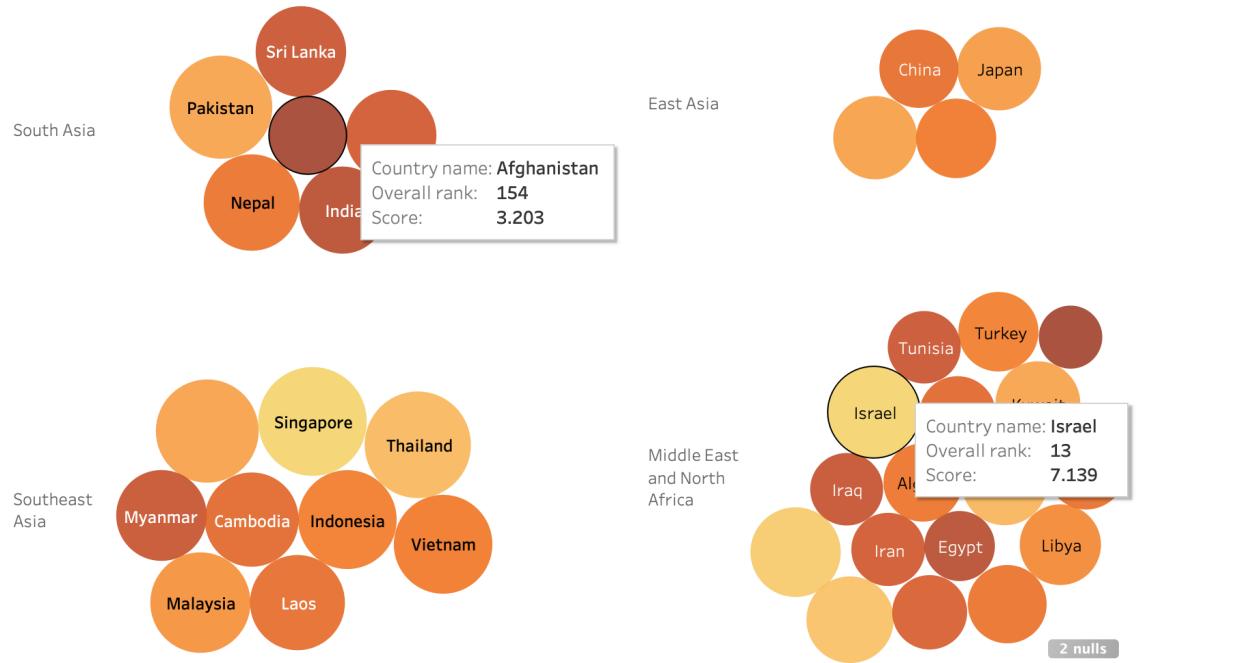
population. For this reason, in the visual & analysis section, perception of corruption was instead chosen to be shown over a 10-year period.

### ***Let's understand the world happiness Index... (Shivangi)***



Overall view of all countries together





2.8

7.7

Western Europe and North America dominated the world happiness report, with the US at 19th global rank.

Western Europe's dominance of world happiness was unshaken by the global pandemic.

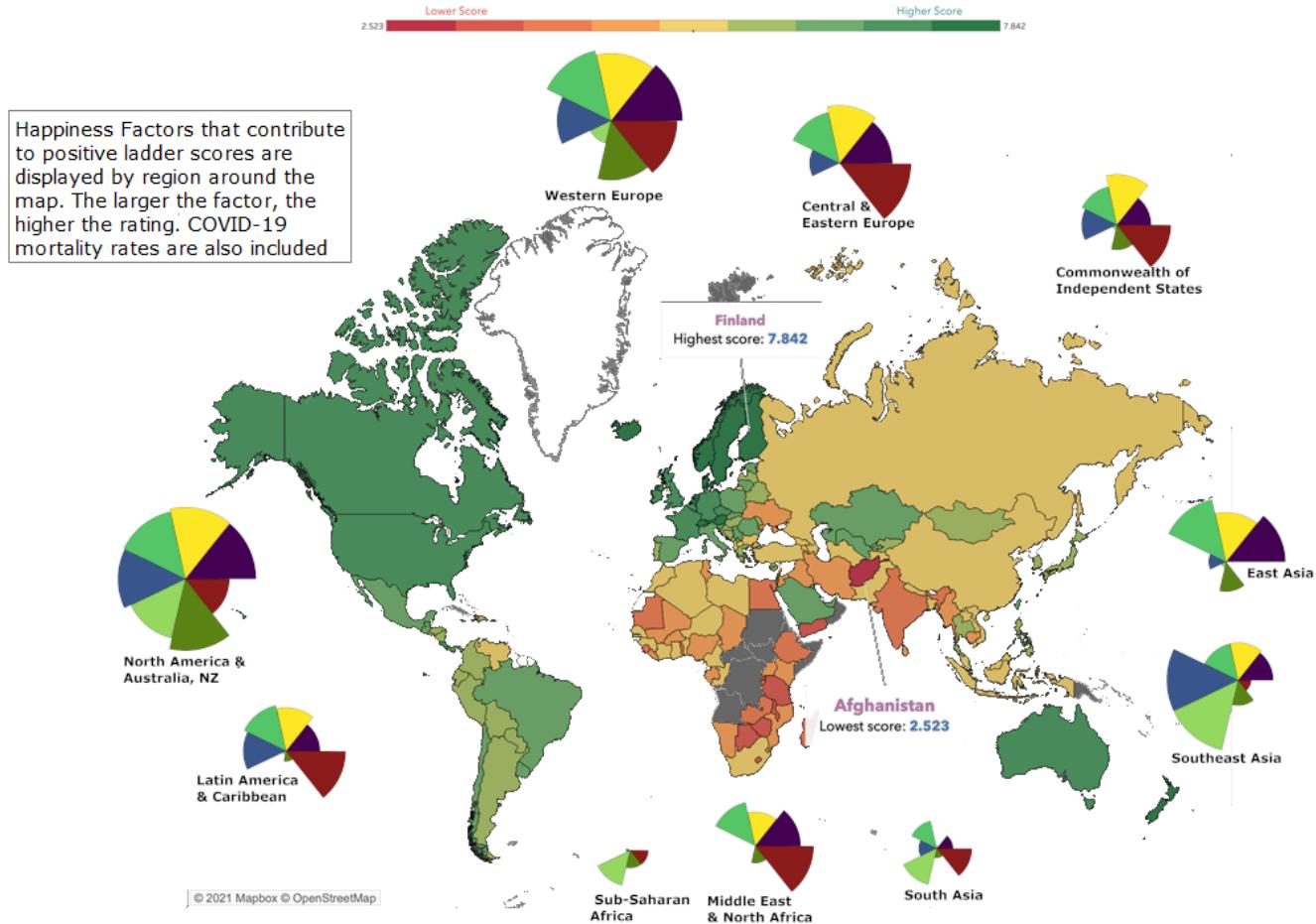
Free healthcare, education, absence of corruption in government, freedom to live your lives and High GDP are the primary factors to evaluate life satisfaction. Countries with higher happiness scores tend to have pretty good performance in these six variables.

Evidently, happiness flourishes in Nordic countries and some Western European countries, and in North America, Australia and New Zealand. The U.S. ranks 19th in the world ranking. Despite having a high GDP, it lags behind in socio-economic and inequality is too high in the U.S.

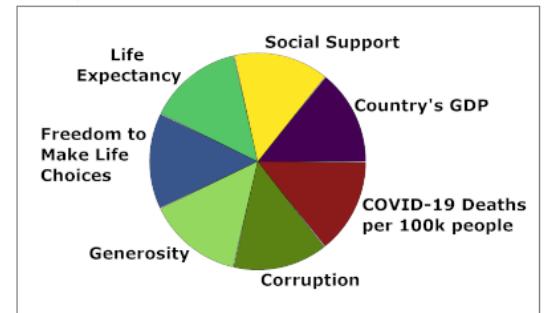
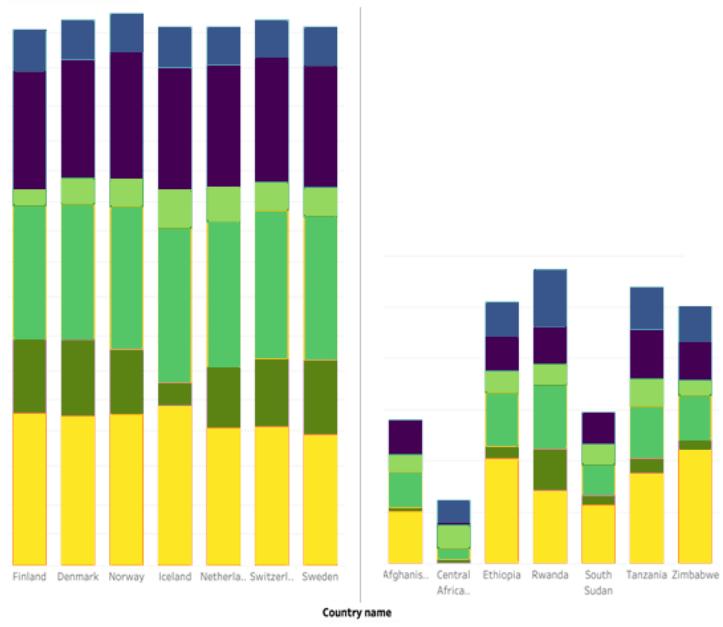
Sub-African, South Asian, Middle Eastern and some Southeast Asian countries are quite unhappy. The least happy country in the World Happiness Report of 2020 is Afghanistan which is in the south asian region. Though, there are few exceptions. For instance, Israel is surrounded by countries with an average happiness score of 4.4.

*Infographic*

## Happiness Ladder Score by Country



### Scores of Happiest Factors from the Happiest vs Least Happy Countries



### Definition of Factors

**Life Expectancy:** based on data extracted from the World Health Organization

**Social Support:** average from a 0 to 1 questionnaire "do you have relatives or friends you can count on to help you whenever you need them, or not?"

**Country's GDP:** purchasing power parity from the October 2020 update of World Development Indicators

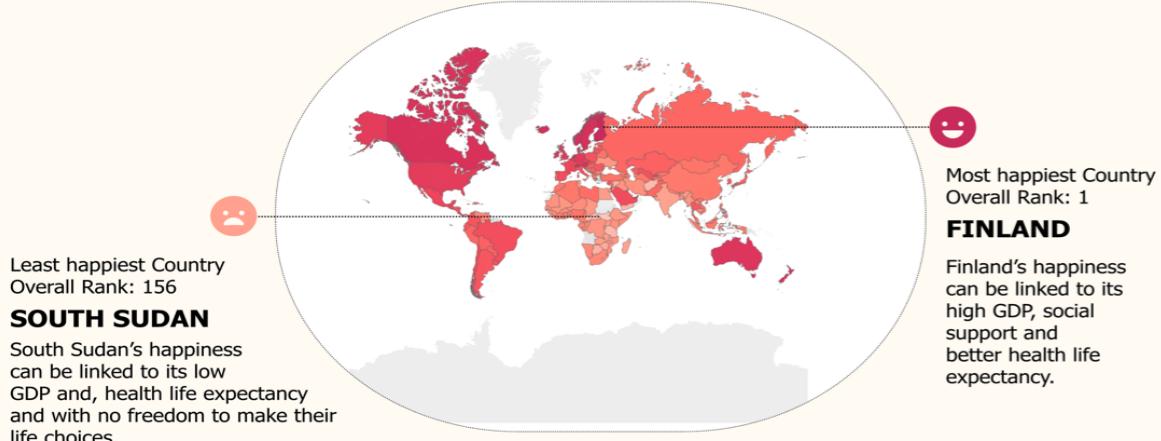
**COVID-19 Deaths per 100k People:** data from the complete Our World in Data COVID-19 dataset for 2020

**Corruption:** average from questionnaire "Is corruption widespread throughout the government or not?" and Is corruption widespread throughout business or not?"

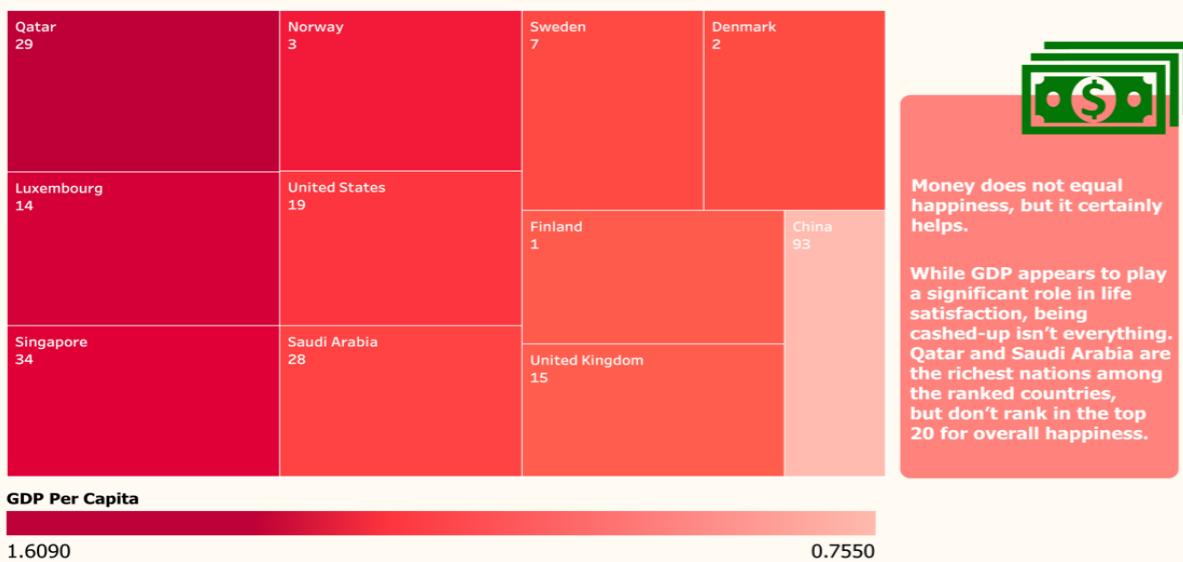
**Generosity:** national average of response to "Have you donated money to a charity in the past month?"

**Freedom to Make Life Choices:** national average of responses to "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?"

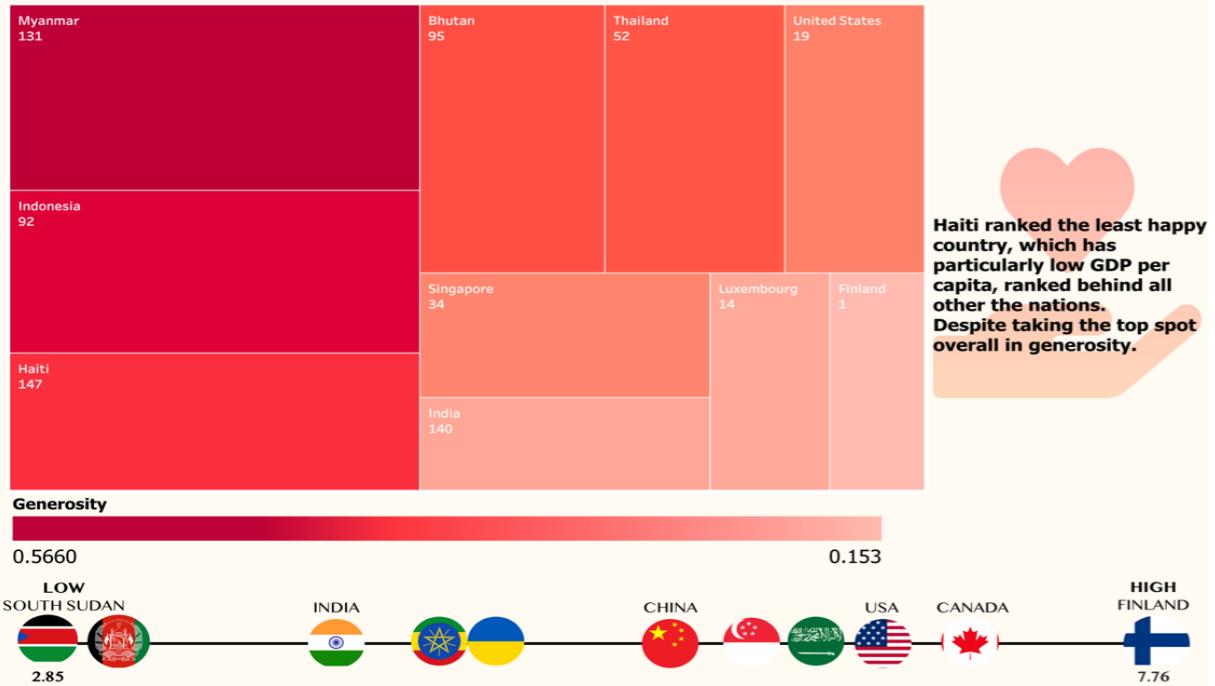
## Does Wealth And Generosity Affects Happiness Index?



### Wealth isn't everything!!



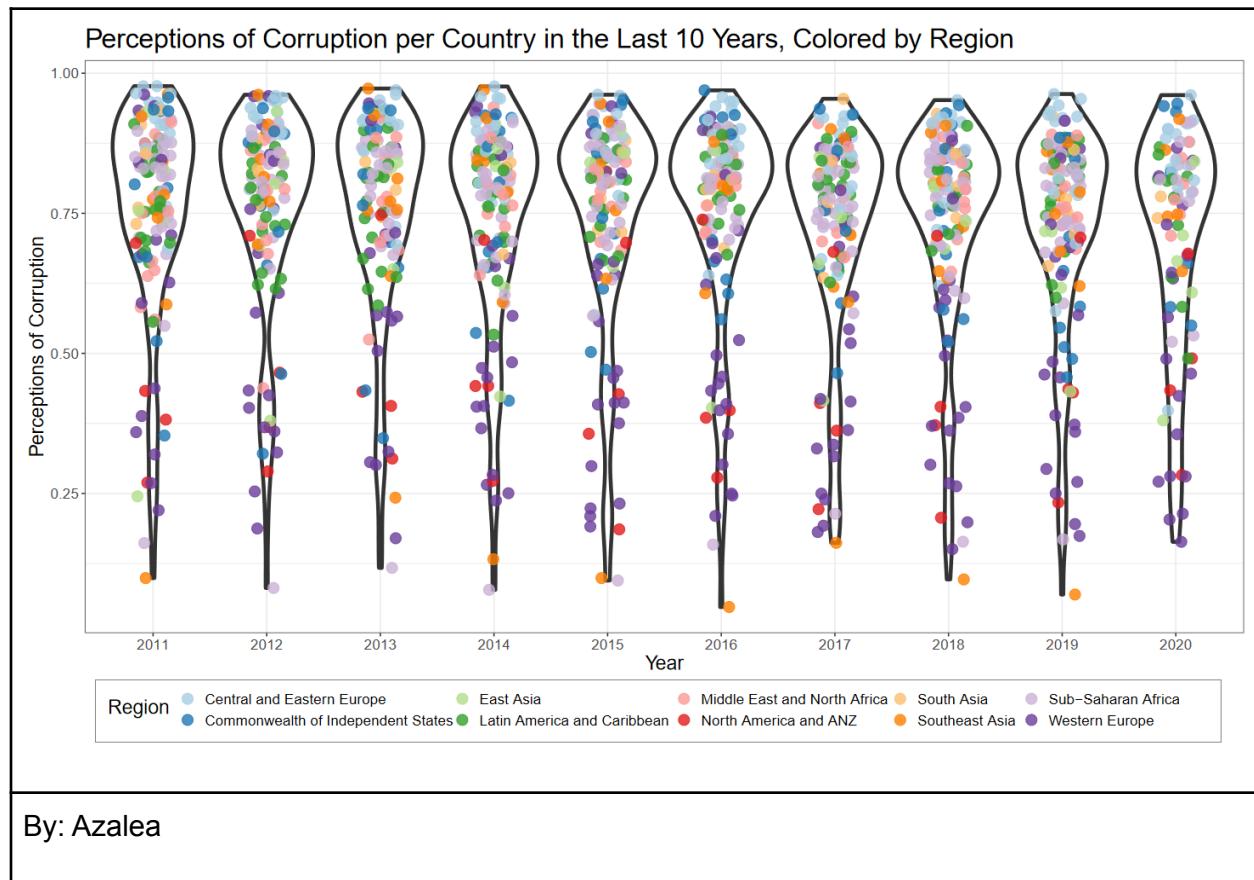
**Generosity is even more valuable currency. Irony is being the most happiest country from last 10 years is lagging in being kind.**



# Visualizations

## Global Perceptions of Corruption

In the final analysis of Perceptions of Corruption, a time series of a global scale proves to be most interesting. Because so much clustering appeared in the 2020 data, and due the large number of data points per year, a violin plot was utilized, with jitter colored by global region laid on top. While it was strongly hypothesized that the global views of corruption would go up in 2020, there was a small difference in corruption perception, and in fact the changes that did occur show the clustering of Perceptions of Corruption shifted downwards slightly as a whole. This shows us that during the time of struggle, perhaps we found a small amount of peace with our government and business leaders.

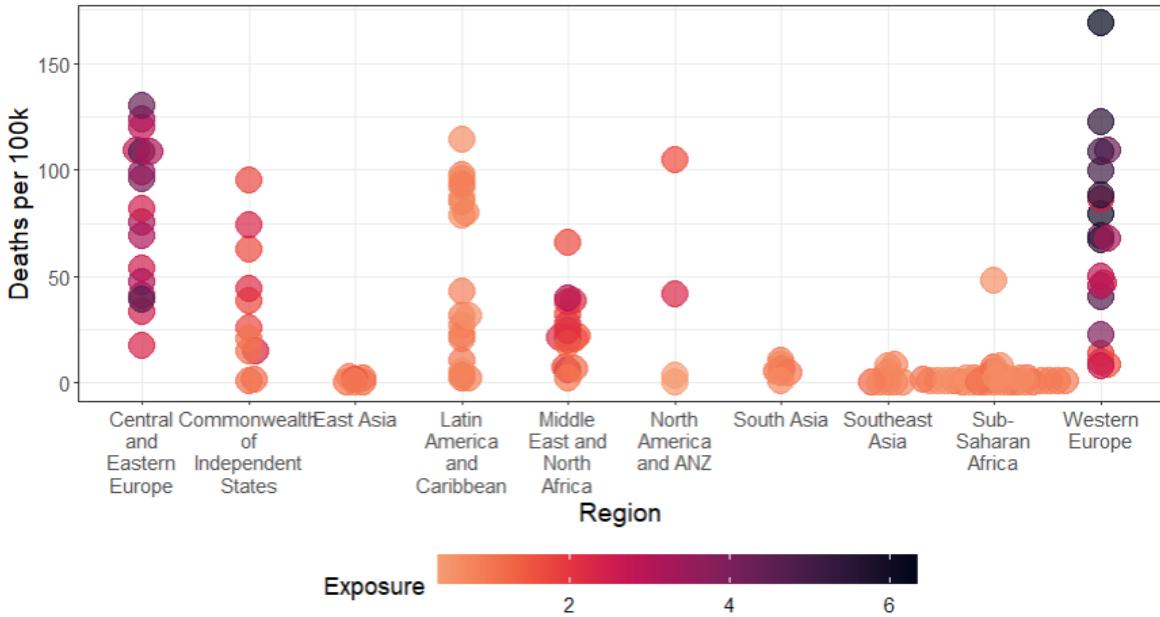


## Covid-19 Exposure Rates, Mortality Rates, and Regions

During the exploratory analysis section and throughout PD2, data comparing COVID-19 Death rates through countries and regions was difficult to explore because the primary factor in mortality rates in 2020 was in the initial exposure. Europe as a whole suffered the most because everyone got sick very quickly, before the region had time to prepare for the virus. This initially

led to the creation of a scatterplot, grouped by region that displayed COVID Deaths by country, with index of exposure as a color factor for each country. This visualization made it very clear European regions had the greatest exposure compared to the rest of the world. However, at this point it was more interesting to see which countries in particular were being shown at each data point. Then there was the creation of this R shiny app, which allows one to click on any point in the graph and receive information about the data point, specifically, the Region, Country Name, COVID-19 Death Rate, and Index of Exposure to COVD-19. Then the second portion of the app allowed the user to scroll through the data, or type in a variable of interest to look up. In this example, "Western Europe " was searched. This lets the user know who, what, where, and a bit of why any country felt the impact of COVID that they did.

### COVID-19 Deaths per 100k Population and Initial Index of Exposure in Every Country, Separated by Region



### Data Table

Show 10 entries

Search: western europe

	Region	Country	Deaths_per_100k	Exposure
7	Western Europe	Austria	69.0841979980469	4.31468200683594
12	Western Europe	Belgium	168.495697021484	6.34330797195435
32	Western Europe	Cyprus	13.5860004425049	1.90779066085815
34	Western Europe	Denmark	22.4093990325928	3.95305609703064
41	Western Europe	Finland	10.125	2.22508335113525
42	Western Europe	France	99.211799621582	5.03089904785156
46	Western Europe	Germany	40.3311004638672	4.83782911300659
48	Western Europe	Greece	46.4162979125977	2.32733154296875
55	Western Europe	Iceland	8.49820041656494	1.64933478832245
60	Western Europe	Ireland	45.3036003112793	3.18810224533081

Showing 1 to 10 of 21 entries (filtered from 148 total entries)

Previous

1

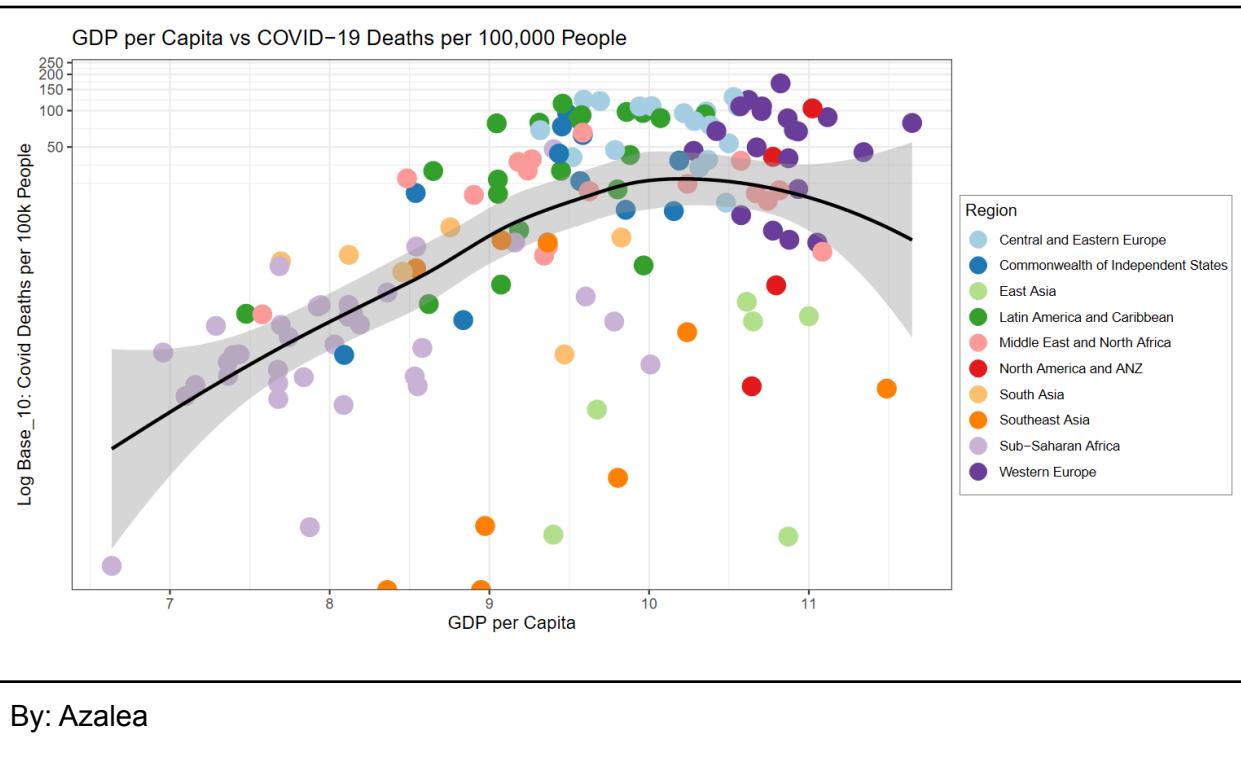
2

3

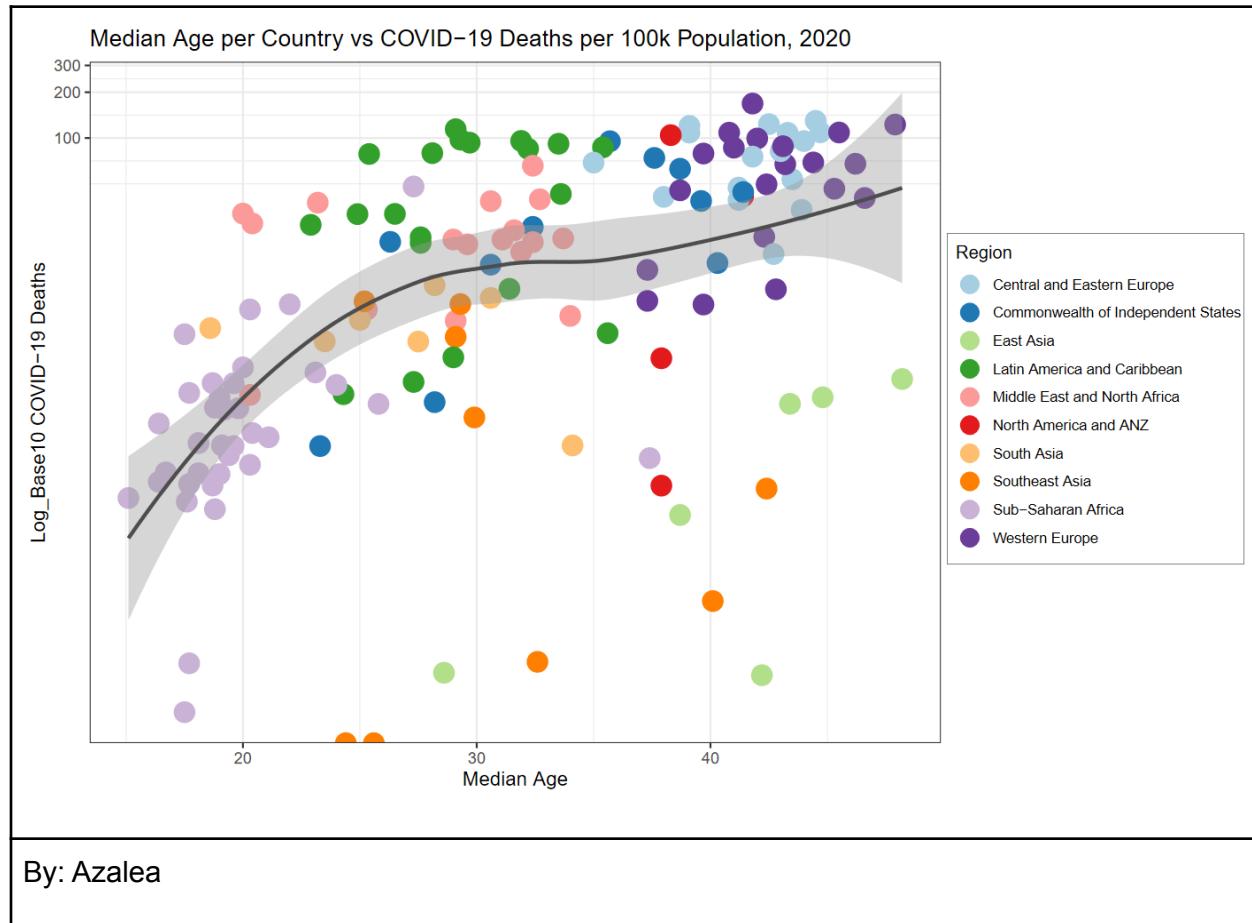
Next

By: Azalea

Now that we are aware of the leading factor in COVID-19, we can take a closer look at other variables in the pandemic. First, GDP per capita. This information is compared in a scatterplot that plots every country in the world, showing the country's GDP per capita against its 2020 COVID mortality rate. As mentioned before, the world was unprepared to handle the virus, and we saw initial surges in countries with a lot of international travel, such as Europe. The mortality rates were dispersed so disproportionately that a log scale was used in order to avoid unreadable clustering. While this visualization highlights an exponential relationship between a country's GDP and COVID mortality rates during 2020, it would be interesting to create a visualization that shows this information in a time series as the pandemic continues.



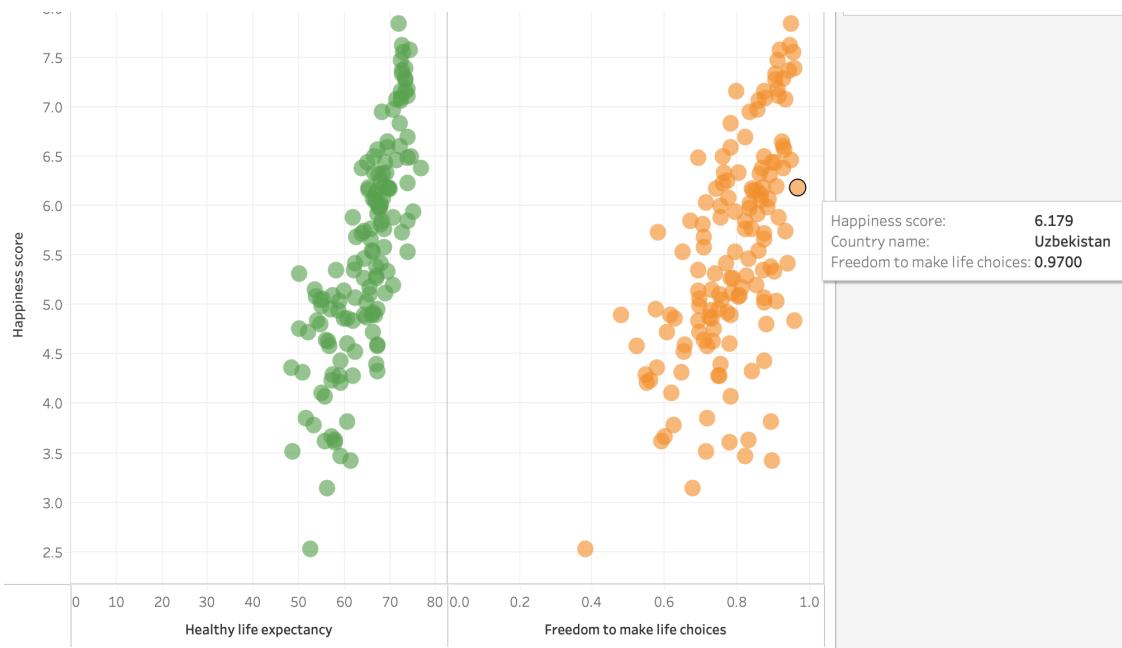
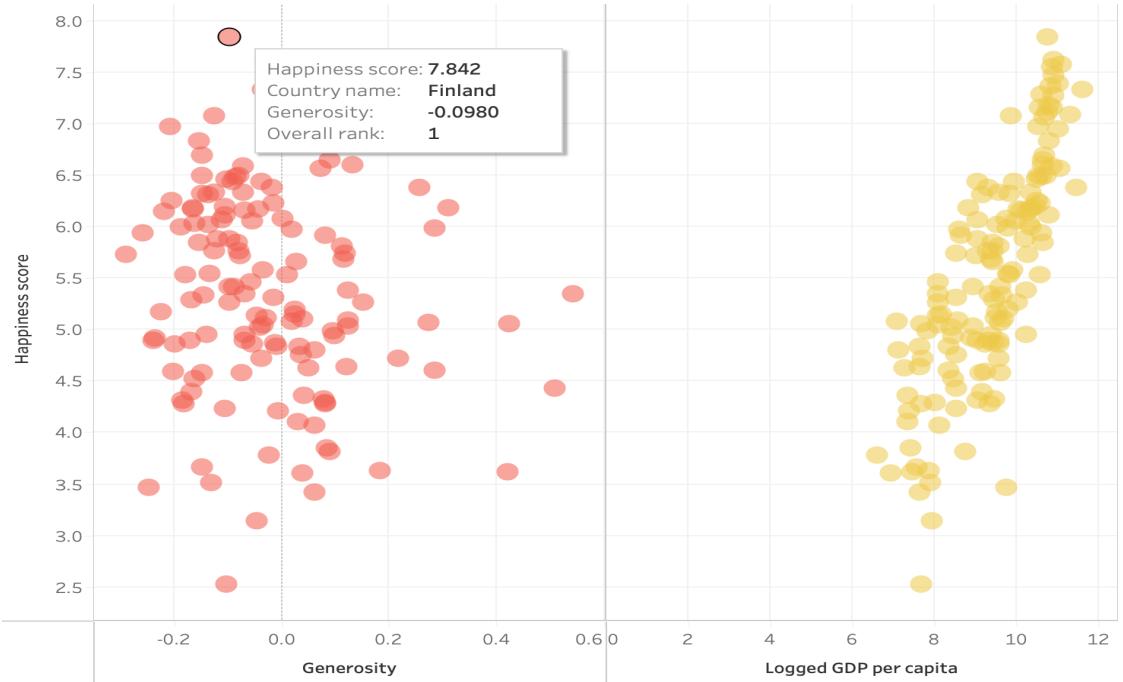
The last graph in this region-colored series looks at the COVID-19 mortality factor that had the highest condemnation rate in 2020: age. This is another scatterplot colored by region, this time with median age of the country on the x-axis and COVID-19 Deaths scaled in log base 10 on the y-axis. The pattern of the plot was again shown using ggplot's geom\_smooth function.

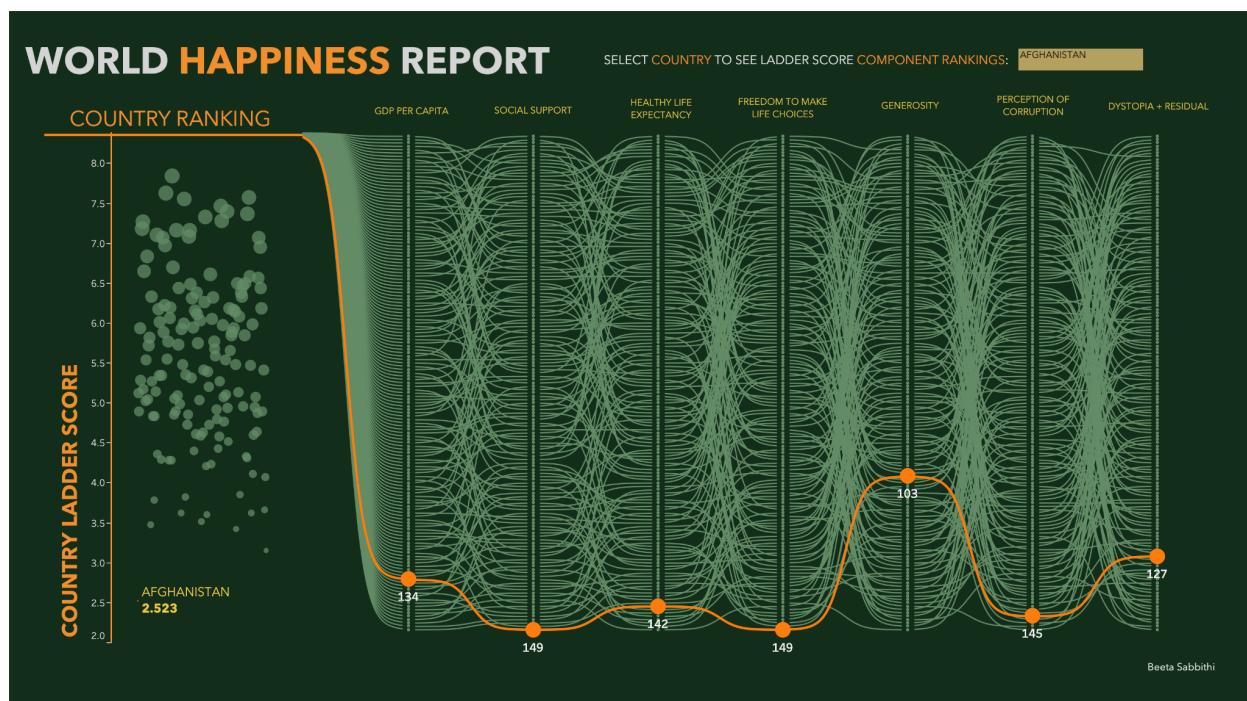
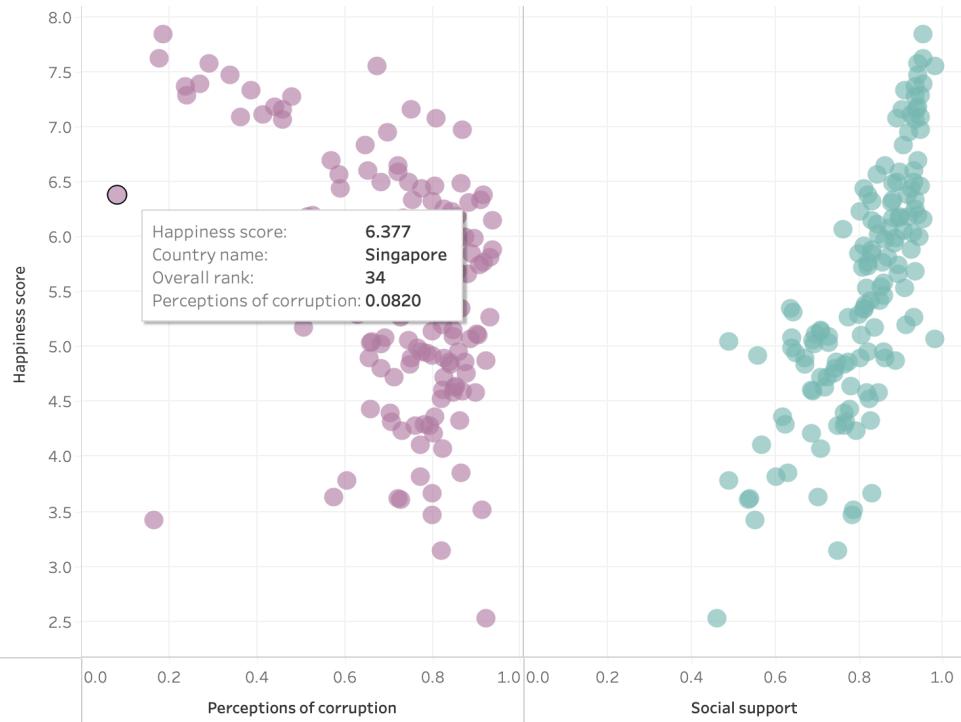


### Does happiness score and the six variables are directly proportional?

Correlation between the six key variables are analyzed with respect to the country's happiness score. Freedom, health, social support, GDP, generosity, and trustworthy governance.

In the GDP per capita graph, it is clearly visible that it is proportional to happiness score. The scatterplot shows the increment in happiness score as it moves towards the higher GDP. The same is with health life expectancy. The higher life expectancy, the happier countries. Though, generosity, freedom to make their life choices and perceptions of corruption are exceptional for few countries. Finland being the happiest country with a negative score in generosity. Singapore has minimal corruption in their governance but still at 34th ranking in the world happiness. Overall, these six variables are the deciding factor of the world happiness report.





For the above graph<sup>4</sup>, each of the six indicators were ranked, in addition to the Ladder Score

<sup>4</sup>

Interactive workbook “SankeyPlot” provided in zip file folder “Beeta”

rankings. By either clicking on a dot to select a country or using the drop-down menu in the top-right of the Tableau dashboard will change the plot to show how each country ranks in each of the indicators. The graph uses a jitter plot of Ladder Score ranks combined with a dot plot of the rankings, with a sankey path to “connect the dots” of each country’s measurements.

Although the Ladder Score is a direct indicator of a country’s happiness, our exploratory analysis revealed that some countries ranked higher in specific factors, which did not have an effect on “happiness,” but revealed other characteristics of a country that would be considered valuable for societal well-being. A deeper analysis via a multi-overlay graph (above graph), was made to see how scores are explained by the six indicators. For example, Afghanistan had the lowest Ladder Score. However, their “Generosity” rank is higher than all top 10 Ladder Score ranking countries. Using the minimal resources they have, the majority of people in Afghanistan are more giving than countries that have the highest wealth. This begs the question: is happiness bought at a price?

## Analysis and Discussion

The visualizations regarding perceptions of corruption were surprising at first, it seems as though most countries do not let their perceptions of corruption in their government and business get in the way of other factors in their lives. This is reasonable however. There are cultural factors that play into other variables, for example a population’s charity contributions can be affected by a financial or cultural obligation. They might rate their freedom to make life choices in a way that involves factors beyond their government. It is certainly possible that countries with high COVID exposure knew their government did the best they could, or vice versa. The overall lowering of corruption perceptions in 2020 compared to the decade previous might show that during the first year of the pandemic, the global population eased their criticisms towards their leaders.

The webpage looking at Exposure and Deaths across global regions hints at deeper factors in regional handling of the pandemic. We can see that Western Europe especially had high exposure and mortality. But other locations, such as Latin America and the Commonwealth of Independent States, had much lower exposure indices yet similar levels of Mortality. Going into the GDP vs COVID Mortality graph, we can see that the countries in both regions cluster in GDP per Capita, suggesting that finances played a heavy part in these regions’ ability to manage the pandemic.

The graph regarding GDP and COVID mortality rates was unsurprising, as we already knew that European countries suffered the highest casualties in the beginning of the pandemic. While these wealthy countries were hit hardest in the beginning, once the vaccines began rolling out, these same countries were able to inoculate their citizens first, leaving scraps of vaccines for poorer countries. As the pandemic continues, I hypothesize that this pattern will begin to reverse shape. It would be interesting moving forward to create time series graphs looking at the pandemic over shorter periods, such as every quarter.

The last visualization in the series shows a clear positive slope between a country’s median age and their COVID-19 Mortality rates. But with the delta variant now spreading across the globe, younger people have been succumbing to the virus at higher rates. It would be interesting to compare the pandemic dataset again with 2021 data.



# Appendices

## Datasets

[Data 2020](#)

[Data Panel](#)

[Mortality Data](#)

## Code

A zip file containing code and interactive graphs of each group member, is included in the submission folder

## Individual Components

### Azalea

My role in the team included teamwork such as setting up the team's slack channel, hosting our weekly Zoom meetings, and contributing to our team's gameplan in completing our project. The infographic uses graphs we originally presented solo for our group powerpoint. These were *Ladder Score of All Countries* by Beeta, *Most Happiest Countries* and *Least Happiest Countries* by Shivangi, and *Average Happiness Perceptions and COVID-19 Death Rates* by me. Since these 4 graphs had relevant relationships to each other, I took them all into GIMP, cleaned the background of Beeta's map visualization, then arranged my star plots around the map, recolored Shivangi's graphs and set them next to each other, and added the text boxes describing the highlighted factors. After this, I created the graphs in the conclusion of the visualization story we've made here, by observing data in regards to COVID-19's effect on the world. To do this, the Mortality Dataset was merged with the Happiness Data using an inner join factored by country name.

All my visualizations presented in this final report were made in R Studio, and edited in GIMP for touch ups. Tableau was also used during the exploratory process, since the ease of click and drag made it possible to explore different variables in rapid succession. This also allowed me to collect regional data for each country, allowing for the color factors in the scatterplots, beeswarm, and violin plots. The star maps, exploratory for the head and tail of countries' COVID mortality rates, and the final visualization overlaid on the map in the infographic were mine.

The Factors by Region visualizations were:

- (exploratory) Happiness Factors of the 10 Countries with the Highest Percentage of Covid Fatalities & Happiness Factors of the 10 Countries with the Lowest Percentage of Covid Fatalities *Star plot*
- (exploratory) Perceptions of Corruption vs ... *scatterplot*
- (exploratory) Perceptions of Corruption vs Having a Female Head of Government *Violin Plot*
- Perceptions of Corruption per Country in the Last 10 Years, Colored by Region *violin plot*

- (interactive) COVID-19 Deaths per 100k Population and Initial Index of Exposure in Every Country, Separated by Region *beeswarm with interactive table*
- GDP per Capita vs COVID-19 Deaths per 100,000 People *scatterplot*
- Median Age per Country vs COVID-19 Deaths per 100k Population, 2020 *scatterplot*

The data exploration I did was mostly centered around the Perceptions of Corruption variable, and the COVID-19 dataset, and comparing these variables against one another to search for any trends or unexpected results.

Reflection:

Coming into this class I had worked with R for data manipulation, not much for data visualization. These visualizations took a lot of time attempting to create clean, easy to understand graphs that did not overload the viewer with information. This was especially tricky with this data, since I wanted to show every country's pandemic data while recognizing how many data points that included per plot. Using R's aspect ratio tool helped in the development of graphs that appropriately filled the space. A lot of time and appreciation was put into setting the size of text, points, and lines making up the data so points were easy to understand.

Most of the graphs were also post edited in GIMP. Though I'd used the program before, the

This was the first time I had created an interactive visualization of any kind, and it awed me how easy it was to create the one displayed here. In fact, everything in R was wonderful to use, and the export to pdf option left all the images looking crisp.

## Shivangi

Summary of work -

My role in the team included teamwork such as weekly Zoom meetings, communicating on Slack about my task status with my team mates, and contributing to outline the project. The second infographic was completely done by me which tells the story of how a country's economy and generosity among people of the country decides the happiness of a people. To do this, I created graphs in the Tableau and designed the infographic on Adobe XD which was actually fun to do.

All my visualizations presented in this final report were made in Tableau. My visualizations for the final paper are:

- Tree maps for the infographic
- Bubble Chart for the exploratory analysis
- Scatterplot, showing correlation between them (complex graph)

Summary of takeaways -

This class has been amazing for me. I got to learn so many things about data; how to analyze, manipulate, visualize and finally tell a story with data. This transition is so fulfilling. I was amazed that this course has taught me a lot which could be utilized in different aspects. The key takeaway is always focus on Data, Message and Audience while designing story/infographic for

the users.

In the second phase, we focussed on exploratory analysis. This phase was to dig into the data and bring something out which can be presented to the audience as a story. I learnt various graphs and interactive/ animated graphs in tableau.

In the final phase, the primary focus of the project involved representing the data story in a technical way. I learnt how to analyze two different rows of a table and tried to show the relation between them.

## Beeta

Summary of work -

For the project, I shared in the responsibilities of setting up meetings, bringing discussion topics to meetings, and communicating on Slack about my task status to the rest of the group.

I set up a folder with project guidelines and initiated a working document for each PD submission. As well as submitted PDF files of our work into the submission folder.

My visualizations for the final paper are:

- Regional Violin Plot
- Correlation Matrix
- Sankey graph (complex graph)

For the regional violin plot, I color-coded the regions. It shows the median Ladder Score of all regions. Because it was important to see the density of each region, I chose a violin plot instead of a box plot. Graph made in R.

For the correlation matrix, the regional indicators were colored using a palette from RcolorBrewer. The “green” hues indicate a correlation, while “yellows” indicate no correlation between the indicators. Graph made in R.

For my complex graph, I combined 3 graphs into one in the dashboard: jitter plot, circle dot chart, and sankey plot. Graph made in Tableau

The jitter plot shows the location of the countries on the Y-axis, which is the ladder score. The circle dot plot is a ranking of each of the 6 indicators, plus the ladder score.

The sankey diagram shows the flow of each country’s ranking.

To set up the dashboard, I applied a parameter to connect the jitter plot to the countries. Doing this allows a user to either click on a point in the jitter plot or select from a drop-down menu. When a user selects a country, a label of the country’s ladder score displays on the jitter plot. I overlaid the circle dot chart on top of the sankey diagram so that each sankey flow line connected to a point on the dot chart.

The color used on the dashboard was chosen specifically for color-blind individuals.

My visualizations for the infographic were made in Tableau (Page 11 only):

- Geomap with the highest and lowest Ladder Score countries marked on map

- Correlation scatter plots of the two social indicators that were identified to have the most influence on Ladder Score.

#### Summary of takeaways -

The final project was a great reflection of how a data visualization “story” is put together in the real world. The first phase of the project encouraged me to learn the datasets before showing others what it revealed. Unless I had a solid understanding, my audience would fall victim to false data, confusion, and incomplete information. An important component of this step was to organize the datasets and identify limitations, outliers, and any factors that could potentially affect the project’s direction. After all, the data is the blueprint.

The second phase focused on exploratory analysis. Because this phase was not intended for the audience, I was encouraged to do many interactions with simple graphs such as scatter plots, violin plot, and line graphs. Seeing the dataset visually showed patterns or facts that simply looking at the dataset would not have shown. This second phase of the project also helped me learn which method of graph works well for the dataset and which did not work. This phase of the project helped define the message of my project.

The last phase of the project involved presenting data to a technical, as well as non-technical audience. With my own knowledge in Human-Computer Interaction, one key point we learn is that “user is not like me.” This phrase applies just as much in data visualization. Learning how to display visualizations for a non-technical audience was more challenging because the message of the data needed to be concise and tell a story. Putting together an infographic helped our team focus on the most important takeaways.

Likewise, presenting data to a technical audience needed to be more complex in order for them to see meaningful insights. It not only helps the audience learn, but provides me with a greater opportunity to receive feedback about observations that others have made that I can apply to improve/showcase my message.

Regardless of which step of the project I was on or in which way my focus shifted, the most important constant throughout has been the mantra, “Data, Message, Audience.”