GEO-TEMPORAL VISUALIZATION FOR TOURISM DATA USING COLOR CURVES

by

In Kwon Choi

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science
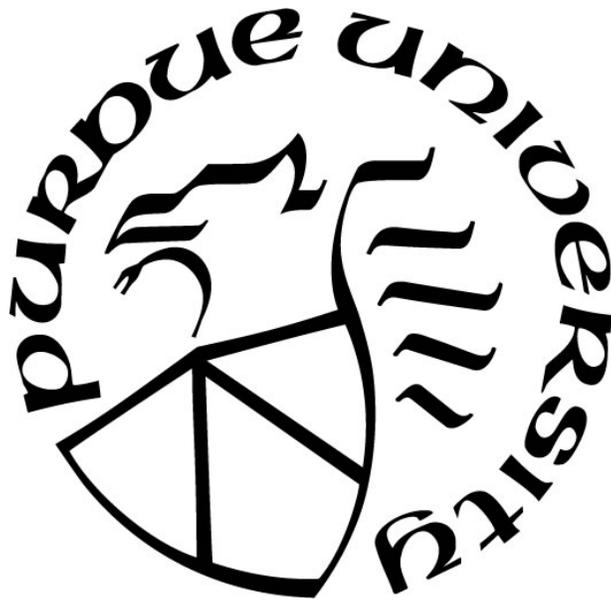
Department of Computer & Information Technology

Indianapolis, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Shiaofen Fang, Chair
    Department of Computer & Information Technology
Dr. Yuni Xia
    Department of Computer & Information Technology
Dr. Jiang-Yu Zheng
    Department of Computer & Information Technology

Approved by:
    Dr. Shiaofen Fang
      Head of the Graduate Program

To my family

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Choi, In Kwon MS
Institution: Purdue University
Degree Received: May 2019
Title: Geo-Temporal Visualization for Tourism Data Using Color Curves
Committee Chair: Shiaofen Fang

For individuals in the tourism industry and other businesses, the department of tourism in the government, or the individuals who are planning a travel, the data of tourist population movement can be a valuable resource that can uncover insights that could bring more profit and more tourists, or make the trip more enjoyable. As visualization is an effective way of conveying information with multiple dimensions, we would like to visualize the geo-temporal floating population data of tourists and residents in Jeju island in the Republic of Korea in two-dimensional space. In this study, we introduce the two methods we have implemented for visualizing the geo-temporal data using color curves as the representation of time dimension. We use the dots as the markers of floating population, and each color of dots represents the 24 hours of a day. In the first method, we plot the colored dots directly on the map, thereby coloring the area the data represents. In the second method, we plot the same dots inside a semi-transparent circle divided into arcs that represent each month of a year. The user can compare the population of tourists and residents between the different times of a day, the different months and the weather conditions to analyze the floating population in the given area.

1. INTRODUCTION

Geo-temporal data, which contains the geographical location data and the time data, is useful resources in a wide range of domains. Because the data can provide the overall trend or changes in certain geographical areas, it is possible to observe the changes in landscapes over time or the volume of traffic throughout a day, for instance. But as Zhong et al (2012) point out, visualizing the geo-temporal data presents several challenges in processing and utilizing large-sized dataset, choosing the right representation form, and combining the methods from different domains such cartography and information science. Therefore, various techniques such as timeline chart, diagrams with arrows, animations of images have been proposed and utilized. In this study, we present the two new methods of visualizing geo-temporal data as addition to the existing techniques.

1.1 Visualization of Floating Population

In the tourism industry and other businesses as well, the information on the movement of population in the neighborhood is a critical resource. Because it provides the information on the popularity of different locations, based on the information, one can make a decision on which place to focus more attention and resources compared to other candidate place. The city officials can also utilize the information to plan the routes of public transportation that can serve the most people, for example, or plan for the traffic management after examining which place is expected to have high traffic in which time of a day. The travelers can also plan their trips based on the floating population data, for instance, to find out when would be the best time to visit a certain tourist spot without being overwhelmed by the crowds. In addition, by plotting the data for multiple years, it is also possible to observe the rise and decline of different areas, which would be useful information for the land-owners or the researchers studying the history of neighborhood.

But the locations recorded in this population data would not be easily combined with the population values in the head without the help of some kinds of visualizing measures, such as a map. Visualization is one of the effective ways to convey data that can be difficult to grasp when given as texts and to provide the quick summary of multiple dimensions combined. As Aoyama et al (2007) state, the goals of visualization include providing the context to the raw data and

uncovering the trends in the data through the graphical representation, which enables the user to have deeper understanding of the data. Then the task becomes finding the most effective method for visualizing the given data. In this study, we have chosen to use 2D visualization as it is hardly possible to avoid the problem of distortion and occlusion of the spatial dimension when visualizing in 3D space.

Though 3D visualization has a definite advantage in its higher capacity for possible graphical positioning of elements compared to the 2D space, we believe our dataset can be represented more accurately in 2D space. For example, in our dataset, there are hundreds of locations to be plotted on the map. Because of the large number, it would be unavoidable to have some locations blocked from the view by other locations. Furthermore, finding out the geographical position may also be difficult as the view is distorted when it is seen from the sides, rather than from the top. Therefore, we have decided to plot the dataset in the 2D space.

In this study, we introduce and compare the results of two methods we have implemented for visualizing geo-temporal data in two-dimensional space using colors. The data contains the floating population of tourists and residents in the various locations in Jeju Island, a popular tourist destination in the Republic of Korea, in 2015.

1.2 Visualization using Color Curve

With the geographical data represented on the map, we have used the color as the indicator of temporal data. In the HSV color model, which represents Hue, Saturation, and Lightness, we are using the hue to distinguish between the data points with different values in the temporal dimension. As seen in the below figure, we use 24 colors which represent an hour in a day. The colors gradually transition from one color to another color as they represent the time as a continuous dimension.
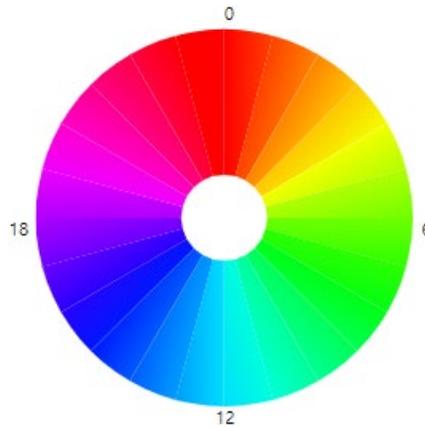
Figure 2.1 Time legend used in the visualization

We have divided the 24 hours to four ranges, six hours each. The colors from the degree 0 to 90 expresses the time between midnight to 6AM. The colors from the degree 90 to 180 expresses the time between 6AM to 12PM. The colors between the degree 180 to 270 shows the hours between 12PM to 6PM. Finally, the colors between the degree 270 to 0 represents the time between 6PM to midnight. We have decided to represent the colors in the form of a circle, because circle is a familiar shape for continuous temporal scale, as in a clock. The difference is that a clock usually represents 12 hours, while we display 24 hours. To avoid confusion, we have added the labels to provide additional information on the hour each arc is representing.

For representing the time dimension in the 2D space, using a color is more suitable for our dataset than adjusting other properties such as sizes. Because we use all 24 hours in plotting the data points, it would be difficult to distinguish the dots by changes in their sizes in 24 levels. Moreover, the advantage of using color is that the hours that are nearby have similar hues. Therefore, a user can notice easily when the distribution of population over time in one area is largely different from another area, for instance, when one location mostly has green dots and the other location has mostly red dots. Then even before looking at the legend, the user can recognize that the two hours represented by two colors are not going to be closely located. Conversely, if the area has mostly blue and light blue dots, the user can understand that those two hours are likely to belong to a same hour range—morning, afternoon, evening, night—as their hues are similar to each other.

In the two methods we have tested, we have used the above time legend as the reference of all colors used in the dots. In the first method, we draw the colored dots representing the number of

population directly on the map. In the second method, we have created the circles with 24 areas which represents each month of a year for tourists and residents. Then the dots with colors are placed inside the circle.

1.3 Thesis Organization

Chapter 1 provides a brief introduction on the study this paper discusses. It provides the background on the nature of data and the graphical dimension of visualization we have chosen for the study. object of the study and the application we have developed. Then it discusses the representation of temporal dimension using colors, which is the focus of the study. Chapter 2 discusses the previous studies in relation to the visualization of the tourist movement data and the visualization of geo-temporal data in a 3D space and a 2D space. Chapter 3 describes the techniques used for the development of visualization, such as generating the map, adding the circles and dividing them into arcs, placing the dots on top of the map in a separate layer, and calculating the number, position, and colors of each dot based on the dataset. It also provides the details of dataset and the normalization strategy we have used to reduce the gap between the highly populated area and the mostly vacant areas. Chapter 4 presents the application interface and describes its functionalities, such as data filters and the hover to examine the details of the arcs. This chapter also provides the analysis results from the visualizations, as well as the comparison between the two methods we have implemented to visualize the dataset in terms of their advantages and drawbacks. Chapter 5 summarizes the study and its result, and it provides the thoughts on the future works for the study.

2. RELATED WORKS

Using colors for depicting the intensity or the volume of an attribute in different geographical areas is a commonly used technique in data visualization, as we often see in a choropleth map or a heatmap. But adding another dimension while preserving the readability of data in other dimensions and not imposing heavy cognitive loads on the viewers can be a challenging task. While there are numerous techniques presented, in this paper, we focus our attention on adding the temporal dimension to the map containing geographical information in 2D and 3D spaces.

There have been many studies that proposed the methods for visualizing geo-temporal data in a 2D dimension and a 3D dimension. For 3D dimensions, some studies have utilized the 3D icons that contain the temporal data in themselves, and some studies have used the vertical position of the markers in the 3D space as the indicator of time. In 2D dimensions, there have been studies that have used a chart, such as a bar chart or a line chart, for expressing the temporal data along with the map. In the studies, the chart is either incorporated in the map or is shown as a distinct portion near the map.

The dataset we have used for the study contains the floating population of the tourists and residents in different locations, in every hour of a day in each month for a single year. Understanding the movement of tourists is important for a number of parties including the businesses in tourism industry targeting the tourists and the local government that seeks to bring more tourists into their regionality. For instance, the business owners who target the tourists can use the population movement information to find out the popularity of the different locations in each time period to effectively advertise their business when there is a large crowd of travelers. The local government can also use the data to generate insight on what kind of places are attracting, or not attracting, the tourists, based on the popular locations they visit in certain hours or seasons.

At the same time, the differences between the places the tourists tend to visit and the residents tend to visit, as well as the paths the tourists and the residents take in the neighborhood are also being widely studied to understand the impact of tourists in the local culture and the sentiments of residents regarding the presence of tourists in their neighborhood. For the rich information they can offer, various approaches have been taken to visualize the movement of tourists.

2.1 Visualization of Tourist Data

One of the most popular topics in the tourist data visualization is visualizing the flow of the tourists by drawing their trajectories or marking the number of tourists on the map. Girardin et al (2007) present a visualization of the paths of tourists based on the photographs the tourists have taken over the years and posted on the web. By extracting the longitude and latitude information from the photographs, they identify the popular venues for the tourists and mark them on the 2D map. Also, they draw the raw trajectories of each tourist as an overlay on the map, so the popular routes are highlighted as they are thicker compared to other less visited routes.

The visualization method of Chua et al (2016) is a different approach in describing the trajectory of tourists. They present a visualization tool *FlowSampler* (figure 2.2) that draws the directional flow of tourists in one of the tourist destinations in southern Italy to study the movement of tourists. By diving the area map into grids, they visualize the movement of tourists between every two region by painting the route in different colors. The stronger the color is, the more popular the route is, and vice versa. In addition, the map can represent the direction of movement by multiple lines with arrows that are pointing to the most frequent directed movements between two areas, for instance, from the south to the southeast. Filtering by time, the user can examine what is the general direction of tourist flow in certain seasons in certain areas. In this visualization, the temporal data is provided separately under the map as a bar chart. A bar in the chart shows the number of people uploading a posting in the selected location on that day. Therefore, it can provide the comparison of the size of population on each day, which is a piece of information not supplied in the main visualization.
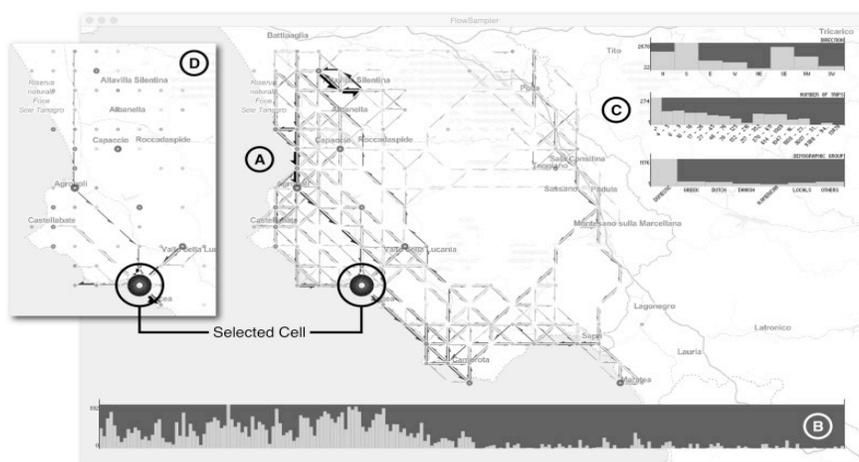

Figure 2.2 *FlowSampler,* a geo-temporal visualization for tourist flow

Instead of visualizing the movement of tourists as routes, Wu and Carson (2008) visualize the spatial distribution of the tourists in each area using the circle markers with various sizes. Each region in the southern Australia has a circle assigned to it, and the size of the circle conveys the information on the percentage of tourists who stayed in the region overnight. Then they compare the percentages at different times by producing several visualizations, as the visualization itself does not display the temporal dimension on the map. While this method does not show the flow directly, it is useful when the user is interested in comparing the number of people in each location.

2.2 Visualization of Geo-temporal Data in 3D space

Andrienko et al (2004) present a visualization based on the Space-Time-Cube by Hägerstrand, which represents the time dimension as the height of the 3D cube and the spatial dimension as the bottom surface of the cube. In Adrienko's visualization, there is a 2D map at the bottom surface area. The occurrences of events are represented as the vertically placed dots. For instance, the most recent event is placed nearest to the map at the bottom, and the most distant event in the farthest point from the bottom. More information about the events such as the number of occurrences can be depicted by changing the size or colors of each dot. Moreover, the 3D cube is linked with a separate map that displays the same area, along with the highlights on the location that is being explored.

Based on the space-time cube, Kwan (2000) have also presented a "space-time-aquarium" (figure 2.3) that plots the trajectories of people of different race over the 2D map at the bottom. Using the visualization, it is possible to see where the population is concentrated in early or late time. Hewagamage et al (1999) also utilizes the space-time cube with different shapes that are overlaid on the spiral inside the cube. The spiral is the trajectory, and the shapes represent different actives such as eating or observing, so that the user can see the start and end of each activity.
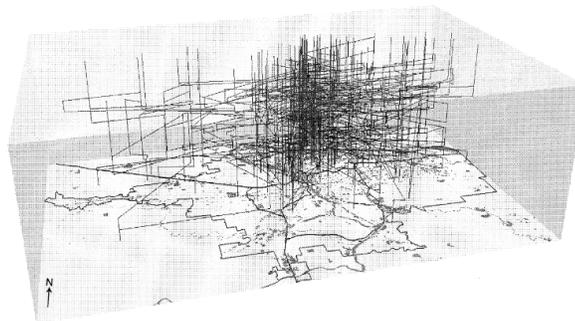
Figure 2.3 Space-time aquarium

Kapler and Wright (2004) also propose a visualization system that uses the height as the representation of temporal dimension. It is different from Adrienko's visualization in that the data points are represented as the points on the line graphs rather than the individual dots. The even-sized dots are connected to each other and form a fluctuating line above the 2D map at the bottom. It is also possible to compare multiple locations at once, by plotting multiple lines with different colors side-by-side.

Tominski et al (2005) propose a 3D visualization of geo-temporal data plotted by 3D icons on a 2D map. In the visualization, the multiple 3D pencil-shaped icons are displayed above the locations being explored. The tip of each icons is pointing to the location they represent. Each facet of the vertical surfaces of the pencil represents a single disease, and the facet itself is a vertical heatmap that shows the temporal pattern of the disease, with the vertical axis of each facet being the temporal dimension. A second type of icon is a helix, which also displays the temporal patterns of multiple diseases at once.

The approach of Thakur and Hanson (2010) is also plotting a pillar-shaped marker above the 2D map (figure 2.3). The marker consists of multiple disks with different width and colors and is vertically placed on the location it represents. Each disk is a representation of a single time period such as a year, and the bigger the value is, the wider the disk is. The color of each disk also represents the size of the value, and a lighter color represents a smaller value. As this marker uses both color and size to represent a size of value, it is more effective than a marker using only one of them in conveying the information. Therefore, by looking at the curve of a pillar and its color, the user can grasp the changes over the years in one look and compare multiple locations at once.
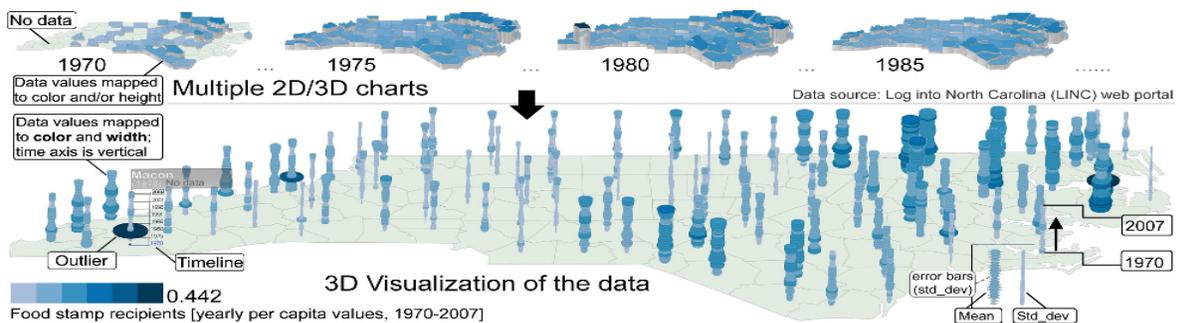

Figure 4.2 3D visualization of geo-temporal data using 3D markers and color gradients

Furthermore, Zhao et al (2013) have proposed a visualization that makes use of multiple visualization techniques. In representation of tourist activity in several regions, they have used

color and height of the 3D bars to mark the sightseeing activity and eating activity, respectively. The bars are also located on the radial chart that represents 24 hours for 3 days. The data is also shown with the geographic information, through the map in the center of radial layout. Also, they have produced another type of visualization that uses space-time cube. In the visualization, the activities are represented as 3D lines vertically positioned, along with the time chart at the side of cube.

2.3 Visualization of Geo-temporal Data in 2D Space

For describing the values within the temporal context in 2D space, one of the approaches is adding a marker that represents the temporal domain to the base map. The markers can be a symbol or a shape that shows the data by the differences in its properties such as the size and the color. In some studies, the researchers have used a chart as a marker itself, and have placed the charts plotting the temporal data on the base map.

Many studies have proposed the methods for juxtaposing a various type of charts directly on the map. In the study to identify popular areas by using the application users' check-in data, Wang et al (2012) present the visualization that displays the changes in the number of check-in users by month as well as the overall popularity of the areas in the map. In the visualization, they first color the areas in the map based on the popularity as a choropleth. Then on the areas that are highly popular, they juxtapose the histograms with three bars each, which show the increases in the check-in user numbers in the certain range of months. For instance, the first bar represents the increase between October and September, and the second bar represents between October and November. Where would be a limitation with the number of months that can be displayed at once, since if the histogram attempts to represent the entire year, the histogram in each location would become horizontally too long and block the data in neighboring locations. But it would be effective in showing how the popularity of multiple locations have changed in the recent few months.

As another type of the chart marker, N. Andrienko and G. Andrienko (2003) put an area chart on each location in the map. In the software CommonGIS, a change of an attribute over time is expressed as multiple small area charts. It is also possible to provide additional information by differentiating the color of each area in the charts. For instance, in the paper, they have assigned a red color to the charts with values higher than the national median, and a blue color to the charts with lower values. Depending on the value, it is possible to have a chart that is partially red and

partially blue. Compared to the above method by Wang et al, it is more scalable as range of time that can be expressed within the set length of a x axis can be as high as the user would like it to be, without the size of the entire chart changing and blocking the view of other areas. But at the same time, it is not as easy to learn in one look which part of the area represents a specific time period.

Instead of placing the temporal data as an additional layer on the base map, Drocourt et al (2011) present the temporal data outside the map using a radial layout. In their visualization of the glacier frontal positions in 10 years, a map is placed at the center of the circle. Then a line is projected from each location of glaciers to the outmost circle. The circle has 10 rings inside, which represent each year. The changes in the position of glaciers are shown as the line segments between each ring, with different colors per distance. The thickness of those lines depends on the absolute value of changes, thereby indicating the amplitude of change. Also, there is another set of markers on the outside of ring of circles, that provides the direction of changes in the glacier position.
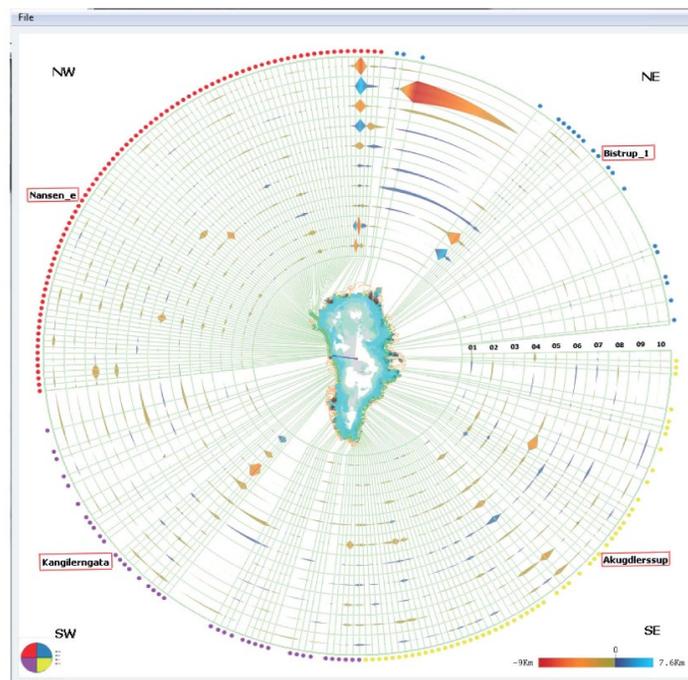
Figure 2.5 Radial layout of geo-temporal data

Some researchers have also used a shape as a marker for the map. Jänicke et al (2012) present a web visualization using the colored dots as markers on a 2D map. They use the dots with different sizes and colors to record the number of crime incidents in each area on the map. The size represents the number of incidents, and the color represents the type of crime incidents. While the idea of using the colored dots for visualizing the data is similar to our visualization method, the

difference is that Jänicke et al did not integrate the temporal dimension with the map. The crime incidents for the time period is displayed as a separate timeline at the bottom of the application, while in our study, the temporal data is represented inside the map by the color of dots.

Zhao et al (2008) have also demonstrated several techniques using a radial layout to represent the movement of people in a day. In the visualization, there is a choropleth map in the center, and the rings outside are divided into 24 hours, which are also colored using the same color scheme as the map to present the movement by time.

The method Shanbhag et al (2005) propose is a combination of several approaches. They plot multiple pie charts and small dots on the map, and also control the saturation, the brightness, and the hue of each area to provide the data on additional dimensions. In the map that visualizes the detail of a single academic plan, in each region, a pie chart shows the ratio of the seats in the school that are filled and the seats that are vacant. The density of the very small dots provides the comparison between each region on the number of students who need free or reduced meal support. The saturation supplements the information on the meal support by showing the capacity of the region regarding the meal support. The hue represents the different schools the students in the neighborhood are assigned to attend, and the brightness represents the average test score.

For the study of Shanbhag et al, there are three methods for incorporating temporal data inside the map, which are wedge, ring, and slice. The wedge method divides the region into three parts from the center, and each part shows the data for a single year, in a clockwise order. The ring method maps the earliest data in the center, then plots more recent year data in the outer ring, thereby creating the number of rings growing outwards. The slice method divides the region vertically with the subregions with the same width. The earliest year data is shown on the far left and the most recent year data is shown on the far right.
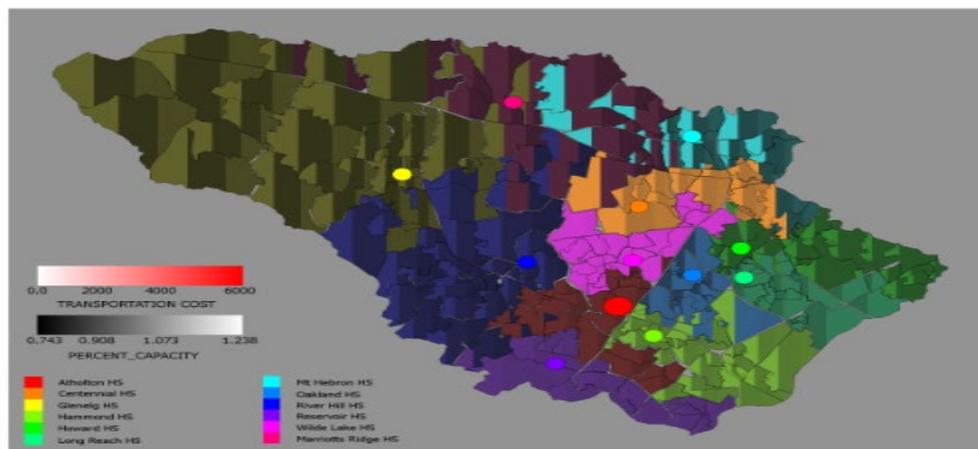
Figure 2.6 Visualization of geo-temporal data in 2D space using time slice

Guo et al (2006) also combine multiple visualization techniques to represent the data, but their approach is different in that they are using multiple separate graphs. They display multiple choropleth map showing each year along with the parallel coordinates, and several heatmaps.

Furthermore, Baskaran et al (2017) presents a method using the three types of time curve that can represent multiple time periods at once. The first type is a time circle, which is a basis for the color legend we have used in the study. In a circle, each degree represents a different hue, which can be used to express a different hour of a day. The second type of time curve is a 2D spiral, which can represent two periods of time, by differentiating saturation of colors of each ring in the spiral. Finally, the third type is a 3D spiral, which draws the spiral in 3D space to show the progress over time vertically. They have used the color curves to draw the trajectories of taxis in a city by hour and a day at the same time.
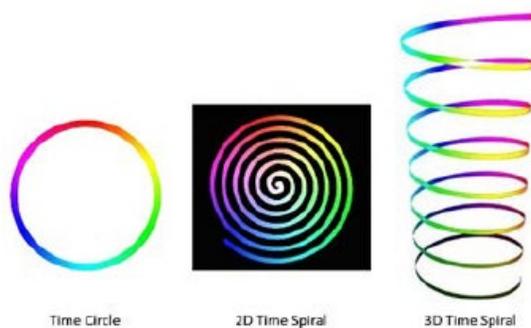
Figure 2.7 Three types of time curves

It would be also possible to display the aggregate of such traffic, as N. Andrienko and G. Andrienko (2008) have demonstrated in the paper. While their work does not involve temporal dimension on the map, they show the trajectories of cars in certain hour range by drawing the lines that express the summary of trajectories.

3. METHODOLOGY

In this study, we have designed the two methods for visualizing the floating population on the map in 2D space. The primary marker we have used is the dots with different colors that express different time in a day. The user can refer to the time legend to find out which hour the dot represents. The time legend, which is in a circular shape, displays a different color in each degree. The number of dots correspond to the volume of the population in the location. The first method places the colored dots directly on the map, and the second method places the dots inside the semi-transparent circles that are drawn in each location. The circles are divided into 24 areas that represent different months and different population types—tourist and resident—, and the dots are placed inside each area based on the month and population they represent.

3.1 Creating the Map

For generating the base map of visualization, we have used Leaflet, which is a JavaScript library widely used for creating interactive maps for web environment. Leaflet map is initiated when the user links it to the existing div that is going to be used as the map container, along with a center of map in longitude and latitude and a source of the map. The actual map is then added to the Leaflet container as an additional layer, which is called a tile layer. For this visualization, we have used a world map from OpenStreetMap. The center of map is set to [33.35, 126.54], which places the Jeju Island at the center of map div.

3.2 Plotting Data on the Map

We have used the D3 JavaScript library to plot the colored dots representing the floating population value in each location, which are added to the SVG canvas on top of the map. On the initial run, the program reads and plots the data in the two CSV files that contain the overall floating population of tourists and residents. When the user selects a different data through the filter, the program reads the matching CSV files. The CSV files consist of the rows that represent a single location, and the columns of each row contains the x and y coordinates of all dots and their color information.

For parsing the data, we have used Papa Parse library to read the csv data row by row. The parsed data is stored as the array of objects, with the location code in the grid (ex: "CB05B75C", which is an English conversion of the original Korean grid system coordinates code) as the key of each object.

Because the files we have used for the visualization contains additional data such as the coordinates and the color of dots, and each file has the data for a single year, not a single month as in the original files, the size of each file have grown to be more than 500 MB. Therefore, instead of loading all files at the beginning, we only read the files that are currently requested. But even so, as the delay is mainly caused by the plotting of dots, we could only alleviate the problem and the delay still remains when using the data filters.

3.2.1 Drawing the Arcs

In the second method we have used for visualization, we draw the circles that are divided into 12 arcs, which are further divided into the tourist and the resident area. The border between the areas inside the circle is drawn by calculating the start, the middle, and the end point in the twelve lines that separate each month. The tourist data is drawn in the outer area, and the resident data is drawn in the inner area of the circles.
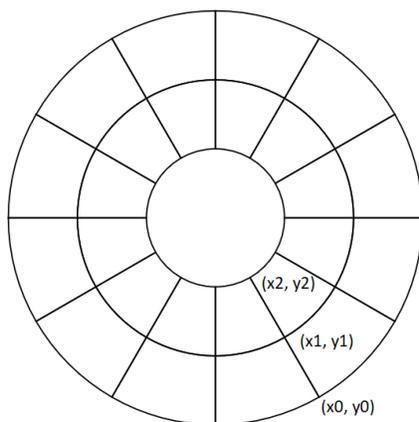
Figure 3.1 Arc used for visualization of tourist and resident data

This design enables the user to compare both population at once, by observing the density and the overall color of each arc. As the number of dots represent the population, the user can see which places are more frequently visited by the tourists or residents. As the color of dots represent

each hour, the user can also find out when the place is crowded or vacant. We use a circle and the arcs inside it as a marker because it is easier to associate continuity with a circle compared to other shapes such as rectangle or triangle. For example, the circular shape is common for clocks, which would be one of the most familiar objects that conveys continuous temporal information. The clockwise ordering of the arcs makes it easy for the user to understand the visualization. The circle itself can be regarded as a clock representing months instead of hours, since the position of each month is in line with the hour in a clock.
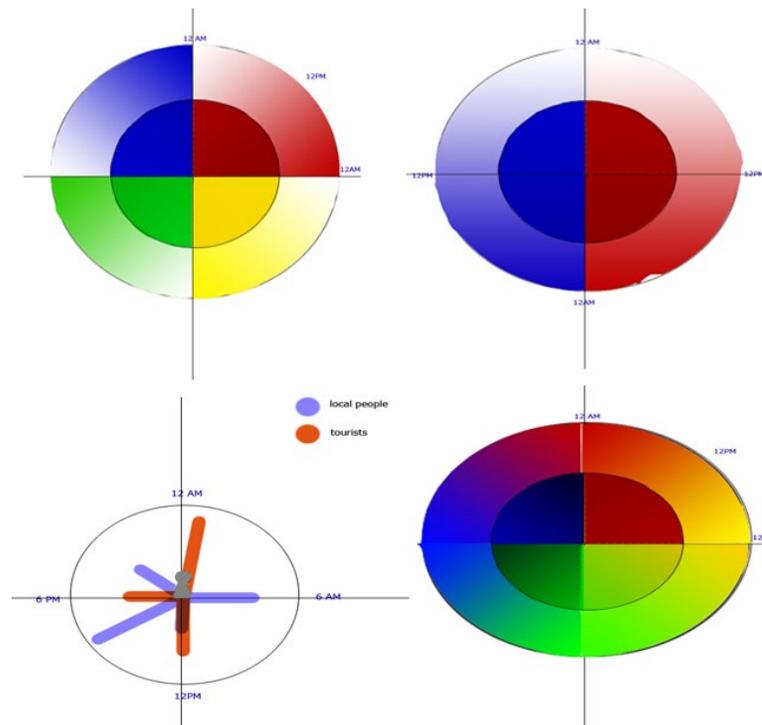
Figure 9 Time legend sketches with clock motif

The program first draws two circles with the common center point. Then the border between the arcs can be calculated as follows:

```
Function calculatePoints()
{
        for i in range (0, 12)
                x0 = circleCenterX + radius * 3 * sin(2 * pi * i/12)
                y0 = circleCenterY + radius * 3 * cos(2 * pi * i/12)
                x1 = circleCenterX + radius * 2 * sin(2 * pi * i/12)
                y1 = circleCenterY + radius * 2 * cos(2 * pi * i/12)
                x2 = circleCenterX + radius * 1 * sin(2 * pi * i/12)
                y2 = circleCenterY + radius * 1 * cos(2 * pi * i/12)
}
```

As a result of the function, we have three pairs of the x and y coordinates calculated for each month's borderline. x0 and y0 refer to the outermost point of the line, x1 and y1 is the midpoint between the tourist and the resident data, x2 and y2 is the innermost point of the line that is the closest to the center of circle. After all the coordinates are found, we draw the border by connecting the two pairs of vertices vertically using D3 path element.

3.2.2 Drawing the Dots

In both methods we have used for visualization, the positions of dots were precalculated and saved to CSV files before being sent to the visualization program to reduce the computation load in the browser. In the program, we use the D3 functions to draw the dots as the circle elements on the SVG canvas.

We have chosen to use the dots to represent population because we wanted to emphasize the context of data, which is about the movement of people in real life. Indeed, the raw values in the dataset are the actual number of people who were present in the location. Therefore, we can say every dot on the map represents a single real person.

In the design stage, we envisioned the dots relocated randomly each time the user uses the visualization. It has a practical advantage as it would assure the same chance for the dots in a highly crowded area to appear on top of other dots. Because it is possible that the dots of certain color appear more often in the front and become overrepresented by chance, having the dots to be relocated in each run can help the users adjust their initial impression that may have been

inaccurate. Nevertheless, calculating the positions and the drawing of several thousands of dots in every run turned out to be a too heavy computation for the program, and we could not implement it in the program.

For the first methods, which concentrate the dots near the center of circle, we have tested a few different representations before settling on the circular-shaped cluster. We wanted to address one of the primary issues we had with the second method, the geographical data being covered by the arcs. Furthermore, we also sought to implement a new method that blends better with the base map.

We first modified the arcs-method by removing the arcs and decided to draw the dots without any additional markers involved. First, we tried drawing a rectangle shape that are concentrated on each location. But the multiple rectangles on the map did not create a map that does not look too artificial and is easy to read, which we sought to create. While it succeeds in showing the most prominent color in each area, as the rectangle is too small, most of data points in the populated areas are not visible.
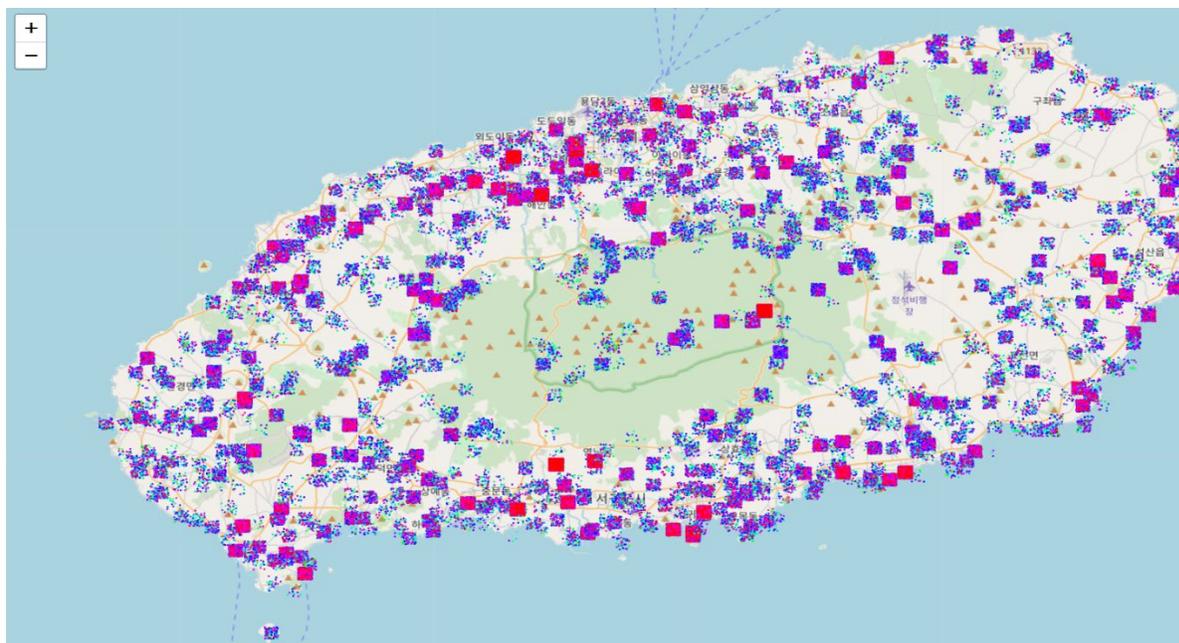
Figure 3.2   Test visualization with rectangle markers

Then we created a ring shape that wraps around the location it represents. But it had the same problems with the rectangle marker because it was still difficult to read the data as the data points were cluttered in the small ring. Moreover, we realized that we do not have to have an empty space in the middle, as most of the locations did not have any symbol or indicator in the base map. Also,

without the empty space, it was still possible to emphasize that it is the center of the entire cluster when we used the following method.

As the problem was caused by the high density in the small area the markers occupy, we decided to set the density to be different in each part of the marker. As a result, in the current method, we use probability to randomly assign the dots while preserving the high density in the center. First, we generate a random number between 0 and 1. If the number is less or equal to 0.6, the distance from the dot to the center is between the radius * 2 and radius * 3. If the number is higher than 0.6 and less or equal to 0.9, the dot has the distance between radius * 3 and radius * 5.5. Finally, if the number is higher than 0.9 and less or equal to 1, the dot has the distance between radius * 5 and radius * 10. As the range of probability becomes smaller as the number increases, the dots have higher changes of being assigned within the nearest area to the center point of circle. Moreover, the size of the area for each probability also grows as the distance increases.

For the second method that assigns the dots at different degrees based on the months they represent, we also use randomization to determine the position of dots inside each area. The distance from the center is calculated as a random number between the radius and the radius * 2 for the local data and between the radius * 2 and the radius * 3 for the tourist data. The degree of each dot is calculated based on its month, which becomes a random number between the value 0f (month – 1) and the month multiplied by 30. The final coordinates are calculated as the cosine and the sine of the degree in radian multiplied by the distance from the center.

3.3 Processing of the Dataset

3.3.1 Dataset Overview: Floating Population of Tourists and Residents

The dataset for visualization was retrieved from the Open Data Portal (https://www.data.go.kr/), a government-run public data portal in the Republic of Korea. The CSV data files contain the average floating population of the residents and tourists for 24 hours in each month in 2015 in Jeju Island, which is a popular travel destination in the Republic of Korea. The recording of population data is conducted by detecting the use of wi-fi signal in each area. Because Jeju island offers a free Wi-fi for the tourists given that they submit a simple form, the tourists and the residents can be recorded separately based on the wi-fi network they are using.

For each month, we have used 18 datasets. 10 datasets are categorized by the weather type, which are sunny, cloudy, rainy, and snowy, with 5 of them for the tourist data and the other 5 datasets for the resident data. Similarly, there are a total of six datasets for the day types, which are the holiday, the day before holiday, and the regular week days. Then there are the composite data that contains all floating populations for all weather conditions and days for the residents and tourists.

The same files are offered in different geographical indicators such as the street address, 300-meter grid code, and 250-meter grid code. In this paper, we have used 250-meter grid metrics for the ease in converting to the longitude and latitude. Each dataset consists of CSV files that contains the records for a single month, and the sizes are around 10MB for tourist data and 20MB for resident data. The original data has approximately 30,460 locations. With the current visualization methods, we could draw about 900 locations at once for the dots-only method and around 160 locations for the method using the dots and arcs.

Table 3.1 Sample data from the floating population dataset

Date	Grid Coor.	00-01	01-02	02-03	…	22-23	23-24
201503	다나 00C95B	0.666129	0.438387	0.315484		1.661935	1.17
201503	나나 84A88D	1.973548	1.735484	1.519677		3.094839	2.371613
201503	다다 09A03C	0.221935	0.195806	0.170645		0.336452	0.264194
201503	다다 10D02A	0.571935	0.487419	0.464839		1.031935	0.782903

3.3.2 Calculation of Geographical locations

In order to visualize the dataset, we had to go through multiple conversions of the geographical indicators to plot them on the SVG canvas, as the original dataset and the leaflet map are based on the different coordinate systems. Then the converted longitude and latitude have to be converted to the pixel locations to be plotted on the screen. As a first step, we used the supplementary CSV file that contains the grid codes and their corresponding x and y coordinates to extract the coordinate information of all locations. Because the original files follow a GRS80 coordinate system, which is a standard coordinate system in the Republic of Korea, the coordinates had to be converted to the WGS84 latitude and longitude that is compatible with Leaflet coordinate system.

For conversion, we have used the Pyproj library in Python, which is a library for the conversion between the different coordinate systems. Next, the resulting latitude and longitude are converted to the x and y coordinate pairs using the conversion function in the Leaflet library. As Leaflet is a JavaScript library, we put the pairs of coordinates into the Leaflet conversion function in the browser and downloaded the console output as JSON file. The last step is not necessary if one is using the Leaflet marker to make annotation. In our case, our marker is dynamically generated on runtime, so we could not use the Leaflet marker which takes an image file as input. Instead, from the conversion function in Leaflet, we have calculated the pixel locations of each coordinates pair to draw the dots. The output is the pair of the x and y coordinates set for the SVG canvas on top of the Leaflet map, based on which the arcs and the dots are drawn.

3.3.3 Calculating the Number of Dots

In the dataset, the values of floating population are given as floats. While the first step would be converting them to integers to use them as the number of dots, there is another consideration that has to be given. Because we are drawing the dots as they are, without further merging or combining between the dots, the value of each location cannot be different by a large margin from each other. The issue is that such large margin could cause some areas to have hundreds of dots when some areas have one or two dots only, thereby making it difficult to observe the areas with comparatively less population. In addition, with the multiple locations that are highly crowded, the total number of dots can exceed the amount the browser can process without crashing.

For example, in the figure 3.3, if we simply multiply all values by 100 and use them as the number of dots, the second location will have more than 100 dots already with those 5 hours out of 24 hours. The problem is exacerbated with the second method displaying both tourist and resident data at once, as then the second location will be having at least 200 dots for five hours' data. Since there are quite a few popular locations, the dots would grow over 1000 dots only after we plot five places that are as populated as the second location. Alternatively, if we choose to multiply by 10 to avoid this issue, the less crowded places as the third location would barely have any dots representing the area, although it did have some visitors. Therefore, simply multiplying by a certain number and using the raw values cannot be a good way of representing our data.

Therefore, the approach we have taken is dividing the values by a same number to reduce the differences between the highly populated and the least populated areas. The common number is calculated for each location, following the algorithm described below.

First, we calculate the median value of floating population based on the 24 hours' values in each month in one location, for the tourist and the resident data separately. Then as an output, we have a single median for each month in both datasets. Next, we calculate the median value again by comparing the twelve medians gathered above. The output value becomes the final median for each of the two datasets. After calculating the final median, for each location, we calculate an average of the final median of the tourist dataset and the final median of the resident dataset. Then the average of two medians become the divider for all values in the location. The final value of floating population is determined by dividing the raw value with the divider, multiplying it by 100 and dropping the decimal parts.

3.3.4   Calculating the Properties of Dots

In our visualization, the position of dots is determined randomly to prevent some dots being always hidden beneath other dots when there is high density of dots. The placement of dots is different for two methods. For the first method that draws the dots directly on the map, we assign a random radius within a threshold and a random degree to each dot. The second method requires the dots in a month to be placed inside the corresponding arcs. The calculation is similar with the first method, except that the degree is also within a range while the degree is entirely random in the first method.

The color of dots is one of the 24 color values that represent a single hour. As the colors are transitioning in a gradient, the colors of the nearby hours have similar colors. For example, the hours around 10PM to midnight have red colors. The hours between 1AM to 3AM have orange to yellow shades. The similarities between the neighboring hours and the disparities between the distant hours provide a cue for the users to distinguish between the areas that are populated at night and at daytime, for instance.

The colors are also preassigned along with the coordinates. As the dataset has a separate column for each hour, the column index is used as the index of the array containing the hex codes to retrieve the matching color code. The hex code is also added to the class name of the dots and is used to display or hide the dots when the user filters by month.

3.3.5   Implementing Mouseover for Arc Visualization

Because the arcs in the current visualization are often overlapping with each other, the data in some of the arcs are not visible if the visualization is static. To resolve this issue, we have implemented a mouseover feature that enables the user to view a larger size of arc and the dots inside of it when the cursor is on one of the arcs.

Each arc consists of the three elements, the bottom ellipse that gives the background color to the arc, the colored dots, and the borders that are drawn on top to make sure the dots do not appear to be out of place. The bottom ellipse is not interactable because it is at the bottom layer, covered by dots. We could have added the feature to the dots, but then it would be unresponsive if the user hovers on the empty area of arcs that only have partial values for some months. Also, triggering the hover function by the cursor on the arc boundary paths would be more inconvenient as the interactive area is too thin. Thus, we have added one emptier transparent ellipse with the same size as the bottom ellipse of the arcs, and implemented hover on the mouseover to the ellipse. This implementation has not caused noticeable difference in the loading time.

When the mouse cursor is on the transparent top layer of arc, a copy of the arc and the dots that are scaled by 3 is displayed next to the current cursor position. After the mouse cursor leaves the arc, the copied arc and the dots disappear. Initially we implemented it as transform, so that the arc that is being hovered upon moves to the sides and comes back after mouse cursor leaves. Nevertheless, the problem with this approach is that the original arc leaves its place during interaction, therefore possibly confusing users who is lost on which area he or she was exploring. As a result, the arcs and dots are now copied, not moved, when interacting with the cursor. Additionally, a highlight is shown at the outer boundary of the arcs that is being interacted.

3.3.5 Missing Values and Data Reduction

In the tourist and resident datasets, 10 months out of 12 months have the data records. In both methods of visualization we have used, the missing values are removed from the visualization by checking if the number of dots is less than 0 or higher than 10000, which is a threshold not going to be met. In the method plotting dots directly, the absence of values for some months is not quite noticeable because the dots for all months form a cluster. Because the missing months are the same

for all locations in a dataset, it does not cause false representations for a subset of locations which may be unlucky and do not have the data for the peak season—summer, for example.

In the second method, however, it may be helpful to fill in the missing values by prediction or based on the values of months nearby. Because each month is represented separately, it may be confusing as the area can be empty due to the missing value or the absence of people in the location. While it is possible to assume the correct meaning by observing the other areas, since often the locations that are popular tend to be popular over all months, suppling the values by prediction or implementing an additional measure that can clearly distinguish between the months without data and the months without data points would be a good addition to the current visualization.

Due to the large number of locations and their dots to be drawn, it was not possible to draw all dots in the browser at once. Therefore, in both methods, the visualization shows 1/100 out of total locations, by using variable for a counter and choosing every 100th location to be drawn only. Because the locations appear only once in the entire data, there is little possibility that this decision distorts the result by the certain locations being selected for drawing more often than others, although there exists the limitation with the certain locations not appearing on the map at all.

4. THE APPLICATION INTERFACE

4.1 Functionality of the Application
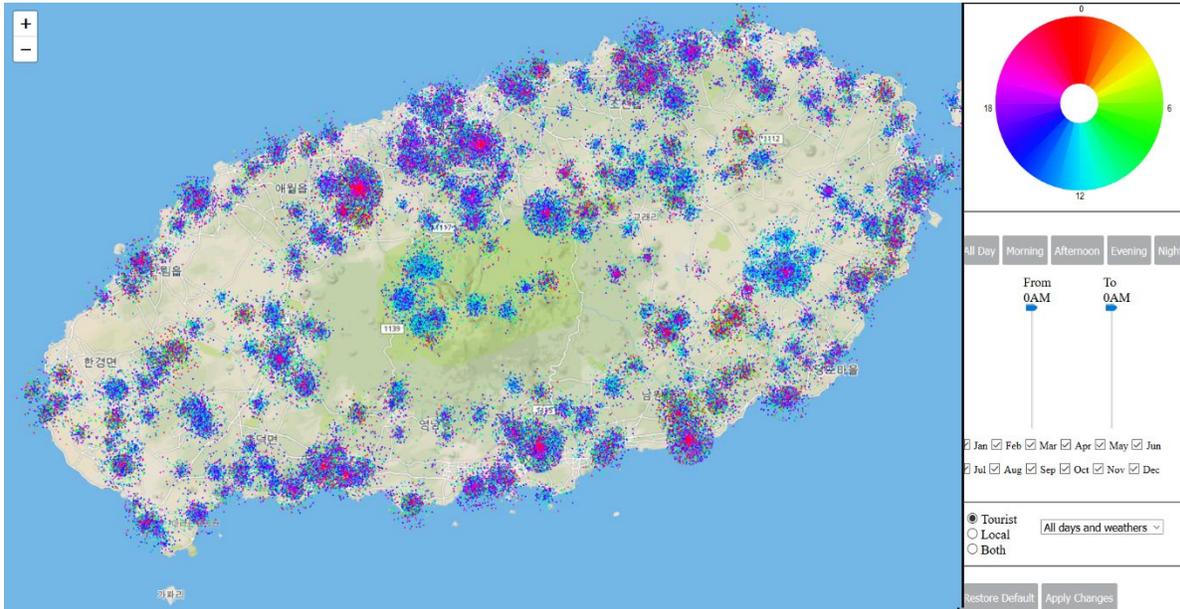
Figure   4.1 Interface of the application

The application visualizes the floating population at each location by hour and month. The user can use the application to make a comparison and explore the specific details by using the data filters and sliders. The figure shows the entire interface as shown in the Chrome browser.

The hour can be filtered by using the two sliders on the right, which represent the starting time and the end time of the hour range. It is also possible to click on the set of time ranges that are expected to be frequently used, such as night and morning. When the slider values are changed, the colored time legend on the top right is changed to highlight the selected hours. The rest of hours are greyed out so that the user can instantly understand which hour of data he or she is currently exploring.
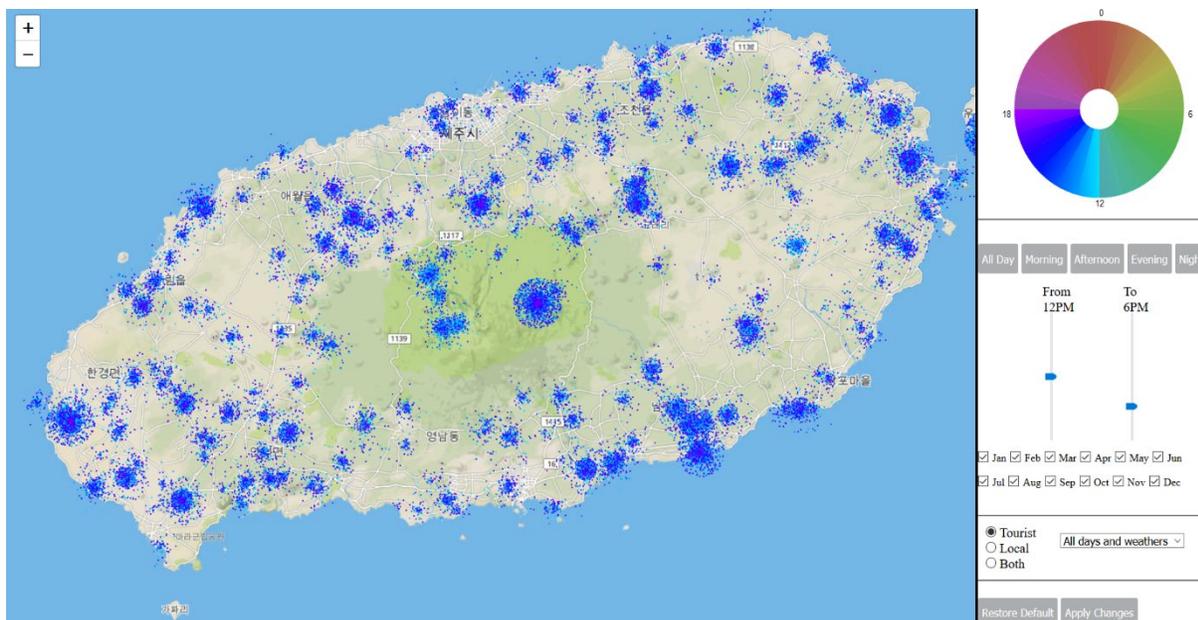
Figure   4.2 Visualization with the time filter used (showing the data from 12PM to 6PM)

The hours that are not in the selected hour range fades to grey as in the figure above, to emphasize which hours the user is looking at. Also, the map will only show the dots that belong to the selected hours.

For months, the user can select the multiple combinations of months to be displayed on the map by using the checkboxes. It is also possible to filter both months and hours at the same time. The user can explore the different data type by selecting on the drop-down list, such as a different weather or a different type of a day.

With the visualization method using arcs, it is possible to examine each arc in more detail by putting the mouse cursor on one of the arcs. Then the hovered arc is highlighted to remind the user the arc he or she was exploring, and the enlarged version of the arc is displayed next to the original arc. This functionality allows the user to read the data from the arcs that are hidden by other arcs in proximity.
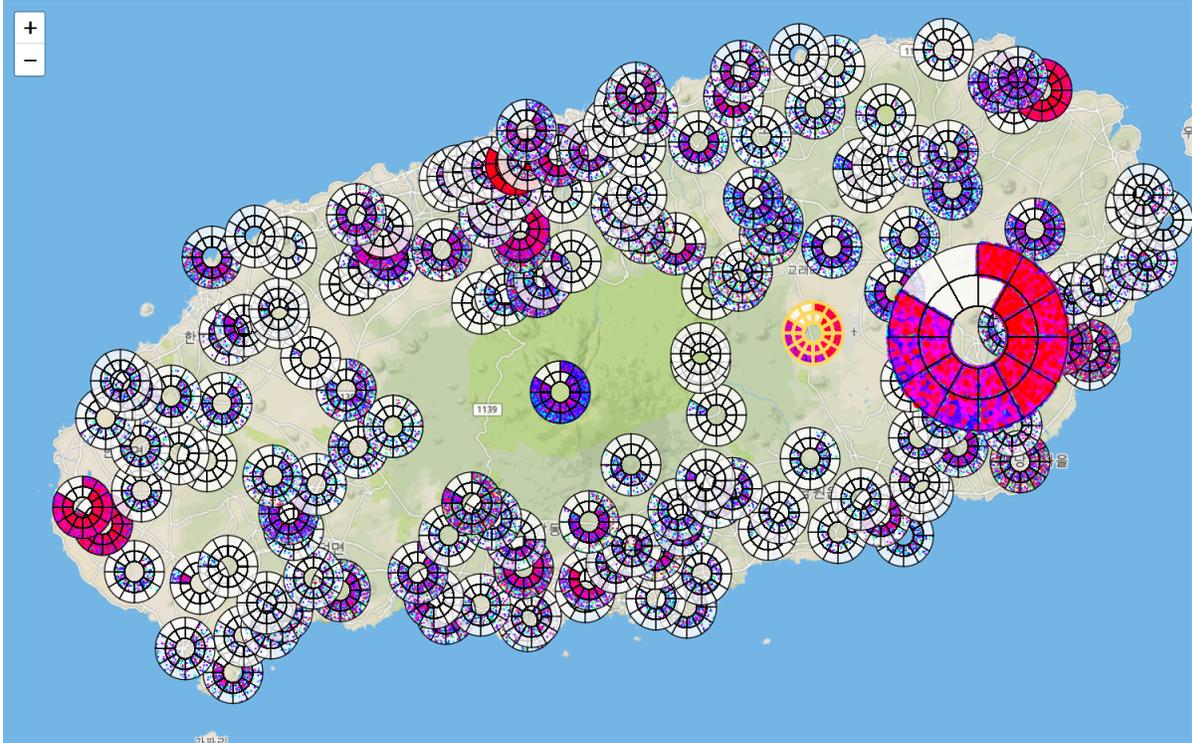
Figure   4.3 Expansion and highlight of the arc on hover

4.2 Analysis of Results

4.2.1 Observations of Floating Population

4.2.1.1. Representation by Colored Dots

Interacting with the visualization, we could observe a number of differences in the visiting places of tourists and residents. For example, the figure 4.4 highlights the areas that have the high floating population of residents. In comparison, the figure 4.5 highlights the areas where the floating population of tourists is high. The area inside the dotted white line is where both the tourist and resident floating populations are similar. From examining the two pictures, we can observe that the population of residents and tourists show differences based on where the location is (The highlights and the labels are not the parts of original visualization). For example, the western part of island is well visited by both tourists and residents; the eastern part is usually visited by the residents; the tourists visit coastal areas in the east side more often than the residents do. Based on this information, the city government can assign more resources for public transportation in the

eastern area, as both population visit there often. Furthermore, as many tourists visit the outer area of the island, the government can plan a bus tour for tourists that has platforms at the popular spots marked in the map.

From the figures, it is also visible that while most locations are the most active from the morning to the afternoon (6AM to 6PM), the locations in the costal areas tend to have more nighttime visitors, between 10PM to midnight. Conversely, for the mountain in the center of island in the figure 4.5, the majority of tourists appear in the morning to afternoon, which is understandable as most people leave a mountain before sunset.
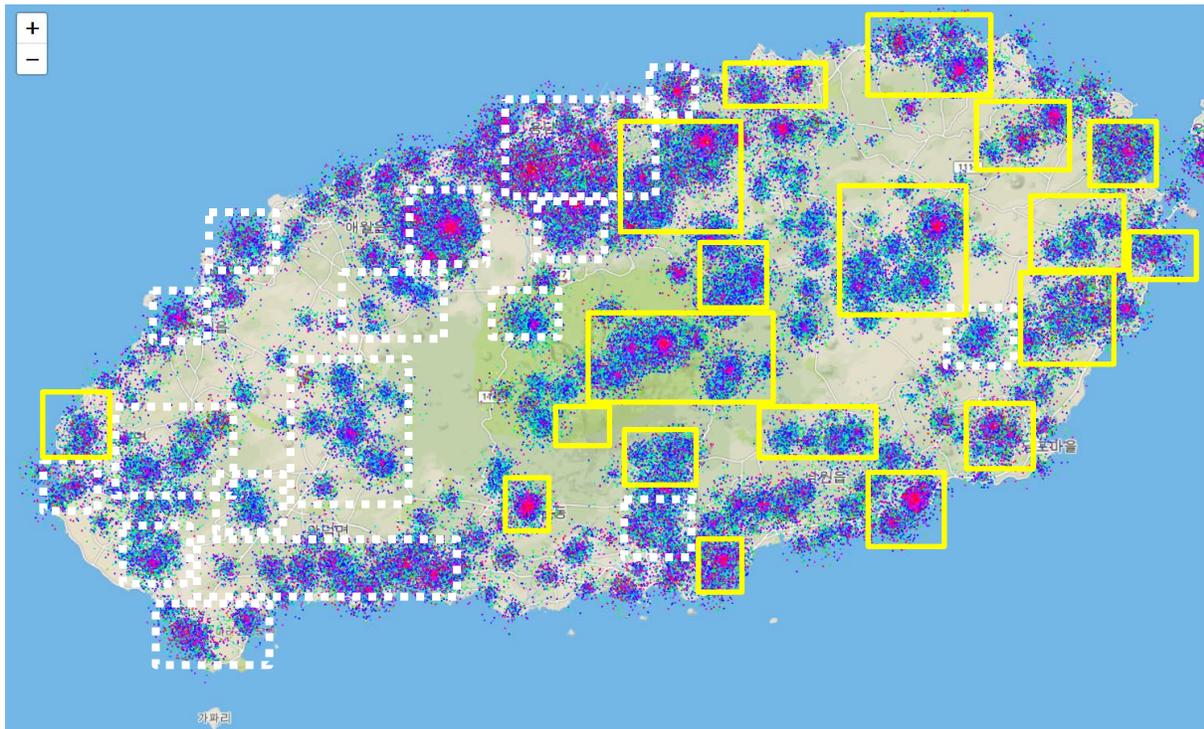

Figure 4.4   Floating population of residents in 2015
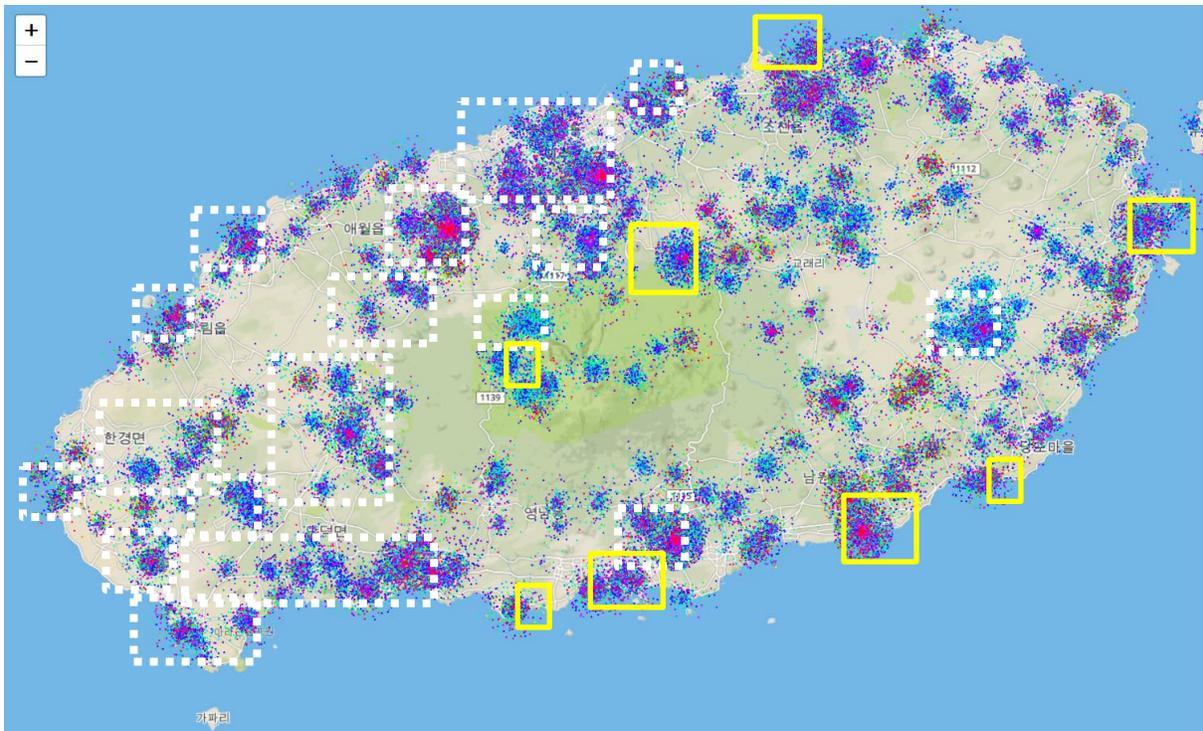
Figure 4.5   Floating population of tourists in 2015

Moreover, there are some observable differences in visiting patterns of tourists by the weather. Between sunny days (figure 4.6) and snowy days (figure 4.7), we can see that there are high concentrations of population on the two locations on the upper part of island on snowy days, unlike sunny days when the tourists are dispersed throughout the island. Between rainy days and sunny days, the locations the tourists visit do not differ much, but the population is less on rainy days, especially after midnight. Therefore, the businesses targeting the tourists as the street vendors or the taxi drivers can use this information to decide where to go in different weather conditions.
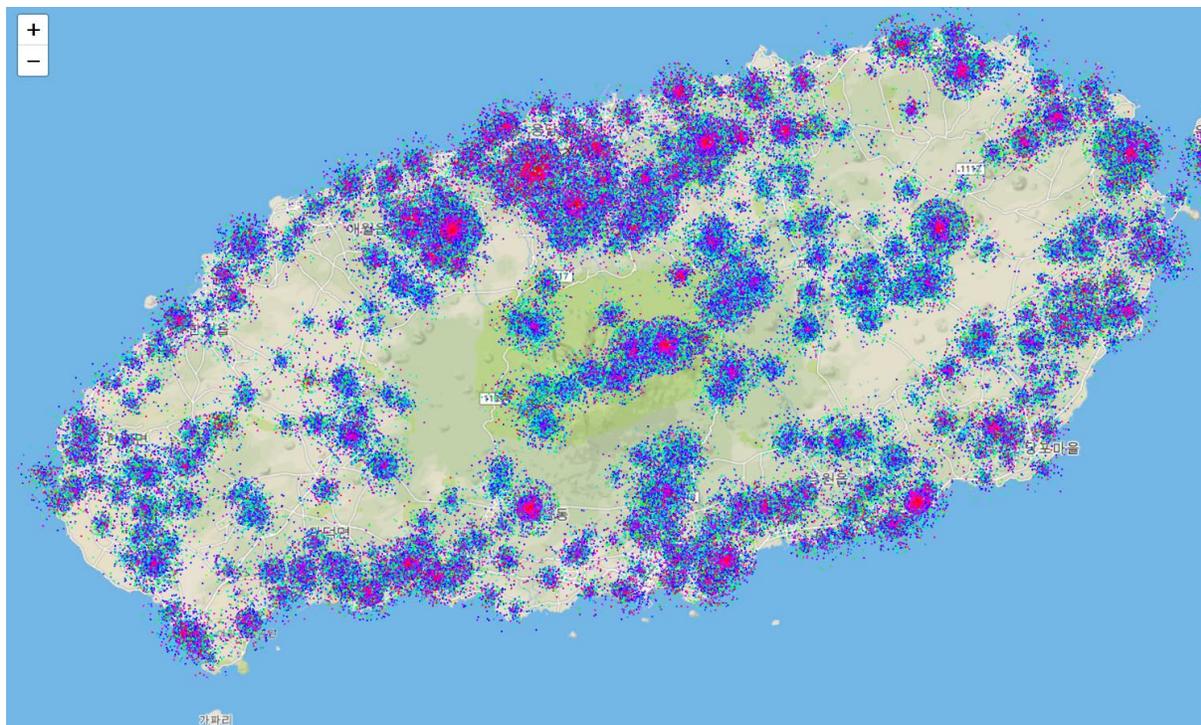
Figure 4.6   Floating population of tourists in sunny days
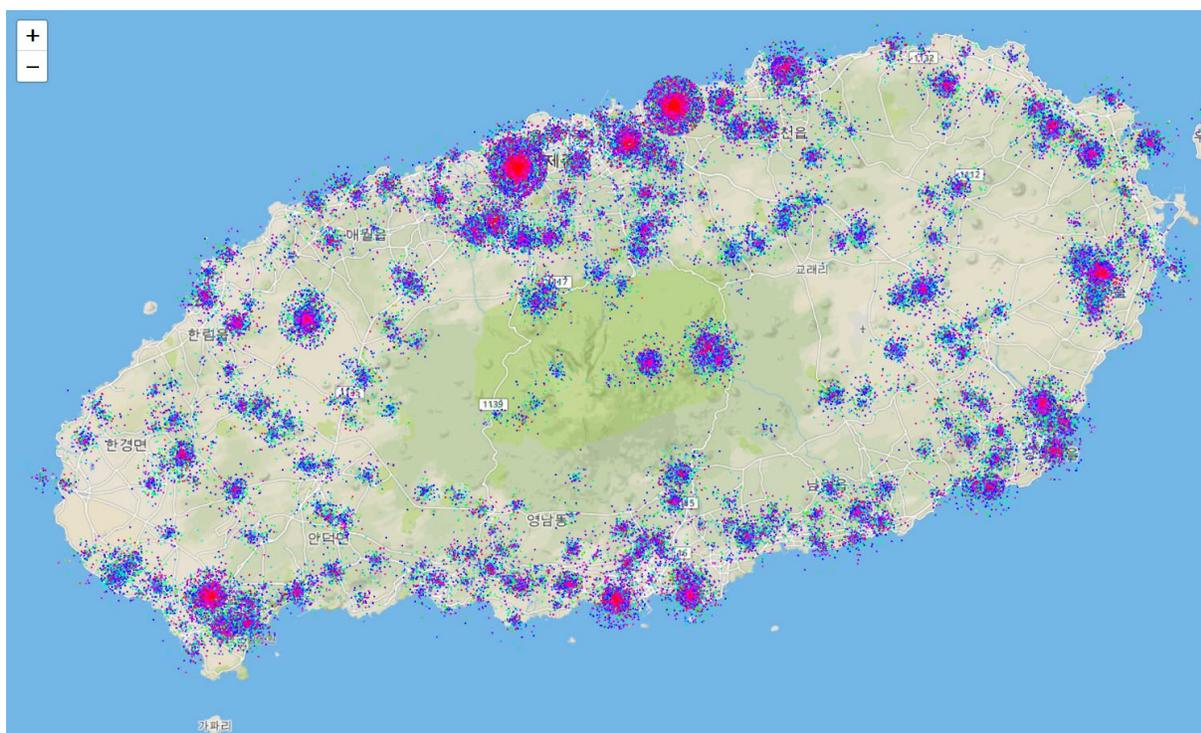
Figure 4.7   Floating population of tourists in snowy days

Moreover, by comparing the population in different times of day, we can observe the trend of floating population in different areas of Jeju island. The figure 4.8 displays the total floating population of tourists for six hours each, with annotations that represent the increase and decrease of the population in the annotated area. The area with red rectangle is where the population is low in the morning and high in the evening; the area with purple rectangle is where the population is high in the morning and low in the evening. This data provides a glimpse into the movements of tourists throughout the day, who are mostly in the middle area in the morning and then flocking at each side of island as time passes. When this data is combined with the data of previous years, or is compared with the different data types, such comparing a rainy day with a sunny day, it would be possible to construct the daily route of tourists that contain the possible sightseeing locations in the proximity they visit.
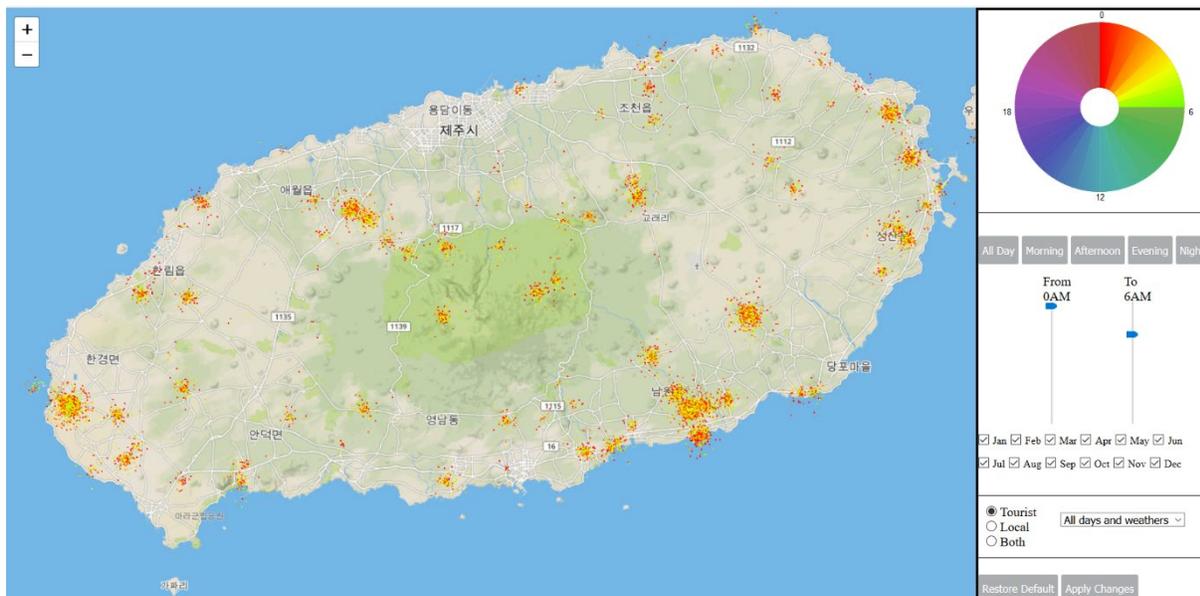

Figure 4.8   Visualization of tourist population at night (from midnight to 6AM)

The visualization of tourists in each time also provides useful information. From the figure above, we identify the areas that have high nighttime tourist population. The police, for instance, can use this information to plan the route for night patrol.

4.2.1.2 Representation by Colored Dots and Arcs

Using the second method that places the dots inside arcs, we could gather similar findings from the first method. We could see that the arcs in most places are colored in blue or red, which indicates they have the largest population between around 4PM to midnight. We could also identify areas that have the most population of tourists and residents in different hours, such as the area A and the area B. In the area A, which is a start and end point of a tracking route to the mountain top, the tourists appear in the later time than the residents; in the area B, a golf club, the tourists appear earlier than the residents. Only in October, many tourists visit the golf club in the morning.

The visualization of the different data, such as different weather types and day types produces the similar result as below, in terms of the distribution of population and the hour the dots represent. As an exception, on a cloudy day, especially in August, the tourists appear earlier than in other type of weathers in the a few areas including J, which is the top of the mountain.
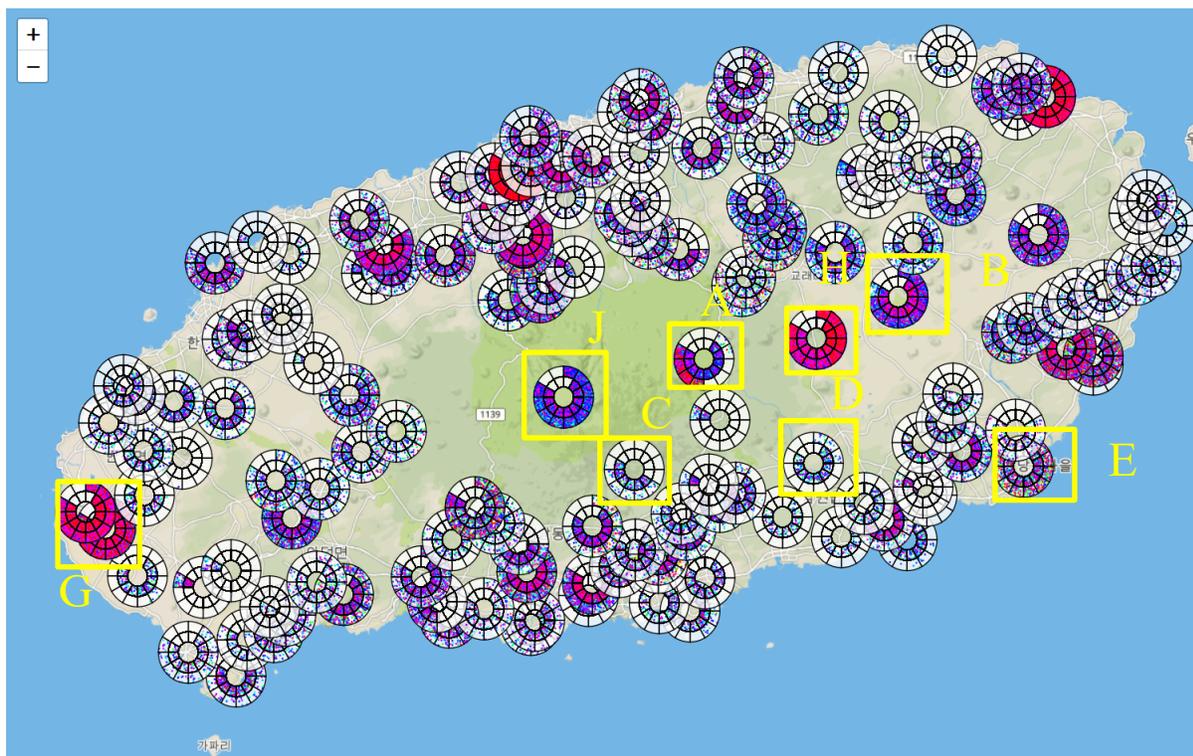

Figure 4.9   Visualization of total floating population of tourist and residents using arcs

While it is not very clear what are the causes of those certain concentration of population, the users can use this information in a useful way. For example, in the area A, as most tourists are present in the area in the late night, the taxi drivers may benefit from driving to the area around midnight to receive more tourist guests from July to October, rather than driving to the outer areas that do not have much population. Likewise, the restaurants or the stores nearby that seek to attract the tourists would better open the store until the late night.

Moreover, it is visible from the visualization that some locations have more tourists during winter and early spring, as the area C. On the other hand, some locations have more tourists between late spring to fall, as the area D. Though the highest population gathers after sunset in the most areas, there are exceptions such as the area E, which has its visitors dispersed throughout a day. As in the previous method, this information can provide more detailed insights on the visiting patterns of tourists for each month.

The visualized data also shows that the most locations have similar level of floating population during all hours except at night, as we can see from the nearly empty arcs in the visualization of night time data (Figure 4.10). But after filtering by afternoon (12PM to 6PM, figure 4.9) and night (midnight to 6AM, figure 4.10), we could identify several areas that are highly busy during night as well. In a few of them, such as the area H and the area G, there are more night time tourist population from January to May than in the later months. For both locations, the difference in the number of tourist population between months are more prudent during the night than during the afternoon. It is interesting because it is usually during the warm season that people stay outside until the late hour. But in these two locations, it is the opposite.

As we learned that the tourists in Jeju island in each season are probably distinct sets of groups using the previous method, we can gain similar understanding through observing each arc and discovering the discrepancies as in the area H and the area G. The capability to display the data of each month separately in one arc makes it easy to detect if there is a location or a month that shows a different pattern from others.

Figure 4.10   Visualization of floating population in the afternoon using arcs
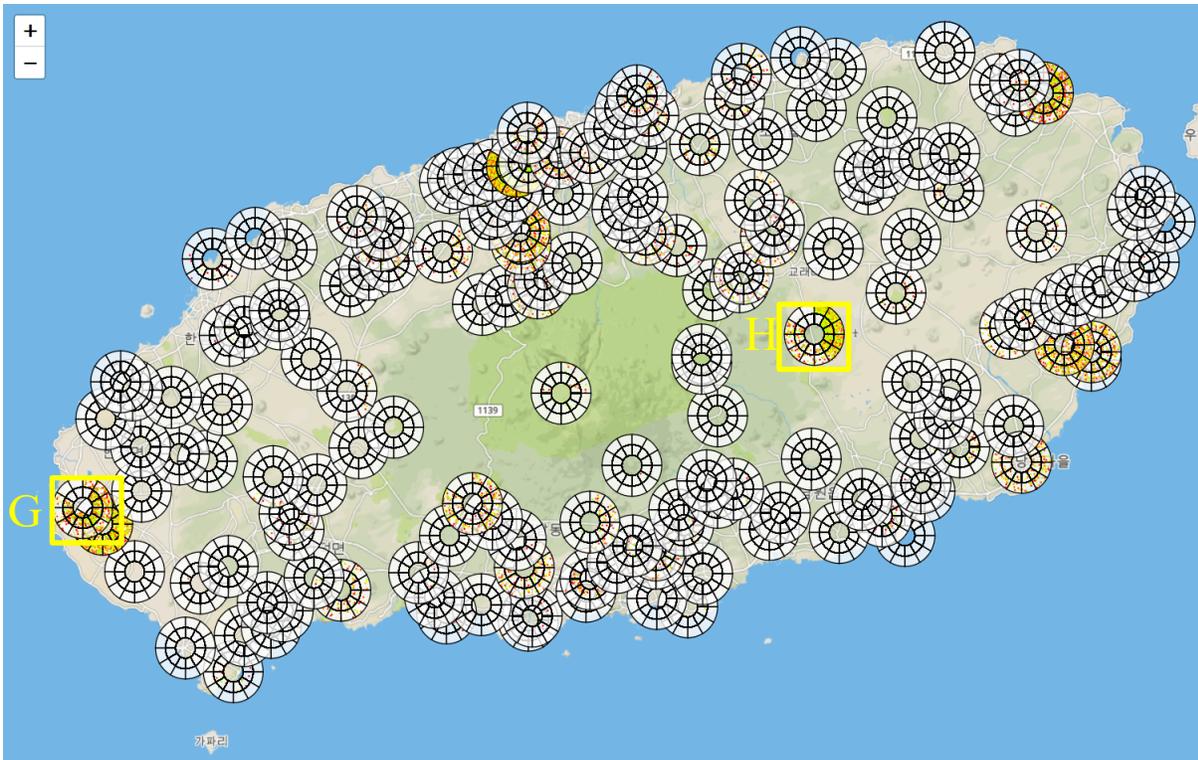
Figure 4.11   Visualization of floating population in the night using arcs

In addition, another location that shows a noticeable difference between months is area F in the figure 4.9, which is a recreational forest. In the area F, the arc for October is the densest arc. Considering the season, it may signify that the forest is worth visiting during autumn when the leaves change color, as more people visit the location during that season. On the other hand, the travelers who would like to relax in a quiet forest can utilize this information to avoid October and choose a different month for a visit.

Using the two methods for visualization, we could gain some insights on the floating population in Jeju island. Exploring different datasets and comparing the results, we have demonstrated that the visualization methods we have implemented are useful in identifying the topics to explore further, such as the difference of tourists visit pattern by each season and the locations that have many tourists in certain month and few tourists in other months, which would be useful supplements in analyzing the tourist behaviors in the Jeju island.

4.2.2 Comparison Between Two Techniques

In this study, we have implemented two methods for representing the floating populations. In the first method, we draw the dots directly on the map. The dots form a circle-shaped cluster for each location. The cluster has the highest density in the center and become sparser as the radius increases. The advantage of this method is that it is easy to compare the size of floating populations in different locations. The more population there is, the denser and the bigger the center of cluster is. Also, because there are no other juxtaposed elements on the map, each location is more clearly visible compared to the second method.

Nevertheless, this method does not support the comparison between the different months or between the tourist and the local resident floating population in a single location, though it is still possible to filter by months or population types to make a comparison between locations. It is also not easy to distinguish the data from two or more locations if they are too close. The problem is not as severe with the locations with high floating population, because then they have dense centers which clearly marks the origin. On the other hand, when the locations are not highly populated, such as when the location have less than 10 dots total, it becomes difficult to distinguish whether a dot is belonging to one location or the other, as it is not clear where is the center for each location.

Shrinking the area that a dot of a single location can occupy, to provide more space between the locations, is one solution. But it causes another problem since the dots in the highly populated

area would become too cluttered to observe the distribution of dots over the hours of a day. Thus, determining the range of each arc can be challenging with this method if the data contain locations with the different level of popularity. But overall, it is effective when the user wants to compare multiple locations with at least moderate popularity, while maintaining the geographical context.

In the second method, we first draw a semi-transparent circle consist of arcs in each location. The circle is divided to 12 arcs that represents a month in a year. The radius is also divided in half to display the tourist population in the outer area of the circle and the local population in the inner area of the circle. The advantage of this method is that the user can compare between the tourist and resident population in different months at the same time without having to switch between them, as they are displayed in one circle together. Moreover, because the data of each month is displayed separately, any changes or diversion from the other moths in the same arc or other arcs can be noticed instantly, without having to try several combinations of the month filters before detecting the interesting diversion.

But the limitation of this method is that the circles and the dots inside are blocking the view of the original map. The number of locations that can be displayed at once is also smaller than the first method, as we have to prevent the circles from overlapping with each other much, to ensure the dots in all circles are visible. The overlapping arcs can be avoided completely by implementing a more sophisticated algorithm, for instance, by setting an interval between any two arcs so that they do not touch each other, but then the total number of displayable locations would become even smaller.

The number of locations is also restricted by the amount of computation required. Compared to the first method, because the second method also draws the arcs which consist of the path element and an ellipse, the computation needed to display the data of a single location is larger with the second method. For instance, the number of locations displayed in the figure 4.5 total, which used the first method, is 977. But with the second method in the figure 4.9, the number of locations is 164.

Another issue is that it becomes difficult to compare the popularity of locations as the number of dots grow. When the arc is fully filled with dots, distinguishing the comparative popularity between a similarly popular area is not possible, because the dots in one arc simply overlap the other dots. Thus, the highly popular place, which has five overlapping layers of dots looks the same as the moderately popular place, which does not have any overlapping dots at all. Therefore,

this method is more suitable if the user would like to explore the details of a single location at a time.

5. CONCLUSION

This paper proposes the two methods for visualizing the geo-temporal data for the floating population of tourists and residents in Jeju island in the Republic of Korea. In the first method, we draw the colored dots representing the number of population directly on the 2D map, so the data can be easily compared by examining the size of each cluster. In order to avoid the dots from the highly populated area becoming cluttered and making it difficult to see each dot inside the cluster, the cluster is drawn to be denser in the center and the sparser near the boundary. Because there is no additional marker as the arc in the second method, the geographical data is also largely preserved.

The second method draws a semi-transparent circle in each location of the same map, which are divided into 24 areas that represent 12 months in a year for the tourist and the resident. The tourist and resident data are distinguished by placing the tourist data in the outer area and the resident data in the inner area. According to the number of floating populations in each month, different number of dots are placed inside each corresponding area. As the arc can display both population types at once, it is possible to compare the population of tourist and resident in one look.

In both methods, the color of dots represents the time of a day, and the number of dots represent the value of floating population. The information on which color represents which hour is provided from the circular time legend next to the map. Using the filters in the interface, it is possible to view the data by the type of weather, such as sunny, rainy, snowy, and cloudy and the type of day such as a holiday, a day before holiday, and a regular day. It is also possible to filter to show only certain hours of a day.

Interacting with the visualization, the user can gain insights on the popular spots of tourists and residents, as well as their movement as a whole in different times of a day. The visualization will supply useful information for the employees in tourism industry, the local government hoping to attract more tourists, and the people who plan to travel in the area, by leading them to discoveries which may not have been reached by reading the raw data, through the combination of geographical and temporal data.

From the visualization, we could observe some interesting patterns in the tourist population in Jeju island. For instance, the primary means of transportation of tourists arriving in the Jeju island

seems to be different for summer and spring. There are more tourists in summer who are using flights, while there are more tourists using ships in spring. We could also observe that a famous tourist spot that is highly popular during summer loses its popularity in spring, which is another sign that the two groups of tourists have distinct interests and travel styles.

The floating population of tourists and residents do not overlap much, but there are a few spots that are popular for both populations. While the popularity of some areas can be explained by gathering more information about the nearby facilities, sometimes the map alone does not provide a reason for the high floating population in certain areas. Therefore, the visualization can point to the regions that are not usually expected to be highly populated.

Between the two methods we have implemented, while the second method using the arcs has an advantage in comparison between the tourist population and the resident population at the same time, the first method is more effective in conveying the information with the smaller loss in the geographical data. Because the first method mostly preserves the view of the map, and the larger number of locations can be displayed at once, it has led to more discoveries while interacting with the data.

This work contributes to the studies of geo-temporal visualization by proposing two new techniques for comparing the population distribution in different times. The two methods integrate the temporal data inside map so that the user can graph the data from both dimensions efficiently. It builds on the previous methods such as the time circle for representing the time in colors to create effective method and demonstrates that the idea can be also utilized to express the size of population, in addition to the trajectories.

5.1 Future Works

As a future work, we would like to add more functionalities and details to the application. For instance, we can add some interactive markers for the famous tourist spots, to supplement the location information in the base map which are not very thorough, and to make it easy for the users to recognize the familiar locations. Also, as in the figure 4.8 that marked the locations that had population grow over time and decrease over time, displaying the information about the changes of population in the area in a day would be useful as it can clearly show the movement of tourists as crowds, though it does not show the route of individual tourists. Additionally, we would like to gather more data over the recent years. The data source only provided the data for 2014 and 2015,

and because the data for 2014 used a different coordinate system stored in image files, we could not extract and convert the location data in time. With data for multiple years, we will be able to visualize how the popularity of each area has changed over the years, which can hint the dynamics and economic state of the neighborhoods in Jeju island.

Moreover, we would like to modify the code to display all data points on the map. In the study, only a portion of data could be used for visualization because the browser application could not draw the dots if the number of dots was too high. One solution could be drawing the different set of dots based on the zoom level. For instance, in the initial run, the maximum number of dots that can be drawn at once is displayed. Then when the user zooms to a point on a map, the dots in the neighboring areas are drawn. Although it still would be not possible to draw all data points at once, the entire data can be utilized without loss.

Furthermore, the second method can also be improved by selecting the arcs to be drawn based on the distance between them, rather than based on the order of appearance in the file as it is now. Then it becomes possible to avoid overlapping of arcs, which would create a clearer and cleaner view of the data on the map.

Finally, we hope to implement a more efficient algorithm for drawing and filtering the arcs and dots to reduce the loading time while interacting with the visualization. In fact, this is the most severe issue we are facing now. Even considering the size of files, the speed of visualization is too slow to be used in real life. The first dots-only method takes 30 seconds on load, with additional loading time in between when the user interacts with the filter. The second method with arcs takes more than 3 minutes on the laptop with 12GB RAM, which certainly calls for refinement of codes.

REFERENCES

Andrienko, N., Andrienko, G. Gatalsky, P. (2003). Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages & Computing*, 14(6), pp. 503-541.

Andrienko, N., Andrienko, G. (2004). Interactive visual tools to explore spatio-temporal Variation. *Advanced Visual Interfaces*, pp. 417-420.

Andrienko, N., Andrienko, G. (2008). Spatio-temporal aggregation for visual analysis of movements. *IEEE Symposium on Visual Analytics Science and Technology*, pp. 51-58.

Aoyama, D.A., Hsiao, J.T., Cárdenas, A.F., Pon, R.K. (2007). TimeLine and visualization of multiple-data sets and the visualization querying challenge. *Journal of Visual Languages & Computing,* 18(1), pp. 1-21.

Baskaran, S., Fang, S., Jiang, S. (2017). Spatiotemporal visualization of traffic paths using color space time curve. *IEEE International Conference on Big Data (BIGDATA),* pp. 3398-3405.

Chua, A., & Servillo, L., Marcheggiani, E., Moere, A.V. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management*, 57, pp. 295-310.

Drocourt, Y., Borgo, R., Scharrer, K., Murray, T., Bevan, S.I., Chen, M. (2011). Temporal visualization of boundary-based geo-information using radial projection. *Computer Graphics Forum,* 30(3), pp. 981-990.

Dübel, S., Röhlig, M., Schumann, H., Trapp, M. (2014). 2D and 3D presentation of spatial data: A systematic review. *IEEE VIS International Workshop on 3DVis*, pp. 11-18.

Hewagamage, K.P., Hirakawa, M., Ichikawa, T. (1999). Interactive visualization of spatiotemporal patterns using spirals on a geographical map. In Proc. *IEEE Symposium on Visual Languages*, pp. 296-303.

Girardin, F., Fiore, F.D., Blat, J. (2007). Understanding of tourist dynamics from explicitly disclosed location information. In Proc. *International Symposium on LBS & TeleCartography*.

Guo, D., Chen, J., MacEachren, A.M., Liao, K. (2006). A visualization system for space-time and multivariate patterns (VIS-STAMP). *IEEE Transactions on Visualization and Computer Graphics*, 12(6) pp. 1461-1474.

Jänicke, S., Heine, C., Stockmann, R., Scheuermann G. (2012). Comparative visualization of geospatial-temporal data. In Proc. *International Conference on Computer Graphics Theory and Applications*. pp.613-625

Kapler, T., Wright, W. (2004). GeoTime information visualization. *IEEE Symposium on Information Visualization*, pp. 25-32.

Kraak, M.J. (2003). The space-time cube revisited from a geovisualization perspective. In Proc. *21st International Cartographic Conference*, pp. 1988-1996

Kwan, M. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: A methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, 8(1–6), pp. 185-203.

Shanbhag, P., Rheingans, P., desJardins, M. (2005). Temporal visualization of planning polygons for efficient partitioning of geo-spatial data. *IEEE Symposium on Information Visualization,* pp. 211-218.

Thakur, S., Hanson, A. J. (2010). A 3D visualization of multiple time series on maps. *14th International Conference Information Visualisation,* pp. 336-343.

Tominski, C., Schulze-Wollgast, P., Schumann, H. (2005). 3D information visualization for time dependent data on maps. *Ninth International Conference on Information Visualisation,* pp. 175-181.

Wang, M., Qin, L., Hu, Q. (2012). Data mining and visualization research of check-in data. *20th International Conference on Geoinformatics*, pp. 1-4.

Wu, C.L., Carson, D. (2008). Spatial and temporal tourist dispersal analysis in multiple destination travel. *Journal of Travel Research*, 46(3), pp. 311-317.

Zhao, J., Forer, P., Harvey, A. S. (2008). Activities, ringmaps and geovisualization of large human movement fields. *Information Visualization*, 7(3–4), pp. 198–209.

Zhao, J., Forer, P., Sun, Q., Simmons, D. (2013). Multiple-view strategies for enhanced understanding of dynamic tourist activity through geovisualization at regional and national scales, *Cartography and Geographic Information Science*, 40(4), pp. 349-360.

Zhong C., Wang T., Zeng W., Müller Arisona S. (2012). Spatiotemporal visualisation: A survey and outlook. *Communications in Computer and Information Science*, 242, pp 299-317.