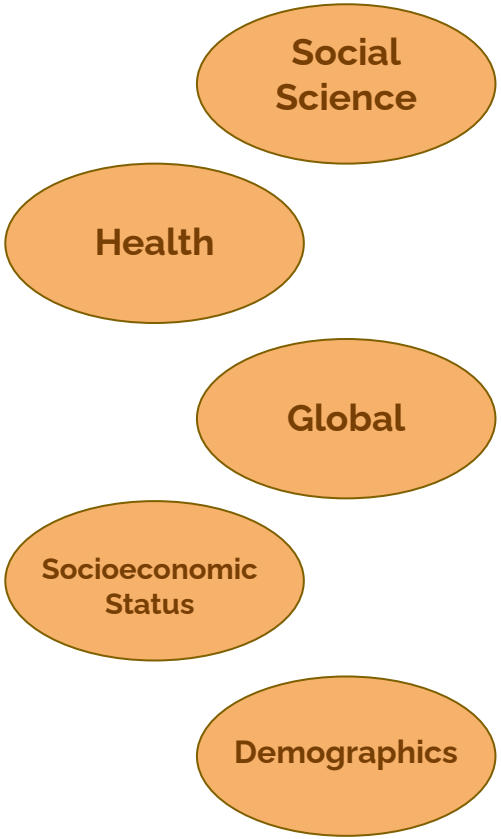
Life Expectancy: A Statistical Analysis

Project 1 - Group 8

Anna Sanchez, Jennifer Reyna, and Shondeen Hechter



Motivation & Summary

Hypothesis: *Average life expectancy amongst countries is affected by socioeconomic factors such as early vaccination, GDP, and schooling.*

Question 1: *Do early life vaccinations affect average life expectancy in developing countries?*

TRUE

Question 2: *Does increased schooling positively correlated with increased life expectancy?*

TRUE

Question 3: *Are average life expectancy and infant deaths affected by GDP in developing countries?*

TRUE

Questions & Data

kaggle

Data compiled by Kumar Rajarshi (2018)
Global Health Observatory (GHO) data repository under World Health Organization (WHO)

Q1: Life Expectancy VS Early Vaccination

mean [Life Expectancy in age] [Country] [Years 2000-2015] [Polio Vaccinations]

mean [Life Expectancy in age] [Country] [Years 2000-2015] [Diphtheria Vaccinations]

Q2: Life Expectancy VS Schooling

mean [Life Expectancy in age] [Country] [Years 2000-2015]

mean [Number of years Schooling] [Country] [Years 2000-2015]

Q3: Infant Death VS GDP Per Capita

mean [Infant Death] [Years 2000-2015]

mean[GDP Per Capita] [Years 2000-2015]

Kumar Rajarshi. (2018). Life Expectancy (WHO)(1)[Dataset]. Available from Kaggle Web site: https://www.kaggle.com/kumarajarshi/life-expectancy-who

Data Cleanup & Exploration

Exploration & Findings:

- Elaborate Data Set
  -22 Columns
  -193 Countries (Unique Values)
  -Data for 2000-2015/Country
  -18 variables for Life Expectancy

- Final Dataset
  -Merged sets from GHO and WHO
  -Four Overall Categories:
    Immunization
    Mortality
    Economics
    Social Factors

Insights & Problems:

- Varying Operationalizations
  -Probability/
  -Deaths/ thousand
  -Cases/thousand
  -Entire population
  -Prevalence (%)
  -Years/

- Varying Data on Diseases
  - Immunization Coverage (%)
  -Deaths per 1,000
  -Reported Cases per 1,000

- Missing Values
  -Missing Years in Countries
  -Missing from less-known countries

Cleaning Data & Solutions:

- Rename and Formatted Columns
  -Made Column Names Descriptive
  -Formatted column values to be reflective of column name (i.e. percentages)
  -Converted per/Thousand to Percentage

- Dropped Null Values
  - .dropna(how='any')

- Converted Data Types
  -Converted all values to float
  -data = data.astype(float)

Data Analysis

Steps:

- Filter data using .loc to get only developing countries
- Began by isolating the values via column manipulation
- .groupby the different variables we wanted
- Gathered means
- Combined into data sets
- Plotted linear regressions and conducted analysis

Figure 1 - Question 1

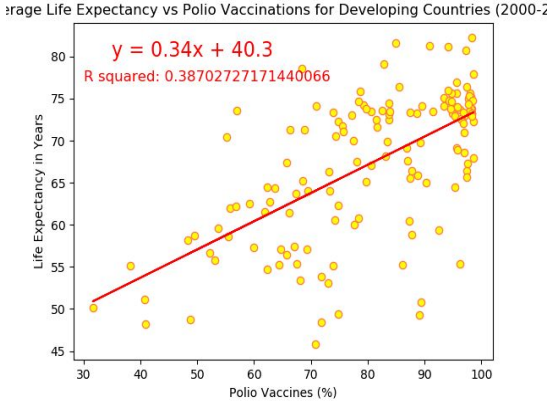


Figure 2 - Question 2

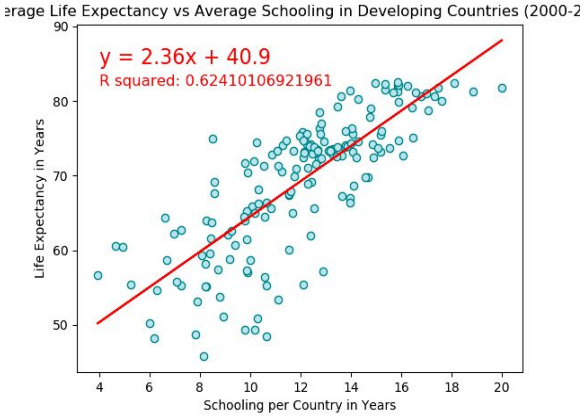
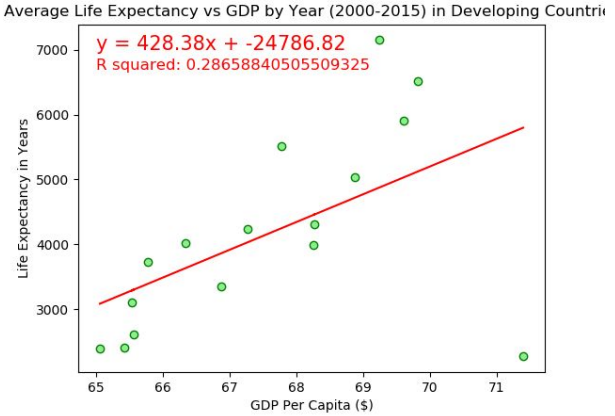


Figure 3 - Question 3



Discussion

Further Analysis:

Life expectancy vs Diphtheria vaccinations:

- Additional regression was ran for Diphtheria vaccinations and results were similar to Polio vaccinations.
- Medical Development such as vaccinations contribute to a greater life expectancy in developing countries.

Infant Deaths by GDP yearly in developing countries:

- Based on Figure 5 infant deaths on average decrease by GDP growth each year.
- Favorable economic factors steadily help provide moderate living and help increase life expectancy at birth.
- The correlation coefficient is negative between the two variables, based on the trend as GDP increases, infant mortality decreases.

Conclusions:

- On average life expectancy is positively correlated with early vaccination, early schooling and GDP growth.
- Indeed, social, economic and preventative factors affect LE in conjunction with biological factors

Figure 4 - Question 1

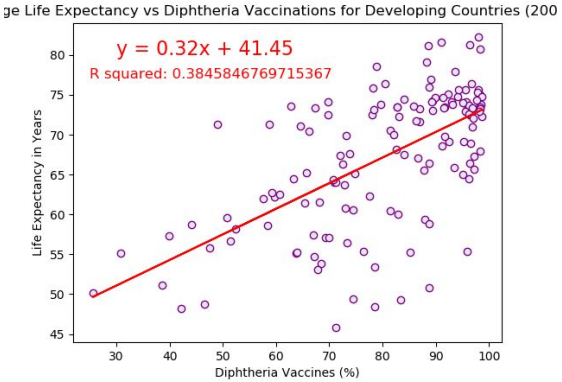
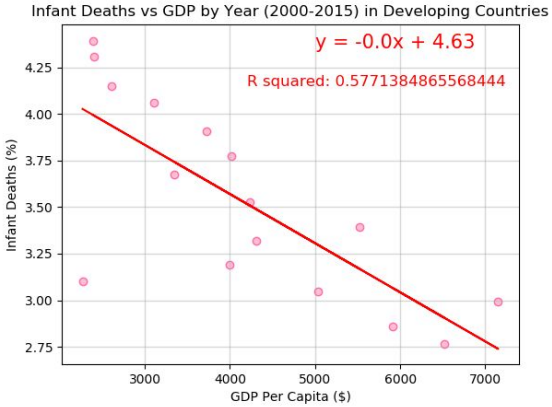


Figure 5



Post Mortem

Difficulties:

- Determining the proper statistical analysis to apply
- Regressions vs Independent T-Test or ANOVA
- Defining a clear hypothesis and subsequent research questions
- Dataset had inconsistent operationalizations, which limited the amount of variable to test at once

Prospective Analysis:

- Further analyze LE over the years to measure any improvements
- Exploring other factors in dataset that affect LE
- Use API and Geo-Mapping for visualizations
- Determine if climates affect factors for LE using weather API
- Identifying individual factors to run further statistical tests in order to accept or reject null and fully support hypothesis

THANK YOU

QUESTIONS - FEEDBACK - COLLABORATIONS

Anna Sanchez
Jennifer Reyna
Shondeen Hechter

DATA ANALYTICS BC- PROJECT 1