NAACP Media Research

CS506 - Spark! Project

Ziyi Jiang, lynnj@bu.edu | Naiwen Xu, xnw1127@bu.edu | Yuhao He, hey1@bu.edu | Manyuan Lu,manyuan@bu.edu

I. Summary & Problem Statement

NAACP, the National Association for the Advancement of Colored People, was founded on February 12, 1909. It is the nation's foremost, largest, and most widely recognized civil rights organization. The vision of NAACP, the National Association for the Advancement of Colored People, is to ensure a society in which all individuals have equal rights without discrimination based on race. Our client, NAACP Boston seeks to understand the coverage of Boston Media in covering Black people and Black communities in Boston.

To accomplish this goal, we are going to define Black communities using census tract level data, collect data from mainly Boston Globe and other printed media sources, analyze the content and sentiment using Natural Language Processing tools. Our analysis will compare the difference between Black people related articles and non-Black people related ones. In addition, we will identify the frequent keywords (e.g. place, name, etc.) in the news about Black people from major Boston media sources. In doing so, we can hopefully give our client a better picture of the coverage of news about Black community and thus eliminate the race discrimination.

II. Methodology & Algorithms

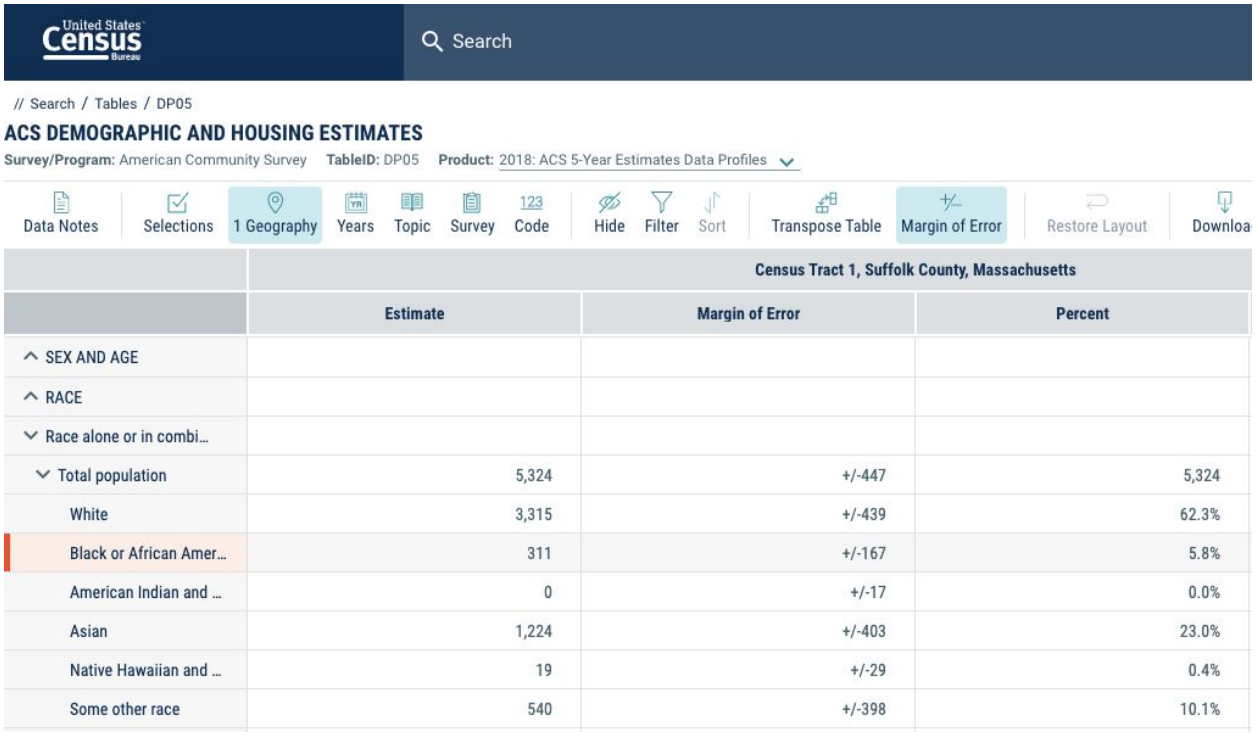
Step 1: Collecting data

As we plan to focus on analyzing the news from the angle of how Black communities are covered based on sub-neighborhoods, the first step is to define which neighborhoods can be considered as Black or non-Black communities. To do so we collected the list of sub-neighborhoods from the United States Geological Survey (see Table 1) and the demographics for each census tract from the United States Census Bureau (see Table 2). Then we used a look-up tool, Census Batch Geocoder which will return the corresponding census tract code if we enter a specific address within one neighborhood, to determine the sub-neighborhoods with the largest concentration of Black people.

Feature Name	ID	Class	County	State	Latitude	Longitude	Ele(ft)	Map	BGN Date	Entry Date
Charlestown	619308	Populated Place	Suffolk	MA	422240N	0710343W	43	Boston North	-	24-FEB-1974
Chelsea	612723	Populated Place	Suffolk	MA	422330N	0710158W	39	Boston North	-	24-FEB-1974
East Boston	612734	Populated Place	Suffolk	MA	422230N	0710221W	13	Boston North	-	24-FEB-1974
Franklin Park	612747	Populated Place	Suffolk	MA	422630N	0710058W	36	Boston North	-	24-FEB-1974
Orient Heights	612795	Populated Place	Suffolk	MA	422315N	0710013W	16	Boston North	-	24-FEB-1974
Revere	612810	Populated Place	Suffolk	MA	422430N	0710043W	10	Boston North	-	24-FEB-1974
Ashmont	612861	Populated Place	Suffolk	MA	421700N	0710408W	82	Boston South	-	24-FEB-1974
Back Bay	617071	Populated Place	Suffolk	MA	422100N	0710513W	13	Boston South	-	24-FEB-1974
Bay Village	1970982	Populated Place	Suffolk	MA	422102N	0710411W	16	Boston South	-	15-JAN-2003
Beacon Hill	619512	Populated Place	Suffolk	MA	422131N	0710404W	92	Boston South	-	27-AUG-2002
Boston	617565	Populated Place	Suffolk	MA	422130N	0710335W	46	Boston South	07-OCT-1931	24-FEB-1974
Cedar Grove	612891	Populated Place	Suffolk	MA	421650N	0710328W	23	Boston South	-	24-FEB-1974
Chinatown	606756	Populated Place	Suffolk	MA	Unknown	Unknown	-	Boston South	-	01-JAN-1992
City Point	612897	Populated Place	Suffolk	MA	422010N	0710143W	26	Boston South	-	24-FEB-1974
Dock Square	1877489	Populated Place	Suffolk	MA	422107N	0710326W	13	Boston South	-	17-AUG-2000

row(s) 1 - 15 of 73

Table 1. List of Sub-neighborhoods from the United States Geological Survey



	Census Tract 1, Suffolk County, Massachusetts		
	Estimate	Margin of Error	Percent
SEX AND AGE			
RACE			
Race alone or in combi…			
Total population	5,324	+/-447	5,324
White	3,315	+/-439	62.3%
Black or African Amer…	311	+/-167	5.8%
American Indian and …	0	+/-17	0.0%
Asian	1,224	+/-403	23.0%
Native Hawaiian and …	19	+/-29	0.4%
Some other race	540	+/-398	10.1%

Table 2. Demographics from the United States Census Bureau

After defining the sub-neighborhoods, the next step is to collect the articles between Year 2014 and Year 2018 from major Boston news outlets. Our team tried to scrape the articles from Boston Globe using the Wayback machine API for Scrapy, but we encountered some problems like anti-scraping techniques that will keep their websites being scraped. Thus, by getting permission, we decided to use the existing data collected by a team who took this project during the last semester.

Step 2: Filtering News

We initially planned to filter news by the addresses mentioned in the article. However, it turns out that it might be too specific and we can not ensure if the whole article is really about that place, or just a location it happened to mention. Then, we moved to a new mechanism – just look up the keywords of the sub-neighborhood it mentioned to determine whether it is about a black community. If an article doesn't mention any

neighborhood, we simply drop that piece of article. If more than three neighborhoods except Boston are mentioned in the article, we also drop it as we can consider the article as not focusing on talking about any neighborhood. Moreover, if the article only contains a single location ‘Boston’, we also remove it from our data because it may point to the Greater Boston Area instead of the neighborhood Boston.

To determine whether a sub-neighborhood is a black community, we calculated the black proportion based on the demographics we collected from the United States Census Bureau. If the black proportion of a sub-neighborhood is greater than 10%, we consider it as a Black neighborhood; otherwise, we labeled it as a non-Black neighborhood.

Step 3: Sentiment Analysis
skip

Step 4: Getting Significant Topics
skip

III. Findings & Observations

Data:

We’ve collected reliable demographics data to determine the Black and non-Black neighborhoods. We also collected all news between 2014 and 2018 from the Boston Globe, in which there are 192,400 pieces of news collected in total and they are classified by year from 2014 to 2018.

Coverage:

We used only neighborhood names to calculate the coverage. For Boston Globe, the coverage (number of articles about black neighborhoods/number of articles about all neighborhoods) is 9.5% and hasn't changed much during the past five years (see Table 3). As we can see, using only neighborhood names to determine the coverage of black people is not enough.

Dataset	# News about black nbh	# News about all nbh	# News	Coverage
globe2014	307	3203	18576	0.09584764
globe2015	416	4474	27000	0.09298167
globe2016	514	5476	35562	0.09386413
globe2017	643	6718	41876	0.09571301
globe2018	1112	11495	69386	0.09673771

Table 3. Statistics of Coverage

Sentiment:

skip

Topics:

skip

IV. Conclusions

skip