The narrative -

Inspirational questions

1. **What kind of story do we want to tell?**
   1. Personal?
   2. Inspirational?
   3. Technical?
   4. Futuristic?
   5. Etc
2. **What is the Hypothesis?**

After listening to you we want to explore what other people have to say about happiness. We want to understand to what degree happiness can be categorized. If you are able to break down elements of happiness, will all of them be created equal? We found the most widely used and reputable measure of happiness by country; we then used k-means clustering to try and see if any of the factors that were measured had an outsized effect on a country's overall happiness score.

We used the **World Happiness Report.** Itis an annual publication of the **United Nations Sustainable Development Solutions Network**. It contains articles, and rankings of national happiness based on respondent ratings of their own lives, which the report also correlates with various life factors. As of March 2020, Finland was ranked the happiest country in the world three times in a row. The World Happiness Report contained a listing of 156 countries with an overall score ranking and a happiness rating by countries.

Of all the factors within the dataset in the 2019 World Happiness report used to calculate the overall happiness score, we believe one or more factors has a larger impact on the overall happiness than the other factors.

Factors:

**The 6 variables provided in the data set are;**

1. **GDP per capita**: Purchasing Power Parity (PPP).
2. **Healthy Life Expectancy**: Time series of healthy life expectancy.
3. **Social Support**: National average of the binary responses to the Gallup World Poll (GWP) question.
4. **Freedom to make life choices**: Freedom to make life choices.
5. **Generosity**: Money donated to a charity in the past month.
6. **Corruption**: Perceptions of corruption’
   1. What is the hypothesis related to?
   2. What is the hypothesis answering?
   3. What is the impact created if the hypothesis is true (or false)?
   4. What is the outcome?
7. **Why is this hypothesis important?**

If you talk to someone long enough you know that happiness is subjective, at least after a point, if you talk further this can lead to arguments. If you continue to argue you will either leave heated and angry or you will reconcile with the other person and look to what you both see as equally important to each of your feelings of happiness.

In that agreement you might find the ideas expressed in the data gathered. This will not cover what makes everyone happy in a specific place but it is a good source on what people around the world can agree on makes them happy when they’re ready to stop yelling.

Commonalities are important for the present but further analysis can help us take what it gathered and look towards the future. If there’s a factor that has a greater effect on a country’s happiness then it stands to reason that’s a good area to focus for others.

* 1. Who do we expect to visit this website?
  2. How do we expect the users to utilize the information provided?
  3. What impact will it have on their lives?

1. **How did we find the data?**

We didn't find it on our own. A little bird told us about it. Seriously though, in our initial project where we created our own happiness index we looked for other attempts and measured happiness by country. This dataset came up frequently.

* 1. Why is this particular data set important?
  2. Why is this data set valid?

1. **Explain each map**
   1. Explain how the map answers the hypothesis question.
      1. Map 1: This map displays the happiness ranks from the 2019 World Happiness Report categorized into 4 groups of roughly 40 countries each (Highest Ranks, Mid-Higher Ranks, Mid-Lower Ranks, and Lowest Ranks). This is a baseline map to explore where the countries lie based on the report’s happiness score.
      2. GDP Plot: The above scatter plot shows the clusters created through K-Means clustering with respect to GDP per Capita. This graph explores each country's happiness score to their respective GDP per Capita based on each cluster. It is evident that each cluster represents a high happiness score, moderate happiness score and least happiness score associated with generally high GDP per Capita, a mixture of high and low GDP per Capita and generally low GDP per Capita, respectively.
      3. Generosity Plot: The above scatter plot shows the clusters created through K-Means clustering with respect to Generosity. This graph explores each country's happiness score to their respective generosity based on each cluster. It is evident that each of the clusters does not have a clear representation based on generosity. It shows that generosity is generally low across the three clusters of happiness score suggesting that the report's 156 countries represent less generosity. Ultimately, this shows that generosity does not impact the happiness score as much as the GDP per Capita.
      4. Corruption Plot: The above scatter plot shows the clusters created through K-Means clustering with respect to perceptions of happiness. This graph explores each country's happiness score to their respective perceptions of happiness based on each cluster. It is evident that the high happiness scores are associated with mostly low perceptions of corruption with a few high representations, whereas the moderate and least happiness scores are associated with generally lower perceptions of corruption. This graph concludes that happiness score is not as impacted by perceptions of corruption as GDP per capita. Additionally, most of the report's 156 countries tend to have lower perceptions of corruption.
   2. Takeaways
   3. Differences
   4. Uniqueness

**What will be included in the website:**

1. **The data set analysis**

**Analysis**:

We reviewed the data set 6 variables and determined that there were no obvious ways that we can distinguish groups through manual inspection and there were no explicit labels within the data that provide logical groupings. We then utilized Tableau K-means clustering to discover patterns in our data set and chose groupings based on 4 clusters. Tableau clustering then divided the population into subgroups based on shared characteristics (see images below).

We further grouped the happiness ranking of the data set into 4 distinct quadrants. The table below shows the grouping labels and ranking. We added the label of the grouped happy rank on top of the cluster to determine if there is a logical grouping of the variable around the ranking.

|  |  |  |
| --- | --- | --- |
| Overall Country Happy Rank | Percent | Label |
| 1-39 | 25% | Very Happy |
| 40-78 | 25% | Happy |
| 79-117 | 25% | Less Happy |
| 118-156 | 25% | Least Happy |

**Conclusion**:

Images of Analysis:

<https://public.tableau.com/profile/farid.hanna4934#!/vizhome/HappinessFinalProject/ClusterwithAvgScore?publish=yes>

Image 1: World Happiness Report (Grouped by Overall Rank of Happiness)

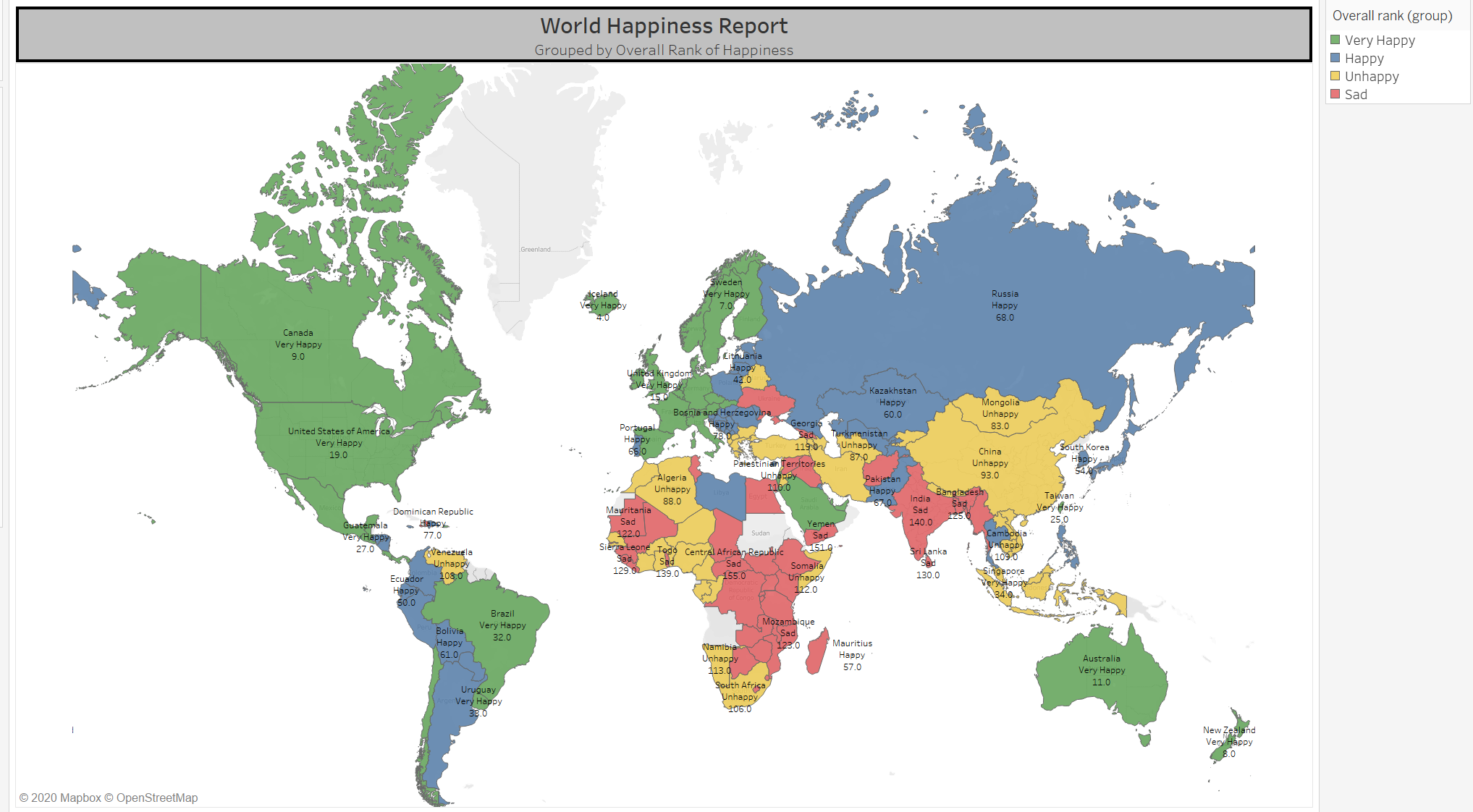


Image 2: Clusters by Country using 6 Variables (1. GDP/Capita. 2. Social support. 3. Healthy Life Expectancy. 4. Freedom to make life choices. 5. Generosity. 6. Perceptions of corruption)

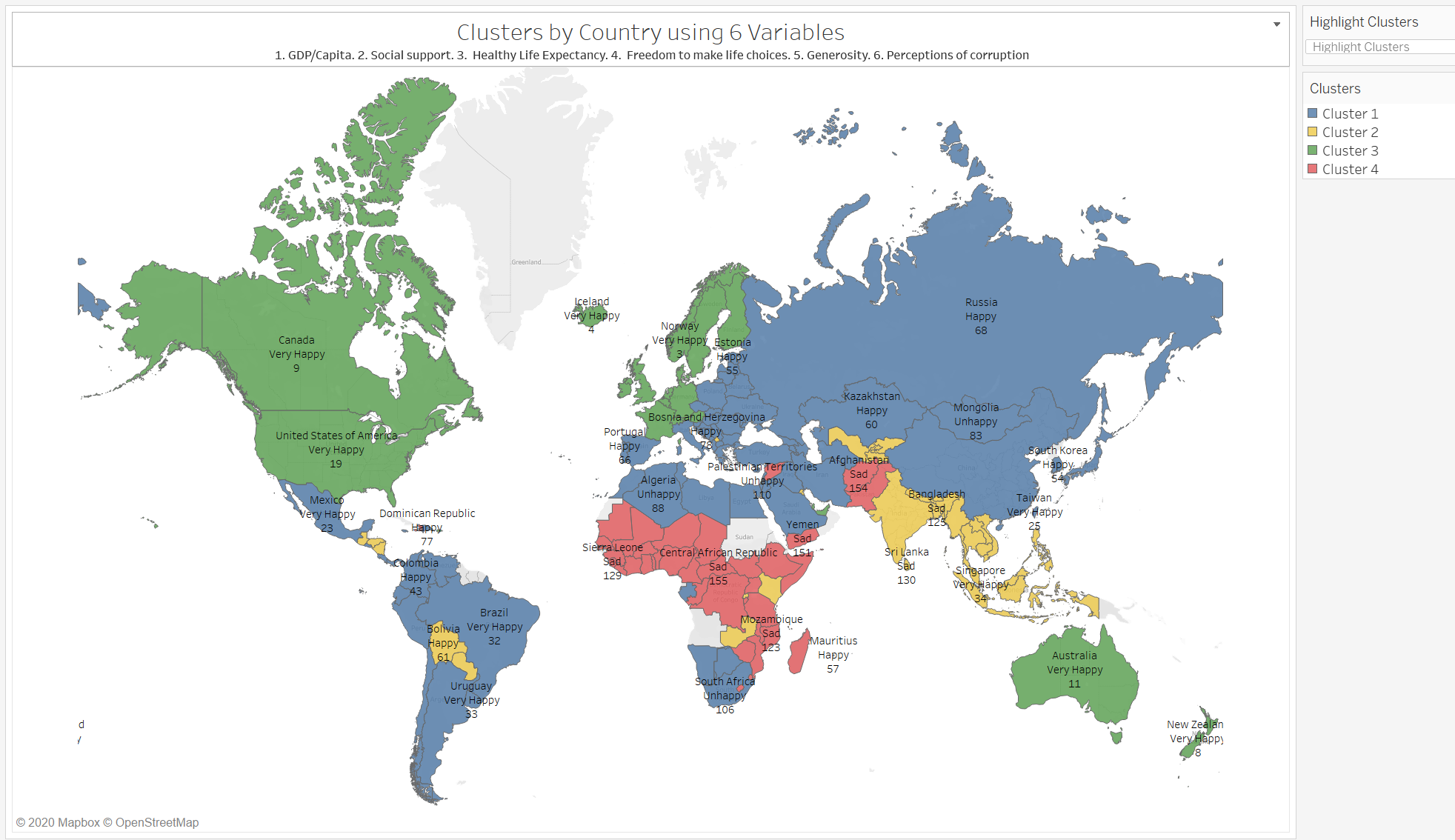
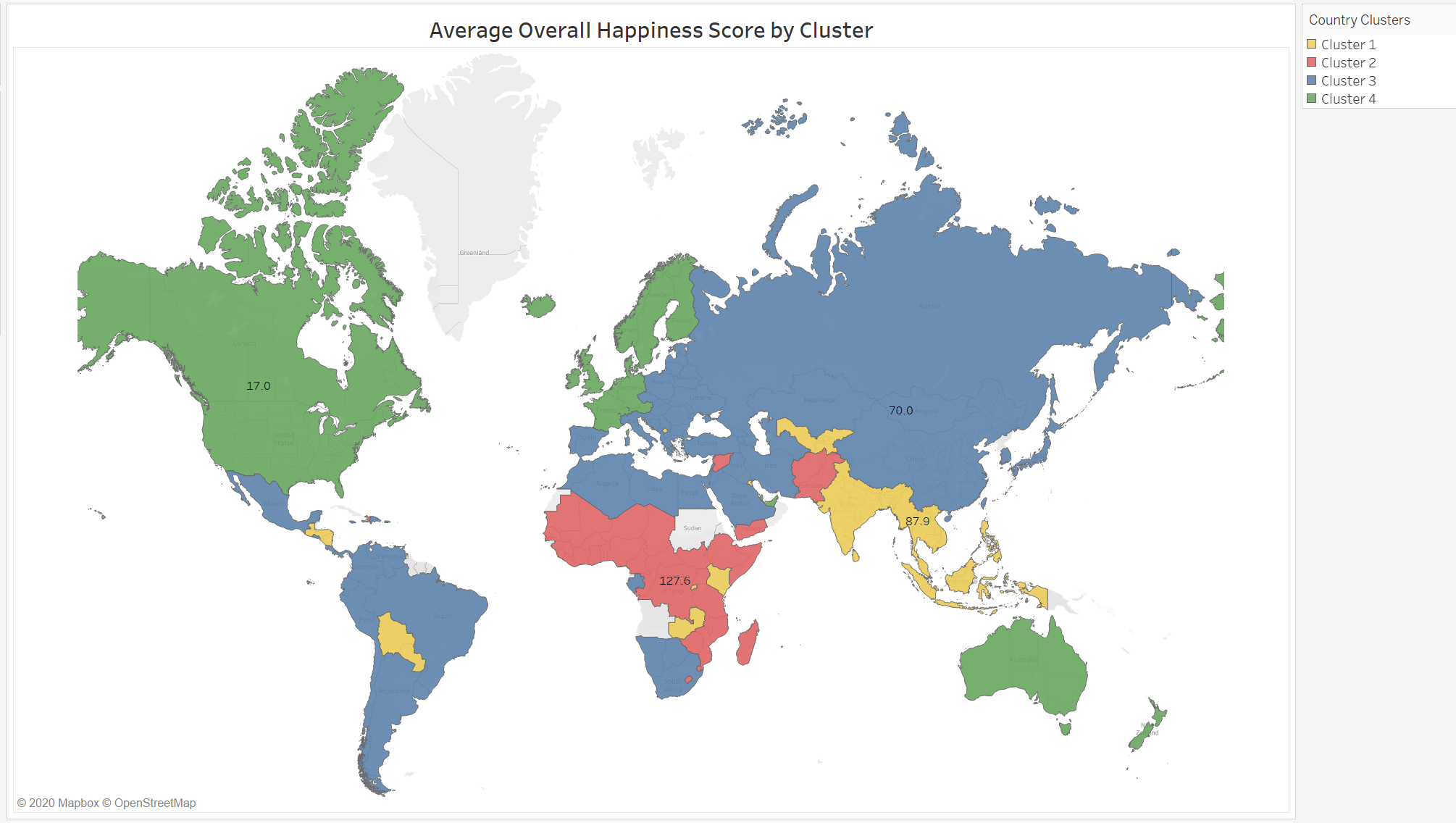


Image 3: Average Overall Happiness Score by Cluster



**Data Sources:** World Happiness Report 2019

<https://en.wikipedia.org/wiki/World_Happiness_Report>

**Description of Data:**

“The World Happiness Report is an annual publication of the United Nations Sustainable Development Solutions Network. It contains articles and rankings of national happiness based on respondent ratings of their own lives, which the report also correlates with various life factors. As of March 2020, Finland was ranked the happiest country in the world three times in a row.”

Data Variables:

From the report itself.

GDP: GDP per capita is in terms of Purchasing Power Parity (PPP) adjusted to constant 2011 international dollars, taken from the World Development Indicators (WDI) released by the World Bank on November 28, 2019. See Statistical Appendix 1 for more details. GDP data for 2019 are not yet available, so we extend the GDP time series from 2018 to 2019 using country-specific forecasts of real GDP growth from the OECD Economic Outlook No. 106 (Edition November 2019) and the World Bank’s Global Economic Prospects (Last Updated: 06/04/2019), after adjustment for population growth. The equation uses the natural log of GDP per capita, as this form fits the data significantly better than GDP per capita.

Healthy Life Expectancy: The time series of healthy life expectancy at birth are constructed based on data from the World Health Organization (WHO) Global Health Observatory data

repository, with data available for 2005, 2010, 2015, and 2016. To match this report’s sample period, interpolation and extrapolation are used.

Social Support: Social support is the national average of the binary responses (0=no, 1=yes) to the Gallup World Poll (GWP) question, “If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?”

Freedom to make life choices: Freedom to make life choices is the national average of binary responses to the GWP question, “Are you satisfied or dissatisfied with your freedom to choose what you do with your life?”

Generosity: Generosity is the residual of regressing the national average of GWP responses to the question, “Have you donated money to a charity in the past month?” on GDP per capita.

Corruption: Perceptions of corruption are the average of binary answers to two GWP questions: “Is corruption widespread throughout the government or not?” and “Is corruption widespread within businesses or not?” Where data for government corruption are missing, the perception of business corruption is used as the overall corruption-perception measure.

Positive Affect: Positive affect is defined as the average of previous-day affect measures for happiness, laughter, and enjoyment for GWP waves 3-7 (years 2008 to 2012, and some in 2013). It is defined as the average of laughter and enjoyment for other waves where the happiness question was not asked. The general form for the affect questions is: Did you experience the following feelings during a lot of the day yesterday? See Statistical Appendix 1 for more details.

Negative Affect: Negative affect is defined as the average of previous-day affect measures for worry, sadness, and anger in all years.