Project 1: Group 4 Proposal

Country Life Expectancy

Introduction

For Project 1, Group 4 will be working with a dataset from Kaggle titled “Country Life Expectancy.” This dataset was chosen due to the size (1365 rows X 19 columns) and amount of numeric data (18 columns). Each collaborator shared interest in a particular subset of the data that we will analyze.

Link to dataset: https://www.kaggle.com/datasets/pedramabdolahi/country-life-expectancy

Research Questions

1. Which countries have the highest death rates related to smoking? What impact do these death rates have on the life expectancy of a country?
2. Do years of schooling increase life expectancy?
3. Is there a correlation between obesity death rates and the GDP?
4. Is there a relationship between deaths by suicide and GDP for all countries? Is there a relationship between deaths by suicide and mean years of schooling for all countries?
5. What is the relationship between government health expenditure and life expectancy for the countries?

Inspiration

We have found one similar notebook on kaggle.com that we might use for guidance with this project:
https://www.kaggle.com/code/philbowman212/life-expectancy-exploratory-data-analysis/notebook#Section-1:-Data-Cleaning

In addition, we have found an article similar to our topic that can be used as inspiration for our project: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9653205/

Visualizations

1. For question 1 above: the plan is to have a line chart showing the rate of change of smoking deaths over time for all countries in the data set.
2. Another visualization for question 1 is a scatter plot of life expectancy vs. smoking death rates for all of the countries.
3. For question 2, a possible visualization is a bar chart (could be horizontal if that is best) showing the years of schooling against the life expectancy.
4. For question 3, there will be a scatter plot to find the correlation between obesity death rates and GDP of the countries.

5. For question 3 again, a line graph will be used to show the rate of change of obesity deaths over time.
6. For question 4, a series of scatter plots should be made to identify the possible relationship between the deaths by suicide and GDP and the deaths by suicide and mean years of schooling.
7. For question 5, we can use a scatter plot showing the government health expenditure vs. the life expectancy and incorporate the GDP values to scale the points.
8. For question 5, another useful visualization would be a geographic map showing the countries are colored according to expenditure and then overlaid with a bubble based on the life expectancy.

Linear Regression

We will be able to do a regression for any of the research questions that we have proposed but the strongest contenders for a linear regression would be: “What is the relationship between government health expenditure and life expectancy for the countries?”

Color Palette

We have found this color palette using coolors.co:
https://coolors.co/09f574-c16200-881600-4e0110-4a7c59

Roles & Responsibilities

Everyone should be participating in the data cleaning; even if it will be done on one computer.

Ethan Tebbe - Question 1 data wrangling and visualizations.
Stephen Ferrier - Question 2 data wrangling and visualizations.
Ayannah Clouden - Question 3 data wrangling and visualizations.
Sara Ruth - Question 4 data wrangling and visualizations.
Spencer Garrett - Question 5 data wrangling and visualizations.

Each person will create their slides for the presentation based on their questions.

Github Link

This is the github link to our project: https://github.com/BramSunner/project-1-group-04