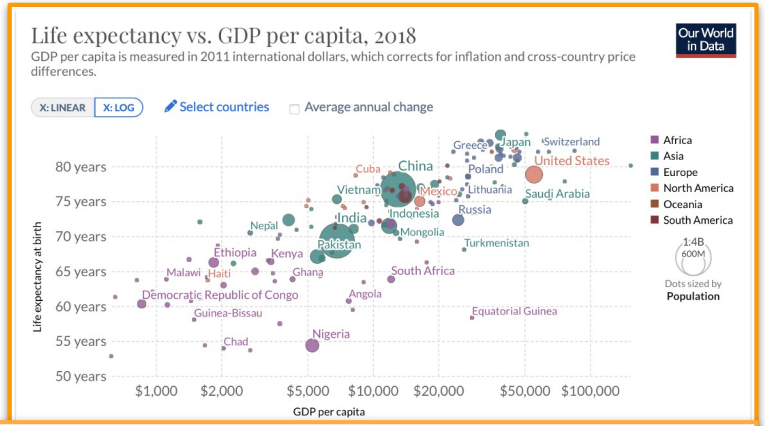
Global Life Expectancy and Related Factors

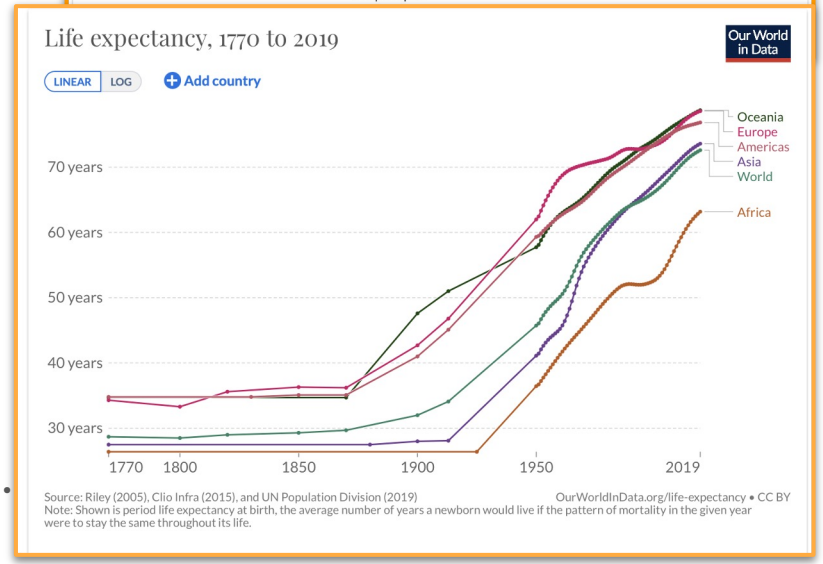
December 1st, 2022

Pooja Kabber, Echo Chen, Dingkun Yang, Andrew Kroening

Introduction | Motivation

- Global life expectancy is increasing as a trend. It is a possible indicator for the overall quality of life in a society.
- Some countries/regions still lag in this metric, and there is some evidence that economics is involved.
- Need more understanding to drive “high-payoff” policy recommendations.
- We seek to determine if we can find areas for countries to invest in to improve quality of life.





Dataset | Overview

- Data span years 2000 - 2015, with one observation per country for 21 variables.
- We chose to omit some missing values where we could not find what we felt was “high-fidelity” replacement data.
- Reduce to two subsets: one each for the years 2013 (out-of-sample) and 2014 (primary).
- Most variables are related to either economics, health or societal well-being.

World Health Organization

kaggle

THE WORLD BANK

Research data sources: Life Expectancy (WHO) | Kaggle & World Bank

Dataset | Variables

- Data span years 2000 - 2015, with one observation per country for 21 variables.
- We chose to omit some missing values where we could not find what we felt was “high-fidelity” replacement data.
- Reduce to two subsets: one each for the years 2013 (out-of-sample) and 2014 (primary).
- Most variables are related to either economics or societal well-being.

Descriptive

Country Name
Year

Health

Life Expectancy
Alcohol Use
Measles Incidence
Deaths under 5 years of age
Diphtheria Immunizations
Thinness among 1-19 yr-olds
Thinness among 5-9 yr-olds
Adult Mortality Rate
Hepatitis B Immunizations
Body Mass Index (Average)
Polio Immunizations
HIV/AIDS Fatalities
Infant Deaths

Social

Average Years of Schooling
Development Status
Population (Est.)

Economic

Income Composition of Resources
Gross Domestic Product (GDP)
Government Expenditure on Health (% GDP)
Government Expenditure on Health (% Total)

Dataset | Missingness

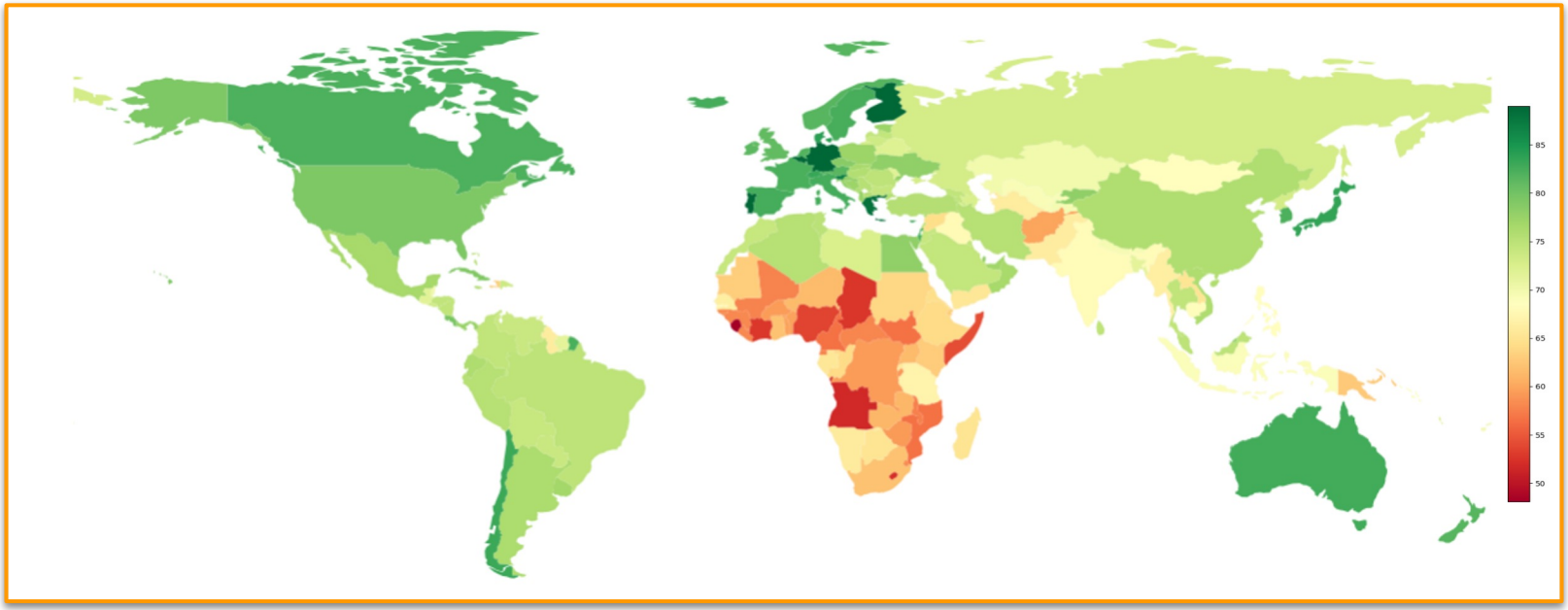
- Original dataset had a significant number of missing population values.
- Also had concerns about the scale and measurement of some population counts in the original Kaggle dataset.

Original Dataset Population Values on Red-Blue Scale

- To compensate, we added population data from the World Bank, which we believe is high-quality

Question #1 | Introduction

"How did major disease, economic, and social factors impact life expectancy around the globe in 2014?"

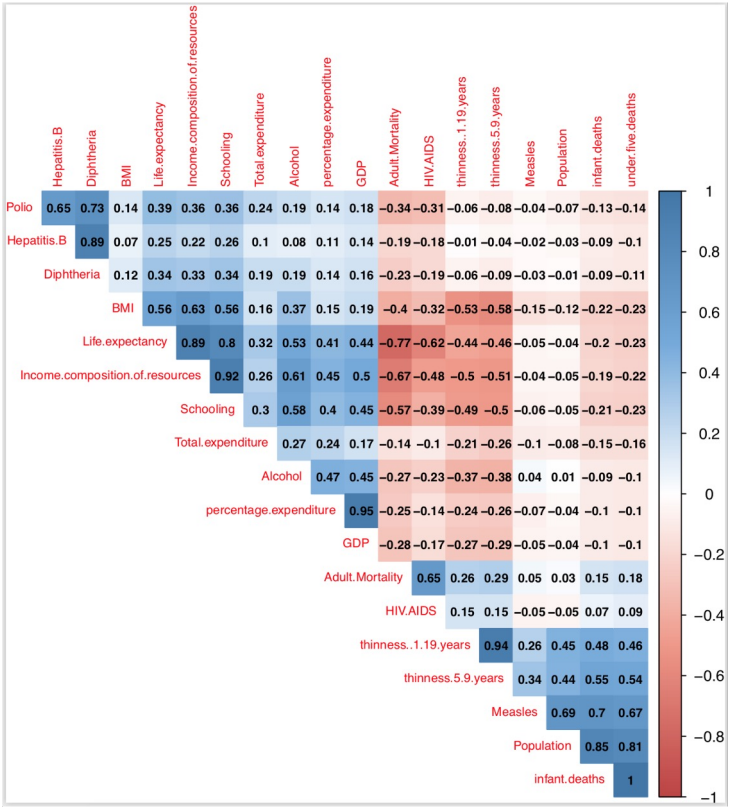


Global Life Expectancies by Country

Question #1 | Feature Selection (A Priori)

Social	Health	Economic
Status	Measles (no. of reported cases per 1000 people)	Total Expenditure (% of govt expenditure on health)
Population	Polio (polio immunization coverage % among 1 year olds)	
Schooling (no. of years of schooling)	HIV (deaths per 1000 for 0-4 year olds)	
	BMI (average BMI of whole population)	
	Adult Mortality (probability of dying between 15 and 60 per 1000 people)	

Question #1 | Correlations



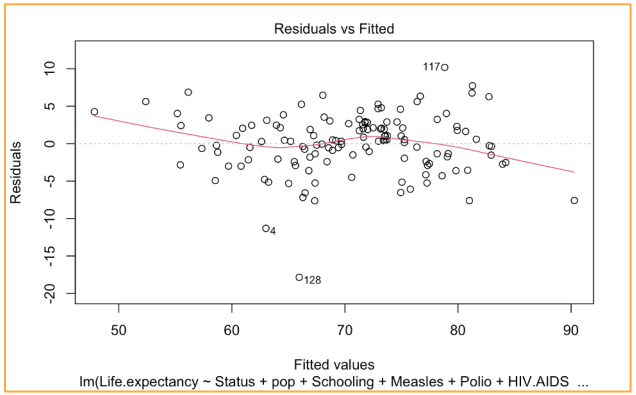
Infant deaths (per 1000 people)	Under five deaths (per 1000 people)
Percentage expenditure (expenditure on health as a % of GDP)	GDP
Hepatitis B (% immunization coverage among 1 yr olds)	Diphtheria (DTP-3% immunization coverage among 1 yr olds)
Thinness (% from ages 10 to 19)	Thinness (% from ages 5 to 9)
Schooling (years of schooling)	Income composition of resources (HDI)

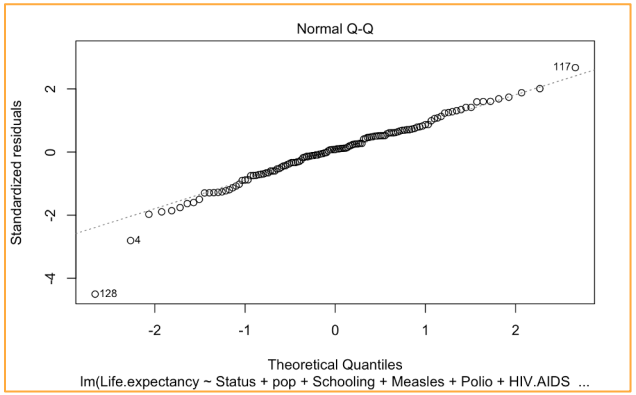
Question #1 | Methods

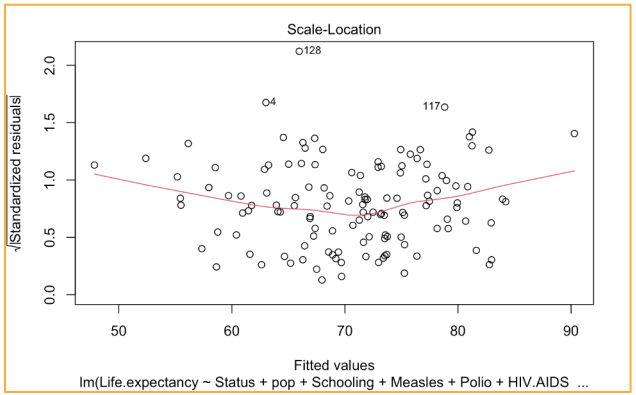
- **Model:** Linear regression with continuous dependent variable Life Expectancy
- **Feature selection:** Combination of a priori variables (variables of interest) and backward stepwise selection
    - Consideration: Used Status instead of HDI (Income composition of resources)
- **Model evaluation and selection:** Adjusted R^2 and BIC

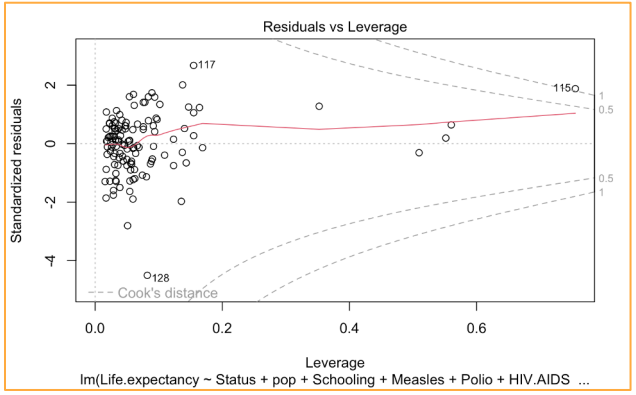
Model	Number of Variables	BIC	Adj.R^2	Assumptions	Number of Significant Variables
Priori Model	9	775.50	0.7683	All met	3
Full Model	20	746.36	0.8615	All met	4
Backward	6	691.28	0.8683	All met	4
Final Model	10	747.08	0.8194	All met	5

Question #1 | Model Validity







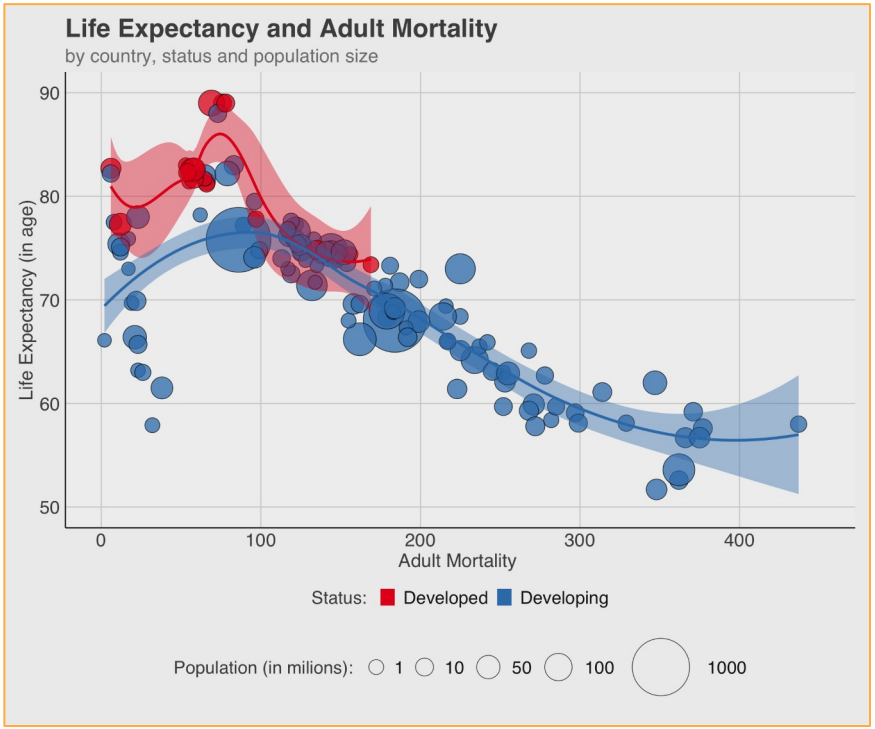


Question #1 | Model Result

Life Expectancy depends on

- **Adult Mortality**
  - probability of dying between 15 and 60 per 1000 people
- **Total Expenditure**
  - % of govt expenditure on health
- **HIV.AIDS**
  - deaths per 1000 for 0-4 year olds
- **BMI**
  - average BMI of whole population
- **Schooling**
  - number of years of schooling

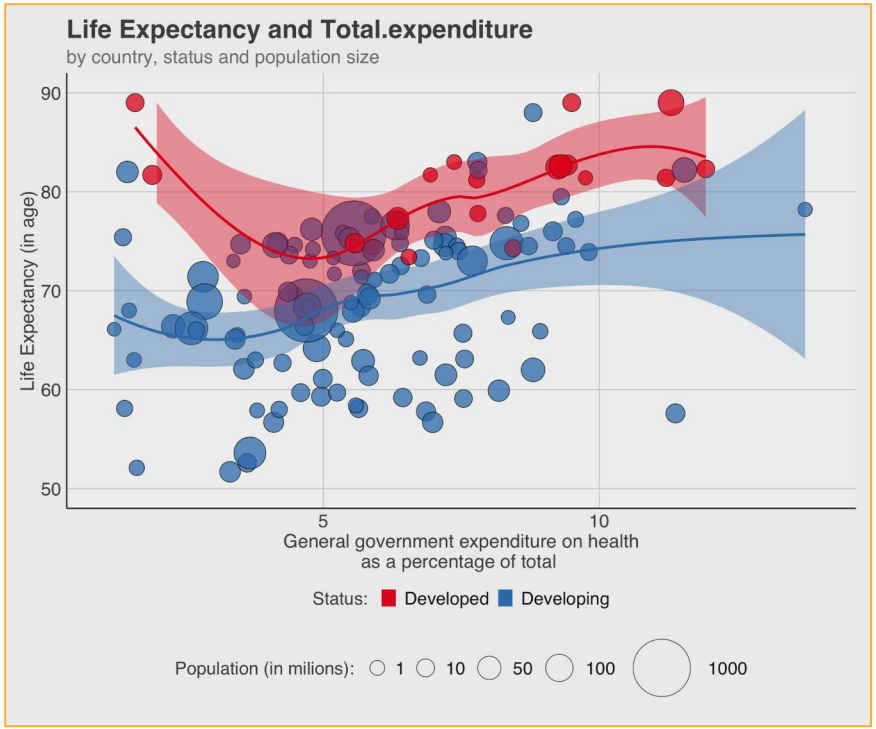
Question #1 | Results



Adults death probability increase 0.1%

Life expectancy 0.03 year (11 days)

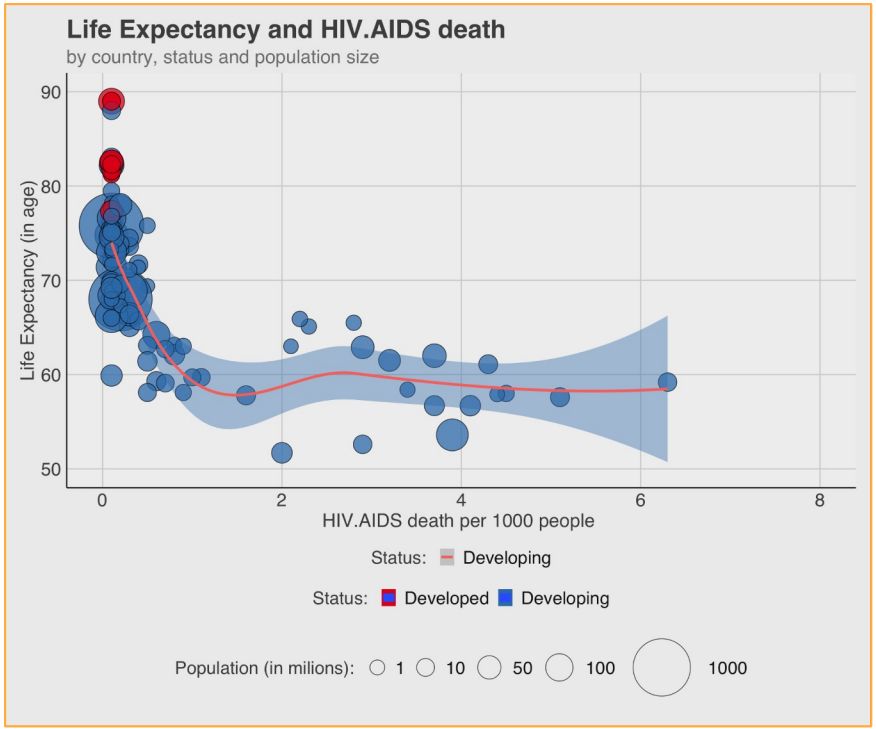
Question #1 | Results



General government expenditure on health Rise 1%

Life expectancy 0.03 year (11 days)

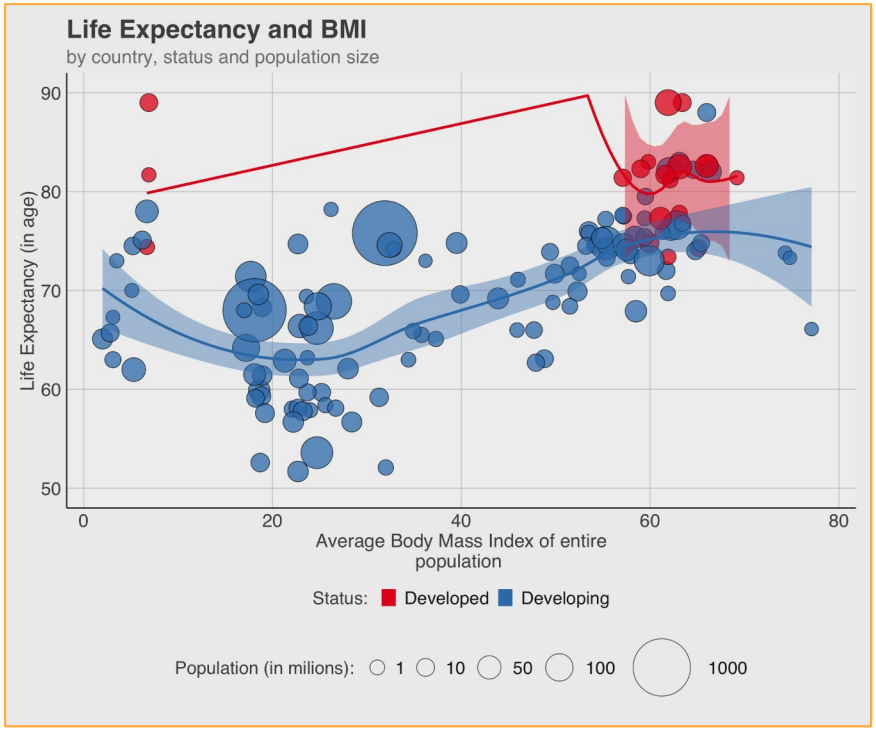
Question #1 | Results



(0-4 yrs) HIV death rates Rise 0.1%

Life expectancy 1.04 year (380 days)

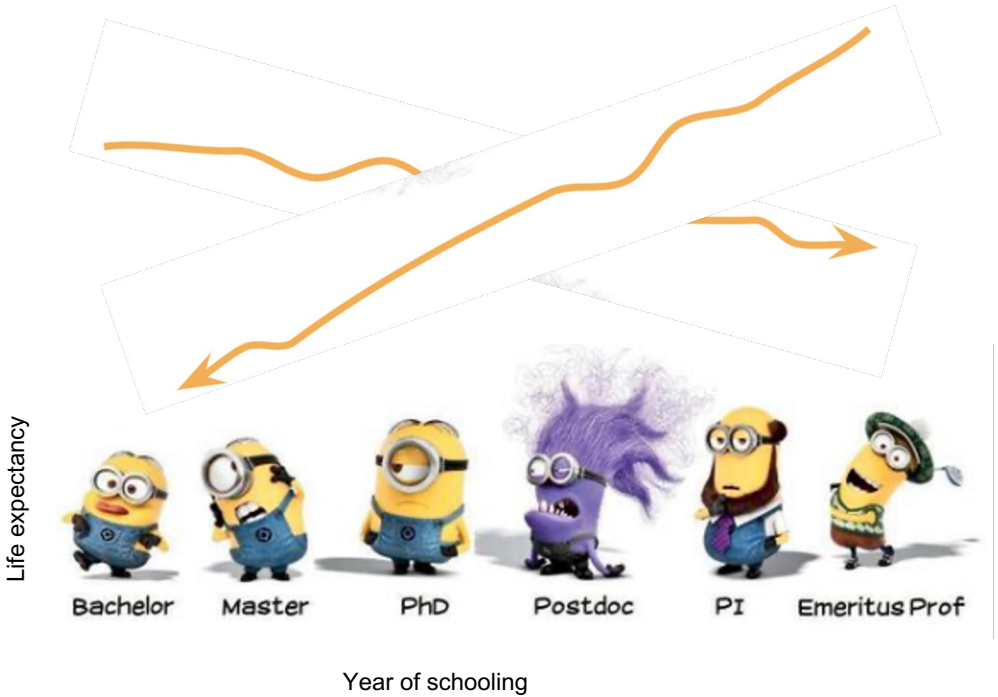
Question #1 | Results



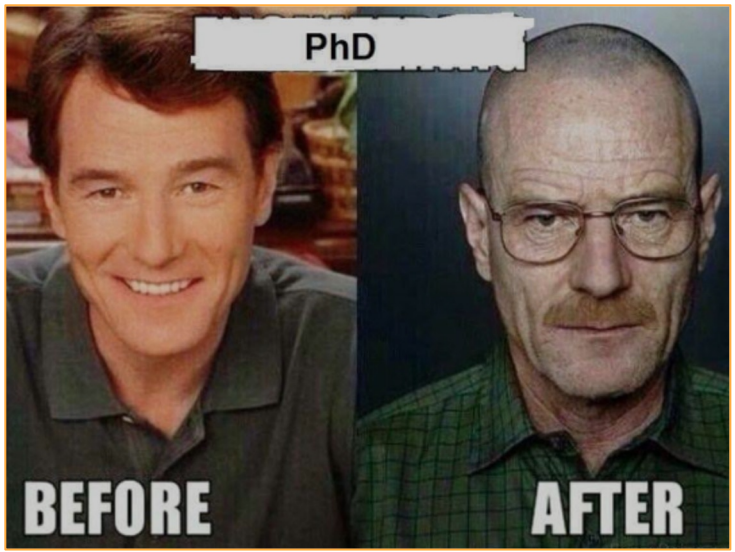
BMI
Increase one unit

Life expectancy
0.04 year (17 days)

Question #1 | Results



Years of schooling rise by 1 year → Life expectancy 0.18 year (64 days)



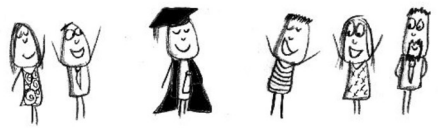
Source : Humandevelopmentreports

Question #1 | Results

Year of Schooling
Rise one year

Life expectancy
1.18 year (431 days)

HAPPINESS IS



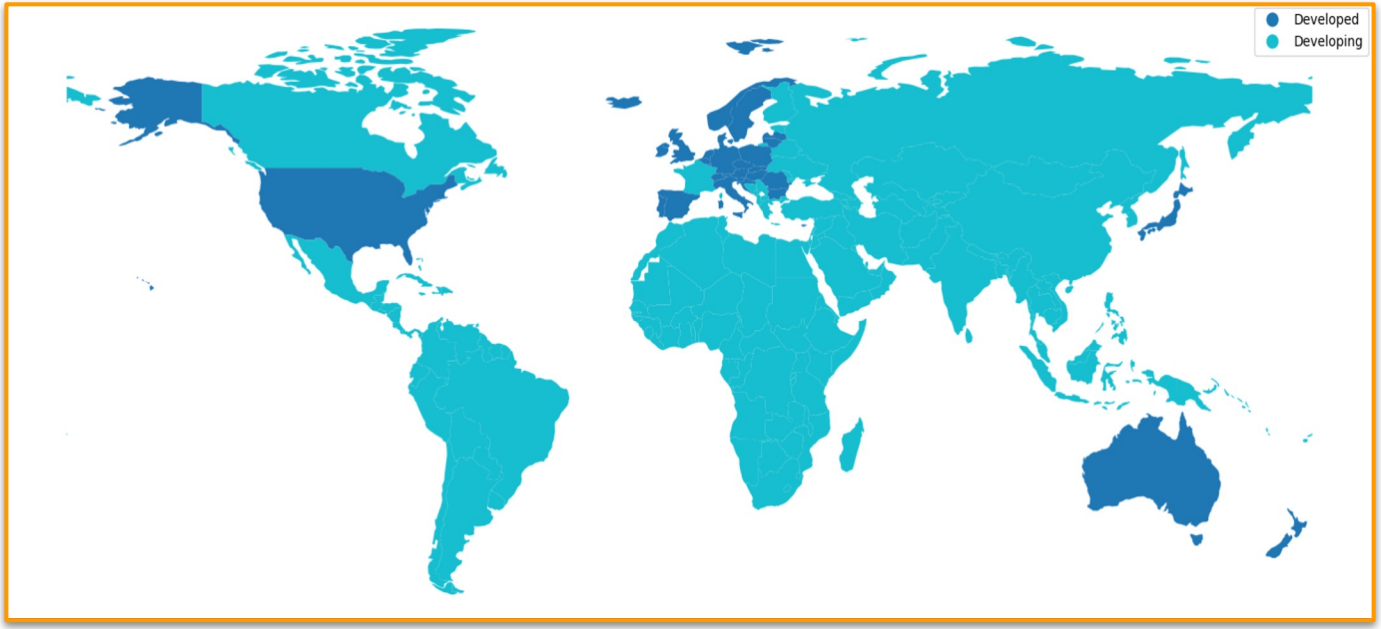
...a PhD.

Question #2 | Introduction

Developed vs. Developing - What are the drivers?

Question #2 | Introduction

"How did disease and mortality rates, along with national economic factors, contribute to a country's development status in 2014?"



Developed (Dark) vs. Developing (Light) Countries

Question #2 | Methods

- Logistic Regression
- Used *a priori* variable selection to target specific areas.

Health

Life Expectancy
Body Mass Index (Average)
HIV/AIDS Fatalities

Social

Average Years of Schooling
Population (Est.)
Development Status

Economic

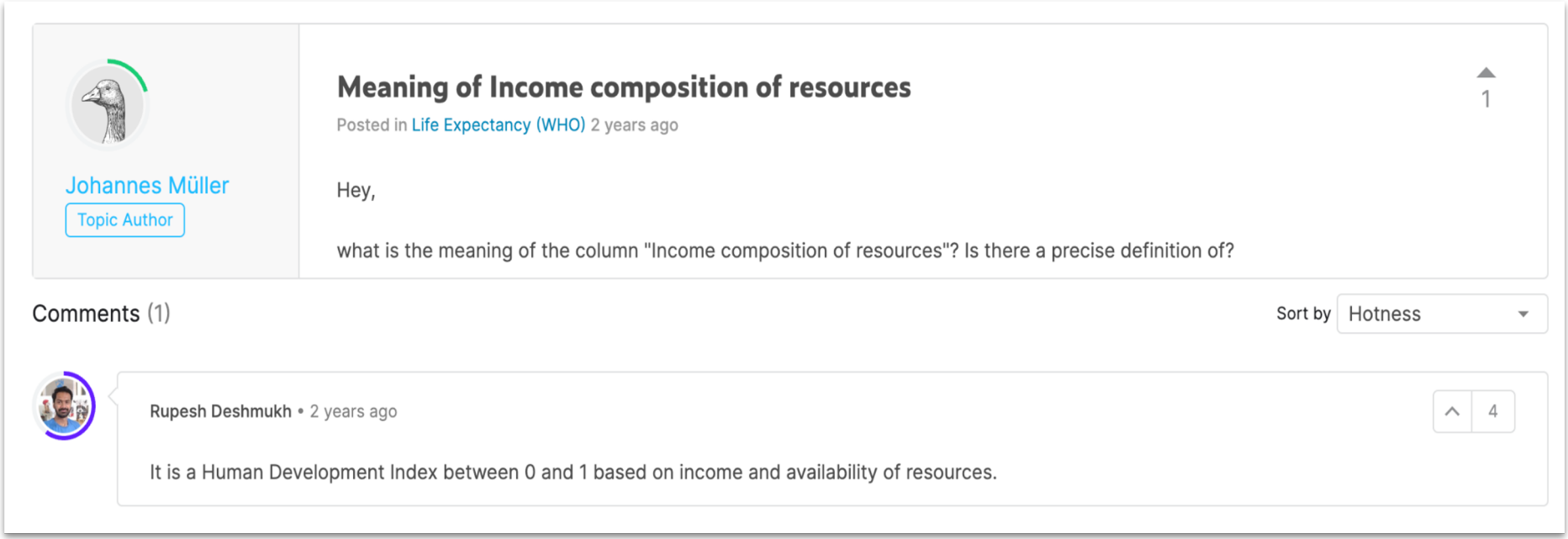
Income Composition of Resources
Gross Domestic Product (GDP)
Government Expenditure on Health (% GDP)

Question #2 | Results

- Of all our predictors, only Income Composition of Resources yielded a significant p-value.
- Overall model results were positive, we found high degrees of accuracy (> 95%).
- We have difficulty establishing exactly what Income Composition of Resources consists of.

Variable	P-Value	Odds Change
Std. Income Composition of Resources ▲	**< 0.01**	**1.55**
Log GDP per Capita ▼	0.14	0.68
Gov. Expenditure on Healthcare ▲	0.17	1.27
BMI ▼	0.23	0.97
Log Population ▼	0.26	0.71
Life Expectancy ▼	0.59	0.93
Years of Schooling ▲	0.59	1.31
HIV/Aids Deaths per 1000	> 0.99	-

Question #2 | Results



Meaning of Income composition of resources

Posted in Life Expectancy (WHO) 2 years ago

1

Johannes Müller

Topic Author

Hey,

what is the meaning of the column "Income composition of resources"? Is there a precise definition of?

Comments (1)

Sort by Hotness

Rupesh Deshmukh • 2 years ago

4

It is a Human Development Index between 0 and 1 based on income and availability of resources.

- We suspect Income Composition of Resources is a re-branded Human Development Index, which would make the results insipid.

https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who/discussion/192631

Question #2 | Final Results

- After triaging the Income Composition variable, we determined it was troublesome
- Dropped this variable and re-ran the model with the remaining predictors
- It turns out education is a great indicator when identify whether the country is developing or developed

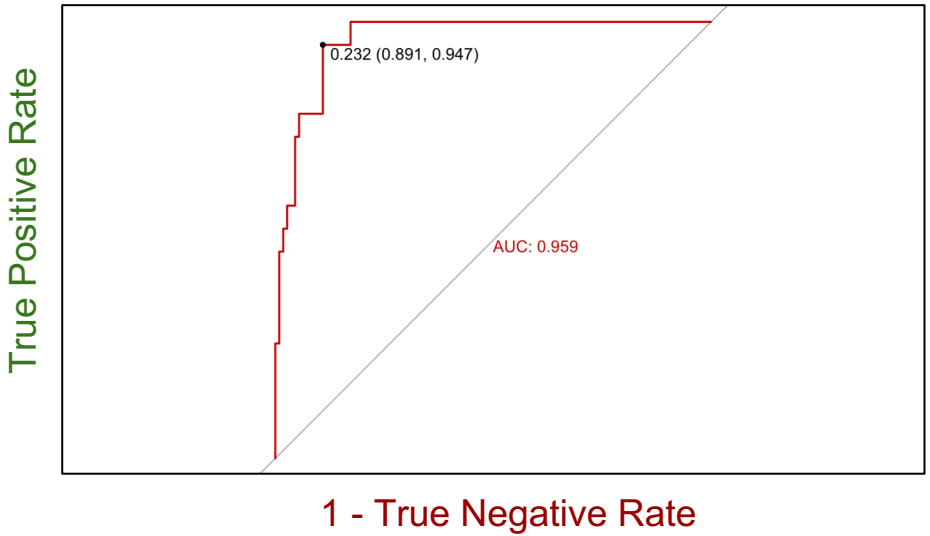
Variable	P-Value	Odds Change
Years of Schooling ▲	**0.02**	**2.39**
Life Expectancy ▲	0.06	1.21
Log Population ▼	0.13	0.68
BMI ▼	0.25	0.97
Gov. Expenditure on Healthcare ▲	0.27	1.17
Log GDP per Capita ▼	0.33	0.77
HIV/Aids Deaths per 1000	> 0.99	-

Question #2 | Assessment

Confusion Matrix

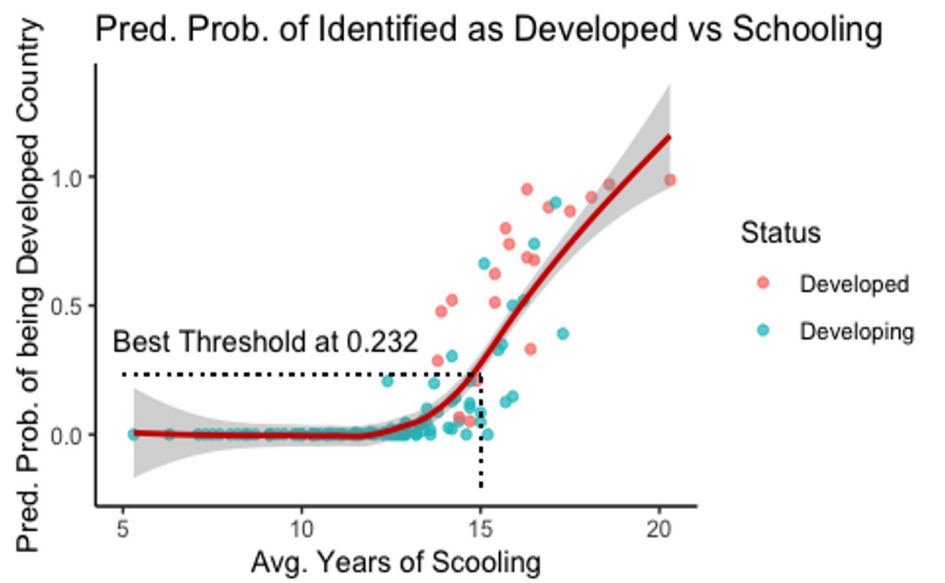
	True Developed	True Developing
Pred. Developed	11	5
Pred. Developing	8	105

95% C.I. of Accuracy: (0.83, 0.95)



Prediction Accuracy Curve for Final Fitted Model with cut-off at 0.232 ("best" option)

Question #2 | Assessment



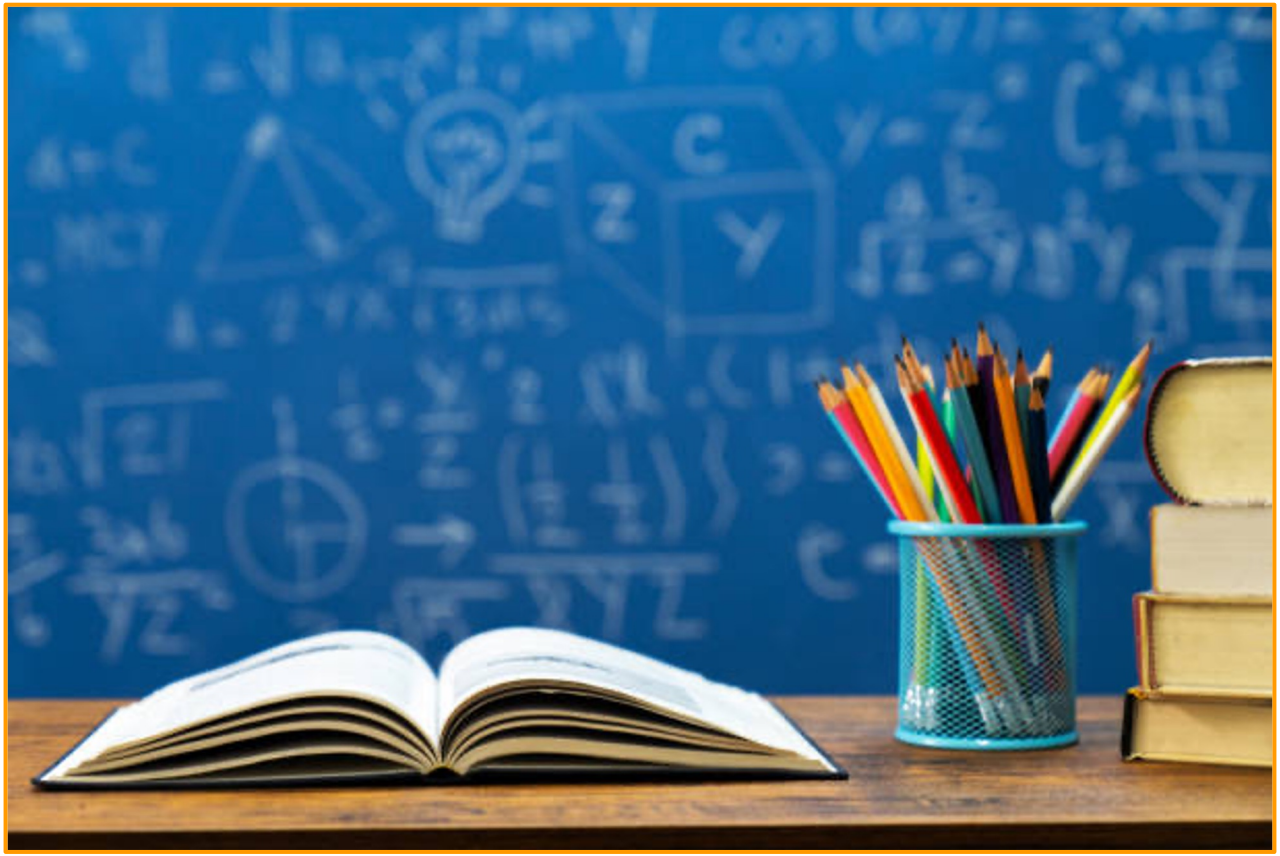
Out-of-Sample Test on Dataset of Year 2013

Confusion Matrix

	True Developed	True Developing
Pred. Developed	16	9
Pred. Developing	3	100

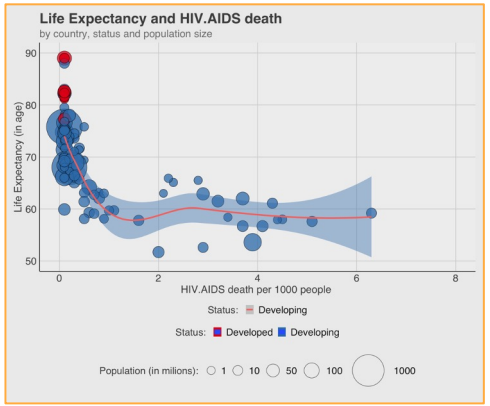
95% C.I. of Accuracy: (0.84, 0.95)

Question #2 | Assessment



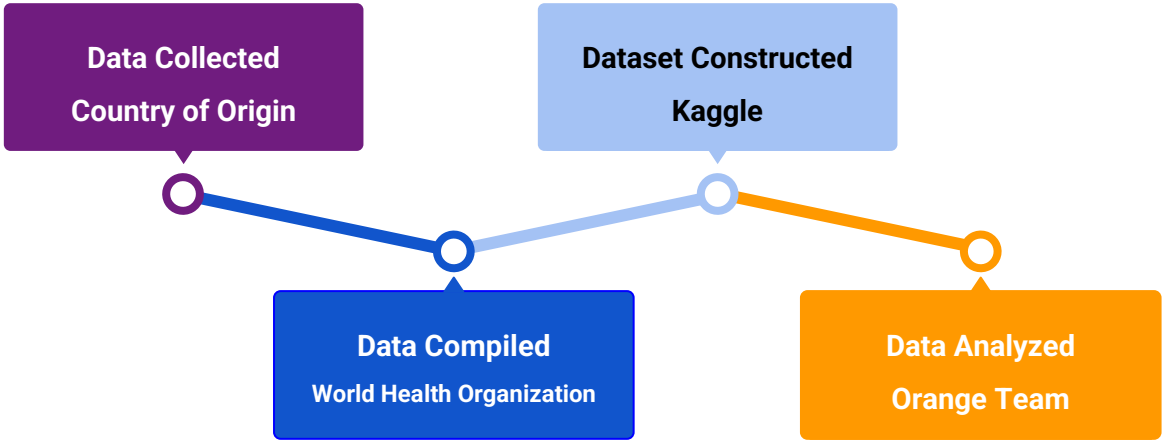
Conclusion | What should we explore further?

- Interaction between Development Status and HIV for Life Expectancy
  - Expected: better medical care and higher life expectancy in developed countries
- Interaction between Development Status and Schooling for Life Expectancy
  - Expected: more schooling may not have an effect on life expectancy for developed countries
- Explore the influence of the region/continent and find a way to include this in a future analysis
- Changes in our categories over time. HDI is going up consistently, but still a lot of developing countries



Conclusion | Limitations

- After triaging the Income Composition variable we determined it was troublesome.
- There appears to be a “chain-of-custody” issue with data. Possibly changes as we go.
- Substantial amounts of missingness, especially population.



Approximation of Data Pathway to Analysis

We established where the data would be sourced from the WHO, but could not verify replacement data in all cases. Appear to be changes en route.

Conclusion | Limitations

- Our dataset had limitations that we did not anticipate prior to the project.
- The biggest, and hardest to resolve, was the political definition of a country, and how that definition can change over time.
- The map background and our dataset had different interpretations, some comical and some potentially confrontational.

Conclusion | Limitations

- Our dataset had limitations that we did not anticipate prior to the project.
- The biggest, and hardest to resolve, was the political definition of a country, and how that definition can change over time.

- The map background and our dataset had different interpretations, some comical and some potentially confrontational.

Conclusion | So What?

- Most important lesson is about data integrity.
- Even from reputable sources, data can still have issues that may not be apparent at the surface
- We make the following recommendations:
  - For policy makers looking to improve average life expectancy, we confidently recommend prioritizing schooling, investing in healthcare, and taking population health measures as more deserving of attention
  - The significance of schooling on development status is noteworthy. We would make policy recommendations that encourage investment in this area.


Questions?