FIT5147 Data Visualization S1

Multifactorial Influences Including Temperature on Influenza Dynamics in Australia

Assignment DVP

Student Name: Fangyu Sun

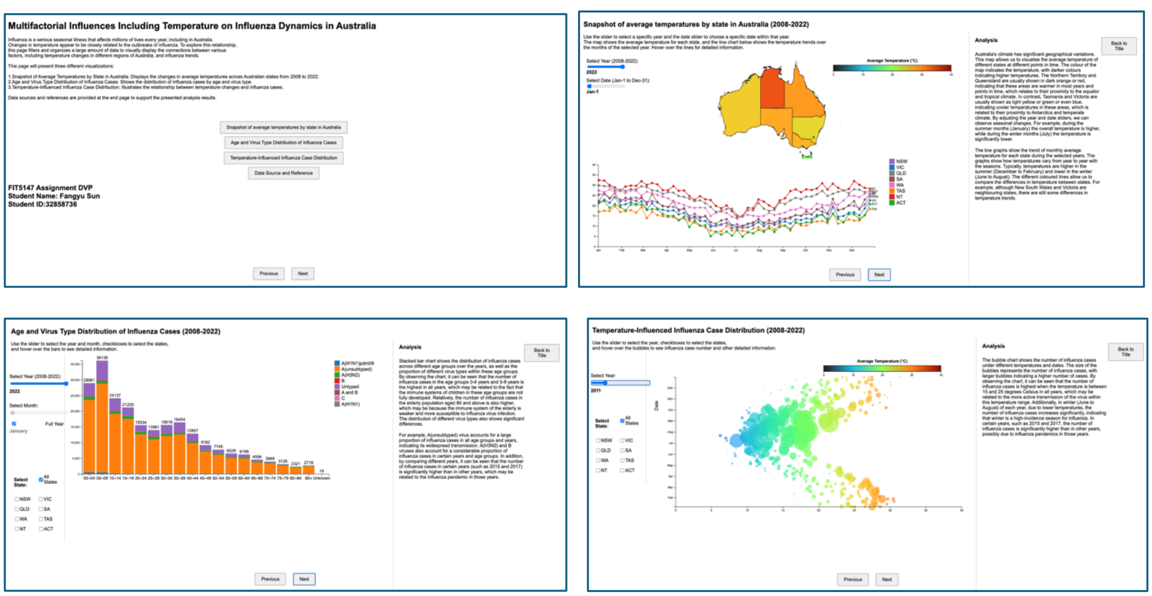
Student ID: 32858736

1. Introduction

The relationship between temperature and the spread of influenza has been a significant research topic in public health. This DVP project aims to use interactive narrative visualization to demonstrate the impact of temperature changes on influenza dynamics in Australia. The study covers data from 2008 to 2022, including average temperatures and influenza cases in various states, to reveal how temperature variations influence the outbreak and spread of influenza.

The target audience for this project includes public health experts, policymakers, and the general public interested in influenza and climate change. Public health experts can use the data and visualization results to develop more effective prevention measures, policymakers can formulate relevant policies based on the findings, and the public can enhance their understanding of influenza and climate change through this information, thus taking necessary preventive measures.

The data for this project comes from the Australian Bureau of Meteorology and the National Notifiable Diseases Surveillance System. By merging, categorizing, and cleaning the data, the accuracy and reliability of the research results are ensured. The study reveals that temperature changes have a significant impact on influenza outbreaks, with a clear correlation between the number of influenza cases and temperature fluctuations in different regions and periods.



(Overview of the Whole Project)

2. Design Process

2.1 Graph Choice

Based on the characteristics of the temperature dataset and the influenza dataset, the DVP adopts heat maps, line charts, stacked bar charts, and bubble charts to present data visualizations.

2.1.1 Heat Map and Line Chart

Temperature and influenza case numbers are continuous data; hence, heat maps can effectively display temperature changes and influenza case distributions across different states over different time periods. Through color variations, users can intuitively see the highs and lows of temperature and influenza cases. Since time is a continuous variable, line charts can effectively show temperature trends over time while distinguishing temperature changes between different states.

2.1.2 Stacked Bar Chart

Stacked bar charts are used to display the distribution of virus types and age groups. Both virus types and age groups are categorical data, and stacked bar charts can intuitively show the number distribution of different virus types over various time periods and among different age groups. Stacked bar charts not only help users understand the spread of different virus types during specific time periods but also show the changes in virus numbers among different age groups. Through this type of chart, users can clearly see which virus types are more prevalent in certain age groups or time periods, aiding in a deeper understanding of influenza transmission patterns and high-risk population characteristics.

2.1.3 Bubble Chart

Bubble charts are used to display the complex relationships between temperature, time, and influenza case numbers. In bubble charts, the x-axis represents temperature, the y-axis represents time, and the size of the bubbles represents the number of influenza cases. Different colored bubbles can be used to distinguish between different regions or virus types, further enhancing the information content and readability of the chart. Bubble charts allow users to intuitively see the impact of temperature changes on influenza outbreaks. For example, larger bubbles may indicate a significant increase in influenza cases within a specific temperature range and time period. Bubble charts can reveal overall trends and help users discover anomalies and potential patterns, such as peak periods of influenza transmission under certain temperature conditions. This visualization method provides users with a convenient way to analyze and understand the dynamic relationships between temperature, time, and influenza case numbers.

2.2 Colour Choices

In this DVP project, different color hues are primarily used to help quickly distinguish between different categories and data features. In the heat map and bubble chart, the variation in color brightness represents temperature levels, with red and yellow colors indicating higher temperatures, and blue and green colors indicating lower temperatures. In the line chart, different colors are used to distinguish temperature changes in different states, ensuring that users can easily identify which state each line represents. In the stacked bar chart, contrasting colors are chosen to differentiate between various virus types, making the chart clearer and easier to read.

2.3 Typographic

All text elements in the project use a uniform font to maintain consistency and readability, ensuring visual coherence throughout the page. The titles are relatively large to facilitate user navigation to different sections. Some images occupy the full width of the page, while others take up half the page width for easy side-by-side comparisons. All images and text are centrally aligned, keeping the page clean and organized, making it easier for users to focus on the content.

2.4 Integration of Design Sheets

All of the above charts are combined from five design sheets (see Appendix). Each chart displays at least one aspect of the design process for this data visualization project. From the initial concept to the final implementation, every step was carefully considered and optimized.

2.4.1 Design Sheet 1

In the preliminary design stage, various possible visualization techniques were considered, including heat maps, line charts, stacked bar charts, pie charts, and Sankey diagrams. Interactive methods such as sliders, checkboxes, and mouse hovers were also considered. These charts and interactions can effectively display the relationship between temperature or other factors and influenza cases.

2.4.2 Design Sheet 2

In further refining the design, the focus was on displaying temperature and influenza data, mainly using heat maps and line charts. The heat map displays the temperature distribution across different states and time periods using color intensity, with lighter colors indicating lower temperatures and darker colors indicating higher temperatures. This allows users to intuitively understand temperature changes in each state. The line chart shows the temperature trend over time, with different colored lines distinguishing the temperature changes in different states, ensuring users can easily identify the data represented by each line. The combination of these two charts helps users fully understand the relationship between temperature changes and influenza spread, improving data readability and user experience.

2.4.3 Design Sheet 3

During the design optimization process, pie charts were initially considered to display the distribution of different virus types across time periods. However, further analysis revealed that pie charts have limitations in displaying complex data relationships, especially when comparing multiple categories and time periods. Therefore, it was decided to use bubble charts instead. The bubble chart uses the x-axis to represent temperature, the y-axis to represent time, and the bubble size to represent the number of influenza cases, with different colors representing different virus types. This design not only increases the information content and readability of the chart but also allows users to more intuitively understand the impact of temperature changes on influenza spread. Ultimately, the bubble chart was placed in the last chart of the design, serving as a summary and display of the overall temperature data and influenza case data. This adjustment not only improved the data visualization effect but also enhanced the user's interactive experience. Additionally, the bubble chart added a tooltip feature, allowing users to see detailed case numbers when hovering over a bubble after filtering by year and state.

2.4.4 Design Sheet 4

Combining elements from the previous design sheets, a more comprehensive design was created, integrating different charts to ensure users can navigate data through simple interactions. The stacked bar chart was further optimized and applied in this stage. The stacked bar chart displays the distribution of different virus types across age groups and time periods, using different colored bars to distinguish virus types, showing not only the total but also the composition of each bar. Users can use interactive features to select specific time periods or age groups and dynamically view the distribution of different virus types. The stacked bar chart design also considered user experience by adding a tooltip function, allowing users to see detailed information, including the specific virus type, case number, and age group, when hovering over a bar. This design not only improves the chart's readability but also increases data interactivity, enabling users to explore and understand the complex relationships in influenza data more deeply.

2.4.5 Design Sheet 5

The final design solution combines the strengths of the previous design sheets, forming a complete interactive narrative visualization. By integrating visualization techniques such as heat maps, line charts, stacked bar charts, and bubble charts, it comprehensively displays the complex relationship between temperature changes and influenza data. Initially, the plan was to display all charts on one page, but in practice, this made the page appear too cluttered, reducing readability and user experience. Therefore, it was ultimately decided to use a slideshow format for the display, with each chart placed on a separate slide, allowing users to switch between slides using navigation buttons and interactive features. This adjustment not only improved the data visualization effect but also enhanced the user's interactive experience, enabling users to explore and understand the dynamic relationships between temperature, time, and influenza cases more deeply.

3. Implementation

3.1 Technical Implementation

In this project, JavaScript and the D3.js library were used to achieve data visualization. The project mainly consists of HTML, CSS, and JavaScript files. The HTML file is responsible for the basic structure and layout of the page, the CSS file is used for style design, and the JavaScript file handles data processing, generates visualizations, and implements interactive functions.

The data was downloaded from the following sources and converted into a format suitable for D3.js processing. Average temperature data was obtained from the Bureau of Meteorology, covering information from 112 stations across Australia, recorded from 1910 to 2022. After cleaning the data, it was categorized based on the states where the stations were located, and the time range was filtered to 2008 to 2022. Influenza case data was sourced from the National Notifiable Diseases Surveillance System, with weekly data released every Friday, covering influenza cases in various states from 2008 to 2022. Each row records information such as date, state, temperature, age group, gender, and virus type. Both files were provided in CSV format, and initial steps included cleaning and standardizing the data.

First, missing values and incomplete records were removed. Then, the date format of both files was standardized to "d/m/yyyy". Next, state names were unified (e.g., VIC, NSW) to ensure consistency between the two datasets. The processed CSV data was converted into JSON format to facilitate reading and processing by D3.js. Specific steps included using JavaScript's built-in date parsing functions to convert date strings into Date objects and abbreviating state names to standard formats. Finally, the processed data was structurally stored in JSON files, providing a foundation for subsequent visualization operations.

```javascript
function loadAndProcessTemperatureData(callback) {
  d3.csv("combined_all_states.csv").then(function (data) {
    data.forEach(d => {
      d.date = d3.timeParse("%d/%m/%Y")(d.date);
      d.state = d.State;
      d.temperature = +d['mean temperature (degC)'];
      if (d.date === null) {
        //console.error("Date parsing error for row:", d);
      }
    });
```

```javascript
function loadAndProcessFluData(callback) {
  d3.csv("combined_flu_data.csv").then(function (data) {
    data.forEach(d => {
      d.date = d3.timeParse("%d/%m/%Y")(d.Date);
      d.state = d.State;
      if (d.date === null) {
        // console.error("Date parsing error for row:", d);
      }
    });
```

3.2 Interactive Narrative Visualisation Implementation

In this project, multiple interactive data visualization components were implemented to allow users to intuitively understand the complex relationships in the data. The implementation process for each chart is detailed below:

3.2.1 Heat Map and Line Chart

The implementation of the heat map involves the following steps. First, load the GeoJSON data containing the boundaries of Australian states and merge it with the temperature data. Next, use D3.js to create an SVG element and draw paths for each state. The fill color of the paths is based on the average temperature of the state on a specific date. Users can select different dates using a date slider, and the colors on the map will dynamically update to reflect the new temperature data.

The implementation of the line chart includes the following steps. First, load and parse the temperature data and filter it by year. Use D3.js to create an SVG element and draw a line for each state, representing the temperature trend for that state in a specific year. Users can select different years using a year slider, and the line chart will dynamically update to reflect the new temperature data.

```javascript
// Function to draw the map on slide 2
function drawMap(data) {
  const width = document.getElementById('map').clientWidth;
  const height = document.getElementById('map').clientHeight;

  d3.select("#map").selectAll("svg").remove();
  const svg = d3.select("#map").append("svg") ⋯
  const projection = d3.geoMercator() ⋯
  const path = d3.geoPath().projection(projection);
  const color = d3.scaleSequential([0, 40], d3.interpolateTurbo);

  // Create tooltip
  const tooltip = d3.select("body").append("div") ⋯
  d3.json("states.geojson").then(function (mapData) { ⋯
  }).catch(function (error) {
    //console.error(error);
  });
}
```

```javascript
function drawLineChart(data, year) {
  const container = document.getElementById('line-chart');
  container.innerHTML = '';  // Clear existing chart
  const width = container.clientWidth;
  const height = container.clientHeight;
  const svg = d3.select("#line-chart").append("svg") ⋯
  const margin = { top: 20, right: 300, bottom: 150, left: 50 }, ⋯
  const chart = svg.append("g") ⋯
  const x = d3.scaleTime() ⋯
  const y = d3.scaleLinear() ⋯
  const line = d3.line() ⋯
```

3.2.2 Stacked Bar Chart

The implementation of the stacked bar chart includes the following steps. First, load and parse the influenza case data and filter it by year and month. Use D3.js to create an SVG element and draw stacked bar charts for each age group, with different colored bars representing different virus types. Users can select different time periods using the year and month sliders, and the chart will dynamically update to reflect the new data.

```javascript
function drawStackedBarChart(data, virusTypes) {
  const container = document.getElementById('stacked-bar-chart');
  container.innerHTML = '';
  const width = container.clientWidth;
  const height = container.clientHeight * 0.8;
  const svg = d3.select("#stacked-bar-chart").append("svg") …
  const margin = { top: 20, right: 200, bottom: 40, left: 40 }, …
  const chart = svg.append("g") …
  const ageGroups = [ …
  ];
  const nestedData = d3.groups(data, d => d.ageGroup).map(([key, values]) => { …
  }).filter(d => ageGroups.includes(d.ageGroup));
  nestedData.sort((a, b) => ageGroups.indexOf(a.ageGroup) - ageGroups.indexOf(b.ageGroup));
  const stack = d3.stack() …
  const layers = stack(nestedData);
  const x = d3.scaleBand() …
  const y = d3.scaleLinear() …
  const layerGroups = chart.selectAll(".layer") …
  const bars = layerGroups.selectAll("rect") …
  nestedData.forEach(d => { …
  });
  chart.append("g") …
  chart.append("g") …
  const legend = chart.selectAll(".legend") …
  legend.append("rect") …
  legend.append("text") …
}
```

3.2.3 Bubble Chart

The implementation of the bubble chart includes the following steps. First, load and merge the temperature and influenza case data. Use D3.js to create an SVG element and draw bubbles based on temperature, time, and case numbers. The size of the bubbles represents the number of influenza cases, and different colors represent different temperatures. Users can filter the data using the year slider and state checkboxes, and the chart will dynamically update to reflect the new data.

```javascript
// Function to draw the bubble chart
function drawBubbleChart(data) {
  // console.log("Drawing bubble chart with data:", data);
  const container = document.getElementById('bubble-chart');
  container.innerHTML = '';
  const width = container.clientWidth || 800;
  const height = container.clientHeight || 500;
  const svg = d3.select("#bubble-chart").append("svg") …
  const margin = { top: 80, right: 30, bottom: 40, left: 80 }, …
  const chart = svg.append("g") …
  const x = d3.scaleLinear() …
  const y = d3.scaleTime() …
  const r = d3.scaleSqrt() …
  chart.selectAll("circle") …
  chart.append("g") …
  chart.append("g") …
  chart.append("text") …
  chart.append("text") …
  const legendWidth = 400;
  const legendHeight = 20;
  const legendSvg = svg.append("g") …
  legendSvg.append("text") …
  const legend = legendSvg.append("defs") …
  legend.append("stop") …
  legend.append("stop") …
  legend.append("stop") …
  legend.append("stop") …
  legend.append("stop") …
  legendSvg.append("rect") …
  const xScale = d3.scaleLinear() …
  const xAxis = d3.axisBottom(xScale) …
  legendSvg.append("g") …
}
document.addEventListener('DOMContentLoaded', function () { …
});
```

3.3 Using the Implementation

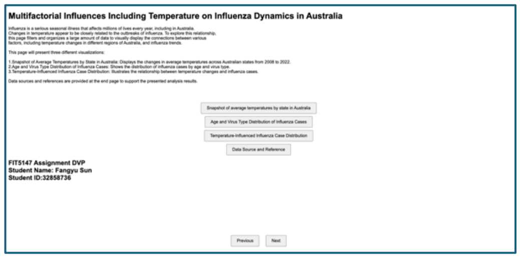
3.3.1 Pre-Configuration Instructions for Project Execution

First, ensure that the data and code are in the same folder. Place all necessary project files, including HTML files and required data files (CSV and GeoJSON files), in the same directory.

To avoid cross-origin request issues, you need to run a simple HTTP server locally before opening the HTML file. Please use the simple HTTP server provided by Python by entering the following command in the terminal or command prompt " python -m http.server 8000 ". After starting the server, you can access the project page through your browser to view and interact with the project.

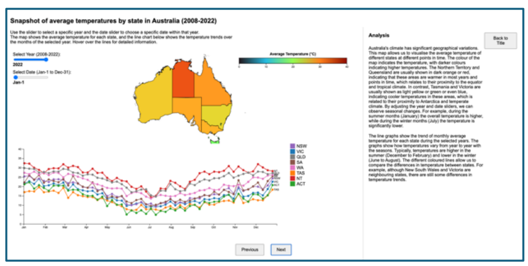
3.3.2 User Interface Instructions

3.3.2.1 Homepage

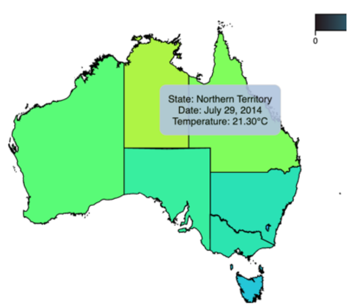


The homepage (Slide 1) includes the project title and an introduction, explaining the theme and purpose of the project. By clicking on the three navigation buttons, users can navigate to different visualization pages. Clicking the buttons at the bottom allows users to view the visualizations in sequence.

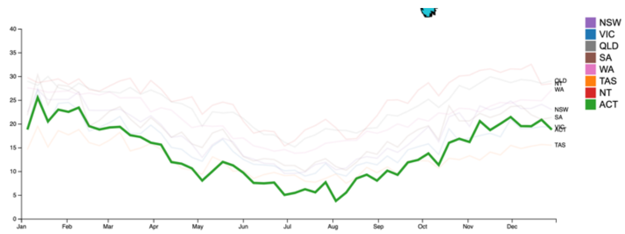
3.3.2.2 Heat Map and Line Chart



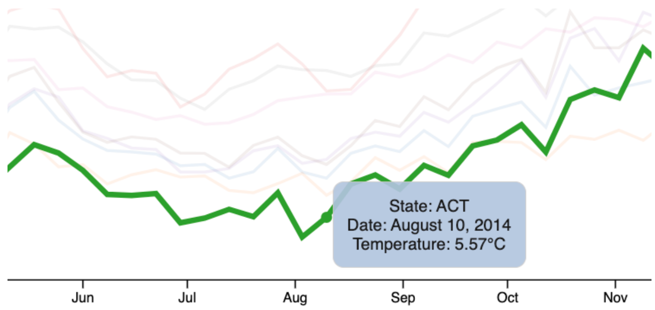
On the Average Temperature Snapshot page (Slide 2), the heat map shows the average temperatures of Australian states on different dates. Users can select specific dates by sliding the date slider, and the colors on the map will dynamically update to reflect the new temperature data. And the line chart shows the temperature trends of each state for the selected year. Users can select different years by sliding the year slider, and the line chart will dynamically update.



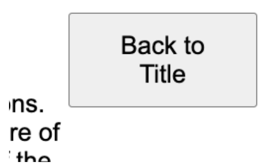
When the mouse hovers over a state's territory on the map, a tooltip will appear next to the mouse, displaying the average temperature (unit in Celsius, to two decimal places) for the selected date and state.



When the mouse hovers over a legend item on the right side of the line chart, the temperature trend for the selected year of the corresponding state will be highlighted.



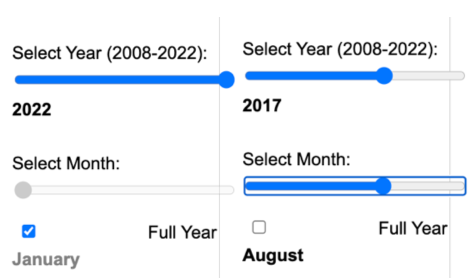
When the mouse hovers over a point on the line chart, a tooltip will also appear, showing the average temperature at that point in time for the state.



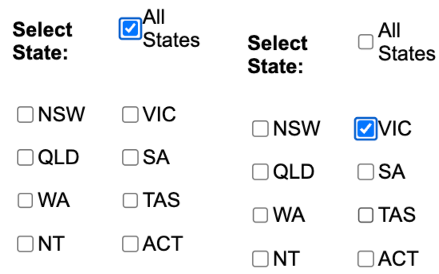
A "Back to Title" button on the right side of the analysis chart allows users to return to the homepage (same for 3.3.2.3, 3.3.2.4, and 3.3.2.5).

3.3.2.3 Stacked Bar Chart

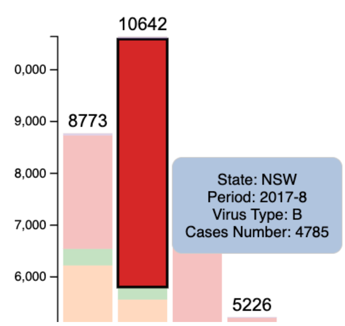
On the Influenza Cases Distribution page (Slide 3), the stacked bar chart shows the distribution of influenza cases by age group and virus type. Users can select specific time periods by sliding the year and month sliders and select different states using checkboxes.



When the "Full Year" checkbox is selected in the time module, the distribution of influenza cases for the entire year will be displayed (the month labels will be grayed out). Unchecking this option allows users to select a specific month's data.



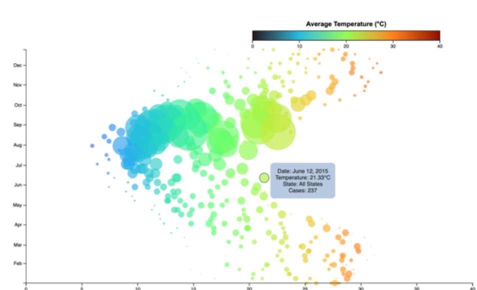
In the state selection module, selecting the "All States" checkbox displays the influenza cases for the entire country. Selecting one or more specific states will automatically uncheck the "All States" checkbox.



Hovering over the bar chart displays detailed information.

3.3.2.4 Bubble Chart

On the Temperature-Influenced Influenza Case Distribution page (Slide 4), the bubble chart shows the relationship between temperature, time, and the number of influenza cases. The size of the bubbles represents the number of influenza cases, and different colors represent different temperatures. Users can filter the data using the year slider and state checkboxes (similar to 3.3.2.3) and hover over the bubbles to view detailed information.



3.3.2.5 Data Source and Reference

The Data Source and Reference page (Slide 5) lists all data sources and references, supporting the analysis results presented.

4. Conclusion

From the visualizations in this project, we can conclude that temperature has a significant impact on the distribution of influenza cases in Australia. Through the heat map and line chart, we can see that during the summer months (December to February), regions with higher average temperatures (typically above 25°C) such as Queensland and Western Australia, show a significant decrease in influenza cases. Conversely, during the winter months (June to August), regions with lower average temperatures (typically below 10°C) such as Victoria and New South Wales, experience a notable increase in influenza cases. Users can select specific dates and years through interactive sliders to observe the dynamic relationship between temperature changes and influenza cases.

The stacked bar chart provides a detailed view of the distribution of influenza cases across different age groups and virus types. The analysis shows that the 0-4 age group and the elderly population (65 years and above) have higher case numbers during peak influenza seasons (such as June to August), especially in Victoria and New South Wales. Additionally, there are significant differences in the distribution of different virus types; for example, type A influenza is widely distributed across all age groups, while type B influenza is primarily concentrated in children and teenagers. Users can select different time periods and regions using sliders and checkboxes to gain further insights into the spread of influenza in specific populations.

The bubble chart further reveals the complex relationship between temperature, time, and the number of influenza cases. The analysis indicates that influenza cases peak when temperatures range between 15°C and 20°C, particularly during the transitional periods of autumn and winter (such as April to May and September to October). In the bubble chart, the size of the bubbles represents the number of influenza cases, and different colors indicate different temperatures. Users can clearly see that influenza cases significantly increase under specific temperature conditions (such as 15°C to 20°C). This visualization method not only provides an intuitive view but also enhances the depth and breadth of data analysis.

Despite the important findings, this project has some limitations. Firstly, the dataset used contains some missing values, which may affect the accuracy of the analysis. Secondly, the dataset's time span is relatively short, covering only from 2008 to 2022, which does not reflect longer-term trends. Finally, the temperature records in the dataset come from different stations, which may have some measurement errors that could impact the analysis results.

Future work can be improved in several ways: firstly, optimizing the data preprocessing and cleaning process to reduce the impact of missing values on the analysis. Secondly, enriching and expanding the data sources by incorporating more meteorological data and influenza case data to enhance the comprehensiveness of the analysis. Lastly, improving visualization techniques and interactive features to allow users to explore and understand the data more conveniently. Through these improvements, it is hoped that future research can provide more detailed data support for public health policy formulation and further enhance the scientific basis and effectiveness of influenza prevention and control.

References

ACORN-SAT. (2018). ACORN-SAT Australia v2 (ongoing): Australian Climate Observations Reference Network - Surface Air Temperature (1910 onwards) | Bureau Data Catalogue. Bom.gov.au;

ACORN-SAT. (2022). ACORN-SAT Station Catalogue. Bom.gov.au.

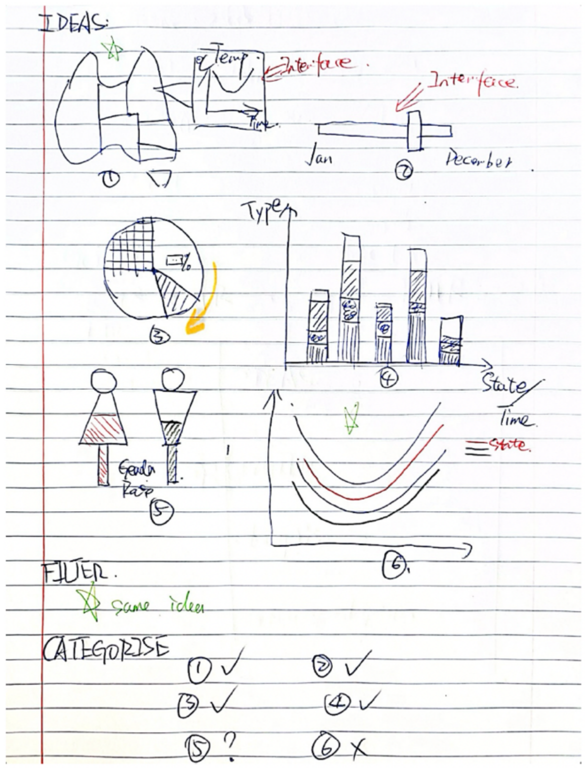
Australian Bureau of Statistics. (2024). National, state and territory population, September 2023 | Australian Bureau of Statistics.

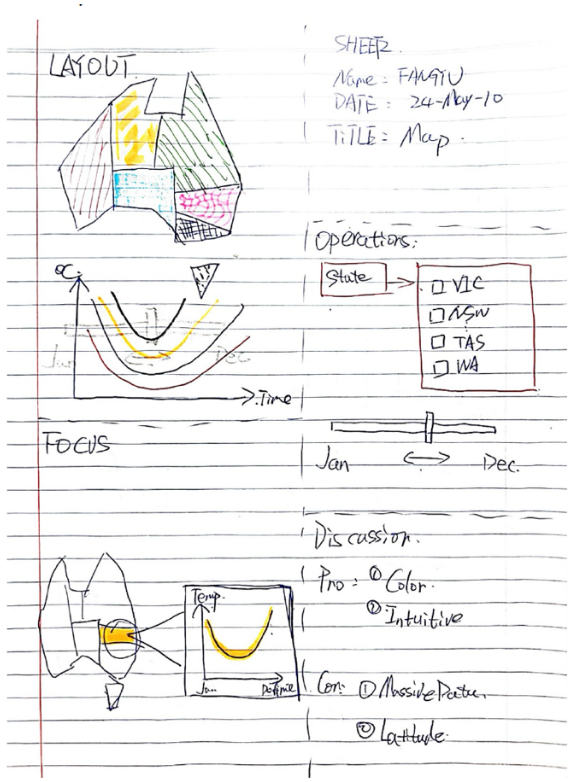
Github. (2024). Five Design-Sheets. Fds-Design.github.io. https://fds-design.github.io/

Health. (2021). National Notifiable Diseases Surveillance System (NNDSS) public dataset – influenza (laboratory confirmed). Australian Government Department of Health and Aged Care. health.gov.au.

Hogan, R. Australian States GeoJSON. Retrieved from https://raw.githubusercontent.com/rowanhogan/australian-states/master/states.geojson

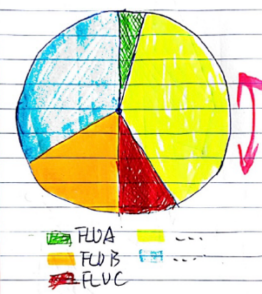
Appendix


IDEAS:
Temp
Interface.
Interface.
Jan
December
Type
State
Time
State
FILTER.
same idea
CATEGORISE


LAYOUT
SHEET.
Name: FANGYU
DATE: 24-May-10
TITLE: Map
Operations:
State
VIC
NSW
TAS
WA
Time
FOCUS
Jan
Dec.
Discussion.
Pro: Color.
Intuitive
Con: Massive Data.
Latitude.
Temp.

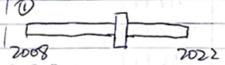
SHEET 3
Name: FANGYU SUN
Date: 24-May-10
TITLE: PIE

LAYOUT.



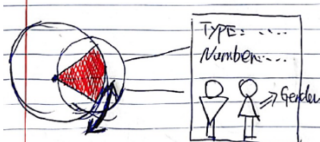
Operation:

①



② o VIC o NT o WA
o NSW o ACT o TAS
o QLD o SA

FOCUS.



Discussion:

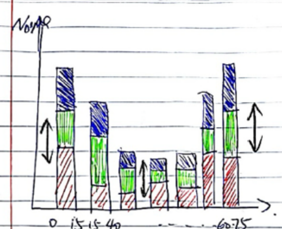
Pro: ① Size
② Color.

Con ① Different State

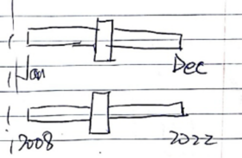
SHEET 4

Name: FANGYU SUN
Date: 24-May-10

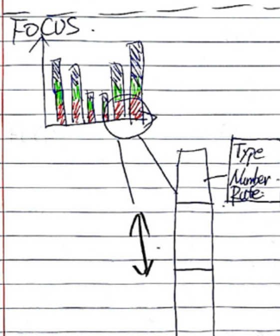
TITLE: LINE/Bar.

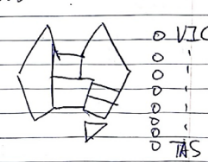


Operation



FOCUS.





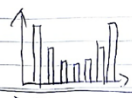
Discussion:

Pro ① Intuitively

Con: ① Can't show rate
② Portion

Layout.

Des1



Des2.

SHEET 5

Name: FANGYU SUN

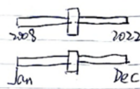
Date 24-May-10

Title: Final Visualization Draft.

MAP.

Desp of Map

Operation:



State

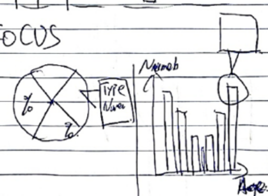
○ VIC ○ NT.

○ NSW ○ SA

○ QLD ○ WA

○ ACT ○ TAS.

FOCUS



DETAIL

① R / PYTHON.
TABLEAU/ Data Cleaning

② 1 week Working Time

③ ~~R~~: Javascript.
D3

④ R ⇒ ggplot, dplyr, tidyr, [illegible]