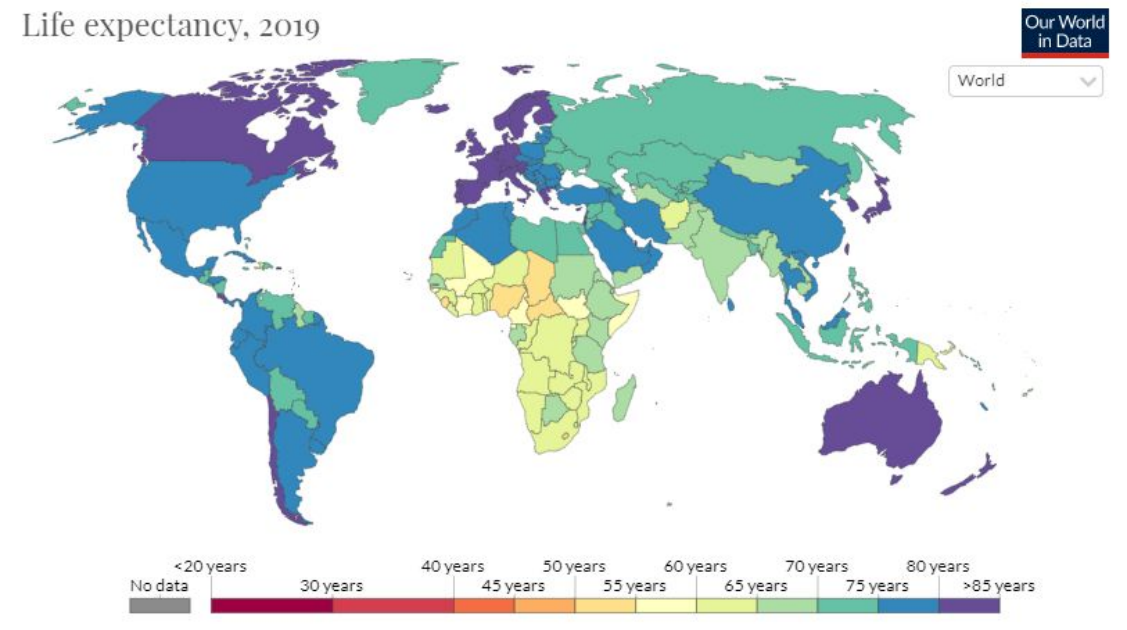
Worldwide Life Expectancy

By: Bryce Hawkins, LeeAnn Markwalter, TJ Augustine

Why Life Expectancy?



- Key metric in assessing overall population health
- Huge variation among countries
  - 47 years: Malwai
  - 83 years: Japan
  - Important to address inequities & disparities
- What factors have the greatest impact on average life expectancy?

Our Data

- “Life Expectancy Data Set” from *kaggle.com*
  - 22 columns, 2,938 rows
  - WHO & United Nations data
- Economic, health, social, political data for 193 countries from 2000 - 2015
- Variables:
  - Country, population, BMI, economic status, percentage expenditure, polio, HIV/AIDS, GDP, schooling, etc.

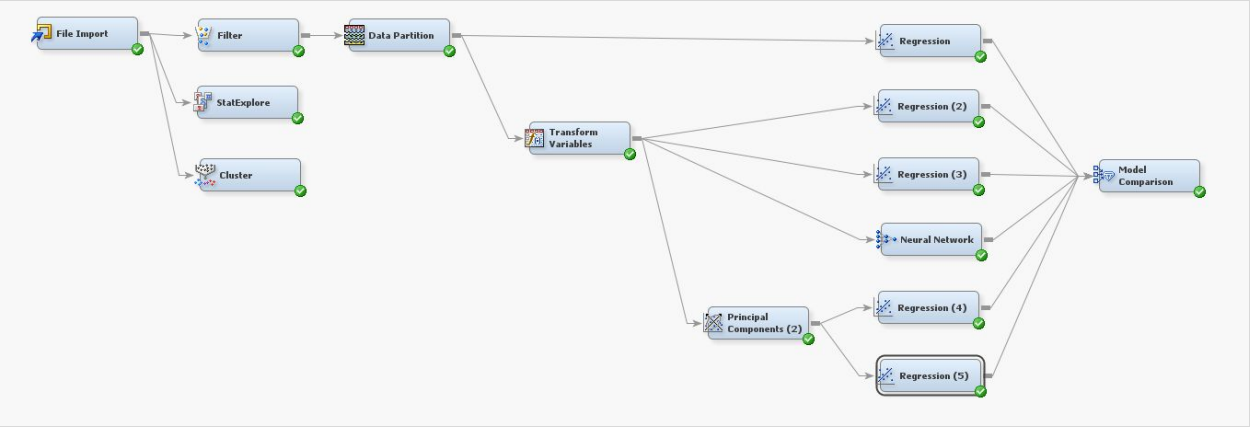
	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis	Measles	BMI	under-five	Polio	Total expe	Diphtheria	HIV/AIDS	GDP	Populatio	thinness	thinness	Income cc	Schooling
2	Afghanist	2015	Developing	65	263	62	0.01	71.27962362	65	1154	19.1	83	6	8.16	65	0.1	584.259	3.4E+07	17.2	17.3	0.479	10.1
3	Afghanist	2014	Developing	59.9	271	64	0.01	73.52358168	62	492	18.6	86	58	8.18	62	0.1	612.697	327582	17.5	17.5	0.476	10
4	Afghanist	2013	Developing	59.9	268	66	0.01	73.21924272	64	430	18.1	89	62	8.13	64	0.1	631.745	3.2E+07	17.7	17.7	0.47	9.9
5	Afghanist	2012	Developing	59.5	272	69	0.01	78.1842153	67	2787	17.6	93	67	8.52	67	0.1	669.959	3696958	17.9	18	0.463	9.8
6	Afghanist	2011	Developing	59.2	275	71	0.01	7.097108703	68	3013	17.2	97	68	7.87	68	0.1	63.5372	2978599	18.2	18.2	0.454	9.5
7	Afghanist	2010	Developing	58.8	279	74	0.01	79.67936736	66	1989	16.7	102	66	9.2	66	0.1	553.329	2883167	18.4	18.4	0.448	9.2
8	Afghanist	2009	Developing	58.6	281	77	0.01	56.76221682	63	2861	16.2	106	63	9.42	63	0.1	445.893	284331	18.6	18.7	0.434	8.9
9	Afghanist	2008	Developing	58.1	287	80	0.03	25.87392536	64	1599	15.7	110	64	8.33	64	0.1	373.361	2729431	18.8	18.9	0.433	8.7
10	Afghanist	2007	Developing	57.5	295	82	0.02	10.91015598	63	1141	15.2	113	63	6.73	63	0.1	369.836	2.7E+07	19	19.1	0.415	8.4
11	Afghanist	2006	Developing	57.3	295	84	0.03	17.17151751	64	1990	14.7	116	58	7.43	58	0.1	272.564	2589345	19.2	19.3	0.405	8.1
12	Afghanist	2005	Developing	57.3	291	85	0.02	1.388647732	66	1296	14.2	118	58	8.7	58	0.1	25.2941	257798	19.3	19.5	0.396	7.9
13	Afghanist	2004	Developing	57	293	87	0.02	15.29606643	67	466	13.8	120	5	8.79	5	0.1	219.141	2.4E+07	19.5	19.7	0.381	6.8
14	Afghanist	2003	Developing	56.7	295	87	0.01	11.08905273	65	798	13.4	122	41	8.82	41	0.1	198.729	2364851	19.7	19.9	0.373	6.5
15	Afghanist	2002	Developing	56.2	3	88	0.01	16.88735091	64	2486	13	122	36	7.76	36	0.1	187.846	2.2E+07	19.9	2.2	0.341	6.2

Data Preparation

- Life Expectancy ➜ target
  - All other variables set as input initially
- Some missing values
  - Population, total expenditure, hepatitis b, alcohol, gdp, percentage expenditure, measles
  - Addressed with impute node in trial #2
  - Transformed variables

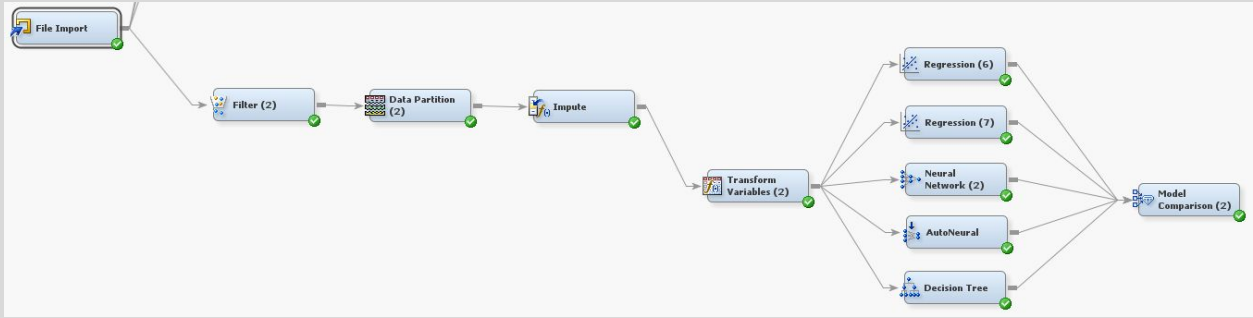
Bahamas	2013	Developing	74.8	172	0	9.42	0	97	0	63.2	0	97	7.5	97	0.1
Bahamas	2012	Developing	74.9	167	0	9.5	0	96	0	62.6	0	99	7.43	98	0.2
Bahamas	2011	Developing	75	162	0	9.34	0	95	0	62	0	97	7.63	98	0.1
Bahamas	2010	Developing	75	161	0	9.19	0	98	0	61.3	0	97	7.44	99	0.2
Bahamas	2009	Developing	74.6	168	0	9.29	0	95	0	6.7	0	97	7.43	96	0.1
Bahamas	2008	Developing	74.5	167	0	10.15	0	9	0	6.1	0	93	7.3	93	0.1
Bahamas	2007	Developing	74.4	167	0	10.75	0	93	0	59.4	0	95	7.8	95	0.1
Bahamas	2006	Developing	74.2	171	0	11.07	0	96	0	58.7	0	94	6.93	95	0.1
Bahamas	2005	Developing	74.1	172	0	10.49	0	93	0	58.1	0	93	5.95	93	0.1
Bahamas	2004	Developing	73.8	174	0	10.1	0	93	0	57.4	0	92	6.2	93	0.1
Bahamas	2003	Developing	73.2	189	0	10.68	0	88	0	56.7	0	93	5.62	92	0.1
Bahamas	2002	Developing	73.1	19	0	10.85	0	89	0	56	0	93	5.26	94	0.1
Bahamas	2001	Developing	72.9	189	0	11.64	0	21	0	55.2	0	98	5.15	99	0.2
Bahamas	2000	Developing	72.6	192	0	12.15	0		0	54.4	0	91	5.21	99	0.1

Diagram Overview



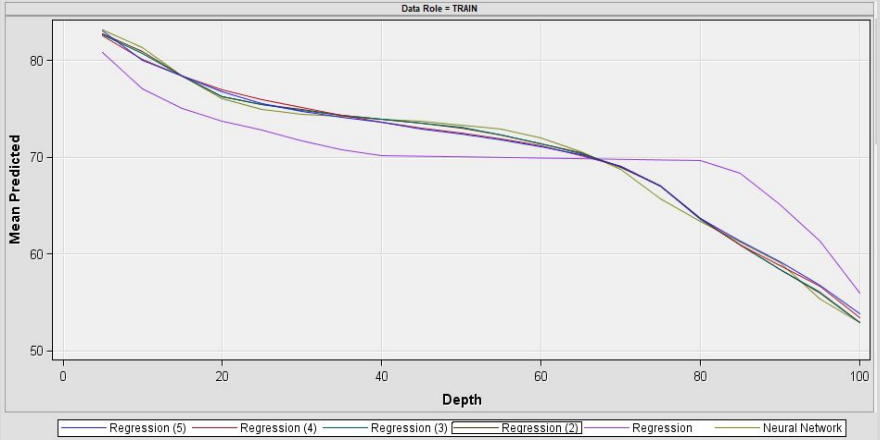
Trial 1

Trial 2



Trial 1 Findings

- Did not impute data
  - Incorporated a Principal Component Analysis
- Most Accurate: Neural Network
  - Based on Average Squared Error
- Significant regression produced adjusted r-squared of .8719
- Variables like Hepatitis B, Measles, and Alcohol were insignificant
  - Alpha level > .05
- The PCA did not produce superior results to the regular
  - Adjusted r-squared was inferior
  - Average squared error was inferior



Trial 2 Findings

- Imputed data but did not use PCA
  - Will the model improve with imputed data?
- Outcome:
  - Model did not show significant improvement
  - Neural network produced best results
  - Adjusted r-squared of regression dropped by .18
- Imputing was not found to show improvement of the models
- Diagnostics:
  - Compute the models with the empty cells not having an impact
  - Use hard data from the countries to update base model

Model Fit Statistics

R-Square	0.8753	Adj R-Sq	0.8708
AIC	2100.3380	BIC	2104.7688
SBC	2259.8280	C(p)	33.0000

Fit Statistics
Model Selection based on Valid: Average Squared Error (_VASE_)

Selected Model	Model Node	Model Description	Valid: Average Squared Error	Train: Average Squared Error	Train: Misclassification Rate
Y	Neural2	Neural Network (2)	10.2886	6.81784	.
	AutoNeural	AutoNeural	10.7511	7.32087	.
	Reg6	Regression (6)	10.9308	8.63061	.
	Reg7	Regression (7)	11.0875	8.95466	.
	Tree	Decision Tree	12.8006	8.84633	.

Significant

- Adult Mortality
- Diphtheria
- Income Composition
- Polio
- Population
- Total Expenditure
- BMI
- Thinness 5-9 Years
- Year
- HIV/AIDS
- Infant Deaths
- Under 5 Deaths
- Status

Insignificant

- Alcohol
- GDP
- Hepatitis B
- Schooling
- Thinness 1-19 Years
- Measles
- Percentage Expenditure

Relationships

- Health metrics
- Deadly diseases
- Child healthcare
- Healthcare spending
- Economic Development Status

Conclusions

- Countries should follow certain procedures to increase life expectancy:
  - Constant monitoring of health metrics and deaths
  - Important resources should be allocated to deadliest diseases
  - Stress importance of healthcare for children and infants
  - Increase expenditures on healthcare on a yearly basis
  - Aid developing countries in providing healthcare resources & necessities to citizens
- Further measures:
  - Continue to manage the diseases that have dissipated
  - Find the right variables to collect and the right way to do it
  - Understand that some variables are still important, just not for lifespan

References

- How different factors have an influence on your life expectancy | by Mubarak Ganiyu | Towards Data Science
- Life Expectancy - Our World in Data
- WHO | Fact file on health inequities
- https://www.kaggle.com/kumarajarshi/life-expectancy-who

Questions?