W13 Project Plan

Tasks

1. Gather national characteristics data from different sources. (Georg Reintam 8h, Taavet Simo 8h)

   The main tool to find the data will be google and once the data is found we will use the pandas library for python to store the data in a way it is easy to work with. The part that makes this task time consuming is finding the characteristics for which many countries have reliable data. When determining a new datasets compatibility with our dataset the first step is to make sure that the country names match with our dataset. The second step is to add the new characteristics to our dataset. Finally we check if the new characteristics columns contain values for each country we have listed in the first column, if there are many empty slots the characteristic column will be removed, if only a few countries are missing then we determine if those countries can be removed from the dataset.

2. Cleaning and evaluating the data. Some additional data gathering may also be required. (Georg Reintam 7h ,Taavet Simo 7h)

   In this step we evaluate the created dataset as a whole using pandas and numpy. Values that appear out of place will be double checked and possibly removed. If the dataset becomes too thin, additional data will be gathered.

3. Analysing and visualizing the cleaned data (Georg Reintam 5h, Taavet Simo 5h)

   We use pandas library and possibly other visualization tools to present the information gathered in our dataset. We will also try to high light interesting information i.e. what percéntage of a nations population watches football or what is the big mac index of each country.

4. Implementing machine learning to find an answer to the research question (Georg Reintam 6h, Taavet Simo 6h)

   In this part we use sklearn to implement a machine learning algorithm that can make a prediction on where someone might live based on our dataset and characteristic values given as test data. We also had the idea of using tableau to make the output more visually pleasing

5. Making conclusions and designing the poster (Georg Reintam 4h, Taavet Simo 4h)

In this task we will design the poster. How we do this is yet to be decided.