Many More Happy Returns! - Life Expectancy and You

CS5346 Group Project

Smitha Sheshadri [redacted], Morath Bastian Dominik [redacted],
Muhammad Huzaifah bin Md Shahrin [redacted], Debjyoti Ghosh [redacted]

April 17, 2019

1 Introduction

"Life Expectancy" (LE) is defined by the World Health Organisation as the average number of years that a newborn is expected to live if current mortality rates continue to apply [1]. Life, as it relates to all beings, is protected by innate behaviors like the 'fight or flight' response that seeks to preserve life in its best possible form, no matter the circumstances. Human beings are no exception; history holds witness to a constant strive for tenable sustenance. Understandably, even mythical creatures with inordinately long lifespans are revered and sought after in literature and popular culture — do Dracula or Thor ring a bell? While immortality continues to be confined to the realms of science fictions, we have, over the years, tried to unlock the secrets to a long and healthy life.

LE serves as one of the major indicators of a population's health and is of interest to policymakers and practitioners alike. In that respect, LE is often used in conjunction with economic measures of prosperity such as *GDP per capita* to indicate the general progress and well-being of a country. Thus, a thorough understanding of the factors that regulate or are impacted by (changing) LE is key to taking major economic and socio-political decisions that in turn, shape a population's well-being and progress.

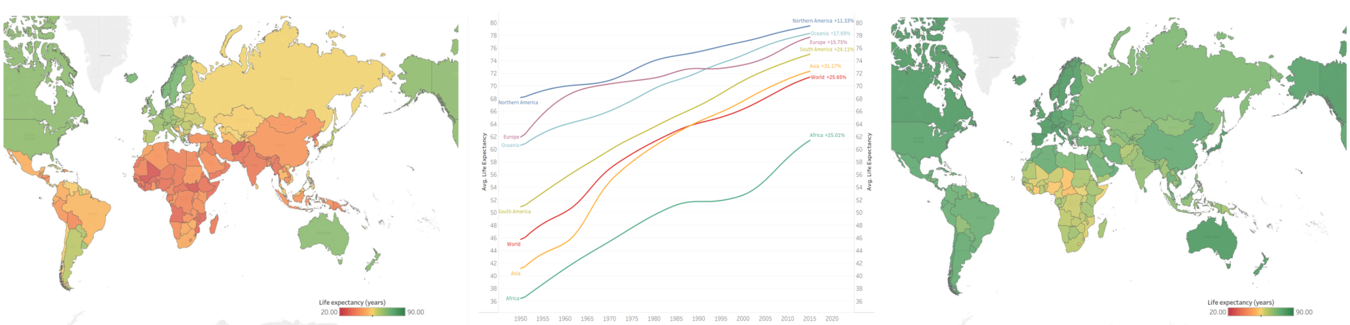


Figure 1: Trends in regional LE from 1950-2015. The left choropleth map shows the LE in 1950 while the right one shows the same in 2015. The general improvement in LE is evident with the shift towards the green colors. The centre line chart provides a breakdown of LE according to regions, where the biggest improvements have come from Asia, Africa and South America.

The last 70 or so years has seen a tremendous improvement in LEs around the world, especially so in developing parts of Asia, Africa and South America, although admittedly Africa still lags behind the world average (see Figure 1). Despite the general increase across the board, the changes to LE are not homogeneous. There is significant variation between countries within continents as Japan, Singapore, and Hong Kong were all ranked in the global top 5 in 2017 while countries like Afghanistan are at the other extreme (see Figure 2).

Several past researches [2] and public-health organizations [3] have linked a number of factors to be governing LE. For the scope of our current work, we chose four of these factors based on their popularity and widespread adoption: Obesity [4, 2], Smoking and Substance Abuse [5, 6], Dietary Habits [2], and Work and Employment [7, 8]. The factors are well distributed over three dimensions that we believe, determines the longevity of a human life: lineage (genetic or hereditary factors), lifestyle choices (dietary choices, addiction, work, etc.) and support (medical, economic, social, and political).

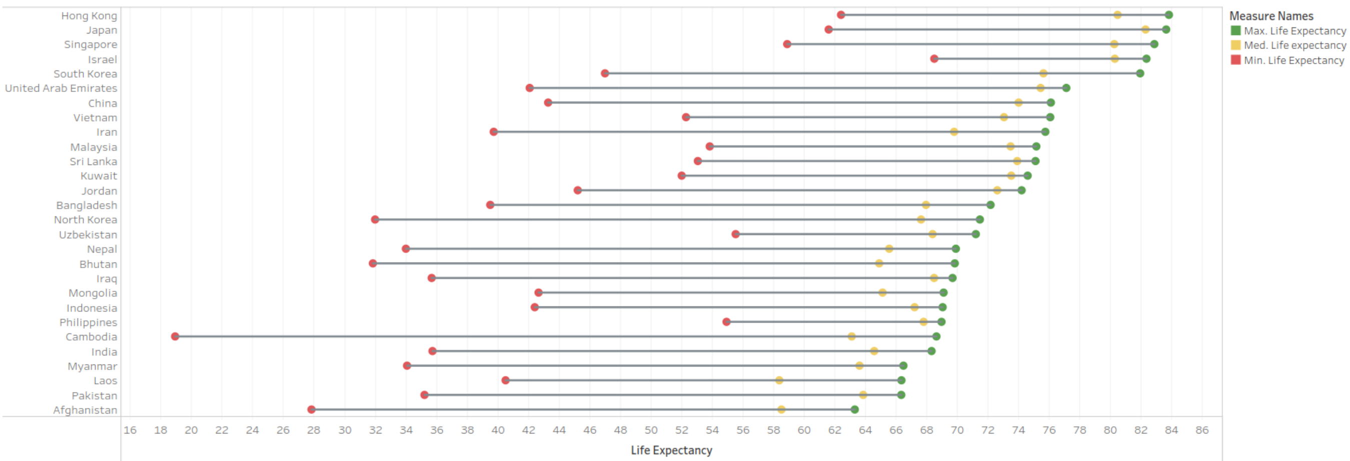


Figure 2: A dumbell chart of LE showing the maximum, minimum and median LE for various Asian countries since 1950. Countries from the same continent/regions can have widely varying LEs.

While establishing causal relationships between the chosen factors and LE was beyond our scope, we explore correlations and global trends in how these factors regulate or are otherwise impacted by LE. With a diverse dataset, drawn from over twelve distinct sources, the data collates information from more than 180 countries and spans more than five decades. With a set of diverse visualizations, we distill awareness on healthier living and emerging social trends, which in turn creates scope for contemplation about an increased vitality across the globe.

1.1 Authors' Contributions

All the authors have contributed equally towards the fulfillment of this report. Table 1 provides a breakdown of each author's individual contributions.

Name	**Contribution**
Smitha	Intro, Datasets, and Sec. 3: Obesity (p. 7)
Debjyoti	Intro, Datasets, Sec. 4: Vices (p. 13), and Sec. 7.1 (p. 31)
Bastian	Intro, Datasets, Sec. 5: Diet (p. 20), and Sec. 7.2 (p. 32)
Huzaifah	Intro, Intro Visualizations, Datasets, and Sec. 6: Employment (p. 26)

Table 1: Breakdown of Authors' Contributions

2 Datasets

The data has been drawn from diverse sources for each of our 4 chosen factors, LE and per capita GDP. For most datasets, the data spans over five decades of information and collates country-level data for over 180 countries. In this section, we detail the datasets we have used and the respective sources.

2.1 Life Expectancy

Life Expectancy

- Dataset: Life Expectancy
- Source: https://ourworldindata.org/life-expectancy
- File: *life_expectancy.csv*
- No. of countries: 209
- Data Collection Period: 1950 - 2015 (A few also have earlier years, which were not used in our analysis)
- Pre-processing: None

2.2 Obesity

Obesity

- Dataset: Years of life lost due to Obesity Related Diseases
- Source: https://www.ncbi.nlm.nih.gov/pubmed/12517229
- File: *YLL.csv*
- Data Collection Period: 1971 - 2003
- Pre-processing: None

Fitness

- Dataset: Positive impact of different intensities of physical activity
- Source: https://www-sciencedirect-com.libproxy1.nus.edu.sg/science/article/pii/S0735109714072489
- File: *MET mortality.csv*
- Data Collection Period: 2000 - 2015
- Pre-processing: None

BMI

- Dataset: Mean BMI of the world
- Source: https://www.who.int/gho/ncd/risk_factors/physical_activity_text/en/
- File: *Mean BMI in 2016.csv*
- Data Collection Period: 1975, 2016
- Pre-processing: None

2.3 Smoking and Substance Abuse

Smoking

- Dataset: Daily Smoking Estimated Prevalence by Country
- Source: https://ourworldindata.org/smoking
- File: *daily-smoking-prevalence-bounds.csv*
- No. of countries: 182
- Data Collection Period: 1980 - 2012
- Pre-processing: None

Alcohol and Substance Abuse

- Dataset: Alcohol and Drug Use Disorder Prevalence by Country
- Source: https://ourworldindata.org/substance-use
- File: *share-with-alcohol-or-drug-use-disorders.csv*
- No. of countries: 182
- Data Collection Period: 1990 - 2016
- Pre-processing: None

2.4 Dietary Habits

Food Index

- Dataset: Food Index
- Source: https://www.oxfam.org.uk/what-we-do/good-enough-to-eat
- File: *food_index.xlsx*
- No. of countries: 125
- Data Collection Period: 2000- 2013 (Note that the statistics are from different years, but all the figures are the most recently available from global data sources. There is only one food index per country though)
- Pre-processing: We only needed the *Data*-Sheet of the .xlsx file and also inverted the scale.

Meat Consumption per country

- Dataset: Meat supply per person
- Source: https://ourworldindata.org/meat-and-seafood-production-consumption#per-capita-trends-in-meat-consumption
- File: *meat_consumption.csv*
- No. of countries: 216
- Data Collection Period: 1990 - 2016
- Pre-processing: None, but note that the dataset indicates the meat supply, so these figures do not correct for waste at the household/consumption level so may not directly reflect the quantity of food finally consumed by a given individual

2.5 Employment

Breakdown of economic sectors

- Dataset: Employment by sector – ILO modelled estimates
- Source: https://www.ilo.org/ilostat/
- File: *MBI_33_EN.xlsx*
- No. of countries/regions: 201
- Data Collection Period: 2000 - 2018, 2019-2020 modelled
- Pre-processing: None, only percentages over the years 2000-2015 were used

Mean weekly hours worked

- Dataset: Hours of work
- Source: https://www.ilo.org/ilostat/
- File: *MBI_8_EN.xlsx*
- No. of countries/regions: 113
- Data Collection Period: 2000 - 2015, note some countries have missing years
- Pre-processing: None

Retirement age

- Dataset: Average effective age of retirement in 1970-2017 in OECD countries
- Source: http://www.oecd.org/els/emp/average-effective-age-of-retirement.htm
- File: *Summary_1970 values.xls*
- No. of countries/regions: 47 OECD and partner countries, several regional blocs
- Data Collection Period: 1970 - 2017, note some countries have missing years
- Pre-processing: Extensive pre-prosessing was done on this dataset. Firstly gaps and irrelevant values were deleted directly in the excel sheet, and column headers added in. Then, using Tableau Prep Builder, the columns for individual years were pivoted and their attributes were standardized to that of a whole number. Thereafter, the new table was saved as a csv file. The whole process was carried out on each of the mens' and womens' datasets, which were on different tabs in the original excel file. An averaging calculation was done in Tableau to then aggregate the values from the two datasets.

2.6 GDP Per Capita

GDP per capita

- Dataset: GDP per capita (Current US$)
- Source: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD
- File: *gdp_per_capita.csv*
- No. of countries: 251
- Data Collection Period: 1960 - 2017
- Pre-processing: The original file had a row for each country, showing the gdp per capita for each different year as a column. We wrote a *Python* script that creates one row for each {*Country/Year*}*-Pair*, showing the gdp per capita, thus reducing the number of columns to 3.

2.7 Continent Dataset

Country and Continent Codes List

- Dataset: Country and Continent Codes List
- Source: https://datahub.io/JohnSnowLabs/country-and-continent-codes-list
- File: *country-and-continent-codes-list-csv_csv.csv*
- No. of countries: 182
- Pre-processing: None

3 Obesity

Adult obesity has been shown to substantially heighten the risk of adverse health outcomes[9] [10]. Several chronic diseases have been associated with obesity, including Type 2 diabetes, cardiovascular disease, and cancer, to name a few. Death rates have been found to have a correlation to body mass index(BMI). The BMI is defined as the body mass divided by the square of the body height, and is universally expressed in units of kg/m^2. Studies have shown that death rates from cardiovascular disease were markedly elevated among individuals with higher BMI.

Table 2 enumerates the visualizations explored in this section.

3.1 Exploration-Visualization Mapping

#	*Exploration*	*Visualization*
E1	Years of life lost due to obesity	*Grouped bar chart*
E2	Mean BMI across the world	*Choropleth map*
E3	Most and least active countries	*Lollipop chart*
E4	Positive effects of different intensity physical activities	*Grouped bar chart*

Table 2: Visualizations for questions explored

E1: Years of life lost due to obesity

Figure 3: Years of life lost due Obesity related diseases

BMI category	BMI value in kg/m^2
Underweight	<18.5
Normal weight	18.5 - 25
Overweight	25 - 30
Class I obesity	30 - 35
Class II obesity	35 - 40
Class III obesity	>40

Table 3: BMI categories

Data	Data Type	Encoding	Note
Age group	Categorical	Position	Each age group is represented along the horizontal axis arranged in ascending order
Obesity category	Categorical	Color	Three different colors are used to represent the three BMI classes with highest BMI as shown in Table 1
Year of life lost	Quantitative	Height	Years of life lost for each age group/BMI category is represented by the height of the bar

Table 4: Visual Encoding for E1

Visualization Description
[11] observed that there is an optimum BMI going under or over which higher risk of mortality. They conducted a longitudinal study and found that the relative mortality risk is already increased at a BMI of 25 kg/m^2 and increases further with increasing BMI. [12] estimated the expected number of years of life lost (YLL) due to overweight and obesity across the life span of an adult using data from Data from the (1) US Life Tables (1999); (2) Third National Health and Nutrition Examination Survey (NHANES III; 1988-1994); and (3) First National Health and Nutrition Epidemiologic Follow-up Study (NHANES I and II; 1971-1992) and NHANES II Mortality Study (1976-1992) We chart and analyse the values of years of life lost as a grouped bar chart.

Insights
The visualization suggests that individuals in the youngest age group who fall under Class III obesity population have the most to lose. [13] suggests that obesity in the youth leads to a very high risk of high blood pressure and high cholesterol, which are risk factors for cardiovascular disease (CVD). [14] observes that obesity also increases the risk of breathing problems, such as asthma and sleep apnea.
These studies predict that youth who are obese are very likely to fall into the same or an even higher BMi category as they grow. These might give insights on why the youngest have the most to lose.

Further analysis on the distribution of BMI categories across the world would uncover the trajectory in BMI the world is following.

E2: Mean BMI across the world

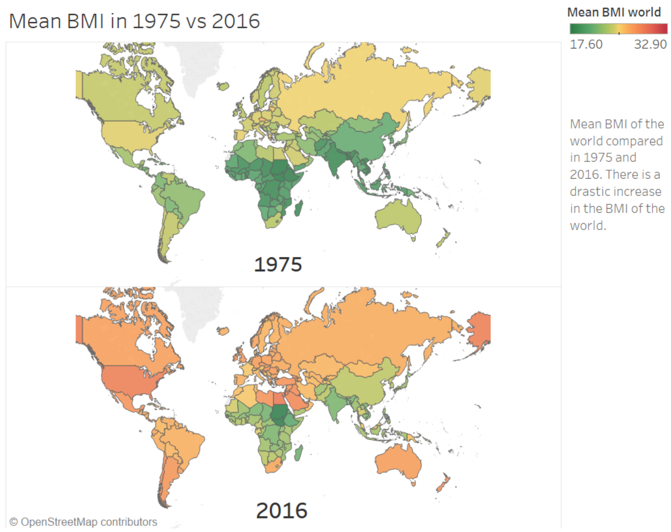


Figure 4: Mean BMI of the world in 1975 and 2016

The mean BMI of the world has changed dramatically over the years from 1975. Most of the countries were low BMI in 1975, on the other hand, in 2016, there is an overwhelming amount of red shades, indicating most regions have approached the high BMI zones. Given that this is the average BMI of the country, this indicates that the high BMI populations have majorly increased in size.

Data	Data Type	Encoding	Note
Country	Categorical	Position	Each country has a distinct position on the map based on its latitude and longitude
Average BMI	Quantitative	Color	A continuous-diverging color scale is used to lay down average BMI values as a choropleth on the map

Table 5: Visual Encoding for E2

Visualization Description

According to [1], there are three forms of malnutrition:

- Hunger and undernourishment
- Micronutrient deficiency
- Obesity or overnourishment

To explain the growing trends of Obesity, [15] reasons that over the past four decades, mean BMI and obesity in children and adolescents aged 5–19 years have increased in most regions and countries. This has lead to an overall increase in the BMI as the young population matures and continues to have a raised BMI. They report that over the 42 years of analysis, the global prevalence of obesity in children and adolescents increased from 0·8% in 1975 to 6.7% in 2016. They also report that obesity increased in every region, with proportional rise being smallest in

[1] https://ourworldindata.org/obesity

high-income regions and largest in Southern Africa (about 400% per decade).

Insights

[15] list a number of reasons for increase in average BMI across the globe and rapid growth of the obese population. In high-income countries, they state that there is a reluctance to use taxes and industry regulations to change eating and drinking behaviours. In developing countries, they reason that there is a lack of policies and programmes to make healthy foods such as whole grains and fresh fruits and vegetables more available and affordable to all. These might be some major contributors to the trend we see in visualization E2.

To get more insight into this trend we look at the most and least physically active countries of the world in 2016.

E3: Most and least active countries

The World Health Organization recognizes insufficient physical activity as one of the 10 leading risk factors for global mortality. They report that people who are insufficiently physically active have a 20% to 30% increased risk of all-cause mortality compared to those who engage in at least 150 minutes of moderate intensity physical activity per week. They give the following statistics in 2016.

- 23% of men aged 18+ years were insufficiently physically active
- 32% of women aged 18+ years were insufficiently physically active
- Over the past 15 years, levels of insufficient activity did not improve (28.5% in 2001; 27.5% in 2016)

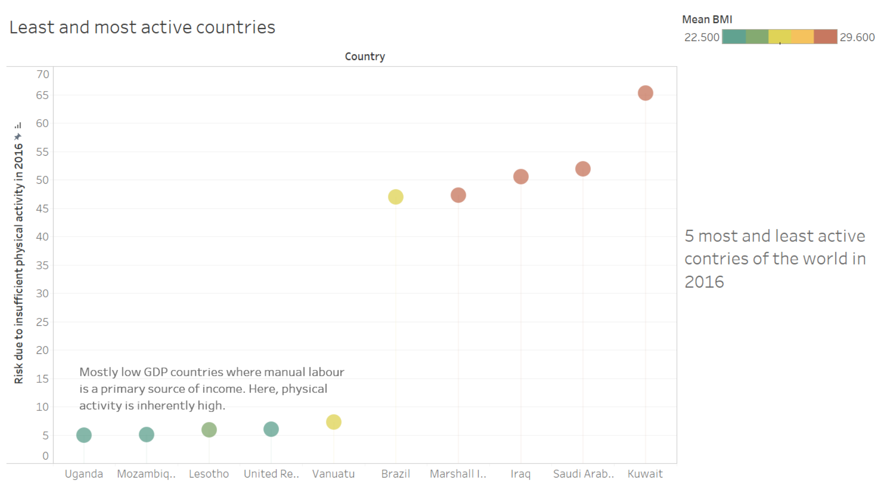


Figure 5: Most and least active countries

Visualization Description
WHO recognized that the prevalence of insufficient physical activity rose according to the level of income.

Insights

In high GDP nations, insufficient physical activity had more than double the prevalence compared to low GDP countries for both men and women. Analysing the high GDP nations, WHO explains high levels of insufficient physical activity by the transition towards more sedentary occupations. Decline in physical labour based occupations is undoubtedly a sign of economic growth, however, WHO recommends that countries dedicate time and effort into

Data	Data Type	Encoding	Note
Country	Categorical	Position	Each country has a distinct position on the chart represented by a circle arranged by rank along horizontal axis
Average BMI	Quantitative	Color	A continuous-diverging color scale is used to lay down average BMI values as a choropleth on the map
Insufficient physical activity prevalence	Quantitative	Position	The vertical axis scales the value of insufficient physical activity

Table 6: Visual Encoding for E3

designing infrastructure which encourages physical activity such as walking and cycling. Singapore has excelled as a country in this regard. The government established the National Health Promotion Board (HPB) in 2001. The organization initiated several workplace health programmes and physical activity programmes targeting weight moderation and reaching of recommended physical activity target. Further, Singapore has designed a Park Connector Network of walking/running/cycling paths that connects the various parks and other green spaces in Singapore, spanning $300km$.

We now look into the effects of different levels of physical activity on obesity related diseases.

E4: Positive effects of different intensity physical activities

Intensity refers to the rate at which the activity is being performed. Different levels of intensity have been found to have varying effects on physical condition of the performer. Physical activity intensity is measured in terms of Metabolic Equivalents(METs). METs are used to estimate the energy expenditure for many common physical activities. It is a unit that estimates the amount of energy used by the body during physical activity, as compared to resting metabolism. The unit is standardized so it can apply to people of varying body weight and compare different activities. At rest or sitting idly, the average person expends 1 MET for that time period.

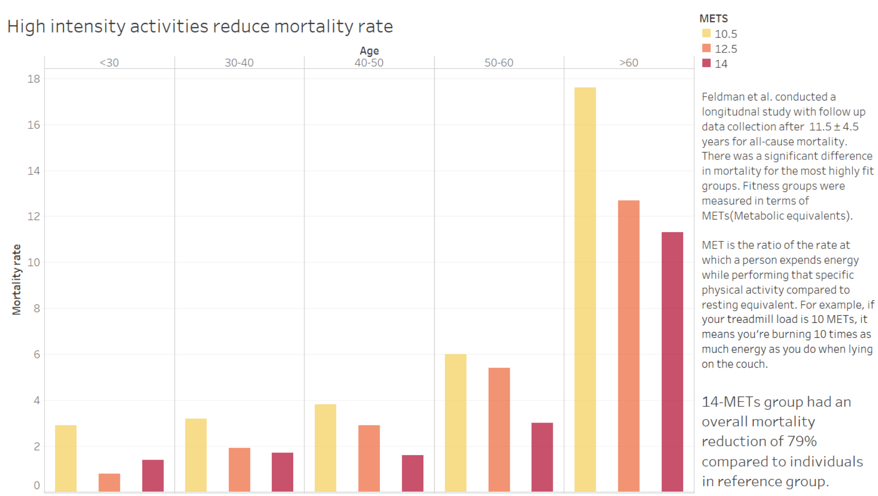


Figure 6: Positive effects of different intensity physical activities

Visualization Description

[4] conducted a study of nearly 70,000 patients. They followed the patients to 11.5 +- 4.5 years for all-cause mortality. They then calculated the crude mortality stratified by age and high-fitness groups (METs 10 to 11, 12 to 13, and >14). Their report shows a graded decrease in mortality across increasing fitness groups after the age of 30 years. They also report a mortality reduction of 51% in 12 to 13 MET groups and a reduction of 79% in >14-METs compared to reference group.

Data	**Data Type**	**Encoding**	**Note**
Age group	Categorical	Position	Each age group is represented along the horizontal axis arranged in ascending order
MET category	Categorical	Color	Three different colors are used to represent the three MET classes
Reduction of mortality rate	Quantitative	Height	Reduction of mortality rate for each group is represented by the height of the bar

Table 7: Visual Encoding for E4

Insights

One pound of fat contains approximately 3,500 calories of energy. Using this, it can be determined how long to perform a given activity to burn the equivalent of 1 pound of fat. For example, one would have to ride his bike at 12 mph, which burns 12.4 calories per minute, for 283 minutes to burn one pound of fat. [4] reports that there is not higher limit to the mortality benefit at high MET levels. This explains why the highest benefits are experienced by the highest MET group.

4 Smoking, Alcohol Consumption and Drug Abuse

There is general consensus even among habitual smokers, drinkers and drug addicts that their act falls under the rubric of '*vice*' — vice which precludes a healthy lifestyle and pre-empts a human life from running its full course. While several past researches put smoking and nicotine consumption as a leading cause of lung cancers and reduced mortality [16, 6], alcohol consumption has constantly been linked with higher risks of stroke, fatal aneurysm, heart failure and early death [5, 17, 18]. All of these evidences point towards a significant impact of the vices on human life expectancy.

In this section, we investigate the effect of SMOKING and Alcohol Consumption and Substance Use (referred to as just ALCOHOL from here on for the sake of brevity) on *Life Expectancy* (LE) across the globe. We explore 4 questions with exploratory temporal and spatial visualizations, and a correlation graph. Table 8 enumerates the questions and the corresponding visualizations used to explore them.

4.1 Exploration-Visualization Mapping

#	*Exploration*	*Visualization*
E1	Geographical distribution of SMOKING, ALCOHOL and LE	*Map Visualization*
E2	How did SMOKING, ALCOHOL-abuse patterns change over the years for individual countries and how did it affect LE?	*Bubble Chart*
E3	Continent-wise distribution of SMOKING, ALCOHOL and LE	*Butterfly Chart*
E4	SMOKING-LE correlation over the years	*Scatter Plot*

Table 8: Visualizations for questions explored

4.2 Visualizations

For every question explored, a screenshot of the visualization is followed by the visual encoding.

E1: Geographical distribution of SMOKING**,** ALCOHOL **and** LE

Data	**Data Type**	**Encoding**	**Note**
Country	Categorical	Position	Each country has a distinct position on the map based on its latitude and longitude
LE	Quantitative	Color, Saturation	A continuous-diverging color scale used to lay down LE values as a choropleth on the map
SMOKING %	Quantitative	Size	Each country represented by a circle with its size directly proportional to the avg. SMOKING %
ALCOHOL-related Disorder %	Quantitative	Color, Saturation	A continuous-diverging color scale used to fill the circle representing each country

Table 9: Visual Encoding for E1

Visualization Description
The viz. explores geographical distribution of LE, laid down as a choropleth on the world map. A continuous red-gold-green diverging color scale is used to map the colors for the choropleth. While lower LEs are represented by shades of red (representative of early deaths), higher LEs are denoted by shades of green (representative of healthier and longer life). To overlay average SMOKING % and ALCOHOL-related disorder % on top of the choropleth, we used a circular mark for each country. For each mark, the size attribute was used to map the SMOKING %, while

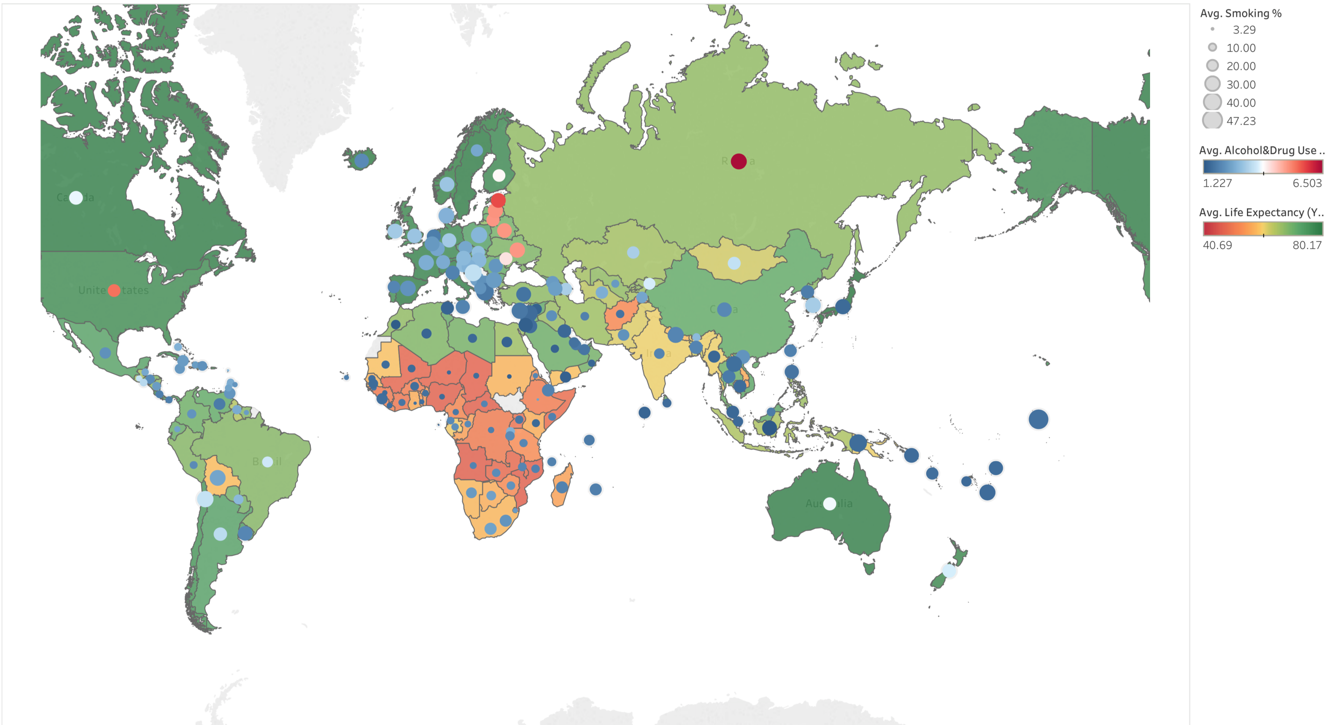


Figure 7: E1 Visualization (*Map Visualization*)

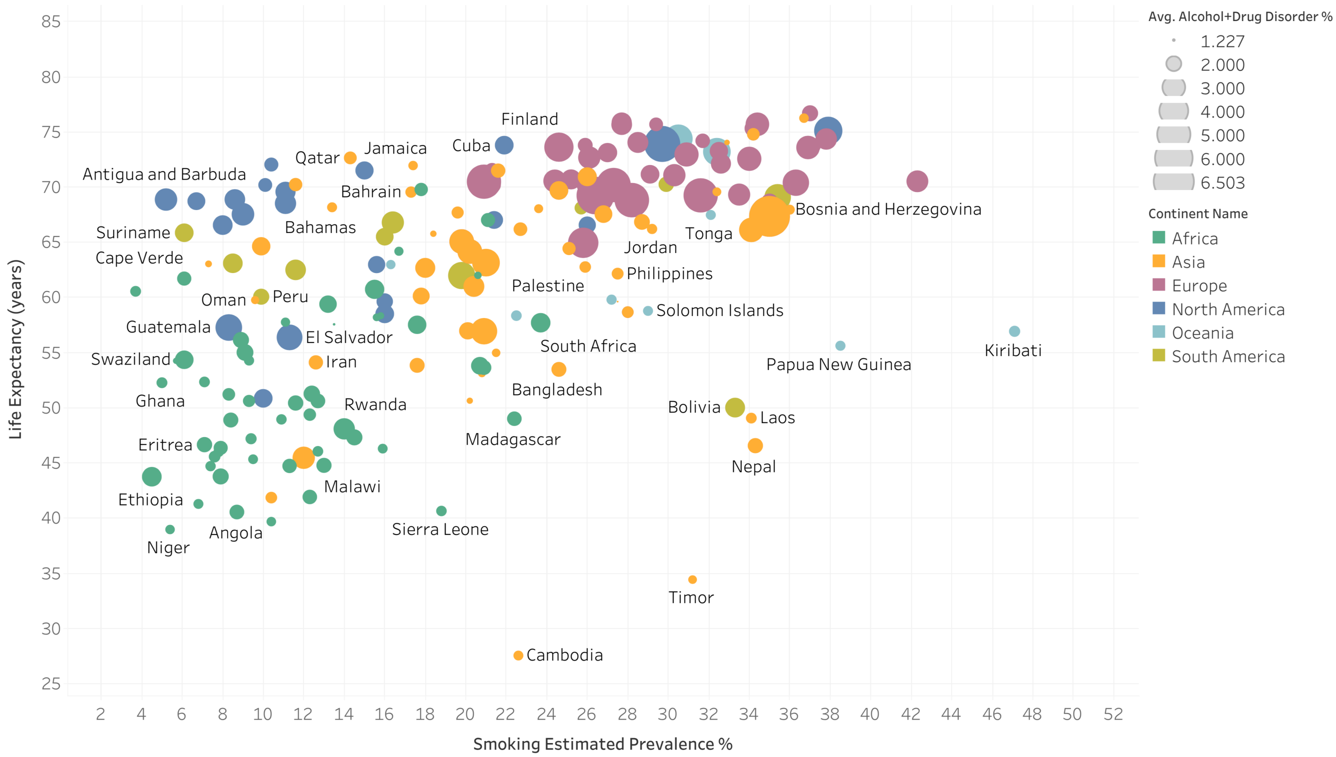
the fill color for each circular mark was drawn from a second diverging color scale (blue-white-red) to denote the ALCOHOL-related disorder %.

The quantitative measurements for SMOKING, ALCOHOL and LE were calculated as an average for individual countries between the years 1980 to 2012.

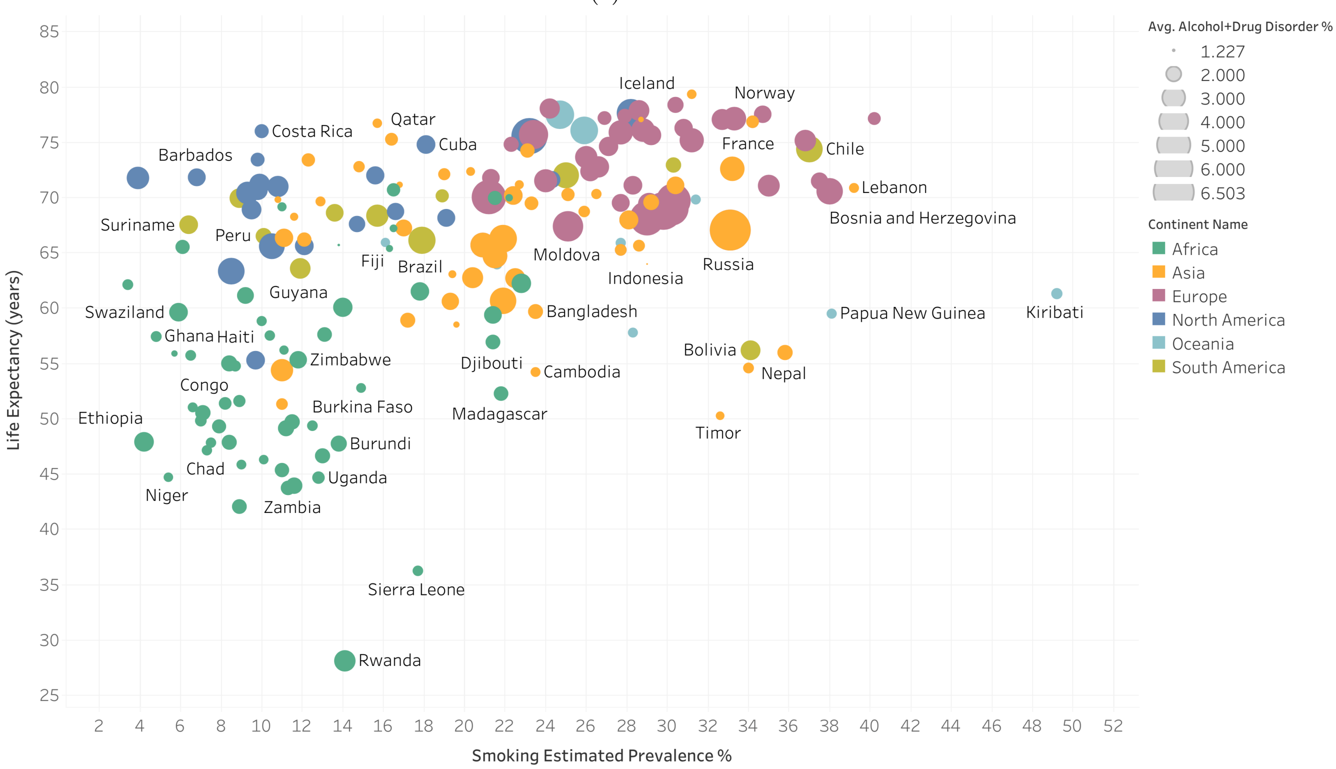
Insights
The visualization shows the global average life expectancy between 1980-2012. A quick glance immediately reveals distinct regional averages — most African countries have low LE, while countries like the USA, Canada, Australia, New Zealand and most of the Europian countries have a high average LE. For most Asian countries, the average LE hovers around the midway mark between the minimum and maximum country averages. The Muslim countries in Africa, in the mid-East and South-East Asia show a very low percentage of disorders resulting from alcohol consumption. This can be attributed to religious tenets that prohibits followers of Islam to drink. Russia suffers from both heavy smoking and very high percentages of alcohol and drug related disorders. In the US, alcohol and drug related disorders pose a greater threat than smoking does.

E2: Temporal patterns in SMOKING and ALCOHOL-abuse for individual countries

Visualization Description
Here, we explore temporal trends in the data between 1980-2012. This visualization was originally developed as an interactive animation that cycles through the years to show yearly trends. For a static representation, we show a snapshot of the data once every 10 years. Each country is represented by a bubble on a 2-D LE - SMOKING% grid. The bubbles are differentiated by colors based on the continent. The size of the bubbles are determined by the percentage of ALCOHOL-related disorders in individual countries. In the ideal scenario, a country gets represented by a small circle (low ALCOHOL%) at the upper left corner (high LE, low SMOKING%) of the 2-D plot.

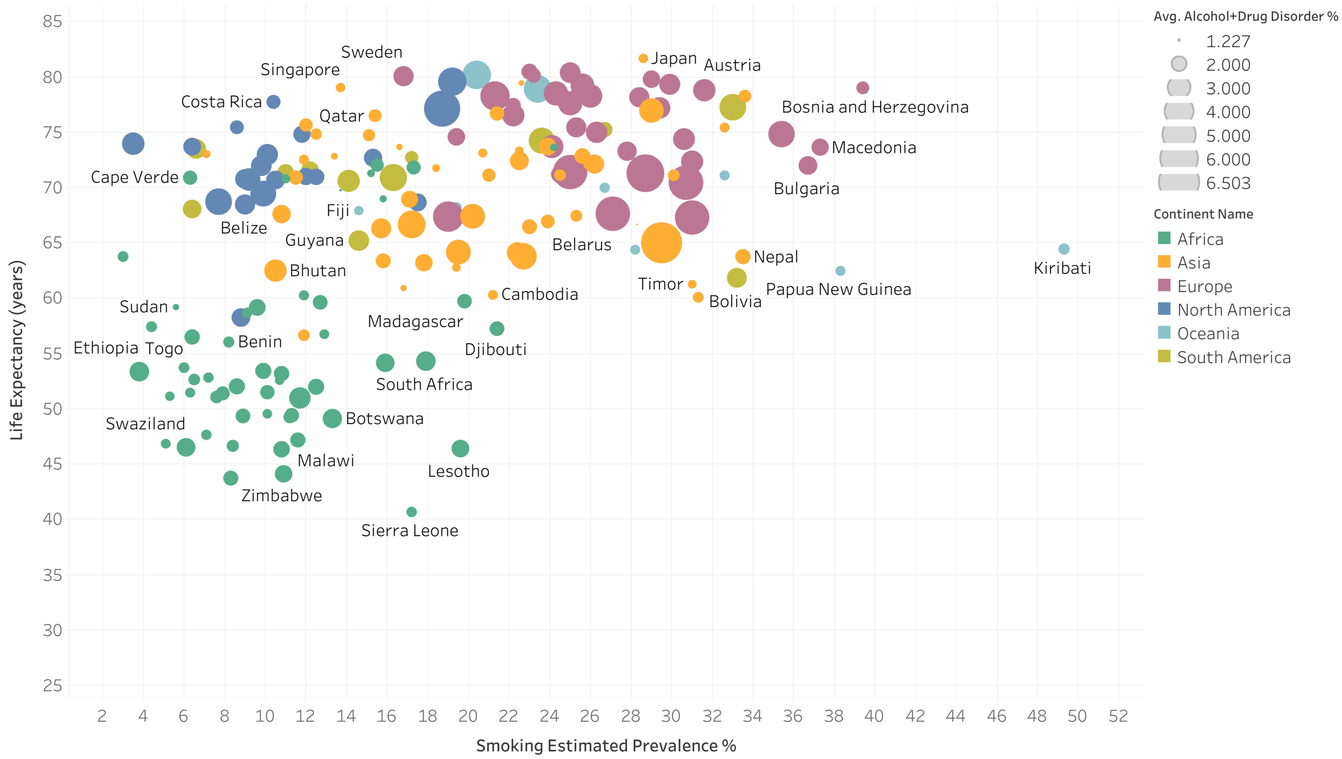


(a) Year: 1980

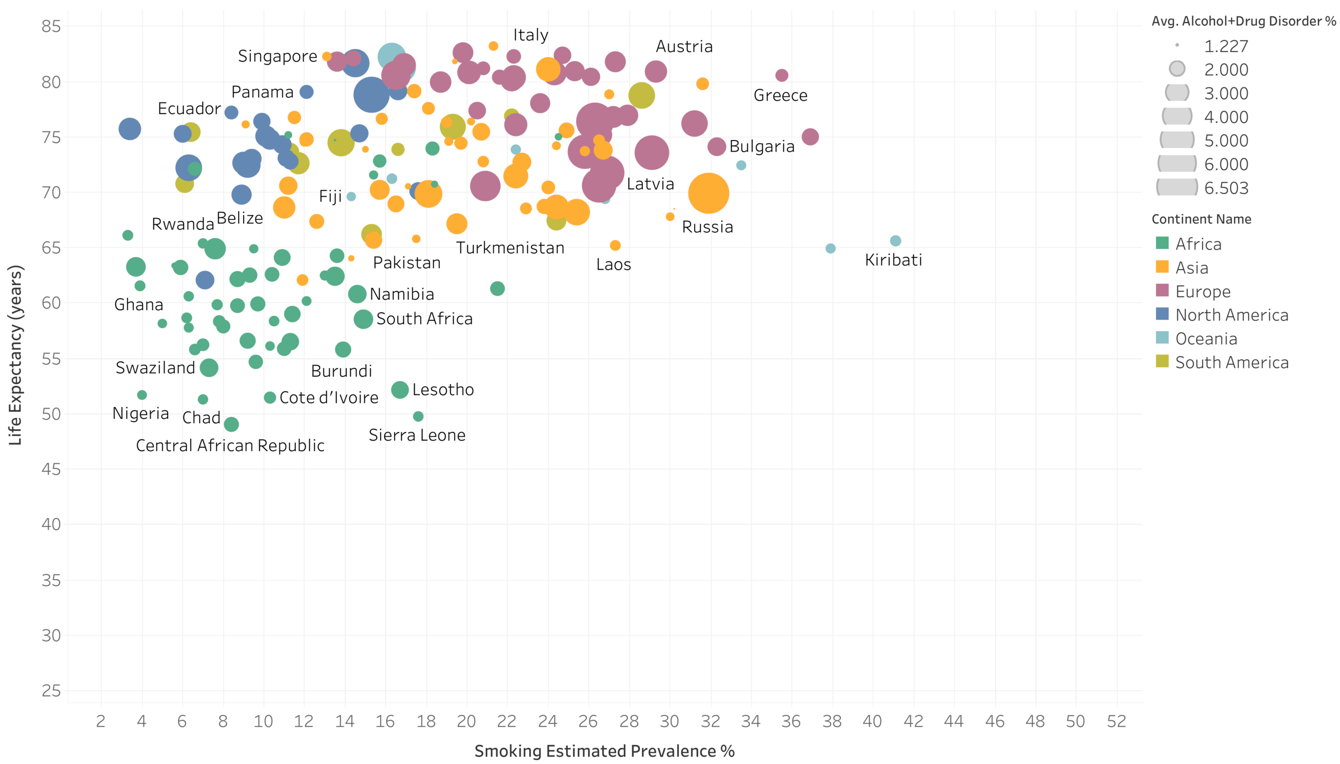


(b) Year: 1992

Figure 8: SMOKING and ALCOHOL-abuse disorders over the years and change in avg. LE (*Bubble Chart*)



(c) Year: 2002



(d) Year: 2012

Figure 8: Smoking and Alcohol-abuse disorders over the years and change in avg. LE (*Bubble Chart*)

Data	Data Type	Encoding	Note
Country	Categorical	Position	Each country gets mapped to a distinct grid position based on its x-coord (SMOKING%) and y-coord (LE)
Continent	Categorical	Color	Each continent is represented by a distinct color
LE	Quantitative	Position	LE along the y-axis
SMOKING%	Quantitative	Position	SMOKING% along the x-axis
ALCOHOL-related Disorder %	Quantitative	Size	Each country represented by a circle with the size directly proportional to the avg. disorder %

Table 10: Visual Encoding for E2

Insights
The visualization shows an encouraging global trend for LE and SMOKING% over the years. Progressing over the 4 subplots of figure 16, it is evident, that the countries tend to get more and more concentrated towards the upper left corner of the 2-D plot area. This means that the global average smoking rate goes down and the global average life expectancy increases. However, the size of the bubbles do not vary much over time, meaning that the average global disorder % due to alcohol and drug abuse does not change much over the years.

With a visualization of this kind, where most of the countries are trying to move towards an improved global norm with higher LE and lower SMOKING%, the datapoints tend to move together in clusters towards the global norm. This makes it easy to detect outliers and anomalies in the data that starkly deviates from the global norm. For example, in this visualization, Cambodia starts off as an outlier in 1980 (see figure 8a), but in 12 years time, by 1992, it has already moved up in the LE-axis towards the global norm (figure 8b. Another example is Rwanda which shows a sudden decline in LE between 1980-1992 (figure 8a and 8b). These anomalies may be explained by war or a sudden calamity that might have claimed numerous lives in a region over a given period of time.

E3: Continent-wise distribution of SMOKING, ALCOHOL-related disorders, and LE

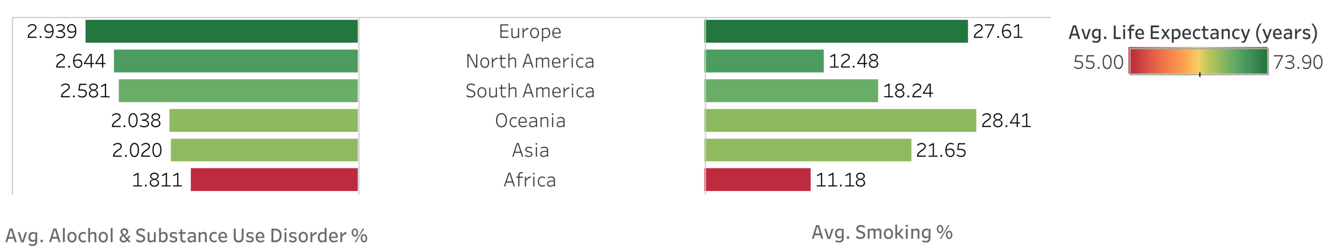


Figure 9: E3 Visualization (*Butterfly Chart*)

Data	Data Type	Encoding	Note
Continent Name	Categorical	Position	Position determined by the sorted order (*desc.*) of LE
LE	Quantitative	Color, Saturation	A continuous-diverging color scale used to map color values of bar-plots for each continent
SMOKING %	Quantitative	Length	Bar-plots along the positive x-axis
ALCOHOL-related Disorder %	Quantitative	Length	Bar-plots along the negative x-axis (absolute values are taken as lengths)

Table 11: Visual Encoding for E3

Visualization Description
We explore continent-wide averages for Smoking and Alcohol-related disorders for the 6 continents between 1980-2012. The visualization is laid out in the form of a butterfly chart. The continents are sorted from top to bottom in the decreasing order of their average LE. Bar-plots along the positive x-axis measures the avg. Smoking% for each continent, while bar-plots along the negative y-axis measure their avg. Alcohol-related disorder%. The bar-plots are colored based on their avg. LE. Consistent with all viz. in our project, a diverging red-gold-green diverging color scale is used to map the values for LE.

Insights
Ideally, one would expect, less smoking and less alcohol-and-drug abuse would contribute to higher mortality and thus, increased life expectancy in the continents. However, our visualization shows it's clearly not what's happening, at least within the scope of the available data. Africa, for example, has much lower LE as compared to Europe or North America. However, both smoking and alochol-and-drug abuse is more prevalent across the latter continents than in Africa. Again, if we analyze between Europe and North America, the average LE of Europe is higher despite higher avg. Smoking% and higher avg. Alcohol-related disorder%. Clearly, there are other factors in play that dictate a country's avg. LE. As we have explored in our project, a country's GDP might be an important factor in determining the average LE.

E4: Smoking% - LE Correlation

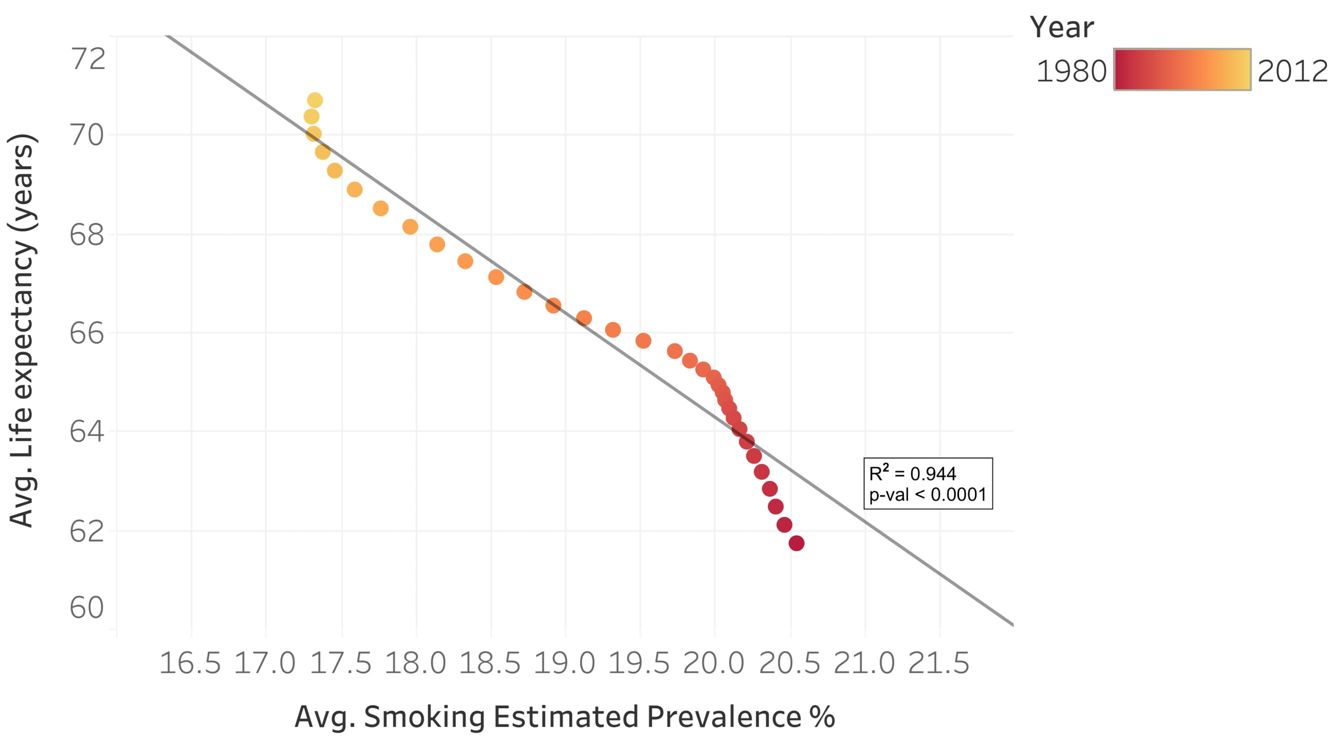


Figure 10: E4 Visualization (*Correlation Plot*)

Visualization Description
This visualization explores a correlation between the global mean of Smoking% and the global mean LE for years between 1980-2012. Each year is a plotted on a 2-D chart area with x-coord as the global mean Smoking% and y-coord as the global mean LE for that year. Note that neither axis starts at zero to make a more effective use of the chart area. A red-gold diverging color scheme differentiates the datapoints for each year — recent years are mapped to lighter shades of gold while past years are represented by darker shades of red. Finally, we draw a trend

Data	Data Type	Encoding	Note
Year	Ordinal	Color, Saturation	A continuous diverging color scale is used; darker shades are used to represent past history
Avg. LE	Quantitative	Position	LE along the y-axis (starts from 60)
Avg. SMOKING%	Quantitative	Position	SMOKING% along the x-axis (starts from 16)

Table 12: Visual Encoding for E4

line to study the correlation between the two variables. The correlation co-efficient (R^2 value) and the corresponding p-value is labeled on the trend line.

Insights
We find a very strong negative correlation ($R^2 = 0.944, p < 0.0001$) between the mean SMOKING% and the mean LE for data between 1980-2012. This shows that the drop in SMOKING% over the years correlates highly with the rise in LE during the same period. Of course, from the data we can at best show a correlation between the two variables and not a causation. However, the fact that the correlation is so strong corroborates to earlier studies that links smoking to increased health-risks and a reduced lifespan.

Additionally, the variable steepness of the curve shows the differential rate at which the SMOKING% dropped over the years. Between 1988-1995, the global average did not drop much, probably pointing towards reduced awareness and that smoking was still in fashion (reminiscent of the 90's hip-hop culture?). However, post 1995, smoking quickly lost traction, probably with the rise of the millennial culture and the internet at the helm of spreading global awareness. Also, the emerging digital era might have converted a sizeable percentage of the global youth to an alternate form of addiction, dictated by technology.

5 Diet

Japanese people are among the longest-living people in the world. This might be due to their healthy diet. The Japanese government proposed a food guide in 2005 and ten years later, researchers [19] in Tokyo investigated how following the food guide affected the country's mortality rate. They found out that people following the guide had a 15 per cent lower mortality rate, suggesting that a healthier diet indeed has a big impact on life expectancy.

In this section, we are going to investigate the importance of a good diet on life-expectancy. For this we use the *Food Index*[2], a measure created by *Oxfam*, a confederation of 20 charitable organizations focusing on the alleviation of global poverty. The Food Index of a country is a score that encapsulates how good the food is. It is a measure consisting of the availability of food (data about undernourishment and number of underweight children), affordability of food (data about price level and price inflation volatility), food quality (data about nutritional diversity and access to safe water) and healthy diet (data about obesity and diabetes). The statistics are from different years, but all the figures are the most recently available from global data sources, such as the World Health Organisation. Each of the sources used different scales in measuring the countries, requiring a process to standardise them so that they could be compared. The standard MIN / MAX rescaling method was used, generating re-scaled values of 0-100 where 0 points is the minimum score (worst) and 100 points is the maximum score (best). The process is based on identifying the countries with the minimum and maximum scores in the original data, scoring them 0 and 100 respectively and then measuring how far every other country is from these maximum and minimum values. The dataset includes 125 countries.

Also, we investigate meat consumption per person to find insights about factors that influence life-expectancy. We have data about how many kilogram per capita of meat was supplied each year from 1961 to 2014 for 216 countries and regions. Note that the dataset indicates the meat supply, so these figures do not correct for waste at the household/consumption level so may not directly reflect the quantity of food finally consumed by a given individual.

Table 13 enumerates the visualizations explored in this section.

5.1 Exploration-Visualization Mapping

#	*Exploration*	*Visualization*
E1	Correlation of Food Index and life expectancy, 2010	*Scatter Plot*
E2	Correlation of Food Index measures and life expectancy, 2010	*Scatter Plot*
E3	Food index breakdown per continent	*Parallel Coordinate Charts*
E4	Correlation of meat consumption and life expectancy, 2010	*Bubble Chart*

Table 13: Visualizations for questions explored

5.2 Visualizations

For every question explored, a screenshot of the visualization is followed by the visual encoding.

E1: Correlation of Food Index and life expectancy, 2010

Visualization Description
This visualization shows the correlation between life expectancy and the food index score of each country in 2010. Colors were used to differentiate among the six continents. Life expectancy data was taken from the year 2010. We

[2]oxfam.com. (2019). Global Food Index :: Oxfam [online] Available at: https://www.oxfam.org.uk/what-we-do/good-enough-to-eat [Accessed 16 Apr. 2019].

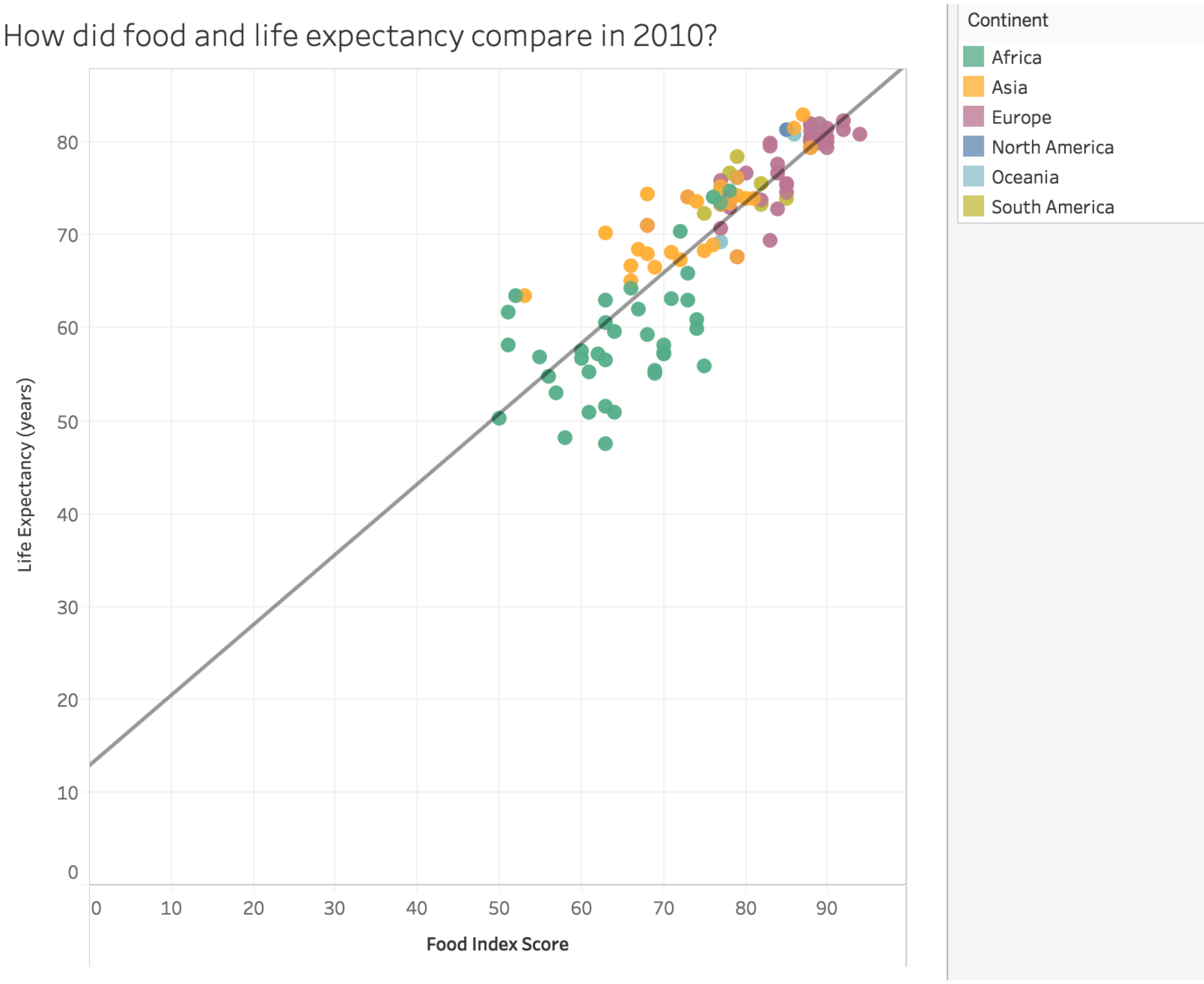


Figure 11: E1 Visualization (*Scatter Plot*)

Data	**Data Type**	**Encoding**	**Note**
Country	Categorical	Position	Each country gets mapped to a distinct grid position based on its x-coord (Food Index Score) and y-coord (LE)
Life Expectancy	Quantitative	Position	LE along the y-axis
Food Index Score	Quantitative	Position	Food Index Score along the x-axis ([0, 100])
Continent	Categorical	Color	Each continent is represented by a distinct color

Table 14: Visual Encoding for E1

drew a linear trend line to study the correlation between the two variables. The correlation coefficient (R^2 value) and the corresponding p-value is labeled on the trend line.

Insights
We find a strong positive correlation ($R^2 = 0.739, p < 0.0001$) between the food index score and the life expectancy in 2010. The fact that the correlation is strong supports studies (such as the one mention earlier about the Japanese food guide) that link a healthier diet to an increased life expectancy.

We also find that that countries in more developed parts of the world, such as Europe, in general have a higher food index than in less developed, mostly African countries. Even within a continent, a higher food index usually comes with a higher life expectancy, as seen on the viz. To investigate this further, in the next section we look at the correlation of life expectancy and the four different factors that contribute to the food index.

E2: Correlation of Food Index measures and life expectancy, 2010

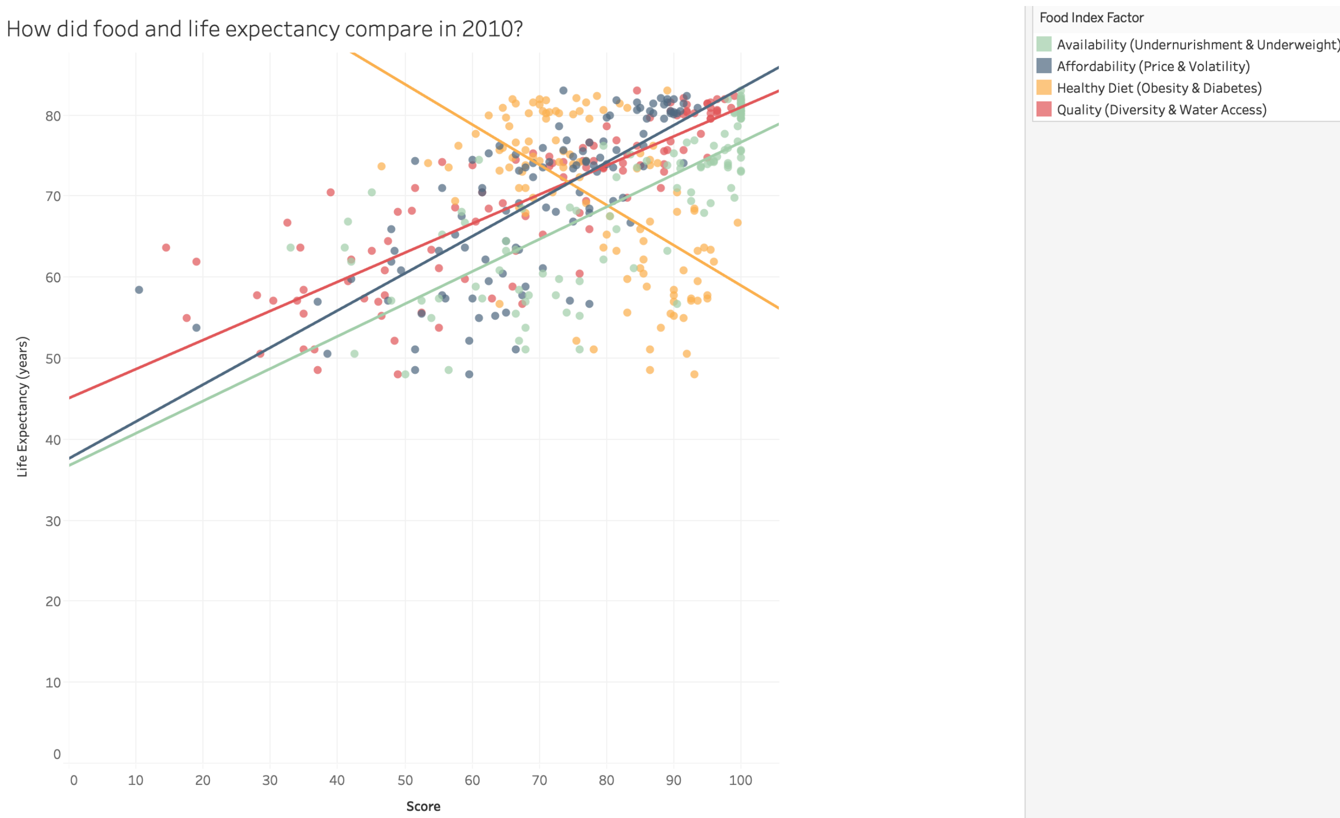


Figure 12: E2 Visualization (*Scatter Plot*)

Data	Data Type	Encoding	Note
Country	Categorical	Position	Each country gets mapped to a distinct grid position based on its x-coord (Food Index Score) and y-coord (LE)
Life Expectancy	Quantitative	Position	LE along the y-axis
Food Index Measure	Categorical	Color	Each of the four Food Index measures is represented by a distinct color
Score	Quantitative	Position	Food Index Score along the x-axis

Table 15: Visual Encoding for E2

Visualization Description
This visualization shows the correlation between life expectancy and the four different Food Index measures of each country. Colors are used to differentiate among the four measures. Life expectancy is taken from the year 2010. Finally, we draw a linear trend line for each measure to study the correlation between the variables. The correlation coefficient (R^2 value) and the corresponding p-value is labeled on each trend line.

Insights
We find positive correlations between life expectancy and quality ($R^2 = 0.712, p < 0.0001$), availability ($R^2 =$

0.595, $p < 0.0001$) and affordability ($R^2 = 0.547, p < 0.0001$) of food, respectively.

We also find a negative correlation ($R^2 = 0.345, p < 0.0001$) between life expectancy and the dietary risk factors of a country. Although this seemingly indicates that life expectancy increases with increasing prevalence of obesity and diabetes, this pattern is probably rather attributable to the higher GDP per capita in these countries. Better health care is provided for in countries with higher GDP. This compensates for the number of years lost due to obesity (see section 3.1), despite the high consumption of (unhealthy) food, which can cause diabetes and obesity [20].

E3: Food Index breakdown per continent

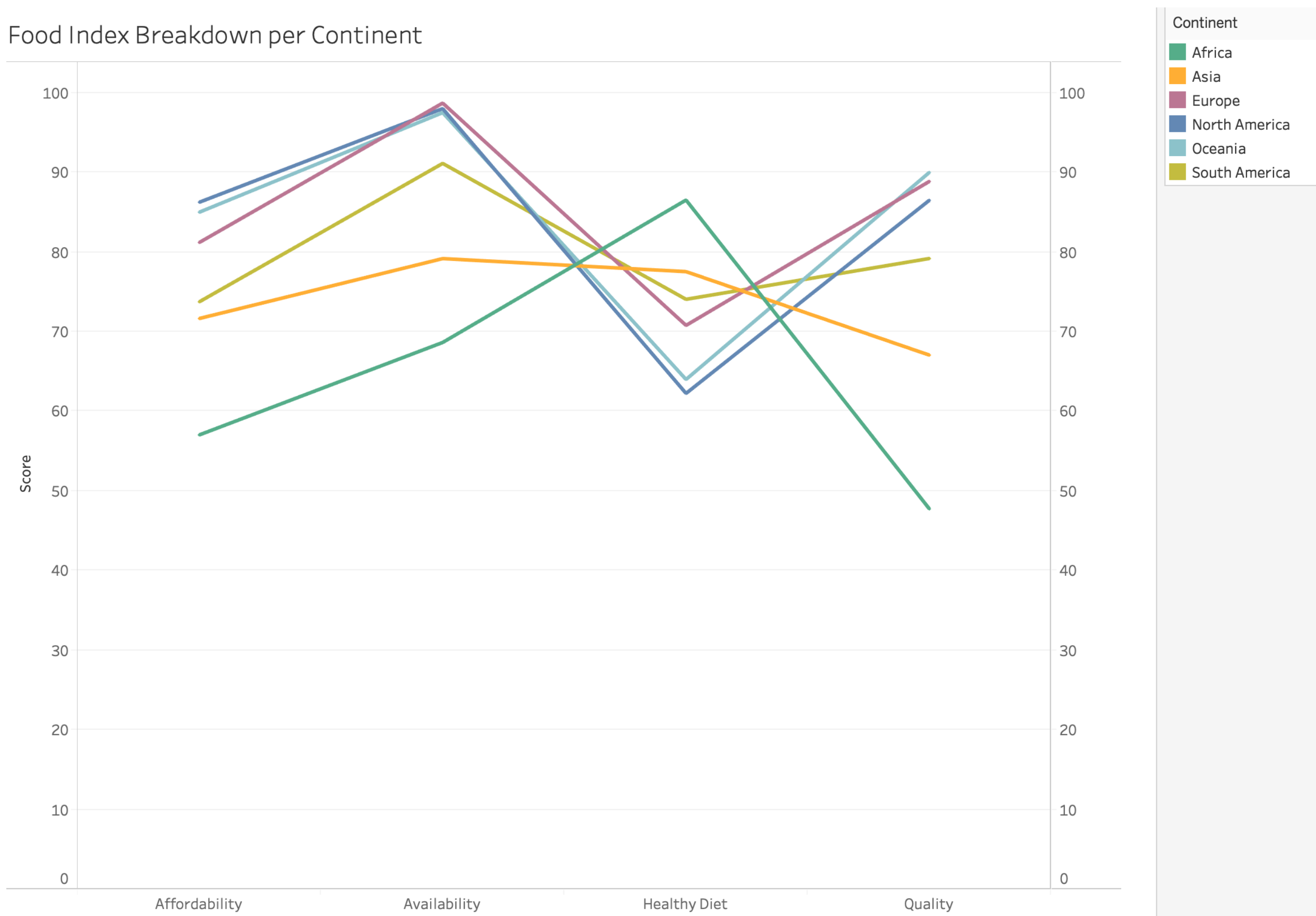


Figure 13: E3 Visualization (*Parallel Coordinate Charts*)

Data	**Data Type**	**Encoding**	**Note**
Continent	Categorical	Color	Each continent is represented by a distinct color. The average of a measure over all countries of a continent is taken
Food Index Affordability	Quantitative	Position	Encoded as the first y-coordinate
Food Index Availability	Quantitative	Position	Encoded as the second y-coordinate
Food Index Healthy Diet	Quantitative	Position	Encoded as the third y-coordinate
Food Index Quality	Quantitative	Position	Encoded as the fourth y-coordinate

Table 16: Visual Encoding for E3

Visualization Description
This parallel coordinate chart shows the average score for the four different Food Index measures for each continent. Colors are used to differentiate among the continents.

Insights
We observe that the gap between the lowest and highest scoring continent is highest for the food quality, and much smaller in affordability and availability. This result could for example help NGO's to decide which food related factor should first be tackled in poorer countries.

It's also worth noting that while the order of continents within affordability, availability and quality is almost the same, it is pretty much reversed in the Healthy Diet category, meaning that poorer countries tend to eat healthier than richer ones.

E4: Correlation of meat consumption and life expectancy

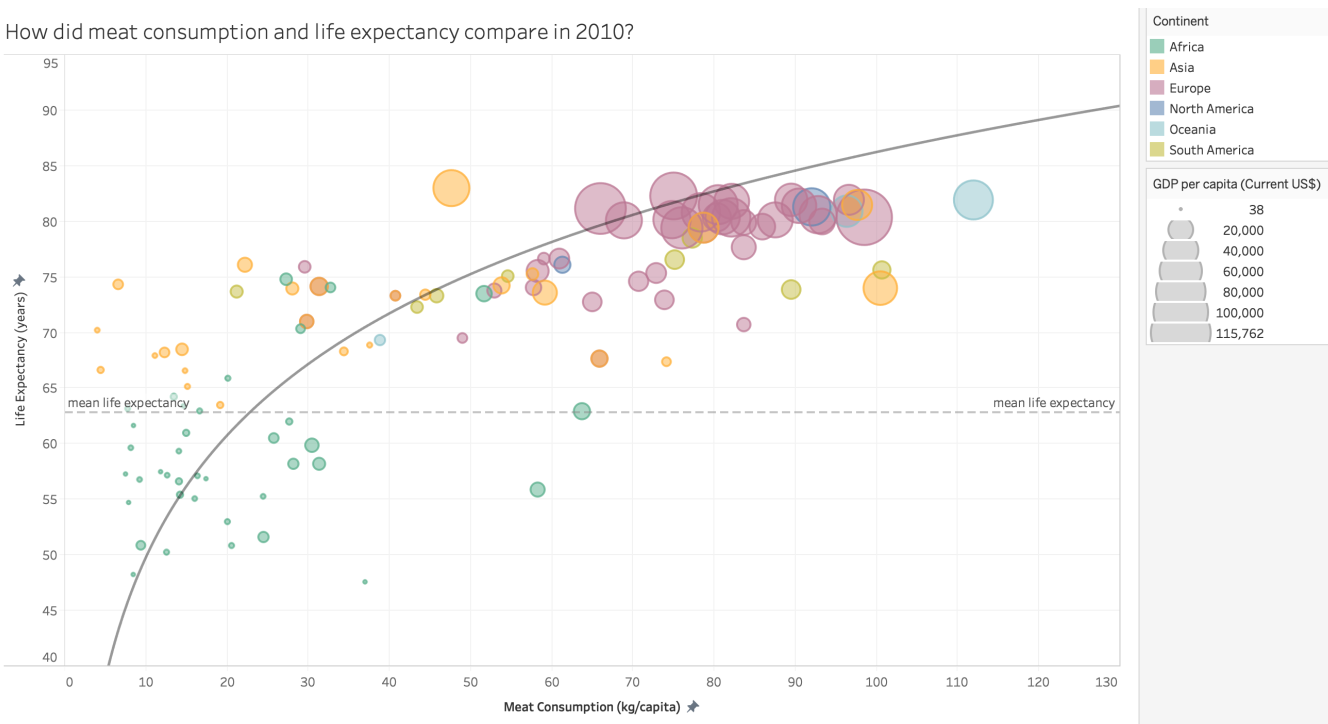


Figure 14: E4 Visualization (*Scatter Plot*)

Data	**Data Type**	**Encoding**	**Note**
GDP per capita	Quantitative	Size	Each country is represented by a circle with the size directly proportional to its GDP per capita (Current US$)
Country	Categorical	Position	Each country gets mapped to a distinct grid position based on its x-coord (LE) and y-coord (Meat Consumption)
Life Expectancy	Quantitative	Position	LE along the y-axis (starts from 40)
Meat Consumption	Quantitative	Position	Meat consumption along the x-axis (kg/capita)
Continent	Categorical	Color	Each continent is represented by a distinct color

Table 17: Visual Encoding for E4

Visualization Description

This visualization explores the correlation between meat consumption, life-expectancy and GDP per capita in 2010. Each country is plotted on a 2D-Chart with the x-axis depicting the meat consumption in kg per capita. The y-axis depicts the life expectancy in years. Note that the axis begins at 40 years in order to increase the visualization appeal. Also, we encode the GDP per capita in 2010 as the size of the bubble and color each country according to the continent it belongs. Furthermore, we add a reference line that shows the average life expectancy. This is especially useful when comparing different years. Finally, we draw a trend line to study the correlation between life expectancy and meat consumption. The correlation co-efficient (R^2 value) and the corresponding p-value is labeled on the trend line.

Insights

The trend line suggests a positive correlation ($R^2 = 0.516, p < 0.0001$) between meat consumption and life expectancy for data in 2010. As discussed earlier, this result does not suggest a causation between higher meat consumption leading to higher life expectancy. The underlying reason might be due to countries with higher life-expectancy usually having a higher GDP per capita, thus having more resources to buy meat in the first place, which is usually rather expensive. This explanation is supported by the fact that the countries with a high meat consumption, depicted on the right side of the chart, also have bigger bubbles, meaning a higher GDP per capita. Thus the chart might rather suggest that meat consumption tends to rise as we get richer.

Looking at the distribution of the continents, we can see clear patches, for example African countries tending to be in the lower-left corner, while European, Oceanian and American countries being in the top right. This result supports earlier analysis suggesting that richer countries tend to have higher life expectancy and also meat consumption.

6 Work and Employment

Given that most of our lives are spent working, it is natural to ask whether our jobs have any bearing on life expectancy. Indeed, there are many ways to analyze employment indices and their relation to life expectancy. In this section, we select several country-level statistics that encompasses different perspectives to answer aspects of this question.

The Gross Domestic Product (GDP) per capita is used as a bridging indicator for the other work-related variables here, and as a way to sort and rank countries. Additionally, its strong correlation with life expectancy, as represented by the so-called Preston curve [7], is fairly well studied. In the following, we link GDP and life expectancy to the economic makeup of a country in terms of its industrial sectors, and also the typical weekly working hours. Finally we investigate how changes in life expectancy have impacted retirement ages around the world.

A variety of visualization techniques were used to explore our analytical questions. They can be found in Table 18.

6.1 Exploration-Visualization Mapping

#	*Exploration*	*Visualization*
E1	How do the GDP and different types of jobs correlate to life expectancy?	*Stacked bar with line plot overlay*
E2	Does the number of working hours have any bearing on life expectancy?	*Scatter Plot*
E3	How has retirement age been affected by life expectancy?	*Adapted Gantt Chart*

Table 18: Visualizations for questions explored

6.2 Visualizations

E1: How do the GDP and different types of jobs correlate to life expectancy?

Data	**Data Type**	**Encoding**	**Note**
Country	Categorical	Position	Countries are ranked from left to right based on its GDP per capita
Job Sector	Categorical	Color	Each sector has a distinct color
% of Job Sector	Quantitative	Position	Sectors make up 100% in total
Life Expectancy	Quantitative	Position	LE is shown as a line plot

Table 19: Visual Encoding for E1

Visualization Description
Countries were ranked by GDP per capita in descending order from left to right in this visualization. For each country, we further provide a breakdown of the major economic activities in terms of agriculture, industry and services, or equivalently primary, secondary and tertiary industries. The prevalence of each sector is given as a percentage, making up 100% in total, and denoted by colours in the stack bar chart. The bar chart is overlayed with a line plot denoting average life expectancy for each country to enable direct comparison to the economic indicators. Note that the range of the axes on both sides are equal so both the percentages and life expectancies can be read off easily on either side.

Insights
It is immediately apparent that GDP per capita has a positive correlation with life expectancy, since the line plot generally slopes downwards from left to right, ignoring the variations between individual countries. Nevertheless,

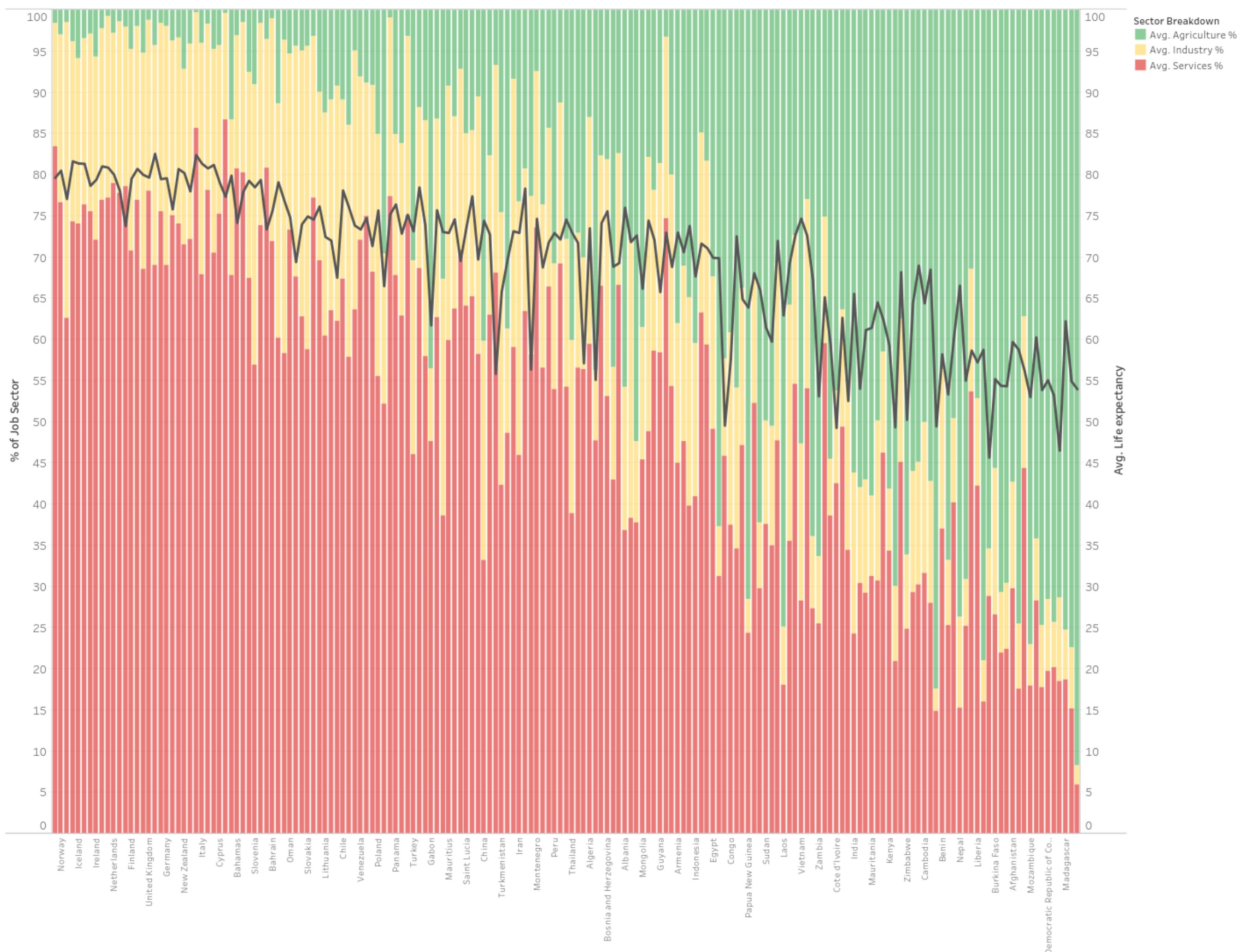


Figure 15: E1 Visualization (*Stacked bar with line plot overlay*)

there are outliers to this trend that can be picked out from the large spikes in the line plot between neighbouring countries. An example is Nigeria having a lower life expectancy than expected for a country of their GDP per capita.

Through the chart, we also see how as a country develops through the economic value chain, it shifts from a primarily agriculture-based economy to more value-added secondary and tertiary sectors. This ultimately results in a higher GDP per capita value. Comparing the extremes, Burundi's economic makeup is over 90% agriculture-based, wheres the country with the highest GDP per capita, Luxembourg, over 83% are services and 15% industries.

E2: Do working hours have any bearing on life expectancy?

Visualization Description
This visualization primarily compares average life expectancy to the average weekly hours worked per employee for different countries over the same time period (2000-2015). For further insight the scatter points were colored based on *log*(GDP per capita). A *log* calculation was used for better differentiation since the range of values was quite large. The points were separated into 4 quadrants based on the average values of each of the two dimensions. Several countries that may be of interest to the reader were marked.

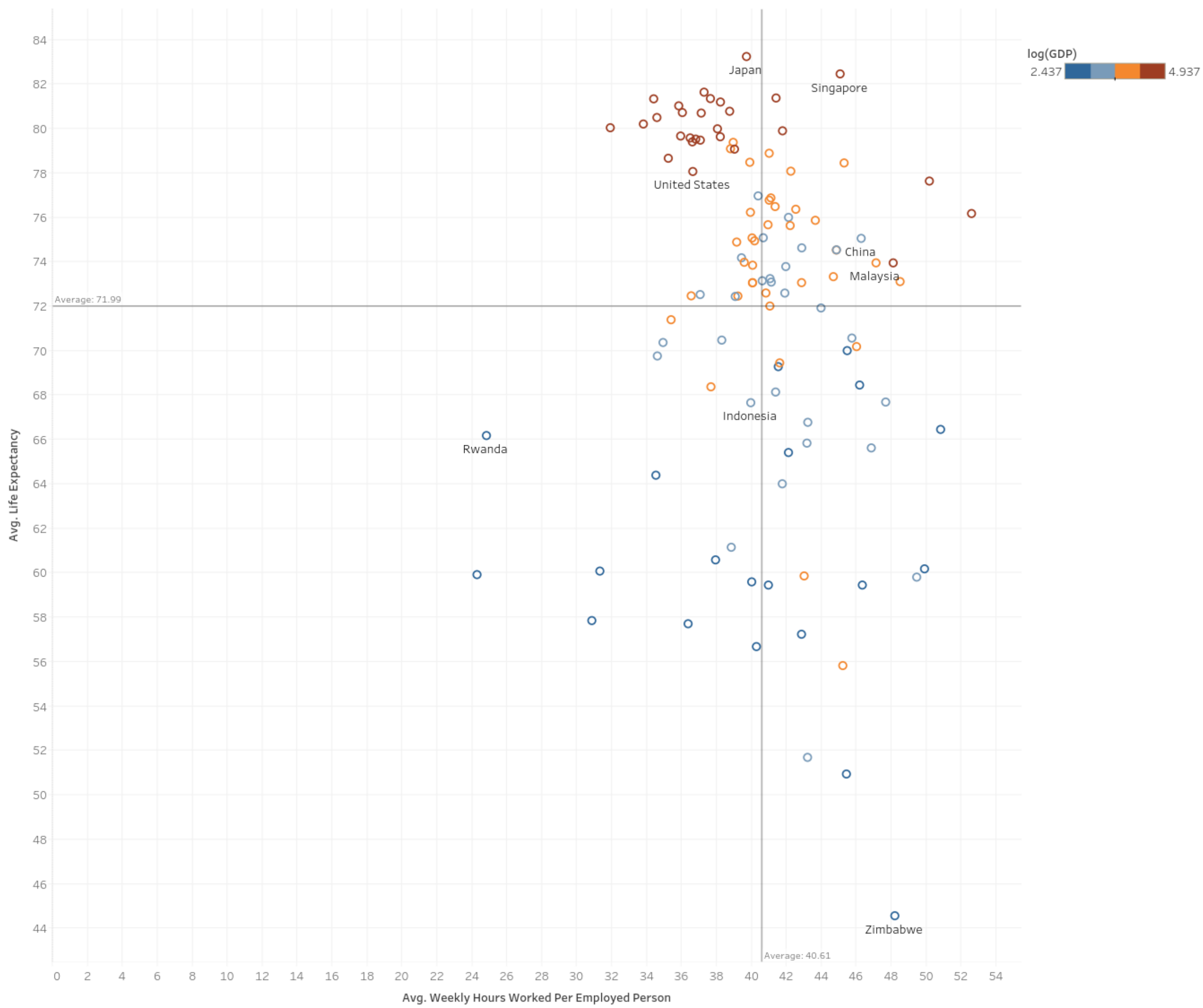


Figure 16: E2 Visualization (*Scatter plot*)

Data	**Data Type**	**Encoding**	**Note**
Country	Categorical	Position	Each country gets mapped to a distinct grid position based on its x-coord (avg. weekly hours worked) and y-coord (avg. LE)
Life Expectancy	Quantitative	Position	LE along the y-axis
Avg. Weekly Hours Worked	Quantitative	Position	No. of hours along the x-axis
log(GDP per capita)	Categorical	Color	GDP was split into four colored bands

Table 20: Visual Encoding for E2

Insights

At a country level there is poor correlation between life expectancy and weekly hours worked. Initially the points were fitted with a regression curve based on Tableau's analytics, but its inclusion was decided against in the final version since it could be misleading in terms of a definite pattern (different curve forms were a possible fit). Instead, the chart reinforces the impact of GDP on life expectancy as most of the richest countries were near the top of the chart. Interestingly, the number of hours worked were much more consistent in the richest countries, with the points tightly clustered in the upper left quadrant, wheres it was much more variable in the countries with lower GDP. Singapore was somewhat of an outlier in this regard as she is found instead in the group of countries with high life expectancy and high number of working hours (upper right quadrant).

Unfortunately within-country data on the impact of jobs and working hours on life expectancy were not readily available as it is possible it would have showed a different picture, as was discussed in, for example [8].

E3: How has retirement age been affected by life expectancy?

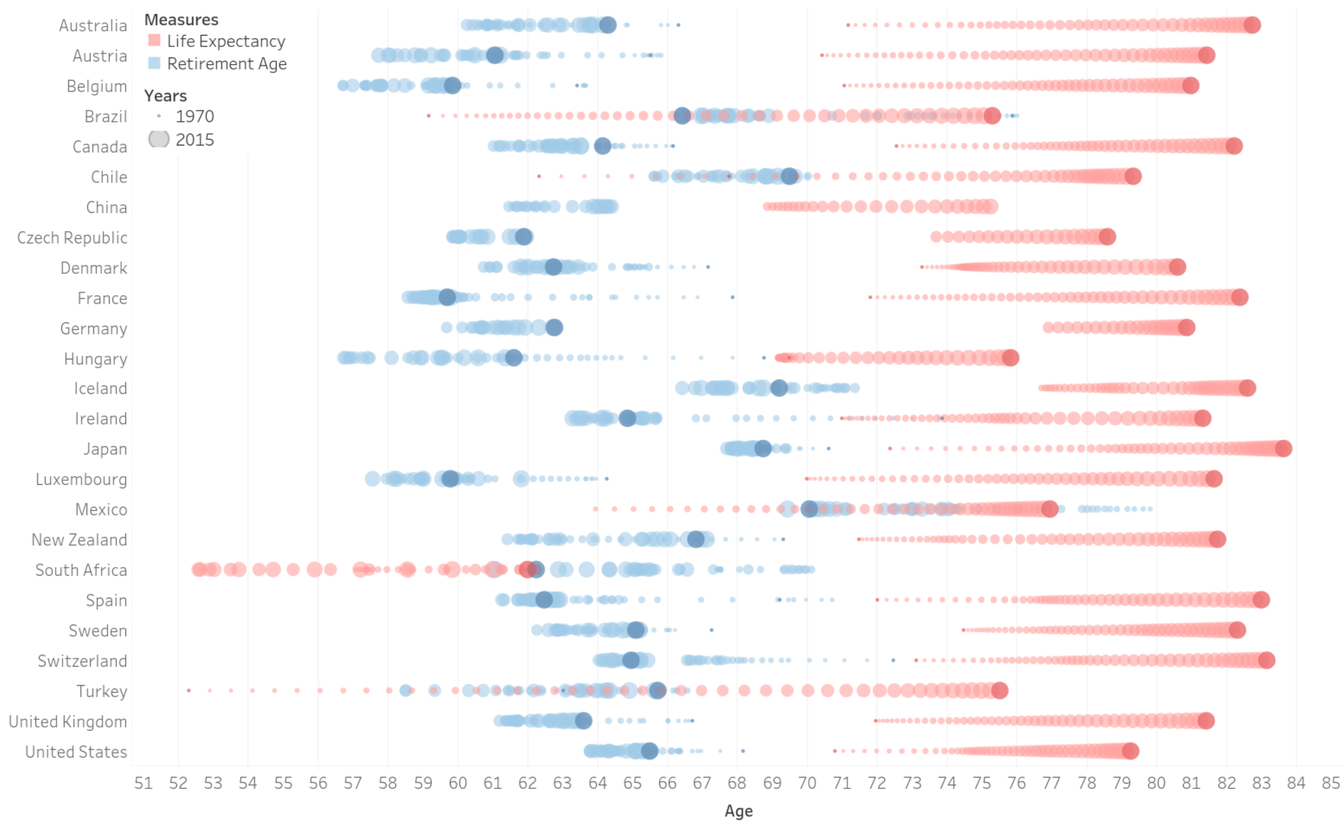


Figure 17: E3 Visualization (*Adapted Gantt Chart*)

Data	**Data Type**	**Encoding**	**Note**
Country	Categorical	Position	Countries along the y-axis
Age	Ordinal	Position	Both retirement and LE share the age axis
Life Expectancy	Categorical	Color	To distinguish from retirement age
Retirement Age	Categorical	Color	To distinguish from LE, aggregated from men and women
Start/End Years	Ordinal	Color	Darker colors used to more effectively highlight the start (1970) and end (2015) points
Year	Ordinal	Size	Size of bubbles gets bigger with time

Table 21: Visual Encoding for E3

Visualization Description
The visualization is in essence using the form of a Gantt chart with two variables – average "effective" retirement age between men and women taken from selected Organisation for Economic Co-operation and Development (OECD) countries and depicted by the blue bubbles, and life expectancy depicted by the red bubbles. Effective retirement

age, loosely defined, is the average age at which the population declares retirement, that may be different from the official retirement age set by the government. Both variables were taken over the period of 1970 to 2015. Here each bubble represents an individual year, which grows with the progression of time, giving a visual cue to the reader on the direction and changes in value of the variable over the time period. The start and end years were highlighted using a darker color for easier tracking of points. The variables were evaluated on a common x-axis of age.

Insights

There is a fairly consistent increase in life expectancies over the years moving from left to right, following the progression of bubble size. It is also clear that the rate of improvement of life expectancies differ between countries according to the length the bubbles are spread out, Turkey being the most impressive. In contrast, retirement ages effectively decrease as of 2015 compared to 1970 in almost all the countries. Intriguingly, retirement ages seem to decrease initially (right to left progression), before a reversal in more recent years (left to right progression). This causes the overlapping pattern seen in the majority of the countries – a cluster made up of bigger bubbles on the left and smaller ones on the right, with the darker blue 2015 point somewhere in the middle. This observation has a real life explanation given that many governments have introduced policies to prolong the number of working years in more recent times as a response to increasing life expectancy. The visualization however shows that the gap has actually grown larger between the effective retirement age and the life expectancy (with the exception of South Africa), indicating that ultimately retirement ages have not kept up with the increased life expectancy. Additionally, while not shown here, the problem is more acute for women (the original data came with a gender breakdown which was combined for this visualization), seeing how they on average retire earlier but live longer.

This problem is symptomatic of an ageing population and has several important social impacts. One major consequence is the rise in Dependency Ratio i.e. ratio of dependents to those of working age, which adds to the tax burden of the working population. As life expectancy increases, people are expected to be productive for a longer period, but it is questionable whether they will be healthy enough to work during their later years.

7 Steps to Re-create the Visualizations

In this section, we list down detailed steps to re-create two of our visualizations. Both visualizations were done using the latest version of Tableau Desktop (ver. 2019.1.2).

7.1 Section 4, Viz E3: Continent-wise distribution of Smoking, Alcohol-related disorders, and LE

The visualization was laid down using a butterfly chart (figure 18). The visual encoding can be found in table 16 (p. 23).

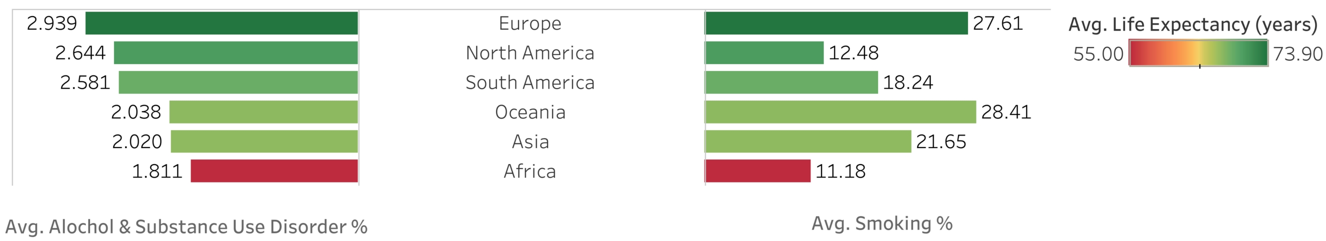


Figure 18: Section 4, E3 Visualization (*Butterfly Chart*)

4 variables were used to create the visualization: *Continent Name*, LE, Smoking%, and Alcohol-related Disorder%. The steps to re-create the viz. are enumerated as follows:

1. Drag *Continent Name* from the *Dimensions* panel to *Rows*.
2. Drag Alcohol-related Disorder% from the *Measures* panel to *Columns* and change the measure to Average.
3. Drag Smoking% from the *Measures* panel to *Columns* and change the measure to Average.
4. On the *Marks* panel, select Alcohol-related Disorder% and drag LE from the *Measures* panel to the *Color* attribute of the *Marks* panel. Change the measure to Average. Also, change the color scale to continuous Red-Green-Gold diverging color scheme.
5. Repeat the previous step for the *Marks* panel of Smoking%.
6. Right-click on the *Continent Name* field on Rows and click sort. Choose sort 'By Field' and sort order as 'Descending'. Choose the field as LE and aggregation as Average.
7. Right-click on the axis of Alcohol-related Disorder% and click 'Edit Axis'. Then find 'Scale' and under it, check the 'Reversed' box. The basic layout of the butterfly chart has now been created. We shall do some additional formatting.
8. Create a Calculated Field in the *Measures* panel and name it 'Zero Axis', in place of the calculation formula, just put a zero.
9. Drag the *Zero Axis* from the *Measures* panel to the *Columns* and place it in between the two variables already placed there. This forms the zero axis in the middle.
10. In the *Marks* panel for *Zero Axis*, drag *Continent Name* to the *Label* attribute.
11. Right-click on *Continent Name* in *Rows* and uncheck 'Show Header'.
12. Right-click on the Avg. Alcohol-related Disorder% in *Columns* and uncheck 'Show Header'.

13. Open the *Format* menu-option from the menu-bar and select ‘Lines’. Under the ‘Sheet’ ribbon, select ‘None’ for dropdowns: ‘Zero Lines’ and ‘Axis Ticks’. Then, under the ‘Columns’ ribbon, select ‘None’ for the ‘Grid Lines’ dropdown.

14. Go to the *Marks* panel for *Zero Axis* and change the mark from ‘Automatic’ to ‘Text’.

15. In the *Marks* panel for ALCOHOL-related Disorder%, left-click on the *Label* attribute and check the ‘Show Mark Labels’ option.

16. In the *Marks* panel for SMOKING%, left-click on the *Label* attribute and check the ‘Show Mark Labels’ option.

7.2 Section 5, Viz E4: Correlation of meat consumption and life expectancy 2010

The visualization was laid down using a scatter plot (figure 19). The visual encoding can be found in table 17 (p. 24).

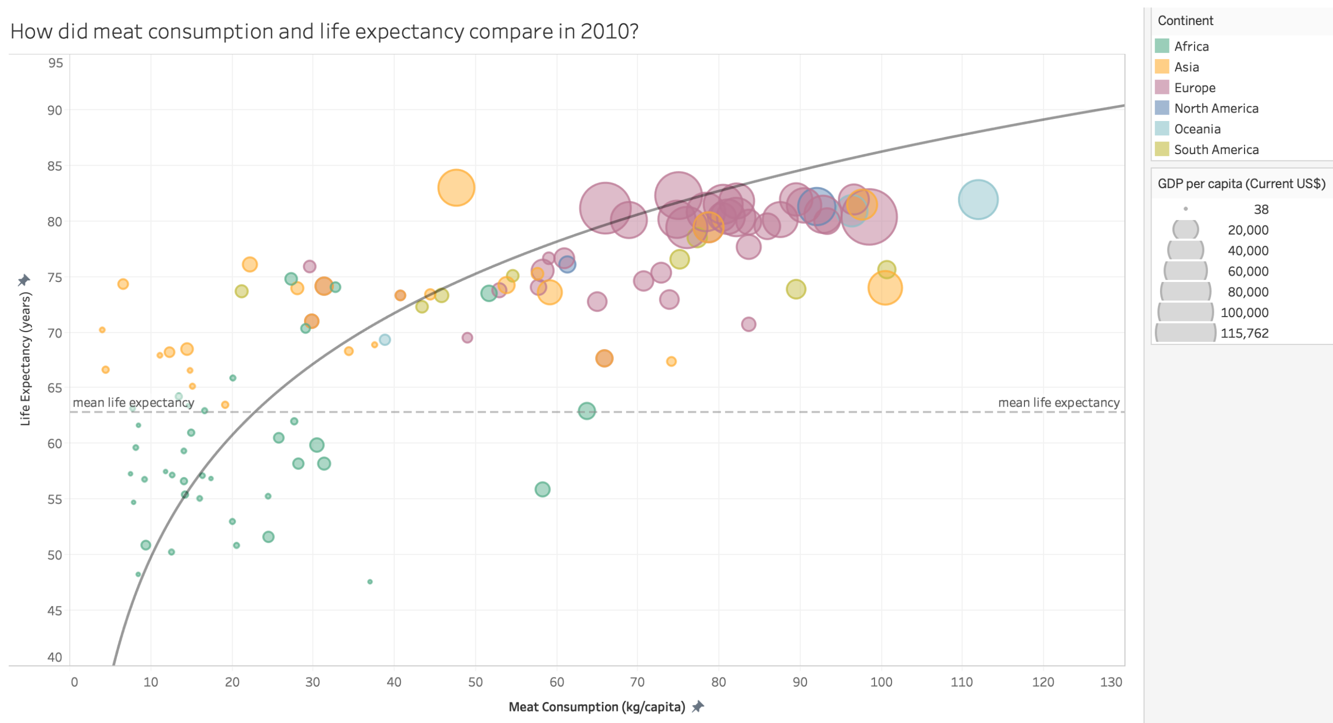


Figure 19: Section 4, E4 Visualization (*Scatter Plot*)

The following variables were used to create the visualization: *continent_name* (String Dimension), *life_expectancy* (Measure), *gdp_per_capita* (Measure), *meat_consumption* (Measure), *year* (Discrete Dimension) and *country_name* (String Dimension). The steps to re-create the visualization are enumerated as follows:

1. Drag *life_expectancy* from the *Measures* panel to *Rows* and change the measure to Average.

2. Drag *meat_consumption* from the *Measures* panel to *Columns* and change the measure to Average.

3. Drag *continent_name* from the *Dimensions* panel to the *Color* attribute of the *Marks* panel. Change the color to the ones displayed in the viz, including changing the transparency to 50%. Change the Shape to filled circle.

4. Drag *country_name* from the *Dimensions* panel to the *Detail* attribute of the *Marks* panel.

5. Drag *gdp_per_capita* from the *Measures* panel to the *Size* attribute of the *Marks* panel. Left-click on Size and increase the size to around 2/3 of the entire slider. Also, right-click on the created SUM(gdp_per_capita)-Size legend that was created on the right hand side and increase the *Smallest*-size a little to match the sizes of the viz above.

6. Drag *year* from the *Dimensions* panel to the *Filter* panel and tick the 2010 field only..

7. In order to create the ring for the bubbles, drag *life_expectancy* from the *Measures* panel to *Rows* again and change the measure to Average. Then right-click on the y-axis of the bottom plot and choose "Dual Axis" and "Synchronize Axis". Select the right y-axis and on the *Marks* panel, change the Shape to circle and remove the *Measure Names* that was created as a color. Right-click on the right y-axis and deselect "Show Header".

8. Double-Click on the y-axis and under *Range*, change to fixed and change the *Fixed Start* value to 40.

9. In order to generate the trend line, make sure you selected the *All*-Marks, then select the *Analytic* section in the top left. Double-click *Trend Line*. Then, right-click somewhere in the plot and choose *Trend Lines -¿ Edit Trend Lines*. Choose *Logarithmic* and untick *Allow a trend line per color*.

10. In order to generate the reference line, right-click on the y-axis and choose *Add Reference Line*. Make sure *AVG(life_expectancy)* and *Average* is selected. In *Formatting*, change the line to a dotted line.

References

[1] World Health Organization. *The world health report 2006: working together for health.* World Health Organization, 2006.

[2] Amit V Khera, Connor A Emdin, Isabel Drake, Pradeep Natarajan, Alexander G Bick, Nancy R Cook, Daniel I Chasman, Usman Baber, Roxana Mehran, Daniel J Rader, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *New England Journal of Medicine*, 375(24):2349–2358, 2016.

[3] Office for National Statistics. What affects an area's healthy life expectancy? https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/articles/whataffectsanareashealthylifeexpectancy/2017-06-28, 2017. Accessed: 2019-04-16.

[4] David I Feldman, Mouaz H Al-Mallah, Steven J Keteyian, Clinton A Brawner, Theodore Feldman, Roger S Blumenthal, and Michael J Blaha. No evidence of an upper threshold for mortality benefit at high levels of cardiorespiratory fitness. *Journal of the American College of Cardiology*, 65(6):629–630, 2015.

[5] PIA Mäkelä. Alcohol-related mortality by age and sex and its impact on life expectancy: Estimates based on the finnish death register. *The European Journal of Public Health*, 8(1):43–51, 1998.

[6] Samuel H Preston, Dana A Glei, and John R Wilmoth. Contribution of smoking to international differences in life expectancy. *International differences in mortality at older ages: Dimensions and sources*, pages 105–31, 2011.

[7] Samuel H Preston. The changing relation between mortality and level of economic development. *Population studies*, 29(2):231–248, 1975.

[8] Paul D Sorlie and Eugene Rogot. Mortality by employment status in the national longitudinal mortality study. *American Journal of Epidemiology*, 132(5):983–992, 1990.

[9] Jeffrey M Friedman. Obesity in the new millennium. *Nature*, 404(6778):632, 2000.

[10] Suzanne M Wright and Louis J Aronne. Causes of obesity. *Abdominal Radiology*, 37(5):730–732, 2012.

[11] JC Seidell and J Halberstadt. Overweight, obesity and life expectancy: do people with a high bmi live longer? *Nederlands tijdschrift voor geneeskunde*, 160:D859–D859, 2016.

[12] Kevin R Fontaine, David T Redden, Chenxi Wang, Andrew O Westfall, and David B Allison. Years of life lost due to obesity. *Jama*, 289(2):187–193, 2003.

[13] Fida Bacha and Samuel S Gidding. Cardiac abnormalities in youth with obesity and type 2 diabetes. *Current diabetes reports*, 16(7):62, 2016.

[14] Sveta Mohanan, Hazel Tapp, Andrew McWilliams, and Michael Dulin. Obesity and asthma: pathophysiology and implications for diagnosis and management in primary care. *Experimental Biology and Medicine*, 239(11):1531–1540, 2014.

[15] Leandra Abarca-Gómez, Ziad A Abdeen, Zargar Abdul Hamid, Niveen M Abu-Rmeileh, Benjamin Acosta-Cazares, Cecilia Acuin, Robert J Adams, Wichai Aekplakorn, Kaosar Afsana, Carlos A Aguilar-Salinas, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128· 9 million children, adolescents, and adults. *The Lancet*, 390(10113):2627–2642, 2017.

[16] Luigi Ferrucci, Grant Izmirlian, Suzanne Leveille, Caroline L Phillips, Maria-Chiara Corti, Dwight B Brock, and Jack M Guralnik. Smoking, physical activity, and active life expectancy. *American journal of epidemiology*, 149(7):645–653, 1999.

[17] J Westman, K Wahlbeck, T Munk Laursen, M Gissler, Merete Nordentoft, J Hällgren, M Arffman, and U Ösby. Mortality and life expectancy of people with alcohol use disorder in denmark, finland and sweden. *Acta Psychiatrica Scandinavica*, 131(4):297–306, 2015.

[18] Vladimir Shkolnikov, Martin McKee, and David A Leon. Changes in life expectancy in russia in the mid-1990s. *The Lancet*, 357(9260):917–921, 2001.

[19] Kayo Kurotani, Shamima Akter, Ikuko Kashino, Atsushi Goto, Tetsuya Mizoue, Mitsuhiko Noda, Shizuka Sasazuki, Norie Sawada, Shoichiro Tsugane, Japan Public Health Center based Prospective Study Group, et al. Quality of diet and mortality among japanese men and women: Japan public health center based prospective study. *bmj*, 352:i1209, 2016.

[20] Ashkan Afshin, Patrick John Sur, Kairsten A Fay, Leslie Cornaby, Giannina Ferrara, Joseph S Salama, Erin C Mullany, Kalkidan Hassen Abate, Cristiana Abbafati, Zegeye Abebe, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*, 2019.