

Mortality Rate Due to Various Types of Cancer in Different Countries

Submitted By :

Ankita Kukrety

Prithviraj Bhosale

The dataset we used for our project is from ourworldind.org. The Cancer dataset which we have chosen is the perfect mixture of measure and dimensional variables. Moreover, it would give us a lot of options to visualize through various plots. There are 38 columns and 6469 rows in the dataset we selected, with most of the data being repeated for different years across different countries. With the given dataset, we can see which types of cancer deaths have affected all groups the most. Moreover, we have merged the datasets of people affected by neoplasm disease as well which is divided by certain age groups. The age groups are under 5, 5-14, 15-49, 50-69 and 70+. Although there are 38 columns, there are variables that are of interest to us as we visualize the data.

1. Prostate Cancer
2. Pancreatic Cancer
3. Neoplasm Age 0-5 Rate
4. Neoplasm Age 15-49 Rate
5. Neoplasm Age 5-14 Rate

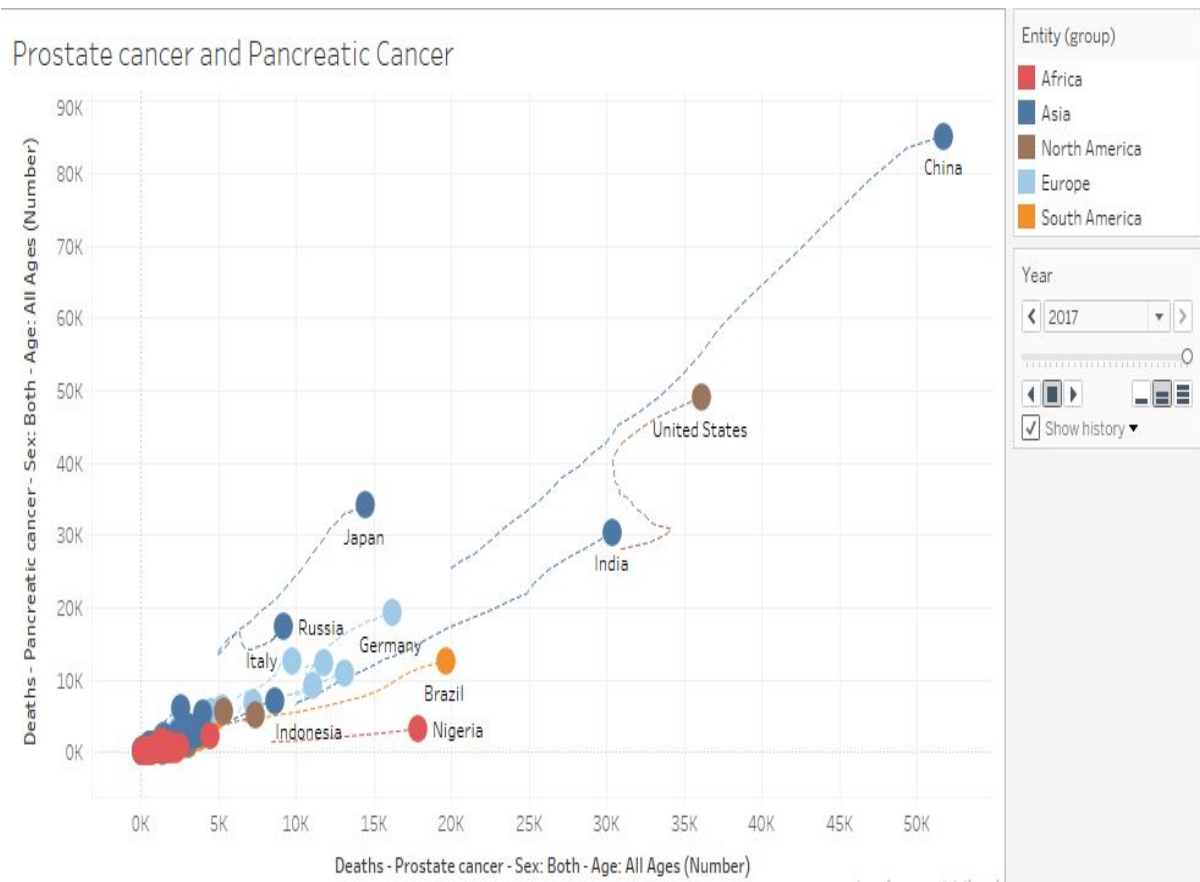
Moreover, there are other variables which interests us like the death by neoplasms which divided by age groups.

The variables which we have are in form as

Death by cancer-: Numerical Form

Deaths by Neoplasm Diseases-: Percentage Form.

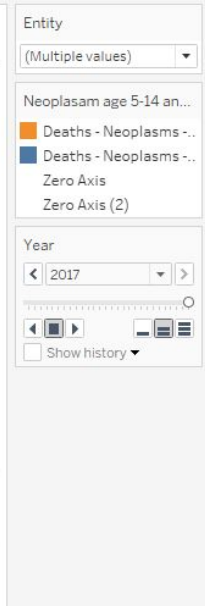
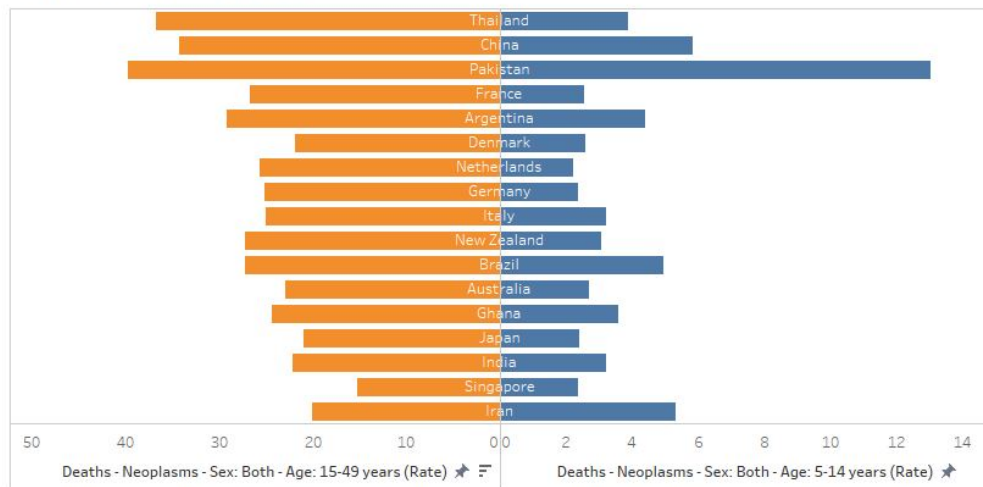
Connected Scatterplot



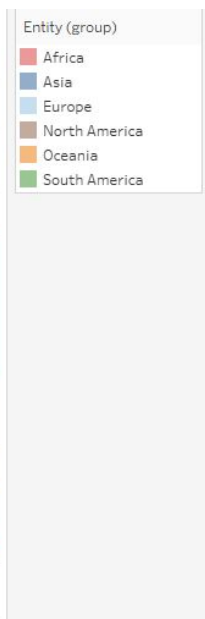
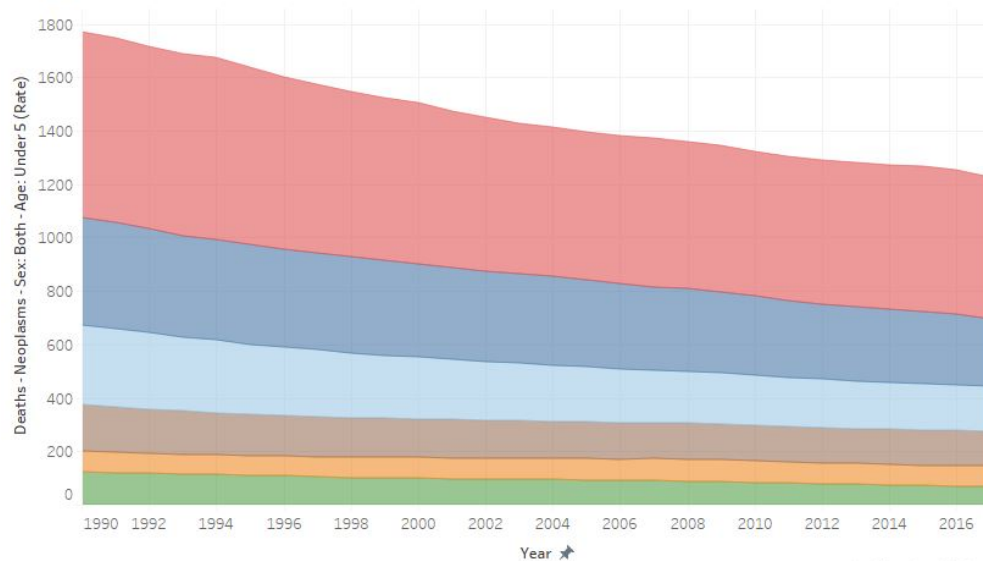
Prostate Cancer and Pancreatic Cancer is caused due to mutations in the BRCA2 gene. The prevalence of prostate cancer and pancreatic cancer is high in the US, China, and India, so it is likely that those with mutations in the BRCA2 gene are also common in those countries. The death rates in Africa are low, with the exception of Nigeria, which might be because it is not as populated as the others.

Neoplasm Plot

Neoplasam age 5-14 and 15-49 - 2017

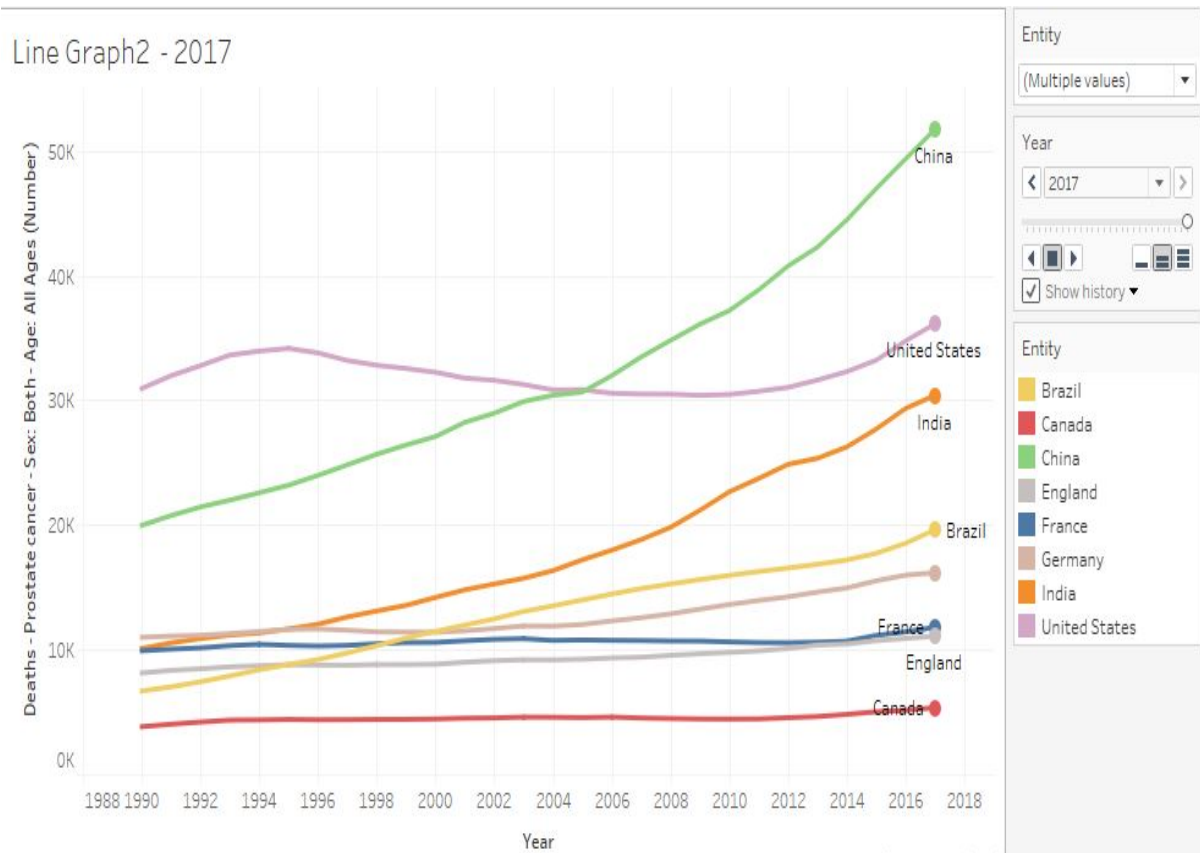


Neoplasam age 0-5



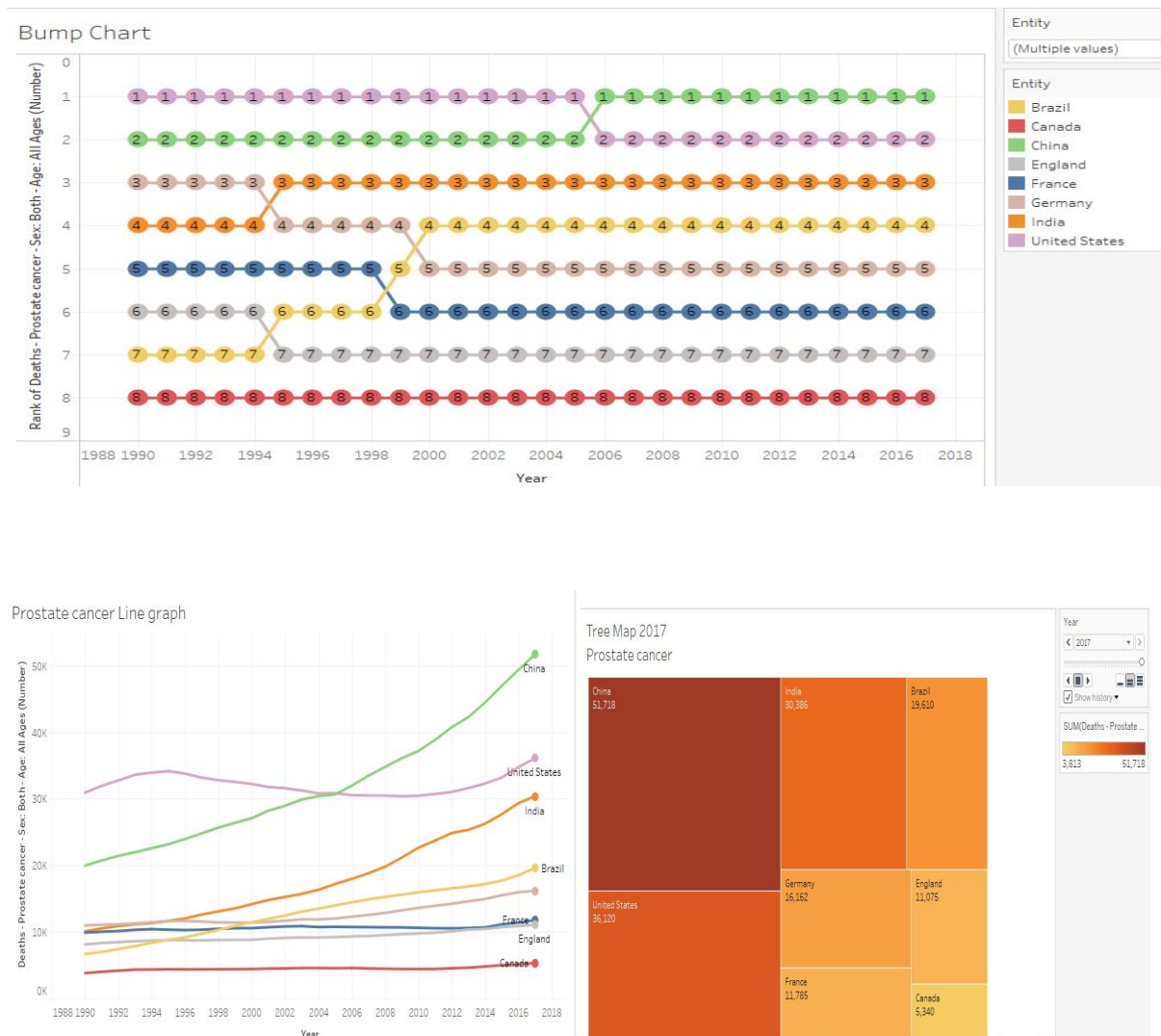
From the above visualization, we can visualize that the mortality rate in Pakistan is high in both categories, so from this we can infer that the infrastructure might not be great. The low death rates in India across both categories may be a result of good infrastructure resources. In the Area chart, Africa has the highest death due to Neoplasm under the age of 5 but, the deaths significantly decrease as the years go by probably due to the development of infrastructure. A possible reason that neoplasm deaths are decreasing all over the world could be that we are detecting it at an earlier stage, giving us a better chance of curing it.

Line Graph



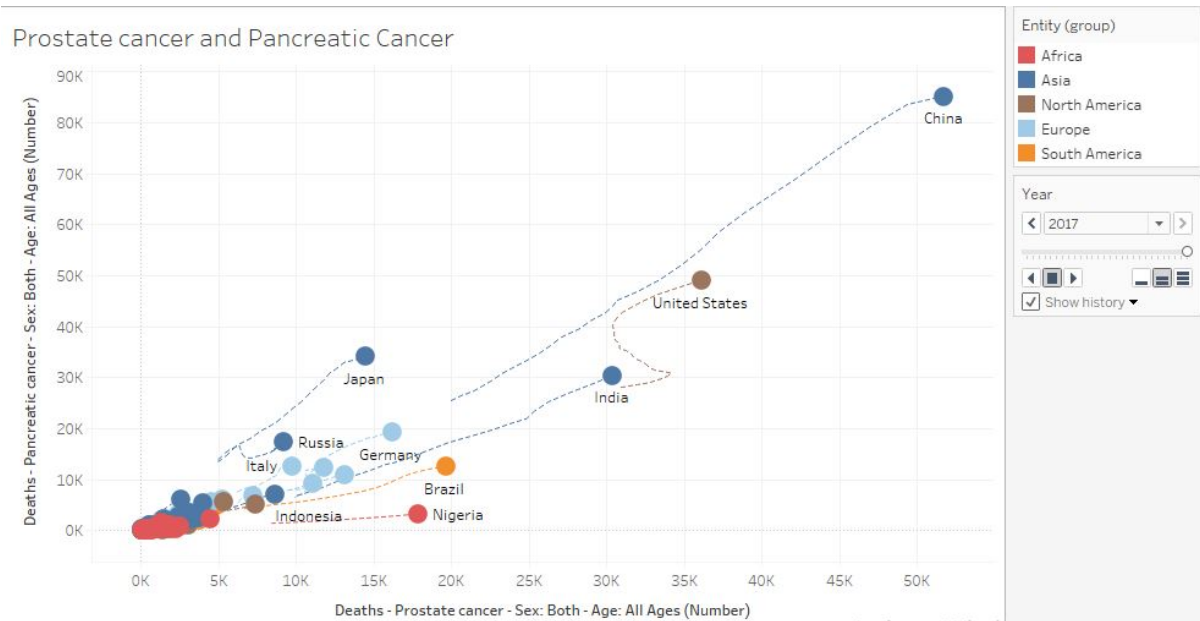
Compared to other countries, China is significantly higher, and it has also increased substantially since 1990. Deaths due to prostate in United States almost remain the same over the period. Canada has the same deaths over this period.

Bump Chart and Tree Map



The graphs are created using data from the Prostate Cancer dataset. Here we have made a bump chart of 8 countries, and they all have been assigned ranks. The United States starts with rank 1 but changes its ranking after 2005. Initially, China ranks second, then moves up to rank one. Brazil begins at rank 7 but climbs up from rank 7 to rank 4 in the span of 6 years. Brazil is one of the highest populated country in the world. In addition, India moves up to 3rd from 4th. Looking at the tree map we can see that in the year 1990 deaths due to Prostrate Cancer is relatively low in China as compared to United States. But, as the years go by, the deaths in China are higher than the US in 2017. We can assume this is due to an increase in population. The number of deaths in Canada is a lot lower compared to other countries which might be due to great medical infrastructure.

Pancreatic Cancer



Tree Map

Death due to pancreatic cancer



Here we have created the graphs based on continents from the pancreatic datasets. Due to its 2/3rds share of the world's population, Asia has the highest number of deaths, while Europe ranks second.

Insights

1. Comparing pancreatic cancer and prostate cancer across all countries, we observe that China has the highest death rate.
2. There are more deaths in China, India, Japan, and the United States than any other country.
3. Pancreatic and Prostate cancer are caused due to mutation in the BRCA2 gene, since China, India, Japan, and United States have a large population the chances of them having a high number of mutations is also high.
4. The Area chart indicates that Africa has a high number of Neoplasm deaths compared to other continents.
5. Additionally, Asia has a high death rate.
6. From 1990 to 2017, the number of deaths decreased significantly in almost all areas.
7. Due to Pakistan's high rates of both types of Neoplasm, we can infer that its medical infrastructure may be lacking.
8. Due to good medical infrastructure, India's death rates from both categories are low compared to its population.
9. The death rate in China was low compared to the United States, but eventually they overtook the United States.
10. Among the pancreatic cancer plots, Asia has the highest number of deaths due to India, China, and Japan, and Europe is second.