

Personality Around the Globe

CS 5630/6630 Final Project

Jordan Pyper; u0363195; u0363195@utah.edu

Hunter Finney; u1319765; u1319765@utah.edu

Jackson Leach; u1110821; jackson.leach@utah.edu

Git Hub Repo: <https://github.com/HunterFinney/PersonalityAroundtheGlobe>

1 Overview

This project explores the data collected across the globe from a psychometric survey that totals to over 1 million participants. Our goal is to take this raw data and transform it into a visualization that is digestible for the public.

The survey scores users based on the Big Five personality traits. The survey will present a question such as: “I often enjoy being the center of attention.” Then the survey will ask the user to respond on a scale from one to five with one being disagree and five being agree. The five core factors are *conscientiousness*, *agreeableness*, *neuroticism*, *openness*, and *extraversion*. The survey records the scores and what country each user takes the survey.

To visualize this massive data set we first must condense the data into a form that is more ready to use than it currently is. This will be done by computing the mean and the median for each personality trait along with other statistical analytics. More detailed information can be found in section 5, Data. Once the data is properly managed, the data can now be transformed into a visualization. Our design focuses on three critical modules that will together serve as our visualization. The first key component is the map view. The map view will show a navigable map of the globe that includes a heatmap of personality intensity based on that chosen personality trait. The next cornerstone of this project is the table view. The table view will display the country level tabular data with each country as a row and each column as a personality factor. The tabular data will be sorted alphabetically at first. The table can be sorted either low to high or high to low by clicking on the column descriptor located at the top of the table. The last critical node will include a country comparison tool using brushes to see how similar countries are grouped.

2 Motivation

a. Background

The Big5 is a very popular social psychology topic that is widely taught in nearly every intro to psychology class across the United States. A quick search of the internet yields no popular visualization to compare personality traits across the globe. There is interest in this subject from the scientific field. There are many papers that analyze differences across the globe in personality tests such as The Geographic Distribution of Big Five Personality Traits Patterns and Profiles of Human Self-Description Across 56 Nations by Schmitt et al (<https://www.toddshackelford.com/downloads/Schmitt-JCCP-2007.pdf>). Papers like these are dense and not easily understood to the general public. These papers are dense, full of numbers, data, and technical jargon. We know

the public is very interested in social psychology due to the large success of the Myers–Briggs test in popular culture. An easy to use tool for the public to not only understand their own country's overall personality but to connect to other people across the globe.

b. Motivation

This project would yield interesting results and allow a lens to be placed on how cultures relate to personality. A lot of questions can arise from this single data set. We can see how volatile personality traits in certain parts of the world. Are there some traits that the majority of people tend to have in common? For example, does the United States have a normal distribution on neuroticism or do we see a grouping where many are either hyper critical of themselves or not? These are interesting questions to pursue, but the most intriguing questions are those that are cross cultural. In a world full of borders and divides, it would be a joy to see how closely related we all are. It would also be interesting to see how widely diverse humans can be. All of these questions should be easily answered without having to dive into scientific research as a lay person. That is where our visualization will fill this gap.

c. Personal Statements

Jordan Pyper

My undergraduate schooling is in sociology and psychology. I've always been interested in the human element of data science and computational research. I wanted to find data that had a human element where people can look at it and understand more about themselves and others. Something that everyone could relate to and find insight in, at least on a human level. I also have a master's degree in statistics and wanted data that I would have to analyze and get my hands dirty in order to understand. One thought was to use principle components analysis or some fun technique to condense information into something usable in a visual. One of the great things about data visualization is a good visualization "is worth 1,000 words". I've been seeking an opportunity to combine effective visual communication with solid statistical analysis. While time constraints won't permit us to try and experiment on all aspects of the visual and/or data that we find interesting, we are hoping to at least demonstrate the harmony between a good analysis and a good visualization.

Hunter Finney

My background is based in computer science, but most of my working experience leans closer to psychology and perception. I spent the entirety of my undergrad in a virtual reality lab running perpetual experiments and interpreting data from those experiments. I often find myself fascinated with how human perception works which bleeds into how visualizations can easily be deceptive or misleading on merely the presentation alone. I aim to produce visualizations that are perceptually clear and

easily digestible even for large amounts of data. This data interests me by being able to parse and draw conclusions from every part of the globe. There is so much data here to enjoy and hopefully reveal things I have yet to learn about people across the Earth. Using the Big 5 data set will leave open many opportunities to look at the data in unique ways. My interest in psychology has kept this popular questionnaire close to home for me, and I am excited to tackle this project.

Jackson Leach

My background is in computer science, but I have always been interested in different personality types, like the enneagram or MBTI tests. I also had the opportunity to study abroad last year in South Korea and was struck by some of the ways that people I met seemed to have similar personality types to myself, but acted differently within their societal context. I am hoping that in working with this data set we will be able to provide some interesting insights on how different cultures can influence personalities.

d. Objectives

Although our backgrounds have led us to wanting to pursue this data in particular, we also wish challenge ourselves and produce an excellent data visualization through conquering these objectives in no particular order.

1. Learn how to provide insightful high-level information while providing the ability to drill down into specifics.
 - (a) Wrangle “big data” and present in a concise and effective way.
 - (b) Describe groups, countries in this case, and make comparisons.
2. Present human centered data that anyone can learn from or find something interesting.
3. Learn how to best present categorical geographic information. Specifically, our data can be visualized geographically, but could be presented differently. Think through best methods for this.

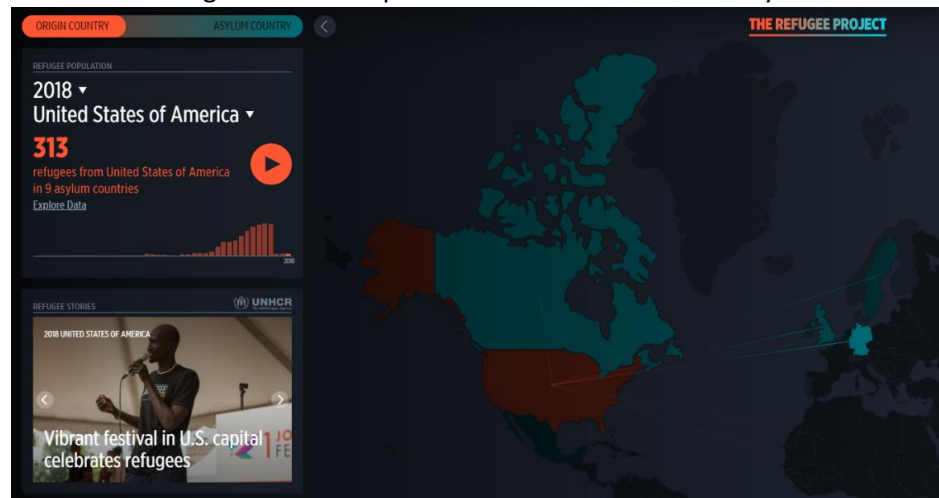
These objectives are the main tasks for this project. We intend to accomplish all of these, while there are other, we could pursue if time permits.

3 Related Work

<http://www.therefugeeproject.org/#/2018>

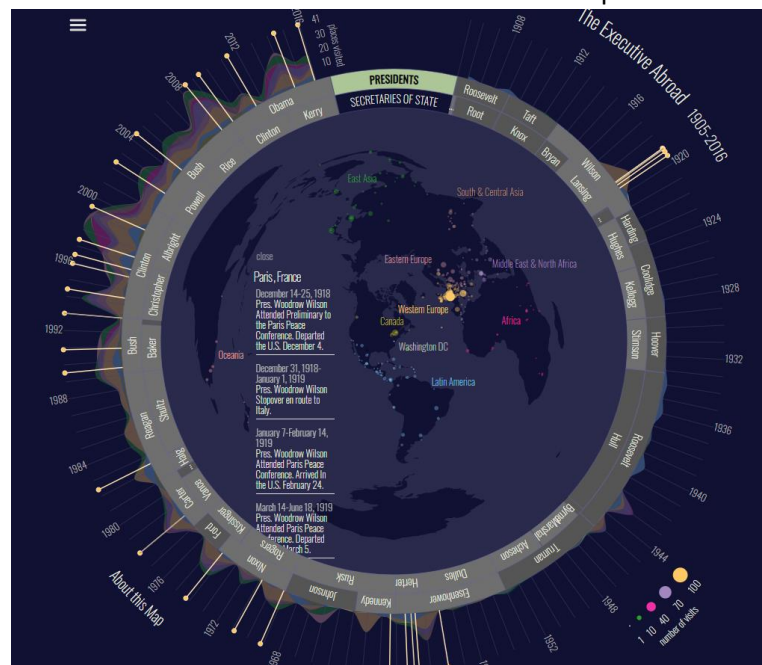
This visualization shows the migration of refugees across the globe by visualizing where they originate from and what is the asylum country. This visualization uses a clickable world map with a detailed column using about one-fifth of the width of the

screen. This work is similar to our general layout we are conceptualizing as well as ideas like clicking the world map for a detailed view of a country.



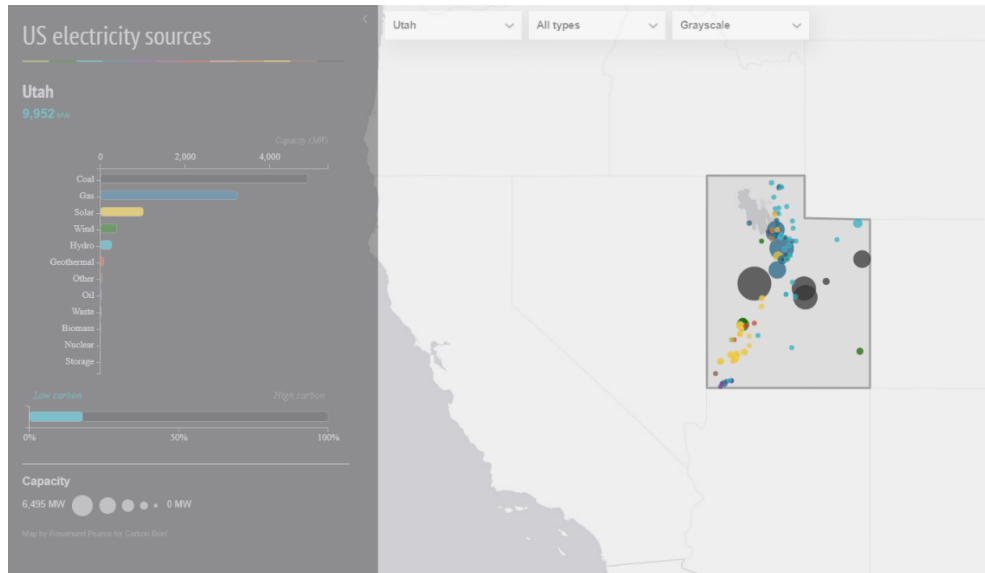
<http://dsl.richmond.edu/panorama/executiveabroad/#lat=45.42&lng=-75.7>

This visualization shows United States foreign affairs from 1905-2016. This visualization had some similarities to our idea of a clickable map, detailed view, along with extra data graphs shown on the circumference of the map. Although similar to our design ideas, this visualization technique seemed ill-suited for our needs. A timeline isn't what we are presenting, so any information presented around the circumference would be difficult to read and compare.



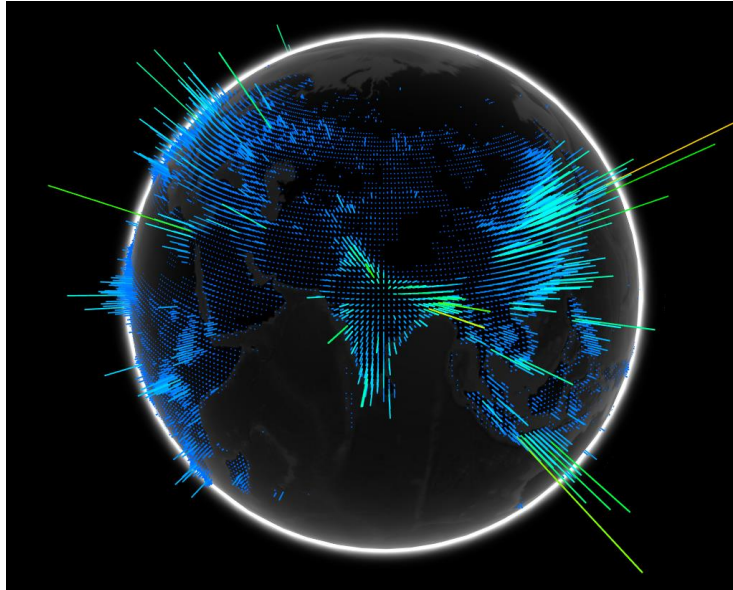
<https://www.carbonbrief.org/mapped-how-the-us-generates-electricity>

This visualization is of the US power resources in 2017. It also includes a map with a detailed view column that allows for detailed information on states. This visualization is interesting since there are separate types of power generation much like our separate pillars of personality. This relationship is rather short lived since our data isn't as location based as this visualization and using circles to indicate our simple numbers wouldn't make sense.



<http://globe.chromeexperiments.com/>

Here is a simple 3D globe showing population density in years 1990, 1995, and 2000. This is our quick dip into 3D to feel that it is still not an effective tool for our dataset.



4 Questions

We aim to answer how do personalities change across the globe over different cultures. Do we see a significant difference in openness when comparing the East to the West? We also are interested to find similar countries as well. What countries trend to have the same personality types and what conclusions can we draw from that? We are also interested in if personalities within countries fall in a normal distribution or will multiple peaks appear?

5 Data

a. Data Acquisition

<https://www.kaggle.com/lucasgreenwell/big5personalitydataset>

This is a very large dataset that has over 1 million respondents from 224 countries. Each row is a single research participant, where each participant was asked 50 questions related to 5 personality factors. Each factor has ten survey questions that correspond to it. While there is more information, such as time completing the survey, this project will focus on the personality responses and country of the respondent.

b. Data Processing

This data requires a substantial amount of processing for the following reasons:

1. The sheer volume of data is prohibitive to good performance of the visualization.

2. The unit of analysis is country, so whatever information is contained will need to be aggregated.
3. The 5 personality factors each have ten separate survey questions. That is, each factor is composed of the responses of ten different questions. Consequently, these Likert scale responses will need to be summarize and aggregated by personality factor.
4. For the ten questions in each factor, some questions are reverse scaled, meaning that the ordinal/interval response scale will need to be inverted before combined with other responses.
5. Only country abbreviations are included in the data set, so we'll need to join full country names to these. Assuming abbreviation convention is standard, this shouldn't be an issue.
6. 5 factors cannot be visualized in 2-D so PCA is needed for any scatterplots.

Jordan Pyper has nearly a decade of experience in cleaning, process, and analyzing data so this is not an issue. He will be using the R statistical software package to carry out the processing and any additional analyses.

6 Exploratory Data Analysis

a. PCA Design

Our PCA Design was inevitably scrapped but the ideas and concepts remained. A PCA plot would show how similar multi-dimensional data are. This idea stuck with us as we designed new and interesting ways to compare countries easily. This led us to incorporating a brush tool in conjunction with parallel coordinates.

b. Sample Size

After looking at which countries vary the most, we found that the outliers are often the ones with a low sample size. We felt this would be deceiving if we didn't disclose the number of respondents as well in each country. Therefore, we added this data to our tables and detailed views. The user is now able to even sort by the number of respondents as well.

7 Design Evolution

a. Minimum Viable Product

1. Display personality profile for each country.
 - (a) This would be for each personality factor.

(b) For the sake of simplicity, this will likely come in the form of aggregated data.

2. Permit comparison of countries.
3. Enable the user to see global patterns, at a glance.

b. Optional Features

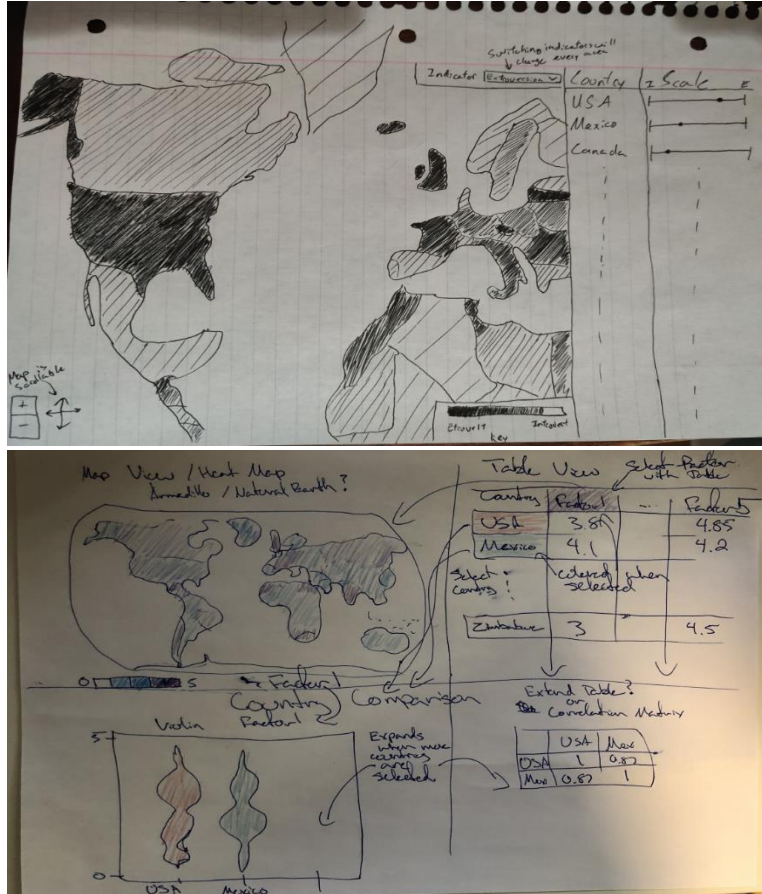
1. There is great diversity in every country. Provide some feature that allows the user to understand this.
2. Provide distribution data for each country to get an idea of variability.
3. Allow the user to focus on a country or countries

c. Initial Designs

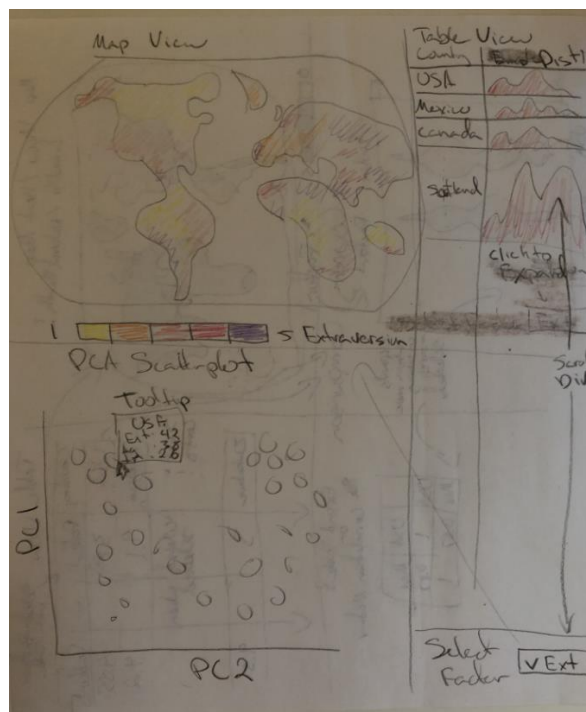
The following are prototypes that we had discussed. The first proposes a visualization with two main features.

1. Map View. This geographic view of the earth includes a heat map that will display a selector personality factor.
2. Table View. This will display the country level tabular data with each country as a row and each column as a personality factor. The average (maybe median or both?) score for the factor is provide in each column. Selection in the map and country comparison are made by clicking on column and row headers.

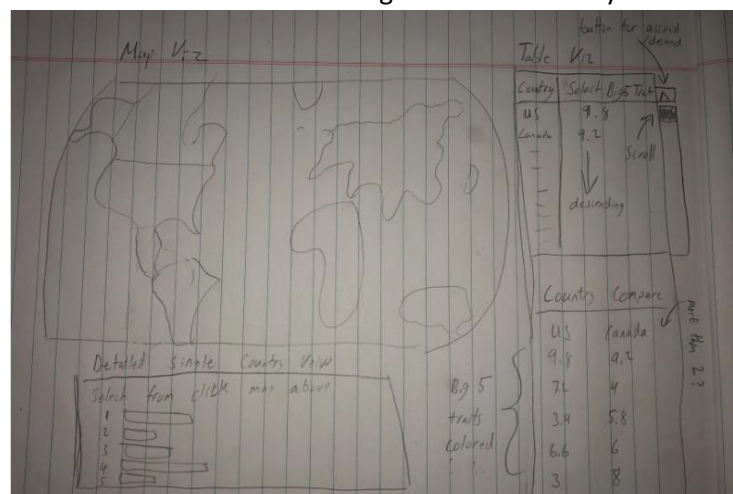
The second sketch is similar, but it adds the option to compare countries. The idea is to allow the user to drill down into a bit finer detail to for a country and compare. Personalities a clearly diverse in all country so we will show a distribution in the form of a violin plot of each selected country. The exact channel in the comparison is not so important as to have an areas where the user can click on some countries and compare their info.



The third prototype combines some features of the first and the second sketches, where the map view is retained and the table view shows country and an empirical density for each country. A scatter plot is provided for a different way of comparing countries. This scatterplot is the of the first two principle components so that we can see separation or grouping in the data. The nature of this plot allows the user to see how similar or different countries are by their position in the scatter plot. A tooltip will provide the personality information for the country so the user can tell which country is which in the plot.



The principal component scatter plot is denser with information and allows the user to see clustering and similarity in the data at a single glance. In general, Euclidean distance can be thought of as a similarity metric for this plot.



The last prototype also includes a map view and table view with a bar graph and table to display country specific information.

Our final version will be best represented by the third sketch with the heat map, table view, and scatterplot.

d. Design Revisions (Peer Feedback)

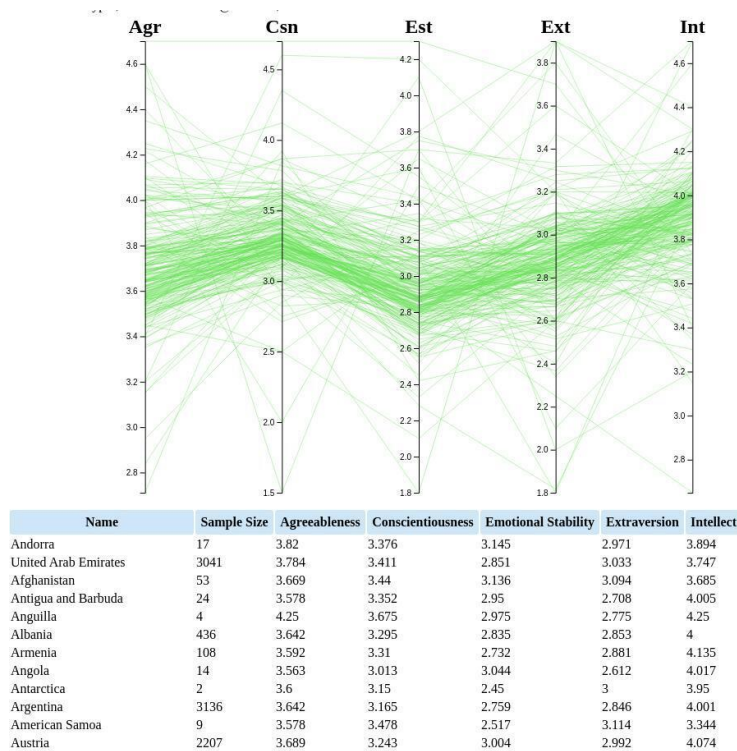
During our first peer review session, we decided that a PCA scatter plot would not be conducive to ease of use for lay people as many have not had exposure to such plots. We have since revised our plot section to use a brush plot technique to show similar countries.

e. Design Revisions (Milestone)

After meets with the TA, we all agreed that we needed to utilize space more in our website. We also needed to add more clarity to our parallel coordinates. To remedy these issues, we colored the lines in the parallel coordinates, so they will not all blur together. We next needed to tackle user clarity. This included storytelling and signposting to give the user a clear idea of what the data is describing.

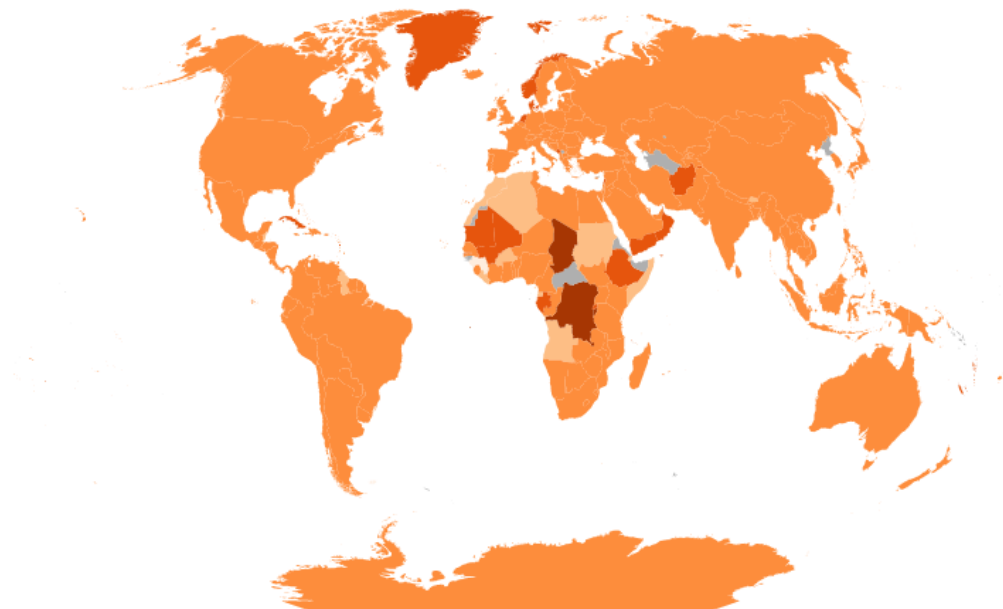
8 Implementation

a. Project Milestone (11/15)



Above is our parallel coordinates view that will allow the user to brush over points and add them to the detailed comparison box below for a further breakdown of what was brushed over.

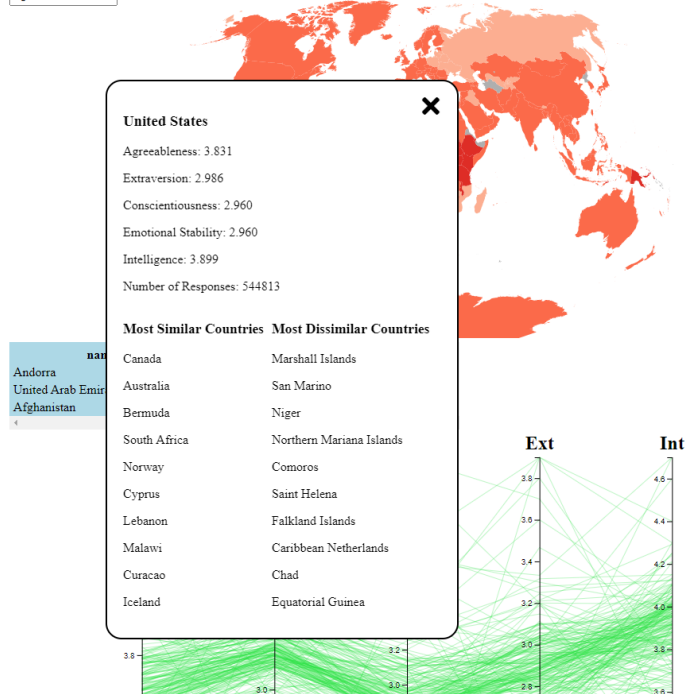
Extraversion



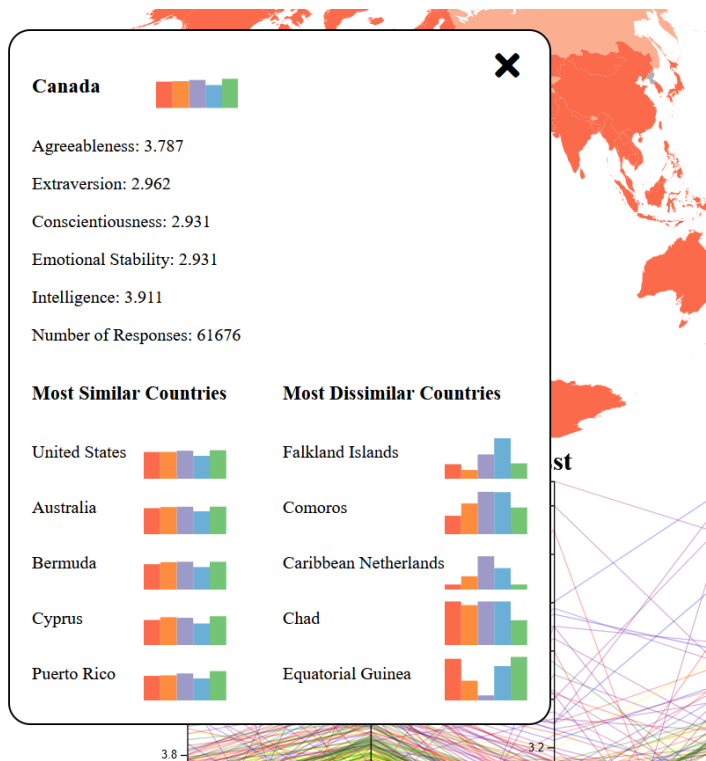
This is our map view in its current state. It is a heat map based on the trait selected in the drop-down menu. The colors will change based on the trait selected and the darker the color the larger the number is in that trait category.

b. Map View Evolution

Agreeableness

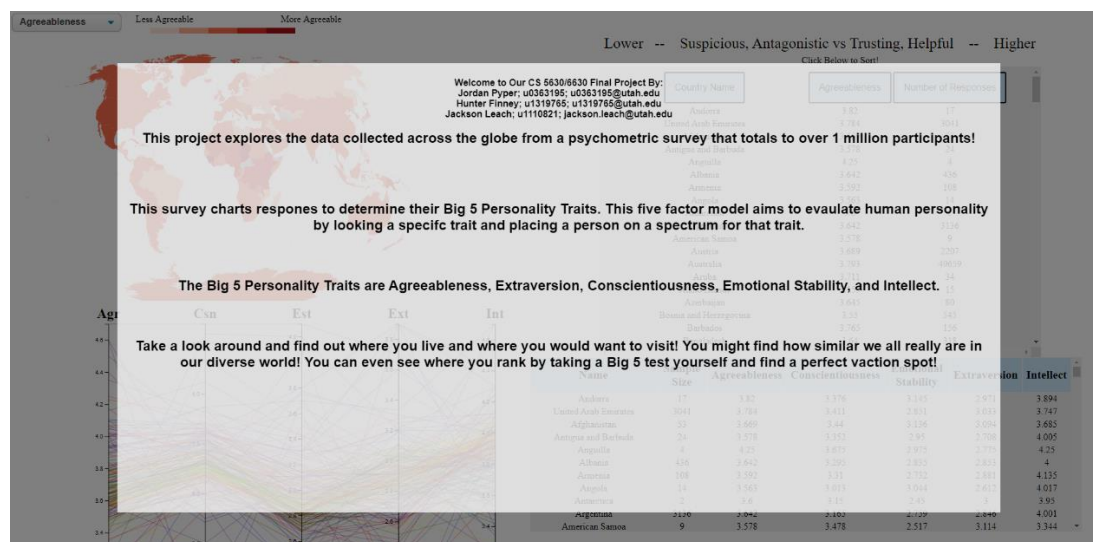


Above is our detailed map view as of November 20th. From here we can see what the specifics of any country focused on and what countries are most similar and most different to that select country.



By November 30th the map view now included bar charts of the countries listed to give the user a quick gist of each country and their average traits.

c. Sign Posting and Storytelling



We next needed to focus on our user experience. To aid in this we added an overlay that will disappear on click to introduce the project briefly. The text above is the final iteration.

Lower -- Suspicious, Antagonistic vs Trusting, Helpful -- Higher

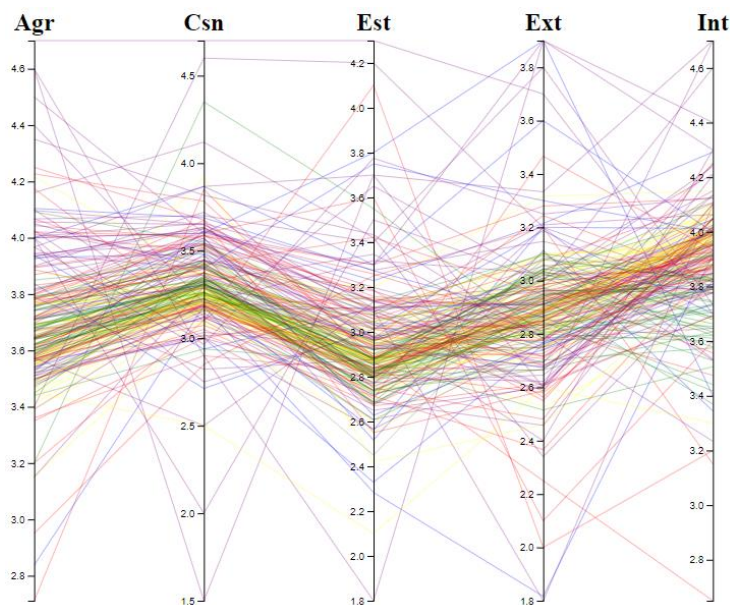
Click Below to Sort!

Country Name	Agreeableness	Number of Responses
Andorra	3.82	17

To remedy our critique of clarity we added signposting above our tables to explain what each trait of the Big 5 means and how to read the data as well as interact with it. Above and below are examples of this signposting.



d. Parallel Coordinates



After our project milestone meeting, our TA recommended us to alternate colors in our parallel coordinates to improve clarity. This is now our final implementation of coloring the parallel coordinates chart. We also increased the area needed to click for the brush to appear.

9 Evaluation

a. Peer Review (11/5)

Alexandra Hurst

-PCA usage felt confusing for those not initiated. If the target audience is general lay people a PCA would not be conducive for this visualization. Multiple comments were made to choose another encoding for information to provide a general look at all dimensions of personality. The large portion of feedback was at this single point. It left most other sections feeling neglected as a result. This has been addressed by Jordan Pyper by reworking the section.

- Mentioned the table that had the empirical density plot in the prototype drawing, and asked about its purpose. They weren't sure what the histograms or empirical densities were supposed to describe. This will be addressed with more clear storytelling and descriptors.

-A general concern was to use not only numbers to represent data. This wasn't directed specifically at any component of our design but a warning of a trap we could fall into. This concern resonated with us as we are looking to use color and other visceral streams of information to connect with the user.

Ky Lamoureux

-The map was generally well accepted. Not much feedback other than the idea was sound and our concepts seem to work.

-The space utilization was greatly praised. This tenant will carry over to our final design and will maintain a priority.

- One theme in their concerns was understanding the purpose of a particular graphic. We might need to make sure it is apparent what the information is describing or add detail to clear up any confusion.

General Concerns

-This peer review was helpful to get to verbalize our concept to critics. Having a third party to tell you where you've gone wrong and where your strengths lie was very helpful. The main problem was that the feedback wasn't directed even though there is a document to guide the feedback. Our map was mostly neglected for feedback and the table view felt pushed aside for criticism on the PCA chart. Most of our feedback was on this chart which ultimately just resulted in scraping the PCA entirely.

b. Self-Evaluation

Through our development we found that many countries are extremely similar on average. The change in personality across cultures did not come across as apparent as we thought it would be. This isn't a failure however. Finding how similar we all are on the inside while being so diverse on the outside is a great result to have. We had great success in wrangling data this large thanks to our backgrounds. We also presented effective visualizations for this data and showed it in a clear way. We also explored different ways to present this data rather than just geographical and succeeded in that as well. We feel proud of our work done here. There are improvements to be made for sure. Added a more directed user experience could be helpful as well as selecting multiple countries to compare directly to each other. We also could do more behind the scenes statistics than we already have to present data such as variability. While more statical data may alienate audience members, more information to the user hardly ever a poor choice.

The results to come of this project is that we found many countries are strikingly similar in personality on average. The change in personality across cultures did not come across as apparent as we thought it would be. This isn't a failure however. Finding how similar we all are on the inside while being so diverse on the outside is a great result to have. We had great success in wrangling data this large thanks to our backgrounds. We also presented effective visualizations for this data and showed it in a clear way. We also explored different ways to present this data rather than just geographical and succeeded in that as well.