Analysis of Life Expectancy

Jianyi Zhang
Chen Wang
Linyu Li

[10 / 17 / 2019]

CONTENTS

Overview

Overall of the dataset;
The questions inspired

1

2

Exploratory Data Analysis

Visualization: tables, plots;
Assessment of transformations

Linear Regression

Multi variable linear regression

3

4

Conclusion

Real life significance

Overview

Overview of life expectancy dataset

🎯 Life expectancy is the average number of years to be lived by a group of people born in the same year, if mortality at each age remains constant in the future.*

🎯 72.0 years was the average life expectancy of the global population in 2016 (latest record of WHO), increased by 5.5 years comparing with 2000.*
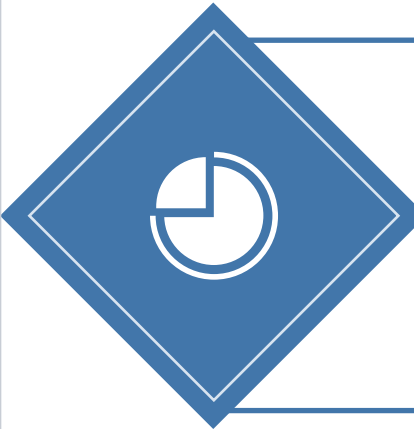
Source

This dataset is downloaded from kaggle. It was collected from WHO and United Nations website.
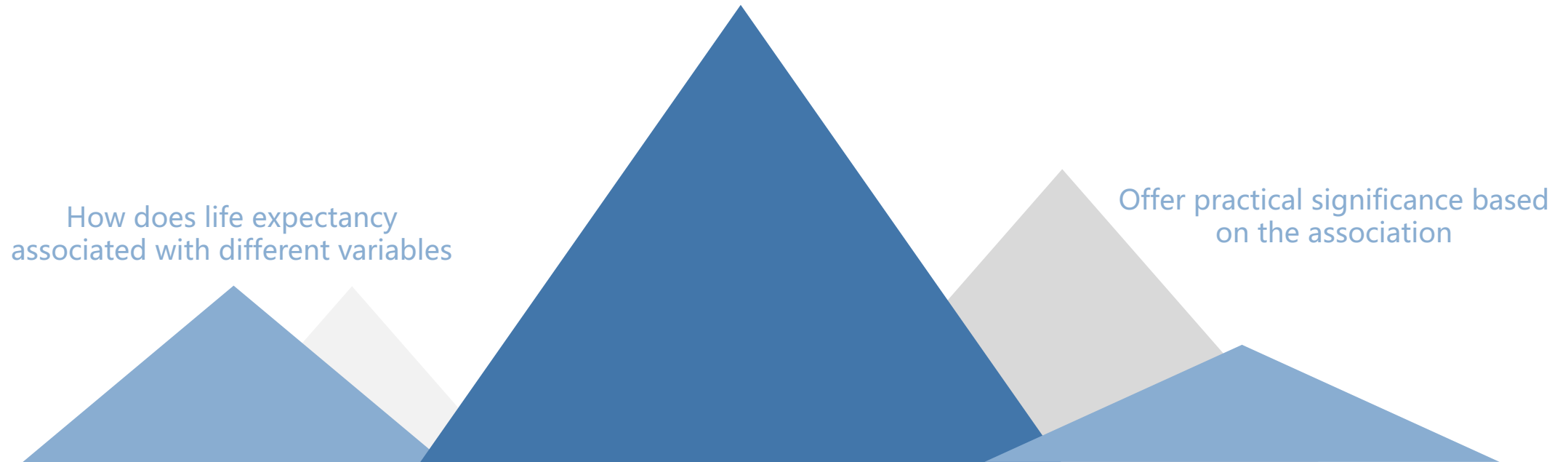
Content

It includes the life expectancy of 183 countries in 2013, as well as factors like healthcare related factors, economic factors, social factors, etc.

- https://www.who.int/gho/mortality_burden_disease/life_tables/situation_trends_text/en/
- https://www.cia.gov/library/publications/the-world-factbook/rankorder/2102rank.html

What variables are associated with life expectancy?

How does life expectancy associated with different variables

Offer practical significance based on the association
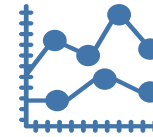
Variable explanation

8 variables

country

183 countries in all

status

Developed or developing

GDP per capita

Gross Domestic Product per capita (constant 2010 USD)

schooling

Number of years of schooling (male aged 25 years and above)

polio

Polio (Pol3) immunization coverage among 1-year-olds (%)

percentage expenditure

Expenditure on health as a percentage of GDP per capita (%)

government expenditure

Government expenditure on health as a % of total government expenditure

life expectancy

In year

Exploratory data analysis

View dataset

	country	status	GDP per capita	schooling	polio	percentage expenditure	government expenditure	life expectancy
1	Belgium	Developed	44302.4843	10.7	99	NA	9.46	87
2	Finland	Developing	45715.6468	10.2	98	NA	9.55	87
3	United Kingdom	Developed	40248.7651	5	95	NA	9.34	87
4	Germany	Developed	44354.7369	13.3	94	NA	11.16	86
5	Greece	Developing	22251.2573	10.4	99	NA	9.26	86
6	Portugal	Developed	21228.0894	8.5	98	NA	9.55	86

Dimension

Dimension	
183	8

Missing value analysis

- There are **185** missing values in all for the dataset.

	country	status	GDP per capita	schooling	polio	percentage expenditure	government expenditure	life expectancy
Number of Missing values	0	0	8	0	0	177	0	0
Percent(%) of missing values (out of 183 counts)	0	0	4.4%	0	0	96.7%	0	0

For GDP per capita: Search GDP per capita online data to find the source or a match, fill the missing value with data from World Development Indicators of World Bank.

For percentage expenditure : Delete the variable/column, as 96.7% of the data are missing, plus failing to find a source or a similar match.
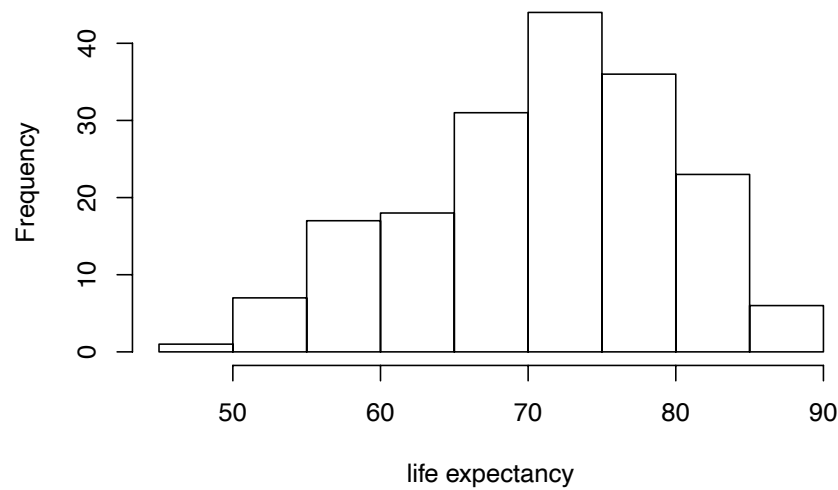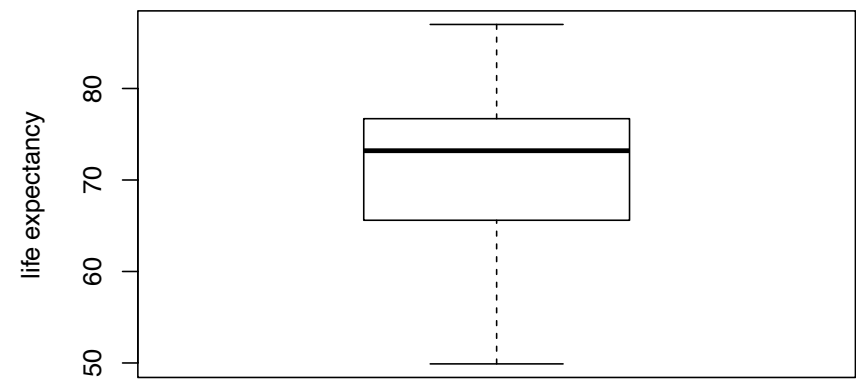
Life expectancy

Summary

Min	1st Quarter	Median	Mean	3rd Quarter	Max
49.9	65.6	73.2	71.24	76.7	87

Plots
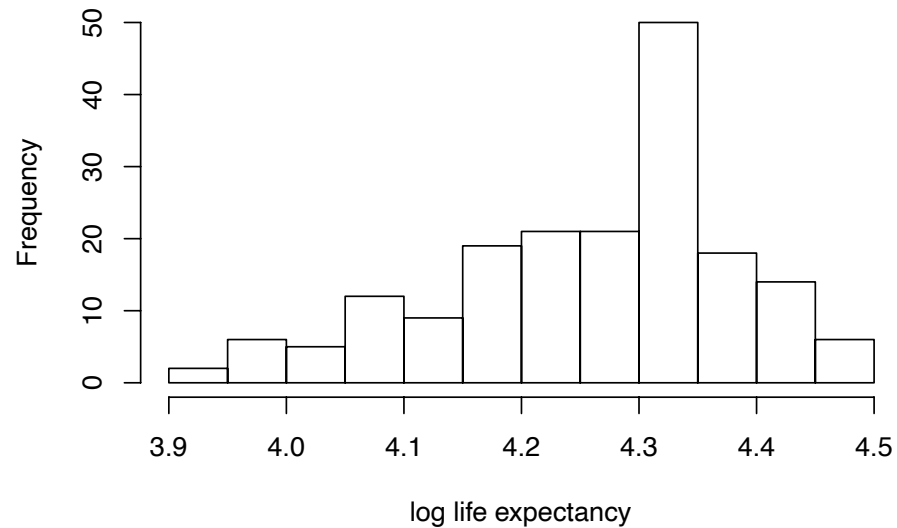

Histogram of life expectancy
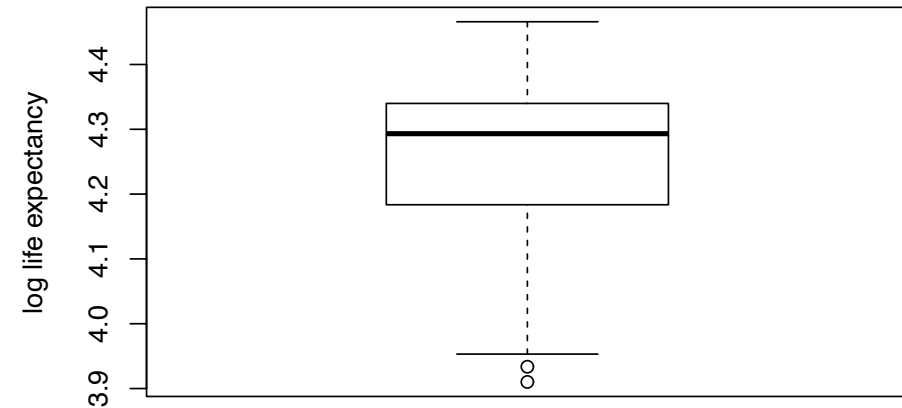

Boxplot of life expectancy

Life expectancy – log transformed

Plots

Histogram of log life expectancy
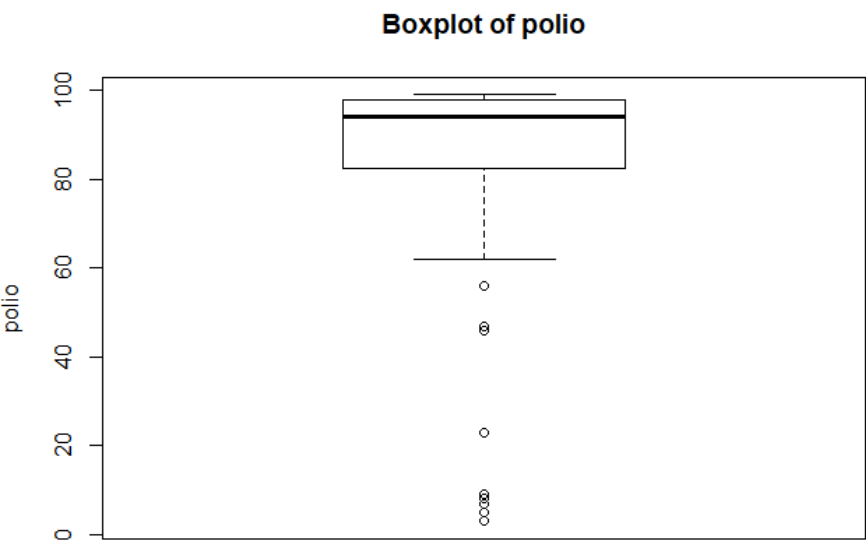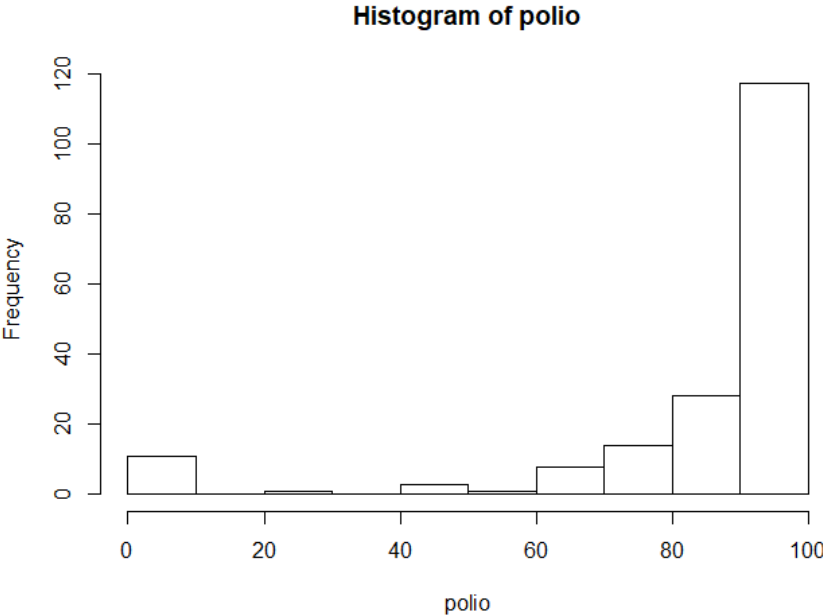
Boxplot of log life expectancy

🎯 We would use the original data of life expectancy because it looks more normal than the log transformed ones

Polio

Summary

Min	1st Quarter	Median	Mean	3rd Quarter	Max
3.00	82.50	94.00	84.73	98.00	99.00

Plots

Histogram of polio

Boxplot of polio

Polio has some extremely small points, but they might be possible.

Government expenditure

Summary

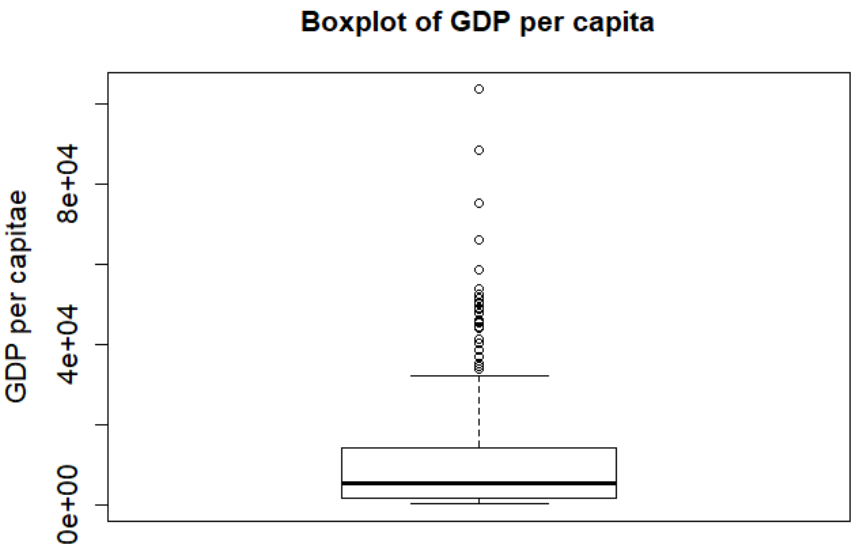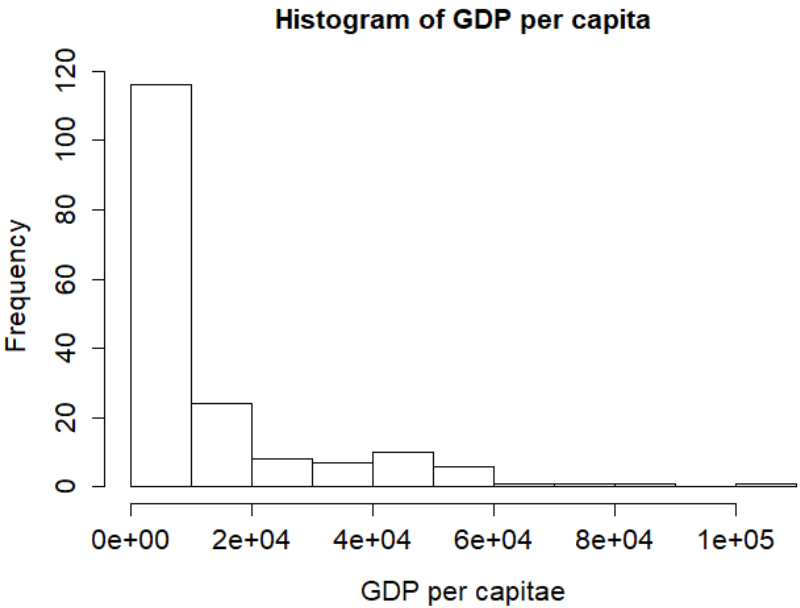| Min	1st Quarter	Median	Mean	3rd Quarter	Max
1.12	4.58	5.96	6.34	7.68	16.9.00

Plots
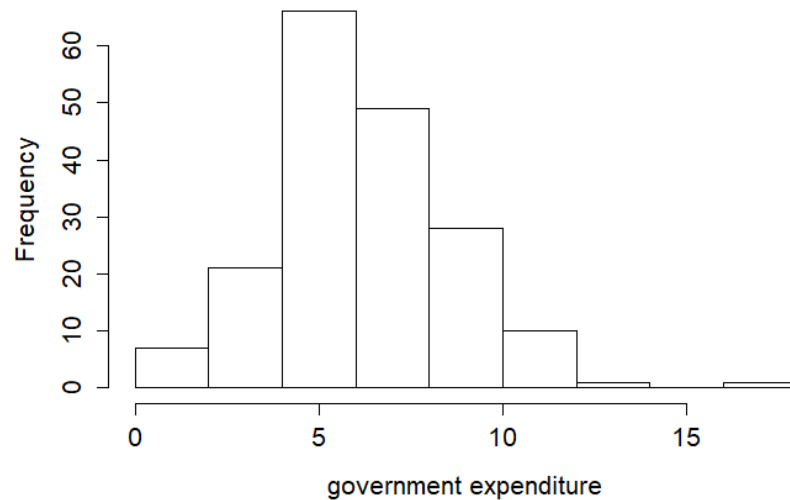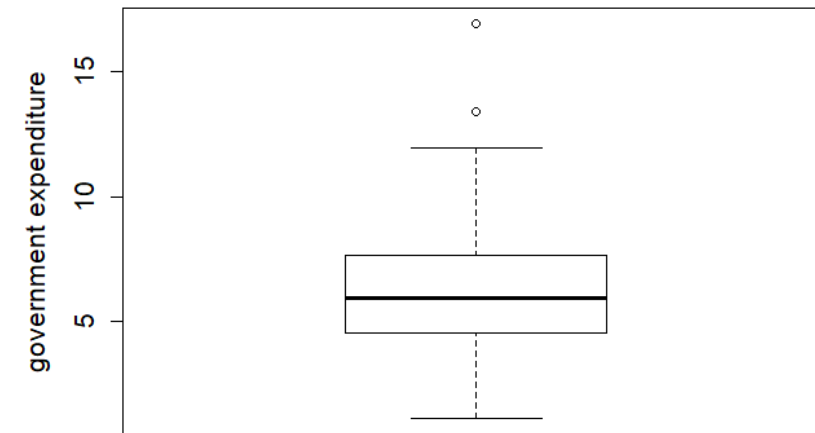
Government expenditure has two extremely large points, but they might be possible.

GDP per capita

Summary

Min	1st Quarter	Median	Mean	3rd Quarter	Max
242.8	1373.6	4969.0	12670.4	14381.5	103721.8

Plots

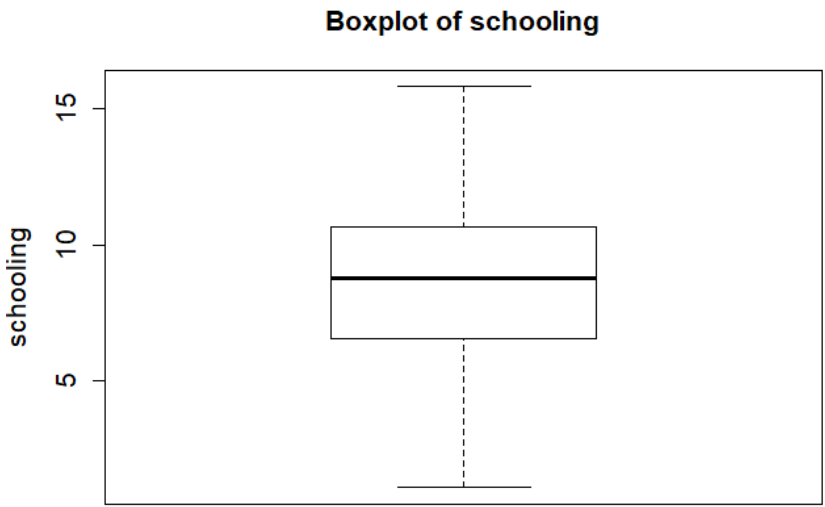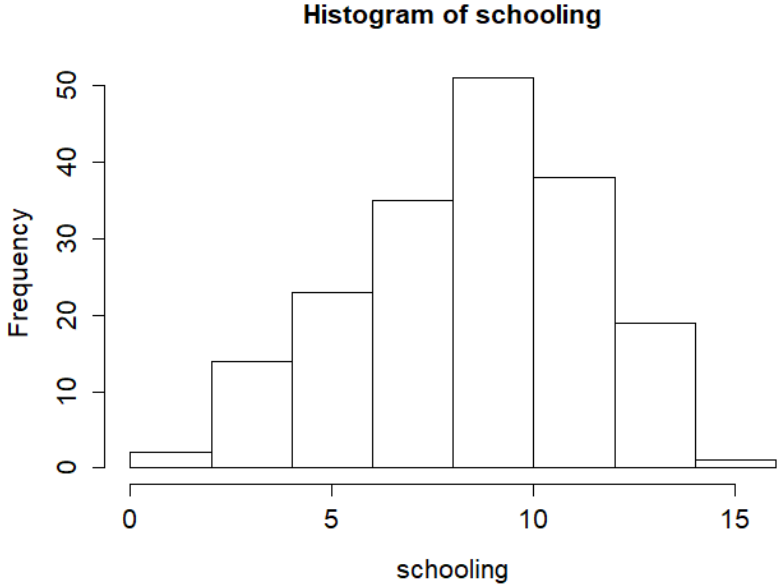Histogram of GDP per capita

Boxplot of GDP per capita

GDP per capita has some extremely large points, but they are economical possible.

Schooling

Summary

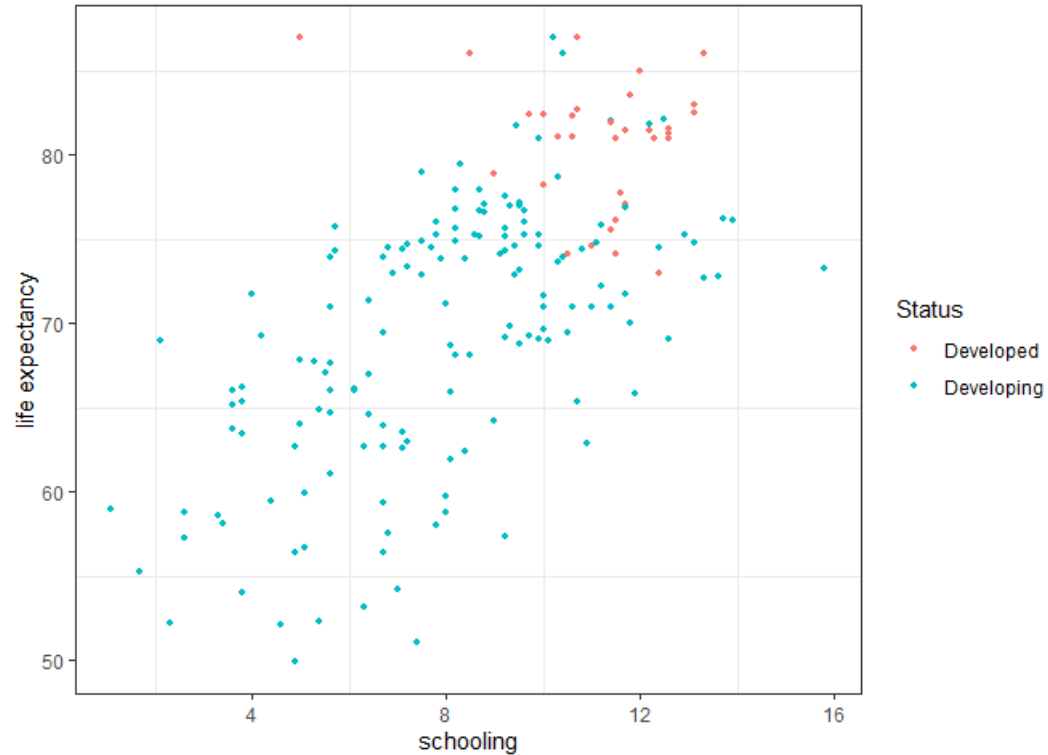| Min	1st Quarter	Median	Mean	3rd Quarter	Max
1.10	6.55	8.80	8.53	10.65	15.80

Plots

Histogram of schooling

Boxplot of schooling

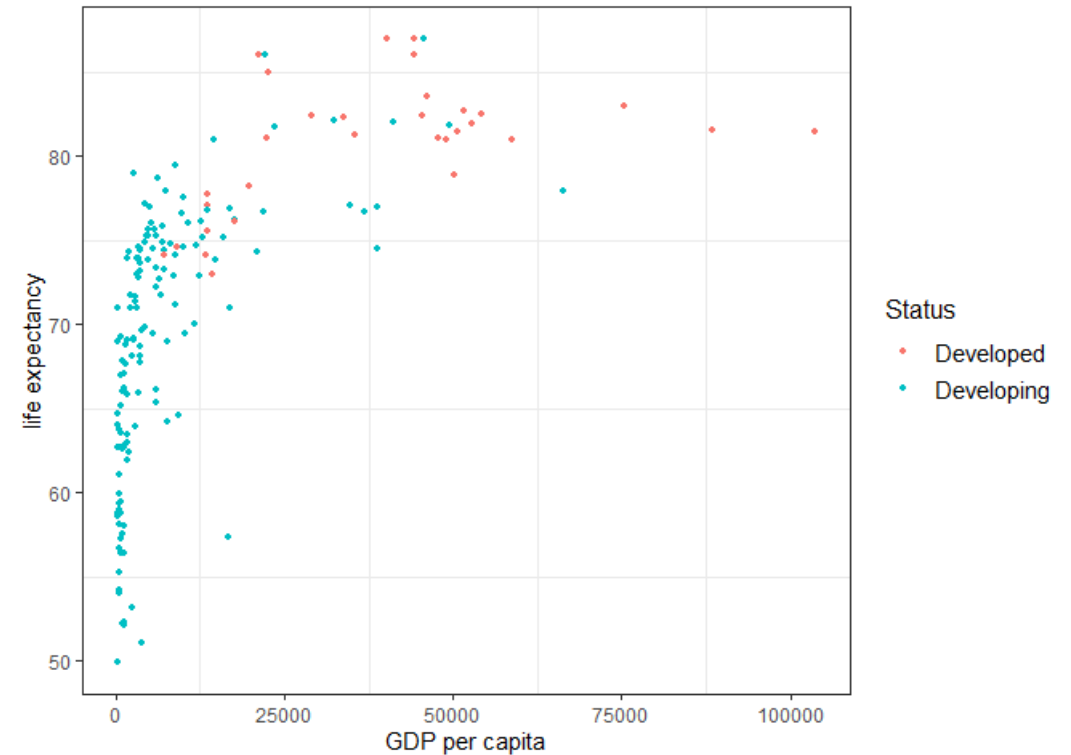No extremely large or extremely small points.

Scatterplots of life expectancy vs. schooling and GDP per capita

life expectancy vs. schooling (color-coding by status)

life expectancy vs. GDP per capita (color-coding by status)

GDP per capita & schooling are related to economics and society, so we expect them to be associated with life expectancy, with possible confounding by developed/developing status.

Linear regression

Model selection

Multivariable linear regression: $E(Y|X_1, \dots, X_P) = \beta_0 + \sum_{j=1}^{p} \beta_j X_j$

Consideration

🎯 Can include continuous response and explanatory variables.

🎯 Can easily include more than one explanatory variable and check association for each explanatory variable adjusting for other variables.

Multivariable linear regression model

	Estimate	P−value	95% confidence interval
Intercept	50.30	$8.71*10^{-66}$	(46.7, 53.9)
Slope for schooling	1.00	$2.22*10^{-9}$	(0.688, 1.31)
Slope for polio	0.075	$6.30*10^{-5}$	(0.039, 0.111)
Slope for government expenditure	0.64	$5.88*10^{-4}$	(0.278, 0.995)
Slope for GDP per capita	0.00014	$5.32*10^{-6}$	(0.0000824, 0.000202)
Slope for status (baseline is developing)	0.96	0.489	(-1.78, 3.70)

E(Life expectancy) = 50.30 + 0.075*polio + 0.64*government expenditure
+ 0.00014*GDP per capita + schooling - 0.96*status

Explanation of multivariable linear regression

💡 The following variables are significantly associated at the 0.05 level in the model we considered: (adjusting for all other variables)

- Schooling (p-value=$2.22*10^{-9}$) : life expectancy increases by 1 year for each additional year of school;
- Polio (p-value=$6.30*10^{-5}$): life expectancy goes up by 0.075 years for each additional percentage of polio immunization;
- Government expenditure (p-value=$5.88*10^{-4}$): life expectancy increases by 0.64 years for each additional percentage of government expenditure on health;
- GDP per capita (p-value=$5.32*10^{-6}$): life expectancy goes up by 0.00014 years for each additional dollar of GDP per capita.

💡 Status (p-value= 0.489) is not significantly associated at the 0.05 level in the model we considered.

The intercept: life expectancy is 50.30 years for a developing country with all polio immunization coverage, government expenditure on health, GDP per capita and schooling years are equals to 0.

Conclusion

Conclusion

How does life expectancy associated with different variables

All the better economic and society status, more health expenditure by government and higher immunization coverage leads to longer live expectancy.

Practical significance

To suggest countries which area should be given importance in order to efficiently improve the life expectancy of the population.

References

[1] Wilmoth, John R. "Demography of longevity: past, present, and future trends." Experimental gerontology 35.9-10 (2000): 1111-1129.

[2] Bongaarts, John. "How long will we live?." Population and Development Review (2006): 605-628.

[3] Ho, Jessica Y., and Arun S. Hendi. "Recent trends in life expectancy across high income countries: retrospective observational study." bmj 362 (2018): k2562.

Q&A

THANK YOU

[10 / 17 / 2019]