

Webpage + Diary Entry

Phua Zong Yao

2023-10-19

WEEK 9

1. What is the topic that you have finalized? (Answer in 1 or 2 sentences)

The topic that I have finalized is social media in Singapore.

2. What are the data sources that you have curated so far? (Answer in 1 or 2 sentences)

I have made use of one of the data sources provided in the examples list. It is a Statista dataset with information on Social Media usage rates, patterns and the overall number of users. I decided to use this source as it contains many graphs and tables, and covers a wide range of data. The link is below.

<https://www-statista-com.libproxy1.nus.edu.sg/study/63698/social-media-in-singapore/> (<https://www-statista-com.libproxy1.nus.edu.sg/study/63698/social-media-in-singapore/>)

WEEK 10

1. What is the question that you are going to answer? (Answer: One sentence that ends with a question mark that could act like the title of your data story),

How has social media usage evolved over the past decade in Singapore, and what are the key trends influencing these changes?

2. Why is this an important question? (Answer: 3 sentences, each of which has some evidence, e.g., "According to the United Nations..." to justify why the question you have chosen is important),

It helps to provide marketers and businesses with important insights about where their target audience spends their time online, which platforms are gaining or losing popularity, and how to effectively engage with potential customers.

It allows us to study the impact of social media on mental health and overall well-being. Identifying trends in social media usage can help mental health organisations and professionals better address the challenges that may be associated with excessive social media use.

Investigating trends in social media usage helps us to identify potential privacy and security concerns. As the way people use social media evolves, so do the associated risks and vulnerabilities.

3. Which rows and columns of the dataset will be used to answer this question? (Answer: Actual names of the variables in the dataset that you plan to use).

The first table I will use is 'The number of social media users in Singapore 2017-2028 (in millions)'. The row I will use will be the years (2017-2028) and the column I will use will be the number of users in millions.

The second table I will use is 'The social media penetration rate in Singapore 2019-2028'. The row I will use will be the years (2019-2028) and the column I will use will be the share of population (in %).

The third table I will use is 'The average monthly hours spent per user on social media apps in Singapore from 2018 to 2021'. The row I will use will be the top 5 most popular social media platforms and the column I will use will be the average monthly hours per user.

The fourth table I will use is 'The size of the social network advertising audience in Singapore as of Jan 2023, by platform (in millions)'. The row I will use will be various social media platforms and the column I will use will be the number of addressable users in millions.

The fifth table I will use is 'The gender distribution of the social network advertising audience in Singapore as of Jan 2022, by platform'. The row I will use will be various social media platforms and the column I will use will be gender (male/female).

The sixth table I will use is 'The average time spent using online media in Singapore in Q3 2022, by activity (in hours per day)'. The row I will use will be various online activities and the column I will use will be the average time in hours per day.

The seventh table I will use is 'The main reasons for using social media in Singapore in 2022'. The row I will use will be various reasons for using social media and the column I will use will be the share of respondents (in %).

Include the challenges and errors that you faced and how you overcame them.

WEEK 11

1. List the visualizations that you are going to use in your project (Answer: What are the variables that you are going to plot? How will it answer your larger question?)

I can use of bar charts. Bar charts are quite effective in showing the comparisons between different categories or items. They are also effective for showing rankings, frequencies and proportions. This is quite relevant for my chosen topic. Some of the variables include years, average monthly hours per user, time spent on social media per day and share of population (as %).

I can make use of line charts. Line charts are useful for showing certain trends or changes over a period of time. They help to reveal changes in data and patterns in the data over a certain period of time. As I want to illustrate how social media usage evolved over the past decade in Singapore, it is important to show how usage rates and trends have changed over time.

I can make use of histograms. For some of the data, histograms may actually be more useful as it can summarise discrete or continuous data that are measured on an interval scale. It is also more convenient to use histograms when dealing with large data sets (such as those with more than 100 observations).

2. How do you plan to make it interactive? (Answer: features of ggplot2/shiny/markdown do you plan to use to make the story interactive)

With ggplot2, I can make use of 'geom_bar()' to create bar charts, and stack bars by filling them with various colours to show the different composition.

With ggplot2, I can make use of 'geom_line()' and possibly 'geom_area()' to display trends over time, along with other ordered variables. Again, various colours can be used to distinguish between the different categories or groups.

I also plan to experiment with features such as igraph and networkd3. These can be found within shiny, and can allow users to interact with the graphs.

3. What concepts incorporated in your project were taught in the course and which ones were self-learnt? (Answer: Create a table with topics in one column and Weeks in the other to indicate which concept taught in which week is being used. Leave the entry of the Week column empty for self-learnt concepts)

Topics Weeks Inserting and adjusting images 1, 2 Bar and Line Graphs using ggplot2 7 Shiny dashboards and interactive graphs 2, 11

Include the challenges and errors that you faced and how you overcame them.

The main challenge was with formatting the quarto project for the github website. In several instances, I could envision what I wanted my website to look like, or how I wanted a particular set of data to be displayed, but I did not know what codes/commands to use to achieve my desired outcome. However, I view it as a good thing in the sense that I got to learn more about using r markdown, shiny and ggplot2 when I used online resources and lecture notes to look for the solution.

WEEK 12

Include the challenges and errors that you faced and how you overcame them (if any)

1. As a new user to both Rstudio and Quarto, I feel that the learning curve was quite steep, even though these software are not viewed as particularly difficult or challenging to pick up. For this, I had to make use of ChatGPT, ask classmates and the TAs for help and just continue to code and learn as I went along.
2. Even after constructing the required graphs, it took me some time to structure my data story in a engaging and overall coherent manner. For this, I mainly went to YouTube and various websites to see possible ways I could display and subsequently explain the data story. That helped quite a bit.
3. On top of selecting the appropriate charts and graphs to convey the story, I feel that further customization of the visualizations to meet the specific needs may require higher levels of coding skills, or more experience, which I currently lack. For this one, I had to ask my classmates, TA and Google for tips on when each type of chart would be most suitable (depending on the nature of the data). After a while, it became somewhat easier to understand.
4. I also encountered challenges in the early stages of adding more interactive elements into my data story and website. Additionally, I also realized that ensuring interactivity runs smoothly across various devices and platforms can also be a challenge. Again, this was mainly trial and error, but thankfully I did not encounter too many disruptions in this area.

WEEK 13

(1) What is the theme of your data story?

The theme of my data story is "How has Social Media in Singapore evolved over time?"

(2) Why is it important to address this question?

It is quite important for several reasons.

a. Communication and Information Flow

Understanding the evolution of social media is crucial for grasping how information and communication have transformed in Singapore. It helps in comprehending the channels through which people access news, interact with content, and communicate with each other.

b. Business and Marketing Strategies

For businesses and marketers, understanding the social media landscape is essential for crafting effective marketing strategies. It involves recognizing which platforms are popular among target demographics, understanding trends, and leveraging the right channels for brand promotion.

c. Government Policies and Regulation

Given the government's active engagement on social media, being aware of the evolution helps in understanding the context within which policies and regulations are formulated. This is especially important as governments around the world are increasingly addressing issues related to misinformation, privacy, and online behavior.

d. Cultural and Social Dynamics

The evolution of social media is closely tied to cultural and social dynamics. It reflects changes in how people connect, share information, and express themselves. Studying this evolution provides insights into societal shifts and trends.

e. Impact on Society

Social media has a profound impact on society, influencing opinions, shaping narratives, and playing a role in public discourse. Understanding this evolution helps in assessing the broader societal impact of social media in Singapore.

(3) Why do you think the data sources that you have curated can help you answer the question?

(4) What are the insights from the data and how are they depicted in plots?

I feel that the various data sources I have curated will be able to provide valuable insights into the evolution of social media in Singapore over time by offering a comprehensive view of various aspects such as user growth, platform preferences, engagement patterns, content trends, and advertising dynamics.

For instance, the data on the number of social media users in Singapore over the years (2017-2028) allows us to observe the trajectory of user growth. This information is fundamental for understanding the overall expansion and adoption of social media platforms in the country.

Also, the data on the average time spent on social media in Q3 2022 offers insights into user behavior and preferences. It allows us to understand the role of social media in the daily lives of individuals, providing context on how it competes with other forms of media consumption.

Likewise, the data on the advertising audience for various social media platforms, segmented by gender, is crucial for advertisers and marketers. It offers insights into the demographic composition of the audience, enabling targeted and effective advertising strategies.

By combining these data sources, we can construct a cohesive narrative of how social media has evolved in Singapore. The user-centric data, coupled with insights into content trends and advertising dynamics, enables a holistic understanding of the changing landscape. This comprehensive view is essential for stakeholders, including advertisers, content creators, businesses, and platform developers, to adapt and tailor their strategies to the evolving preferences and behaviors of social media users in Singapore.

(5) How did you implement this entire project? Were there any new concepts that you learnt to implement some aspects of it?

I started off with some data loading and cleaning. That involved importing the provided dataset into RStudio using functions like `read.csv` or other appropriate data reading functions. Then, I made use of summary statistics and visualization techniques to explore the characteristics of the dataset. From there, I began exploring trends and patterns in the data to gain initial insights.

Afterwards, I experimented with `ggplot2` to create various visualizations, such as line plots for user growth, bar charts for reasons for using social media, and pie charts for time spent on social media. I also took some time to explore concepts like faceting, theme customization, and color scales in `ggplot2` to enhance visualizations. Alongside `ggplot2`, I also enhanced visualizations with interactive features using `plotly`. These would include concepts such as adding hover effects, zooming, and panning for a more engaging user experience.

Lastly, I also attempted to develop a Shiny web application to present the insights in an interactive and user-friendly manner. While doing so, I started learning concepts like reactive programming, UI design, and server logic in Shiny. To conclude, I deployed the shiny app on a dedicated server. This was because of a deployment consideration where shiny could not be run on quarto, which was deemed a static markdown page.

(694 words)

References

Howe, S. (2023, March 29). Social Media Statistics in Singapore [Updated 2023]. Meltwater.

<https://www.meltwater.com/en/blog/social-media-statistics-singapore>

(<https://www.meltwater.com/en/blog/social-media-statistics-singapore>)

NUS Libraries: Library Proxy: Login. (n.d.). Login.libproxy1.Nus.edu.sg. Retrieved November 24, 2023, from

<https://www-statista-com.libproxy1.nus.edu.sg/study/63698/social-media-in-singapore/> (<https://www-statista-com.libproxy1.nus.edu.sg/study/63698/social-media-in-singapore/>)