UNIVERSITY OF HERTFORDSHIRE

School Of Physics, Engineering & Computer Science

Name: SANDRA BINU

ID : 22029960

7PAM2000-0901-2023 - Applied Data Science 1
2023-2025

Assignment 1 : Visualisation

Date: 10 November 2023

Contents:

Abstract

Data visualisation is the visual representation and exploration of data with the help of graphs, charts and maps, telling a meaningful story and showing the relationship between the variables.

"A picture is worth a thousand words," as the saying goes. And the phrase has never been more true than it is in the big data era when enterprises are overwhelmed with information from both on-premises and cloud-based sources, as well as a variety of data kinds.

It's getting harder to sort through information and determine what is important and what isn't. Visuals provide the capacity to quickly and easily assess what is important. They also make analysis much simpler and faster.

Visualisation helps to understand the data, its patterns and relationships which is might not be obvious from the huge numerical data. It helps to compare and analyse trends in the variables over time. It also helps to identify the outliers and anomalies in the data more easily.

About Data

The data used in this assignment is the Cancer Types and Deaths in All Countries over the period of 1990 to 2019. The data shows the death count caused by 29 types of cancer. This dataset is taken from Kaggle.

- Link to the data: cancer-and-deaths-dataset

- Link to the GitHub repository: Data-Visualisation-GitHub-Repository

1. Line Plot

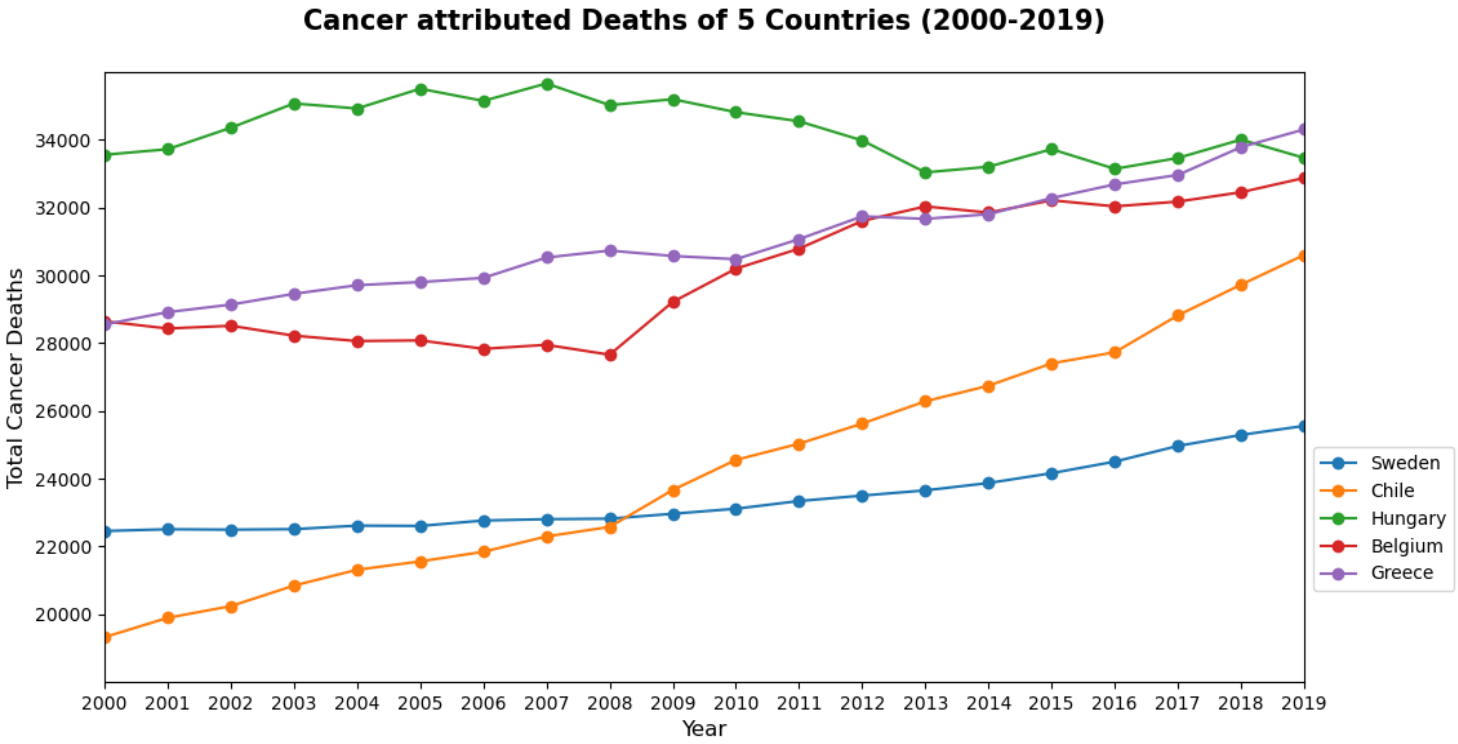

Figure 1

Here the multiple Line plot shows the time series variation in total death caused by cancer in 20 years irrespective of its type in the 5 developed countries Sweden, Chile, Hungary, Belgium and Greece.

Line charts are used to illustrate patterns or variations in the number of deaths within a specific variable, such as years. It helps illustrate how the number of cancer caused deaths in each country varies over time.

Insights:

- The graph shows the total death counts from 2000-2019.
- From 2000 Hungary had a slight increase in cancer deaths until 2007 after which it can be seen to have a steady decline in the death count.
- Belgium and Greece show a very close count at the beginning of 2000 followed by a steady decrease in the count in Belgium by 2008. From 2010 they show similar pattern in their cancer deaths with almost same number of cancer deaths by 2015. After 2015 Greece showed a higher rate of cancer deaths compared to Belgium.
- Chile can be seen to have a uniform rate of increase in cancer deaths over the years.
- Similar is the case of Sweden but the rate of increase is much lower, reaching a lowest death cases in 2019 as compared to the other 4 countries.

2. Stacked Bar Chart

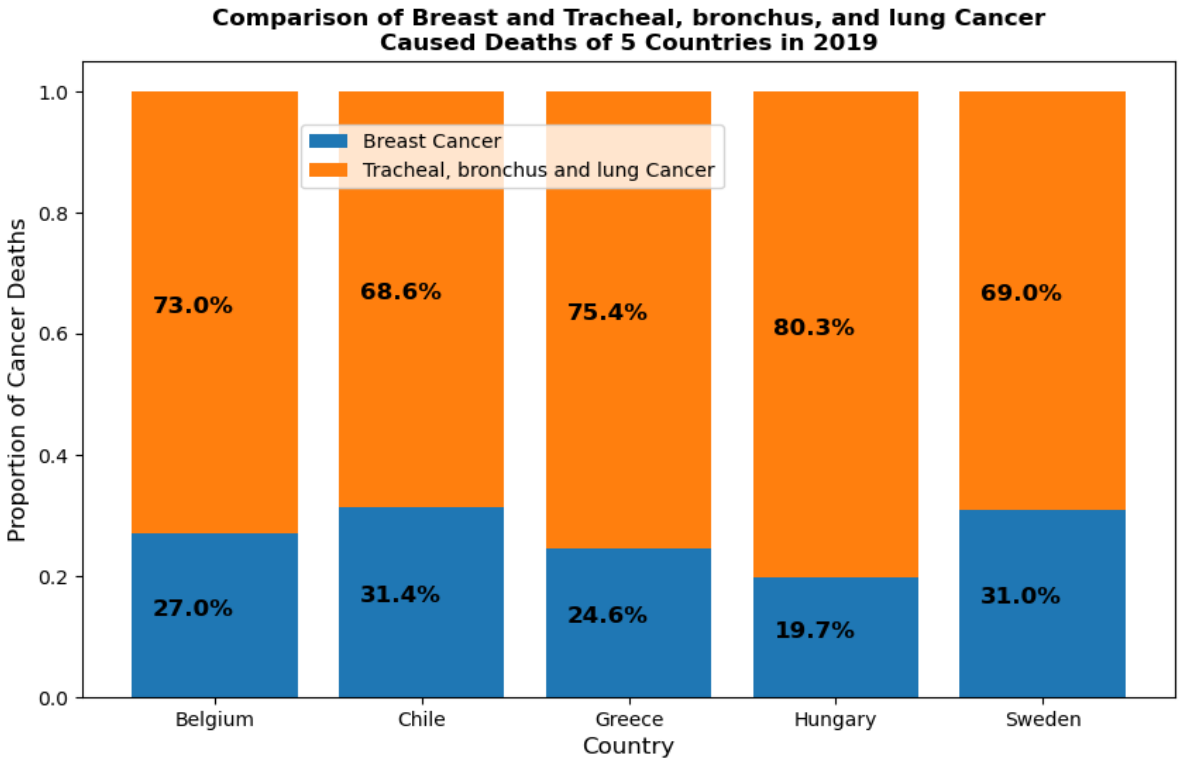

Figure 2

Here the plot is a proportion stacked bar plot which shows a comparison between the two among the most fatal cancer types, breast cancer and tracheal, lung and bronchus cancer deaths in the selected five countries in the year 2019.

A proportional stacked bar chart shows the size and proportion of the data at the same time. A proportional stacked bar chart helps us to compare between the proportional distribution of these cancer types in each country. It helps to visualize the composition of deaths of selected types in a year in each country.

Insights:

- In every country taken, the proportion of tracheal, bronchus and lung cancer is higher as compared to breast cancer.
- The distribution of breast and tracheal, bronchus and lung cancers are almost similar for Chile (31.4%) and Sweden (31.0%).
- Hungary is seen to have a lower proportion of breast cancer deaths with 19.7% whereas the tracheal, bronchus and lung cancer stands top with 80.3%.
- Belgium has a proportion of 27% breast cancer and 73% tracheal, bronchus and lung cancer.

3. Pie Chart

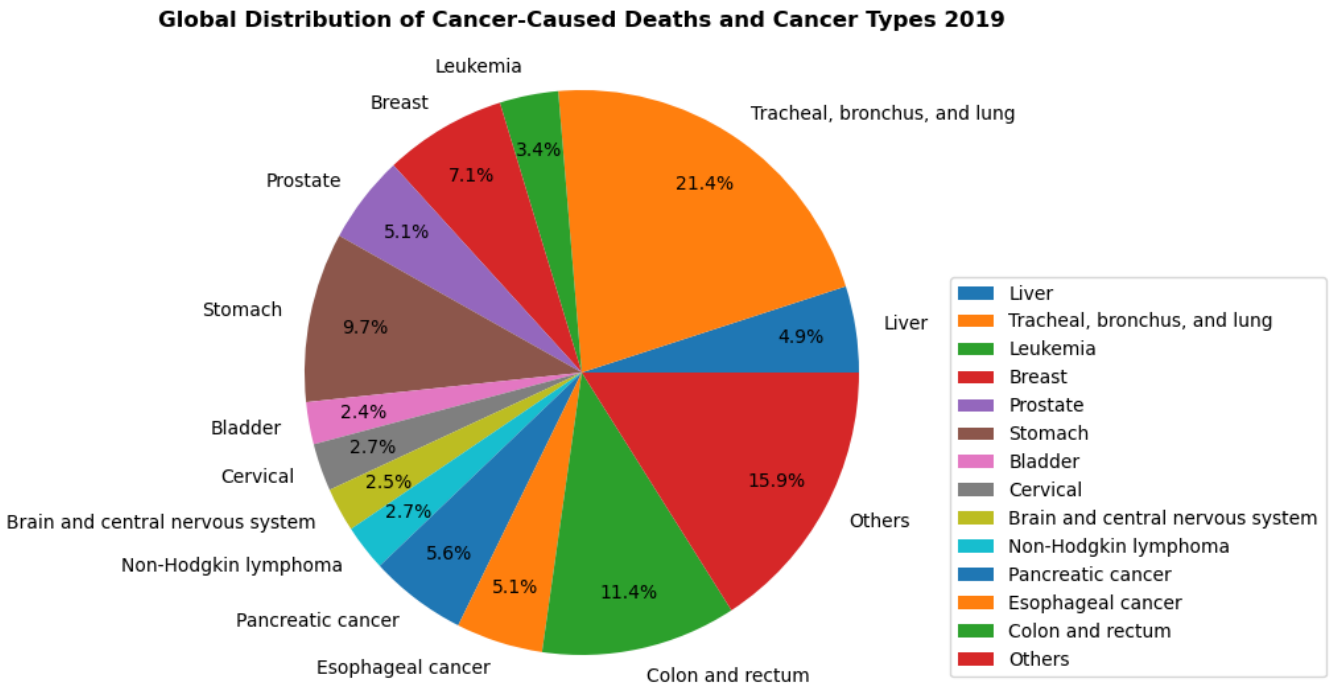

Figure 3

Here the global cancer death count and its types in 2019 are shown as a pie chart.

Pie Chart shows the relative proportion of each cancer type deaths to the global cancer death count. It summarizes a huge data set in a visual format.

Insights:

- The chart has 14 pies each representing the proportion of a cancer type except the 'Others'
- 'Others' represents a set of cancer types (16 types – Kidney, Lip and Oral, Larynx, Gallbladder and biliary tract, Malignant skin melanoma, Hodgkin lymphoma, Multiple myeloma, Other neoplasms, Thyroid, Uterine, Ovarian, Testicular, Nasopharynx, Other pharynx, Non-melanoma skin, Mesothelioma) which only covers 15.9% of the global cancer deaths in 2019.
- Tracheal , bronchus and lung cancer contributed to about 21.4 % cancer deaths in 2019 globally. This was followed by colon and rectum, stomach and breast which accounts for 11.4, 9.7 and 7.1% cancer deaths respectively.

Conclusion

From this assignment, it is understood that the cancer trend tends to increase over time in most countries and the death caused by different types of cancer varies drastically. These plots ensures the clear representation of how cancer deaths varies with time in the selected developed countries, how is the proportional composition of the most popular breast and tracheal, bronchus and lung cancer deaths in these countries in the year 2019 and also the global distribution of different types of cancer deaths in 2019. It is also very clear that the tracheal bronchus and lung cancer contributes a major part of the cancer fatality cases.

References

- *Matplotlib line*. (n.d.). https://www.w3schools.com/python/matplotlib_line.asp
- GeeksforGeeks. (2020, October 1). *Stacked Percentage Bar Plot In MatPlotLib*. https://www.geeksforgeeks.org/stacked-percentage-bar-plot-in-matplotlib/
- GeeksforGeeks. (2021, November 30). *Plot a pie chart in Python using Matplotlib*. https://www.geeksforgeeks.org/plot-a-pie-chart-in-python-using-matplotlib/
- WCRF International. (2022, April 21). *Global cancer data by country | World Cancer Research Fund International*. https://www.wcrf.org/cancer-trends/global-cancer-data-by-country/