Final Submission

jingying

2023-11-24

Questions

Week 9

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

The topic I have finalised on is on Starbucks in California, US, and its correlation to race and background of inhabitants in specific cities within its branding.

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

The data source I have obtained is for free, and originates from the following link: https://data.world/alice-c/starbucks

The data set contains information on the number of Starbucks in each city, combined with census data from ca.gov. From the same link, I also selected the dataset for country so that there is an option to check out the global distribution of Starbucks.

```
# Data set
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
Starbucks_Cali <- read.csv("Starbucks in California - City Stats.csv")

# First 10 rows of the data set
head(Starbucks_Cali,10)
```

```
##                                     ID            City              County
## 1    Monterey Park - Los Angeles County   Monterey Park  Los Angeles County
```

```
## 2         Milpitas - Santa Clara County        Milpitas  Santa Clara County
## 3         Rosemead - Los Angeles County        Rosemead  Los Angeles County
## 4        Hercules - Contra Costa County        Hercules Contra Costa County
## 5         Cerritos - Los Angeles County        Cerritos  Los Angeles County
## 6        Inglewood - Los Angeles County       Inglewood  Los Angeles County
## 7  Rowland Heights - Los Angeles County Rowland Heights  Los Angeles County
## 8          Daly City - San Mateo County       Daly City    San Mateo County
## 9           Walnut - Los Angeles County          Walnut  Los Angeles County
## 10          Carson - Los Angeles County          Carson  Los Angeles County
##    Number.of.Starbucks Starbucks.per.million.inhabitants
## 1                    1                                17
## 2                   10                               150
## 3                    3                                56
## 4                    2                                83
## 5                    4                                82
## 6                    3                                27
## 7                    3                                61
## 8                    5                                49
## 9                    2                                69
## 10                   8                                87
##    Starbucks.per.10.sq..mile Median.Age Median.Household.Income
## 1                       1.30       43.1                 $51,736
## 2                       7.36       36.1                 $94,589
## 3                       5.81       38.1                 $47,964
## 4                       3.22         39                 $94,493
## 5                       4.58         44                 $87,853
## 6                       3.31       33.4                 $44,021
## 7                       2.29       40.2                 $63,750
## 8                       6.53       38.3                 $75,399
## 9                       2.22       43.1                $101,358
## 10                      4.27       37.6                 $70,416
##    X2010.Population Percentage.of.white.population Land.Area..Sq..miles.
## 1             60269                             19                  7.67
## 2             66790                             21                 13.59
## 3             53764                             21                  5.16
## 4             24060                             22                  6.21
## 5             49041                             23                  8.73
## 6            109673                             23                  9.07
## 7             48993                             23                 13.08
## 8            101123                             24                  7.66
## 9             29172                             24                  8.99
## 10            91714                             24                 18.72
##    Number.of.Starbucks..Rank. Starbucks.per.million.inhabitants..Rank.
## 1                           1                                        1
## 2                         376                                      334
## 3                         190                                       96
## 4                         113                                      183
## 5                         247                                      177
## 6                         190                                       15
## 7                         190                                      112
## 8                         285                                       78
## 9                         113                                      133
## 10                        352                                      200
##    Starbucks.per.sq..mile..Rank. Median.Age..Rank.
```

```
## 1                                68                364
## 2                               380                205
## 3                               350                258
## 4                               210                277
## 5                               304                374
## 6                               216                132
## 7                               132                299
## 8                               363                263
## 9                               125                364
## 10                              281                245
##    Median.Household.Income..Rank. Percentage.of.white.population..Rank.
## 1                             125                                    1
## 2                             374                                    2
## 3                              93                                    3
## 4                             373                                    4
## 5                             358                                    5
## 6                              67                                    6
## 7                             222                                    7
## 8                             288                                    8
## 9                             391                                    9
## 10                            267                                   10
```

Week 10

What is the question that you are going to answer?

Does Starbucks everyday association and correlation to (the white) race and socioeconomic status offer any insightful reflections on consumer culture, social stratification, and branding? Otherwise, is it all just a matter of stereotyping and coincidence?

Why is this an important question?

With the implications of the common myth touted by Australian millionaire Tim Gurner that "When I was trying to buy my first home I wasn't buying smashed avocados for 19 bucks and four coffees at $4 each", his claims reflect the identification of Starbucks and its drinks to be commonly associated with affluence in its 'pricey' coffee to be significant to one's savings. According to public perception, Starbucks is a brand commonly associated with white Americans and westerners with leanings towards middle to upper class. Furthermore, as implied by Geraldine Rosa Henderson and Kathy Zhang in The Oxford Handbook of Consumption (2019), Starbucks is a white-aligned brand as a cited case of two men of colour being denied service due to implied ethnoracial prejudice, which also provides some reflections of the social racial dynamics, interactions, and inequality at that time in the US.

Which rows and columns of the dataset will be used to answer this question?

I will use the columns: ID, County, Number of Starbucks, Starbucks per million inhabitants, Median Age, Median Household Income 2010, Population Percentage of white population

As the last two rows of the data set are NA values, as of the writing of this rmd doc, I intend to take into account the rows of 1-431 and work from there.

Week 11

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'm personally looking at Percentage of white population, Median household income, Median age, Number of Starbucks, Starbucks per million inhabitants, and Starbucks per 10 sq mile. In particular, I want to pay special attention to Starbucks per million inhabitants as a way for me to relate the differing populations of the different geographical locations to give me the most apt proportion of Starbucks distribution.

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

I will repurpose the Shiny app I fiddled with in Week 8 to allow the user to plot their own graphs based on the variables of my data sets, so that they can also explore the different plots with me :)

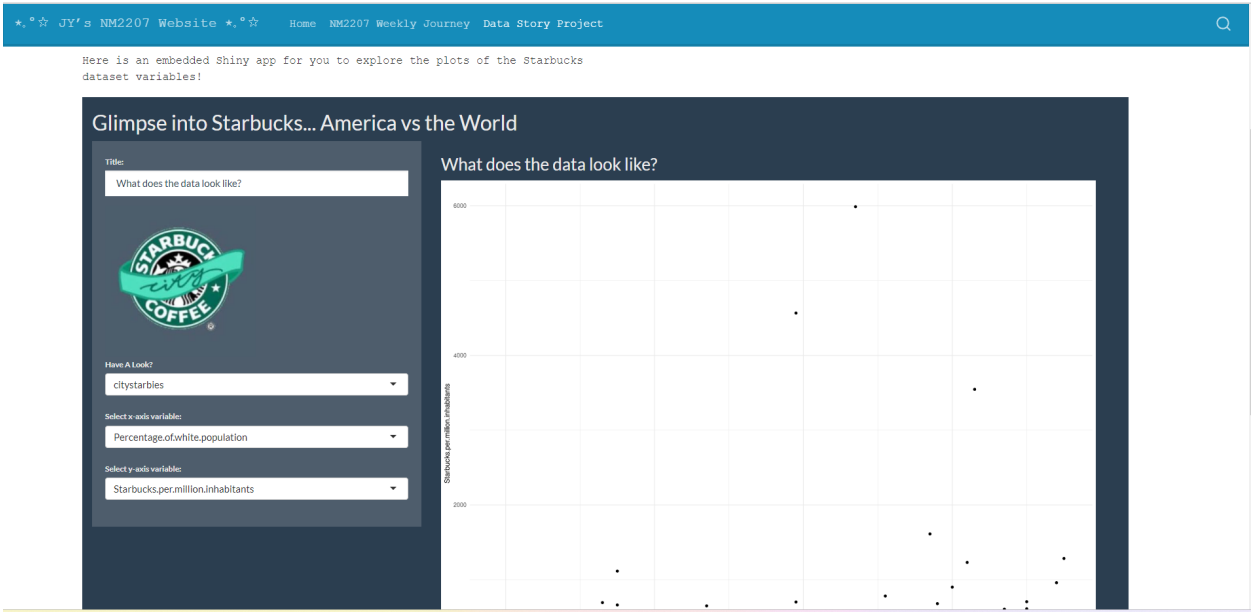

Figure 1: Here is a screenshot of my webpage where the interactivity will be.

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
ggplot	2,7
basics of Shiny app	8
Quarto website set-up	9 (I'm only listing this as 9 because the Canvas content for this is not under any specific week, but was accessed during week 9)

Topics	Weeks
Integration of Shiny dashboard with Quarto	9 (I'm only listing this as 9 because the Canvas content for this is not under any specific week, but was accessed during week 9)
conditional outputting different images and choosing different variables from a curated list on Shiny	
Adding a blog feature on my website to showcase my writing in 'posts'	

Week 12

Progress?

I added a blog page section to my website! I felt that my website was quite lacking in content and I could have added more to it. My website was barely scratching the bare minimum, which I felt was a problem and an area of improvement.

Hence, I self-learned with the wonderful help of Google to add a blog to my existing Quarto website. In the website, I added four entries of creatvie writing I did during my junior college days for my school's creative writing club! That way, more information about me can be included on my website along with the data story project to better personalise this website.

link referred: https://samanthacsik.github.io/posts/2022-10-24-quarto-blogs/

Week 13

What is the theme of your data story?

The theme of my data story is customer profiling and stereotypes in branding and target audience orientation.

Why is it important to address this question?

Studying stereotypes related to a company is crucial for branding and business marketing because it provides insights into how the company is perceived by its target audience, future potential customers and other stakeholders, and the wider market. By understanding these stereotypes, businesses can shape their brand image to meet customer expectations or identify possible disadvantages or potential controversies and problems detrimental to their image or interests. Studying this would hence allow businesses to address and mitigate negative stereotypes, fostering a positive brand image.

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

I selected data sources based on Starbucks, a prominent multinational chain of coffeehouses.

My hypothesis before exploring the data was that places with greater proportion of white, upper-middle class, and higher numbers of young-to-middle-aged populations will be more likely to have more Starbucks branches open in that vicinity to cater to such profiles that are stereotyped to be their most common customers. If such assumptions are correct, I would expect to see a positive correlation between the number and distribution of Starbucks branches and the variables of population that I have listed. This is due to two main reasons:

- Common expectations and recognition of Starbucks' target audience assume them to be those of relatively unrestrained spending habits, modern urbanised seekers of convenience, and those who tend to be American millennial and younger Generation X people as the stereotyped urban hipster overpriced coffee place (Quain, 2019).
- It is my personal assumption that Starbucks is also associated with white customers, with this assumption born out of the numerous racial controversies Starbucks had undergone against people of colour in recent years.

Hence, by exploring the data sets and comparing the demographics of cities in California, a notable American state, as well as noting the difference in Starbucks branch frequencies between different countries, these data sources will help me examine any possible trends (or lack thereof!) that can allow me to understand more about the correlation and relationship between Starbucks distribution and the profiles of populations in their vicinities.

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

From my own exploration of this plot visualisation app, I personally observe that there doesn't seem to be a strong correlation between the proportion of Starbucks distribution and most of my variables of concern. The scatter plots do seem quite sparsely separated in the generated graphs, and a generally weak positive correlation was observed in some of the variables of interest.

In the graph for age versus Starbucks distribution, the points cluster broadly and mostly between the early 20s to early 50s. At first glance, it is telling of the general demographic of Starbucks to tend towards those of the millennial generation and Generation X (with points bearing higher number of Starbucks per million inhabitants occurring between the mid-20s to late 40s). This also aligns with general ceonceptions and stereotypes of Starbucks as a millennial luxury or indulgence.

Interestingly, the plotted graph between "median household income" and "Starbucks per million inhabitants" shows no correlation unlike what I expected given the common perception that Starbucks is considered an 'everyday luxury' or 'treat'. Hence, I conclude that the real issue is definitely more nuanced than simple "higher SES = more Starbucks patronage".

Within the international, a scatter plot was also used, this time with Country as the y-variable and the numbers for Starbucks outlets as the x-variable out of ease of viewing. This data set and graph was included for the user to take a step out of the US to get an idea of global distributions in comparison to other countries, and draw possible conclusions or ideas based on ideas and associations to global race distributions and stereotypes in Starbucks outlet distribution.

Ultimately, the weak correlation tells me that the association between the frequency of Starbucks branches and outlets in a particular geographical location and factors such as population whiteness, age, and SES is weak. However, the observable positive correlation in the plots between "percentage of white population and Starbucks per million inhabitants" is apparent, which at least suggests that the stereotype that Starbucks leans towards a customer profile and target of satisfaction that may be prominently White does have some effect in the patronage demographics of Starbucks as well as meeting stereotyped assumptions on the target audience of Starbucks. Furthermore, my first concern was on Starbucks as a 'white-biased' brand, which is helpful in identifying at least some weak relationship between Starbucks, their brand and identity, and the racial profile of their customers. However, it is not a strong determining factor in deciding the frequency of distribution of Starbucks branches in a US Californa location.

The exact screenshot graphs of specific mapped variables can be found on my webpage with a similar write-up as this answer for my website visitors to view.

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

I planned to integrate my data project with the interactive segment of my website project, hence I had to customise my Shiny app (which its basics were covered in class) to adopt conditional displays and changes in options for users to explore the two data sets I have included with me. This customisation in how changes in the interface and available variables of interest is a new concept I had to learn and figure out to keep the interaction element in my Shiny app.

As for the personal element, I included my hobby for creative writing into the project by incorporating a blog-post display element into the webpage, with a categorised mix of old poetry and prose I wrote available for viewing. I also added some personal details in the webpage display backgrounds, as well as the different dataset logos in my Shiny app, to add further personalisation in my personal website.

*[Week 13 entry: 948 words]

References

Henderson, G. R., & Zhang, K. (2019). *Race and consumer inequality* (p. 397). Oxford, UK: Oxford University Press.

Quain, S. (2019). What Is the Target Market for Coffee? Small Business - Chron.com. Retrieved November 17, 2023, from https://smallbusiness.chron.com/target-market-coffee-71600.html