

Project GapMinder

Roli

June 12, 2019

In this Exploratory data analysis assignment, the Gapminder Dataset is used for analysis. In Data Description (description of each variables). In Data exploration section it is regionwise life expectancy and visualization (boxplots, histograms) of the different variables present in the data. We will be omitting missing values and applying clustering.

1. QUESTIONS :

- 1.1 Find the high-income countries with the lowest life expectancy?
- 1.2 Find the three countries with the longest life expectancies and United States?
- 1.3 Find the Life expectancy by region?
- 1.4 Find the High income countries with low life expectancies?
- 1.5 what is the Life Expectancy over years?
- 1.6 For the year 2015, what are the top three countries with the largest income?
- 1.7 What was the percent growth (or decline) in income in 2015 for the United States?
- 1.8 income vs year for United States?

2. Data Description:

GapMinder is the dataset which has income across the countries collected over years dating 1800 to 2015

Following are the variables in the dataset:

1. Country : Names of the countries and is a Factor w/ 197 levels
 2. Year : year for which this observation is collected is int
 3. Life : Life Expectency for people in that country is num
 4. population : Population of country is Factor w/ 15260 levels
 5. income : It is the Income of that country is int
 6. region : continent of that country is Factor w/ 6 levels
- the number observations is 41284
 - the number of variables are 6
 - types of variables 1. Country : is a Factor w/ 197 levels 2. Year : is int 3. Life : is num 4. population : is Factor w/ 15260 levels 5. income : is int 6. region : is Factor w/ 6 levels

Summary of gapminder

3. Data Exploration:

Data summary statistics

3.1) Mean and Median of life expectancy Region wise

- Europe & Central Asia life mean is 41.74110
- America life mean is 35.37370
- Europe & Central Asia median is 48.79419
- America life median 44.54065

3.2) Five Num Summary Life Expectancy Region wise

Q.Find the three high-income countries with the lowest life expectancy

Ans.We found that three high-income countries with the lowest life expectancy were

- Lesotho life as 48.50, income as 2598, region as Sub-Saharan Africa
- Swaziland life as 51.50, income as 6095, region as Sub-Saharan Africa
- Guinea-Bissau life as 53.40,income as 1386, region as Sub-Saharan Africa

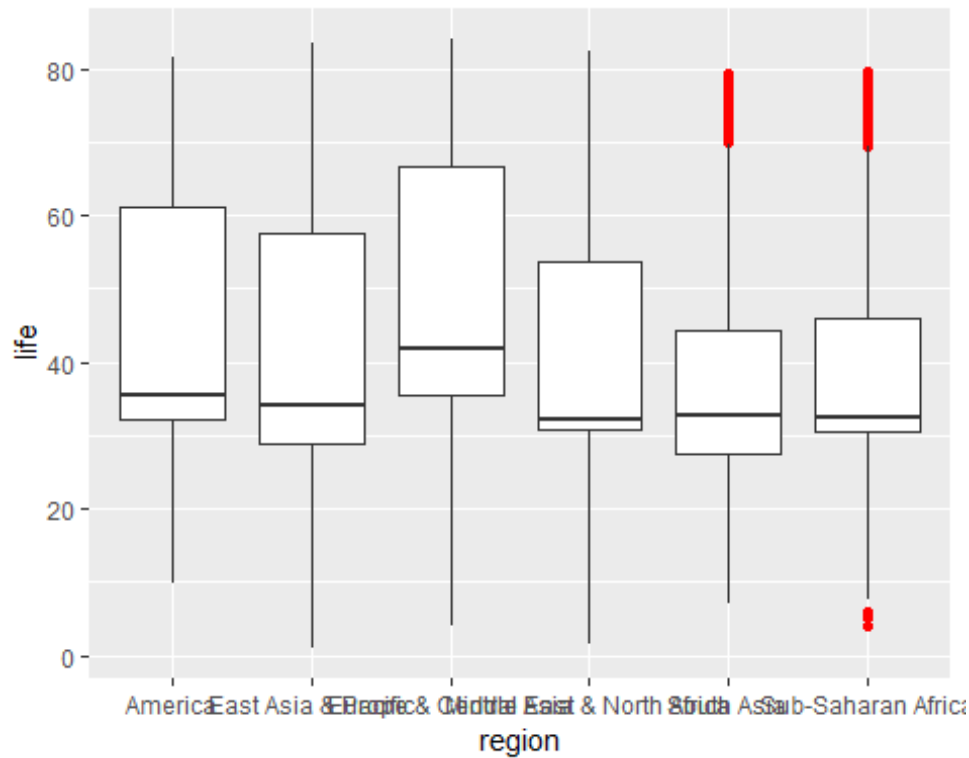
Q.Find the 3 countries with the longest life expectancies and United States

- Andorra life as 84.1, income as 46577, region Europe & Central Asia, rank 1,
- Japan life 83.5,income as 36162, region East Asia & Pacific, rank 2,
- Switzerland life 82.9, income as 56118, region Europe & Central Asia, rank 3,
- and lastly for United States life 79.1, income as 53354, region America, rank 36

Longevity Summary

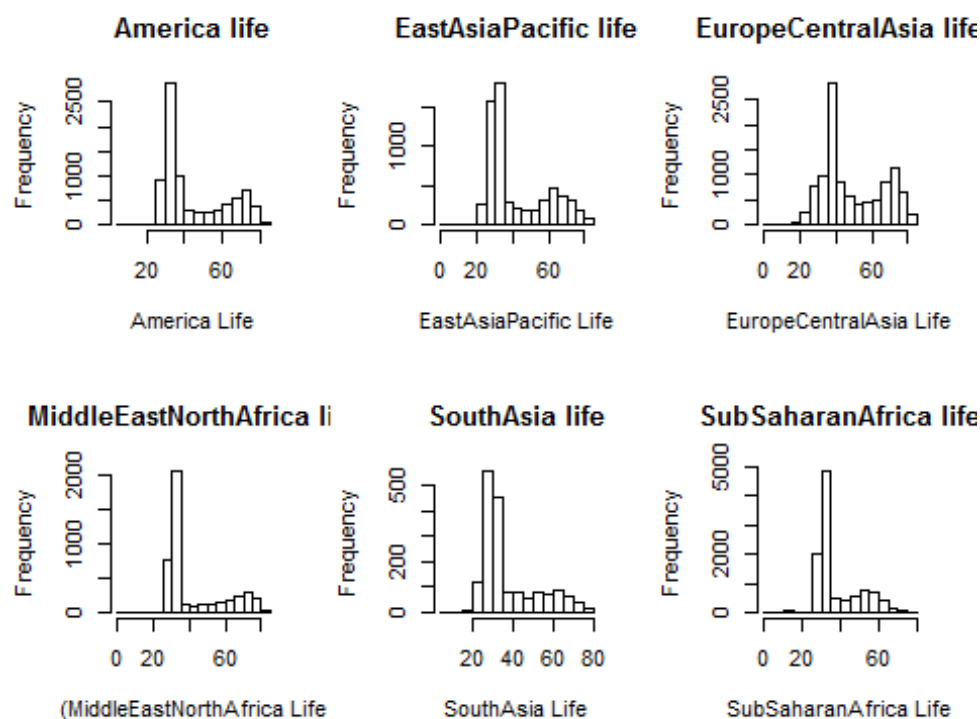
4. Data Visualization:

4.1 Boxplot of life expectancy by region

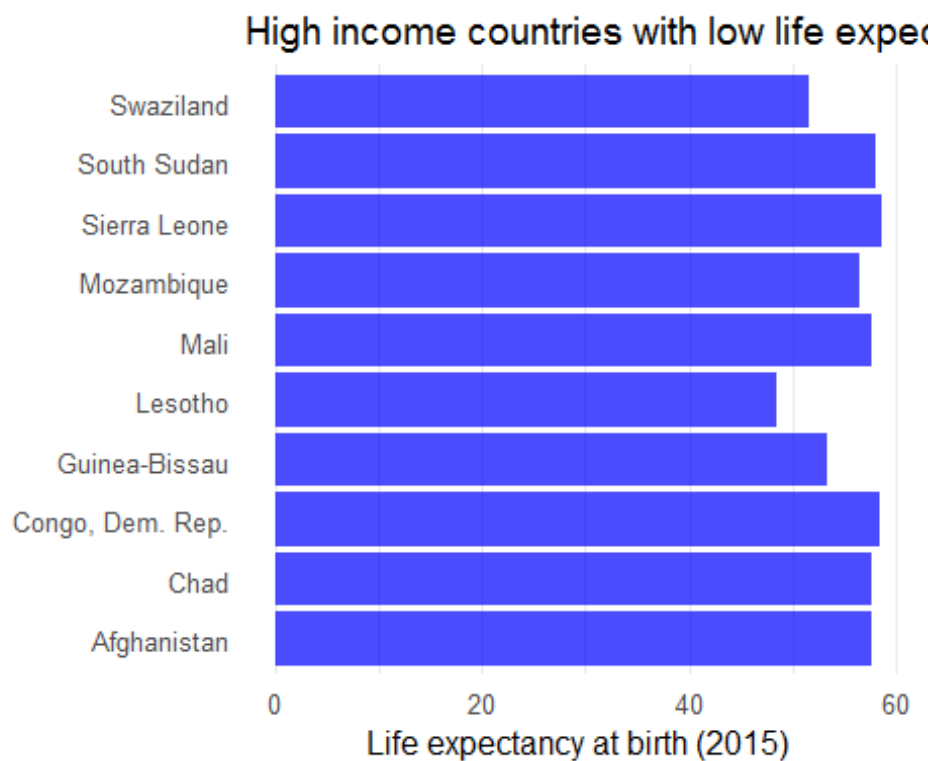


- Some interesting observations: Europe is substantially higher than all continents. Asia is also VERY spread out with a few countries at the top of our distribution. it is seen that Europe and Central Asia's life expectancy is substantially higher than all continents. Asia is also VERY spread out with a few countries at the top of our distribution.

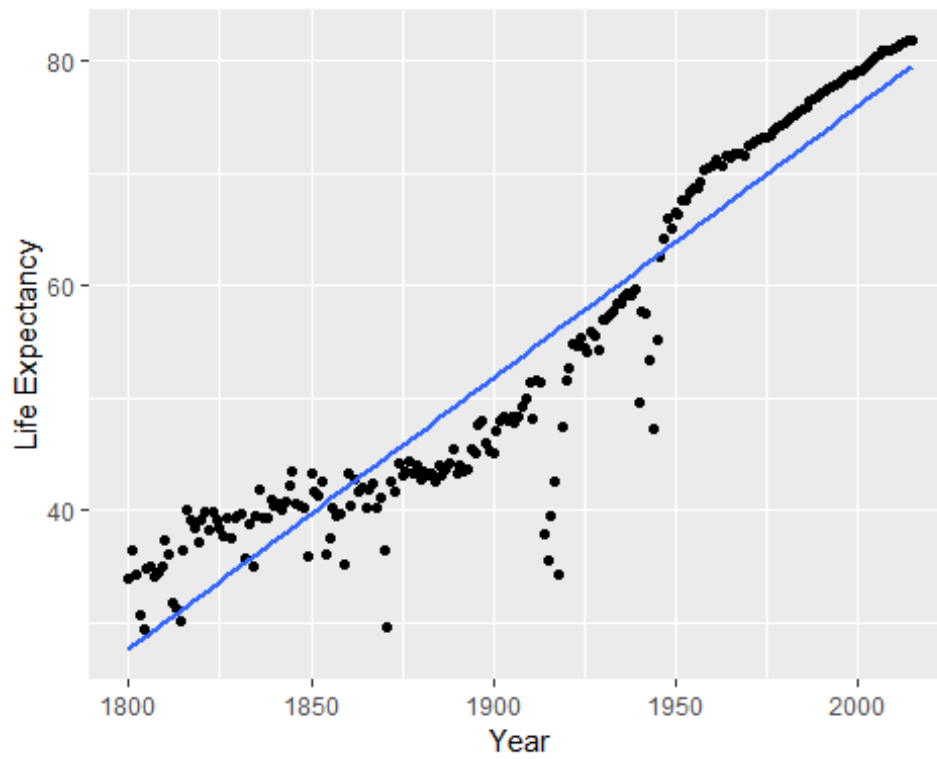
4.2 Histogram of life expectancy by region



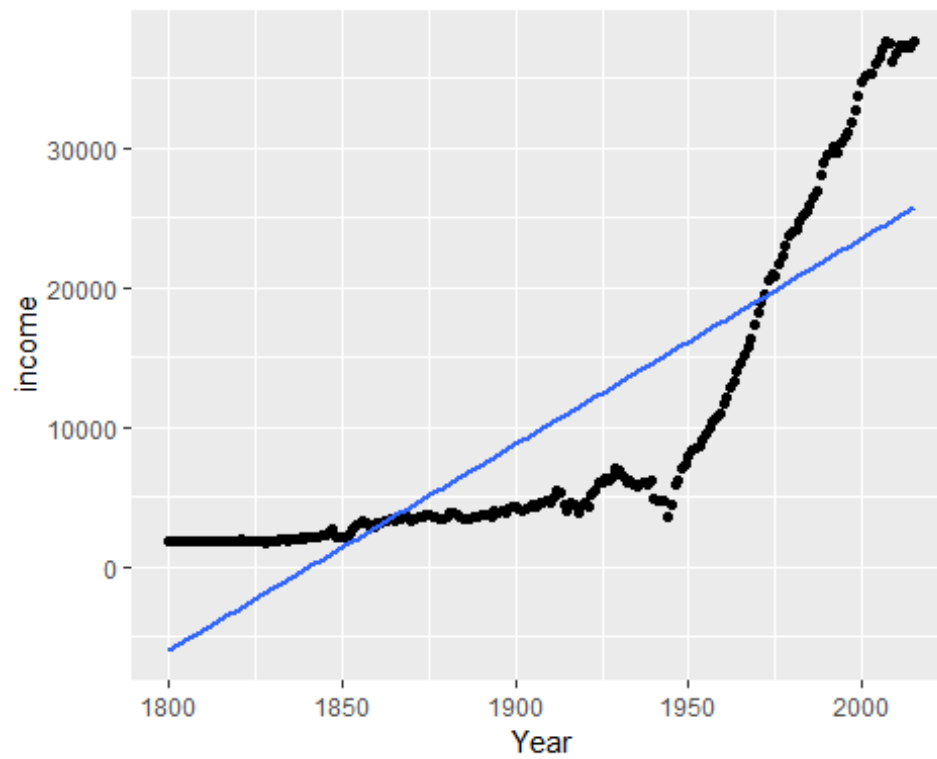
High income countries with low life expectancies



Life Expectancy over years



- This diagram shows Life Expectancy by Year



Q.For the year 2015, what are the top 3 countries with the largest income?

- Ans .were Qatar with income as 132877, Luxembourg with income as 88314, Kuwait with income as 82633

Q.What was the percent growth (or decline) in income in 2015 for the top 3 countries with largest income?

Ans. Kuwait income was 82633, and percent growth is 44.30318

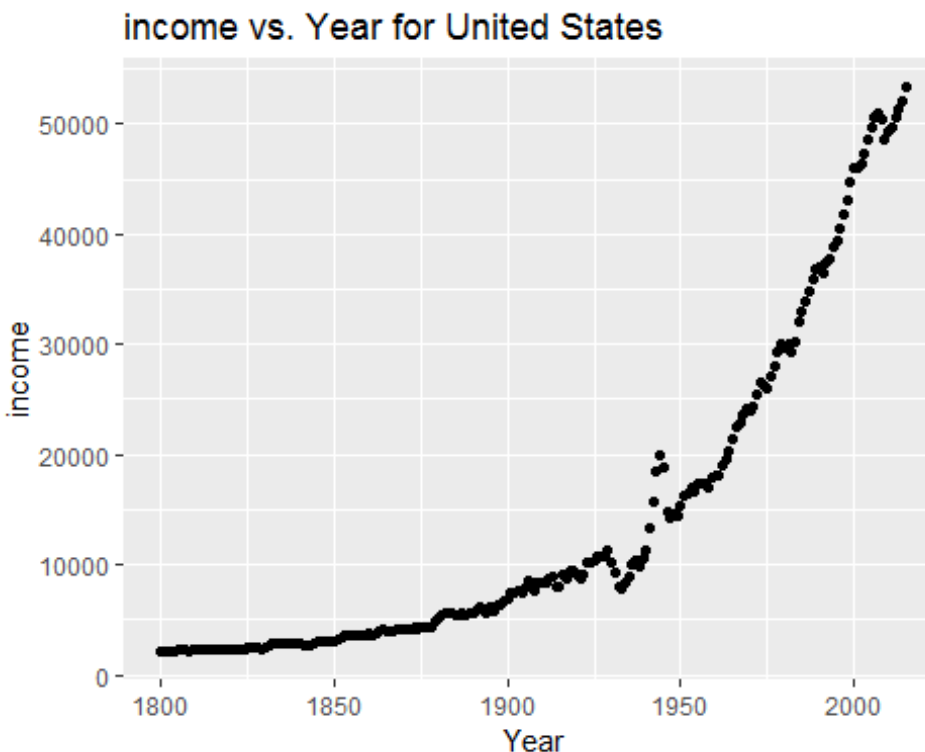
Singapore income was 80794, and percent growth is 37.75012,

Malaysia income was 24320 and percent growth is 29.43805

What was the percent growth (or decline) in income in 2015 for the United States?

- 1.United States in year 1942 life was 64.5332 and income 15702 and percent growth is 0.1860412 in 2015 We had a 0.18% growth which is not so good.

income vs year for United States



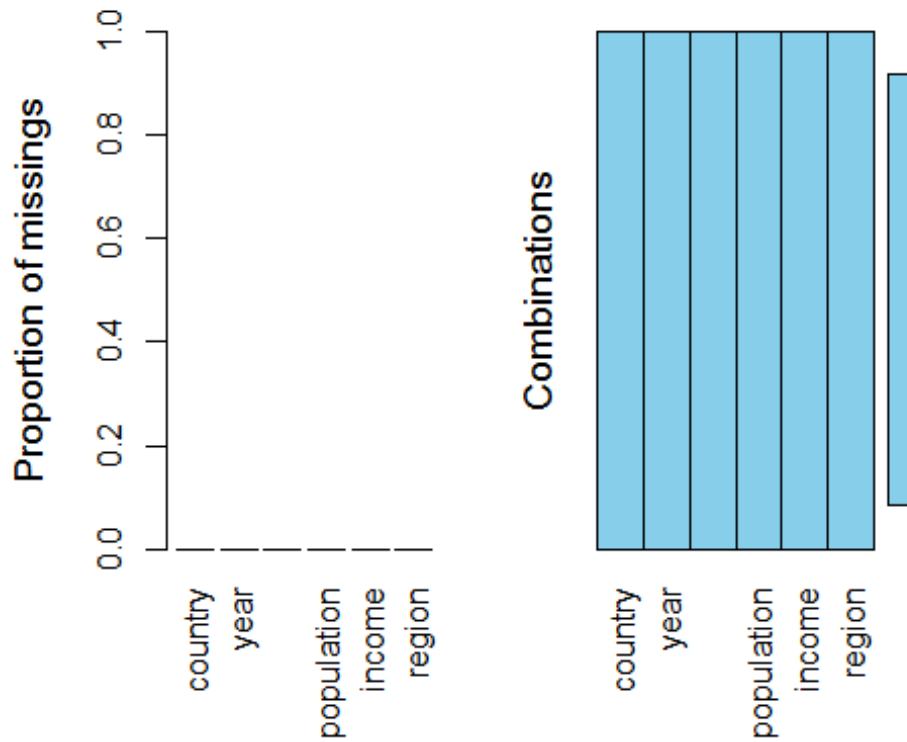
- There's been a steady increase with a few crashes here and there for the income of USA over years.

5. Hierarchical Clustering

It's advisable to gather some basic useful information about the dataset like its dimensions, data types and distribution, number of NAs etc. You will do so by using the `str()`, `summary()` and `is.na()` functions in R.

5.1 Dealing with Missing Values

using Imputation to deal with missing values, used mice method based on random forest to impute.

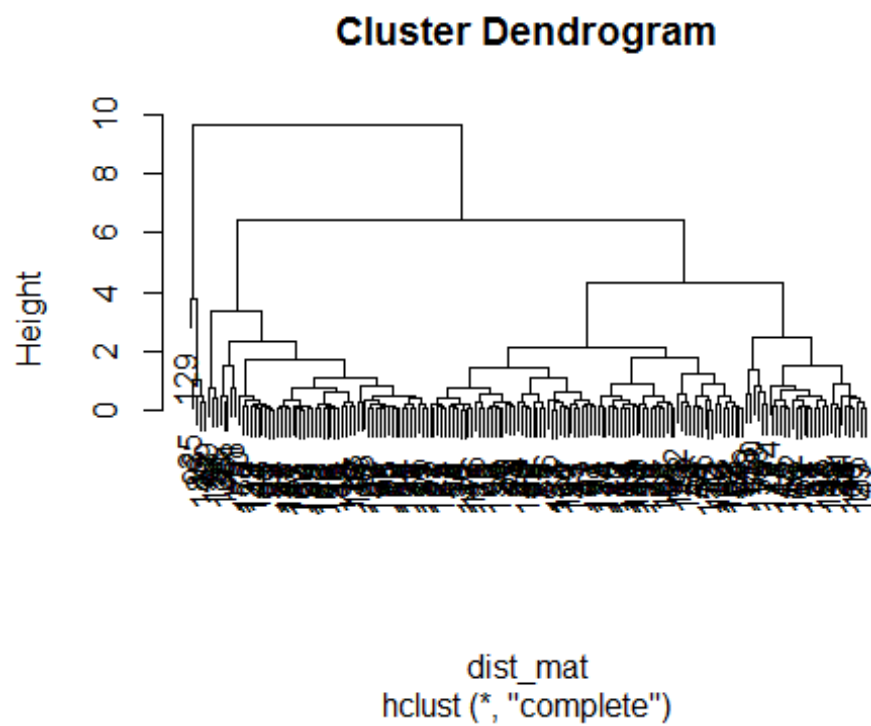


5.2 K-Means Clustering

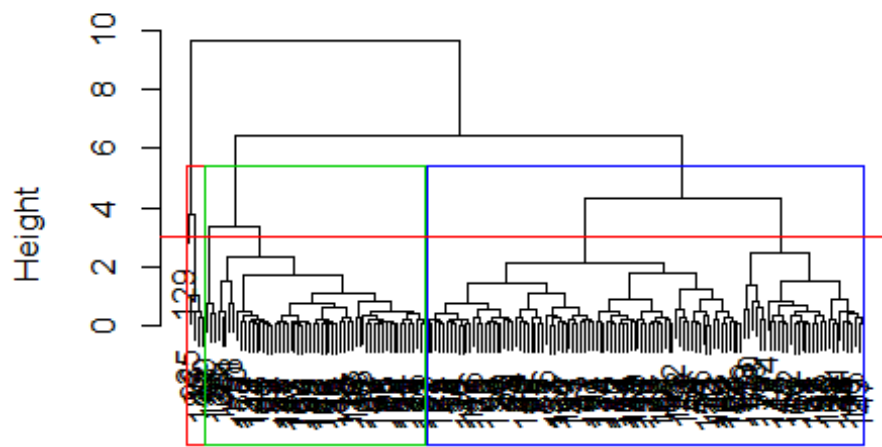
no missing values



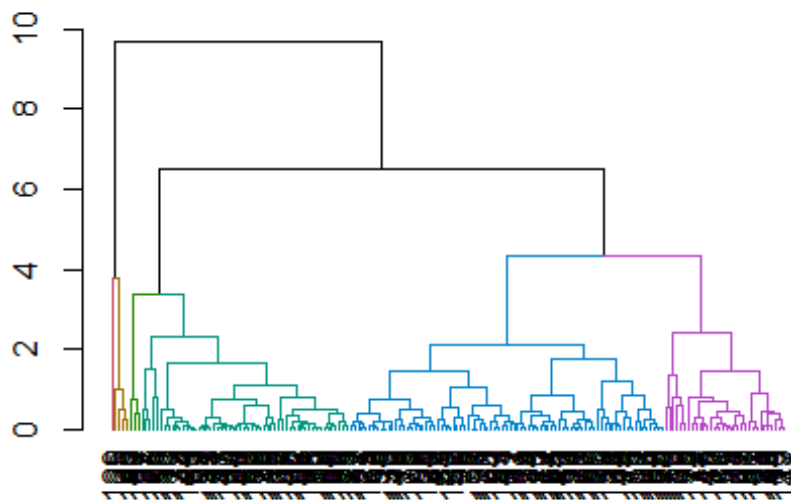
Agglomerative coefficient



Cluster Dendrogram



dist_mat
hclust (*, "complete")



SUMMARY

We found that three high-income countries with the lowest life expectancy were Lesotho life as 48.50, income as 2598, region as Sub-Saharan Africa Swaziland life as 51.50, income as 6095, region as Sub-Saharan Africa Guinea-Bissau life as 53.40, income as 1386, region as Sub-Saharan Africa

We found that the the 3 countries with the longest life expectancies and United States

Andorra life as 84.1, income as 46577, region Europe & Central Asia, rank 1, Japan life 83.5, income as 36162, region East Asia & Pacific, rank 2, Switzerland life 82.9, income as 56118, region Europe & Central Asia, rank 3, and lastly for United States life 79.1, income as 53354, region America, rank 36

Europe and Central Asia 's life expectancy is substantially higher than all continents. Asia is also VERY spread out with a few countries at the top of our distribution.

we saw the graph for high income countries with low life expectancies and saw that zimbabwe , zambia,swaziland,south sudan etc are in the list.

we saw that Life expectancy improved over years especially after 1950.

we saw that in the year 2015, the top 3 countries with the largest income were

were Qatar with income as 132877, Luxembourg with income as 88314, Kuwait with income as 82633

we saw that the percent growth (or decline) in income in 2015 for the top 3 countries with largest income were Kuwait with income 82633 and Percent Growth as 44.30318, Singapore with income 80794 and Percent Growth as 37.75012, Malaysia with income 24320 and Percent Growth as 29.43805

we also saw that the percent growth (or decline) in income in 2015 for the United States was 0.18% which is not so good. There's been a steady increase with a few crashes here and there for the income of USA over years

References

1. <https://www.r-project.org/help.html>
2. <https://www.rdocumentation.org>