# Marriage: is it an outdated concept and if so, why?

# 1 Aims, objectives and background

## 1.1 Introduction

The concept of marriage originated many years before recorded history, but the earliest record of this tradition dates back to about 2350 BC in the Far East [8].

This tradition was then accepted into many different cultures accross the world as time progressed. The concept of marriage has evolved from it's origins (which was one man being married to multiple women) to include mainly monagomy as the guiding principle for Western marriages and then later, same-sex marriage.

It is known from history, that the main purpose for marriage originally was to be an alliance between families (with a political or economical goal in mind). Marriages were mainly arranged , and in some cultures they still are arranged.

Although marriages can by polygamous, same-sex, monagamous, arranged or out of love, it is still widely recognized that one of the main caveats of marriage is a bond between two people that involves responsibility, commitment, legalities and challenge.

Marriages have evolved as a concept to be more inclusive, however as time goes on it can be seen that less and less young people are deciding to get married. I aim to show this data cohesively in this project. Furthermore, I aim to identify trends between economics, social, political and other factors that correlate to the downward spiral in this tradition.

## 1.2 Aims and objectives

I would like to explore the following:

- Are marriage rates decreasing throughout the world over time?
  - Measuring marriage trends accross different countries over time
  - Finding a measure to explore the spread of data
- If so, why?
  - What educational correlations are there?
  - What economic correlations are there?
  - Are there any other important factors to take into consideration?

For this project, my aims are as follows:

1. Decide how much data is needed to sufficiently investigate marriage trends over time
2. Decide which data to use for analysis
3. Collect data using by accessing CSV files from government websites
4. Clean and store the data
5. Carry out exploratory analysis on the data and display any trends that are found in a graphical format
6. Find correlations and infer causation from the data sets

### 1.2.1 Scope

The scope of this project includes mainly looking at countries that have a Western culture. It also includes looking for correlations and trying to make inferences about causation. This project will not be going into detail about any one specific country and will not be going into deep detail about the possible socio-economical reasons for the marriage rates being how they are.

## 1.3 Data

### 1.3.1 Data requirements

One main dataset was used for this project. This dataset is from [1] it includes a large amount of data from different countries and continents around the world about how their Crude Marriage Rate (per 1000 inhabitants) changes over several years. The original heading for the last column was Crude Marriage Rate per 1000 inhabitants, I changed this to CMR in the .csv file to make this easier to read and use in dataframes. I am limiting my line graphs to only a few western countries and some asian countries. This is because the dataset is too large to plot on one graph. I am also comparing richer countries to less rich countries to see if there is possibly a correlation there.

### 1.3.2 Choice of dataset

The dataset from [1] was chosen because it includes a large amount of data about many countries around the world, these datasets are also recorded over many years and other statisticians have been able to draw successful conclusions about this data.

### 1.3.3 Limits and constraints of the data

***Not all the countries have a complete dataset***

Some of the countries in this dataset only have one or two records. This can skew the data in later stages of the project when I begin to workout the slope of the data. This also means that they can cause division by zero in calculations of the slope which reult in some countries having a NaN slope which has to be removed from the dataset. This means that certain countries are not accurately represented in this report. There also is an issue where the country names are not all exactly the same as the country names contained in geopandas. Thus some data cannot be output onto the map and thus some countries could lack representation in this project.

***The most extensive data record for countries are mainly for rich countries***

The dataset mainly has lots of data for the richer countries throughout the world such as the United Kingdom, United States, Europe etc. This means that the data is more accurate for richer countries but may be inconclusive for less rich countries which have a signifiantly lower amount of records.

### 1.3.4 Methodology

The methodology for this project is as follows:

- The dataset with the most general worldy data about marriage is selected
- From there the dataset is analysed and output to graphs to affirm initial hypothesis
- The data is then taken through a more rigorous analysis to find the deeper correlations between specific countries and their marriage rates
- The data is visualized into a pie chart and maps to make visualizing spread easier

# 1.4 Ethical Considerations

## 1.4.1 Impact of using data to propose analysis

It is important to acknowledge that this report is mainly relying on one dataset, and this dataset mainly represents the richer countries in the world. This means that inferences made from this dataset could be inaccurate and could inaccurately represent different communities. Thus it must be stated that in this project I am limiting the potential harm to others with the way I analyse this data by:

- This project is only focused on analyising the objective features of the dataset, this project is not intending to make any judgements on any country/person, but rather aims to find correlations objectively within the dataset
- The conclusions that are found in this project do not claim to be a full analysis on all the marriage data in the world, and it is understood that this is a sample that could already contain biases that could affect the results of this project

## 1.4.2 Usage of the data

The data used for this project was obtained from "Our World in Data" which is governemnt website. There is no disclaimer stating that I cannot use the dataset. It is also stated on the website that all visualizations, data and code used on this website are completely open access under the Creative Commons BY license. The website states that I am able to use, distribute and reproduce this data in any medium provided that the sources are credited.

This applies to any person who would like to recreate this project using the same dataset that is used here.

## 1.4.3 Usage of images

Images used for this project are referenced and were taken from government sources where image use is allowed based on the correct referencing of any image or data used.

# 2 Initial General Data Analysis

# 2.1 Import Libraries

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df = pd.read_csv("marriage-rate-per-1000-inhabitants.csv")
```

## 2.2 General First Review of Dataset

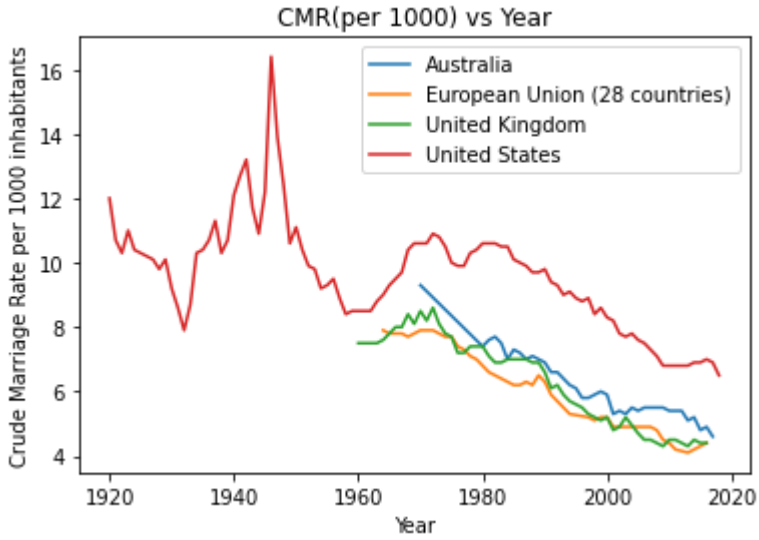
We will output some data to graphs to get a general idea of some of the trends of countries we are interested in, we will first evaluate some richer countries and then some poorer countries (that have enough records to be accurate).

#### Graphs of rich country crude marriage rates over several years

#### Wealthy western countries

In [2]:

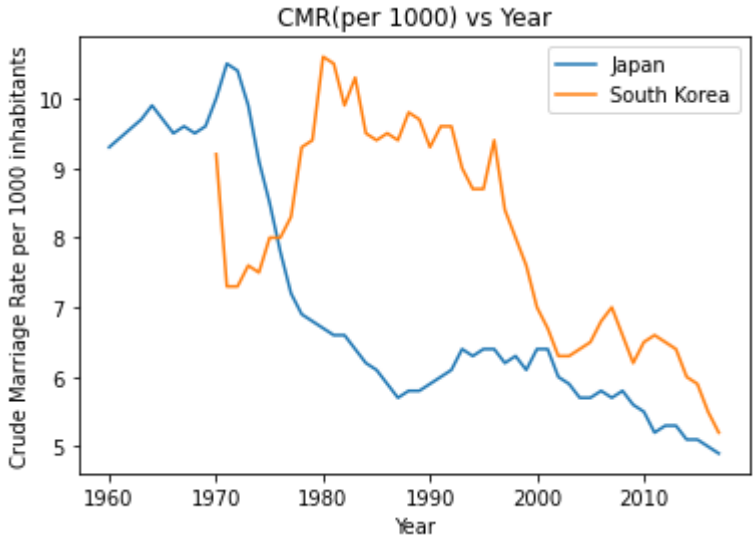
```
# Clean the data, there is too many countries to show on just one graph
# We are focusing our study on a few western countries for this graph section

# I learned how to do this from [2]
dfAUS = df[df['Entity'] == 'Australia']
dfAUS.set_index('Year', inplace = True)
dfAUS.groupby('Entity')['CMR'].plot(legend=True)

dfEU = df[df['Entity'] == 'European Union (28 countries)']
dfEU.set_index('Year', inplace = True)
dfEU.groupby('Entity')['CMR'].plot(legend=True)

dfUK = df[df['Entity'] == 'United Kingdom']
dfUK.set_index('Year', inplace = True)
dfUK.groupby('Entity')['CMR'].plot(legend=True)

dfUS = df[df['Entity'] == 'United States']
dfUS.set_index('Year', inplace = True)
dfUS.groupby('Entity')['CMR'].plot(legend=True)


plt.xlabel("Year")
plt.ylabel("Crude Marriage Rate per 1000 inhabitants")
plt.title("CMR(per 1000) vs Year")

plt.show()
```




**Wealthy asian countries**

In [3]:

```
dfJAP = df[df['Entity'] == 'Japan']
dfJAP.set_index('Year', inplace = True)
dfJAP.groupby('Entity')['CMR'].plot(legend=True)

dfKOR = df[df['Entity'] == 'South Korea']
dfKOR.set_index('Year', inplace = True)
dfKOR.groupby('Entity')['CMR'].plot(legend=True)


plt.xlabel("Year")
plt.ylabel("Crude Marriage Rate per 1000 inhabitants")
plt.title("CMR(per 1000) vs Year")

plt.show()
```

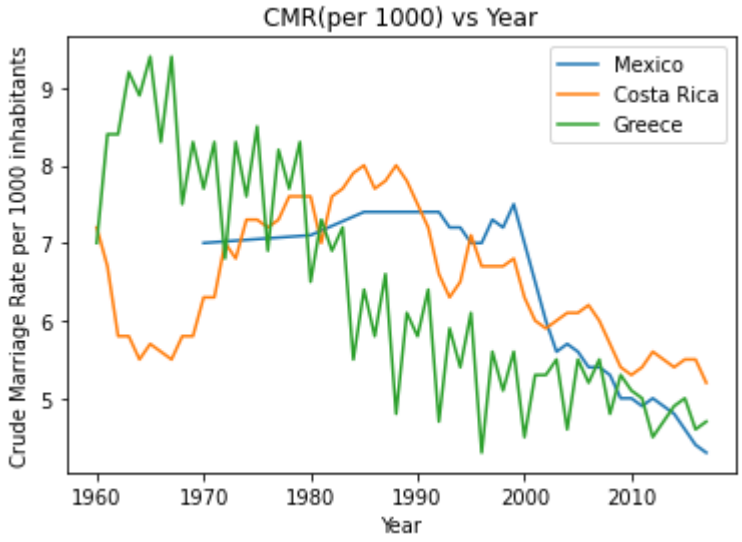



The analysis here is of the richer countries in the world to see if this is correlated to the marriage rate declining. We can see by the graph that the number of people getting married over the time period of 1960 until 2026 has steadily decreased in Australia, the European Union, the United Kingdom and the United States (the United states has more data than any other entity in this dataset). It can be seen that for the United States in 1920, which was shortly after the first World War, there were almost double as many marriages per 1000 inhabitants than there is now. In the 1930s, which was the time of the Great Depression, the rate sharply dropped. Then in 1946 , the year after the Second World War ended, marriages sharply peaked. This decline started in the 1970s, and since then marriages in the US have seen a steady decline by almost 50%. The UK and Australia have also seen marriage rates declining for decades.

**Graph of poor country crude marriage rates over several years**

In [4]:

```
# Clean the data, there is too many countries to show on just one graph
# We are focusing our study on a few western countries for the graph section

# I learned how to do this from [2]
dfAUS = df[df['Entity'] == 'Mexico']
dfAUS.set_index('Year', inplace = True)
dfAUS.groupby('Entity')['CMR'].plot(legend=True)

dfEU = df[df['Entity'] == 'Costa Rica']
dfEU.set_index('Year', inplace = True)
dfEU.groupby('Entity')['CMR'].plot(legend=True)

dfUK = df[df['Entity'] == 'Greece']
dfUK.set_index('Year', inplace = True)
dfUK.groupby('Entity')['CMR'].plot(legend=True)


plt.xlabel("Year")
plt.ylabel("Crude Marriage Rate per 1000 inhabitants")
plt.title("CMR(per 1000) vs Year")

plt.show()
```




As outlined above, the data for non-rich countries is sparse. However after looking at the dataset for non-rich countries that had a reasonably large amount of data records describing them, I chose Mexico, Costa Rica and Greece. It can be seen that these countries are also seeing a decline in marriages throughout the years

# 3 Specific data and correlation analysis

## 3.1 Check the entire dataset to see if we can make the assumption that in general the crude marriage rate is decreasing over time

In the analysis below, I will be looking for a way to generalize whether each entity is either increasing their crude marriage rate (per 1000 inhabitants) over time, or if they are decreasing over time. I have decided to do

this by calculating the slope between the earliest record for an entity and the latest record of an entity. I am using this to create a trend line and from there calculate whether the slope is increasing or decreasing. I then need to do a deeper analysis on the spread of this data accross the world to see if it matches any other common data like wealth spread and educational spread. If there is no correlation there, then I will need to conduct more research on what could be causing the data to be spread the way that it is.

**Calculate Slope**

In [5]:

```
# There will be warnings due to divides by zero but these are not necessary for display
# from the dataset
import warnings
warnings.filterwarnings('ignore')

# The slope of all the entities need to be calculated to determine
# if we can make the assumption that in general marraiage rates are going down accross

# This is a function to determine the gradient of the slope
def slopeCalc(x1,x2,y1,y2):
    x = (y2 - y1)/ (x2 - x1)
    return x

# All the unique entities are being found here to be examined
Entities = df['Entity'].unique()
slopeTable = pd.DataFrame(columns=['Entity','Slope'])

# The unique entities are being looped through to find the slope points, we are looking
# at the slope from the earliest time point to the latest time point and finding the as

for entity in Entities:
#   Here we create a new dataframe just for each unique entity
    dfTemp = df[df['Entity'] == entity]

#   Then we are finding the minimum year and the maximum year
    maxYear = dfTemp['Year'].max()
    minYear = dfTemp['Year'].min()

#   Now we find the assosciated CMR for each of these years
    dfTemp2 = dfTemp[dfTemp['Year'] == minYear]
    CMR1 = dfTemp2['CMR'].values[0]

    dfTemp2 = dfTemp[dfTemp['Year'] == maxYear]
    CMR2 = dfTemp2['CMR'].values[0]

#     print('Entity ' + str(entity) + ' MaxYear ' + str(maxYear) + ' MinYear ' + str(mi

#   We calculate the slope now for each of these values
    slope = slopeCalc(CMR1,CMR2,minYear,maxYear)

#   Add the data to another data frame
    new_row = pd.DataFrame({'Entity':[entity], 'Slope': [slope]})
    slopeTable = pd.concat([slopeTable, new_row], ignore_index = True, axis = 0 )


# some of the data elements had only one entry so they returned a NaN value,
# we need to cleanse our data of these values
slopeTable = slopeTable.dropna()

# Now we have a database of slopes associated with Entities
slopeTable

```

Out[5]:

| | Entity | Slope |
|---|---|---|

| | Entity | Slope |
|---|---|---|
| **0** | Albania | 570.0 |
| **1** | Algeria | 8.163264 |
| **2** | American Samoa | -4.137931 |
| **3** | Andorra | -25.714287 |
| **5** | Anguilla | -2.560976 |
| **...** | ... | ... |
| **175** | Uruguay | -7.659575 |
| **176** | Uzbekistan | -13.47826 |
| **177** | Vatican | -2.941177 |
| **178** | Venezuela | -15.652174 |
| **182** | Yemen | 9.999998 |

169 rows × 2 columns

Now we need to find out how many of the records have a negative slope so we can see if there is a correlation present

In [6]:

```
1 len(slopeTable)
```

Out[6]:

169

**Calculate distribution of slope**

There are 169 Entities, each with a valid slope calculation, now we need to find out how many of these are negative and how many of these are positive.

In [7]:

```
sumNeg = 0;
sumPos = 0;

# create a list object to store the names of the countries that have increased in marri
PosList = [];

# create a list object to store decreasing values
NegList = [];

# iterating over the dataframe to find how many marriage rates have increased or decrea
for index, row in slopeTable.iterrows():
    if row['Slope'] < 0:
        sumNeg = sumNeg + 1
    else:
        sumPos = sumPos + 1
        PosList.append(row['Entity'])

print('The total amount of countries where marriage rates have declined is ' + str(sumN
print('The total amount of countries where marriage rates have increased is ' + str(sum

# print(PosList)
```

```
The total amount of countries where marriage rates have declined is 108
The total amount of countries where marriage rates have increased is 61
```

This is an interesting result, I originally thought that the amount of countries decreasing in marriages over time would be higher. The dataset does however include non-western societies which follow different cultural norms that originally wasn't in the scope of the report.

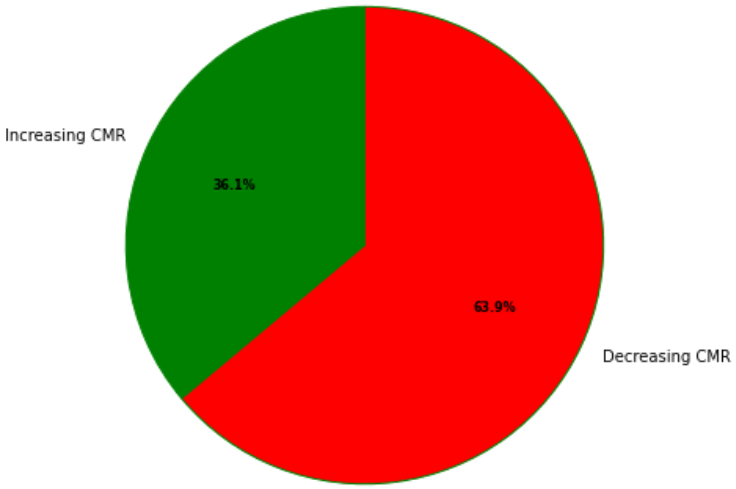
**Visualize Spread with a Pie Chart**

The result can be visualized with a pie chart

In [8]:

```
# green for increasing and red for decreasing
colors = ( "green", "red")

# Wedge properties
wp = { 'linewidth' : 1, 'edgecolor' : "green" }

percentages = [sumPos/len(slopeTable) * 100, sumNeg/len(slopeTable) * 100]

labels = ['Increasing CMR', 'Decreasing CMR']

# # Creating autocpt arguments
# This creates the labels on the wedges of the graph
def func(pct, allvalues):
    absolute = int(pct / 100.*np.sum(allvalues))
    return "{:.1f}%".format(pct, absolute)

# The settings for how the pie chart should be displayed
fig, ax = plt.subplots(figsize =(10, 7))
wedges, texts, autotexts = ax.pie(percentages,
                                  autopct = lambda pct: func(pct, percentages),
                                  labels = labels,
                                  shadow = False,
                                  colors = colors,
                                  startangle = 90,
                                  wedgeprops = wp,
                                  textprops = dict(color ="black"))

plt.setp(autotexts, size = 8, weight ="bold")
ax.set_title("Pie Chart showing the proportion of entities with a declining or increasi

plt.show()
```

Pie Chart showing the proportion of entities with a declining or increasing Crude Marriage Rate (CMR) per 1000 inhabitants




## 3.2 Visualize the data on a map and analyse correlations

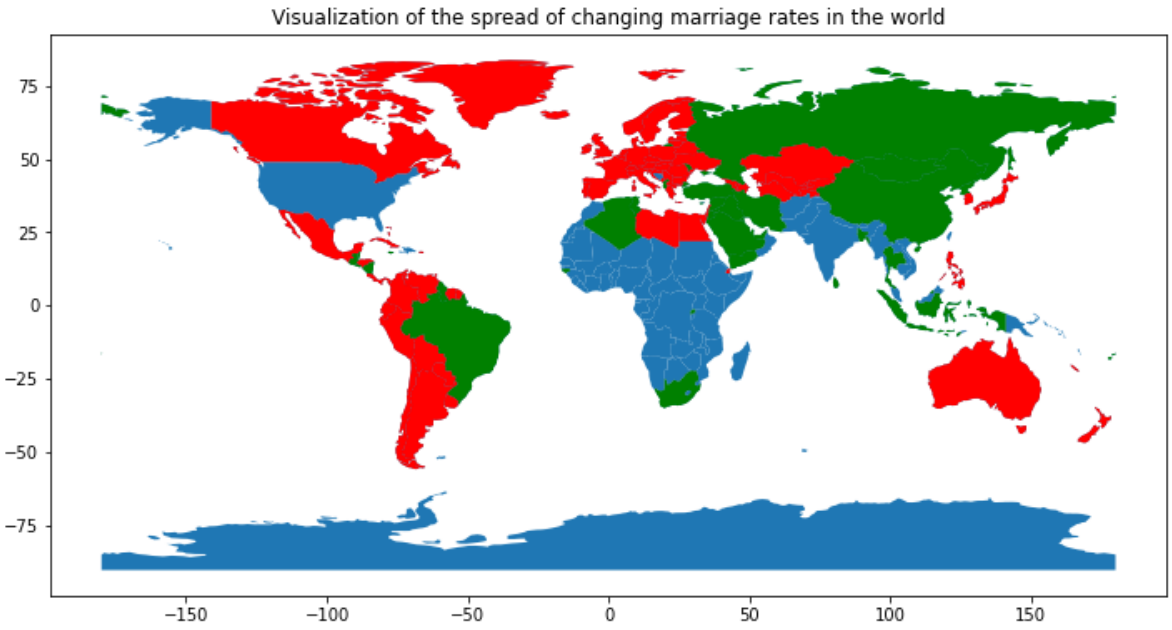
Show this data on a world map so we can see the geographic locations of decreasing or increasing marriage

rates

### Visualize Spread of marriage data on a map

In [9]:

```
# Install geopandas using the command conda install -c conda-forge geopandas in the Ana
# This code was partially adapted from [2]
# You may need to install some dependencies if you are on windows, you need to acces th
# You will need to install GDAL, Pyproj, Fiona, Shapely and Geopandas as shown in [5]

import geopandas

# get the path of the geopandas standard map
world = geopandas.read_file(geopandas.datasets.get_path('naturalearth_lowres'))

# Increase the size of the map
f, ax = plt.subplots(1, figsize=(12, 12))
ax.set(title='Visualization of the spread of changing marriage rates in the world')
ax = world.plot(axes=ax)


# some of the values for this will not show on the map just because the country name ma
# match the country name specified by geopandas
for index, row in slopeTable.iterrows():
    if row['Slope'] > 0:
        try:
            world[world.name == row['Entity']].plot(color='green',ax=ax)
        except:
            print(row['Entity'] + ' is not in the geopandas country names')
    if row['Slope'] < 0:
        try:
             world[world.name == row['Entity']].plot(color='red',ax=ax)
        except:
            print(row['Entity'] + ' is not in the geopandas country names')
plt.show()


```

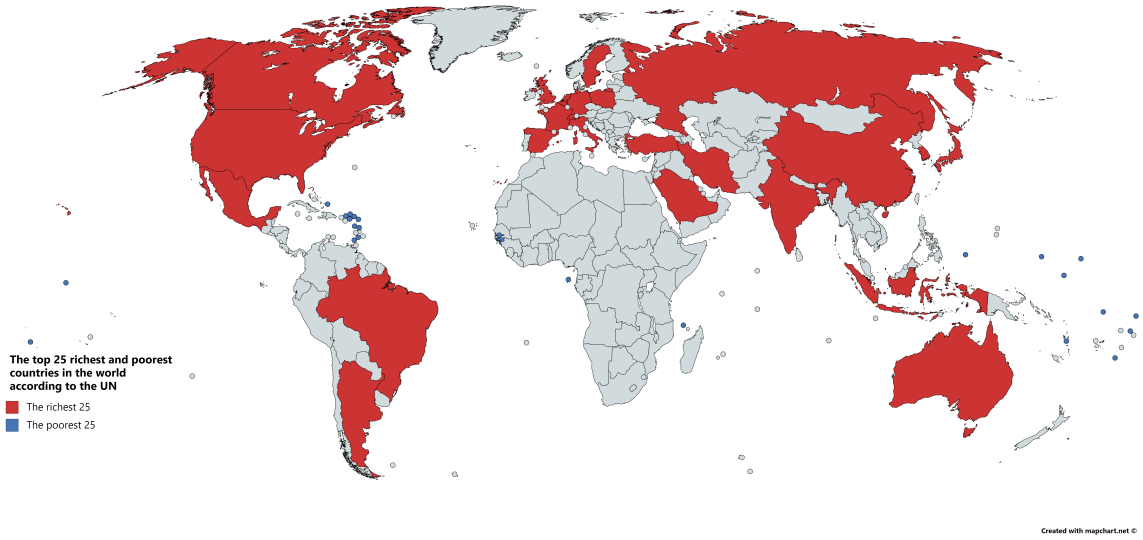


We can see that the red countries have a decline in marriage rates, and the green countries have an increase in marriage rates over time. It must also be noted that all the countries in the dataset are not displayed here because their country name may not match the internal Geopandas country names database. We now can see whether the data correlates to the wealth distribution of countries. The visualization for the wealth distribution was sourced from [6].

### Import an image of the richest countries and compare

In [10]:

```
import IPython

from IPython.display import Image, display

display(Image("RichestCountries.png"))
```



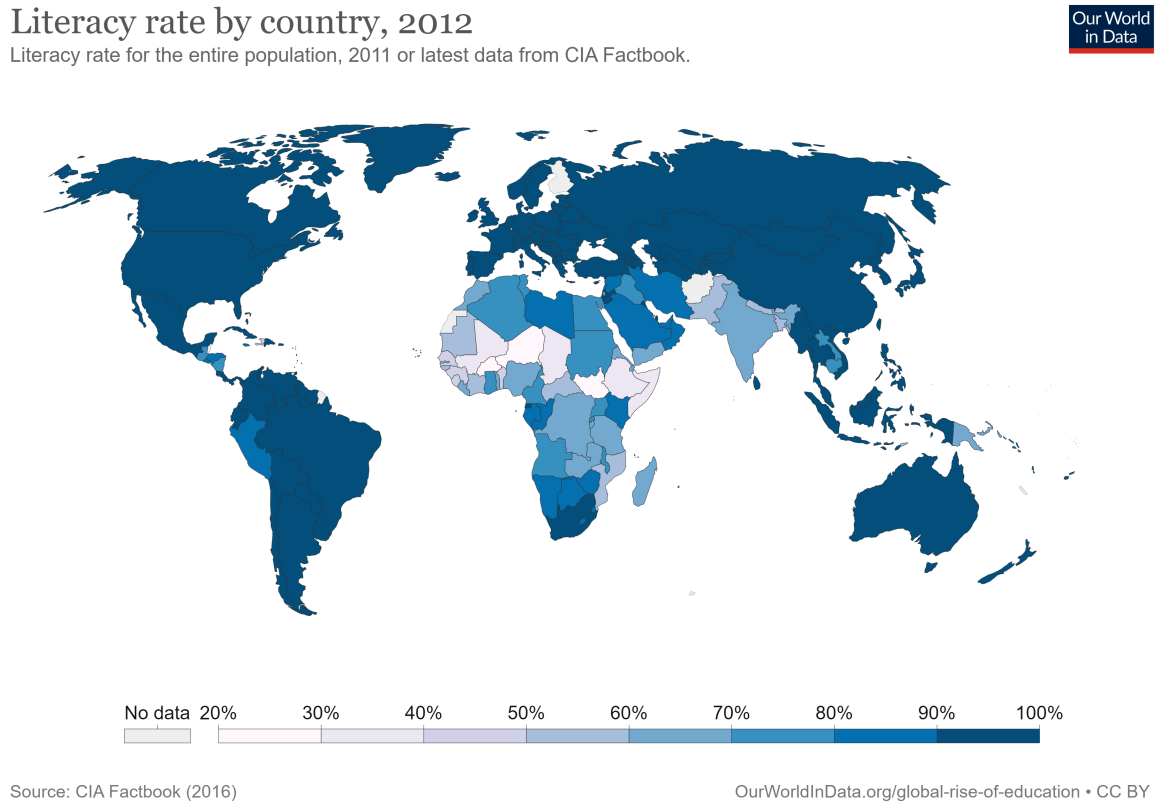

This visualization of the datasets shows that there isn't a visible correlation between the wealth of a country and the marriage rates declining.

### Import an image of the literacy levels of countries and compare

The next investigation that must be conducted is whether we can see a similarity between the eductaion maps and the marriage rates. The visualization below was sourced from [7]

In [11]:

```
import IPython

from IPython.display import Image, display

display(Image("Education.png"))
```




We can see by comparing this map to our map that the drop in marriage rates is happening in highly educated countreis and not highly educated country. This means that the drop is more complex than just education and finances like I originally hypothesized.

# 4 Summary

## 4.1 Conclusions

Through the analysis we can make the statement that the marriage rate is generally decreasing over time. We have also been able to see that this is not caused by the wealth of a country or the general education of its inhabitants. This means that the interaction between time and marriage rates dropping is more complex and will require more research to fully understand why this is occuring. The next step that could be added to this project would be to include more datasets and conduct more specific research in specific countries to design a study that can be more conclusive on the causation that may be taking place here. However my assumption after my analysis is that the reasons behind this reduction in marriage rates is due to social factors, which can be difficult to extrapolate graphically.

## 4.2 Summary of prepared data

***slopeTable dataframe:***

In [12]:

```
1 slopeTable
```

Out[12]:

| | Entity | Slope |
|---|---|---|
| **0** | Albania | 570.0 |
| **1** | Algeria | 8.163264 |
| **2** | American Samoa | -4.137931 |
| **3** | Andorra | -25.714287 |
| **5** | Anguilla | -2.560976 |
| **...** | ... | ... |
| **175** | Uruguay | -7.659575 |
| **176** | Uzbekistan | -13.47826 |
| **177** | Vatican | -2.941177 |
| **178** | Venezuela | -15.652174 |
| **182** | Yemen | 9.999998 |

169 rows × 2 columns

The slopeTable describes the data slopes per Entity, it also shows the degree to which the data has changed over time, higher values generally mean a steeper line and more of a difference over time.

# 5 References and Resources

[1] Ortiz-Ospina, E. and Roser, M., 2022. Marriages and Divorces. [online] Our World in Data. Available at: https://ourworldindata.org/marriages-and-divorces (https://ourworldindata.org/marriages-and-divorces) [Accessed 27 June 2022].

[2] Onelinerhub.com. 2022. Python Matplotlib: How to fill countries with colors using world map - OneLinerHub. [online] Available at: https://onelinerhub.com/python-matplotlib/how-to-fill-countries-with-colors-using-world-map (https://onelinerhub.com/python-matplotlib/how-to-fill-countries-with-colors-using-world-map) [Accessed 27

June 2022].

[3] Data Science Parichay. 2022. Pandas - Delete rows based on column values - Data Science Parichay. [online] Available at: https://datascienceparichay.com/article/pandas-delete-rows-based-on-column-values/ (https://datascienceparichay.com/article/pandas-delete-rows-based-on-column-values/) [Accessed 27 June 2022].

[4] Lfd.uci.edu. 2022. Archived: Python Extension Packages for Windows. [online] Available at: https://www.lfd.uci.edu/~gohlke/pythonlibs/ (https://www.lfd.uci.edu/~gohlke/pythonlibs/) [Accessed 27 June 2022].

[5] Montoya, S. and Montoya, S., 2022. How to install Python - Geopandas on Anaconda in Windows - Tutorial — Hatari Labs. [online] Hatari Labs. Available at: https://hatarilabs.com/ih-en/how-to-install-python-geopandas-on-anaconda-in-windows-tutorial (https://hatarilabs.com/ih-en/how-to-install-python-geopandas-on-anaconda-in-windows-tutorial) [Accessed 27 June 2022].

[6] Reddit.com. 2022. [online] Available at: https://www.reddit.com/r/MapPorn/comments/cucx6i/the_top_25_richest_and_poorest_countries/ (https://www.reddit.com/r/MapPorn/comments/cucx6i/the_top_25_richest_and_poorest_countries/) [Accessed 27 June 2022].

[7] Our World in Data. 2022. Literacy rate by country. [online] Available at: https://ourworldindata.org/grapher/literacy-rate-by-country (https://ourworldindata.org/grapher/literacy-rate-by-country) [Accessed 27 June 2022].

[8] Incegd.com. 2022. The history of marriage | Ince | In any case. [online] Available at: https://www.incegd.com/en/news-insights/family-matrimonial-history-marriage#:~:text=It%20is%20widely%20agreed%20that,B.C.%20in%20the%20Far%20East. (https://www.incegd.com/en/news-insights/family-matrimonial-history-marriage#:~:text=It%20is%20widely%20agreed%20that,B.C.%20in%20the%20Far%20East.) [Accessed 27 June 2022].