Life Expectancy Across Continents

IOWA
Department of Mathematics

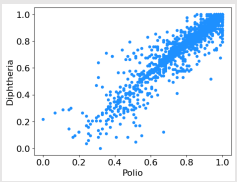
Jessa Rhea

Applied Mathematical and Computational Sciences, University of Iowa

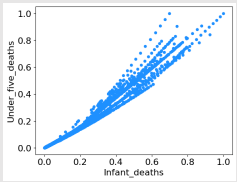
Introduction

- Kaggle → Life Expectancy World Health Organization (WHO) data
  - Already cleaned
- Three main regions
  - European Union
  - Africa
  - Asia
- 13 features of normalized data + 1 outcome feature

Highly Correlated Features

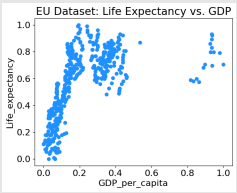


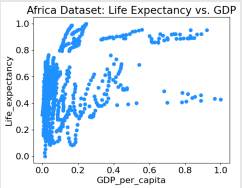
R^2: 0.909
Correlation: 0.95

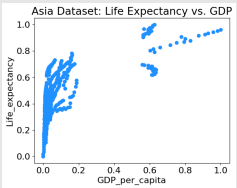


R^2: 0.972
Correlation: 0.99

Life Expectancy vs. GDP per Capita of each Region

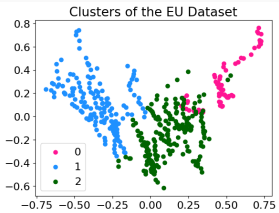


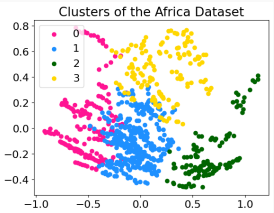


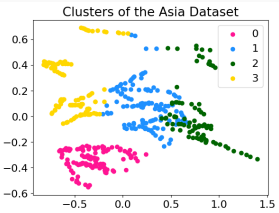


Results

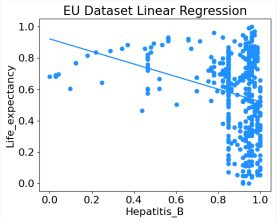
K-Means Clustering



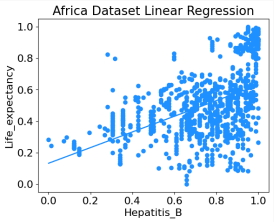




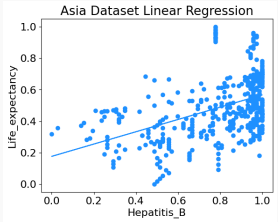
Simple Linear Regression



p-val: 9.02e-09
R^2: 0.074

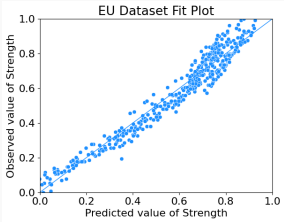


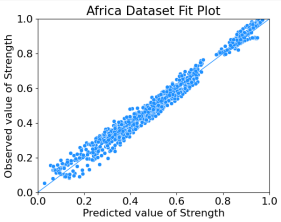
p-val: 7.16e-23
R^2: 0.202

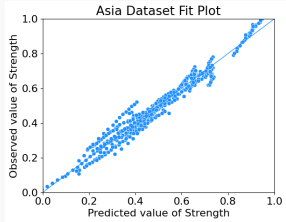


p-val: 8.16e-48
R^2: 0.228

Multiple Linear Regression







Conclusions

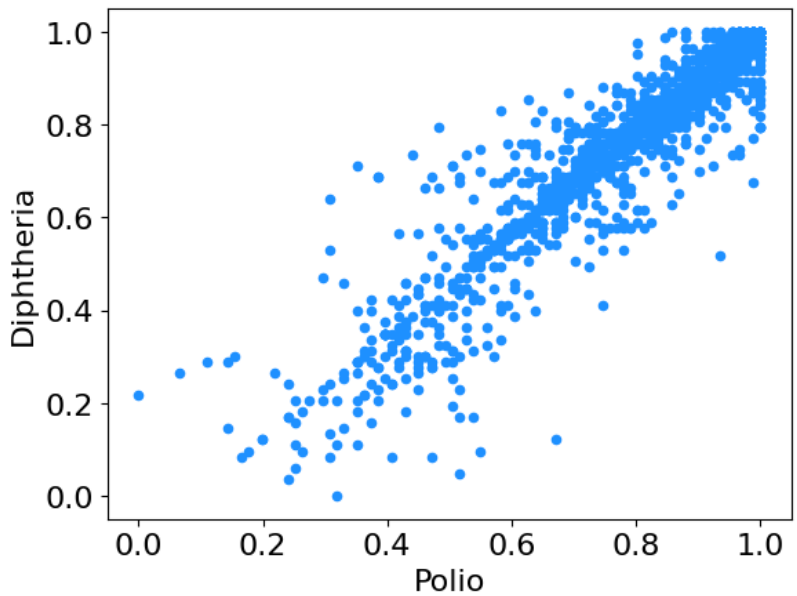
- K-means clustering shows that Africa and Asia may have a more involved model for life expectancy than the European Union
- Then numbers may indicate that linear regression might be statistically significant, but R^2 indicates otherwise
- Be aware of confounding variables, especially in the EU dataset

- Multiple Linear Regression provides the best model in this preliminary exploratory analysis
- Important features
  - All Datasets: Infant deaths, adult mortality, alcohol consumption, HIV incidents, GDP
  - EU: Measles, population
  - Africa: Hep-B, measles, BMI, Polio, schooling
  - Asia: BMI, population, schooling

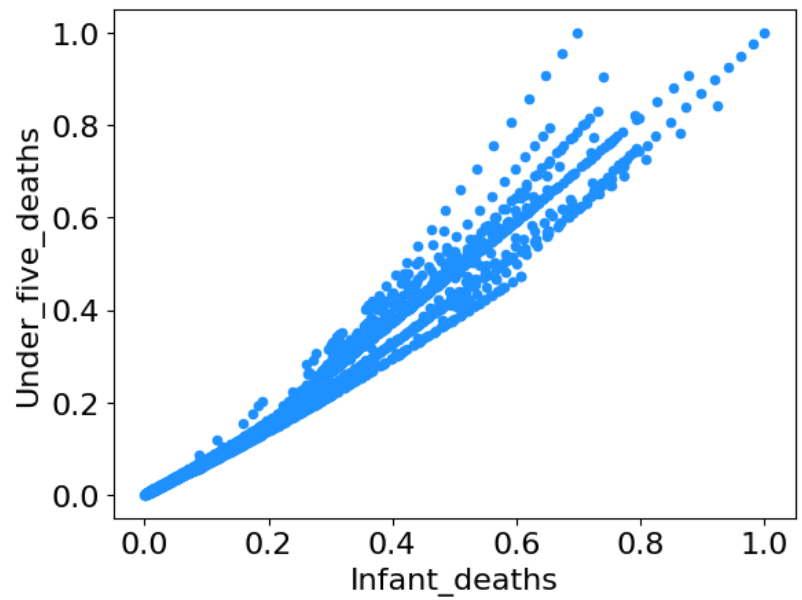
Introduction

- Kaggle → Life Expectancy World Health Organization (WHO) data
- Already cleaned
- Three main regions
  - European Union
  - Africa
  - Asia
- 13 features of normalized data
- + 1 outcome feature

Highly Correlated Features

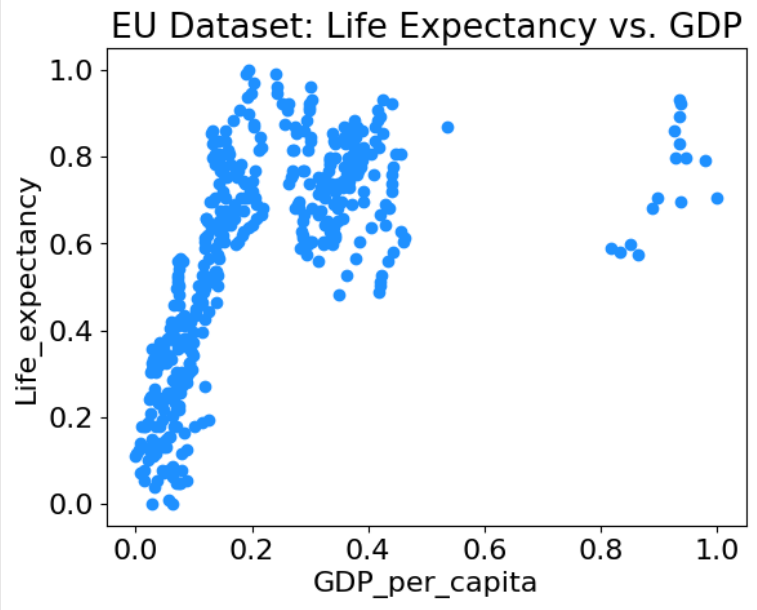


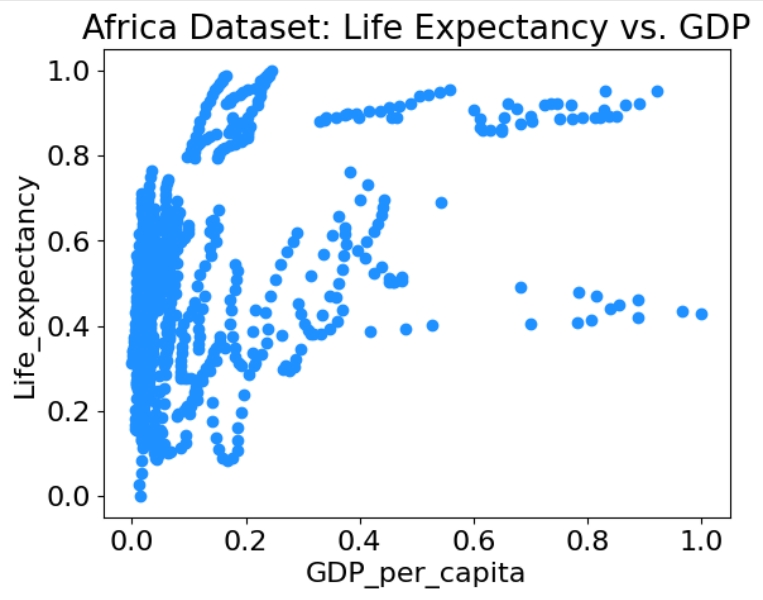
R^2: 0.909
Correlation: 0.95

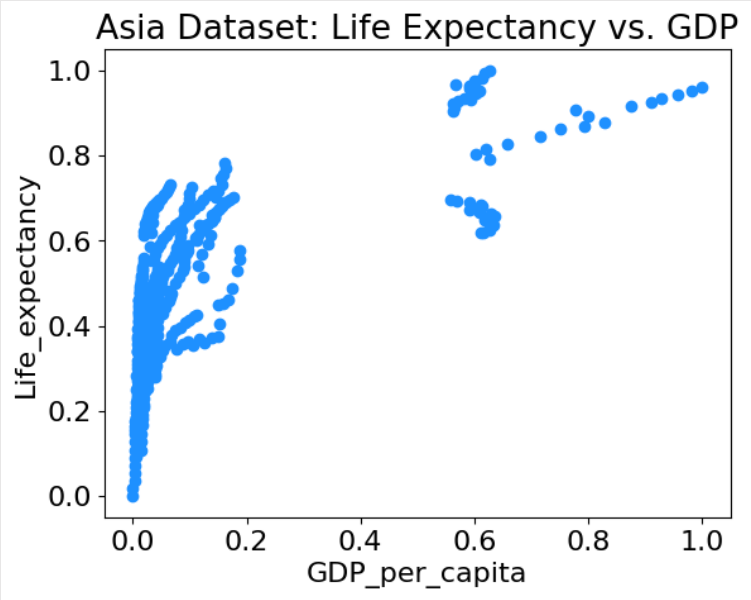


R^2: 0.972
Correlation: 0.99

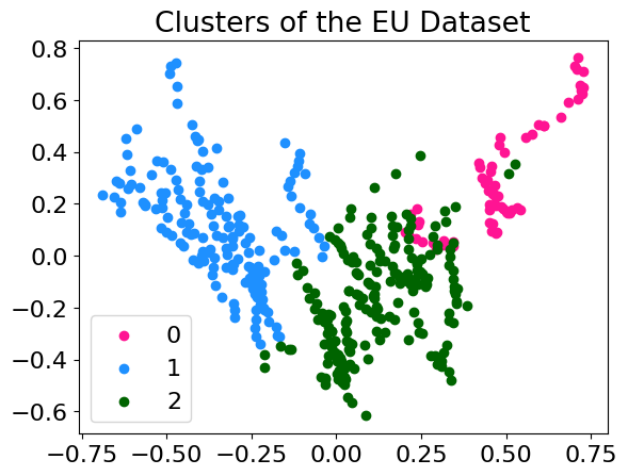
Life Expectancy vs. GDP per Capita of each Region

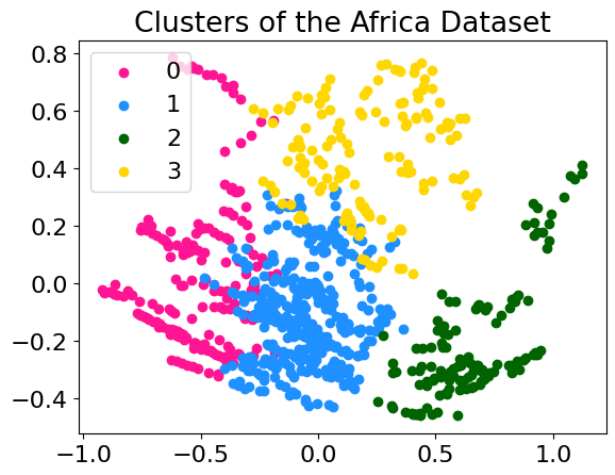


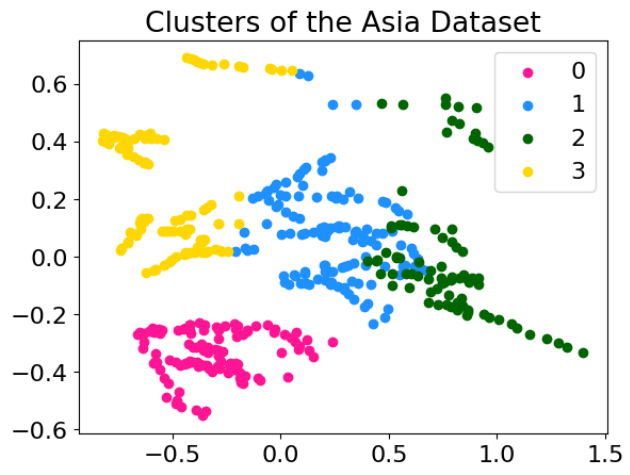




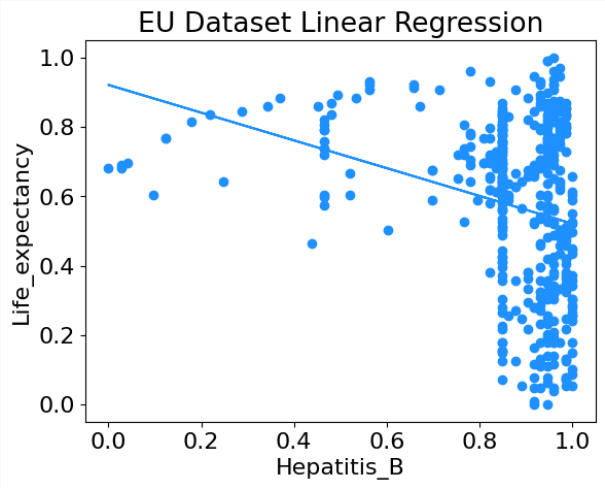
K-Means Clustering



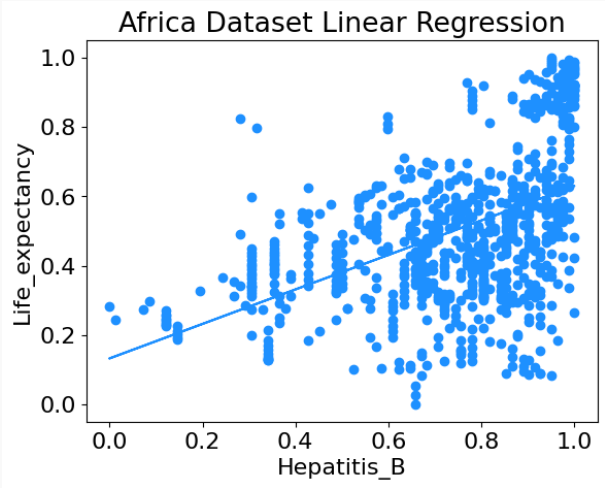




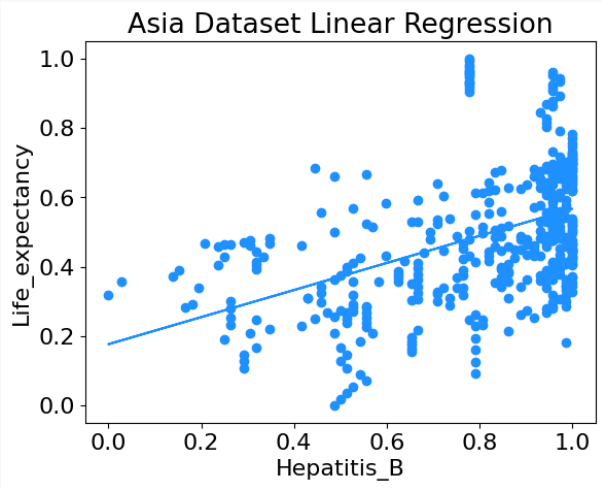
Simple Linear Regression



p-val: 9.02e-09
R^2: 0.074

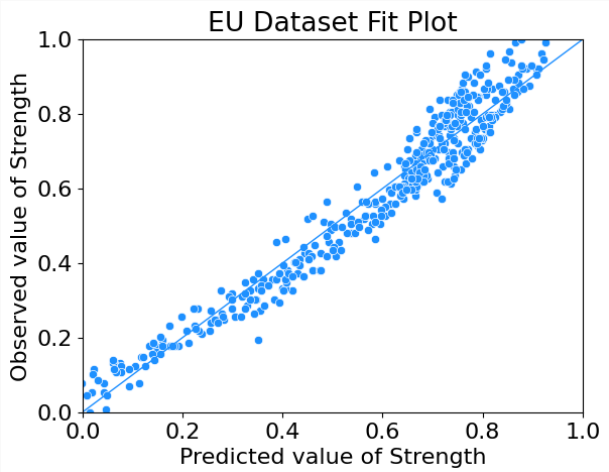


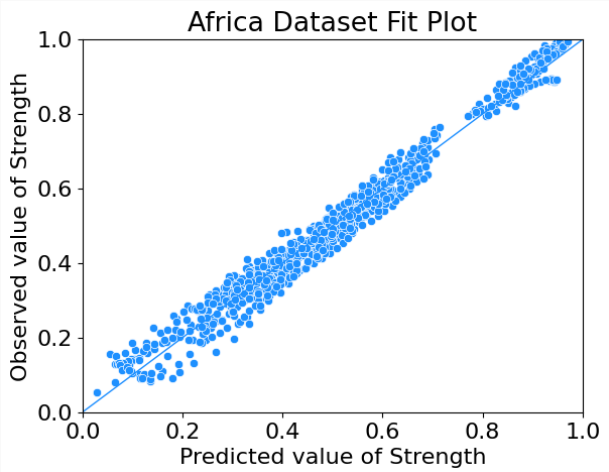
p-val: 7.16e-23
R^2: 0.202

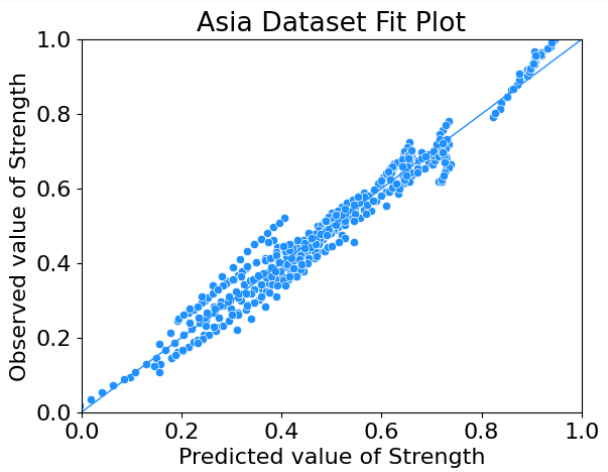


p-val: 8.16e-48
R^2: 0.228

Multiple Linear Regression







Conclusions

- K-means clustering shows that Africa and Asia may have a more involved model for life expectancy than the European Union
- Then numbers may indicate that linear regression might be statistically significant, but R^2 indicates otherwise
- Be aware of confounding variables, especially in the EU dataset

- Multiple Linear Regression provides the best model in this preliminary exploratory analysis
- Important features
  - All Datasets: Infant deaths, adult mortality, alcohol consumption, HIV incidents, GDP
  - EU: Measles, population
  - Africa: Hep-B, measles, BMI, Polio, schooling
  - Asia: BMI, population, schooling