Factors Influencing Life Expectancy

Abel Mitiku	**Jaiden Atterbury**	**Bethany Quevedo**	**Megan Yam**
Pre-Science (Int. CS)	Statistics: Data Science	ACMS	Pre-science
abemit@uw.edu	jatter41@uw.edu	bquev@uw.edu	meganyam@uw.edu
AD	AA	AA	AC

What is your **topic**?

Write 1-2 paragraphs here explaining your topic, its societal relevance/importance, and why visualization may be a powerful tool to explain the topic.

Our initial plan for this project was to solely focus on environmental topics, in particular our main focus was on air quality around the globe over the years. However, due to the limitations of the air quality datasets we did find, we had to further extend our analyses to air quality as well as life expectancy and the factors associated with it. The main complication with combining these two topics is going to be finding a way to combine the datasets so that they don't act as two independent sources, but can instead be worked with in unison to make for a more comprehensive report.

Life expectancy is arguably the most important topic to the human race, as without life itself no other problems are relevant. In particular, the overall life expectancy of a specific community can be used as a way to gauge the overall health of said community. Finding the causes of trends in life expectancy can lead to better quality of life, reduced mortality rates, and socio-economic benefits such as higher work productivity for a longer period of time, as well as increased innovation. Air quality is also important, as without keeping our planet healthy and safe, humans would have no place to flourish. Air quality also plays a role in life expectancy, which is one of the main research topics this project seeks to explore.

Visualizations are a great way to explore the topic of life expectancy and air quality because they allow readers to better understand trends and what factors are creating these trends. Furthermore, complicated ideas such as regression and modeling can be better understood through visualization than they are through equations and numerical summaries.

Who are your **readers**?

Who are your readers? What are your assumptions about them? Why are they reading your page?

Our readers could be immunocompromised people who are looking for a new place to live in, or want to know if the current area they live in has relatively good air quality. It's important for immunocompromised people to be in an area with good air quality, as inhaling air pollutants can increase susceptibility to respiratory infections, pneumonia, and other serious illnesses. These same people can also use the information provided about life expectancy and immunization rates to further aid their decision on where to live.

What **datasets**?

What dataset(s) do you plan to use? Why are these datasets legitimate/authoritative? Please link to all datasets.

As mentioned above in the topics section, we intend to work with two primary datasets; the first being a life expectancy dataset from Kaggle, which is taken from data provided by the World Health Organization (WHO). The second dataset is an air quality dataset (we are using the Excel Version), which is also sourced from data provided by the WHO. These datasets are legitimate/authoritative as they come from the WHO, which is an organization formed to "promote health," "keep the world safe," and "serve the vulnerable." Based on these goals of the WHO, it seems as if they are a trustworthy source for data related to health and safety.

What are some possible **questions** to address via visualization?

List some possible questions that you think are amenable to being addressed via interactive visualization and why?

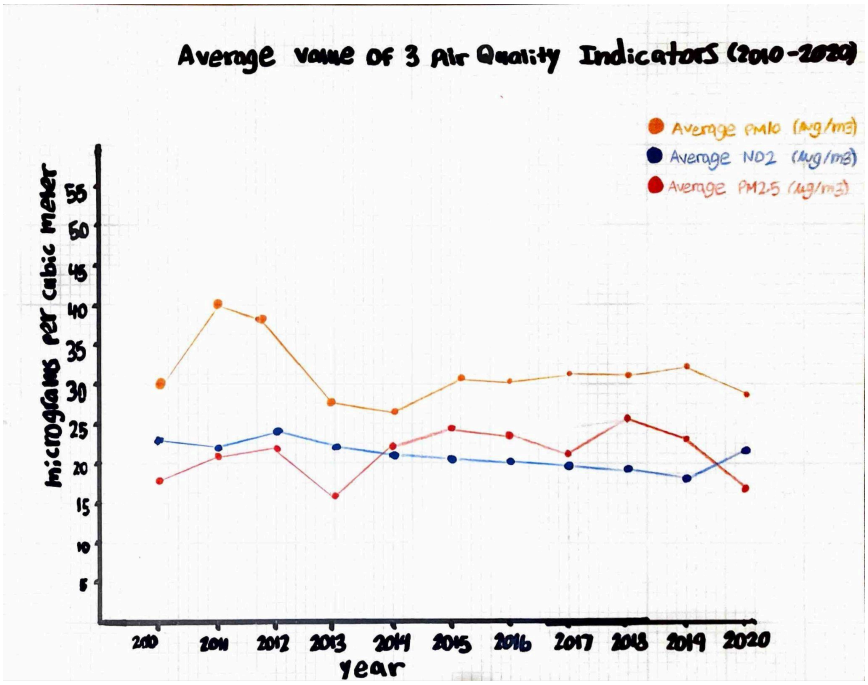
Some possible questions that we could explore through the use of interactive visualizations include but are not limited to:

- What factors/variables (such as certain disease prevalences) can be used to predict the life expectancy of a country in a given year?
- How big of an impact does a country's developmental status and education system have on its overall life expectancy?
- Which regions and/or countries have the highest life expectancy, and has this changed over time?
- Is there a relationship between certain air quality indicators (i.e. PM2.5, etc.) and life expectancy?

These questions are good candidates for being addressed via interactive visualizations as they have many variations of visualizations that can be made from them. What we mean by this is that through interactivity, users can make different visualizations of the same data to answer these questions. For example, in order to answer the question about which factors have a linear relationship with life expectancy, we can make multiple scatterplots of life expectancy versus different factors for several different years, this can be controlled through interactivity such as scroll down menus.

Please include at least two sketches—manual sketches either drawn with paper/pencil or a digital sketching app—to portray initial ideas.

First sketch: Average value of the three air quality indicators between the years of (2010-2020).



Second sketch: Scatterplot of life expectancy versus adult mortality in the year of 2015. This same plot could be replicated for a plethora of different factors, and through the years 2000-2015. This could be done through interactivity.