MAT 259: MySQL Knowledge Discovery in the SPL Database - The Interest in Race

Lara Floegel-Shetty

January 2021

1 Concept

For my project, I took interest in seeing if I could uncover a correlation between major events surrounding violence against people of color, or more specifically the black community, and the public's interest in the type of resources they check out of the library. I wanted to first take a general look at the check out counts of relevant materials to see if any trend appears before taking a closer look to see if any increases in the number of checkouts aligns with the dates of killings that made national headlines, protests, or transfer of powers between people in office. The question I ask is **Are the rising racial tensions during major controversial events reflected in the materials people check out from libraries?**

2 Process

As there is no general category in the Seattle Public Library that focuses solely on racism towards the black community, I had to take a bit of liberty in choosing what type of data I was looking for. I first looked at books that discussed topics similar to what I was interested in and found a list of popular books that centered around what I wanted. The source of the list can be found here.

I then looked at common words amongst the books in the list and used them as key words to filter out a general collection of books from the SPL database. This does introduce the potential for books irrelevant to my topic or books discussing discrimination towards non blacks, but it is a bit difficult to only pick out relevant books without using much more meticulous methods. I then used those words to filter through the SPL database to obtain a general count of book checkouts throughout the years.

3 Query 1: Overview and Analysis

Listing 1: Python example

```
SELECT
    YEAR(cout) AS years,
    COUNT(cout) AS counts
FROM
    spl_2016.outraw
WHERE
        ((LOWER(title) LIKE '%new_jim_crow%')
        OR '%how_to_be_an_antiracist%'
    OR '%so_you_want_to_talk_about_race%'
        OR '%black_lives_matter%'
    OR '%the_end_of_policing%'
    OR '%police_brutality%'
        OR '%white_fragility%'
    OR '%white_supremacy%'
    OR '%antiracism%')
        AND YEAR(cout) <= '2019'
GROUP BY years
ORDER BY years ASC
```

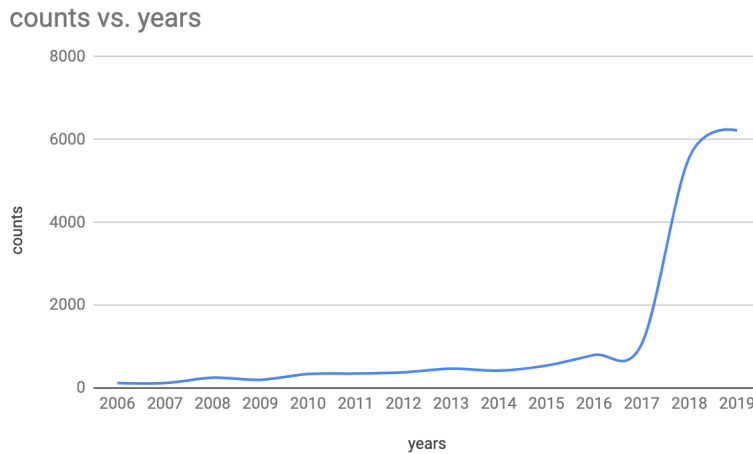| years	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
counts	121	118	248	198	337	348	378	466	418	541	797	1062	5576	6220

Figure 1: Number of books checked out annually from 2006 to 2019

Query 1 provides a broader overview of the number of relevant resources checked out annually from 2006 to 2019. From an initial on look, there is an overall increase in resources discussing racism towards blacks being checked

2

out each year. The first rise we see is around the time period Barack Obama was elected President of the United States, the first colored person to take that position in US history. Whether there is a direct correlation between him taking office and a small rise in the resources discussing issues surrounding the black community cannot be established with certainty, but it is interesting to see an overall rise in these topics throughout his years of presidency.

In the dates around police killings of black folkx that made national headlines, such as Eric Garner (2014), Michael Brown (2014), Tamir Rice (2014), Walter Scott (2015), Alton Sterling (2016), and Philando Castile (2016), there is no significant increase in checkouts of materials that are relevant to the issues that were discussed during those periods. When taking a closer look at the data, not included in this report, that is dated around the times of the deaths, there are some slight increases in the number of checkouts, but nothing that could be considered abnormal from the general rising interest.

In 2016, however, there is a abnormal rise as the number of increases rose from the average hundred to over two hundred checkouts. Looking at major events that occurred during this time, we can see the occurrence of the 2016 Presidential Election during which controversial candidate Donald Trump ran for office and issues regarding racism towards people of colors became a part of the mainstream conversation. Although it cannot be said to be directly correlated, there is reason to believe that the discussion of racism becoming a common talking point influenced the choices of materials people picked from the SPL, as reflected by the query above.

The most abnormal aspect of the query can be found beginning in the year 2018 where the checkout numbers increased by almost five times. To see the cause behind this, we would need to look at mainstream media during those years. In these years, topics such as white supremacy, police brutality, and black lives matter rapidly gained traction in society's discussion, the cause most likely can be attributed to the actions of Trump during his presidency, the effects of it examined more closely in the next section.

4   Query 2: A Closer Look

Listing 2: Python example

```
SELECT
    YEAR( cout )  AS  years ,
    MONTH( cout )  AS  months ,
    COUNT( cout )  AS  counts
FROM
    spl_2016 . outraw
WHERE
        ((LOWER( title )  LIKE  '%new_jim_crow%')
        OR '%how_to_be_an_antiracist%'
    OR '%so_you_want_to_talk_about_race%'
```

```
        OR '%black_lives_matter%'
    OR '%the_end_of_policing%'
    OR '%police_brutality%'
        OR '%white_fragility%'
    OR '%white_supremacy%'
    OR '%antiracism%')
        AND YEAR(cout) >= '2018'
        AND YEAR(cout) <= '2020'
GROUP BY years, months
ORDER BY years ASC
```

years	months	counts
2015	1	34
2015	2	40
2015	3	22
2015	4	33
2015	5	32
2015	6	22
2015	7	39
2015	8	41
2015	9	35
2015	10	40
2015	11	36
2015	12	36
2016	1	47
2016	2	43
2016	3	51
2016	4	47
2016	5	39
2016	6	49
2016	7	58
2016	8	64
2016	9	55
2016	10	55
2016	11	58
2016	12	57

Figure 2: Number of books checked out monthly from 2015 to 2016

Query 2 provides a closer examination of key events that have occurred over the past years. The SQL code is the same as the one used to compile query 1, except with the added grouping of months to allow a more detailed insight into any changes that may have occurred around controversial events.

In Figure 4, the data examines checkouts that occurred from the years 2015 to 2016, years that coincide with the change of presidency from Obama to Trump. Before Trump took office, the average number of resources check outed by the public is 30; however, after Trump takes office, the average increases to around 50. In July 2016, Philando Castile was killed by police, an event that sparked some protests. There is a consistency in the increased checkouts as it stays around above 55.

In Figure 5, the data looks at the counts from the years 2018 to 2020. In March 2018, Stephon Clark was shot several times, a police killing that sparked major protests. This increase in the public interest could be reflected in the data

years	months	counts
2018	3	216
2018	4	517
2018	5	490
2018	6	500
2018	7	553
2018	8	587
2018	9	201
2018	10	866
2018	11	761
2018	12	776
2019	1	678
2019	2	445
2019	3	475
2019	4	337
2019	5	328
2019	6	257
2019	7	247
2019	8	683
2019	9	679
2019	10	692
2019	11	621
2019	12	540
2020	1	565
2020	2	423
2020	3	276
2020	8	17
2020	9	113
2020	10	144
2020	11	194
2020	12	208

Figure 3: Number of books checked out monthly from 2018 to 2020

as the number increases from 216 in February to 517 in March and remaining in the 500s or even increasing in the following months. There is a drop in the checkouts in 2019 but then a sharp rise again but I was not able to find a specific reason for this occurrence.

In 2020, the Black Lives Matter Movement reached an all time high as protests were held nationwide and filled national news feed for weeks. These protests followed after the deaths of Breonna Taylor and George Floyd. Unfortunately, the data during these momentous times was lost due to libraries being closed due to COVID-19. But if we were to look at the previous patterns around other killings that sparked nationwide conversations, we can predict a high, perhaps even the largest, increase in public interest in learning more about issues surrounding the black community.

5 Conclusions

The data presented is far from perfect due to the number of outside influence and biases. It is difficult to only get resources that specifically pertain to racism towards black folkx so there is the potential that resources discussing racism towards other people of color communities could be among the compiled data.

Additionally, important information during the year of 2020, which some could consider the peak of the conversation about black issues, was lost due to COVID-19. But if we were to look at the data overall, there is positive affirmation that public interest in education about police brutality and like topics is increasing and major national events do influence a rise in checkout books that discuss these events.