Diary entry submissions and Final submission

Merkayla Wong

2023-11-24

Week 9 (1) What is the topic that you have finalized? (Answer in 1 or 2 sentences) Upon scrolling data.gov, I was interested in finding out more about resale flat prices, particularly since I am graduating soon and will be looking for a place in Singapore. My final topic is Exploring HDB Resale Prices in Singapore for the year 2023. In question 2 I will describe why I chose this data set.

2. What are the data sources that you have curated so far? (Answer 1 or 2 sentences) https://beta.data.gov.sg/datasets/d_8b84c4ee58e3cfc0ece0d773c8ca6abc/view (https://beta.data.gov.sg/datasets/d_8b84c4ee58e3cfc0ece0d773c8ca6abc/view) The data set was chosen as it is very comprehensive. It is up to date spannig from 2017 till now and also, alot of data is being collected on this particular topic (it comprises of 11 columns).

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)

Do regions nearer the city center anticipate higher median resale prices for HDB flats in 2023, driven by enhanced convenience and accessibility, rendering them more comfortable and desirable living spaces?

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),

The question of whether regions closer to the city center have higher median resale prices is crucial for prospective homebuyers and policymakers to understand housing dynamics in Singapore. Evidence from the data story reveals that spatial patterns in median resale prices provide insights into the accessibility and convenience factors influencing property values. Furthermore, understanding anomalies in pricing, such as certain central areas being priced lower is essential for refining housing policies and informing potential homebuyers about nuanced considerations in different regions.

-Affordability: According to the Straits Times Singapore in 2021, affordability for housing are key issues for young Singaporeans.New prospective buyers often have budget constraints. For eg, understanding the factors that influence 3-room flat prices in a specific location can help them determine whether Ang Mo Kio is an affordable choice.

-Property Selection: Insights into factors like flat size or location can help buyers decide which property suits their needs best.

-Market Trends: Buyers need to be aware of market trends and fluctuations to make informed decisions about the timing of their purchase.

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

The rows i am using are those that are in the year 2023. As for columns, these are the column that I will be using “month”, “town”, “flat type” and lastly, “resale price”. Some challenges includes filtering the year to 2023, as it was together with the month under the “month” column. The median resale price must also be calculated for the the year 2023 to get median resale price.

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?),

The first visualization involves an interactive map of Singapore, where regions of the city are plotted as dots. By hovering over these dots, users can access information on the median resale prices of flats in each region during the year 2023. This spatial representation allows for a nuanced exploration of how housing prices are distributed geographically, providing a contextual understanding of the overall market landscape. The second visualization takes the form of a non-interactive bar graph, where each bar corresponds to a specific region in Singapore. The variable plotted is the median resale price of flats in each region for the year 2023. This static graph serves as a quick reference for users to see the median resale prices across different areas. The third visualization is an interactive dropdown menu graph. The variables plotted include the x-axis representing regions of Singapore, the y-axis representing resale prices of different flat types (2 room, 3 room, 4 room, multigenerational, executive, one room), and the bar heights indicating the total number of flats sold in each region for the selected flat type. This dynamic graph empowers users to focus on specific flat types, exploring how resale prices vary across regions. Additionally, the inclusion of the total number of flats sold offers insights into the demand for different types of flats in different areas.

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),

For the Interactive map, users can hover over the map to view different median resale prices across Singapore. In the dropdown option, i can make it interactive by including user input where users can view different flat types.

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)

```
Table <- data.frame(Concept= c("ggplot","shiny","Creating new columns and extracting data","g
gplot2 visualising data", "Integration shiny into quarto","Choosing what are the suitable gra
phs for the data presented","Self learnt leaflet by watching YouTube and searching online for
examples","Learnt to troubleshoot when upload to GitHub") , Weeks =c("Week 2","Week 2","Week
4","Week 7","Week 8","Week 10","NA","NA"))
Table
```

```
##                                                                  Concept
## 1                                                                ggplot
## 2                                                                 shiny
## 3                                  Creating new columns and extracting data
## 4                                                ggplot2 visualising data
## 5                                             Integration shiny into quarto
## 6              Choosing what are the suitable graphs for the data presented
## 7 Self learnt leaflet by watching YouTube and searching online for examples
## 8                                Learnt to troubleshoot when upload to GitHub
##     Weeks
## 1  Week 2
## 2  Week 2
## 3  Week 4
## 4  Week 7
## 5  Week 8
## 6 Week 10
## 7      NA
## 8      NA
```

Week 12 (1) When I was deciding to plot the interactive map, I realised there was no geographical data for me to reference alongside with the town data in the dataset. Hence I needed to source the geographical data of each town on my own to find the various latitude and altitude values that were required in the map plot. After finding the various values, I had to manually key into a dataset as I could not find the file for such data, I could only get raw values from the internet.

While I was plotting the map, I was also unfamiliar with the use of leaflet and it took me a while after watching a few YouTube videos. There were also a few errors that I have encounter along the way with the map plot such as missing reference variables when plotting the map. In the end I found out the mistake was due to an extra filter function that was unnecessary and it cause some data to be filtered out, leaving the data that was used to plot having less entries than expected.

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

The theme of the data story revolves around understanding and exploring the dynamics of the housing market in Singapore in the year 2023. Through interactive visualizations, the story aims to provide insights into the spatial distribution of median resale prices, regional comparisons, and variations in resale prices for different types of flats. The theme aims to help users gain a comprehensive view of the housing landscape, enabling them to make informed decisions or observations regarding the real estate market in Singapore.

2. Why is it important to address this question?

Understanding the housing market is crucial for various stakeholders, including potential homebuyers, real estate investors, and policymakers. Housing is a fundamental aspect of people's lives, and its affordability and accessibility impact the well-being of communities. By examining median resale prices across regions and different flat types, the data story can contribute valuable information for individuals looking to invest in property, prospective homebuyers making informed decisions, and policymakers seeking insights to address housing challenges.

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

Taken from data.gov and curated from HDB, the dataset is comprehensive.including information on median resale prices, flat types, regions, and total sales, providing a comprehensive overview of the real estate landscape. The one that I selected offers a detailed, structured, and quantitative representation of the housing market dynamics, enabling meaningful analysis and visualization. The reliability and currency of the data are critical to ensuring that the insights derived accurately reflect the current state of the market in 2023.

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

The data analysis provides valuable insights into the resale property market in Singapore. Firstly, the interactive map shows clear spatial patterns of resale prices, demonstrating that regions in closer proximity to the city center generally command higher median resale prices. This is evident through the varying colors and sizes of dots representing different regions, providing a visual representation of the geographical distribution of property values.

Secondly, the data story highlights anomalies in certain central areas. Contrary to the common hypothesis that central regions would uniformly have higher prices, certain areas defy this trend and are priced lower. The investigation attributes these anomalies to estate maturity, shedding light on the nuanced factors influencing property prices within central districts. This insight is prominently depicted in both the interactive map and the dropdown graph, offering a comprehensive view of the intricacies in pricing patterns.

Furthermore, the analysis reveals a compelling correlation between flat size and location. Larger flat types are found to be more prevalent and affordable in regions farther from the central area, challenging preconceptions about the desirability and comfortability associated with central locations. This correlation is visually depicted in the interactive dropdown graph, where the prevalence of larger flat types is linked to regions situated at a distance from the city center.

After determining the types of visualizations needed for my project, I sought tutorials to guide the implementation process and ensure suitability. Data cleaning involved handling NA values and extracting essential variables, with guidance from chatpgt in resolving any encountered errors. Reference slides aided in enhancing the visual appeal of my plots. The deployment onto GitHub posed challenges, particularly on shiny.io, attributed to directory issues within the NUS One Drive.

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

During the planning of the interactive map, a lack of geographical data alongside town data in the dataset prompted the sourcing of latitude and longitude values for each town. These values were manually entered into a dataset as the raw data was available but not in file format. While grappling with the use of leaflet for mapping, additional effort was invested after watching instructional videos to familiarize myself with the tool. Challenges surfaced, including missing reference variables, which were later traced back to an unnecessary filter function causing data discrepancies in the plotted map.