Visualization of COVID-19 Data in South Korea

Rui Zhang, Ruixin Tang, Yidan Cai

Abstract—The spread of COVID-19 has raised worldwide concerns from epidemiologists, researchers, policy makers and the wider public. A proper organizing and visualization of existing multi-sourced data with multiple dimensions and linkages to show the causal dependencies will enable people to exploit more knowledge about the virus. In this paper, an interactive tool is built to visualize the COVID-19 data in South Korea. With this tool, users can have multiple views about the evolution of the virus, and carry on exploratory analysis using the interactions with ease.

Index Terms—COVID-19, South Korea, Visualization

1 Introduction

The COVID-19, first identified in January 2020 in Hubei province, China, has become a worldwide pandemic with over 100 million people affected [1]. Governments are working on joint efforts to get through this pandemic and sharing valuable experience and data to better learn about and fight against it. Visual tools are crucial to facilitate researchers to get insights into its development. However, many current visualizations of the COVID-19 statistics only focus on the daily trends of the cases, and the geographical density of the cases [2] [3] [4], or lack the interaction for users to establish links between observed phenomena [5]. In this work, we aim to develop a visualization tool using the Data Science for COVID-19 (DS4C) Patient Policy Province Dataset published by the (South) Korea Centers for Disease Control and Prevention (KCDC), to provide handy tools for researchers to investigate the epidemiological features and of COVID-19, and for governments to take effective and responsive actions.

1.1 Problem Description

The general goal of this work is to visualize the transmission, evolution trends, and demographic distributions of the spread of COVID-19 in South Korea. We will dig deeper into the transmission networks among groups of cases, and the composition of gender, age, and region of the cases. Visualization techniques will be used to illustrate the trend and compare the distribution with respect to different dimensions, and track the transmission through the links in the network. Through establishing links among different dimensions of the data (time, patient features, geographical information, external factors, etc), users can have a direct understanding of how the epidemic spreads and explore possible demography and external environmental features of the spread.

2 Data Analysis

2.1 Domain Data Specification

The DS4C dataset are consistently and reliably constructed based on the report materials of KCDC and South Korea local governments, due to the transparency of the quick information announcement of KCDC. The information includes the obligatory data of regional and demographic daily newly confirmed/deceased/released cases, de-identified patient information and group infection cases, geographic information and education/aging/medical levels of cities, and the list of policies enforces by the Korean government since the outbreak.

Due to its complex nature and information focused on multi-perspectives, cleaning and merging of domain-related forms are of our first concern. The first group of tables we consider as related are the total number of cases of different age, gender and provincial groups, recorded by days from 02-Mar-2020 to 30-June-2020 (provincial data from 20-Jan-2020), and the policies enforced by the Korean government in the first two quarters in 2020. From these we try to illustrate the overall trend and some general distribution of the infection and death tolls of COVID-19 among different groups of people, with policies combined, the effectiveness of the policies enforced by the Korean government could be reflected inline with the trends.

The other group of inter-related tables is the most detailed two files which recorded 5165 de-identified individual patients and 174 group infection cases information. They will be used to establish the network and drill down into the spread of the pandemic. They have a lot of missing values in their records in different columns. But since we are interested in investigating the spread between individuals, some individual patients info can be traced using the group information of the group case they belong to. In this way, we try to combine a maximal network with these records.

2.2 Data Abstraction: What

Time-related trends data: This includes 5 time-series tabular files, which depict the total confirmed, released, and deceased cases in South Korea, and the statistics of cases in different gender, age, and province groups. These are quantitative data that show the amounts as an evolution over time. As external effects, the policies imposed by the Korean government and the periods are also provided. The policies data are categorical data of different social sectors. All of these data are provided in tabular forms, and tables are multi-dimensional tables with geographical information and social sectors being the additional keys. A shared external key of these time series data is the timestamp (by dates), which could be used to join the tables together and explore the evolution over time.

Epidemiological Data of Individuals and Groups: This relates to the information from three tables. The *Case.csv* table records 174 cases and their concerned amount of people. Each case ID represents an infection case with a specific case name. They are classified into group and non-group infections, and the geographical information forms a hierarchical structure of *province - city - longitudes and latitudes*.

The *PatientInfo.csv* table records 5616 individual cases with the ordinal data age, the categorical data of gender and geographical information of city and province, and the symptom onset/confirmed/released date of the patient. Each entry represents the information of a patient with a unique patient ID, which is the key of this flat table. The ID starts with the region code and ends with a specific number given by KCDC. This table also records whom the patient is infected by (if can be traced), which enables us to construct the infection network to trace its spread and distinguish possible superspreaders. Infection cases is a shared external key with the table *Case.csv*, from where we could complete some missing values from the group information if the patient belongs to a group infection.

The *Region.csv* table contains the geographical longitudes and latitudes of the cities in South Korea. It is a flat tabular-formed data where code or the combination of province+city can be the unique key to lookup the information of a region. Geographic data can be aggregated hierarchically, from latitude and longitude to city then to province. It also gives the number of kindergartens/elementary schools/universities/nursing homes and the ratio of the academy and elderly as quantitative data. These data indicate the population of students and elderly people of a region, which tend to easily become clusters of infection in many countries.

3 Task Analysis

3.1 Domain Specific Tasks

Our tool mainly targets at helping epidemiologists and researchers to gain insights from different attributes of the South Korean data, allowing them to query and make comparisons among different groups, carry on descriptive analysis in terms of the trends and distribution, and trace the spread across the infection network. It also aims to assist policy makers to understand the effects of enforced policies with an overview from an strategically advantageous position, and to provide ordinary people with a straightforward impression on the spread and trend of COVID-19 in South Korea. Specifically, a set of domain-related tasks and questions are brought up:

1. Could the evolution of COVID-19 in South Korea be described properly by the graphs?

2. Do the infections and deceased numbers follow certain demographic/geographical/temporal patterns, based on the selected attributes?

3. Is it possible to distinguish causal infections and track the spread of the virus among the population, or distinguish some super-spreading events [6]?

4. Do certain policies exert remarkable impacts on curbing the spread of COVID-19?

The first task requires to present the data properly, which will further allow users to generate hypotheses about the distribution over different attributes, query and compare the data and trend, discovery certain patterns and examine the hypotheses, as are essentially required by the second and third task. The fourth task is to present the decision makers with the story, and all these tasks enable users to make summaries and conclusions from proper interactions. Focusing on these tasks, we will organize and select appropriate tools to encode and visualize the data, and deliver user-friendly Tableau [7] dashboard to end users.

3.2 Task Abstraction: Why

Using a consistent set of generic terminologies, as introduced in [7], our proposed domain tasks can be transformed into visualization tasks. Our first task is about presenting, searching and querying, to firstly provide an overview of the data, enable users to browse and explore it, and get a vivid understanding of the evolution process. The second task is to provide people with a visualization tool to analyze, search, and query the data in their interested aspects, for users to discovery and generate hypothesis about the demographic/geographical/temporal distributions of COVID-19 in South Korea, and further verify or reject existing hypotheses. It is also about searching and querying for certain groups of infections, to carry on comparisons with the identified cases with specific goals that are intimately related to the hypotheses. The third task is similar to the second task about analyzing, searching and querying, while the different aim leads to it being performed in different visualization forms. The last task is about presenting the policy with the trend, and provide querying functionality for specific points, to guide the decision makers through a series cognitive operations and their underlying influences, and delivering knowledge for wiser decision making process.

4 Visualization and Interaction Design: How

Focusing on our tasks, we will derive the visualization techniques based on our data abstractions, and the interaction designs to help users carry on analysis.

4.1 Numerical Evolution of Daily COVID-19 Cases

An area chart is an extension of a line graph. Since the data of each province will bleed into another area, the stacked area chart [8] is chosen in this case. The number of daily increased cases is encoded as height. Using area charts. The fully-stacked height of the topmost line will correspond to the total newly(or accumulative) confirmed(released/deceased) cases when summing across all the province groups. The heights of each segment allow getting a general idea of how each province group compares to the other in their contributions to the total. The province can be added as a selection function to show more specific detailed information within different provinces.

A line chart or a bar chart is also a valid choice, but it is more suitable for national data since plotting stacked lines for each province makes it cluttered and confusing to read. The bar chart uses a line mark and encodes the confirmed value as the bar heights. However, the bar chart suits for the case of searching for individual values. Thus, a highlighted line chart is added at the top, which expresses the daily number of confirmed cases value with aligned vertical position and point marks, ordered by date. By using the connection marks between dots, a line chart emphasizes the implication of trend relationships, which is more suitable for exploring the abstract trends of the confirmed cases [7].

As the data is recorded by the confirmed date, a filter on the time dimension can be implemented to show the state on a certain date or the trend within a period.

4.2 Policy timeline

A timeline is a graphical way of displaying a list of events in chronological order [9]. In this way, it is reasonable to build a sequential timeline when dealing with a list of policies with the announced date. The main function of this timeline is to convey to the user the time-related information, presenting a view of released policies visually. The type of policy in different areas is encoded by distinguishable colors.

4.3 Geographical Information of Confirmed Cases

The geographical distribution of patients can be represented using a Choropleth map [10]. The number of patients in each region is encoded as the color on the map. The regions are shaded in similar color and the darkness indicates value of an attribute. Each color is mapped to a value or an interval of value. This allows users to compare the confirmed numbers across the regions clearly. Each case can be represented as a point located on the map using its corresponding longitude and latitude. The size of the map indicates the number of a specific case. The map should also be used as an interactively filter allowing users to select a province and highlight or filter the information in this province. The Choropleth should also be able to be filtered by the time frame to see the temporal change of the infected cases. It should be possible to filter the regions according to the number of confirmed cases to find the regions which have patients exceed or below a specific number.

4.4 Demographic and provincial Composition of Confirmed Cases

To present the demographic and provincial composition of confirmed and deceased cases in different groups, pie charts [7] could be the most common choice. Using pie charts, each single attribute values (e.g. an age group in the attribute age) can be encoded each single with area mark and the angle channel represents their respective percentage. Bar charts can be more accurate than pie charts since they encode the absolute values as the height of bars, but the scale can become a problem when the relative difference of different groups tend to be large. Therefore, we choose to use the bar chart only on the provincial data to present the ranking and pie charts for showing the composition of each group.

The Sankey Diagram [11] can also be a good choice to show the composition of different attributes. It is a type of flow diagram where the width of each flow is proportional to the flow rate [12]. The flows show the internal links among different attributes, which illustrate the same data from different dimensions at the same time. Compared to pie charts, it encodes the amount into the width channel, which avoids the indistinguishable differences when two groups have similar angles. So it could be a better technique for our tasks.

As the data is also recorded the confirmed data of the cases, a selection function for the time dimension can be add to show the state on different dates.

4.5 Infection Network

To investigate the spread path of the infections among people, the directed graph can be used to represent the relationships. Each patient with an ID is encoded as a node, while the causal infection is encoded as a directed edge pointing from the spreader to the infected person. The group infections information can also be encoded as the color mark for the nodes, which will help users quickly distinguish certain group cases.

4.6 Regional Information

The regional information that is of our interest to show for each city contains the elderly people ratio and student ratio, and the counts of related facilities (e.g. the amount of kindergartens, universities, and nursing homes). As they do not split a certain total amount, bar charts can be a proper way to show the ratios and amounts of each city.

5 Realization

Our visualization tool is built using Tableau [7]. It is a powerful platform which can take in many kinds of data and plot elegant graphs with speed and ease. It is also capable of establishing physical or logical links among multiple data files, and provides underlying support to help us focus on visualization rather than front-end techniques.

In our implementation, the multiple graphs are organized in two dashboards: the first one shows the overall information about the national trends and compositions, while the second dashboard shows the part of data that are recorded with detailed information, of their density distribution, the infection networks, the information of group cases, etc.

5.1 Numerical Evolution of Daily COVID-19 Cases

A bar chart is also added, as it gives a good indication of the ranking of infections in each province on a specific date or within a selected period.

Four stacked area charts are implemented with the function of province and duration filters. A list of the province name will be provided for choosing a single province to view. A small map with administrative districts will also be employed as a filter. When hovering over the district, the corresponding segments of the stacked areas in each chart are highlighted, together with the related part in the bar chart for the rank of provinces and pie chart for the composition of provinces. When a district is selected, only the related provincial data, instead of a stacked one, will be presented in four area chart.

As for the selection of duration, a slider, together with type boxes, allow users to select a period that is interested and necessary to investigate. This filter works on both the stacked area charts and the combined charts for national trends of daily confirmed cases. By selecting the starting and ending date, the charts will automatically be trimmed by the time and then adjusted to fit the given view window. Select a single date using the bar of that day on the combined chart is also allowed, which is convenient to show the state on an exact day.

5.2 Policy timeline

By plotting policy along a time axis at when they are released, a set of dots are drawn on the timeline. When hovering over the dot, the date will be highlighted, and a floating window will become visible, showing the type, date, and policy details. The three most critical policies related to the type Alert, Education, and immigration are used as reference lines with different colors in the national confirmed cases daily charts, indicating the date of the announcement of the policy.

5.3 Choropleth Map of Confirmed Cases

The Choropleth map is implemented with the aforementioned time filters and number filters. The filters use sliders as well as type box. The map can be chosen to show per province or per city. The cases are shown as points on the map and these points and can be chosen to set as visible or not. When hover over a point, it shows the confirmed number of the case and the name of the case. The opacity of the points are set as 80% for better visual quality of overlapping points. The map allows zoom in and zoom out to see the detailed and overall information. The map is linked to the bar chart and the Sankey diagram on the same page with date as their external key.

5.4 Demographic and provincial Composition of Confirmed Cases

Though our first choice to realize the visualization of this composition is the Sankey diagram, but only 5616 cases are recorded with their detailed information of each attribute (age, gender, city, confirmed date, etc), with some null values included. So for the detailed data dashboard, this information is shown using the Sankey diagram. But for the other dashboard, even though we have the daily confirmed cases of each group (each attribute value), each attribute is recorded in separate files with only the number of cases for each attribute values, making it impossible to make connections between these attributes. Thus, there is no way to identify each case (a single patient) and its values for each attribute in order to distinguish its flow. Therefore, we have to use three pie charts to show the decomposition of the same population as for different attributes, which is also informally referred to as ”donut charts” in many places. A bar chart is also added as it gives a good indication of the ranking of the number of infections in each province on a specific date or within a selected time period.

The Sankey diagram is built using a Tableau extension ”ShowMeMore” [13]. In order to keep as much data as possible, we did not remove the data with partial null values for certain attributes. These flows will directly crossing the missing attributes, without joining into any group of this attribute.

5.5 Infection Network

The infection network is also realized using the extension ”ShowMeMore”. We use the force-directed graph [14] [15] to visualize this network relationship. Patients from the same group are encoded with the same color in order to distinguish different group cases. Other individual patients which do not belong to a group are encoded with individual colors. Repulsion forces exist for nodes from different groups and attraction forces for nodes in the same group. Spring forces on the directional edges from the source node (by whom the patient is infected) to the target node (the patient) keep causal infection cases stay close to each other. In order to have a maximum network, the patients with missing information of the person he is infected by, but belong to a certain group infection, is added to the same group with a random link to a person in that group infection (representing he is infected by a person in that group infection).

5.6 Regional Information

The regional information is separated to two bar charts. The upper bar chart shows the statistics of kindergartens, universities and nursing homes. The lower bar chart shows the elderly alone ratio and elderly population ratio. The y axis indicates the number of quantity and the x axis indicates the province name. When click in a province in the map it shows the statistics of the corresponding province. When the ”show city” in the selection menu is chosen, both bar charts show the statistics of the whole country.

6 Use Cases

To demonstrate how our developed visualization tool can be used to perform the analysis, we will present examples to deal with the questions stated in the task analysis.

Could the evolution of COVID-19 in South Korea be described properly by the graphs?

Fig. 1 is a comparison of the cases in different time lines on the map. It shows the first case appeared in Incheon. Then next case appeared in the nearby province Seoul in a few days, followed by the south-west of the country, However, the largest case group is located in Daugu province. Overall the COVID-19 cases mainly spread in two regions that have darker color, one at the north west and the other at the east.

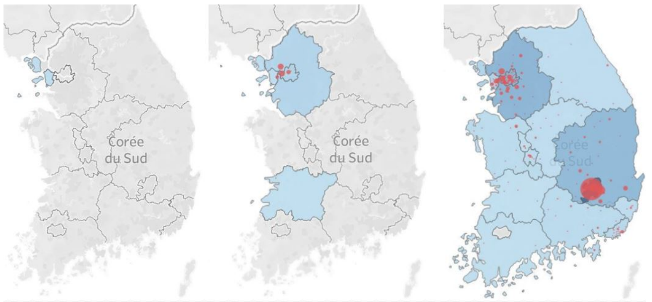

Fig. 1: The spread of COVID-19 over time. (Left: 2020/01/20. Middle: 2020/02/01. Right:2020/06/29.)

From the line chart of the national daily confirmed cases (Fig. 2 left), users can observe that most of the cases were confirmed between February and March. It can be clearly observed that the virus broke out in the second half of February, peaked at the end of February and the beginning of March, and then declined rapidly before remaining in a relatively stable state. From the stacked area chart (Fig. 2 right), the same information is also conveyed as the Choropleth map. Switching to the area chart of the accumulated confirmed cases by province, users can distinguish the three provinces with the highest number of infected people, which are Daegu, Gyeongsangbuk-do, and Seoul, corresponding to the two regions marked with darker blue on the map.

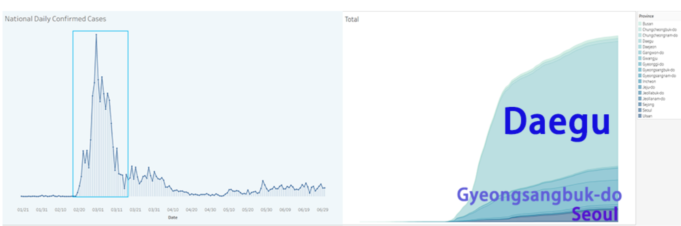

Fig. 2: Left: National daily new confirmed. Right: Stacked area for confirmed cases in each province.

Do the infections and deceased numbers follow certain demographic/geographical/temporal patterns, based on the selected attributes?

The two provinces (Seoul and Gyeonggi-do) have the highest numbers of kindergartens and elementary schools also have high number of confirmed patients. Fig. 3 shows the province Gyeonggi-do has the highest number of kindergartens and elementary school on the upper bar chart. Kindergartens and elementary schools may be the places highly likely to transmit virus, and the number of kindergartens and elementary schools can also in some way indicate the population. The province that have highest elderly alone ratio and elderly population ratio is the fourth most infected area. Another reason that this province(Gyeongsangbuk-do) has a large amount of patients could be that it surrounds the most infected province. Same as the province Gyeonggi-do that surrounds Seoul.

From the Sankey graph Fig. 9 (appendix) we can see the two largest infected age groups are 20s and 50s. While in the most infected province Dague, the two largest age groups are 20s and 40s (Fig. 10 in the appendix).

Is it possible to distinguish causal infections and track the spread of the virus among the population, or distinguish some super-spreading events?

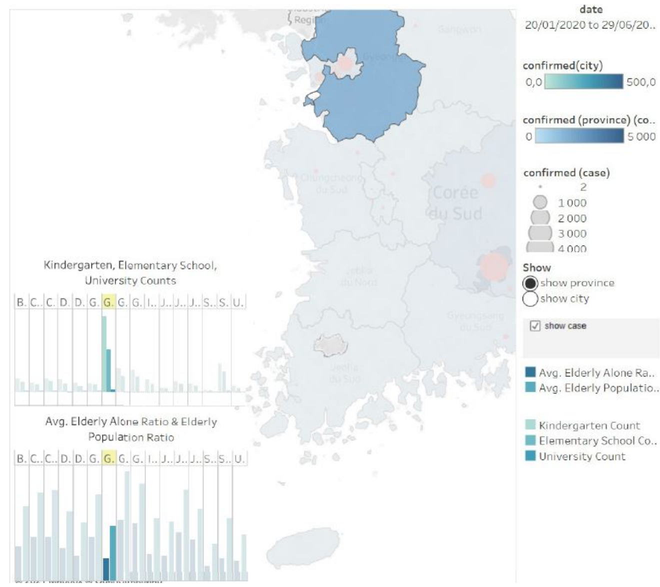

Fig. 3: Two bar charts with highlighting on Gyeonggi-do province.

The map with the timeline shows the spreading of COVID-19 cases geographically, but the interaction between cases are not shown and it is not certain how the virus gets spread. From the infection network some transmission chains of the disease could be discovered, and the spreading across multiple cases can be traced.

Fig. 4 shows the force-directed infection network graph constructed from the recorded data, with the left one being the cases on the first day of the records, and the right one the complete network. Different clusters are pushed away because of the repulsive forces among nodes, and nodes with connecting edges or belonging to the same group stay together by the spring force on the edges. By applying the date filter to the graph, the infection network on different dates can be shown.

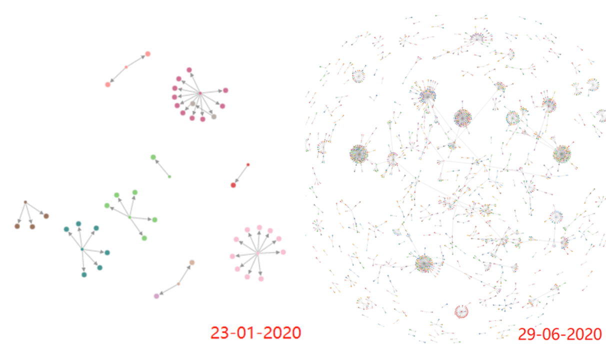

Fig. 4: The infection network graph. (Left: 2020/01/23. Right: 2020/06/29.)

Zooming into the network and with the selection and filtering functions, one can distinguish the interesting clusters. For example, Fig. 5 shows the group infection called Guro-gu Call Center, in Gangseo-gu city, Seoul, with the center patient being selected. Even though it is unclear how he was infected, the close contacts that were infected by him have been recorded and shown around this center node. And if we set the date to be a week after, further chains from this group infection will also appear as is shown by the graph on the right of Fig. 5. The two nodes circled by blue can be immediately distinguished as the key spreaders in further infections (super-spreaders who have infected many people), and by checking the node tool tips with the patient ID and the confirmed city, it is possible to establish the transmission path of the virus on the map, carry on postmortem analysis on the intermediate failures in monitoring, and gain experience in taking proper measures to curb further transmissions.

In addition, the network provides epidemiologists a tool to track the successors of the virus during the spread, therefore enables them to quickly trace back the precedent cases from the first occurrence of certain variants.

With the network, it is easy to spot some clusters of a high concentration of linkages. These group cases include churches, companies, clubs and other social organizations, which alerts policy makers and the public of the occasions with high dangers.

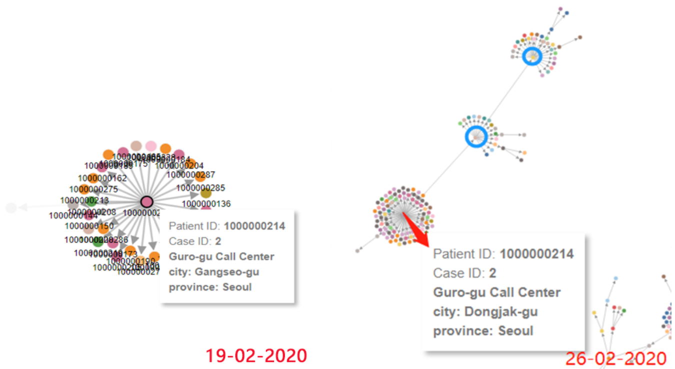

Fig. 5: The group infection "Guro-gu Call Center" in the network.

Do certain policies exert remarkable impacts on curbing the spread of COVID-19?

To evaluate whether the announcement of the policy has an impact on the infection, it is necessary for users to observe the trends right after the policy is released. This can be done by selecting a period of 2 weeks, with the date of the announcement as the starting date. Three policies on the combined chart will be the focus.

Fig. 6: Trends of daily confirmed cases with relative policies.

As the highest infectious disease level(4) was released on 02/23, the user can select the duration from 02/16 to 03/01. From the trends, it can be clearly observed that this alert hardly had any impact on infection control since it turned up after the alert was announced. Especially in Daegu, the peak number of infections in a single day was reached 6 days(02/29) after the alert.

For the date when all the schools were closed(03/02), nationally, the new confirmed cases were dropped after the closure of schools. After selecting the duration, users can see that on 03/03, the percentage of the infected 10s and 20s age groups were 4.01% and 29.32% respectively. The 20s group in Fig. 7 reached a peak of 30.07% on 03/07 and then dropped to 28.08% on 03/16 and kept at approximately 27% afterwards. However, for the 10s group, it raised to 5.25%. This will give an insight for the users who would like to investigate whether the closure of schools has impact on the infection rate among students.

Lastly, the immigration regulation of mandatory 14-day quarantine for all the overseas countries was finally released. The national trend dropped quickly after the policy. Seoul, which is surrounded by Gyeonggi-do, went to second place in Fig. 8. Since Seoul is the capital of South Korea, it has an international airport, which must be a focused target for incoming overseas cases. Thus, it is reasonable to see the infection increased dramatically in Seoul since all the travelers from outside the countries will be quarantined in Seoul or the surrounding province, Gyeonggi-do and Incheon. Fig. 8 shows users that the number of infection declined. For the people studying the necessity

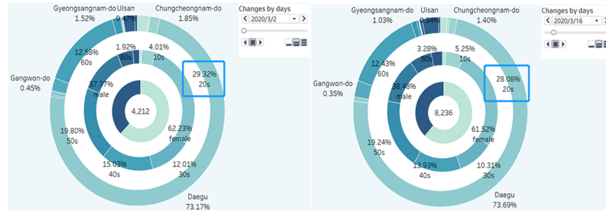

Fig. 7: Demographic composition of different province, age, and gender groups. Left: 2020/03/02. Right: 2020/03/16.

of quarantine, it can be an evidence to prove that after the mandatory quarantine, the government succeeded in controlling the number of infections.

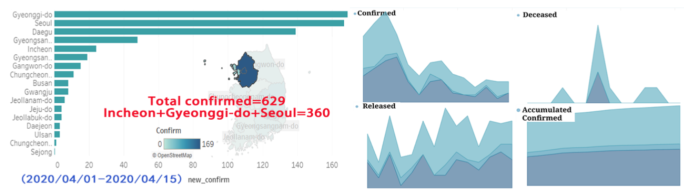

Fig. 8: Left: The rank of provinces. Right: The trend of new cases. Duration of 2020/04/01 - 2020/04/15

7 Discussion and Conclusion

In this work, an interactive dashboard is built with multiple visualization techniques implemented, including the bar chart, the line chart, the donut (/pie) chart, the force-directed network, the Sankey graph and the Choropleth map. Using the interactions, some geographical, temporal and demographical patterns of the COVID-19 cases can be discovered and exploratory analysis can be made. The transmission of the virus can also be discovered using the network graph. Various interaction functionalities including selection, highlighting, tooltips, and filters on numbers, dates and specific regions are implemented for higher level demands for exploratory analysis. Overall, the visualization allows users to explore and analyze the data, and answer relevant domain questions.

Our visualization tool can be used to dig into the reasons that affect the number of people getting infected, analyze the correlations of the factors and establish models in making predictions and analysis. The network allows researchers to trace the spread and evolution chains of the virus, and policy makers to carry on postmortem analysis on the intermediate failures in monitoring, and gain experience in taking proper measures to curb further transmissions. Our tool also helps users to see the effectiveness of the policies on the curbing of virus across the timeline. It allows users to observe not only the overall trend, but also the impact of the policy on the target groups by study the changes of the group distribution among population. Making connections between the type of group cases and the policy announcement would be a further function, since there are policies that are related to the limitation of the size of activities or the oversea regulation.

References

[1] Statistics of covid-19 coronavirus pandemic. `https://www.worldometers.info/coronavirus/`. Accessed February 01, 2021.

[2] Interactive data visualization: See trends. `https://ec.europa.eu/eurostat/web/covid-19/visualisation`. Accessed December 30, 2020.

[3] Laboratory-confirmed covid-19-associated hospitalizations. `https://gis.cdc.gov/grasp/COVIDNet/COVID19_3.html`. Accessed December 30, 2020.

[4] Development of covid-19 in graphs. `https://www.rivm.nl/coronavirus-covid-19/grafieken`. Accessed December 30, 2020.

[5] Covid-19 coronavirus data dashboard. https://informationisbeautiful.net/visualizations/covid-19-coronavirus-infographic-datapack/. Accessed December 30, 2020.

[6] Wikipedia: Superspreading events. https://en.wikipedia.org/wiki/Superspreading_event. Accessed February 02, 2021.

[7] Tamara Munzner. *Visualization analysis and design*. CRC press, 2014.

[8] A complete guide to area charts. https://chartio.com/learn/charts/area-chart-complete-guide/. Accessed 2019, M Yi.

[9] Phong Nguyen, Kai Xu, Rick Walker, and B.L. Wong. Timesets: Timeline visualization with set relations. *Information Visualization*, 15, 10 2015.

[10] Waldo R Tobler. Choropleth maps without class intervals. *Geographical analysis*, 5(3):262–265, 1973.

[11] Henry Riall Sankey. The thermal efficiency of steam-engines. In *Minutes of the Proceedings of the Institution of Civil Engineers*, volume 125, pages 182–212. Thomas Telford-ICE Virtual Library, 1896.

[12] Wikipedia: Sankey diagrams. https://en.wikipedia.org/wiki/Sankey_diagram#cite_note-2. Accessed February 02, 2021.

[13] Infotopics: apps for tableau - showmemore. https://appsfortableau.com/shop/extensions/show-me-more/. Accessed February 02, 2021.

[14] Thomas MJ Fruchterman and Edward M Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.

[15] Stephen G Kobourov. Spring embedders and force directed graph drawing algorithms. *arXiv preprint arXiv:1201.3011*, 2012.

APPENDIX

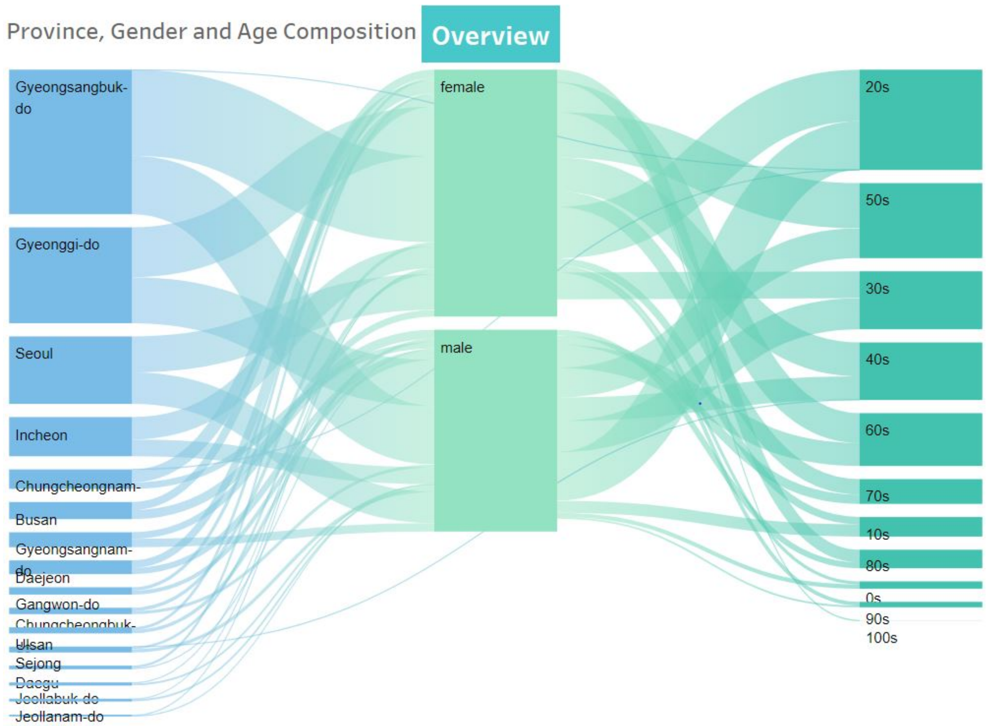

Fig. 9: Sankey graph - demographic composition of confirmed patients.

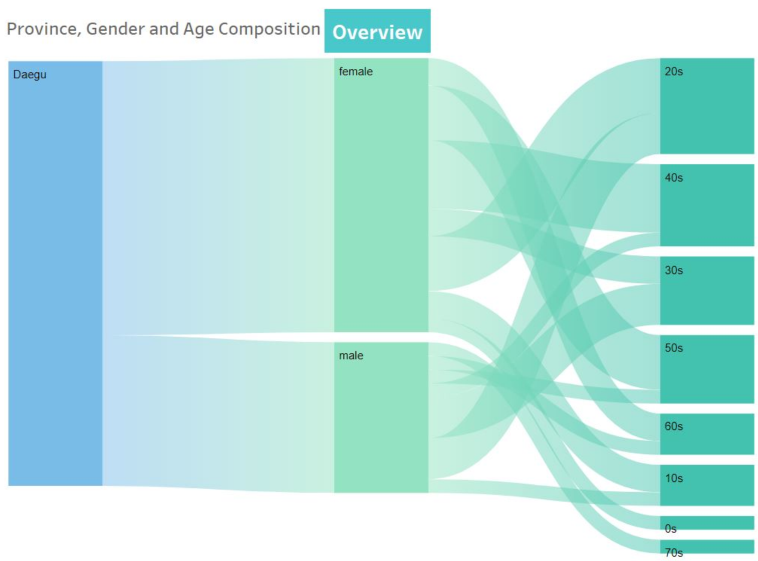

Fig. 10: Sankey graph - demographic composition of confirmed patients in Dague.