Rise of the Dragons

MSIS2629: DASHBOARD'S, SCORECARD'S AND VISUALIZATION
AMEYA GHATPANDE – W1185595
INDIVIDUAL PROJECT

Table of Contents

Project Overview

Life expectancy is a statistical measure of the average time an organism is expected to live, based on the year of their birth, its current age and other demographic factors like sex, access to health care services, living standards etc.[1]. Humans are not immortal (other than superman). Every one of us must say goodbye to this beautiful world one day. Average Human life expectancy (around the world) in 2016 is 78.8 years[2]. If we see the trend over last 100 years, then life expectancy has increased by about 20 years. Long life is a blessing and no one can predict your life. Life expectancy in humans mainly depends on following factors[3]

- Child Mortality
- Access to Health care services
- Epidemics/diseases
- Humans destructions - Wars/genocides
- Pollution and environmental conditions
- Lifestyle and living standards
- Diet habits

This project explains the life expectancy in different regions (countries and continents) over a period last 50 years. It will mainly cover visual stories about following four things

- Life expectancy in Asia (and its growth)
- Star performer (Japan) over period of 50 years
- American life expectancy and high cost of health care
- Comparison between American and Japanese health care systems

Inspiration for selecting this topic

Hans Rosling was a Swedish statistician, educator and a dynamic speaker[4]. His visualization "200 countries, 200 years, 4 minutes - The Joy of Stats[5] explains the life expectancy over 200 years and few parameters affecting it. This short video is very powerful and informative. Few days back I got a chance to see his video and was highly impressed about his data driven storytelling technique. This inspired me to explore more about the life expectancy data and its trends.

---

[1] https://en.wikipedia.org/wiki/Life_expectancy
[2] http://www.npr.org/sections/health-shots/2016/12/08/504667607/life-expectancy-in-u-s-drops-for-first-time-in-decades-report-finds
[3] https://ourworldindata.org/life-expectancy/
[4] https://en.wikipedia.org/wiki/Hans_Rosling
[5] https://www.youtube.com/watch?v=jbkSRLYSojo

"Life is beautiful" is a classic 1997 comedy-drama Italian movie where a father enacts an imaginative play to protect his 7-year-old son in a Nazi camp[6]. This is one of my favorite movies and it shows the vision of a father to protect his son and his life's sacrifice. This movie also tells us how violent and unpredicted the life was back in early 1900's to 1940's. My generation people are lucky since we have not experienced world wars and atomic bombings. Though no one can predict human extinction, wars, genocides and epidemics can surely affect the life expectancy badly[7].

I am highly impressed by these two above references and I have visualized life expectancy over the period of last 50 years for different continents/countries.

Audience

Anyone who is interested in knowing the human life expectancy trends will find this project interesting. The complexity of the visualizations has been designed keeping into consideration audience of varying skills. However, for 3rd and 4th story, where this project becomes more specific to US life expectancy trends and dives deep into "Dollar value of life" in USA, main audience for this project will be the health care officials of the US government. Looking at the visualization, exploratory as well as actionable insights can be derived. This will help the US government to rethink on the health care spending's per person per year. The data shown about other countries can also be considered while implementing new policies.

Process for completing the project

I followed the following steps in the project

Deciding the topic

Even though life expectancy was always back of my mind, since this is vast subject, I narrowed down the scope of project to only following topics
- Asia's growth story
- Japan and its rise
- US health care spending
- Comparison between US and Japanese health care systems

---

[6] https://en.wikipedia.org/wiki/Life_Is_Beautiful

[7] http://www.cnn.com/interactive/2014/10/health/epidemics-through-history/
http://www.worldatlas.com/articles/the-10-worst-epidemics-in-history.html

Data collection and exploration/wrangling/cleaning

I have collected data from multiple sources and merged/joined/blended (In tableau language) them together so that I can generate the insights. My data sources are -

Data Name	Explanation about data	Reference (Data source)
life-expectancy.csv	Average life expectancy for all countries over 200 years	https://ourworldindata.org/life-expectancy/
Countries.xlxs	Country and the continent it belongs to	http://www.worldatlas.com/cntycont.htm
life-expectancy-vs-healthcare-expenditure.csv	Country wise health care spending over the years	https://ourworldindata.org/grapher/life-expectancy-vs-healthcare-expenditure
Health-parameters.csv	Comparison between US and Japan health care systems	http://www.nationmaster.com/country-info/compare/Japan/United-States/Health/Quality-of-health-care-system

Since the data sources are collected from multiple websites, they require cleaning and transformations like

1] New column creations

I created a new column "Life-Threatening-events" mapping the major life loss because of wars/epidemics and the corresponding countries which had the maximum impact. This was used in a worksheet to show these effects on life expectancy

```
In [45]: def worldEvents(Entity):
            if Entity == "Japan":
                return "Hiroshima Nagasaki Atomic Bombing"
            elif Entity == "Spain" or Entity == "United States" or Entity == "United Kingdom":
                return "The Great Influenza Epidemic"
            elif Entity == "France" or Entity == "Russia":
                return "The nightmare of Tuberculosis"
```

```
In [46]: df["Life-Threatening-events"] = df.Entity.apply(worldEvents)
```

```
In [54]: df.columns
         df[((df.Entity == "Russia") | (df.Entity == "Spain"))]
```

```
Out[54]: Index([u'Entity', u'Year', u'Country code',
                u'Life Expectancy at Birth (both genders)', u'Life-Threatening-events'],
               dtype='object')
```

Out[54]:

	Entity	Year	Country code	Life Expectancy at Birth (both genders)	Life-Threatening-events
10417	Russia	1896	RUS	31.900	The nightmare of Tuberculosis
10418	Russia	1900	RUS	30.500	The nightmare of Tuberculosis

2] Null removals

There were "NaN" for health care spending of some countries for early years. I filtered them out as follows

```
In [3]: lifeAndHealth = pd.read_csv("life-expectancy-vs-healthcare-expenditure.csv")
```

```
In [5]: len(lifeAndHealth)
        lifeAndHealth[lifeAndHealth.Entity == "Afghanistan"]
```

Out[5]: 12441

Out[5]:

	Entity	Year	Country code	Health Expenditure per capita (WDI)	Life Expectancy (at birth) World Bank data
0	Afghanistan	1960	AFG	NaN	31.580049
1	Afghanistan	1961	AFG	NaN	32.095976
2	Afghanistan	1962	AFG	NaN	32.611878
3	Afghanistan	1963	AFG	NaN	33.127317

```
In [14]:  lifeAndHealth = lifeAndHealth[lifeAndHealth["Health Expenditure per capita (WDI)"].isnull() != True]

In [15]:  len(lifeAndHealth)
Out[15]:  3543

In [16]:  lifeAndHealth.head(5)
```

Out[16]:

	Entity	Year	Country code	Health Expenditure per capita (WDI)	Life Expectancy (at birth) World Bank data
42	Afghanistan	2002	AFG	77.562874	55.671878
43	Afghanistan	2003	AFG	91.867101	56.107561
44	Afghanistan	2004	AFG	90.123345	56.569732
45	Afghanistan	2005	AFG	94.045016	57.058439
46	Afghanistan	2006	AFG	98.527584	57.570659

3] Join from different sources

After loading the data into Tableau, I used the right combination of joins, to get the relevant data as follows
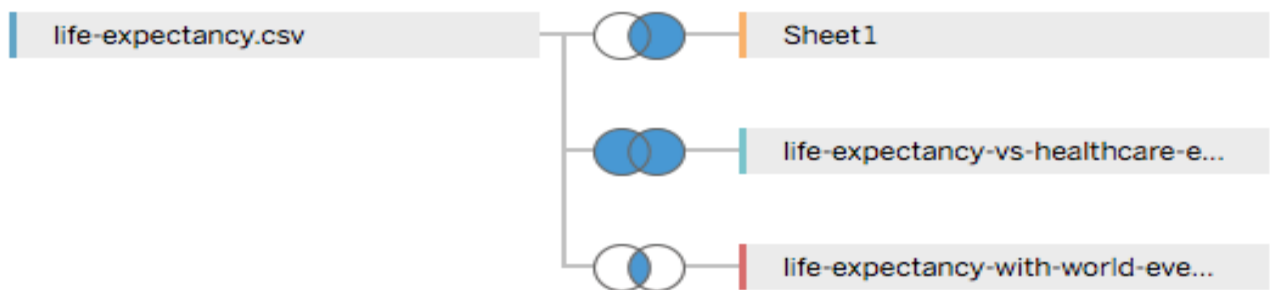

4] Calculated fields

Dollar Value of Life - I created this calculated field to identify how much each country is spending on the health of everyone to the expected life expectancy. Smaller the ratio indicates better investment in health spending's.

Dollar Value of Life = Health spending per person per year / life expectancy in that country for that year

Example] US has 128$ for year 2015 which is the highest among all the countries

5] Sets

I created few sets from the large/original columns so that my filters can work more effectively. In the below set called "G7 nations", I selected 7 of the G7 nations as follows. Sets are useful, since they can be used as filters in the dashboard and unnecessary values are not shown to the user. (Only selected values are shown to user to choose from)
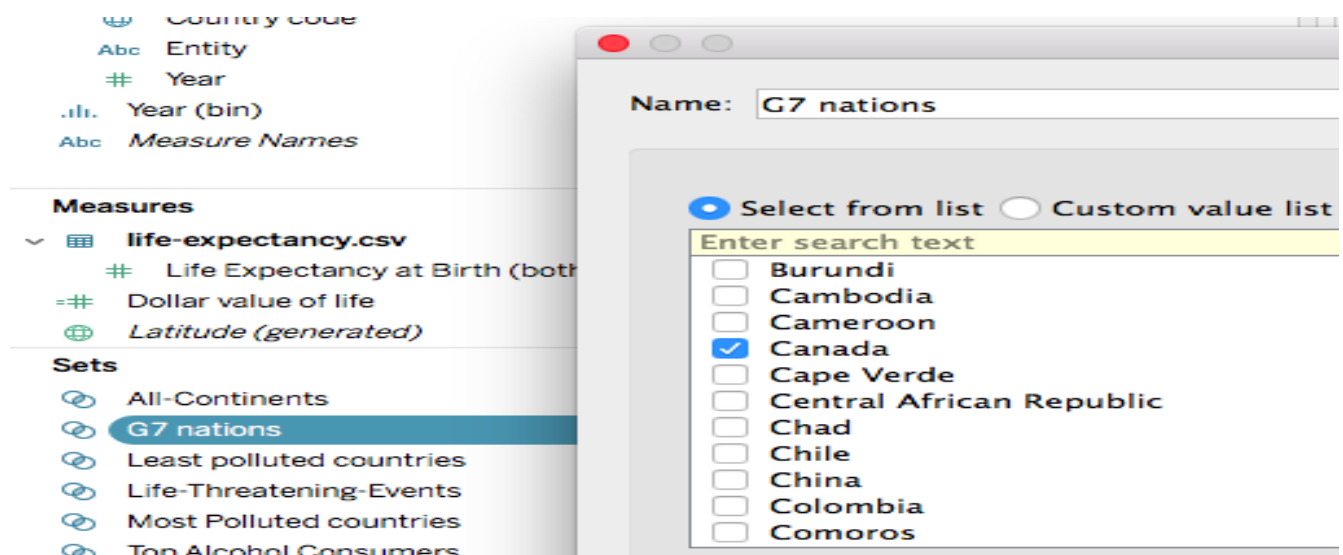
6] Bins

I created a group of 20 years combined, so that I can show the change in life expectancy for different continents over last 60 years. Bins helped me average the life expectancy for 20 years
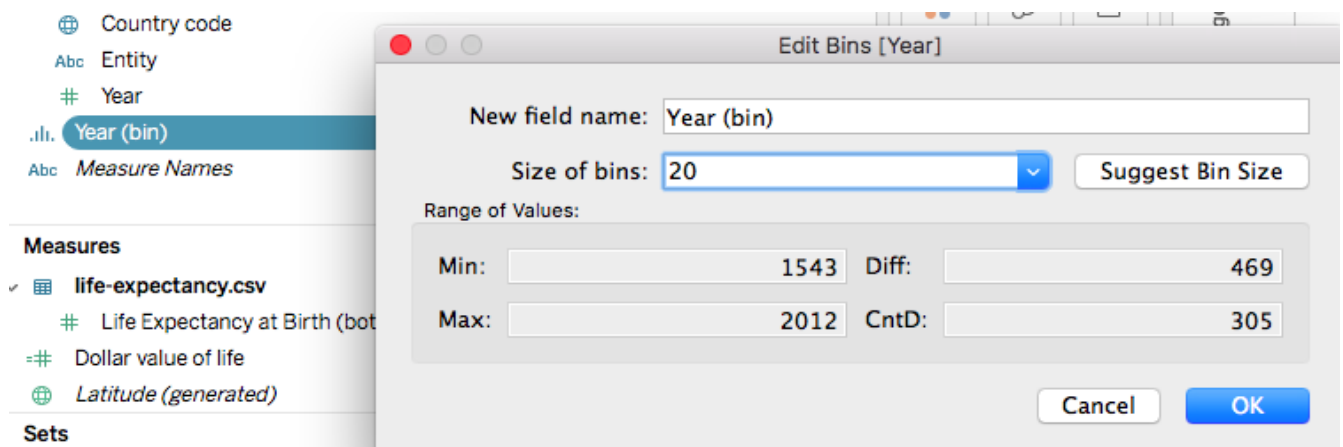
Creating a storyline

Once the data was cleaned, I created multiple visualizations to support the claims. This step involved eliminating excessive visualizations which were not helping the story and finalizing the worksheets. Then all the worksheets were arranged so that flow of the story was complete. This dashboard has following flow

- **Exploratory / Perception approach** - The first visualization tells how life expectancy in Asia is increasing. This just gives the figures without going into country level details. This sets the stage for bigger picture.

  Then the next visualization shows the details about Japan and how it helps in bringing up the life expectancy of Asia. This is one level deeper as compared to first worksheet

- **Comprehension approach** - The third visualization plots the "Dollar value of life" for all the G7 nations. This clearly shows US as an outlier compared to others

  Projection approach - The last visualization goes into details of health care system between US and Japan. Direct comparison is made because Japanese health care system is the best in the world. Corrective actions can be taken by the US health care authorities.

Visualizations

1] Rise of the dragons

Claim - Asia has seen an increase of 40% in life expectancy over last 50 years

Explanation - During 1940's, life expectancy was poor in Asian countries as compared to the west. Main reasons for poor life expectancy were bad economic health, wars, poor living conditions and less access to medical facilities. In 1940's Asia was 17 years behind in terms of life expectancy as compared to leading continent(Europe). Today (till 2010) the gap is just 7 years. Asia is still 2nd last (only above Africa which has life expectancy of only 56 years) in all continents. However, with fast developing countries like India (which still has low life expectancy 65 years in 2016) and other Southeast Asian countries, Asia can improve the rank in next 20 years. In the last 50 years, following factors have helped Asia immensely
  - After World war 2, Japan has grown and prospered a lot. In fact, Life expectancy is highest (83 years) in Japan in 2016 amongst all the countries.

- Independence to India in 1947, and the industrial revolution and development led to the prosperity and better living conditions. Access to health care and improved diet
- China becomes a manufacturing giant and rise of industrial revolution improves the working conditions as well as living standards
- No major epidemic outrage in last 50 years in Asia (which killed more than 20k people in Asia)
- Eradication of yellow fever, polio and high improvements in child mortality among developing countries like India, Bangladesh

Process of creating dashboard
- Map the countries into corresponding continents. Done by joining a different data source
- Create year bins of 10 years each from 1940 onwards
- Calculate the average life expectancy for all countries in a continent for a given year
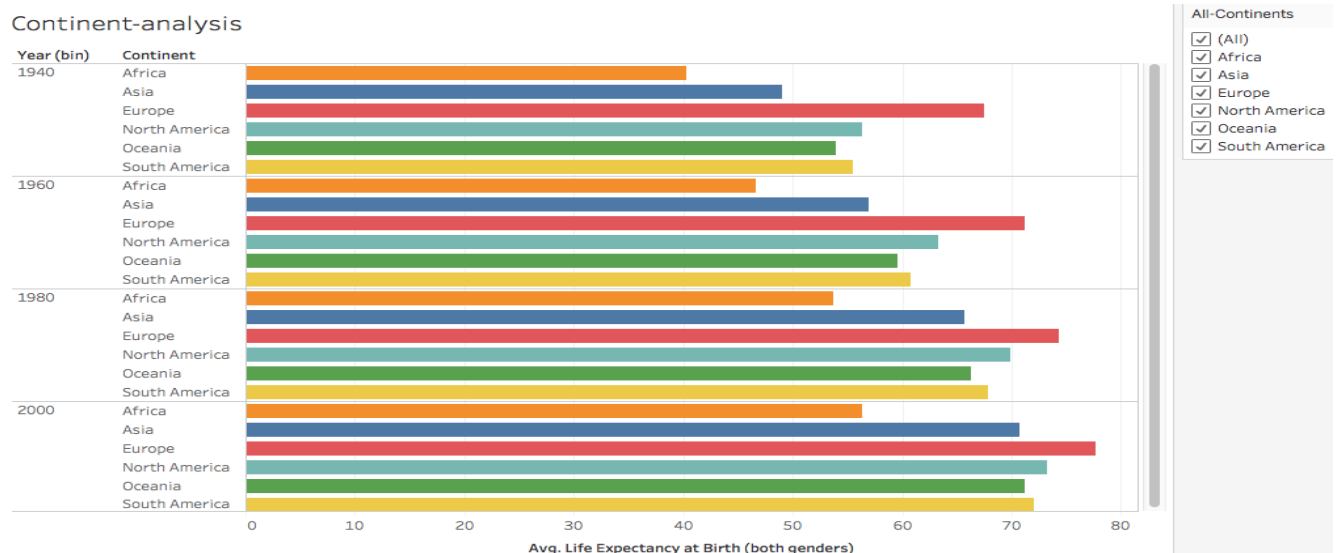
Principles of Visualization used -
- The trend in life expectancy growth of Asia over the years is shown as direct comparison with other continents (Bar graphs)
- **Aesthetics** - Different colors for continents helps in separating them easily
- Grouping of years (bins) helps reduce the number of entries. This makes the worksheet more readable and convenient to user

Warrant for the claim -
http://www.businessinsider.com/how-has-life-expectancy-changed-throughout-history-2015-6
http://visual.ons.gov.uk/how-has-life-expectancy-changed-over-time/

2] Japan life expectancy

Claim - Life expectancy grew as fast as a bullet train in Japan

Explanation - Japan is famous for its high-speed bullet trains. It can also be seen how quickly Japan has risen the ranks in life expectancy. This small Asian country which is known for its quality of products is also home to one of the costliest cities in world (Tokyo). Japan has suffered a lot during its lifetime. Let it be the Atomic bombings, the Yen stock market crash[8], or the constant fear of earthquakes (with major ones in 1923, 1995, and the most recent in 2011 etc.[9]). Japan had suffered an atomic bombing in 1945, which brought their life expectancy down from 48 years in 1935 to 30 years in 1945. Every Time Japan has risen like a phoenix from the ashes and is number one country with life expectancy of 83 years as of 2015.

Process of creating dashboard -
- Create a "Set" for years. Select 1960 and 2010 as two years
- Calculate the average life expectancy in the world for 1960 and 2010.
- Plot the life expectancy of only Asian countries. Use the Set of years to choose either 1960 or 2010. For any selected year, show the average life expectancy of the world in that selected year. Observe Japan and its growth above the average line from 1960 to 2010

Principles of Visualization used -
- Context is set. Here the average life expectancy is shown which acts as a context. This clearly shows how Japan is way above the world life expectancy average in 2010
- **Aesthetics** - Considering wide range of audience, the visualization is kept simple. Size of the circles (representing countries) is increased for better visibility

Warrant for the claim – (Articles/ References which support the claim)
http://www.nbcnews.com/news/asian-america/japan-has-highest-life-expectancy-any-major-country-why-n130871
http://pulptastic.com/10-reasons-japanese-people-get-really-old/
http://www.agewatch.org.uk/secrets-of-longevity/japanese-longevity/

---

[8] https://en.wikipedia.org/wiki/Lost_Decade_(Japan)
[9] https://en.wikipedia.org/wiki/List_of_earthquakes_in_Japan

Asian countries life expectancy
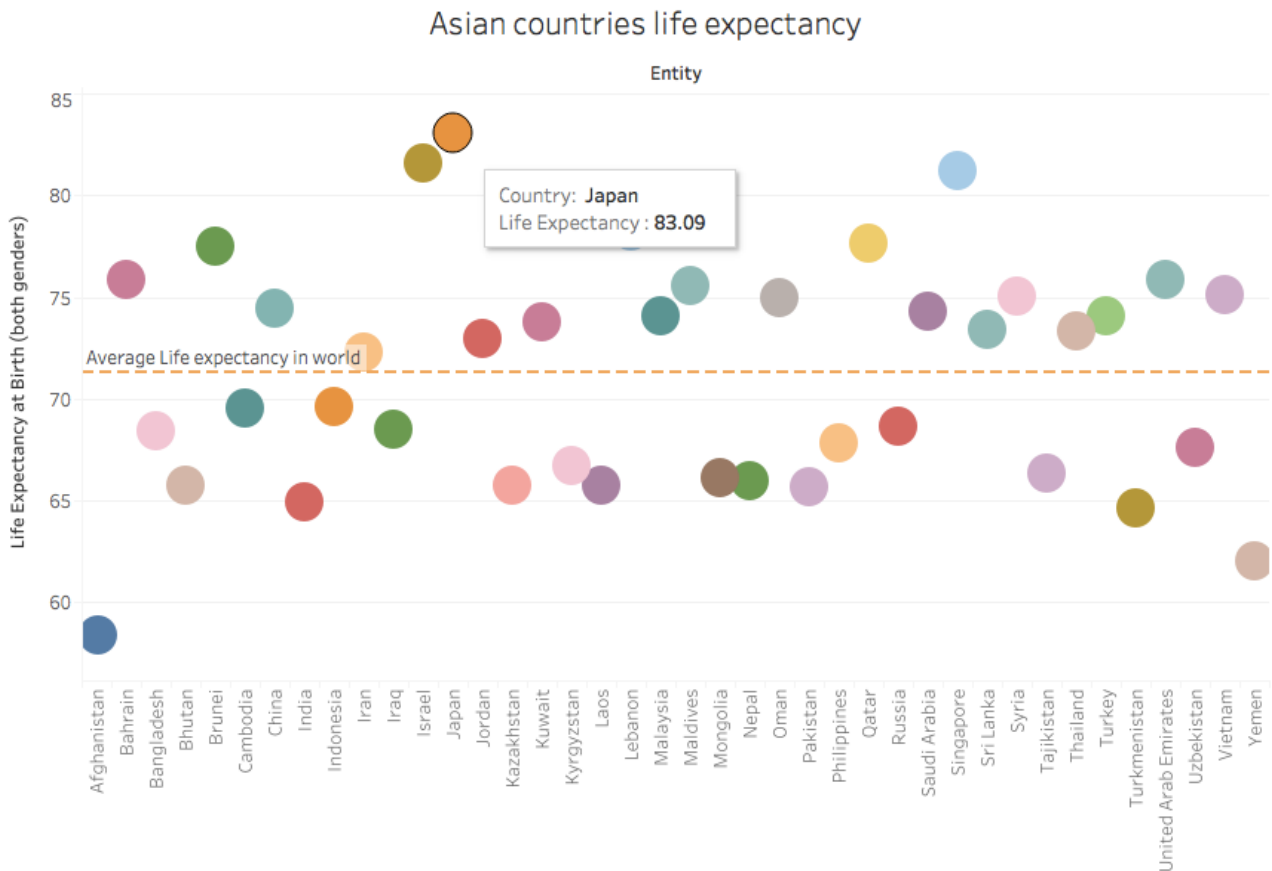
3] US Health care spending's are not very effective

Claim - US is an outlier in health spending with less success in increasing life expectancy[10]

Explanation - The per capita spending on Health care is highest in US. In the last 10 years, the per capita spending's have increased from 4800$ in 2000 to 8500$ in 2010. However, the life expectancy has not increased significantly. It is almost constant at 78 years for last 10 years

From the data, I wanted to find out the most productive/effective health care spending country. This means the lowest money spent per person per year and the highest life expectancy achieved. Now if we consider all the countries of the world, there is a lot of diversity in economic conditions, living standards, food, water, climate and pollution levels.

---

[10] http://economistsview.typepad.com/economistsview/2017/03/life-expectancy-and-health-expenditure.html

So, I have restricted the comparison to only the G7 nations. I considered these nations, since they are all developed countries with comparable access to health care, living standards and economic growth.

Process of creating worksheet -
- Create a new Set of G7 nations
- Create the calculated field "Dollar value of life" which is explained above.
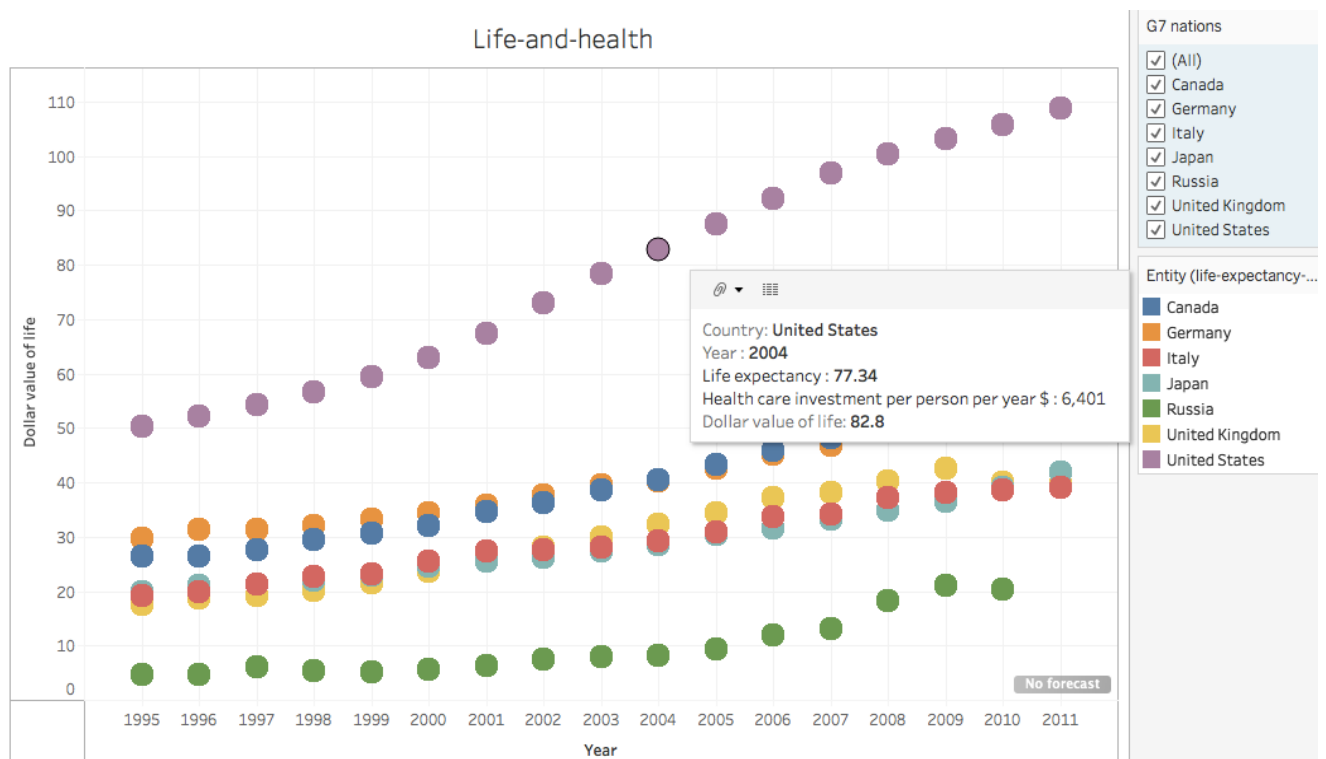
Principles of visualization used -
- Time series (line graph) is used to show the trend in healthcare spending. USA can be easily identified as an outlier
- **Aesthetics** - Effective use of tooltip clearly gives the necessary information. Audience is not kept guessing about "Dollar value of life"

Warrant for the claim - (Articles/ References which support the claim)
http://www.visualcapitalist.com/u-s-healthcare-system-global-outlier-not-good-way/
https://ourworldindata.org/the-link-between-life-expectancy-and-health-spending-us-focus
https://www.brookings.edu/opinions/with-health-care-costs-the-u-s-is-a-huge-outlier/

4] Japanese healthcare system is best in the world

Claim - US healthcare system need lot of improvement

Explanation - I wanted to understand what makes the Japanese healthcare system so effective. Effectiveness of health care system will be dependent on cost, skills of medical staff, convenient locations of hospitals etc. I found a comparison between US and Japanese healthcare system and found out that US lags Japan in almost all the parameters of healthcare system[11]. The only parameter in which US leads Japan is "Modern equipment". This can again boil down why US healthcare system is so costly.

Process of creating worksheet -
- Map different parameters of healthcare index for Japan and USA
- Create table showing comparison of US and Japan

Principles of visualization used
- Complex is not always better - Do not complicate things. Keep things simple. Here a table shows a direct comparison between USA and Japan for different health care parameters. No fancy visualizations are used to show simple information

Warrant for the claim - (Articles/ References which support the claim)
http://www.asaging.org/blog/can-japan-serve-model-us-health-and-long-term-care-systems
http://www.nationmaster.com/country-info/compare/Japan/United-States/Health

Parameter	Japan	United States
Accuracy and completeness in filling reports	86.11	74.00
Convenient location	86.11	80.10
Cost	85.00	45.81
Friendliness and courtesy of staff	83.00	74.00
Healthcare system index	85.00	69.00
Modern equipment	83.00	95.00
Short waiting times	83.00	60.00
Skill and competence of medical staff	87.00	74.00
Speed of delivering examinations and reports	88.00	69.00

---

[11] http://www.nationmaster.com/country-info/compare/Japan/United-States/Health/Quality-of-health-care-system

Challenges

- **Deciding the scope of the project** - Since life expectancy is a big subject and is dependent on multiple parameters, I found it tough where to start my research
- **Finding different data stores which will support my claims** - There was no single dataset which satisfied my claims. I had to read a lot of articles and find data from multiple sources
- **Understanding different parameters in life expectancy** - Since this topic is new to me understanding the parameters was tough at the beginning. I had to read and compare the health care systems of US and Japan for one the visualizations

Critique

- I had to restrict myself to only few selected parameters of life expectancy. I had created other worksheets showing relation between
    - Pollution levels in countries and life expectancy (see Appendix 1)
    - Life threatening events in the world (see Appendix 2)
    However, the claims/relations from these worksheets were deviating from my main story. Because of time constraints, I could not work on these worksheets in more details and so did not include them in final workbook
- I could derive an insight that health care spending per capita in US is the maximum as compared to other G7 countries. However, considering the complexity of health care systems, I was not able to completely understand the cost structure as well as why US cannot reduce the health care costs.
- Availability of abundance of data led me to different directions.
- I could not find much information about **rebuttal warrants**.

Deliverables

GitHub – https://github.com/ameyaghatpande/Rise-of-the-Dragons

Tableau file- https://public.tableau.com/profile/ameya.ghatpande#!/vizhome/Life-expectancy/Dashboard2
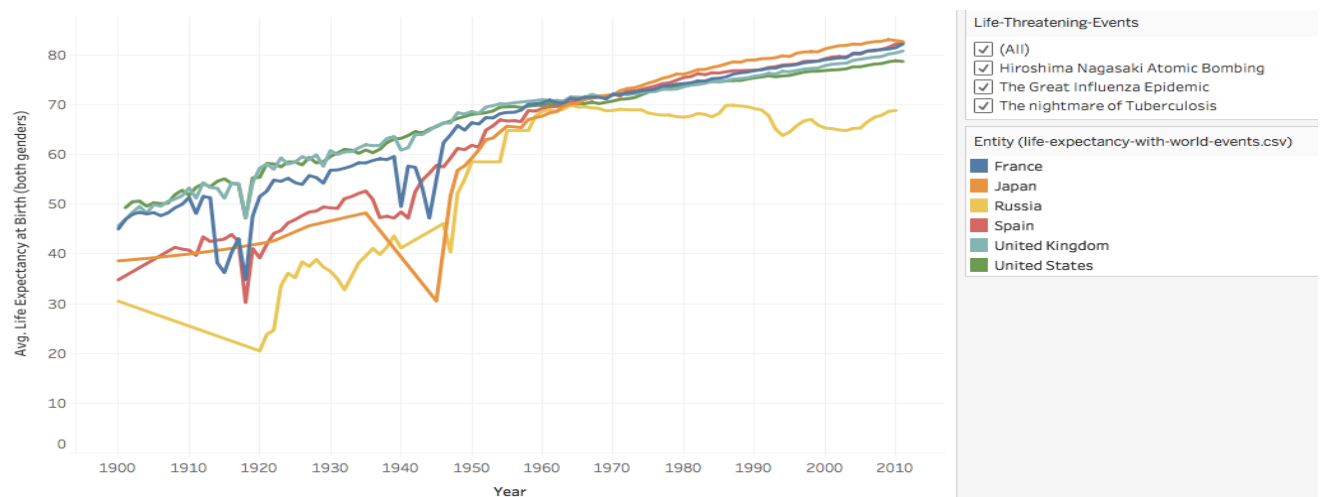
Appendix

Appendix 1] This visualization shows the impact of pollution and life expectancy. I searched the highest least polluted countries and their life expectancy. My analysis was life expectancy

increased over the period of years. Since this visualization deviated the flow of my story, I did not add it in final workbook.
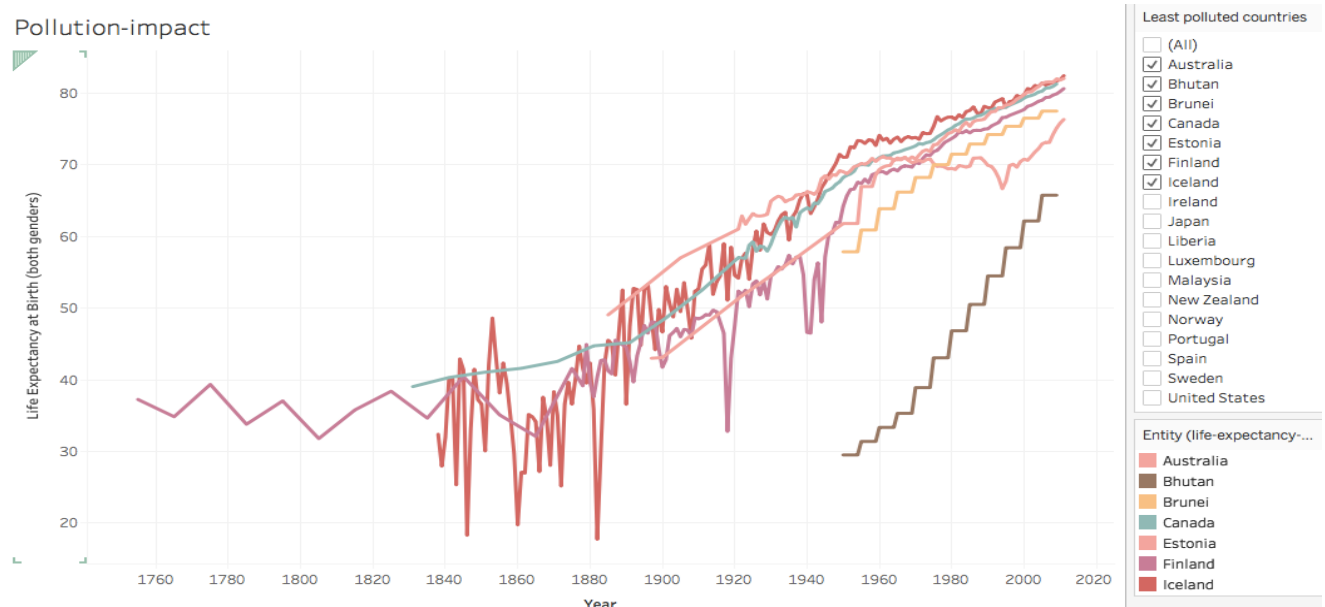
Appendix 2] This visualization shows the impact of wars/epidemics and life expectancy. I searched the major epidemics and world wars and the life expectancy relation. My analysis was the countries which suffered these epidemics/atomic bombings have improved life expectancy rapidly over the period of years. This was interesting story, I did not add it in final workbook because of time constraints and scope of project

Thank you