**Data Analysis and Visualization**

Hijab Tahira Beg 22i-2071

Syeda Zunaira Ahmed 22i-2075

# 

# Submitted to:

Ma’am Zonera Anjum

# 

# **Interactive Data Visualization Dashboard Report**

## 

## **Introduction**

This report outlines the insights derived from an interactive data visualization dashboard created using D3.js. The dashboard uses a cleaned dataset from COVID analysis in South Korea, focusing on relationships between geographic locations, infection types, and timestamps. This document provides an explanation of the visualizations included, their purpose, and the key findings.

## **Visualization Descriptions**

### **Force-Directed Graph**

The Force-Directed Graph maps relationships between provinces, cities, and infection types. Each node represents an entity - province, city, or infection type, and edges represent relationships between them. The graph is interactive, allowing zoom and drag capabilities to explore connections.

This visualization highlights clusters of related provinces and cities. It identifies patterns in infection spread, showing which cities are most central in the network of cases.

**Zoom and Pan**: Users can zoom in/out and pan across the graph to explore specific regions of the network.

**Tooltips**: Hovering over a node displays details about the entity, such as its name.

**Drag and Drop**: Nodes can be dragged to reposition them dynamically, enhancing exploration.

**Click Actions**: Clicking on a node provides additional information or interaction feedback.

**Filters**:

* Users can toggle visibility of specific relationship types (e.g., "Province-City" or "City-Infection") using checkboxes, making it easier to focus on relevant data subsets.

### **Geographic Map**

The Geographic Map visualizes the spatial distribution of cases in South Korea, showing locations and clusters using colored markers. Each marker represents a city, with size proportional to the number of confirmed cases.

This map helps pinpoint regions with high concentrations of cases. Clusters of cases in specific provinces suggest localized outbreaks or hotspots.

**Zoom and Pan**: Users can zoom and pan across the map to focus on specific regions.

**Tooltips**: Hovering over a marker shows detailed information about the city, including the infection type, number of cases, and whether it belongs to a cluster.

**Click Actions**: Markers highlight related data points or trigger additional exploration options.

**Filters**:

* **Region Filter**: A text input allows users to filter locations by region name.
* **Cluster Filter**: A checkbox toggles the visibility of clustered or non-clustered markers.

### **Timeline Visualization**

The Timeline Visualization provides an animated view of confirmed cases over time. It uses a slider and play-pause functionality to explore how the pandemic evolved.

Temporal trends reveal peaks and troughs in the spread of the infection. It highlights significant events in the timeline of the outbreak, such as sudden surges or declines.

**Key Features**

1. **Trend Visualization**:

* A time-series line chart displays the number of confirmed cases across provinces.
* Each province is represented by a uniquely colored line for differentiation.
* The trend updates dynamically as users interact with the timeline slider or animation.

1. **Interactivity**:

* **Tooltip**: Displays detailed information (confirmed, released, and deceased cases) for the hovered data point.
* **Legend**: Allows users to filter the visualization by selecting individual provinces or viewing all provinces collectively.
* **Play/Pause Animation**: Animates the trend over time, showing the progression of cases.
* **Slider**: Lets users manually adjust the timeline to view data for specific dates.

1. **Scales and Axes**:

* X-axis: Represents time (dates) with a linear time scale.
* Y-axis: Represents the number of confirmed cases, defaulting to a logarithmic scale for multiple provinces and a linear scale for individual provinces.

**Columns Used**

1. **Date**: represents the timeline for the trend visualization.
2. **Province**: Identifies the geographical region where the data was recorded.
3. **province\_confirmed:** Shows the number of confirmed COVID-19 cases per province.
4. **province\_released** and **province\_deceased**: Included in the tooltip for detailed data exploration.

**Visualization Details**

1. **Tooltip**:

Displays the following details for the hovered province at the current timeline:

* Province name
* Confirmed cases (province\_confirmed)
* Released cases (province\_released)
* Deceased cases (province\_deceased)

1. **Legend**:

* Allows users to toggle visibility of individual province trends.
* Clicking a province in the legend highlights its trend and switches to a linear y-axis for focused analysis.

**Filters and Controls**

1. **Filters in the Legend**: enables filtering by province, allowing comparison or focused analysis.
2. **Time-Based Filter (Slider)**: provides the ability to focus on specific dates and dynamically updates the visualization.
3. **Animation (Play/Pause Button)**: automates the visualization, showing the temporal progression of cases.

**Dynamic Scales**

* **Logarithmic Scale**: Default for displaying multiple provinces, efficiently handling large disparities in data.
* **Linear Scale**: Automatically applied when a single province is selected for detailed analysis.

### **Sunburst Chart**

The Sunburst Chart shows hierarchical relationships among provinces, cities, and gender-specific data. Inner rings represent higher-level categories, while outer rings detail subcategories.

This chart helps identify patterns in demographic data and provides an intuitive understanding of case distribution across different regions.

**Zoom on Click**: Clicking on any segment zooms into that category for a detailed view.

**Tooltips**: Hovering over segments reveals their hierarchical path and associated values.

**Breadcrumb Navigation**: Users can navigate back to higher levels of the hierarchy using intuitive breadcrumbs.

**Animations**:

* **Transition animations** smooth the zooming experience as users navigate deeper into the chart.
* **Dynamic updates** to labels and arcs ensure a seamless exploration experience.

### **Tree Map**

The Tree Map visualizes the hierarchical distribution of cases across provinces and cities, with nested rectangles representing different categories.

This visualization emphasizes the relative contribution of each region to the overall case count, making it easier to identify the most and least affected areas.

**Features**

**1. Hierarchical Data Representation**

* Data is organized into a three-level hierarchy: **Province > City > Infection Case**.  
    
  **first level hierarchy:**  
  A screenshot of a computer screen

  Description automatically generated

**Second level hierarchy:**

A screenshot of a computer

Description automatically generated

**Third level hierarchy:**  
A screenshot of a computer

Description automatically generated

* Each level visualizes the maximum confirmed cases for the respective category using a tree map.

**2. Interactive Elements**

* **Zoom and Pan**: Users can zoom in and out to explore details at different hierarchical levels.
* **Breadcrumb Navigation**: Displays the current navigation path, allowing users to quickly navigate back to specific levels.
* **Reset Button**: Restores the default view and resets the zoom state.

**3. Tooltips**

* Displays detailed information, including:

1. **Name** of the hierarchy level (Province, City, or Infection Case).
2. **Total Cases** at the node.
3. **Percentage Contribution** to the parent node.

**4. Legend**

* Includes a gradient color legend to represent the range of confirmed cases.
* Dynamically scales based on the minimum and maximum case values in the dataset.

**Columns Used:**

The visualization uses a CSV file, cleaned\_covid\_map\_data.csv, containing the following fields:

1. Province
2. City
3. infection\_case
4. province\_confirmed (numeric)

## **Insights Derived**

### **1. Geographic Clusters and Case Distribution**

* Provinces like Seoul and Daegu have significantly higher case densities, as seen in the Geographic Map. These regions serve as epicenters for the outbreak, possibly due to higher population densities or major transit hubs.
* Targeted interventions, such as strict lockdowns or prioritized vaccination in these regions, could effectively control the spread.

### **2. Temporal Trends in Case Growth**

* The Timeline Visualization shows sharp peaks corresponding to specific dates, indicating outbreaks or superspreader events. For example, significant case surges align with holidays or major public gatherings.
* Policies to restrict large-scale gatherings during key periods could prevent similar spikes.

### **3. Gender Dynamics**

* The Sunburst Chart highlights a gender-skewed distribution of confirmed cases in some regions. In some provinces, males or females are disproportionately affected, suggesting varying exposure risks or healthcare-seeking behavior.
* Public health campaigns could be tailored to target specific demographics based on their risk exposure.

### **4. Relationship Mapping Between Provinces and Infection Types**

* The Force-Directed Graph reveals that certain infection types, like "Clustered Community Outbreaks" are more prominent in urban centers, while "Imported Cases" are scattered across provincial hubs.
* Different regions required customized response plans, focusing on community transmission in cities and monitoring travel-related cases in smaller towns.

### **5. Hierarchical Contribution of Regions**

* The Tree Map shows that a small number of provinces account for the majority of cases. Sub-regions within these provinces have highly variable contributions, with cities like Daegu dominating.
* Resource allocation should be weighted toward these high-contribution regions, focusing on healthcare infrastructure and testing capabilities.