**Do Health Status and Mortality Causes Have a Correlation with Demographics and Location?**

**Spring 2019 - Business Analytics for Managers BYGB-7975-001**

**Final Report**

**05/01/2019**

**Team Wolverine**

Hayley Teng

Ila Srivastava

Katie Cao

Sravya Katta

Victoria Cleveland

Hayley (Jie) Teng – OECD Health Status Snapshot

A screenshot of a cell phone

Description automatically generated

Figure 1. OECD Health Status Snapshot

* The Figure 1 summarize the overall OECD Health Status and the main indicators used in the study.
* At the country level, United States, Iceland and Israel have the best overall health status compared with the rest of countries. While female’s life expectancy at birth is higher than male’s life expectancy at birth, female’s good/very good health status population percentage is overall lower than male’s good/very good health status population percentage for each country.
* The snapshot not only give the big picture perspective of the dataset, but also implies the positive correlation between the population percentage of good/very health status with high education and number of deaths caused by all diseases, which gives the potential of the study for deeper analysis.

Victoria Cleveland – Average life expectancy by gender

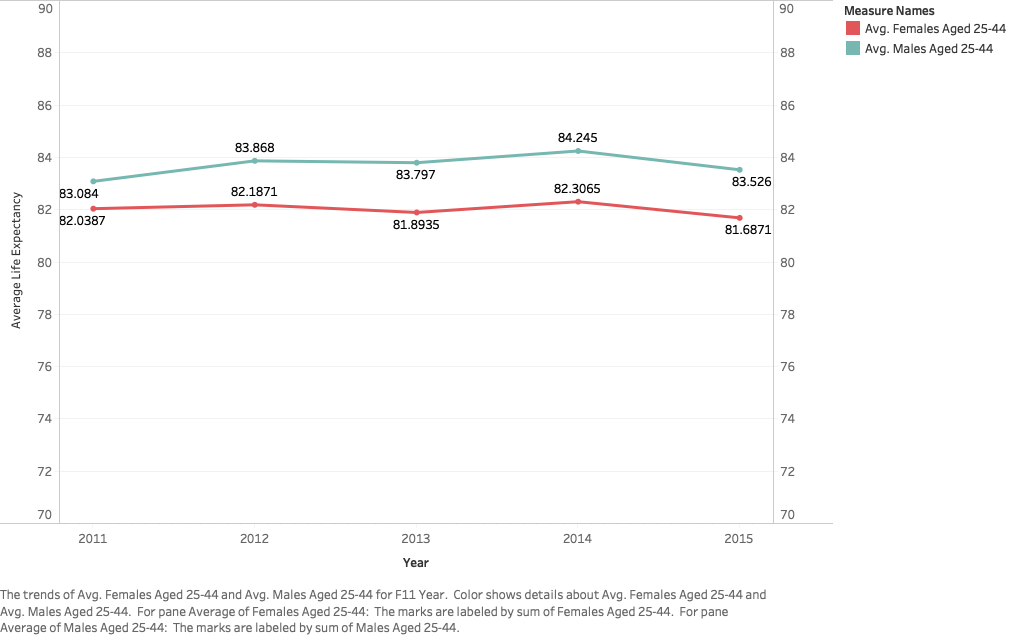


Figure 2. Line chart of Average Life Expectancy by Gender from 2011-2015

* This chart compares the trend in average predicted life expectancy of males and females aged 25-44 between the years of 2011-2015.
* The insight here is that the general trend in life expectancy goes up and down for females and males at the same rate, however, males between the ages of 25-44 are predicted to have higher life expectancies than females of the same age.
* The implication here is that both males and females can expect their average life expectancy to fluctuate at the same rate, however females of this age group should be more mindful of factors that could lower their life expectancy since theirs is generally lower than males at this stage in their lives.

Ila Srivastava – Life expectancy by country

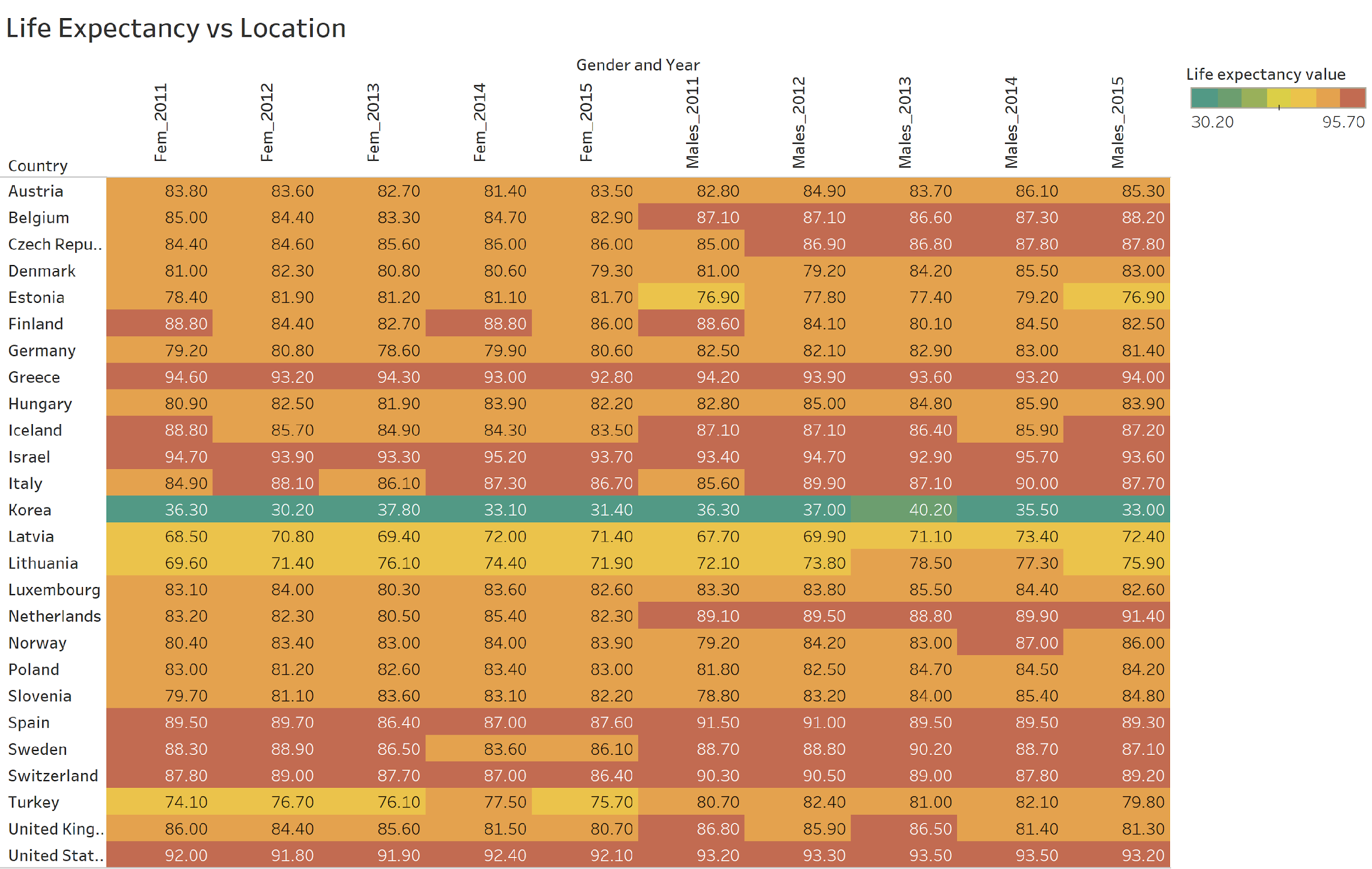


Figure 3. Heat map detailing life expectancy over time of 25-44-year-old males and females across all the locations

* This heat map chart indicates life expectancy for a period of 5 years (2011-2015) for males and females in age range of 25-44 for all locations.
* The insight gained from this chart is that overall Greece and United States(Dark orange) have higher life expectancy with males having higher values than females whereas Korea(Green) have lower life expectancy and need to carefully look at their health status and other demographic factors affecting their life expectancy in order to bring improvement.
* The implication here is that Government should examine the factors leading to decrease in life expectancy in countries with lower values (Green and yellow ones specifically).

Hayley (Jie) Teng – Life Expectancy at Age 65 Gender Comparison

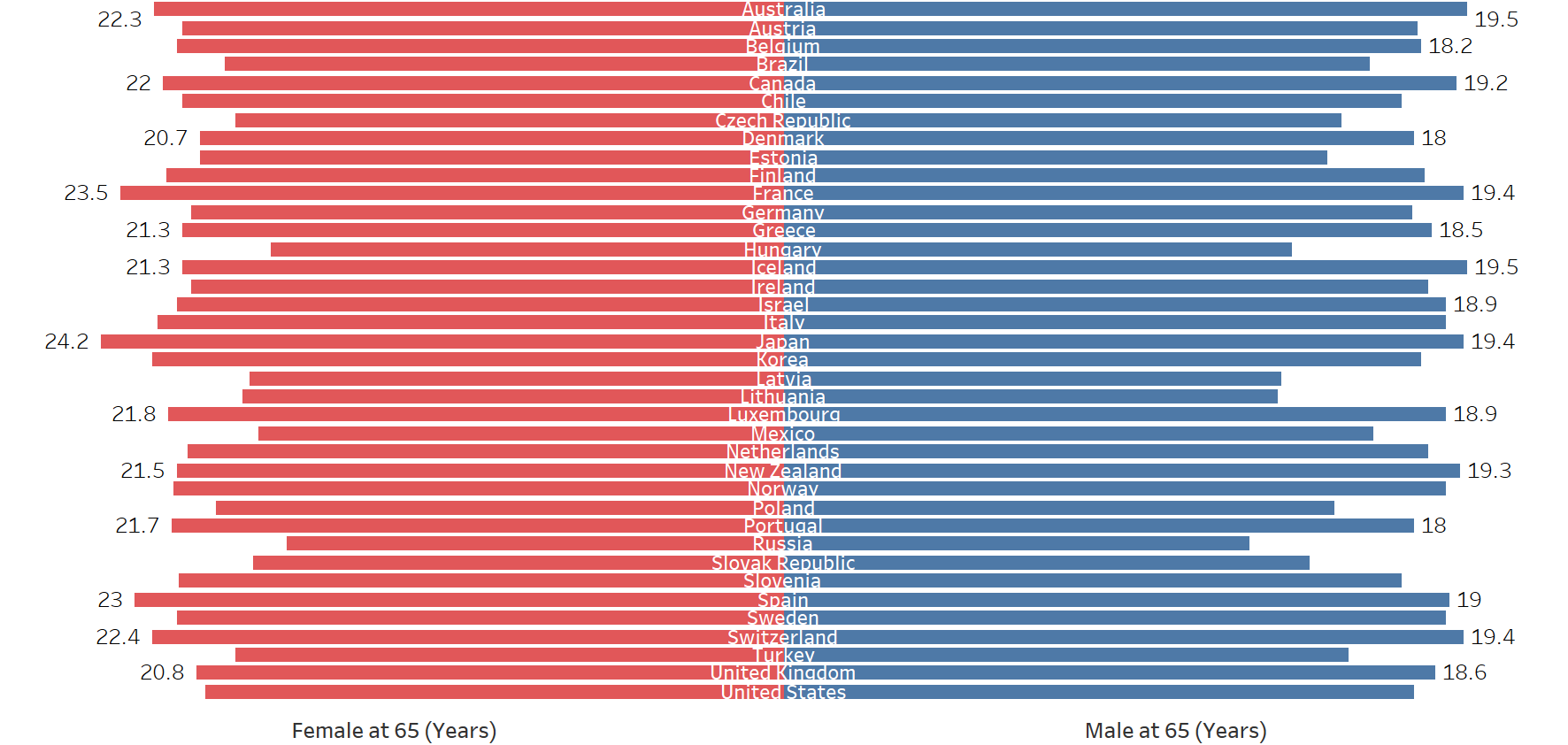


Figure 4. Butterfly Bar Chart of Life Expectancy at Age 65 for Gender Comparison

* The Figure 4 demonstrates the gender difference of life expectancy at age 65 for each country.
* Overall, for each country, the female’s life expectancy at 65, which measured by years, is slightly higher than the male’s life expectancy at 65.
* The Butterfly bar chart’s analysis result looked into the aging population whose age starts at 65 and the insights matched with the life expectancy from birth for both gender.

Ila Srivastava – Life expectancy variation for males of 60 and 80 years

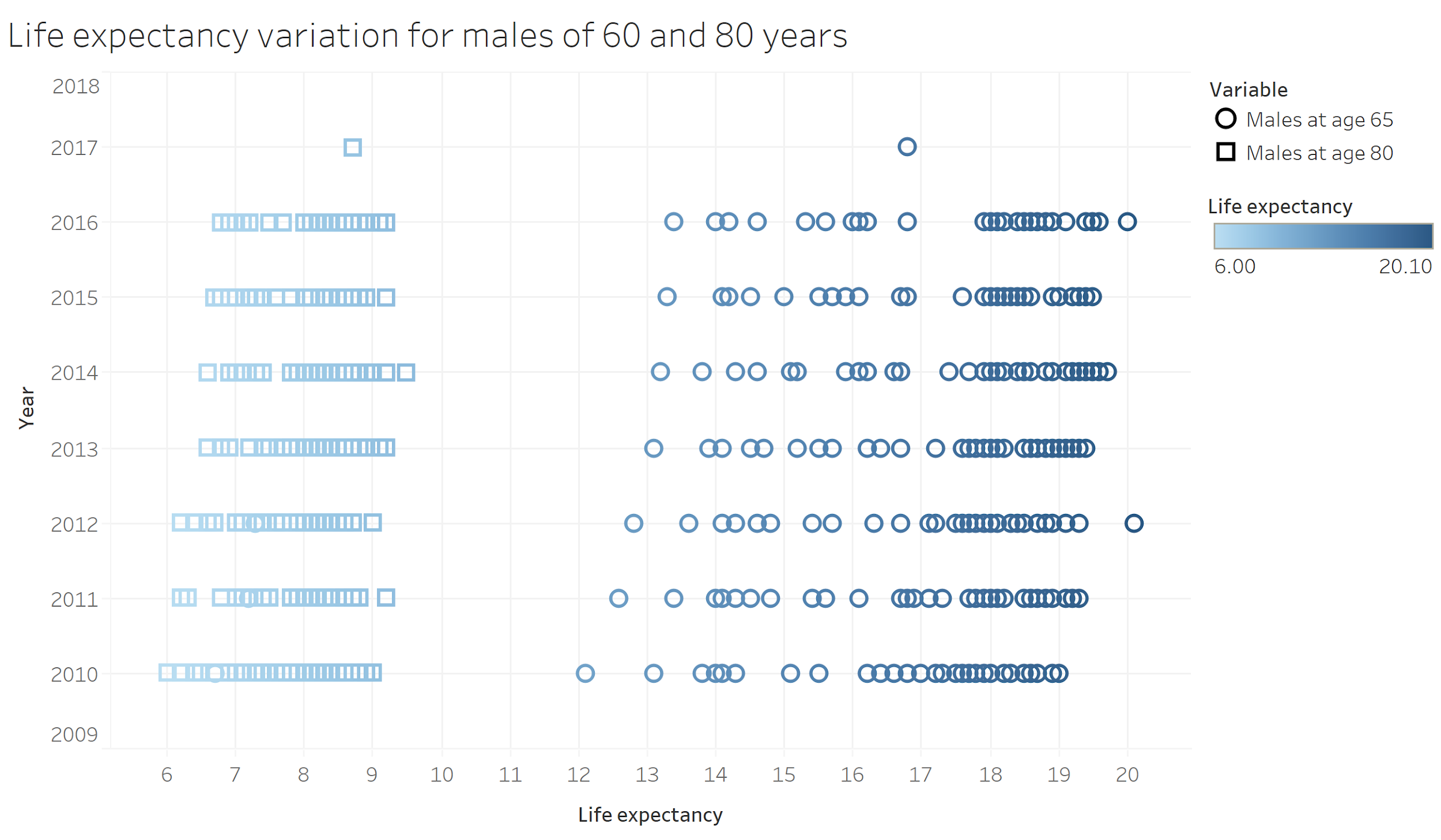
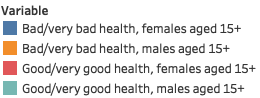


Figure 5. Scatter plot for life expectancy for males aged 65 years and 80 years over the years

* This scatter chart indicates life expectancy for a period of 7 years (2010-2017) for males aged 65 and 80 years.
* The insight gained from this chart is that males having age 65 have higher life expectancy independent of other factors affecting their life expectancy whereas 85-year-old males have expectancy varying from 6 to 9 years. Over the years the life expectancy seems to have improved for both categories of males.
* The implication here is that over the years factors that might be affecting their life expectancy have been improved thereby improving overall life expectancy.

Victoria Cleveland – Trend in health status by gender



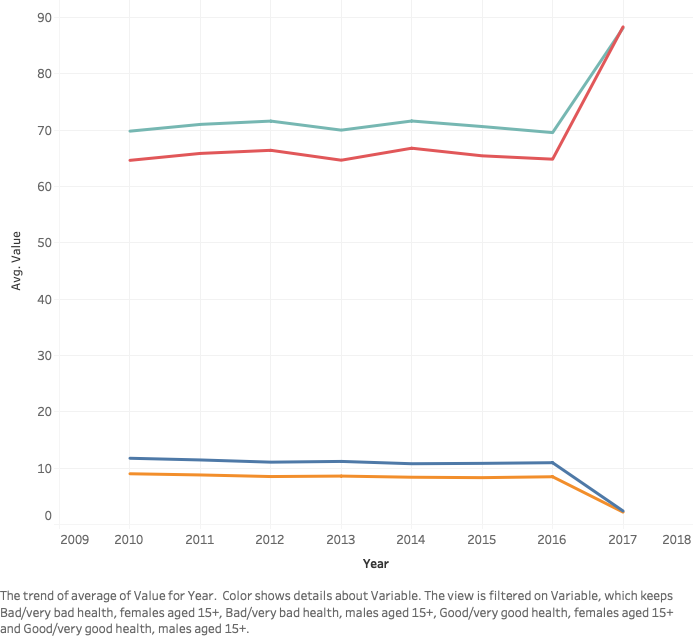


Figure 6. Line chart showing the trend in health status by gender 2010-2017

* This chart depicts the trend of both males and females in very bad health and in very good health.
* The insight here is that from 2016-2017, those with bad health sharply declined in health status, becoming worse, and in the same time period those with good health sharply increased in health status, becoming better.
* The implication here is that those in bad health have to pay extra attention to their health status in order to improve it to the good health status.

Katie Cao - Predicting life expectancy at birth using percentage of population in good health, 2011-2015

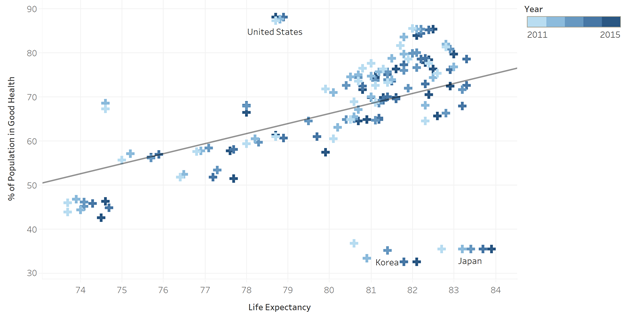


Figure 7. Scatter plot with the linear trend line of predicting life expectancy at birth using percentage of population in good health, 2011-2015

* The scatter plot depicts the relationship between the life expectancy of total population (by years) and the percentage of population in good health, incorporating the linear regression trend line to predict the relationship between the two variables.
* % of Population in Good Health is positively correlated with the life expectancy of the total population. There are outliers on the middle top (those are data of United States) and the ones at the bottom right (Korea and Japan).
* It would be interesting to further study other data of outlier countries (United States, Korea, and Japan) to understand why the trend isn’t the case for those countries. For instance, why is it that in the States, high percentage of population in good health does not imply higher life expectancy compared to other countries?

Victoria Cleveland – Average number of deaths per 1000 births

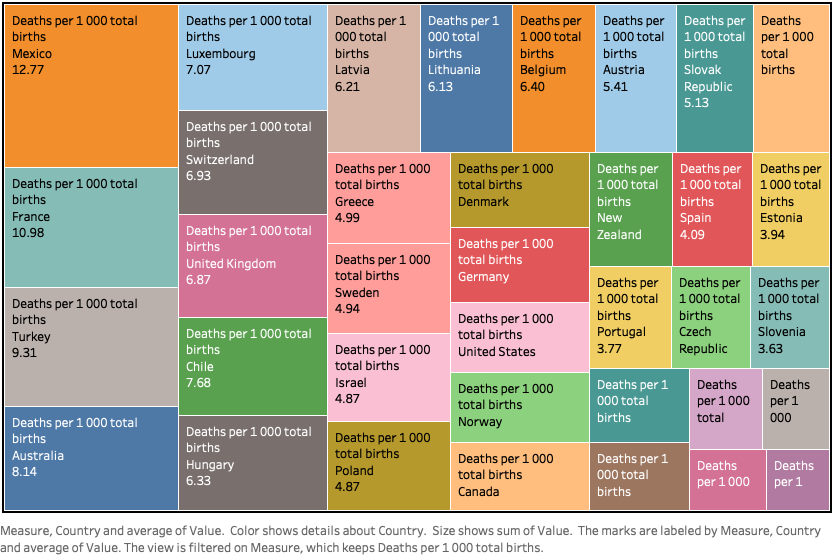


Figure 8. Treemap of average number of deaths per 1000 births

* This chart shows the average number of deaths per every 1,000 births, sorted by highest to lowest.
* The insight here is that Mexico, France, Turkey, and Australia all have the highest number of deaths per 1,000 births.
* The implication here is that countries with the highest death rate per births must put extra effort and attention into their maternal care sectors.

Sravya Katta – Relationship between life expectancy and drug use disorders

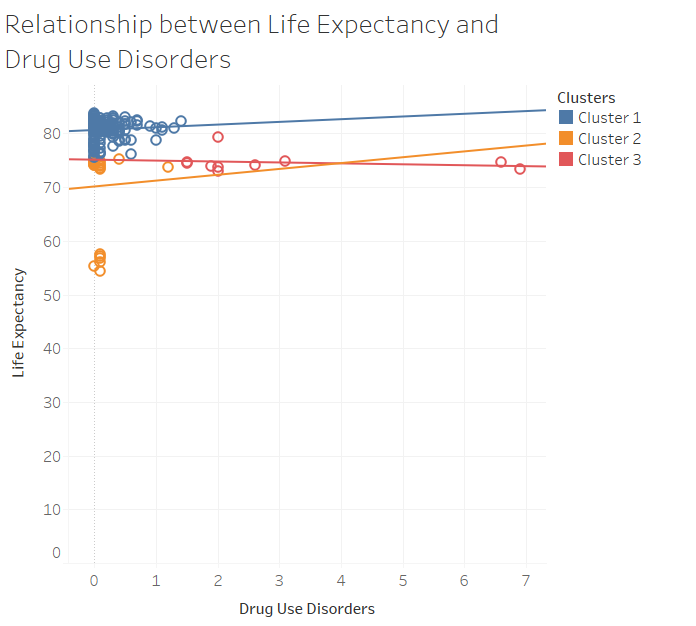


Figure 9. Relationship between life expectancy and drug use disorders

* Figure is a scatterplot that shows the relationship between Drug Use Disorders vs. Life Expectancy. Color shows details about Clusters which categorizes drug usage deaths.
* This shows that when there is high drug usage, there also exists a decrease in life expectancy as per the trend line. But, where there is low or moderate drug usage, we can see that there is increase in life expectancy.
* Drug/substance usage disorder is very prevalent across the globe. In order to improve the life expectancy and health status, government has to improve counselling and rehabilitation facilities and make them easily accessible to help people with substance addiction.

Ila Srivastava – Death per 100 000 population for accident types 2010-2016

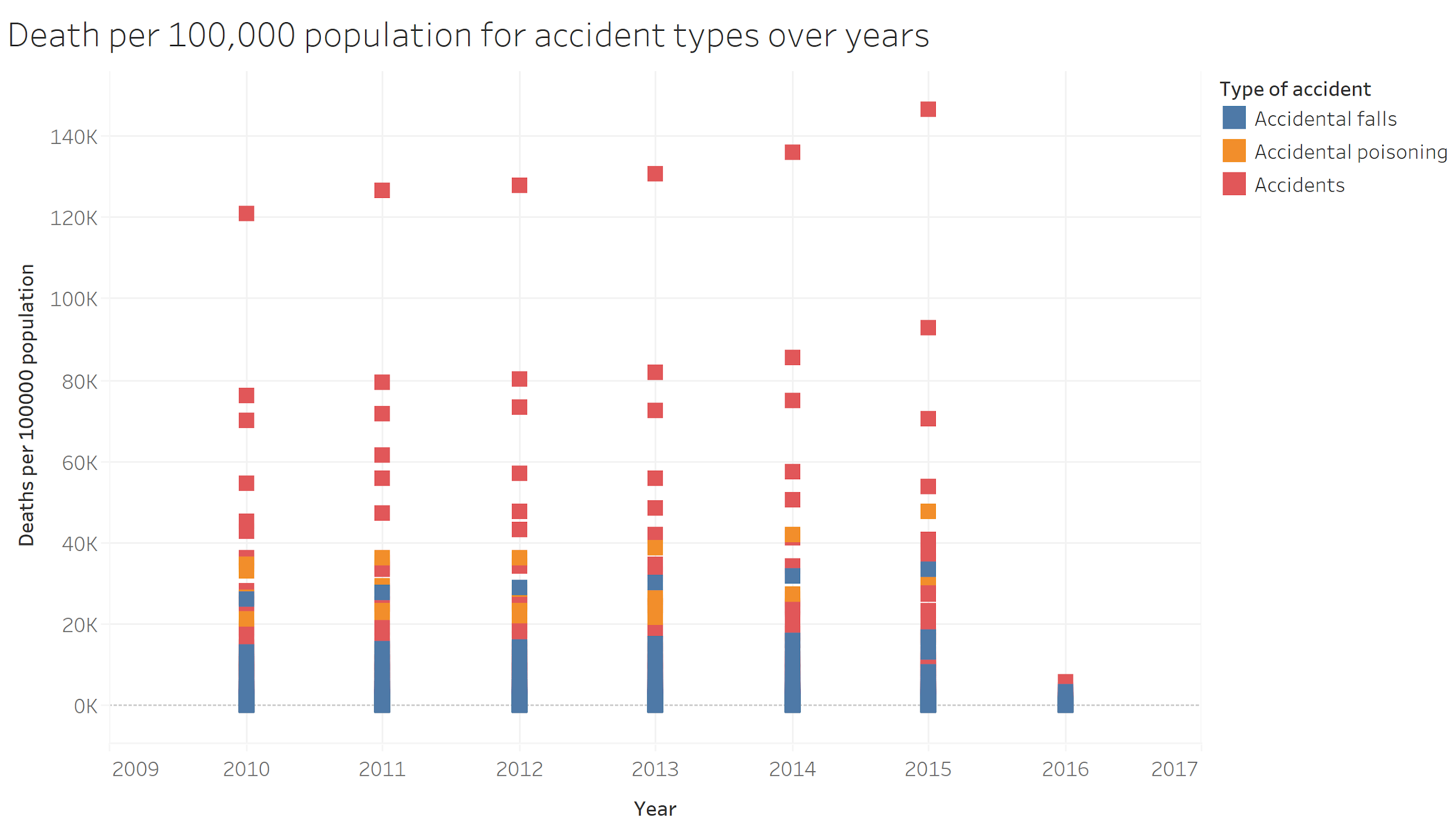


Figure 10. Scatter plot to show deaths per 100,000 population for accident type through years 2010-2016

* This scatter chart indicates accident types and how that affect deaths per 100,000 population through years 2010-2016 affecting the life expectancy.
* The insight gained from this chart is over the years the major reason for death is accidental falls and accidents. Also, general accidents is the top reason for deaths and chart indicates that accidents have increased over the years causing higher number of deaths.
* The implication here is that over the government need to improve and incorporate good policies for improving rules and regulations to prevent any types of accidents and keeping a regular check on regulatory status to prevent deaths.

Sravya Katta – Deaths per 100 000 population due to digestive diseases

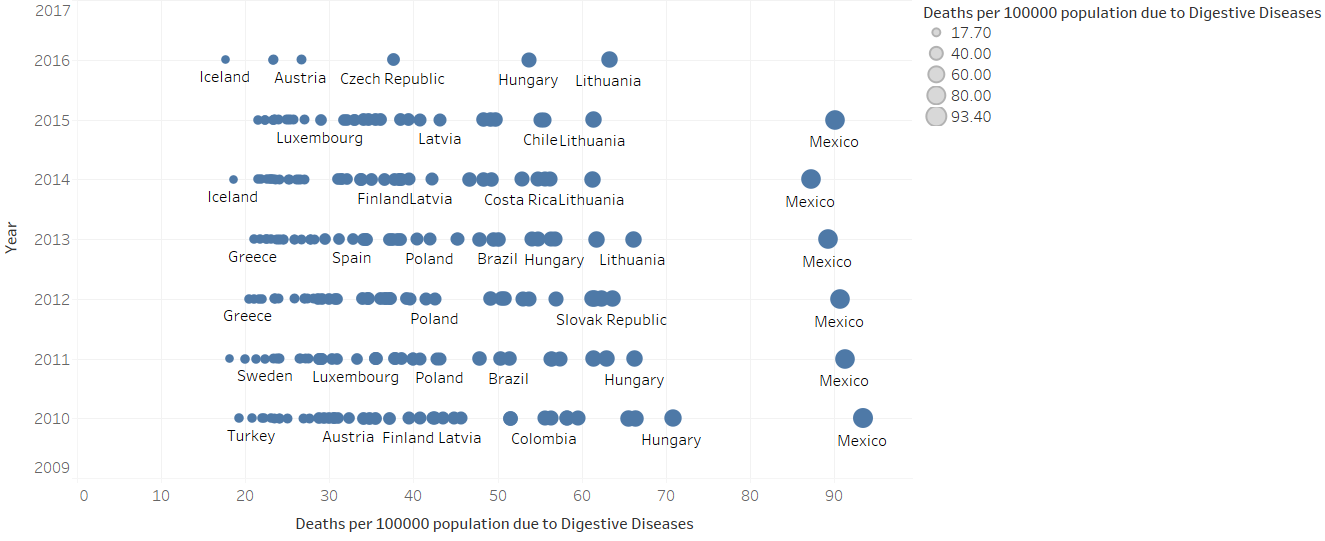


Figure 11 Deaths per 100 000 population due to digestive diseases

* The above Gantt plot is a of sum of Deaths per 100 000 population due to Digestive Diseases for Year. Size shows sum of Deaths per 100 000 population due to Digestive Diseases. The marks are labeled by Country.
* We can analyze that countries with low/moderate per capita income such as Mexico, Lithuania, Hungary, Columbia also tend to have higher number of deaths due to digestive diseases. This pattern can be observed across the time span of 2011 to 2015. It can be noted that these countries also face socio-economic issues.
* Countries facing socio-economic, security issues should take special measures to tackle frequently occurring diseases such as digestive diseases in order to control the increasing mortality rate from it.

Katie Cao – Gender gap variance of mortality caused by tuberculosis, 2010-2015

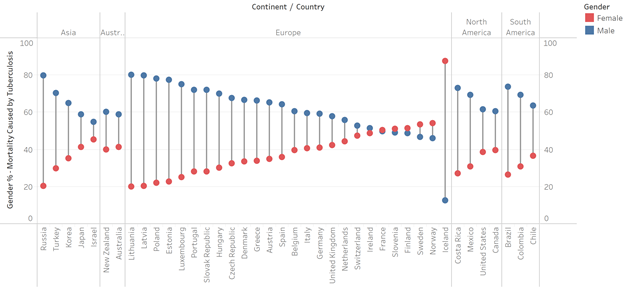


Figure 12. Barbell chart showing the gender variance of Mortality caused by Tuberculosis (TB) by Country, 2010-2015

* The barbell chart shows Gender variance by percentage of Mortality caused by Tuberculosis disease by Country, arranging in descending order of the gender gap per each continent.
* There is a variance in gender when it comes to mortality caused by TB, most significantly in Russia, Lithuania, Latvia, Poland, and Iceland. In most countries, Males have a higher percent of mortality caused by Tuberculosis than Females. The exception are some European countries, namely Iceland, Norway, and Sweden, where more females died of this disease than males.
* Tobacco is known to be increasing the risk of developing tuberculosis. This gender analysis in mortality rate by TB can raise further research question that males are likely to die from TB because of their tendency to smoke. Studies in smoking rate among genders in Iceland, Norway, and Sweden, for example, could be done to compare why this isn’t the case for those countries.

Katie Cao – Mortality rate caused by transport accidents 2010-2015

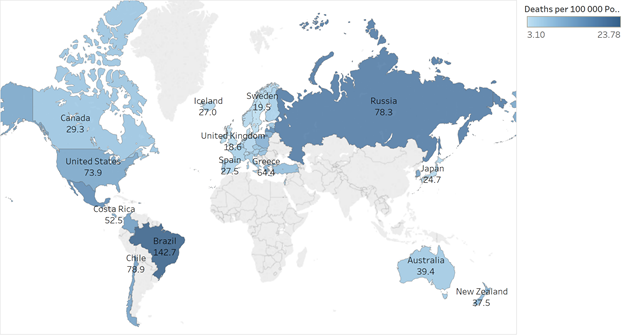


Figure 13. Map showing mortality rate caused by Transport accidents per 100 000 population, Average 2010-2015

* Above map shows the mortality rate caused by transport accidents per 100 000 population across countries with color ranging from light to dark, representing low to high number of deaths.
* Brazil has the highest mortality rated caused by Transport accidents per 100 000 population. Chile, Russia, and United States follow. Generally, European countries have low mortality rate caused by transport accidents.
* Countries in South America and Russia should learn from policies in European countries in how they ensure safety of their citizen on the road. Another reason could be better investment towards infrastructure and better education on transportation.

Katie Cao – Mortality caused by three major incurable diseases by continent, 2010-2015

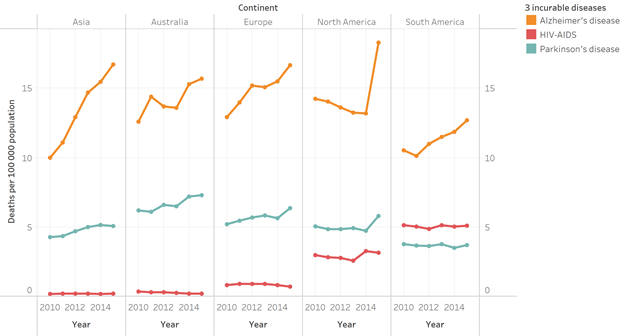


Figure 14. Line chart showing mortality per 100 000 population caused by three major incurable diseases by Continent, 2010-2015

* The line chart shows the number of deaths per 100 000 population caused by Alzheimer’s disease, HIV-AIDS, Parkinson’s disease by Continent from 2010-2015.
* Interestingly, we observe an increase in mortality cases of all threes for all continents that the dataset has available data, except HIV-AIDs go down for Europe from 2014 to 2015. Both Asia and North America sees a sharp increase in the cases of Alzheimer’s disease mortality cases. HIV-AIDS and Parkinson’s disease mortality cases do increase, but not at a fast pace as Alzheimer’s.
* As such diseases remain incurable, they seem to be evolving and cause more mortality across continents. Governments should allocate more funds towards researching the cure for such disease in order to maintain the wellbeing and keep the mortality rate down.

Hayley – Female Mortality Population Percentage of Selective Two Disease

A screenshot of a cell phone

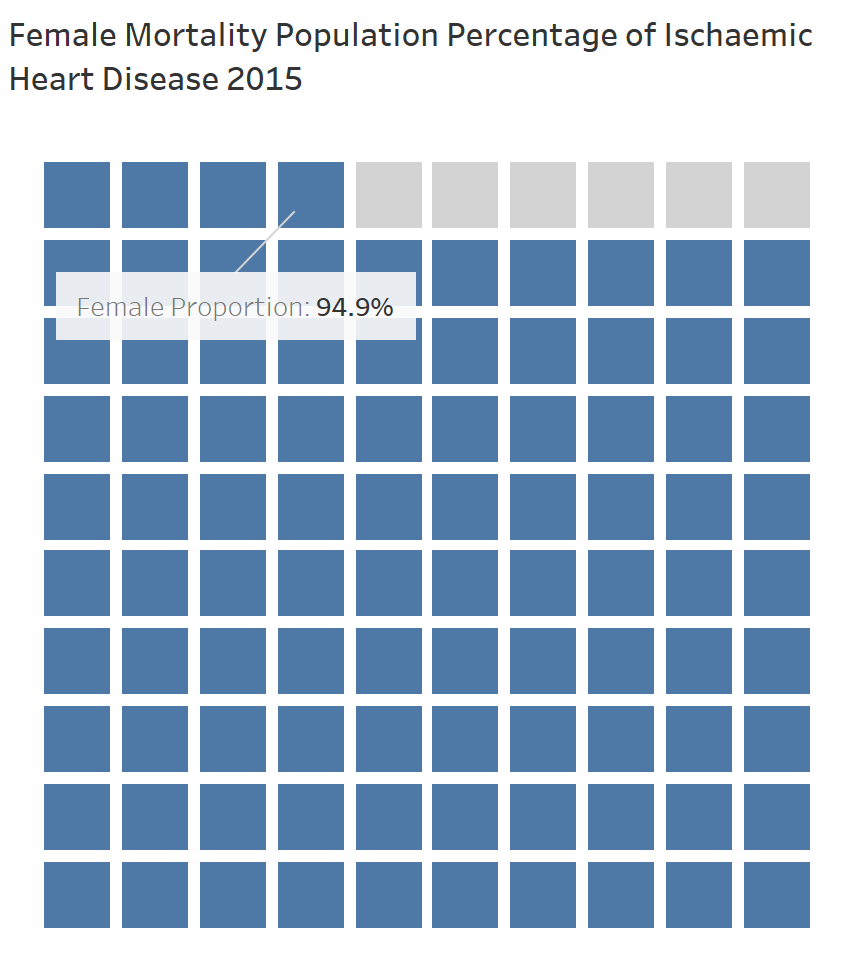
Description automatically generated

Figure 15. Compare Female Mortality Population Percentage

* The Figure 15 compares the female number of deaths as the percentage of total number of deaths in two diseases, Malignant Neoplasms of Trachea, Bronchus and Lung and Ischaemic Heart Disease.
* While both diseases remain as one of the highest causes of mortality in the dataset, female’s number of deaths in both diseases account high percentage. Female mortality population of Malignant Neoplasms of Trachea, Bronchus and Lung contributes 67.7% and Ischaemic Heart Disease contributes 94.9%.
* Since both disease play the significant roles, the study will include both diseases into the machine learning model.

Hayley (Jie) Teng: Asthma Mortality Predict Life Expectancy

A close up of a map

Description automatically generated

Figure 16. Asthma Mortality (# of deaths) Predicts Life Expectancy of Total Population at Birth (by Years)

* The Figure 16 use scatterplot and linear trend line to display the predictive analysis using Asthma Mortality which measured by number of deaths to predict the life expectancy of total population at birth.
* The downward sloping linear regression trend line, which generated by Tableau algorithm, suggested the negative relationship between Asthma Mortality and Life Expectancy of Total Population at birth.
* With the P-value lower than 0.0001 and R-square of 0.51, the predictive analysis implies that Asthma Mortality can explain around 51% of Life Expectancy with confidence level of 95%. The significant contribution of Asthma Mortality to Life Expectancy direct the study to include the independent variable into machine learning model.

Sravya Katta – Relationship between health status and education

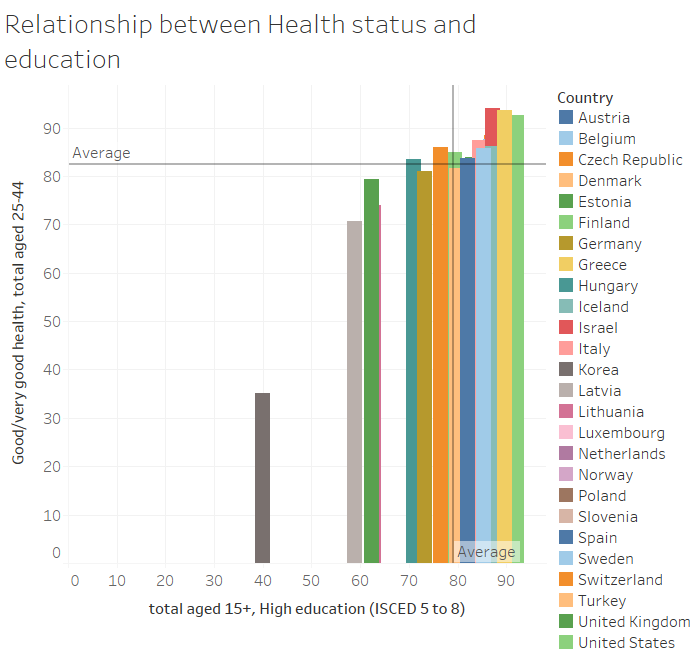


Figure 17. Dual Combination chart showing the association between average health status expectancy of 25 – 44 year olds with good health and education level globally

* The above figure is a bar chart that shows the trends Total aged 15+, High education (ISCED 5 to 8) vs. Good/very good health, total aged 25-44. Color shows details about Country.
* We can understand from the above chart is that health status expectancy for healthy 25-44 years old across the globe follows the education level pattern. As the education level for people aged 15+ remains stable over the years, we identify that health status also correlates with the education level with minor fluctuations.
* The implication here is that health status expectancy for healthy individuals follows the education level. Hence, Governments need to undertake additional efforts to improve the education level of individuals in order to see an improvement in their respective health status.

Victoria Cleveland – Education status of people in good health by country

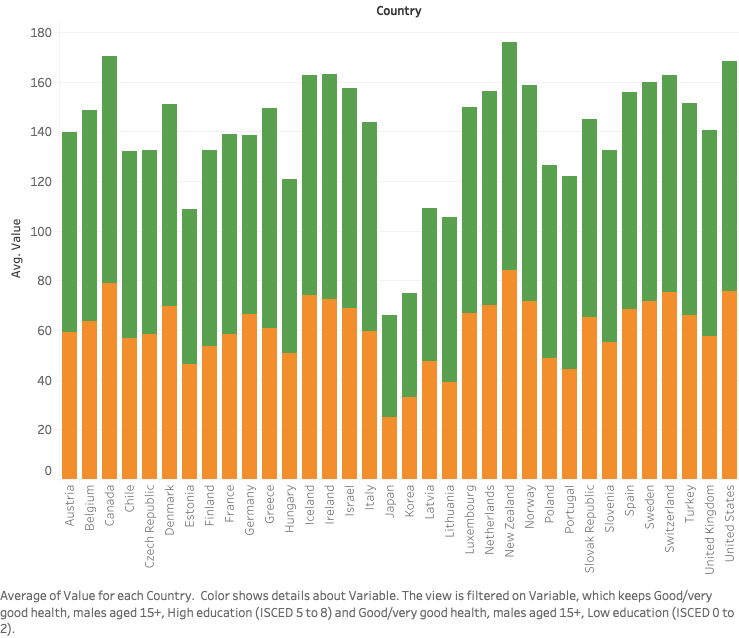


Figure 18. Bar chart of education status of people in good health by country

* This chart depicts the education status of those in good health in countries around the world.
* The insight here is that there are more people with high education in good health status, and less people with a low education in high health status.
* The implication here is that those who are more highly educated are more likely to be in a higher health status

Sravya Katta – Health status comparison across various countries and gender

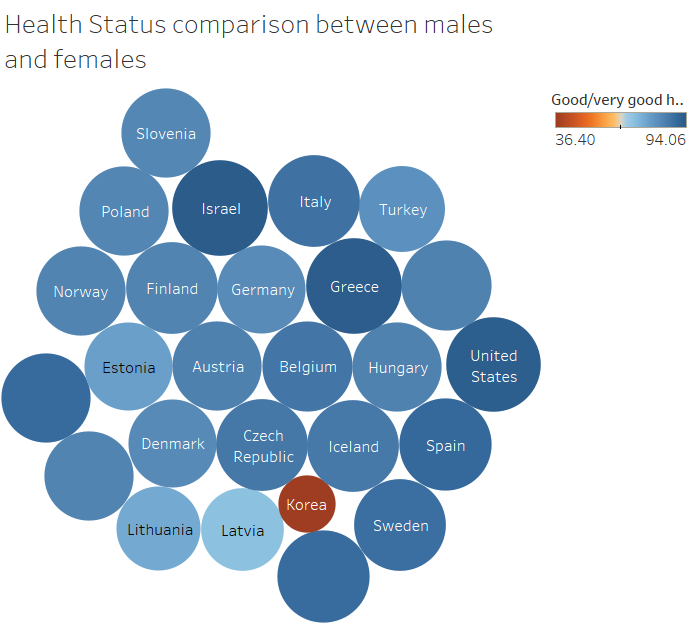


Figure 19. Bubble chart of comparison between male and female health status

* Figure is a bubble chart where Color shows Good/very good health, males aged 25-44. Size shows Good/very good health females aged 25-44. The marks are labeled by Country.
* We can infer from the above chart that although there is some difference between health status of males aged 25-44. But, females of similar age do not show much difference across several countries. The trend follows that in countries where males have good health, females also tend to have good health.
* Significant importance needs to be given to improving overall health across low performing countries such as Korea and Latvia. Understand the difference between health status in males and females to improve the citizens’ health.

Ila Srivastava – Female life expectancy expressed as % of population for various education status

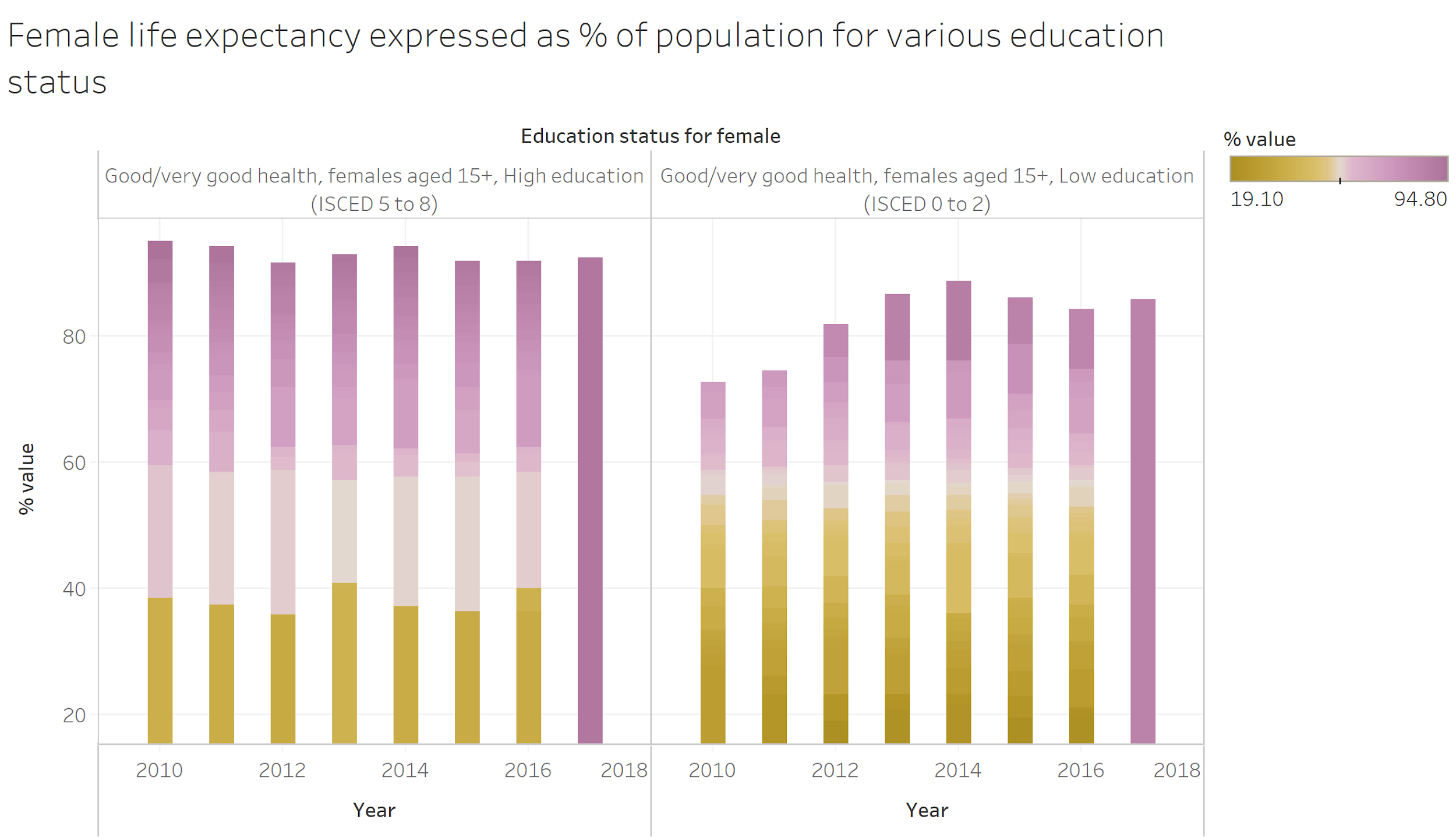
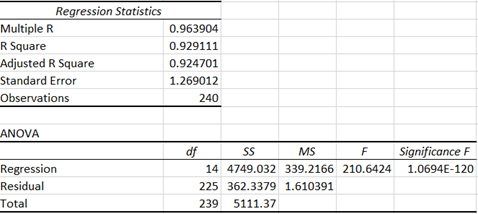


Figure 20. Bar chart depicting the female’s life expectancy as a percentage value for population for low and high level of education status

* This bar chart indicates life expectancy for a period of 7 years (2010-2017) for females based on high and low levels of education status for females aged 15+.
* The insight gained from this chart is that females that have low levels of education status have lower life expectancy. The yellow shaded bars indicate that low level of education is quite prevalent and is a concern for reducing life expectancy values. However, females with high levels of education have higher values for life expectancy and therefore indicates education status does impact life expectancy.
* The implication here is that over the government need to improve and incorporate good policies for improving education status in a region as it directly impacts the life expectancy.

H2: Number of deaths caused by different diseases has a negative correlation with life expectancy

Statistical Model: Multiple Linear Regression



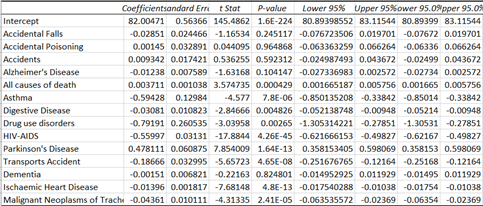


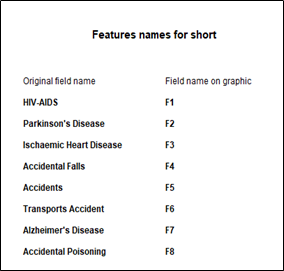
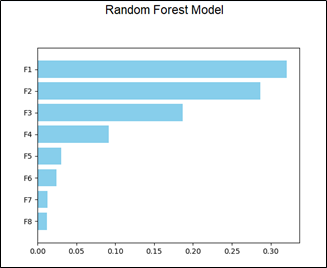
Figure 21. Multiple Linear Regression Model representing the correlation between Causes of mortality and Life expectancy

* This is multiple linear regression model representing the correlation between mortality rate caused by different diseases and Life expectancy.

* Adjusted R squared is 92.47% which indicates that change in Life Expectancy can be explained by the dependent variables to a large extent. Standard error is low at 1.26. Null hypothesis can be rejected as P-value is lower than 0.05 indicating that the model is statistically significant. Variables such as Asthma, drug use disorders, HIV, Parkinson’s disease are predictors of importance. High P-values (> 0.05) associated with the T statistic is an indication that independent variables related to accidents, Alzheimer’s, dementia may not be needed in the regression model. As coefficients are negative for most of the variables, this satisfies our hypothesis that number of deaths caused by different diseases has a negative correlation with life expectancy

* Particular attention needs to be given by Government and health administrators to reducing mortality due to drug use disorder, HIV-AIDS, asthma as this impacts the life expectancy of the population. Rehabilitation, prevention and promotional activities needs to be undertaken to tackle these issues.

Machine Learning Model: Random Forest



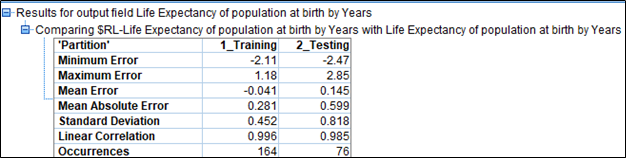


Figure 22. Predictor importance of features and analysis node generated by Random forest model

* The predictor chart displays the importance for each of the predictors contributing to higher or lower life expectancy. Analysis node output displays the life expectancy of population for a country and how well was it predicted by Random forest model.

* The insight gained from this chart is that that feature F1 and F2 are the most important predictors for predicting the life expectancy of a country. F1 and F2 signify HIV-AIDS and Parkinson’s disease respectively. Since this is a regression problem, we will look at Standard deviation and Linear correlation values for the testing dataset. Standard deviation is 0.818 which signifies that model is a good fit as the value is less and correlation value is quite high which again indicates model is good fit and significant.
* The implication here is that over the government need to specifically incorporate policies to cure and prevent diseases like HIV-AIDS, Parkinson’s disease and Ischaemic Heart disease and understand causes and add medical treatments to eradicate these diseases.