

Project Proposal

Torin Frever, Gabe Cuhan, Shane Martin, Sanjay Unni

Project Description

The dataset we are working with was found on Kaggle.com, called the “World Happiness Report”. The data was collected through the Gallup World Poll. It has been collected annually since 2012, and we are using the 2015 dataset, as it has the most data available.

This report has gained attention from experts in various fields across the world. It can be of interest to economists, psychologists, policymakers, and more. This data and the analyzation of it could have a major effect on the prioritization of certain issues in a country by their leaders. We’re going to target the leaders of countries.

If a leader of a country were to look at reports made from this data, they may realize that the overall level of happiness is dissatisfying, and they could see the attributing factors to said unhappiness and attempt to fix these problems. Potential questions that may be of interest to world leaders include:

- Do nations with higher GDP have generosity contribute more towards the overall happiness score?
 - What about the “family” contribution?
- Which one factor should a nation focus on in order to boost its overall happiness score the greatest?
 - What specific changes would a nation have to make in order to effect this factor?
- How does happiness regarding freedom and trust vary across different regions?
 - Do more free nations have freedom contributing a larger amount to their overall happiness score, and vise versa?

Technical Description

We plan on using Shiny to format our final product. The data we are reading in is a static .csv file, and we will use libraries such as tidyverse, ggplot2, tidyr, and more to filter, arrange, summarize, and visualize our data.

We will be answering simple questions such as which country has the highest average GDP, or which country consistently had one of the highest happiness scores, and then tie them to more complex questions, like if there are any noticeable relationships between those statistics and other items of measurement (like trust in the government or life expectancy).

The biggest challenges we will face include selecting meaningful comparisons that will help us analyze the issues facing different countries today. We will also have challenges narrowing our data between counties and regions, much like in the “any_drinking” data frame, where each state had many counties. This complication will force us to choose strong questions and sort the data carefully and in a organized way.

We will have to use the libraries listed above well in order to display the information well.

Project Set-up

We made a GitHub repo that everyone was invited to, and set up 5 different issues:

- Uploading the 3 .csv files we plan on using.
- Creating a pages branch for uploading our Shiny report.
- Deciding what information we want to select.
- Deciding what format we want to use to present our data.
- Assigning roles to everyone for manipulating the data.