ANALYSING CANCER TRENDS

FINAL PROJECT REPORT- DATA WAREHOUSING AND BUSINESS INTELLIGENCE

CONTENTS

NO.	TOPIC
1	INTRODUCING THE PROBLEM
2	WHERE DID THE DATA COME FROM?
3	DATA CLEANING AND DATA PRE-PROCESSING
4	ANALYSIS OF CANCER PREVALENCE/MORTALITY WORLD-WIDE
5	ANALYSIS OF CANCER PREVALENCE / MORATLITY IN THE UNITED STATES
6	CANCER AND ANNUAL INCOME – WORLD-WIDE TREND
7	CONCLUSION
8	CHALLENGES FACED
9	OUR AUDIENCE
10	DASHBOARDS

THE PROBLEM

Developing cancer is now more common than getting married or having a first baby, research suggests. We are holding a tough war against cancer. Troubling as this sounds, Cancer is, by far, a harder problem — a condition deeply in-grained in the nature of evolution and multicellular life. Given that obstacle, cancer researchers are fighting and even winning smaller battles: reducing the death toll from childhood cancers and preventing — and sometimes curing, cancers that strike people in their prime. But when it comes to diseases of the elderly, there can be no decisive victory. This is, in the end, a zero-sum game. Measuring from 1990, when tobacco had finished the worst of its damage and cancer deaths were peaking, the difference is somewhat less pronounced: a decline of 44 percent for heart disease and 20 percent for cancer. The diseases that once killed earlier in life like the bubonic plague, smallpox, influenza, tuberculosis — were easier obstacles. For each there was a single infectious agent, a precise cause that could be confronted. Even AIDS is being managed more and more as a chronic condition. But Cancer is a complex disease that can grow out of a variety of risk factors, which makes it as difficult to manage as it is to explain.

Through the medium of visualization using tableau we understood the prevalence / mortality rate of different types of cancer(like liver, larynx, kidney, breast, thyroid, Stomach, Uterine, Ovarian, Bladder, Prostate, Cervical, Pancreatic, Esophageal, Testicular, Nasopharynx, Colon & rectum, Non-melanoma skin cancer, Lip & oral, Brain & nervous system, Tracheal, bronchus, & lung cancer, Gallbladder & biliary tract cancer) for the population all over the world over a 25-year period. We also gained insights about trends of different types of cancers over a period of time, country-wise cancer trend, all this over a period of 25 years, world-wide. Finally, a study of cancer in the United States in specific was also done. Same kind of prevalence /mortality analysis, and trend analysis was done state-wise for a period of 25 years. As a final result of these visualizations, we have a better understanding of which are the top 5 countries/states which have highest rates of cancer prevalence & mortality, which cancers are most common, what is the cancer picture with respect to United states and which are the most life-threatening cancers in the US.

WHERE DID THE DATA COME FROM?

- Data Sources - Center for disease control and prevention, International Agency for Research on Cancer (IARC).
- The world-wide data has country wise statistics. The US data has state wise statistics. Both these data sets have data related to around 25 different cancer types.
- Data features included –
    - World-wide & United states data – both had details specific to race, age, gender, prevalence of different cancers, mortality associated with the different cancers. We are specifically interested in the mortality and prevalence of the different cancers in different countries/states.
- Data Spread – 1990 – 2015
- All the data used considers data for both men and women.

- All the data has been standardized as a Rate per 100,000 total population to handle the population differences among the different countries & states.
- It was a large volume of data with more than 10,000-30,000 rows.

DATA CLEANING AND DATA PRE-PROCESSING

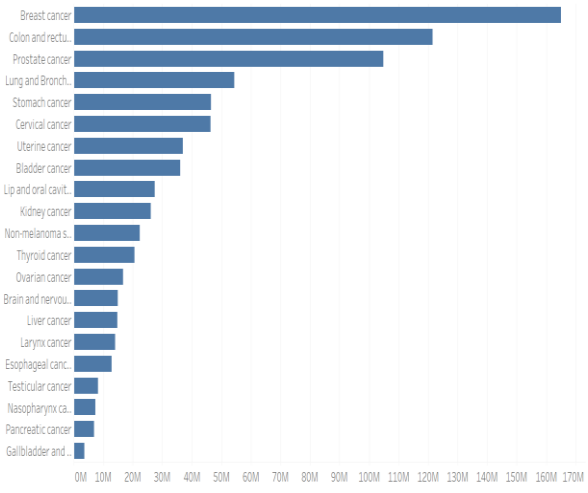
Filtered the data to include only the mortality and prevalence data for all genders, for all countries and states. Added a calculated field to calculate the death rate and incidence rate to standardize the population as a rate of 100,000 total population.

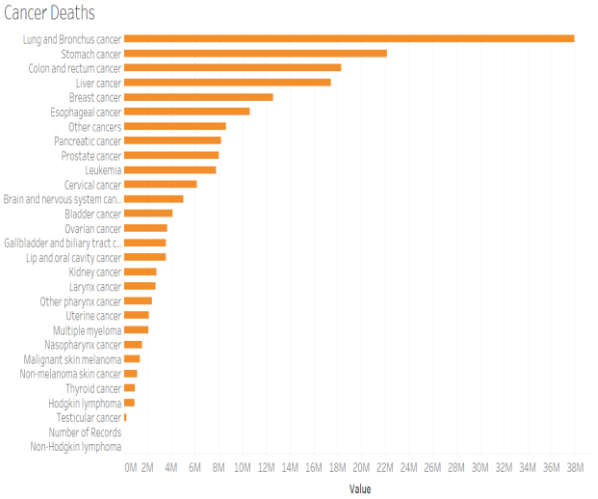
Since the data contained information for all the countries/ states and 25 different varieties of cancers, we decided to study in detail about the top 5 cancers/countries/states only in order to provide a qualitative analysis. This helped understand the most notorious cancers better.

ANALYSIS OF CANCER PREVALENCE/MORTALITY IN THE WORLD

Cancer Prevalence in the World –

- For all countries
- For all Years
- **Top 5 Cancers(Prevalence)** – Breast Cancer, Colon Cancer, Prostate Cancer, Lung Cancer, Stomach Cancer
- **Top 5 Cancers(Mortality)** – Lung Cancer, Stomach Cancer, Colon Cancer, Liver Cancer, Breast Cancer





From the dashboard it can be observed that for the top 5 cancers,

These are the top 5 countries with the highest prevalence of these cancers.

- Breast Cancer – United States, China, India, Germany, Russia
- Colon and Rectum Cancer – China, United States, Japan, Germany, Russia
- Prostate Cancer – United States, Germany, China, Italy, Japan
- Lung & Tracheal Cancer – China, United States, Japan, Germany, Russia
- Stomach Cancer - China, Japan, India, Russia, United States

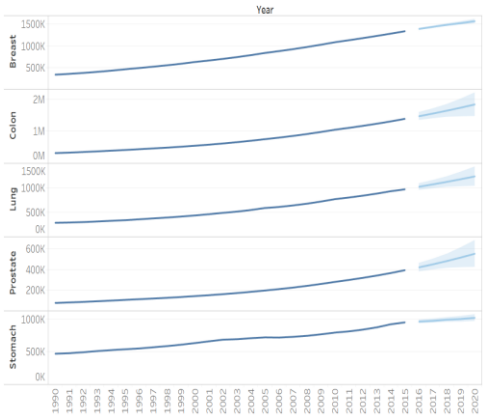
These are the top 5 countries with the highest Mortality due to these cancers.

- Breast Cancer – China, United states, India, Russia, Germany
- Colon and Rectum Cancer – China, United States, Japan, Russia, Germany
- Liver Cancer – China, Japan, India, United States, Nigeria
- Lung & Tracheal Cancer – China, United States, Russia, Japan, Germany
- Stomach Cancer - China, Japan, India, Russia, United States

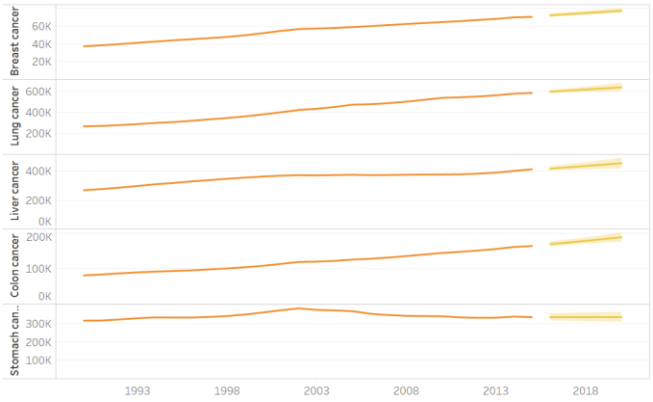
Taking a closer look at the different cancer prevalence/mortality trends in different countries, We pondered upon the following questions –

Why are the cancer trends on an increase in China?

Cancer Prevalence in the China



Cancer Mortality in the China

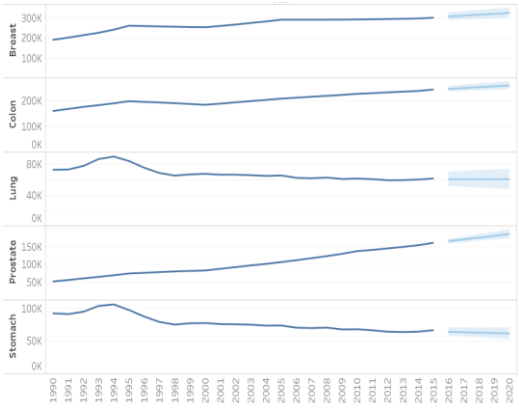


The forecast also shows an increasing trend of the cancer. Studies show that Increasing trends are attributed to

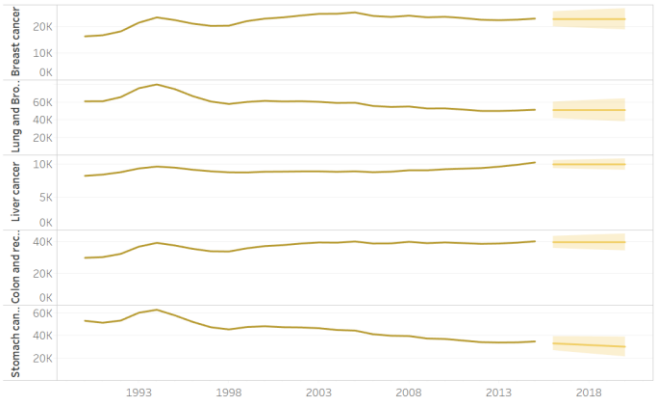
- Increase in air pollution, increase in smoking habits.
- lifestyle changes, stress is among the other reasons for increase in breast cancer cases.
- Having more than one child lowers breast-cancer risk. With the one-child policy in place since 1979, most women – especially if they worked in the city – had to strictly follow the policy in order to avoid being fined. Although the one-child policy rule was replaced in 2015 with a two-child policy rule, the possible benefit on breast cancer incidence will probably take 15 to 20 years to show.

Why Russia has a decrease in trends of cancer?

Cancer Prevalence Trend in Russia



Cancer Mortality Trend in Russia



- Decrease in the number of Lung & Stomach cancer cases.
- Forecast – shows a decreasing trend.

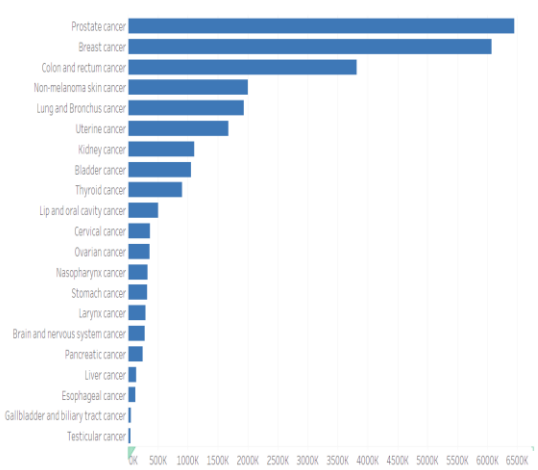
Studies show –

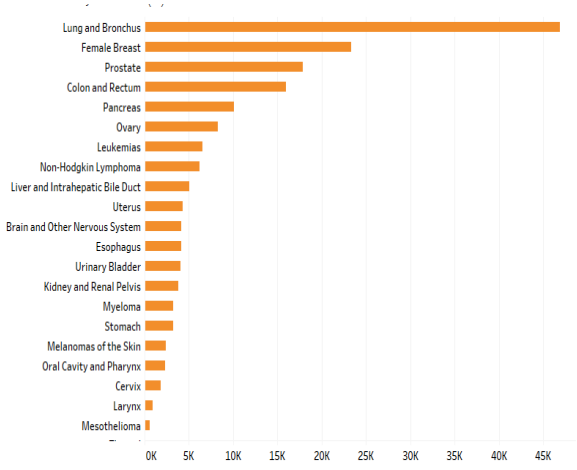
- Lung cancer trend is attributed to decrease in smoking, consumption of low tar & nicotine cigarettes, increased use of filter cigarettes.
- Stomach Cancer Trend attributed to decrease in alcohol consumption.

ANALYSIS OF CANCER PREVALENCE/MORTALITY IN THE UNITED STATES

Cancer Prevalence in the US –

- For all States
- For all Years
- Top 5 Cancers(Prevalence) – Prostate Cancer, Breast Cancer, Colon Cancer, Lung Cancer, Uterine Cancer
- Top 5 Cancers (Mortality) – Lung and Bronchus Cancer, Breast Cancer, Prostate Cancer, Colon and Rectum, Pancreatic Cancer





- As seen from the dashboard – United States has highest prevalence rates of Prostate and Breast Cancer and has highest mortality rates of Lung and breast cancer.
- State- wise top 5 Cancers(Prevalence) –
    - Prostate Cancer– Delaware
    - Breast Cancer – Maine
    - Lung and Bronchus Cancer – West Virginia
    - Colon and Rectum Cancer – West Virginia
    - Uterus Cancer – Maine
- State- wise top 5 Cancers (Mortality) –
    - Lung and Bronchus Cancer – West Virginia
    - Breast Cancer – Pennsylvania
    - Prostate Cancer – Florida
    - Colon and Rectum Cancer – West Virginia
    - Pancreatic Cancer - Pennsylvania
- Though the prevalence of Prostate and Breast Cancer is high in the US, the survival rate is quite good

Taking a closer look at the different cancer prevalence/mortality rates in different states, We pondered upon the following questions –

Why is the prevalence of cancers highest in Maine?

- Cancer rates vary by ethnicity. In that aspect, Maine has 90% White's population who are more prone to breast and prostate cancer.
- Aging is also considered a reason for cancers. Maine is one of seven states where the median age is over 40, making Maine a state with high cancer prevalence.

- In addition to that, 75-85 percent of the cancer diagnoses stem from preventable risk factors such as smoking, diet, exercise, obesity and sun exposure.

Why does West Virginia have high cancer prevalence and mortality rates?

- West Virginia has high rates of air Pollution & Less effective Tobacco control measures in place which can be attributed to the high numbers of cancer prevalence and mortality.
- West Virginia's 2018 "State of Tobacco Control" report grades

"F" for Tobacco Prevention Program Funding;

"C" for Smokefree Air;

"F" for Tobacco Taxes;

"D" for Access to Tobacco Cessation Treatments;

"F" for Tobacco 21 Laws.

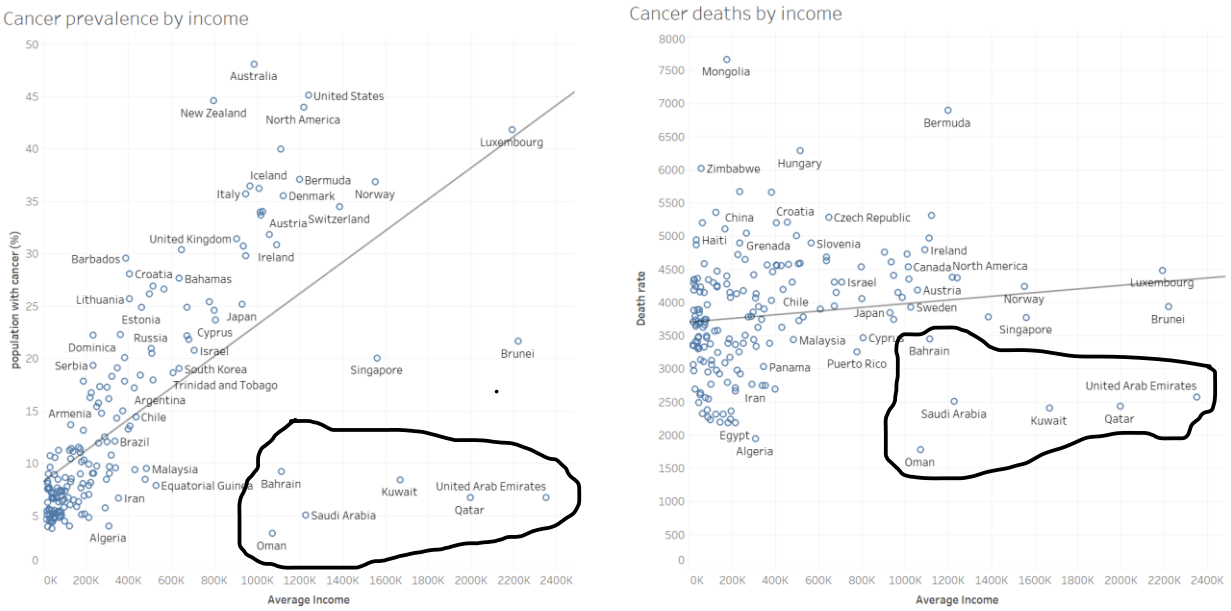
- Radon is the 2nd leading cause of lung cancer,
- **Radon gas** is a colorless, odorless, tasteless radioactive **gas** that's formed during the natural breakdown of uranium in soil, rock, and water.
- **Radon** exits the ground and can seep into your home through cracks and holes in the foundation.
- **Radon gas** can also contaminate well water.
- Some geographical areas have naturally higher prevalence of radon levels than others.
- 20 counties of West Virginia are considered Zone 1 with predicted average indoor radon screening levels > 4 pCi/L.(picoCuries per liter)
- 29 counties of West Virginia are considered Zone 2 with predicted average indoor radon screening levels - 2 to 4 pCi/L.
- To fix this problem radon mitigation system can be installed in the houses.

Why is the cancer mortality high in Pennsylvania?

- Smoking, diet and population age are among the factors contributing to Pennsylvania's high ranking
- Shortly after the accident at the Three Mile Island (TMI) nuclear power plant some area residents believed an increase in cancer had occurred, possibility that stress and other psychosocial factors may affect the carcinogenic process, perhaps through the action of

corticosteroids on immune function or of stress hormones like prolactin on regulation of neoplastic cell growth. Thus, causing cancer.

CANCER AND ANNUAL INCOME – WORLD-WIDE TREND



CONCLUSION

Interesting and valuable insights were drawn from visualizing the cancer data for all the countries and states of the US. Below are the key highlights of this study.

- Developed and Developing countries are equally facing high cancer numbers.
- Prevalence of breast cancer in China is lower than US but a rapid increase in incidence of the disease is witnessed, the same rapid increase is being witnessed in India too. This trend is cause for great concern.
- Russia shows decreasing trends with respect to many types of cancers and many studies attributes this to having heathy lifestyle practices, educating the population on decreasing alcohol consumption and quitting smoking. These are some take-aways from the Russia cancer story which can be a solution to the increasing trends.
- In the United States, Like for example, West Virginia's high cancer numbers – on one hand, are attributed to not catching Cancer at early stages, while on other naturally occurring gases like Randon, are also considered as culprits in causing cancers.
- Countries in the middle east are doing a good job of managing cancer.
- Thus, it can be concluded that demographic, ecological, environmental, cultural and genetic variables contribute to the heterogeneity of cancer incidence. By controlling the cancer incidence, we can get the cancer related mortality under control.

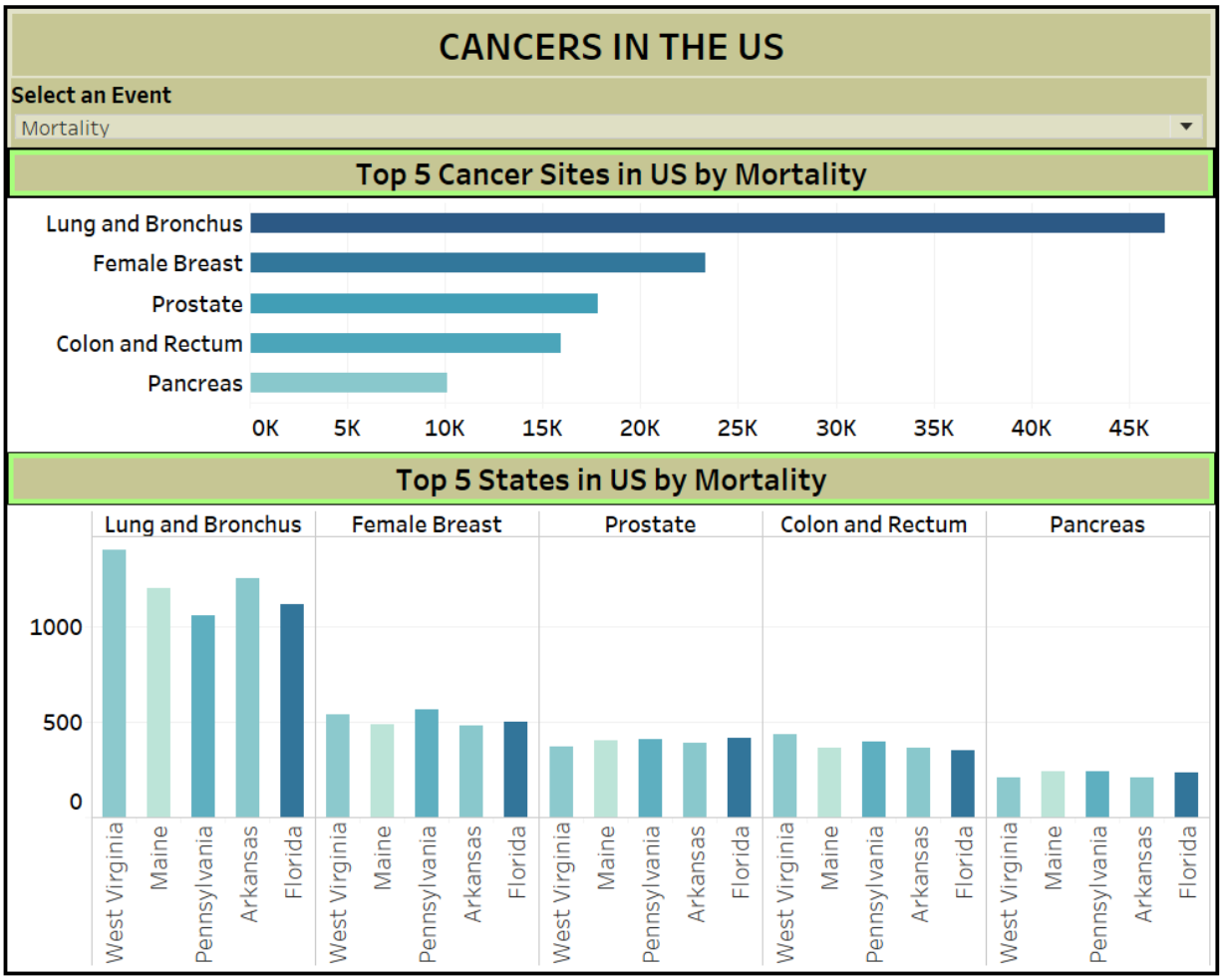
CHALLENGES FACED

- We had difficulty in accumulating the large volume of data into one csv.
- We had to filter and extract data according to our needs to avoid crashing our laptops as the number of rows of data available was huge.
- As the population of all the countries and states is different, we had to standardize the numbers to get in a fair comparison.
- To create dashboards, we had to combine multiple sheets of data to depict all the insights in a concise way.

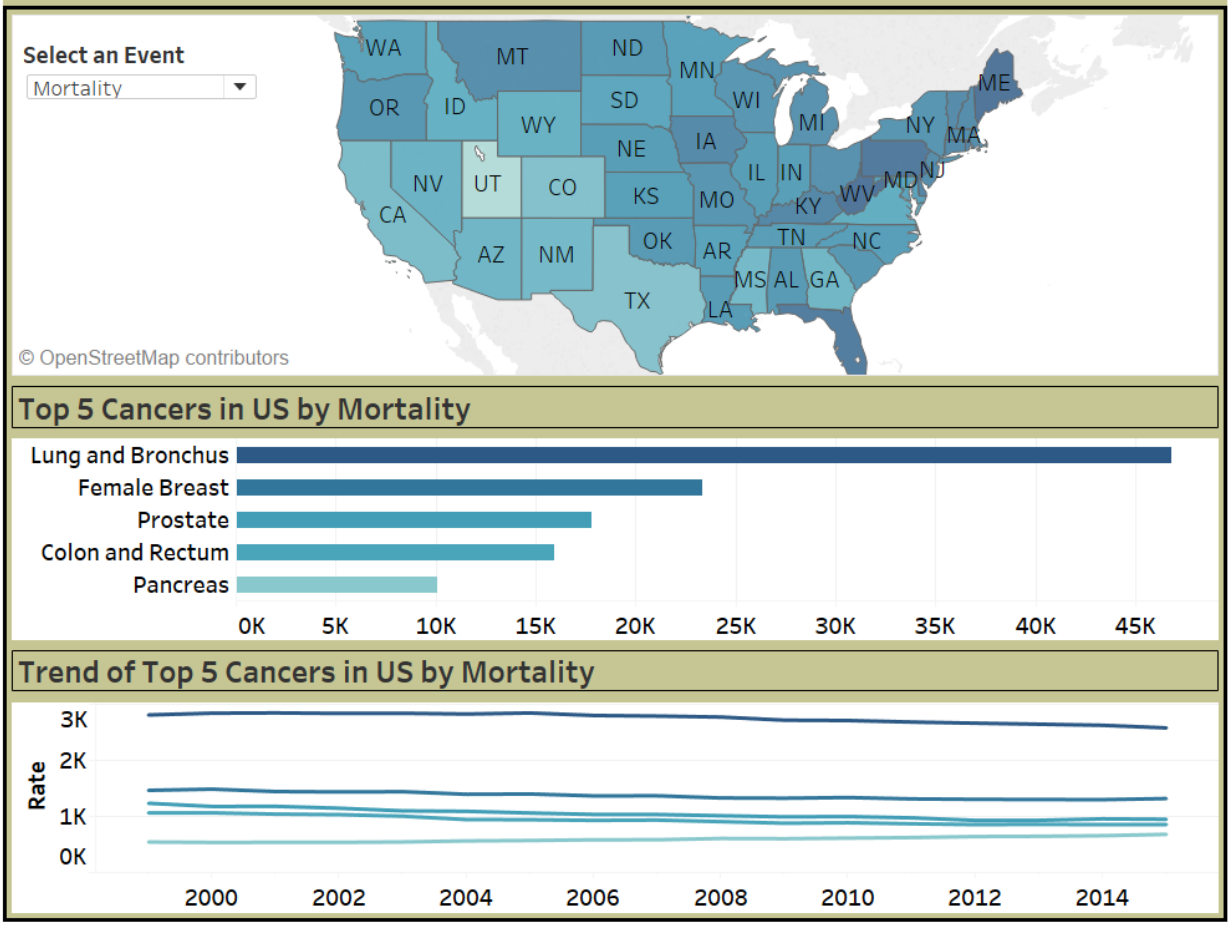
OUR AUDIENCE

- We have created basic dashboards, which can cater any level of audience.
- The visualizations give very detailed insights into the different cancers and their prevalence and mortality stories.
- Here is a screen shot of the dashboards.

DASHBOARDS


CANCERS IN THE US
Select an Event
Mortality
Top 5 Cancer Sites in US by Mortality
Lung and Bronchus
Female Breast
Prostate
Colon and Rectum
Pancreas
0K
5K
10K
15K
20K
25K
30K
35K
40K
45K
Top 5 States in US by Mortality
Lung and Bronchus
Female Breast
Prostate
Colon and Rectum
Pancreas
1000
500
0
West Virginia
Maine
Pennsylvania
Arkansas
Florida

CANCER TREND IN US AN OVERVIEW



Top 5 Cancers in US by Mortality

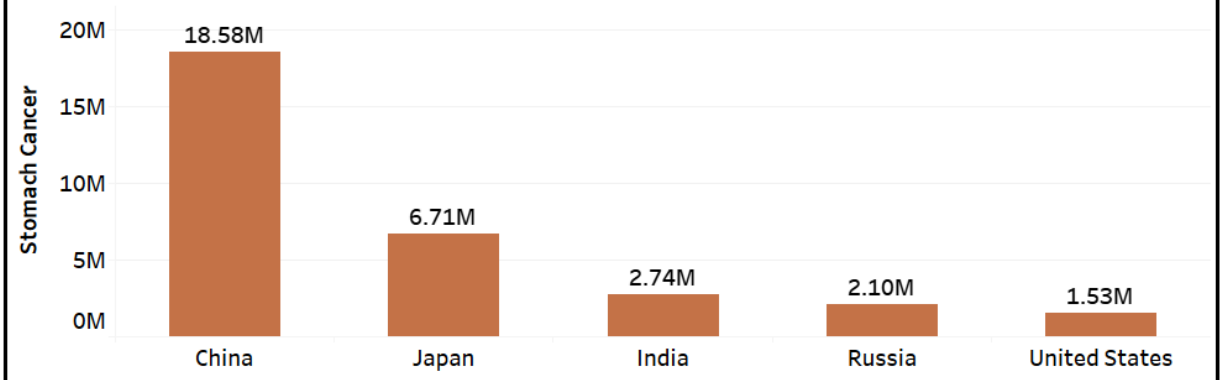
Trend of Top 5 Cancers in US by Mortality

WORLD WIDE CANCER PREVALENCE

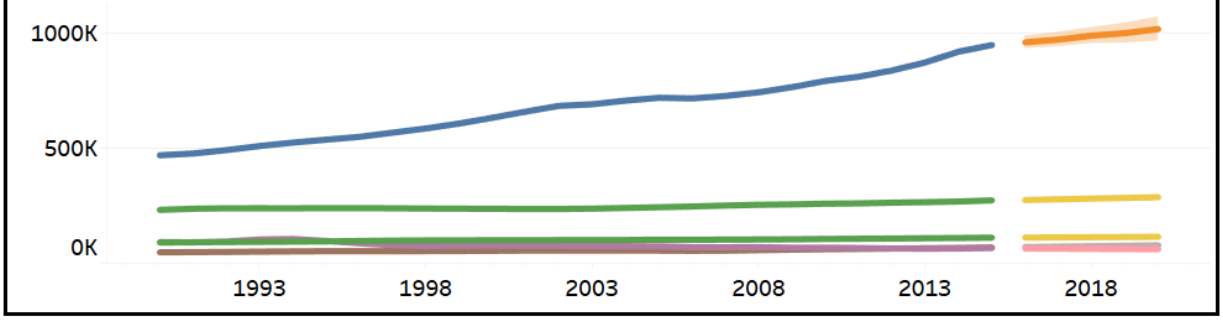
Select a Cancer Type

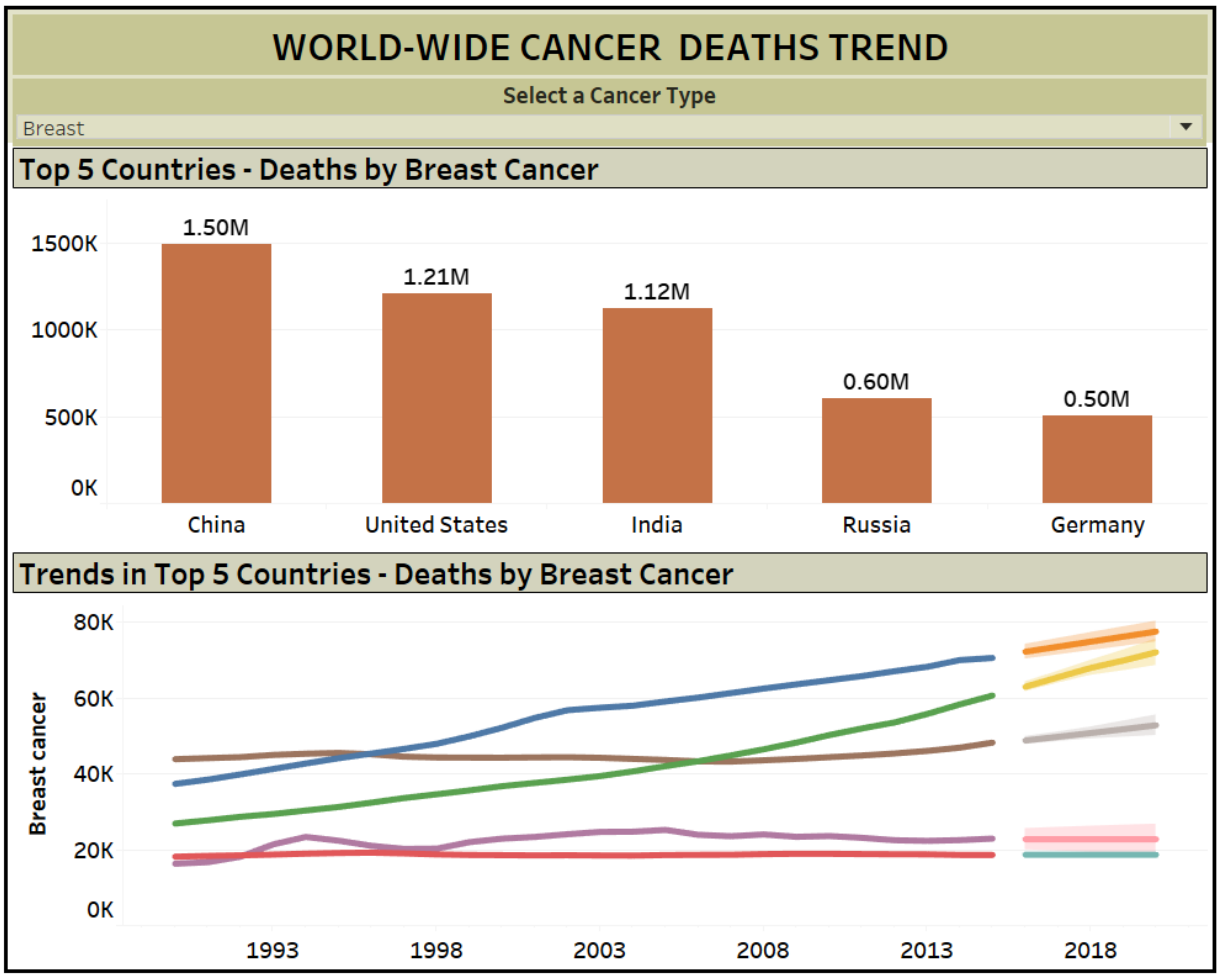
Stomach

Top 5 Countries - Deaths by Stomach Cancer Type

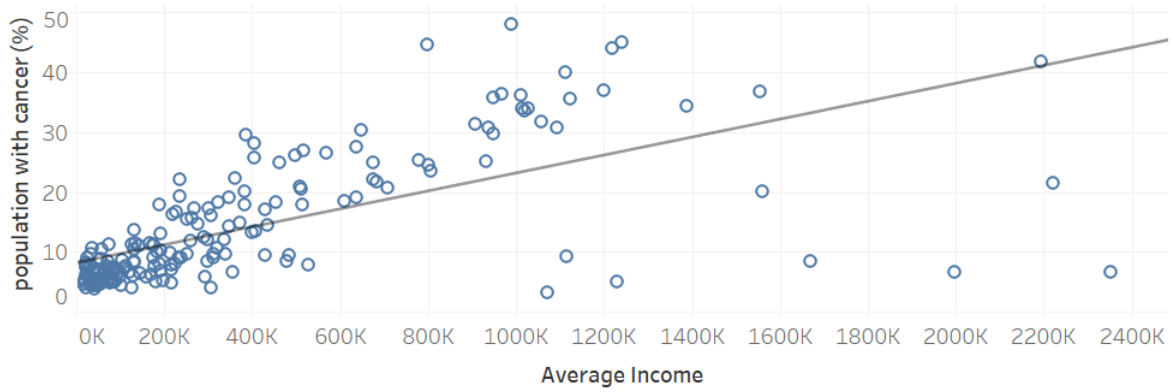


Trends in Top 5 Countries - Deaths by Stomach Cancer Type




WORLD-WIDE CANCER DEATHS TREND
Select a Cancer Type
Breast
Top 5 Countries - Deaths by Breast Cancer
1500K
1000K
500K
0K
1.50M
1.21M
1.12M
0.60M
0.50M
China
United States
India
Russia
Germany
Trends in Top 5 Countries - Deaths by Breast Cancer
Breast cancer
80K
60K
40K
20K
0K
1993
1998
2003
2008
2013
2018

Cancer prevalence by income



Cancer deaths by income