Name Carla J Bradley

Title : Happiness Index Analisys

Date: March 5 2021

I am glad I found this data set. I learned to implement the concepts I have learned in this class and implement correlation calculation to discover correlation on societal variables how this societal correlate to each other and varies depending on the country's culture and political reality.

I confirmed a strong correlation between life expectancy and happiness, which is quite normal, and I also confirmed that the happiest countries are also the richest countries.

Therefore, their health facilities are more advanced than the poor countries. Another reason for people to be happy live long!. I also determined that a sense of corruption in this study was not really accurate. At least it didn't adjust to the reality I see on the news. Therefore, it is important to gather data from reputable sources and always ensure the data source is trustworthy.

**What do you feel was missed during the analysis?**

I perceive that the reliability of the "sense of corruption" variable was potentially off and not adjusted to reflect the realities of citizens within countries, such as Russia and China. According to the data set, it was denoted with a low perception of corruption within these countries.

As an example, I correlated, ‘perception of corruption’ vs. ‘freedom to make like choices’ variables and I noticed a high contrast in their values. This indicates a significant difference in these two societal values expressing an overall high trust in government—but still provides a reduced sense of freedom, which is not logical in my opinion.

An interesting contradiction in the data indicates that ‘perception of corruption’ is very low in Russia and China versus higher in the United States, but Chinese and Russian citizens said to trust their government. The freedom to make choices is lower than the United States influencing the conclusion of the analysis within this area.

**Were there any variables you felt could have helped in the analysis?**

One of the most significant limitations of this data set was the integrity and authenticity of the data since there is a sense of corruption value. It is possible that the dataset did not reflect several countries' current reality and true influences that possibly the impacts the analysis and conclusion, in particular the correlation connections.

Another interesting variable to be added within the dataset would have been the addition of value representing “sense of faith” it would have been great to run a correlation with the rest of the variables, especially with societal as life expectancy, social support, and generosity.

**Were there any assumptions made you felt were incorrect?**

I was hesitant to address the zero values in my data set since I am not sure whether they are missing data or outliers. I considered replacing the zero value with a value close to previous values year or represented as "NA.,” but since there were only (8) zero values and some of them were possibly real values as the example below, I decided to overlook those values.

*Example:*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| year | Country or region | Regional indicator | Ladder score | Social support | Healthy life expectancy | Freedom to make life choices |
| 2019 | Afghanistan | South Asia | 3.203 | 0.517 | 0.361 | 0 |

**What challenges did you face, what did you not fully understand?**

One of the most significant challenges that I faced while working with this analysis was confidently identifying the type of distribution to use within this specific data frame.