Suicide Rates Overview 1985 to 2016

Global suicide trends and analysis

Joyce(Yuqing) Jian, Zara(Ruozhuo) Wang, Sherry Huang, Sherry(Yanhuan) Huang

UNIVERSITY OF TORONTO

OVERVIEW

Our project is to uncover trends and patterns in suicide rates in 1985 to 2016. We will examine relationships between suicide rates and age, gender, climate, GDP; trends in suicide rates over the years; and related questions, as the data admits.

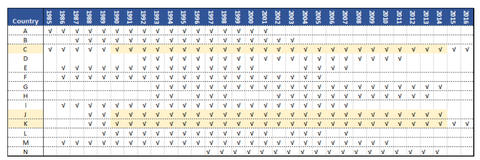
DATA DESCRIPTION

This compiled dataset pulled from four other datasets linked by time and place, and was built to find signals correlated to increased suicide rates among different cohorts globally, across the socio-economic spectrum.

The datasets includes the following information that we used to analyze factors related to suicide rates:

- Years
- Age groups
- Gender
- Country/Location
- GDP

DATA CLEANING

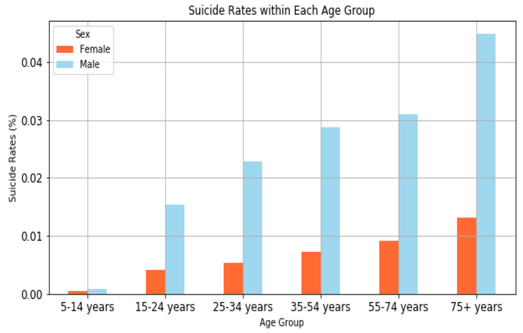


- Dropped unnecessary columns in the DataFrame
- Renamed columns to a more recognizable labels
- Removed missing and duplicate data
- Set up several bins (groups) for analysis

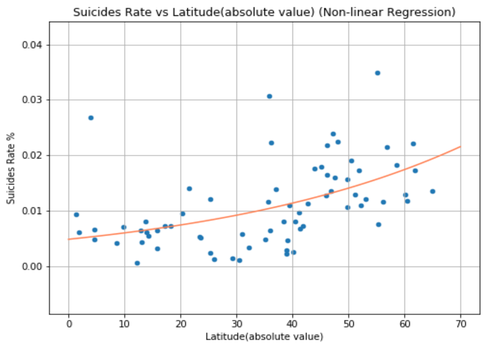
Since a lot of data are missing in 1985-1989, and 2015-2016 in the raw dataset, we used data in 1990-2014 from 38 countries to analyze suicide rates VS Years/Age group/Gender/GDP.

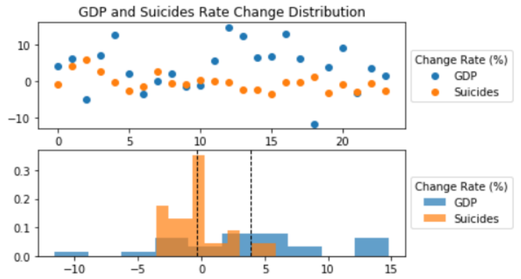
To have the most countries' continuous data, we decided to use the data in 2010-2014 from 71 countries for analyzing suicide rate VS Counties/Location.

VISUALIZATION



Google Heatmap by Suicide Rate (Europe)





EVALUATION

- Over 1990-2014, the number of suicides for female in all countries is significantly less than that of male (1081921 vs 3625386, t-test statistics ≈ -18.77, P-value ≈ 1.458e76 ≈ 0). This difference is also in suicide rates (0.006% vs 0.022%, t-test statistics ≈ -51.72, P-value = 0).
- The correlation coefficient between the absolute value of latitude and suicide rate, also known as R-value, is 0.4727, which means they have a moderate positive relationship.
- Both linear and non-linear regression analysis have been conducted to describe the trend above. Based on the R-value, the exponential model better explains the data.
  - *Linear model*: Suicide Rate = 0.0002017 * Latitude(abs) + 0.003861
    R-square = 0.2235.
  - *Non-linear model*: Suicide Rate = 0.004839 * exp^(0.02134* Latitude(abs))
    R-square = 0.2492.
- To test the correlation of suicide rate and GDP figure, we pull out a T-test on the following hypotheses.
  - *Null Hypothesis (H_0):* GDP is affecting suicide rates
  - *Alternative Hypothesis (H_1):* GDP has no influence on suicide rates
  - T-test result gives a P-value = 0.07841685 (>0.05)

METHODS USED

Pandas, Numpy: Restructured raw data and performed scientific computing for analysis
Matplotlib: Visualized analysis results and plotted findings using line chart, bar plot, scattered plot and Google Heatmap
API, JSON, Python Requests: Retrieved coordinates from Google Map using API and imported into the DataFrame
Data Modeling: Analyzed output using line of best fit (linear/non-linear regression model) and statistical testing

CONCLUSION

- Total suicides rates, including all gender and all age groups were decreasing over the years.
- Suicide rates of age group 75+ dropped the most because of healthcare and pension system improvements.
- Either from suicide rates or suicide counts perspective, suicide is correlated with age.
- Either from suicide rates or suicide counts perspective, suicide is correlated with gender, with female having significantly less suicides counts comparing to that of male.
- Countries located further away from equator have higher suicide rates than countries with lower longitude.
- Suicide rates are higher as the latitude increases.
- Statistically saying, the suicides growth is not significantly related to GDP growth to the extent of a confidence level of 95%.

REFERENCE

https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016