BRAIN STROKE

Stroke is a medical condition in which **poor blood flow** to the **brain** causes **cell death**.

There are two main types of stroke: **ischemic**, due to lack of blood flow, and **hemorrhagic**, due to bleeding. Both cause parts of brain to **stop functioning properly**.

The major risk factor for stroke is **high blood pressure**. Other risk factors include **high blood cholesterol**, tobacco **smoking**, **obesity**, and **diabetes** mellitus.

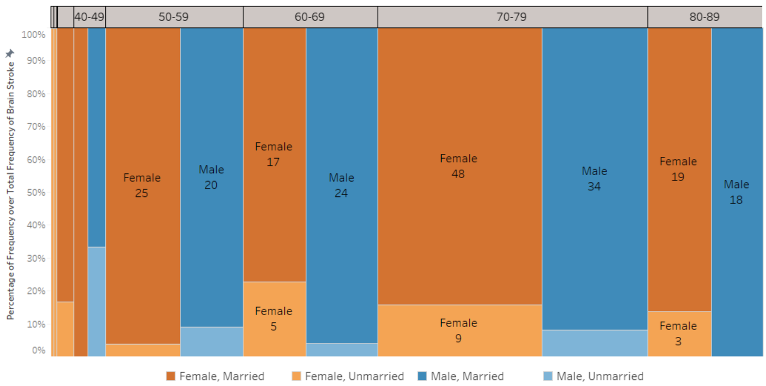
The dataset selected includes the personal data of patients in the hospital to investigate the potential **factors** affecting brain stroke occurrences.

---

1 DISTRIBUTION OF BRAIN STROKE PATIENTS OF DIFFERENT GENDERS AND MARITAL STATUSES ACROSS DIFFERENT AGE GROUPS

First and foremost, let us take a look at the overall distribution of brain stroke patients in the dataset over their gender, marital status and age.

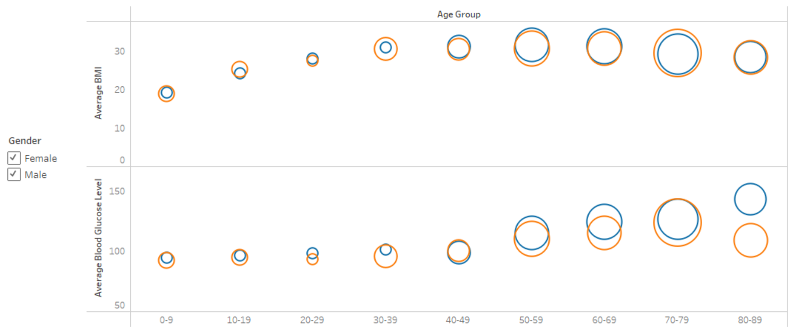
We can clearly observe that the age group of **70-79** had the **most** brain stroke patients. Also, **most** of the brain stroke patients were **married**.



2 RELATIONSHIP BETWEEN FREQUENCY OF BRAIN STROKE PATIENTS WITH AVERAGE BODY MASS INDEX (BMI) AND AVERAGE BLOOD GLUCOSE LEVEL IN MALE AND FEMALE ACROSS DIFFERENT AGE GROUPS

Now we look into certain leading factors of brain stroke such as obesity and diabetes mellitus, which are related to BMI and blood glucose level respectively.

Overall, the frequency of brain stroke patients is relatively **greater** in the **older age groups** if compared to the younger ones, as depicted by the **size** of bubbles. Another point to note is that **male** patients had **significantly higher average blood glucose level** than that of female in the age group of **80-89**.



3 IMPACT OF RESIDENCE TYPES (RURAL AND URBAN) ON BODY MASS INDEX (BMI) AND AVERAGE BLOOD GLUCOSE LEVEL OF PATIENTS

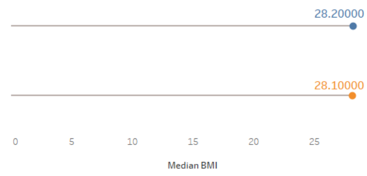
Since obesity and diabetes mellitus can lead to brain stroke, we could find out whether or not the residence type had an effect on the BMI and blood glucose level among patients, perhaps due to different lifestyles.

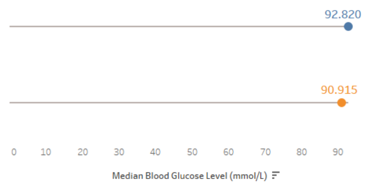
However, it was depicted that residence type **did not** seem to be a **significant factor** causing difference in BMI and blood glucose level among the patients based on their median values.

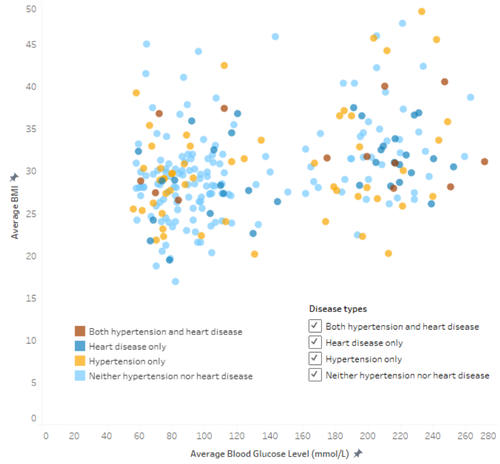
4 RELATIONSHIP BETWEEN BODY MASS INDEX (BMI) AND AVERAGE BLOOD GLUCOSE LEVEL

We would be curious to know if the BMI and average glucose level are correlated to each other.

According to the chart shown below, there **did not** seem to exist a **strong positive linear relationship** between BMI and average blood glucose level of patients. Nevertheless, they are still **slightly or moderately associated** as we can see a **slight increasing trend** in the scatter plot.



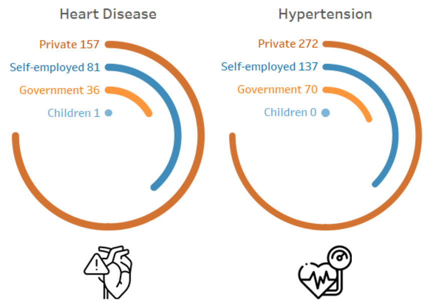




5 INFLUENCE OF JOB TYPES ON THE FREQUENCY OF HYPERTENSION AND HEART DISEASE PATIENTS

Besides, we can investigate if the job type of patients has an impact on their probability of suffering from heart disease and hypertension (caused by high blood cholesterol and high blood pressure respectively).

It can be observed that **most** of heart disease and hypertension patients were working in **private sector**. This could tell that employees under private sector were experiencing too much work stress if compared to the others.

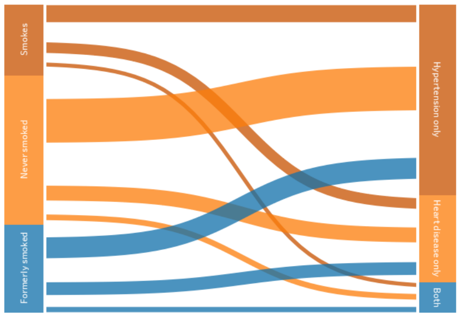


6 DISTRIBUTION OF BRAIN STROKE PATIENTS IN BETWEEN DIFFERENT SMOKING STATUSES AND DISEASE TYPES

Is smoking another factor leading to heart disease or hypertension?

We can notice that there is a **significant large proportion** of **non-smoking** brain stroke patients suffering from **hypertension** only, whereas three of the smoking statuses contributed fairly to the other two disease types.

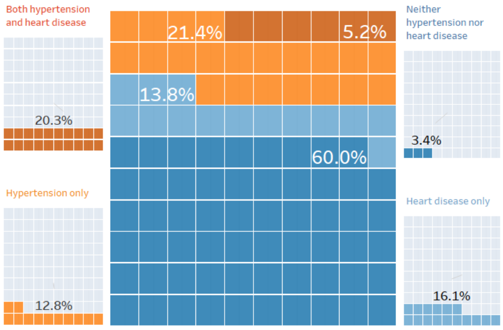
It tells that smoking is probably **not a direct cause** of hypertension. Instead, hypertension could be a result of **stress overload** as indicated previously.



7 PART-TO-WHOLE RATIO BETWEEN DIFFERENT DISEASE TYPES TO THE TOTAL FREQUENCY OF BRAIN STROKE PATIENTS

Smaller waffle charts indicate percentage of patients of different diseases suffering from brain stroke. Larger waffle chart displays the overall proportion of brain stroke patients over different disease types.

Patients suffering from **both hypertension and heart disease** had the **highest probability** of suffering from brain stroke, despite that they occupied the **smallest proportion** over the total brain stroke patients.

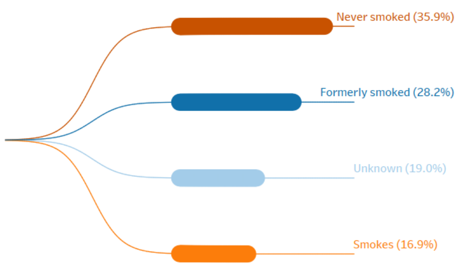


8 DIFFERENT SMOKING STATUSES AND THEIR PERCENTAGE OF BRAIN STROKE PATIENTS

Also, it was mentioned that tobacco smoking is a factor causing brain stroke.

Nonetheless, in this dataset we found out that **most** of the brain stroke patients were actually **non-smokers** as depicted in the chart below.

This could potentially be due to the reason that there existed **other factor(s)** causing the non-smoking patients to suffer from brain stroke **more significantly** if compared to solely tobacco smoking.



Visualisation by Chong Ming Sheng (Student ID: 32202792) for FIT3179 Data Visualisation Project 1, Monash University Malaysia.

URL of Visualisation

https://public.tableau.com/app/profile/ming.sheng.chong/viz/DataVisualisationProject1_16618011818780/FinalDashboard?publish=yes

Domain

My visualisation focuses mainly on health issue, particularly the main risk factors affecting brain stroke occurrences.

Why

My visualisation mostly depicts the relationship between some leading factors of brain stroke and the actual frequency of brain stroke patients. This visualisation can be used to predict probability of an individual suffering from brain stroke based on his/her current living lifestyle as indicated in the dataset. Therefore, perhaps individuals could take extra precautions to improve their health in order to prevent themselves from suffering from brain stroke.

Who

My visualisation aims to raise awareness among the worldwide population about the major risk factors leading to brain stroke.

What

The table dataset chosen was published by Jillani on Kaggle website (Jillani, 2022). It is a clean and real dataset which had been preprocessed properly before uploading to Kaggle. It includes the personal data of patients in hospital.

Why and How

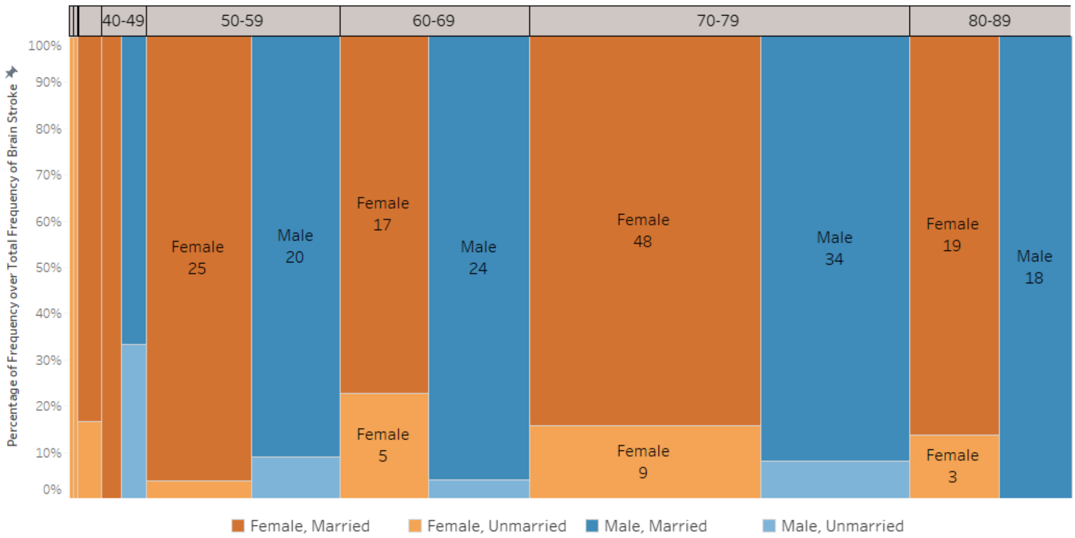


Figure 1: Marimekko chart.

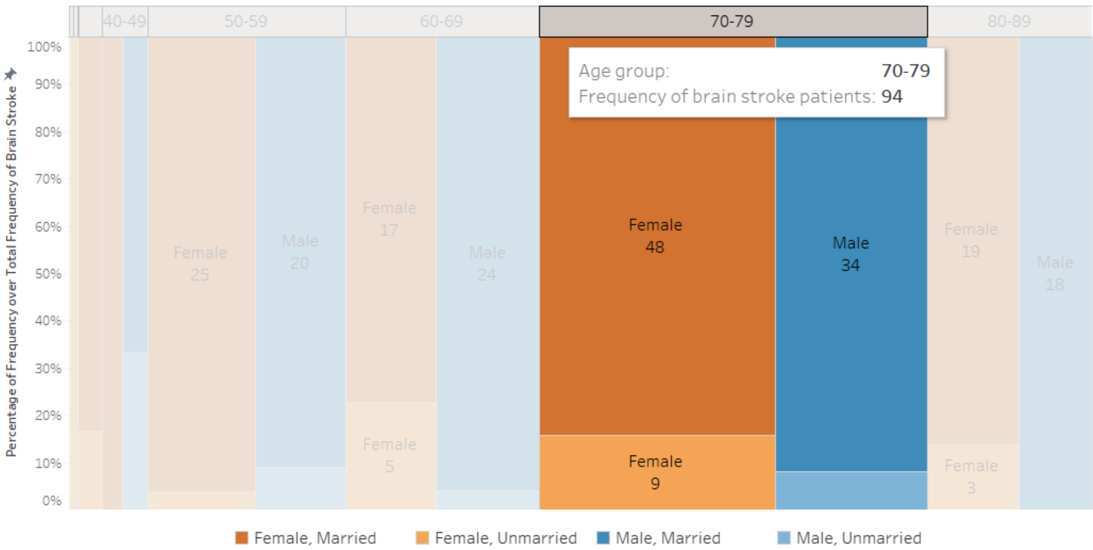


Figure 2: First special feature of Marimekko chart.

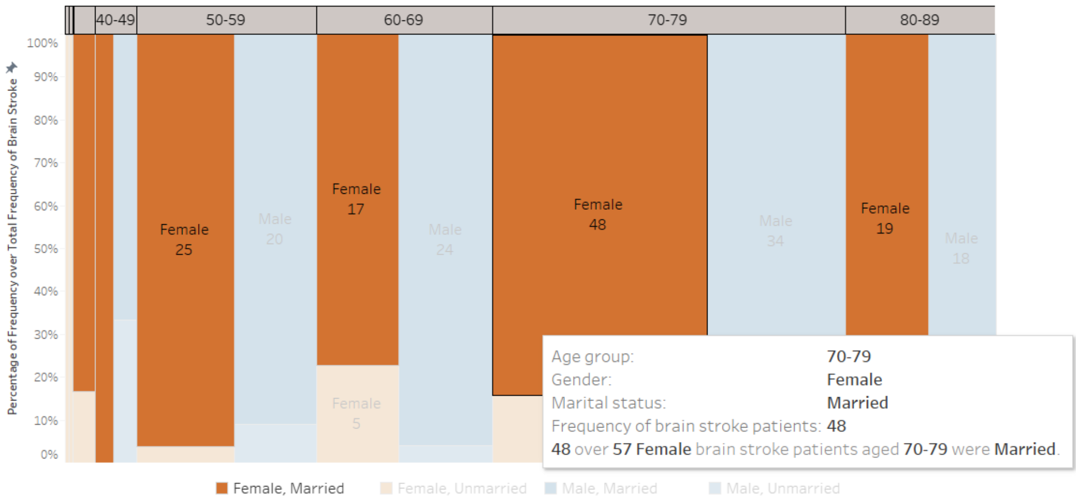


Figure 3: Second special feature of Marimekko chart.

The idiom in Figure 1 was designed with reference to "How to build a Marimekko chart in Tableau" (n.d.). It allows readers to see the part-to-whole relationship between both married and unmarried male and female brain stroke patients to the total frequency of brain stroke patients in different age groups. Readers can easily compare the proportion occupied by different categories of brain stroke patients within or across specific age group(s).

This Marimekko chart used areas (rectangles) as the mark, whereas horizontal spatial region, vertical position, area and colour hue as the channels, which respectively encode age groups, percentage of frequency of brain stroke patients over total frequency, frequency of brain stroke patients, and categories of brain stroke patients.

Based on Figure 2, hovering over any age group will highlight only the brain stroke patients within the particular age group. Based on Figure 3, hovering over any category of brain stroke patients will highlight all data under this particular category across all age groups.

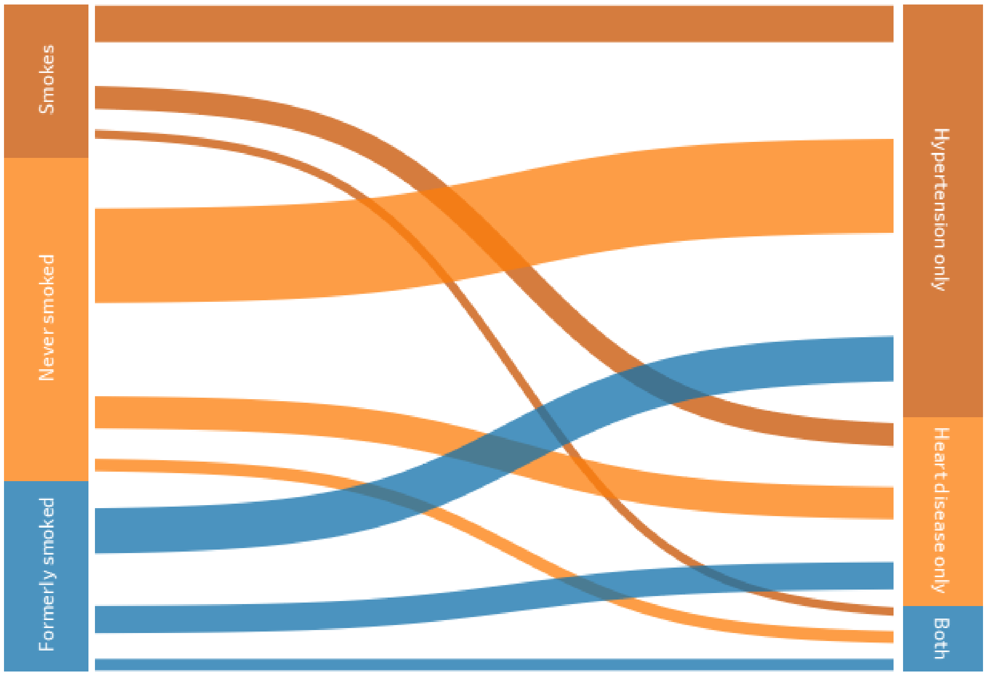


Figure 4: Alluvial diagram.

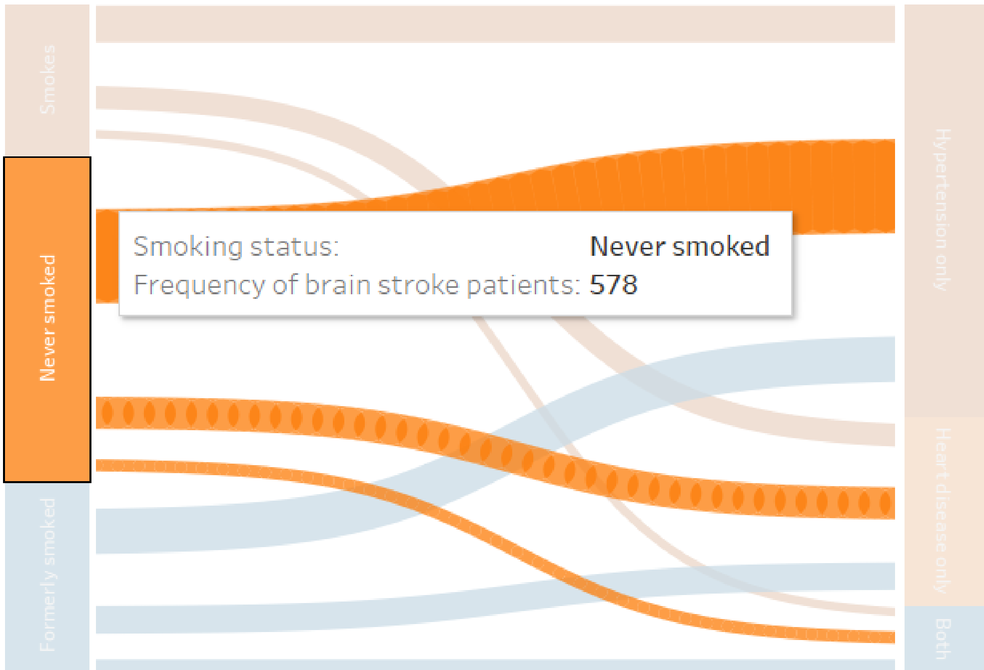


Figure 5: Special feature of Alluvial diagram.

The idiom in Figure 4 was designed with reference to TheDataQ (2020). It clearly displays the part-to-whole ratio between different smoking statuses and disease types among patients. Readers can visualise easily which smoking status occupied the largest or smallest proportion of specific disease type and vice versa.

This Alluvial diagram used lines and nodes as the marks, whereas size of connection lines and colour hue as the channels, which respectively encode frequency of brain stroke patients and smoking statuses as well as disease types.

Based on Figure 5, hovering over any smoking status or disease type will highlight its corresponding connection lines. Thus, readers could get the key information easily. For instance, non-smokers had the highest probability of developing hypertension, but the lowest probability of developing both hypertension and heart disease in themselves.

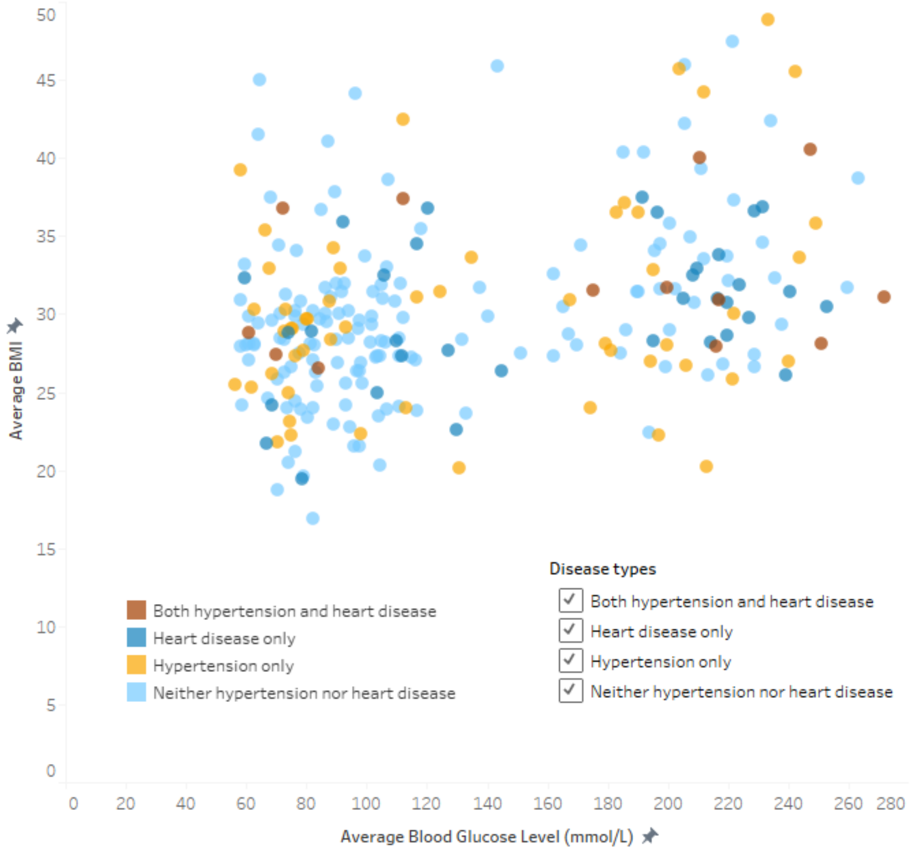


Figure 6: Scatter plot.

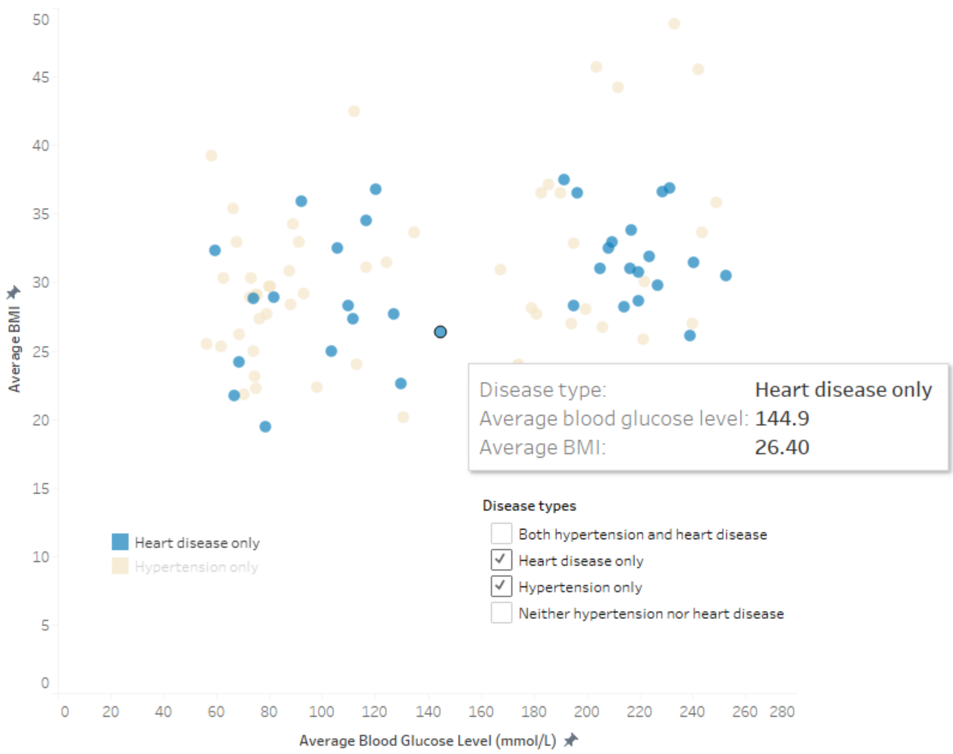


Figure 7: Special feature of scatter plot.

The idiom in Figure 6 depicts the correlation between average blood glucose level and average BMI among patients, plus the underlying trend in between them.

This scatter plot used points as the mark, whereas colour hue and both horizontal and vertical position as the channels, which respectively encode disease types and both quantitative attributes as indicated in the scatter plot.

Based on Figure 7, readers can utilise the filter feature to select or remove certain data points to be displayed, which improves readability of visualisation and reduces cognitive load. Also, hovering over any data point will highlight all data points that are of the same disease type. This allows readers to inspect the distribution of data points of specific disease type(s) clearly.

Design Layout

My visualisation consists of apparently six rows and a range of one to three columns. Rows and columns were consistently applied to separate different charts via whitespaces, reduce sight lines, and make the entire visualisation symmetrical and balanced. This creates proximity where readers could visually group relevant elements together easily. Whitespaces were also used to create a border for my visualisation.

Design Colour

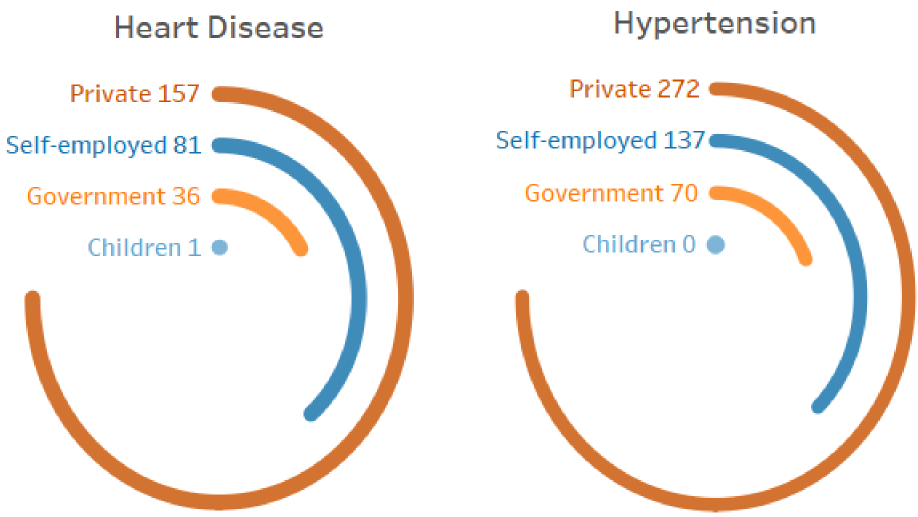


Figure 8: Consistent use of blue, pale blue, red and orange colours.

Colour palette that is visually friendly for colour blindness was selected so that the colour blind people could easily distinguish different colours used in my entire visualisation.

Blue, pale blue, red and orange colours from the colour-blindness palette were consistently applied to every chart in my overall visualisation to encode specific qualitative attributes where appropriate. One example was shown in Figure 8, in which the idiom was designed with reference to Data Embassy (2019).

Design Figure-ground

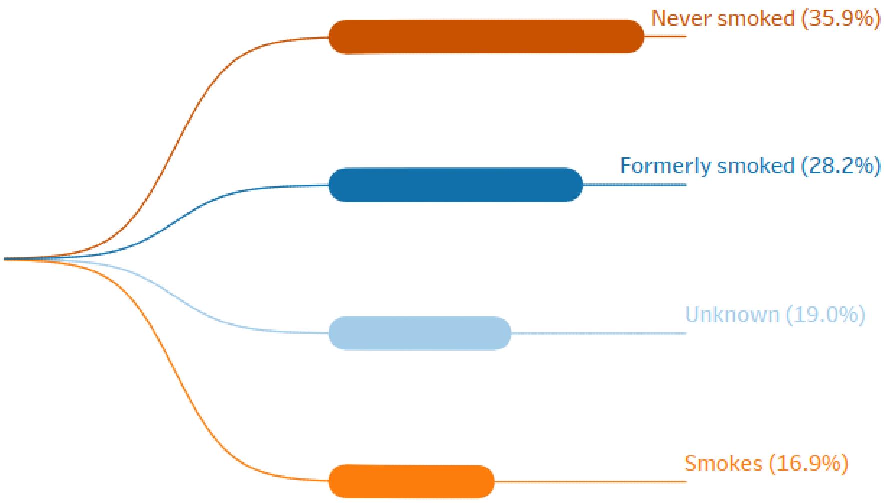


Figure 9: Desaturated colour encodes unimportant data.

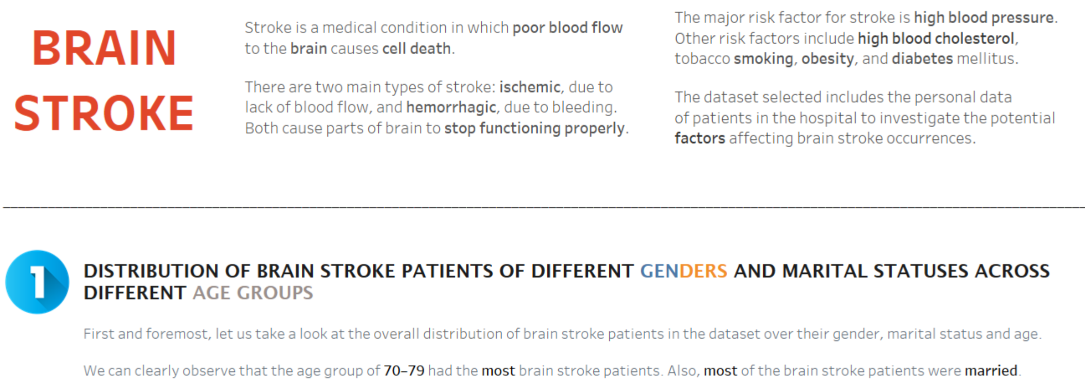


Figure 10: Application of typography to create visual hierarchy.

As shown in Figure 9 in which the idiom was designed with reference to Narang (2020), a relatively desaturated colour such as pale blue was used to encode unimportant data which is the “unknown” smoking status. This makes the other three important data to be more stand out apparently.

Furthermore, bolder, larger or capitalised texts were used to depict more important information such as visualisation title, chart titles and key messages in texts, as shown in Figure 10. This creates visual hierarchy to the texts.

Design Typography

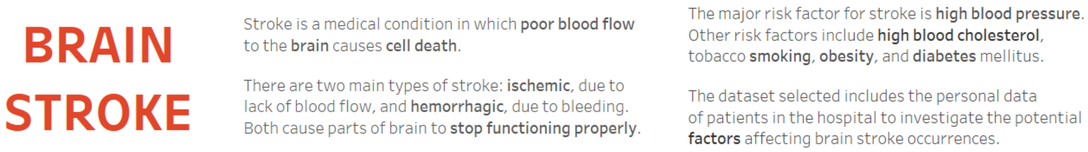


Figure 11: Two blocks of texts for caption.

Three different typefaces were used in my visualiastion, which are Tableau Bold, Tableau Book and Lucida Sans, because they are all non-standard typefaces and are very readable without serifs.

Tableau Bold makes the visualisation title extra bold and the most stand out. Lucida Sans differentiates chart titles from the captions that were written in Tableau Book. This was shown in Figure 10 previously.

Referring to Figure 11, the caption was divided into two blocks of texts so that each line does not contain more than ten words, improving readability of texts.

Besides, horizontal text layout was applied for captions below chart titles to minimise sight lines.

Design Storytelling

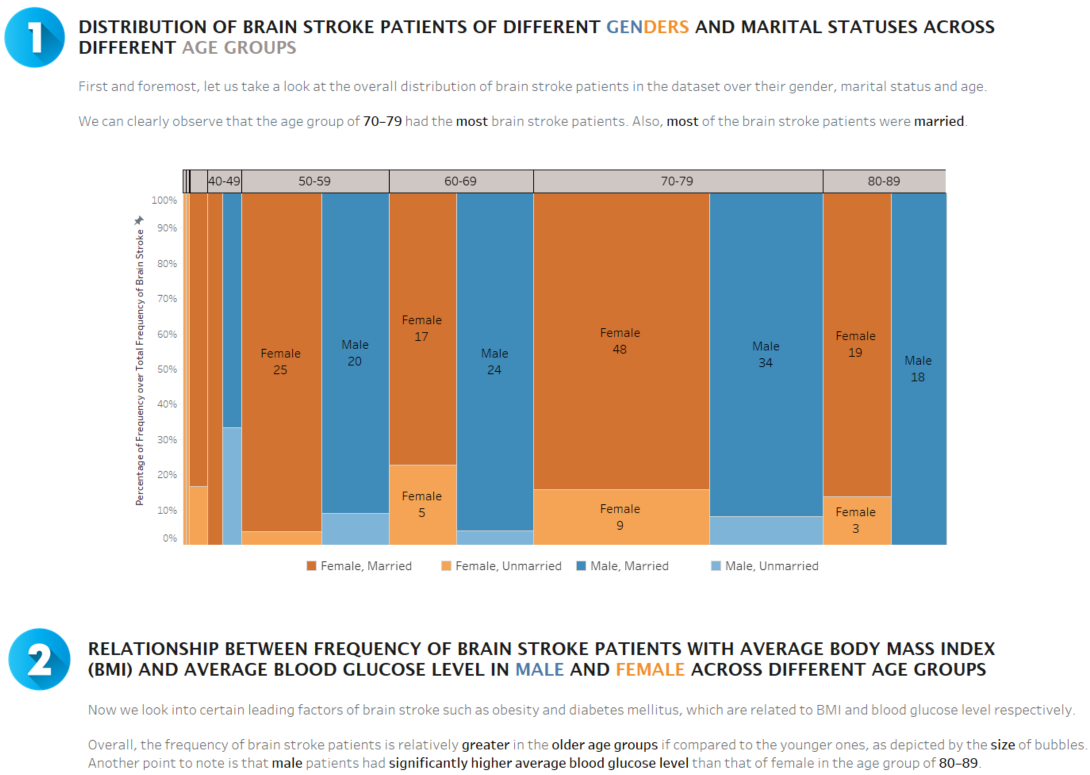


Figure 12: Number icons indicate flow of charts.

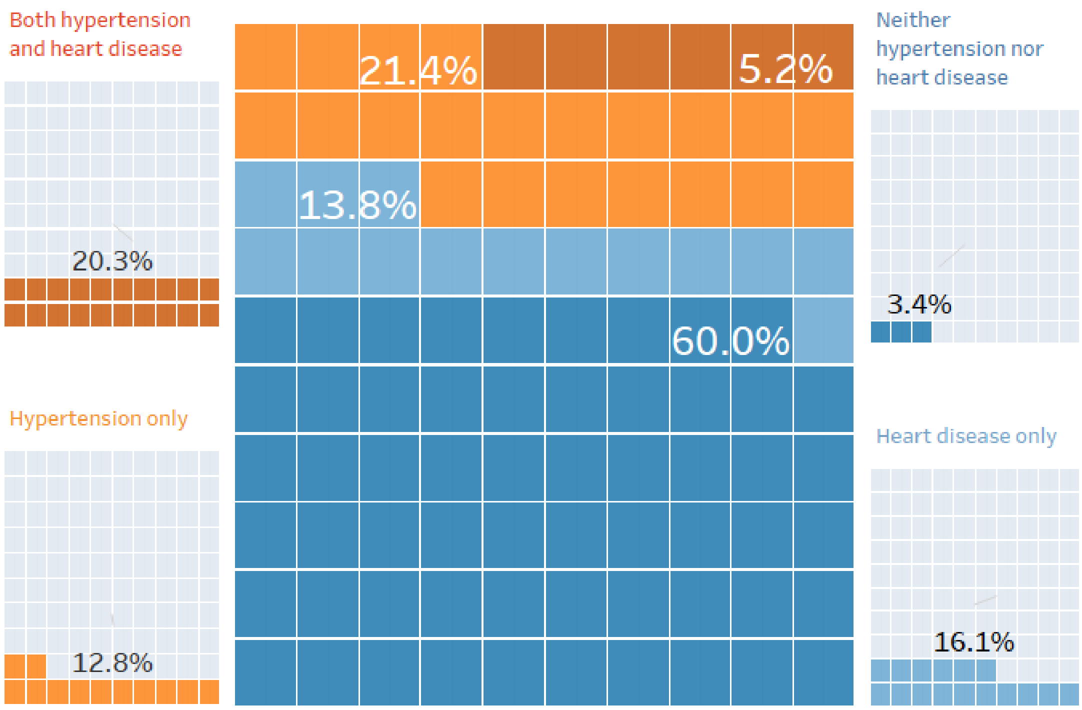


Figure 13: Annotations convey messages to readers attractively.

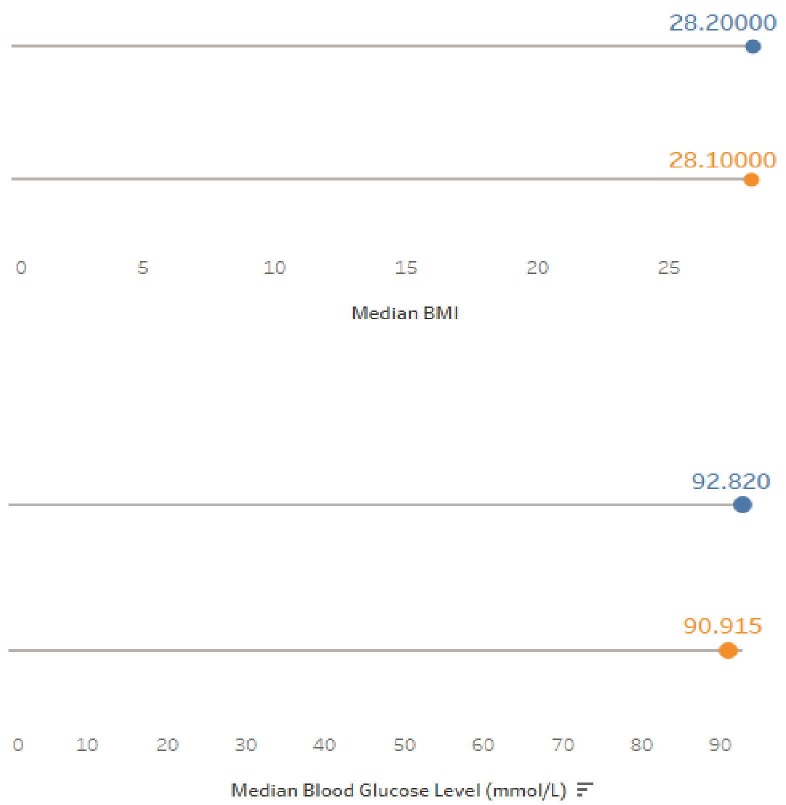


Figure 14: Display of data in specific order.

As shown in Figure 12, number icons were used to indicate the flow of charts in the visualisation to guide readers through my visualisation by viewing the charts based on specified order.

Moreover, based on Figure 13 in which the idiom was designed with reference to Kriebel (2017) and Krishna (2021), some annotations were provided to convey information to readers in a more appealing way, so that they could be guided through the annotated visualisation.

Lastly, some charts in my visualisation display the data in a specific order based on the value of their encoded quantitative attribute, as shown in Figure 8, Figure 9 and Figure 14 in which the idiom was designed with reference to Data Coach (2021). This could intuitively guide readers to read the data from top to bottom, which is from the highest to the lowest value of their encoded quantitative attribute.

Bibliography

Data Coach. (2021, January 16). Why is it the Cleveland dot plot? | Build this viz | Tableau tutorial with Data Coach [Video file]. Retrieved from https://www.youtube.com/watch?v=ICeRn3RE94A

Data Embassy. (2019, April 10). Tableau mini tutorial: Radial bar chart [Video file]. Retrieved from https://www.youtube.com/watch?v=JO4Wm2SLpKE&list=PLfq8fijvKB1E5ShIotyKnlhCHfwxOjIXp

How to build a Marimekko chart in Tableau [Web blog post]. (n.d.). Retrieved from https://www.tableau.com/about/blog/2016/8/how-build-marimekko-chart-tableau-58153

Jillani. (2022). *Brain stroke dataset* [Data file]. Retrieved from *https*://www.kaggle.com/datasets/jillanisofttech/brain-stroke-dataset

Kriebel, A. (2017, January 17). How to create waffle charts in Tableau [Video file]. Retrieved from https://www.youtube.com/watch?v=wRpBkl6-uhU&list=PLfq8fijvKB1E5ShIotyKnlhCHfwxOjIXp

Krishna, H. (2021, July 11). Creating waffle charts in Tableau [Web blog post]. Retrieved from https://medium.com/analytics-vidhya/creating-waffle-charts-in-tableau-a6a8f9e31898

Narang, M. (2020, October 16). How to create dendrogram chart in Tableau [Video file]. Retrieved from https://www.youtube.com/watch?v=J5AzsPW0c5k&list=PLfq8fijvKB1E5ShIotyKnlhCHfwxOjIXp

TheDataQ. (2020, March 1). Create Sankey chart in Tableau 2020 (done in 10mins with superstore data) [Video file]. Retrieved from https://www.youtube.com/watch?v=tXzj-mCaDcQ&list=PLfq8fijvKB1E5ShIotyKnlhCHfwxOjIXp

Appendix

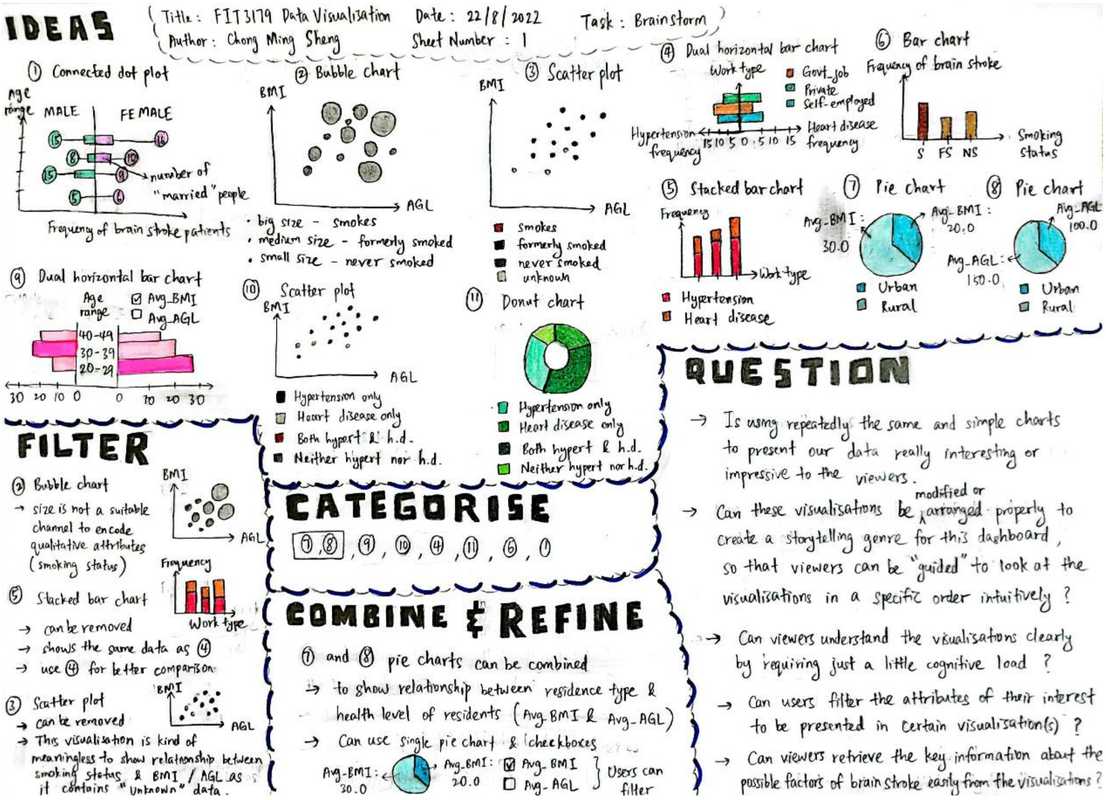


Figure 15: 5DS Methodology Sheet 1 (Brainstorm).

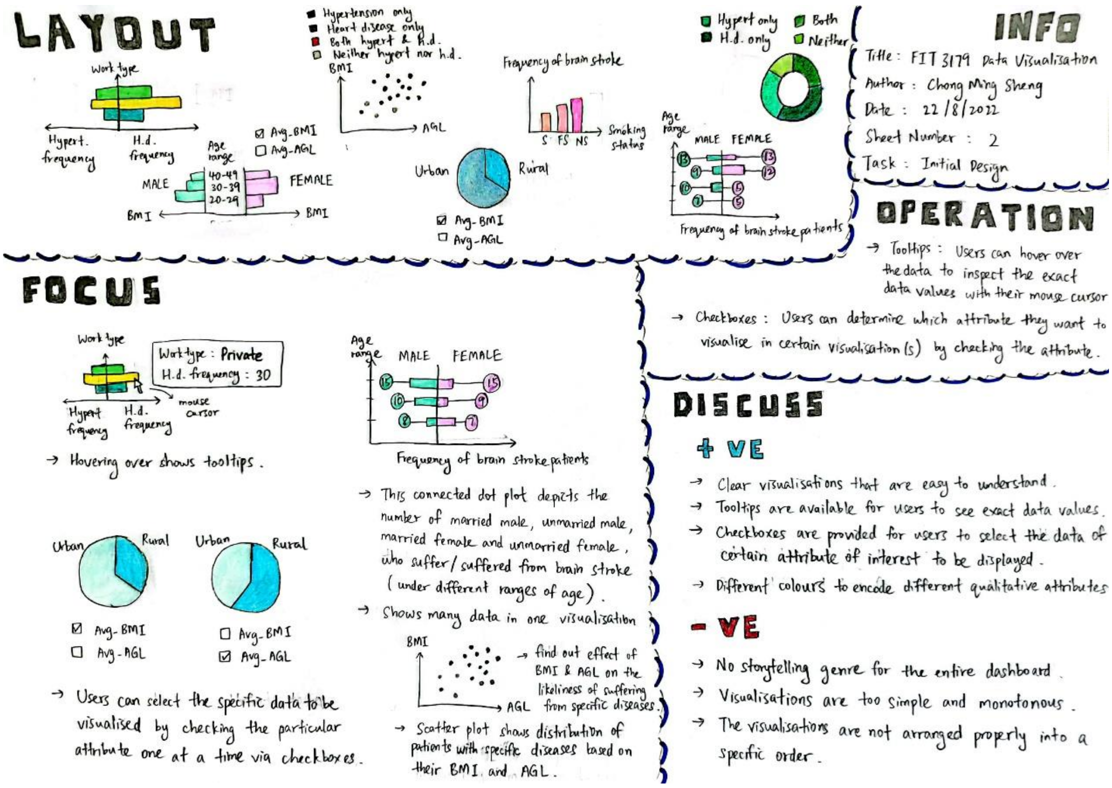


Figure 16: 5DS Methodology Sheet 2 (Initial design I).

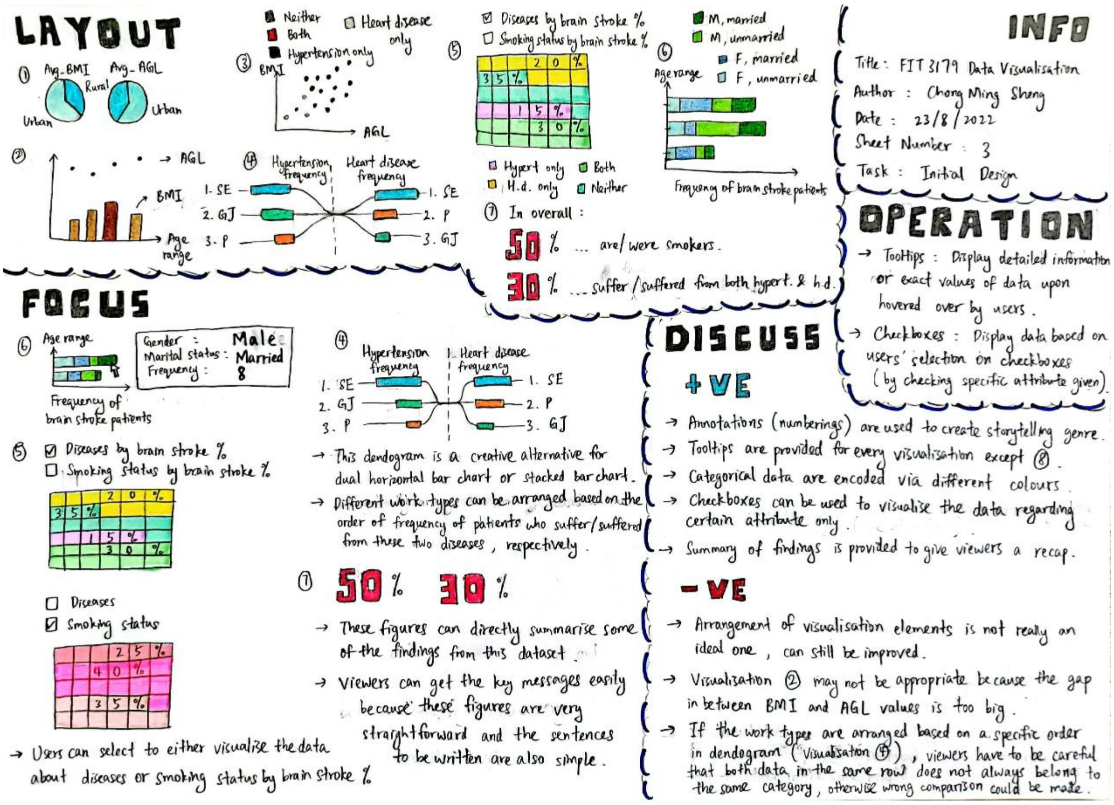


Figure 17: 5DS Methodology Sheet 3 (Initial design II).

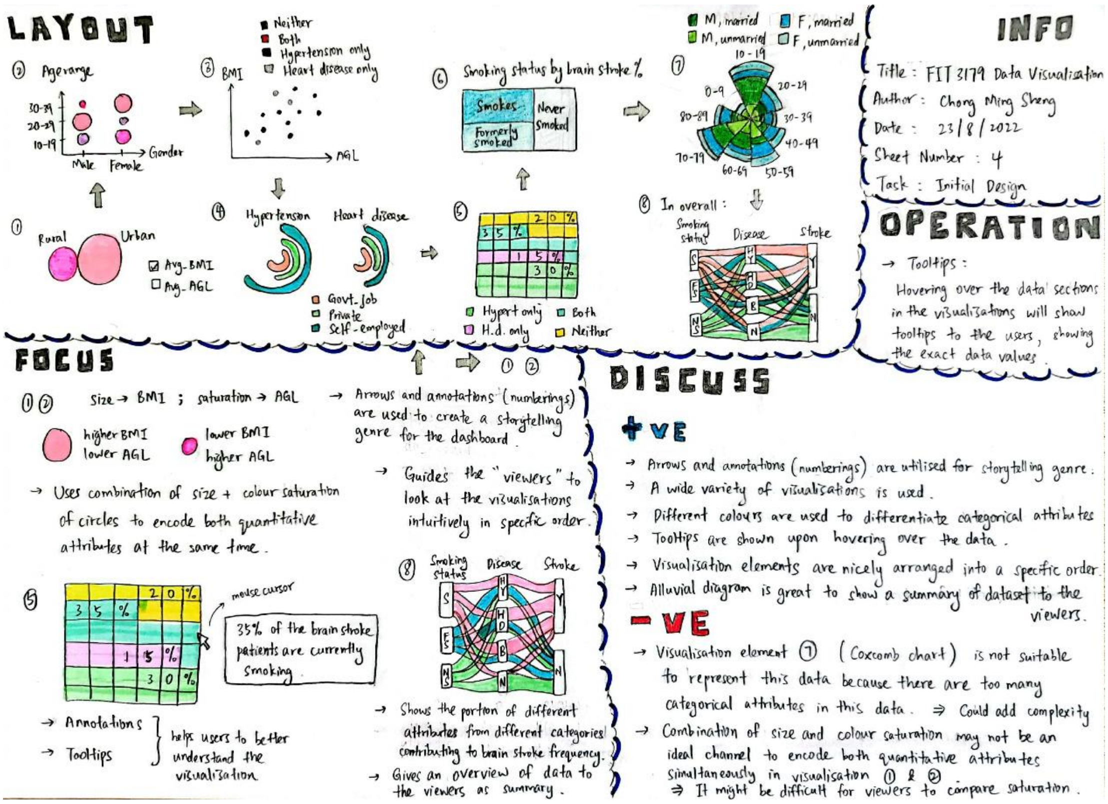


Figure 18: 5DS Methodology Sheet 4 (Initial design III).

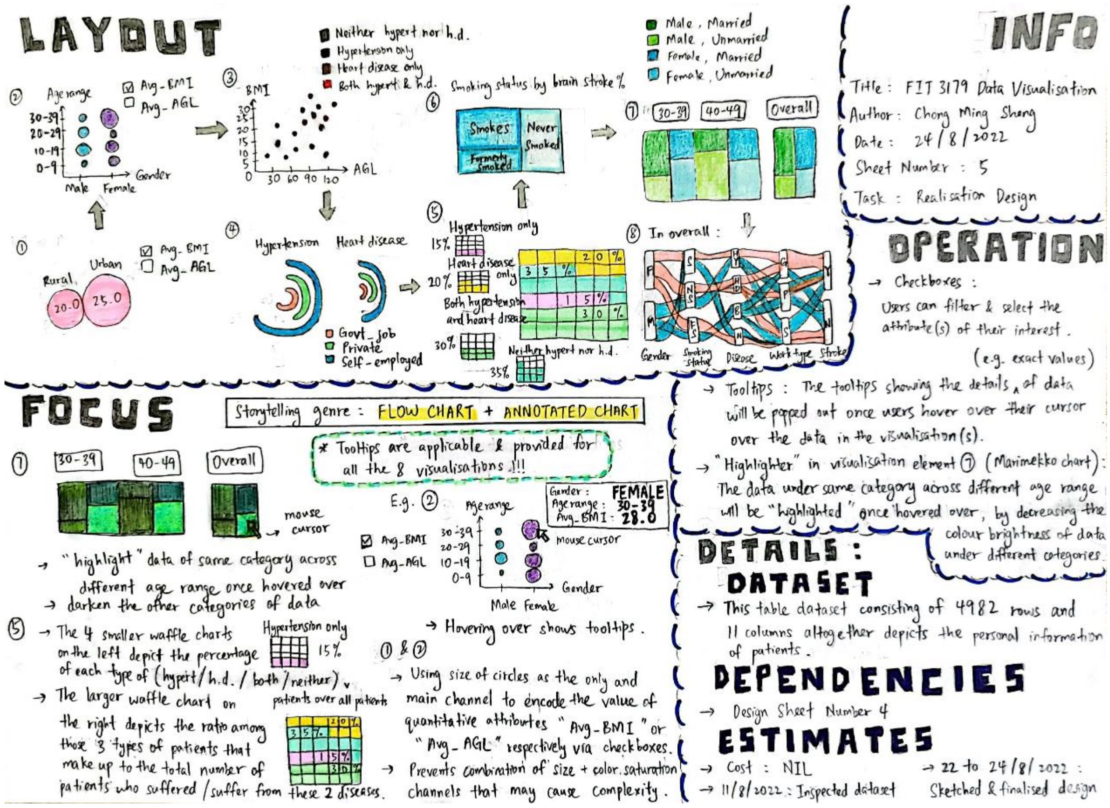


Figure 19: 5DS Methodology Sheet 5 (Realisation design).