Final Project

# Introduction

Cancer is one of the leading causes of death worldwide, as per the World Health Organization. Cancer is a generic term for a large group of diseases and can affect any part of the body. Cancer occurs when cells in the body grow rapidly, and without control. The abnormal cells may start invading or spreading to other parts of the body.

There are many types of cancer. Some of the most common cancers are breast, lung, colon, etc. Many factors can cause cancer such as radiation, carcinogens or toxins, genetic factors, or exposure to certain bacteria or viruses. Cancer may be prevented by implementing evidence-based strategies and avoiding risk factors.

Cancer prevention strategies include:

* Not using tobacco
* Maintain a healthy body weight
* Healthy diet
* Exercise
* Avoiding alcohol
* Getting certain vaccines such as HPV and hepatitis B
* Using sunscreen

Many cancers have a high chance of getting cured with early detection, and quality healthcare. Early detection is possible from regular screenings and early diagnosis. Cancer mortality can be reduced from early detection.

A country’s GDP per Capita is an economic metric that can be used to determine the economic prosperity of a country and the financial wealth of citizens. As per Investopedia “Gross domestic product (GDP) per capita is an economic metric that breaks down a country’s economic output to a per-person allocation.” GDP per capita is calculated by dividing the GDP of a nation by its population. A higher GDP per capita of a country may indicate a higher standard of living, and a higher net-worth for citizens of the country compared to countries with a lower GDP per capita.

The usual advice to prevent cancer is to eat a healthy diet, exercise, avoid toxins, and not smoke. I believe having a high salary or net worth may indirectly affect all of those factors. E.g., A person with a high salary, may be able to afford a gym membership which keeps them healthy, and in turn, decreases their risk of cancer. A country with a high GDP per capita may have better healthcare, and easier access to healthcare. The country may have access to the latest research, and the best medicines to cure cancer. Citizens of a country with a high GDP per capita may also be able to afford regular screenings for early detection. I believe citizens of a country with a higher GDP per capita should have lower cancer incidence, and lower mortality from cancer, as opposed to a country with a lower GDP per capita.

I wanted to explore the following research questions:

* What is the cancer incidence in different countries?
* What is the cancer incidence rate in different countries?
* What is the mortality from cancer in different countries?
* What is the mortality rate from cancer in different countries?
* How does a country’s GDP per capita affect the cancer incidence rates and cancer mortality rates?

By analyzing data and creating visualizations, I hope to explore how the GDP per capita of a country affects the cancer incidence and mortality rates. This study can help identify trends that may be useful for healthcare planning, and creating policies in different parts of the world.

# Methodology

For this analysis, I used data from multiple sources to explore the relationship between cancer incidence, mortality rates, and GDP per capita. It is important to note that only data for 2022 was included. Year 2022 was the latest data available for cancer incidence and mortality rates, as well as a country’s population size and GDP per capita. Ideally, data from many different years should have been included, but that would have broadened the scope of this analysis. Some years also had missing cancer data for incidence and mortality, which could have led to an incomplete or inconclusive analysis.

The primary datasets were sourced from:

* World Cancer Research Fund: Provided data for each country on total cancer incidence, cancer mortality, and cancer incidence ASR (Age-standardized rates) across countries.
  + https://www.wcrf.org/preventing-cancer/cancer-statistics/global-cancer-data-by-country/#global-cancer-incidence-both-sexes
* World Health Organization (WHO): Provided total cancer mortality and ASR mortality data for different countries.
  + https://gco.iarc.fr/en
* Worldometer: Provided GDP per capita and population size data for each country.
  + https://www.worldometers.info/gdp/gdp-by-country/

The data was collected from publicly available database sites. Once I downloaded the datasets, I checked them for inconsistencies and missing values to ensure accuracy. I had to review country names across the datasets to ensure they matched. Some countries did not have data for all the metrics analyzed, but those countries were still included to ensure as many countries as possible were analyzed. E.g. Some countries did not have GDP per capita or population size. Countries with incomplete data may have missing data in some of the visualizations or dashboards. I imported the cleaned data into Tableau for visualization and analysis.

To compare the rates of cancer in different countries, ASR incidence and ASR mortality data were used. ASR stands for Age-standardized rate.

* ASR Incidence: The number of new cancer cases, standardized for age, per 100,000 people.
* ASR Mortality: The number of mortalities from cancer, standardized for age, per 100,000 people.

Age is an important risk factor for cancer. According to the National Cancer Institute, the incidence rate for cancer climbs steadily as age increases. Cancer Research UK also indicates that mortalities from cancer increase with age, with the highest mortalities occurring in older people. Countries with a higher proportion of older people would be expected to have the highest cancer incidence and mortality rates. It is therefore necessary to adjust the data for age to do a proper comparison between countries.

Countries with a high population size would also be expected to have a high total cancer incidence and mortality. Example, India is currently the most populated country and should have a high total cancer incidence and mortality. ASR uses a standardized population proportion of 100,000, which allows a comparison of countries with different population size. The ASR data was obtained directly for each country from the data sets, and was not calculated.

Another helpful metric is MIR – Mortality to incidence ratio for cancer. As per the paper “Mortality-to-incidence ratios by US Congressional District: Implications for epidemiologic, dissemination and implementation research, and public health policy” in Science Direct, the MIR can be used to compare cancer mortalities to the number of cancer diagnoses. For this analysis, the MIR was calculated using the formula below.

* MIR = ASR mortality / ASR incidence.

Since the MIR was calculated using ASR mortality and ASR incidence, it will be standardized for age. It would also be expressed in terms of 100,000 people. A higher MIR, or MIR close to 1 would suggest a higher mortality rate and a lower cure rate for cancer incidences.

To analyze the data, I connected the different datasets in Tableau:

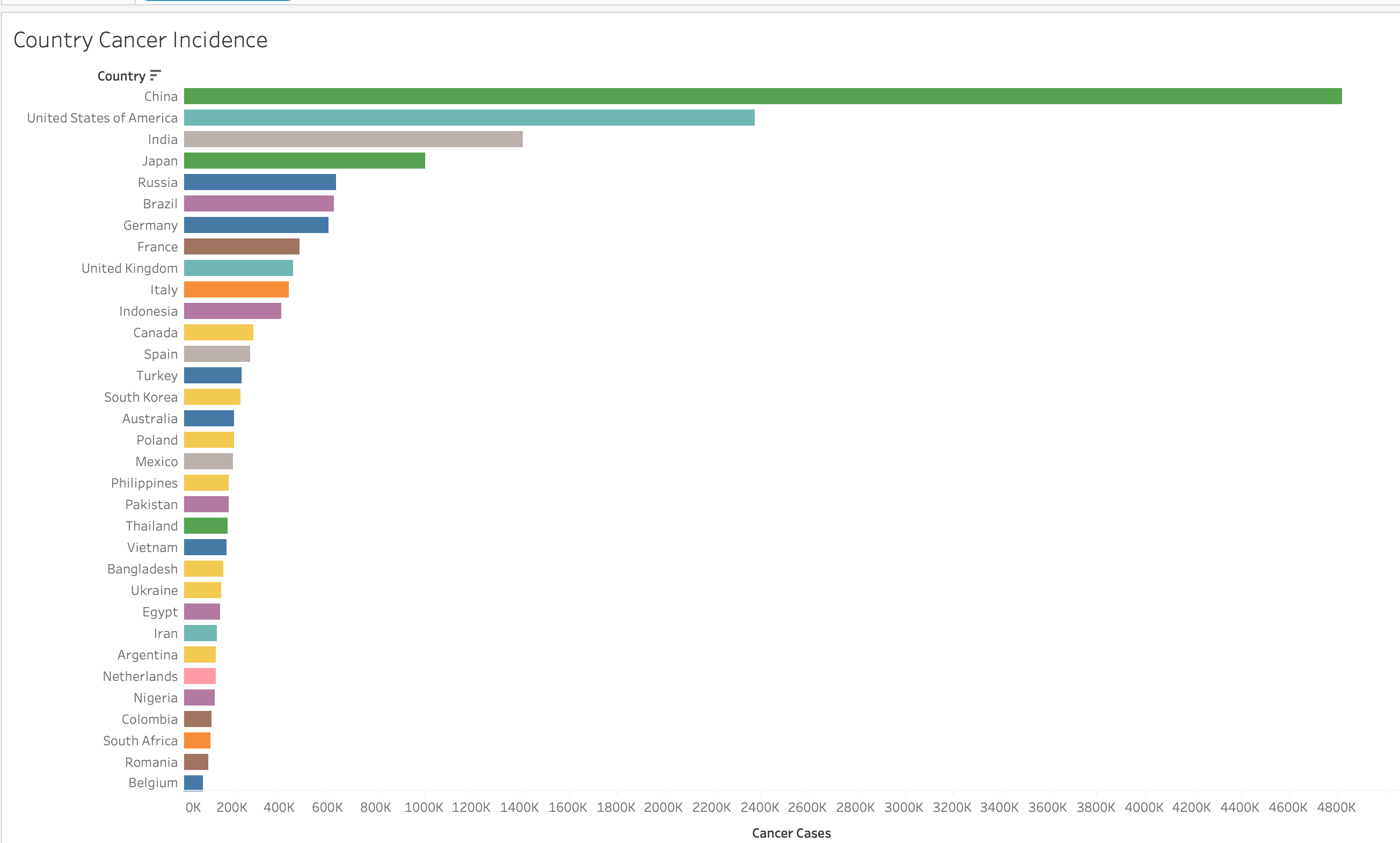
* Cancer incidence data was linked to mortality data using the country field.
* GDP per capita and population data were also matched with the cancer datasets using the country field.

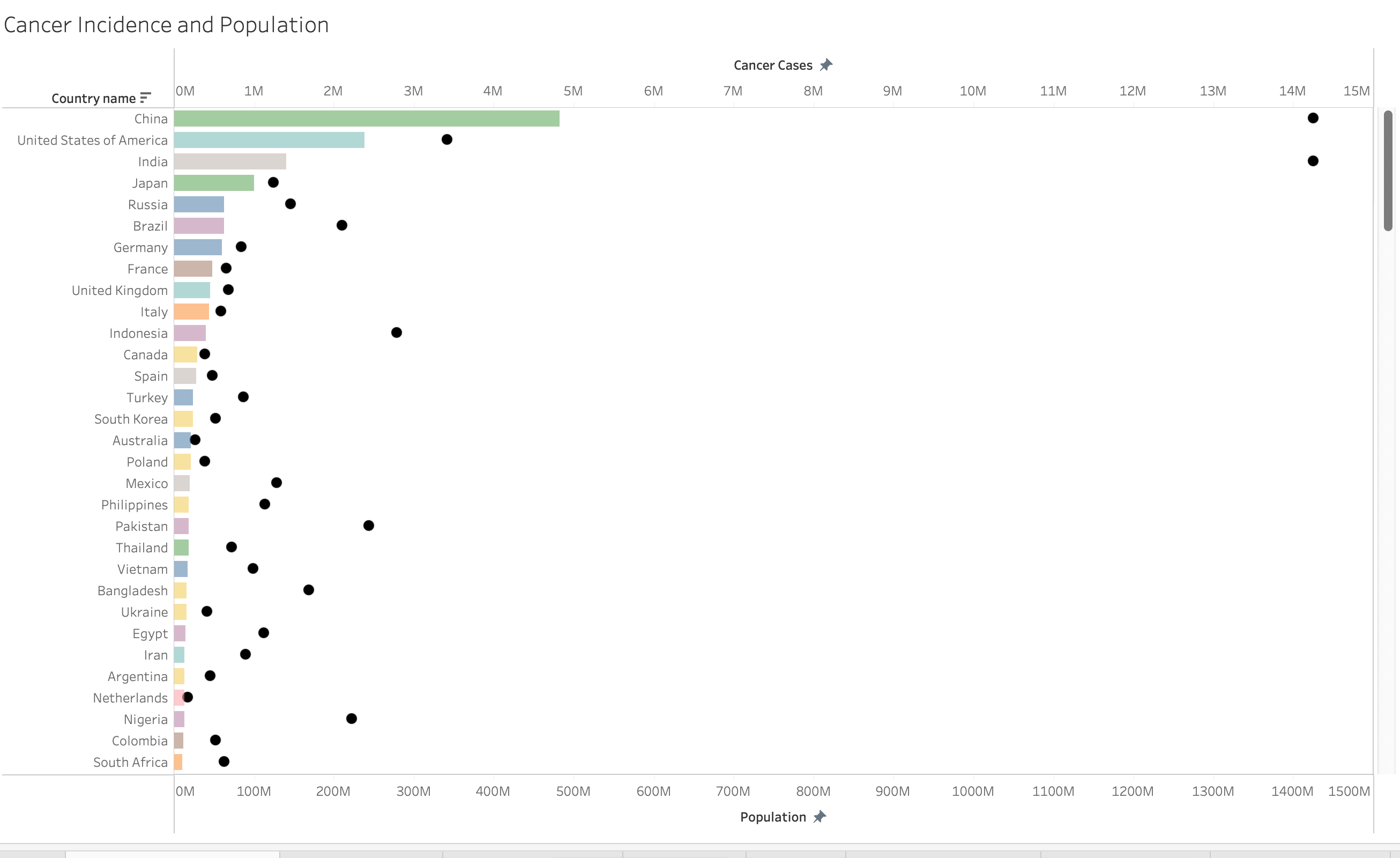
# Analysis

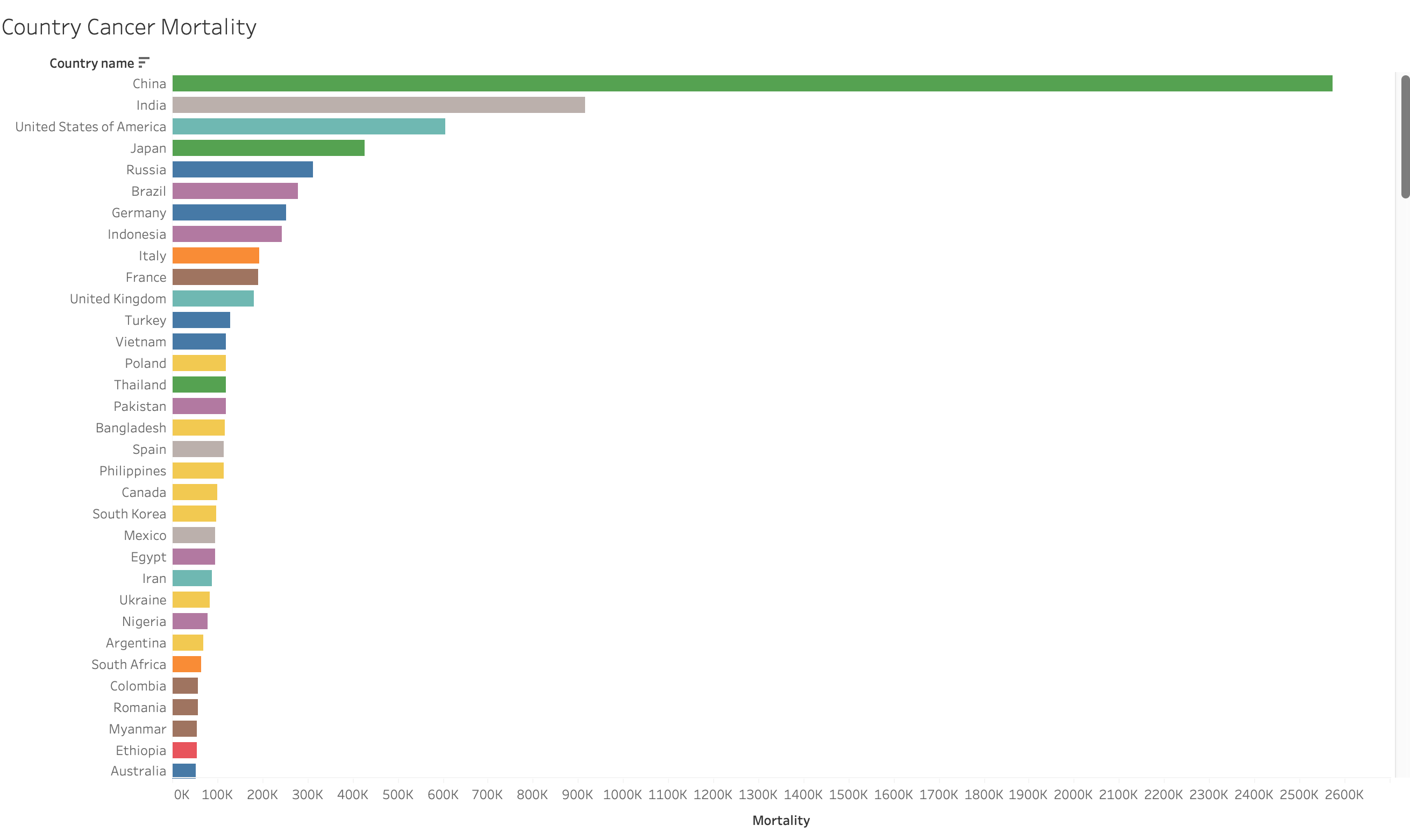
Several visualizations and two dashboards were created to analyze the data and answer the research questions.

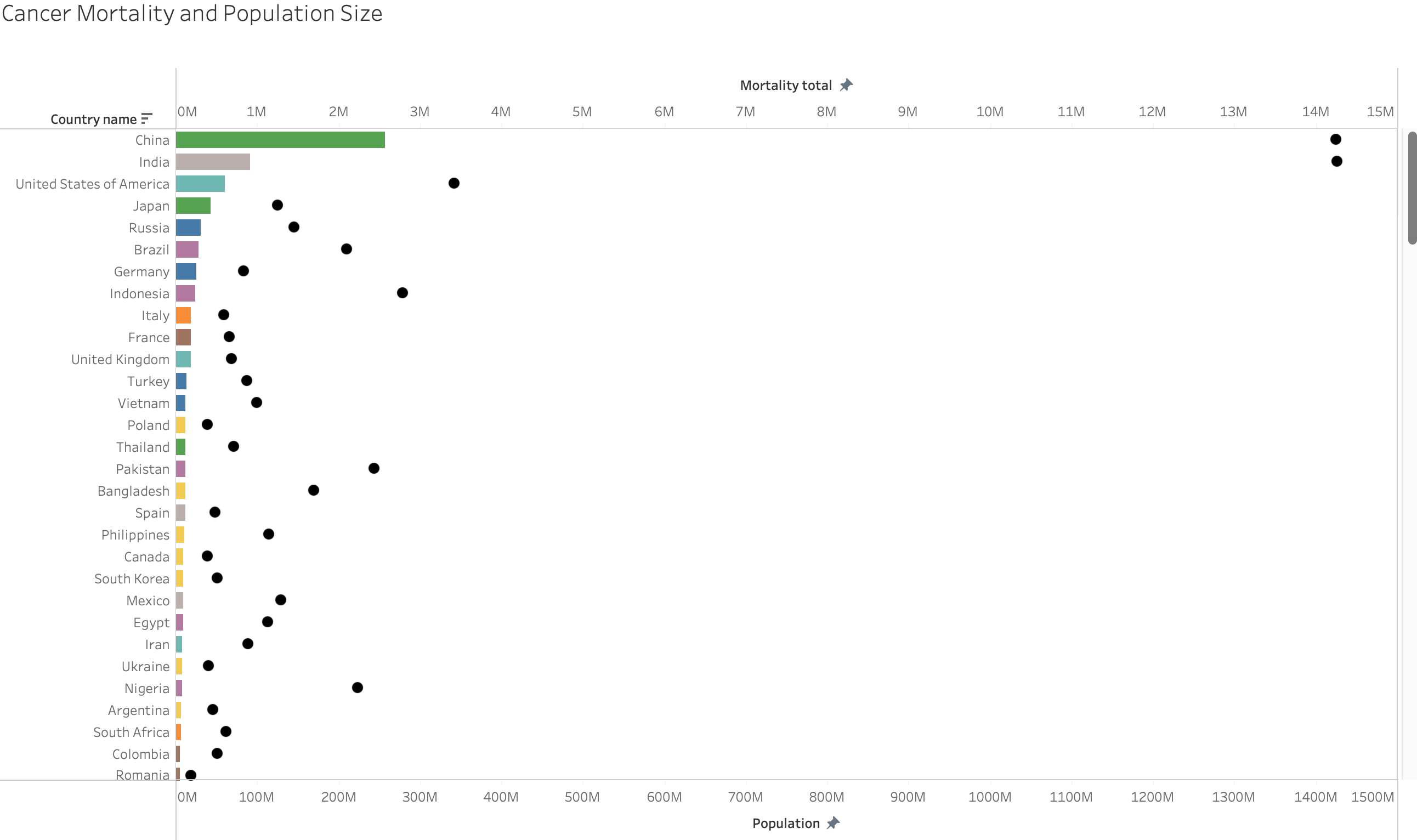
The following research questions were answered using the visualizations below:

* What is the cancer incidence across different countries?
* What is the mortality from cancer in different countries?









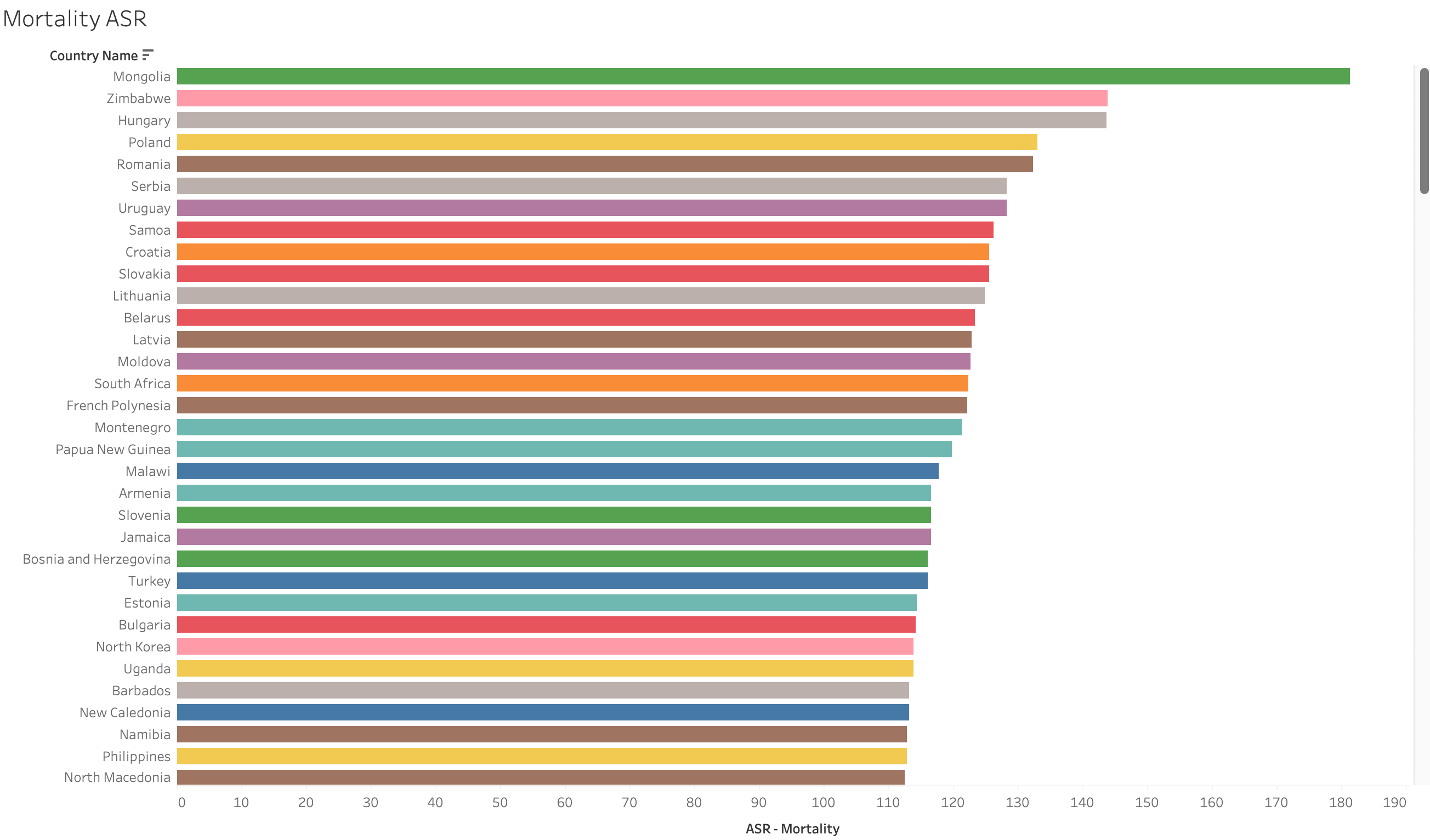
The cancer incidence and cancer mortality data for each country were visualized using bar charts above. A bar chart is an appropriate visualization for comparing the cancer incidence and mortality in each country. I created separate bar charts to show only cancer incidence and mortality totals, and separate bar charts to show cancer incidence and mortality along with the population size. Initially, I only created bar charts for incidence and mortality with population size, but it did not allow a fair comparison between countries as the cancer incidence and mortality total is much lower for each country and the bars were hard to see, especially for countries with low cancer totals.

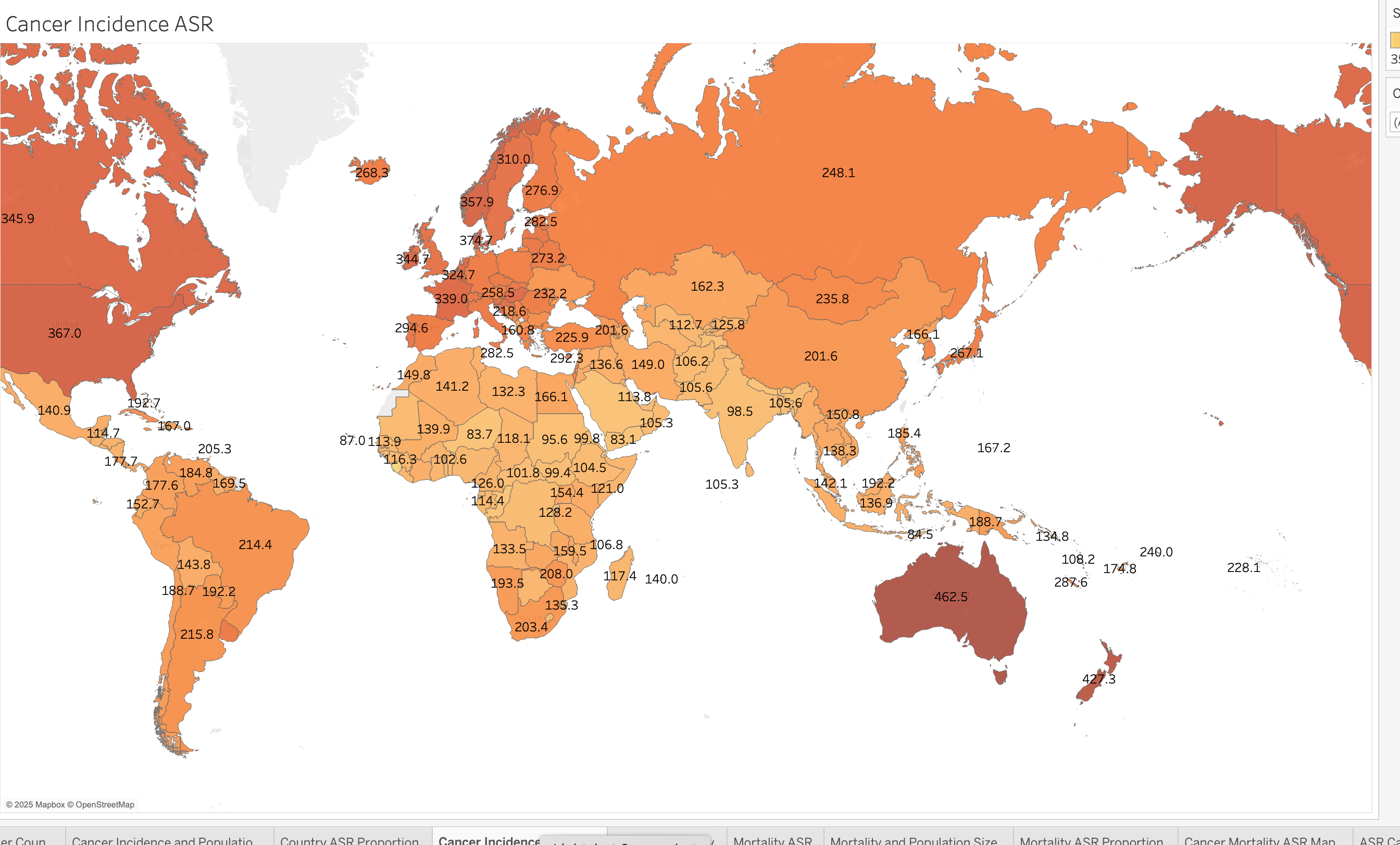
China, USA, and India had the highest total incidence and mortality, but this doesn’t take population size into account. Since these countries have a large population size, high totals would be expected.

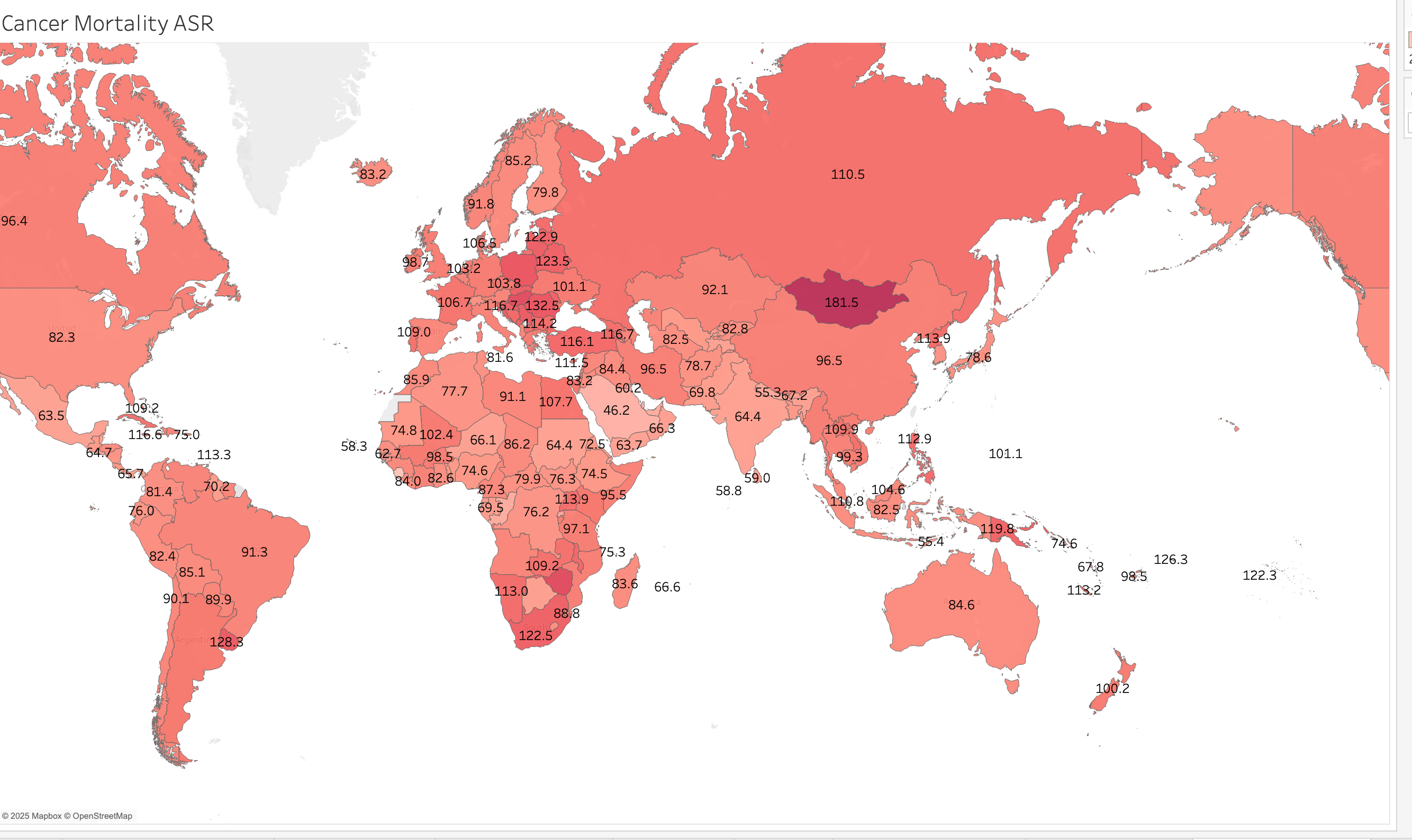
The following research questions were answered using the visualizations below:

* What is the cancer incidence rate in different countries?
* What is the mortality rate from cancer in different countries?







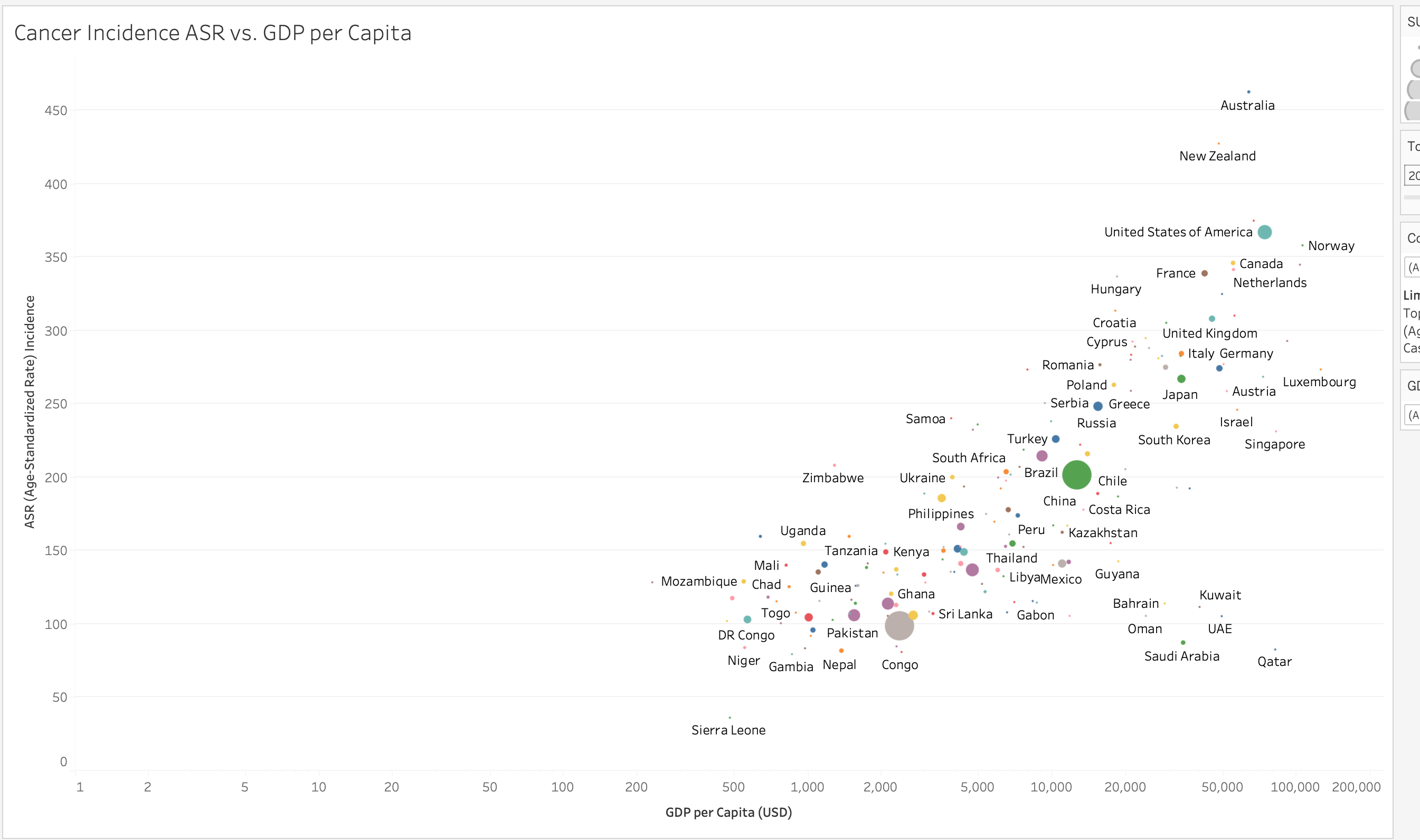


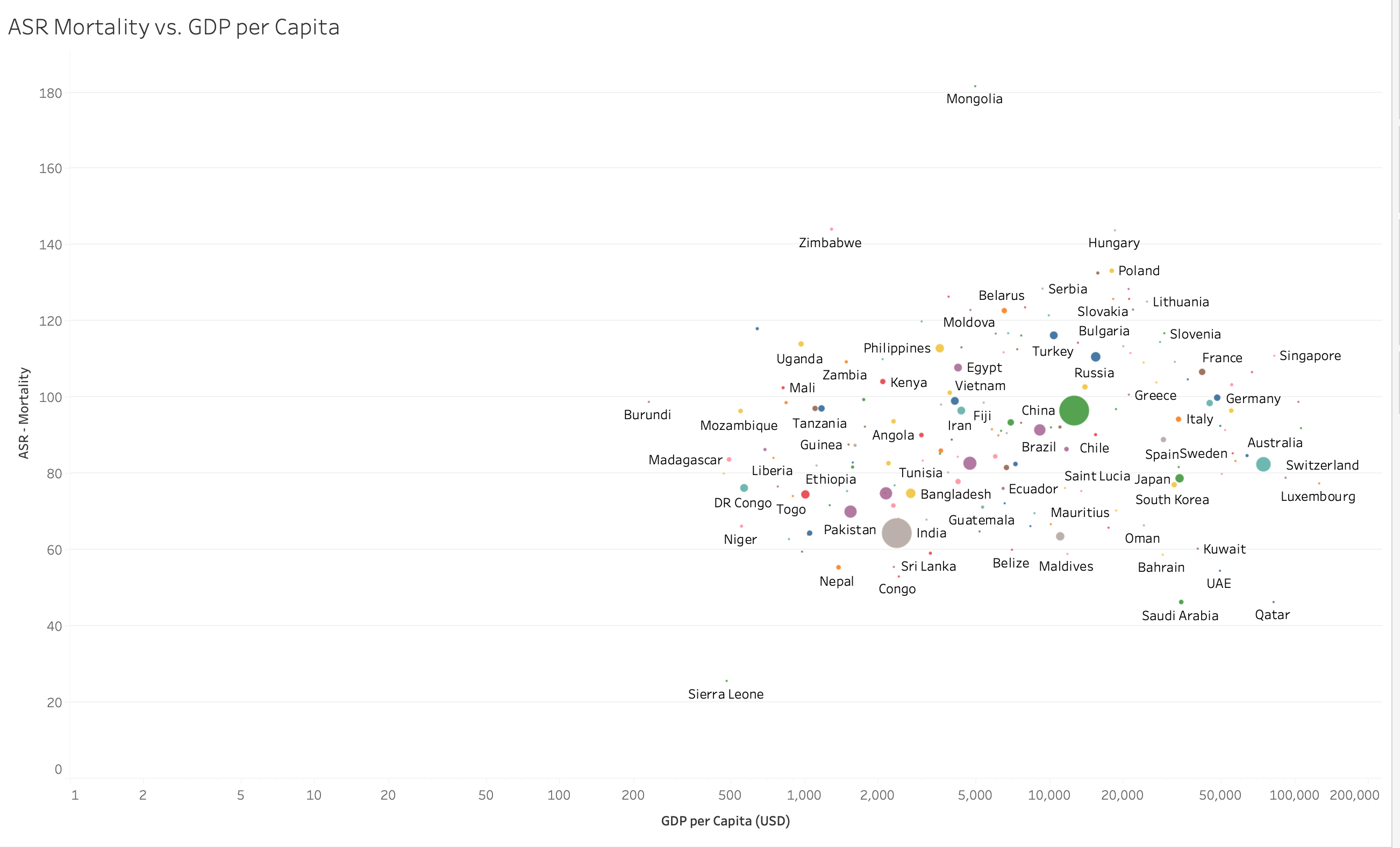
The cancer incidence and cancer mortality rates for each country were visualized using bar charts above. ASR was utilized to compare the rates of cancer. A bar chart is an appropriate visualization for comparing the cancer incidence and mortality rates in each country. The ASR rates were also visualized on a Map to show ASR rates across countries, geographic areas, or continents.

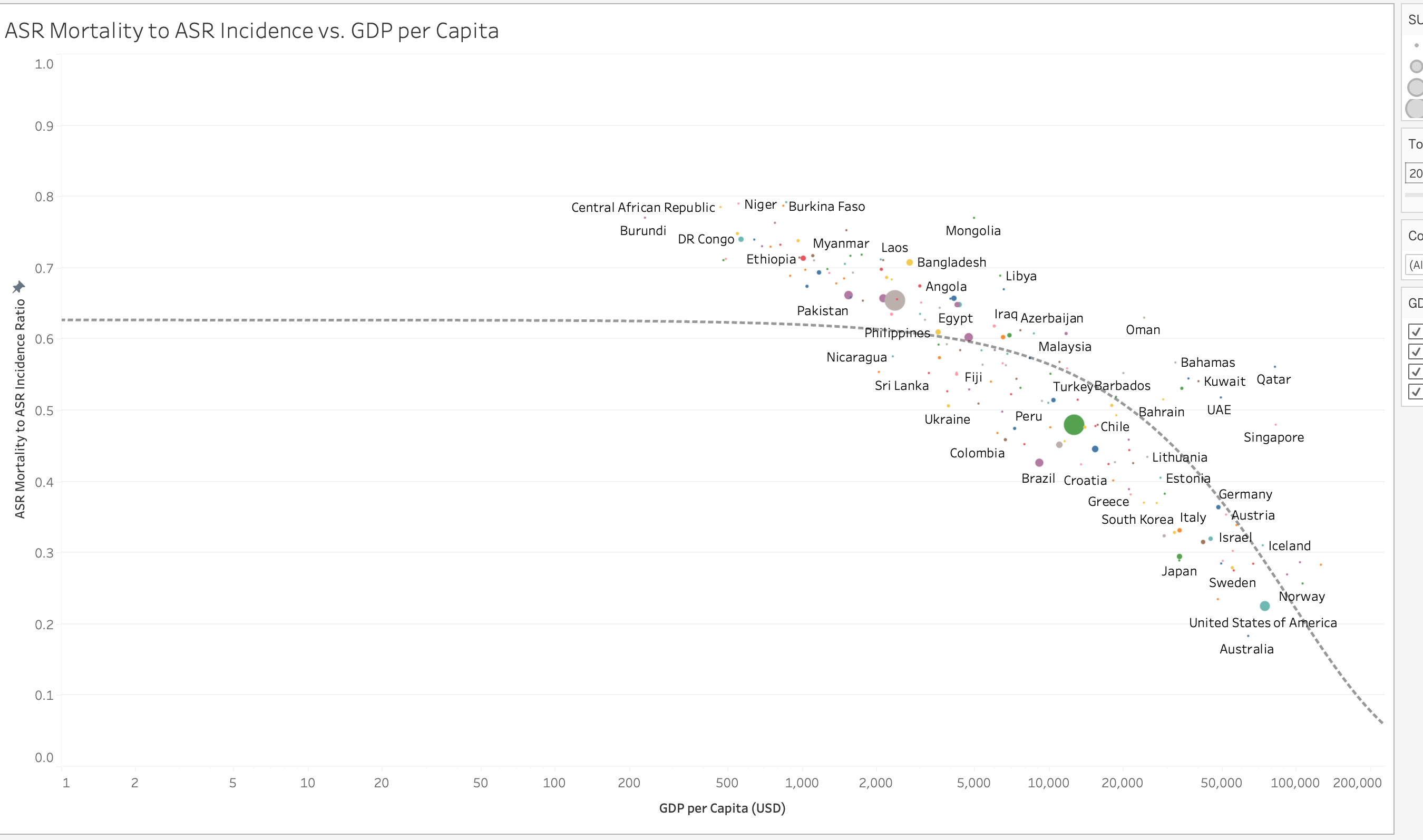
ASR incidence was the highest in Australia, New Zealand, and Denmark. ASR mortality was the highest in Mongolia and Zimbabwe.

The following research questions were answered using the visualizations below:

* How does a country’s GDP per capita affect the cancer incidence rates and cancer mortality rates?







Scatter plots are appropriate visualizations to show relationships, and review for possible trends. ASR for incidence and ASR mortality were plotted against GDP per Capita for each country on the scatter plots. MIR was plotted against GDP per capita for each country as well on a scatter plot.

A trend line was not added in the scatter plot for ASR incidence rate vs. GDP per Capita of a country, as the data points were spread out. The data points did seem to suggest a trend where the Cancer incidence increased when the GDP per Capita of a country increased, but some outliers did not follow that trend. Qatar had a low ASR incidence but had high GDP per Capita, as compared to Australia, which had a very high ASR incidence with a high GDP per Capita. A deeper statistical analysis could be performed to look for a possible trend and add a trend line, but I refrained from doing that as my experience in statistical analysis is minimal.

A trend line was not added in the scatter plot for ASR mortality vs. GDP per Capita of a country, as it seemed the data points were distributed across a large area. Mongolia seemed to be an outlier with a high ASR mortality compared to other countries. This scatter plot did not suggest a trend for the ASR mortality as the GDP per Capita of a country changes.

A trend line was added in the scatter plot for MIR vs. GDP per Capita. The data points on that scatter plot, along with the trend line, suggest a possible relationship with economic prosperity and lower cancer mortality. The MIR seemed to decrease as the GDP per Capita of the country increased.

The Global Cancer dashboard, and the GDP per Capita impact on Incidence ASR and Mortality ASR dashboard were created to address all the research questions. The dashboards are appropriate to quickly filter data for each country, or view the data for all countries.

# Conclusion

Based on the analysis, there were some important trends uncovered.

* There seemed to be a trend where the ASR incidence increased as GDP per Capita of a country increased. My initial hypothesis was that wealthier people may be able to live a healthier life, and thus have a lower cancer incidence rate, but that does not seem to be the case. The ASR incidence could be higher in wealthier countries as people may have better access to healthcare and regular screenings of cancer, thus leading to a higher reporting of cancer. People in countries with a lower GDP per capita may have lower access to cancer screenings, and cancer may never get reported, thus incorrectly showing a low ASR incidence in those countries.
* The relationship between ASR mortality and GDP per Capita of a country was not very clear. Mongolia had a very high ASR mortality rate, but it is a middle income country. Singapore and Qatar have a high GDP per Capita, but Singapore ASR mortality was much higher than Qatar.
* When reviewing the MIR data against the GDP per Capita of a country, there seemed to be a clear trend. The MIR decreased as GDP per Capita of a country increased. This suggests that wealthier countries may better be able to treat cancer and prevent mortalities from cancer once diagnosed. Those countries may have the best healthcare systems, and may be utilizing expensive treatments and drugs to decrease mortality. Countries with a higher GDP per capita may have better access to healthcare, leading to early detection of cancer and thus decreasing cancer mortality as well.
* Countries with a lower GDP per Capita seemed to have both high cancer incidence rates and high mortality rates, which may indicate that people in those countries may not be able to lead a healthy life due to financial pressures, and challenges in cancer treatment and early detection. Zimbabwe, Uganda were examples of that trend.

While there were some trends as discussed above, further research is needed. This analysis only included data for the year 2022. The cancer rates and mortality rates could change as data from more years are included. As more and more years pass, new treatments and drugs are available for cancer, possibly decreasing mortality. The GDP per Capita of a country may change over time as well, allowing to further explore the relationship between Cancer incidence and mortality rate vs. GDP per capita of a country.

Cancer incidence and mortality rates may also be affected when further variables are added such as gender, types of cancer, genetics of a population, or geographical area of a country. E.g. Countries located in areas with high sunlight UV radiation could have a higher rate of skin cancer. Some cancers may be more prevalent depending on the gender, such as breast, or lung cancer. Future research could account for these variables as well.

The quality of healthcare, access to healthcare, and whether healthcare is free or paid may affect the cancer incidence rates and mortality as well. A country with a low GDP per Capita may have free government-sponsored healthcare, but the quality of the healthcare could be worse, which may increase the mortality rate.

Future research could explore these questions:

* How does the cancer incidence rate and mortality rate change when compared to the GDP per Capita of a country for different years?
* How do variables like gender, type of cancer, genetics of a population, or geographical area of a country affect cancer incidence rate or mortality rate when compared to GDP per capita of a country?
* How does access to healthcare, and the quality of healthcare, affect cancer rates and mortality rates when compared to GDP per Capita of a country? What is the relationship, if healthcare is free or paid?

By investigating these factors further, policymakers and healthcare organizations can develop strategies to improve cancer care and reduce mortality worldwide. I hope this analysis was helpful to explore the relationship between cancer and GDP per Capita of a country.