Assignment 5 - Covid-19

Concept Document

Group-11

Members: Abdul Ahad Ayaz, Sayalee Chavan, Suganthi Jaganathan, Bhagyashree Sanjay Borade

This document provides the overview of our project, containing the description of dataset, the End Users that we are targeting, visualization techniques that we are going to use and in the end how can we explore the visualized data using different interaction techniques.

DATA

There are two multivariate datasets that we are going to use **covid_de.csv** and **demgraphics_de.csv**. First dataset **covid_de.csv** provides the daily update of Covid-19 cases of different states and counties of Germany based on gender and age group. Second dataset **demgraphics_de.csv** is the supporting dataset for the first dataset as it contains the population of different states of Germany based on gender and age group. The reason of choosing these dataset upon other dataset is that it satisfies the requirement of End User described below.

Data Source

- **covid_de.csv**

  - https://www.kaggle.com/headsortails/covid19-tracking-germany?select=covid_de.csv (covid_de.csv is refined dataset based on RKI data)
  - https://npgeo-corona-npgeo-de.hub.arcgis.com (Collected by [Robert Koch Institute](#))
- **demographic_de.csv**

  - https://www.kaggle.com/headsortails/covid19-tracking-germany?select=demographics_de.csv
  - https://www-genesis.destatis.de/genesis/online/data?operation=sprachwechsel&language=en (Collected by [Statistisches Bundesamt](#))

Dataset Type

Dataset	Spatial vs Non-Spatial Data	Point, Scalar, Vector Data
covid_de.csv	Non-Spatial Data	Point Data
demographic_de.csv	Non-Spatial Data	Point Data

Dataset Characteristics

The first dataset **covid_de.csv** consist of following columns:

Column Name	Data Type	Description
state	Nominal	It contains the name of German states that are total 16 in number. Example Baden-Wuerttemberg, Nordrhein-Westfalen, Hessen, Bayern etc.
county	Nominal	It contains the name of counties of each state label with LandKries(LK)/StadtKries(SK). Example SK Koeln, SK Hamm, LK Paderborn, LK Guetersloh in Nordrhein-Westfalen state.
age_group	Ordinal	It consists of different age categories such as 0-4, 5-14, 15-34, 35-59, 60-79, 80-99 and empty field NA with age unknown.
gender	Ordinal	It consist of gender M for male and F for female and NA for gender unknown.
date	Nominal	It consist of date when the data is updated.
cases	Quantitative	It consist of per day count of confirmed cases of Covid-19.
deaths	Quantitative	It consits of per day count of death due to Covid-19
recovered	Quantitative	It consists of per day count of recovered patients from Covid-19.

The Second file **demographics_de.csv** that we are going to use for our project consist of following columns:

Column Name	Data Type	Description
state	Nominal	It contains the name of German states that are total 16 in number. Example Baden-Wuerttemberg, Nordrhein-Westfalen, Hessen, Bayern etc.
gender	Ordinal	It consist of gender such as male and female.
age_group	Ordinal	It consists of different age categories such as 0-4, 5-14, 15-34, 35-59, 60-79 and 80-99.
population	Quantitative	It consists of population categorized by state, gender and age.

Data Modeling

- **covid_de.csv** is described as Entity E_7^P, where P represents that dataset has Point Data and 7 represents the dimensions.

- **demographics_de.csv** is described as Entity E_3^P, where P represents that dataset has Point Data and 3 represents the dimensions.

USER AND TASK

Users are the ones for whom the application or the concept has been designed. Meanwhile, tasks are used to express the target of our visualization. They describe the purpose of an application or a project.

User

The potential user would be the North-Rhine-Westphalia (NRW) State government, Germany.

Task

Sophisticated tasks will be:

- To visualize the current scenario of Covid-19 in NRW for all over the state with the major details like age-group, demography, gender, recovered and active cases ratio on daily basis. This provides an overview on the existing situation.
- To visualize the infection rate in NRW with age group and location using covid_de.csv and demographic_de.csv respectively. This helps the user to control the Covid-19 spread in particular county.
- To visualize the age group, gender and active cases then highlight the age group such as "0 to 14". This helps the user to monitor the children's health conditions and analyze the impact of operation of schools (Kindergarten).
- To visualize and highlight the increase/decrease in Covid-19 infections with respect to population, active and recovered cases. This helps the user to decide upon the operation and further plans to open/retain schools, universities, shopping malls and workplaces.
- Compare the current situation of Covid-19 spread of NRW with other states to make policies like lockdown etc.

VISUALIZATION TECHNIQUES

The preprocessed data will be visualized using various interactive visualization techniques to have a broader overview of correlation amongst data which will help to deduce significant information.

Techniques

The visualization techniques that will be used are as follows:

- **Time-oriented visualization technique**
  - Multi-series Line chart (for multivariate data) : x-axis represents "days" and y-axis represents "cases". Graphical attribute (glyph) is a colour which denotes categories of cases (confirmed, recovered, death).
- **Region-based technique**
  - Bar chart(3D): x-axis represents "age group" and y-axis represents "No. of cases".

- Grouped Bar chart(3D) (for multivariate data): x-axis represents "Counties" and y-axis represents "cases". Graphical attribute (glyph) is a colour which denotes categories of cases (confirmed, recovered, death).
- **Multivariate data visualization technique**

  - Scatter plot (for multivariate data): x-axis represents "age group" whereas y-axis represents "No. of cases". Graphical attribute (glyph) is a colour which denotes gender (male, female).
  - Bubble chart: x-axis represents population and y-axis represents "No. of cases". Size of the bubble depends on the count of cases.

INTERACTIONS

Interaction in visualization techniques provides more understanding of visual representation. User can do the selection, filtering for switching between different visualization techniques or choosing data variables. Interaction for visualization techniques specified as following:

- **Filtering/ Range slider:** With the use of range slider, the user can select a particular range for days mentioned on the x-axis as well as able to see cases in the range of selected days. This technique can be used in Multi-Series line chart.
- **Selection:** Selection can be used for visualizing different data variables by selecting/unselecting legend-based checkboxes and selection would be one or multiple. This interaction operator can be used in Multi-Series Line Chart where the user can do selection based on "No. of confirmed cases", "No. of recovered cases" and "No. of deaths". For Bar Chart(3D) and Grouped Bar Chart(3D), the user can select between Bar Chart and Grouped Bar Chart in dropdown. For Grouped Bar Chart(3D), visualization can be customized by selecting checkboxes represents "No. of confirmed cases", "No. of recovered cases" and "No. of deaths". For Scatter plot multi-series, User can customize the visualization by selecting the checkbox in legends based on gender(male/ female). For Bubble Chart, User can customize visualization by selecting the checkbox which consists of "Total confirmed cases", "No. of recovered cases", "No. of deaths" parameters.
- **Hover:** This interactive technique reveals detailed data point information by moving the mouse cursor over data value space which will be included in all visualization techniques mentioned above. For eg, hovering on scatter data points reveals information about no of cases represented by y-axis and age group represented by x-axis.