
title: "Project Topic and Relevance" author: "Mara Kage" date: "September 29, 2017" output: pdf_document:
default html_document: default

Project Topic and Relevance (~ 300 words)

Describe what interested you in the project you've chosen to work on this quarter. What is sociologically relevant about the topic you've chosen and why do you think this is an important area of investigation? One of the main social justice topics within the Asian community in the US is the effort to disaggregate the Asian racial categorization. Although the category represents 48 different countries and over 300 languages, many discourses, assumptions and policies are made based on an aggregated racial notion. In consequence, this racial categorization largely misrepresents the complexities of identity, experience and the diversity of life outcomes, particularly masking the struggles of many communities under this umbrella term. Pervasive narratives have been part of the experience of Asian American in the context of white supremacy racial domination since the arrival of the Chinese immigration mid 1850's. Consistent with the treatment of other communities of color facing discriminatory policies and practices, such as denial of citizenship and civil rights, a turning point was the Chinese Exclusion Act of 1882 blocking Chinese immigration in the US. At that time, pervasive narratives that Chinese were taken all the jobs and "yellow peril" were already part of the social consciousness and discourse shared by the white dominant population. Furthermore, these narratives extended to all other Asian identified communities. Another turning point in the Asian American history was the incarceration of 120,000 Japanese and Japanese descent in 1942-1945 to what is recognized today a result of race prejudice, war hysteria, and lack of political leadership. The main point to note is that at that time, most of the American population approved the incarceration largely based on those same pervasive narrative fears that have plagued the Asian communities since its immersion in the American society. At face value, when the Model Minority myth theory was introduced in the 1960s by sociologist William Petersen, referring to minority communities, appeared positive. It claimed that despite discriminatory challenges, these communities have thrive in education and socioeconomically by "just hard work". These narratives have furthered the misunderstanding of the Asian community, in addition to perpetuate practices and policies that justify white narratives to push against affirmative action, particularly against the Black community, and to further marginalize the struggles of the Asian community. I have chosen to study the diversity of the Asian community in King County in understanding that these pervasive narratives are still playing very loudly, parallel to the general ignorance of who is the Asian category and what is the reality of the members of this racial category members. Concurrently, I recognize my own lack of knowledge of my community, and as a result, in this study I seek to better understand the geopolitics, historical and socioeconomic variables.

Finding Data (~ 350 words) Describe how you found the data used for your project. This should include a discussion of the resources you used and the reasons you selected the dataset you did. If you changed datasets at any point in the quarter explain why you made the decision to switch and what effect(s) you think this had on your final analysis. What theoretical questions did you set out to answer with your data? How did these change as the quarter progressed? Looking at literature, I have heavily leaned from the beginning on CARE's study "The Hidden Academic Opportunity Gaps Among Asian Americans and Pacific Islanders: What Disaggregated Data Reveals in Washington State. This study heavily based on Prof. Takeuchi's early 2000's study on Disaggregated Data on the Asian Community in WA State. I have contacted Prof. Takeuchi inquiring about the data utilized in his study, and he kindly described they have relied part on Census, and part on non-public data, and suggested to use Census – American Fact Finder for my research. Due to the particular level of Asian subgroups I wanted to capture, and Prof. Takeuchi's work, under your guidance I have captured the same Asian subgroups described in his study: Cambodian, Chinese, Filipino, Hmong, Japanese, Korean, Laotian, Vietnamese, and Taiwanese. These groups are both accessible data wise and fairly captured the diversity of the community. I have based on datasets from Census 2010 as newer ACS surveys didn't display data on all the subgroups in question. Due to the reasons stated on the project topic and relevance question, I felt strongly throughout the research to utilize the same datasets throughout the research as my focus for this quantitative analysis have not changed. I have focused on the tracts in King County and curious to disaggregate the Asian population in combination with King County. With the mapping, I looked for understanding where each population lives, in which areas do they concentrate or disperse; in which areas do they overlap and which specific groups overlap. Are they in perceived wealthier neighborhoods, or working

class neighborhoods. Are they mostly located in historically neighborhoods of color or have integrated in other areas in the city and at what rate? Who are the subgroups represented in these areas? In answering these questions, I hoped to unfold my main question of the diversity and complexity of the Asian population.

Data Structure & Data Munging (~ 500 words)

Reflect on the original structure of your data. Explain the ways in which you had to clean, transform, or merge the data to do the following: • Initially organize your data • Create meaningful visualizations • Run statistical analyses What challenges did you encounter in any or all of the aforementioned activities? Did anything surprise you about this aspect of research? Explain. In all honesty, I don't think there was any part of it that was not very challenging and in this sense, I can't remember not being surprised. The first main challenge was that I didn't understand well what organization meant in computer terms and what the consequences of not organizing would lead to. I have unknowingly saved my work in different directories and drive and as a result, from Lab 8, I have painstakingly encountered several issues in having RStudio not recognize previous work and data as the subsequent labs built off from each other. I can only say that I have learned about data organization the hard way. My first epiphany, after the first week being utterly lost in the language and what was going on, came on the form of 5 main points: set working directory, install package (install.packages- Tidyverse), download package(library), import files (.read_csv), manipulate files(functions). I have essentially created all my data manipulation from one main "Asian Alone" dataset. The data came fairly sorted from Guided Search via American Fact Finder. (Guided search – people – origins – ancestry – geographies – Census Tract – WA –King - Race and Ethnicity – select all "Alone subgroup" desired categories). It already generated the population count data per tract and per Asian subgroup. Merging, selecting and filtering the columns and rows were the main functions to manipulate this table and tidy the data frame to answer specific questions and generate the first visualizations of plot, scatterplots and histograms via select(), merge(), filter() %>% piping specific rows and columns of the data that I wanted to combine the information, or omit in a new object. That was and still is extremely challenging.

Visualization was the most appealing aspect to me. I didn't play with ggplot2 as much as I hoped due to battling with basic technical skills constrains. But once the plots, histograms, charts, bar charts, and later, thanks to your coding, the maps came, it was amazing to see. Particularly the maps as it was a deep curiosity and something I've been longing to see the whole quarter. Thank you for the coding. The information that I wanted to see could not be clearly seen and understood with other types of visualizations. I had tried and the closest that conveyed the information was geom_point, facet_maps, facet_grids scatterplots of tract/Asian subgroup population. Statistical analysis was the latest part of the course, and in my journey, after the process of attempting to learn R and coding, I had little to focus on meaningful questions and analysis. In the last part of the course, I focused on the actual project and poster presentation, generating the visualizations necessary. My project mainly consisted in understanding more a theoretical reasoning and contextual understand of race categorization and the Asian American experience. I believe although it lacked statistical complexity, my efforts have been in the importance of descriptive analysis through research on race theory, Asian American history, Asian American identity and contemporary issues, in addition, elaborating what problem this project seek to address, in combination gave the depth it needed to be a meaningful research question.

Reproducible Research (~ 350 words)

As you understand it, what is the value of reproducibility for research? Include a description of the ways you employed principles of reproducibility in your own project. Which techniques seem most useful (i.e. what might you want to learn more about/incorporate into your research methods going forward)? As I understand, once the learning of the language takes place, to better store, organize, manipulate and share the research so it can be analyzed by others for the purpose of accountability and duplication. In terms of methods, I can see that it is particularly helpful and convenient if utilizing large amounts of data. Projects that require team contribution also benefits with this format as it is convenient to virtually share in places like GitHub as we have used. I see it at the moment as both a great point of accessibility, for reasons above, such as contributors living far apart and being to easily share and communicate the research, but I also, as a beginner, can see a great deal of accessibility barriers as it demands not only physical/financial access to your own computer, a computer that can perform these functions, but mostly because of the learning the skill aspect. But then I understand that statistical analysis might not be in the forefront of necessities for the general population. As I am not familiar with other methods of data visualization, manipulation and statistical analysis for comparison, it is hard to tell the real effectiveness and

benefits of this rather large amount of skill building that demands a lot of time and energy at minimum, but I am determined to learn it now as I recognize the effectiveness and future demands. Once again, I believe I have to relearn this class, practice, retain functions and be more comfortable with general coding skills as much of the frustrations of figuring the constant interruptions that derived from general coding, outside of R in itself. However, I can see that If I get to the point of being fluent in this language, it can be very helpful and I can imagine project contributions being particularly exciting, in addition to playing with aesthetics of visualizations.