Exploring the Popularity of Chinese-Related Media in Seattle Public Library

—

By `Zeyu Wang`

Introduction

This project aims to explore the popularity of Chinese-related media in the Seattle Public Library by analyzing checkout numbers for various item types, such as books, CDs, and music, from 2006 to 2022. Using SQL to query data, this study provides a detailed examination of trends and category percentages over time, as well as an analysis of the Dewey Decimal System classification of Chinese culture. Through data visualization, this project aims to gain a comprehensive understanding of the representation and popularity of Chinese-related media in the Seattle Public Library, which can provide valuable insights for cultural studies, library management, and Seattle Public Library users. The project will compare Dewey Classification Checkout Numbers for Chinese-related books in 2006 and 2022, which can give an idea about how Chinese culture is represented and how it has evolved in the Seattle Public Library.

Data Retrieval, Analysis, and Interpretation

Data Cleaning

The SQL query utilize an inner join to combine the subject and outraw tables and filter for subject entries containing the keywords "%china%" or "%chinese%". This query retrieves relevant data but also includes some duplicate records. The query employs the "distinct" keyword for deduplication to eliminate duplicates. The use of barcodes as a means of deduplication was not chosen because the project focuses on the popularity of Chinese-related media as indicated by checkout records rather than individual items. Additionally, the selection of 'subject.subject' was not included as some books may belong to multiple subjects, which would still result in duplicated records.

Comparison of Queries with and without 'subject.subject'

Query1:

```
SELECT distinct id, barcode,itemNumber, item type, title, cout, deweyClass
FROM spl_2016.outraw
inner JOIN spl_2016.subject ON outraw.bibNumber = subject.bibNumber
where subject.subject like '%chinese' or subject.subject like '%china%'
```

This query produces a base dataset that will be used as the foundation for further analysis and queries in the project.

Query2:

```
select
YEAR(cout) AS years,
count(IF(itemtype like '%bk%', 1, NULL)) AS 'Book',
count(IF(itemtype like '%cas%', 1, NULL)) AS 'Audio Tape',
count(IF(itemtype like '%cd%', 1, NULL)) AS 'CD & DVD',
count(IF(itemtype like '%vhs%', 1, NULL)) AS 'Video(VHS)'
from
(SELECT  distinct id, barcode,itemNumber, itemtype, title, cout, deweyClass
FROM spl_2016.outraw
inner JOIN spl_2016.subject ON outraw.bibNumber = subject.bibNumber
where subject.subject like '%chinese' or subject.subject like '%china%') as
A
where YEAR(cout) not in (2023)
GROUP BY YEAR(cout)
ORDER BY YEAR(cout) DESC;
```

Checkout Numbers for Various Item Types in Chinese Related Books(2006-2022)

years	Book	Audio Tape	CD & DVD	Video(VHS)
2022	16240	0	5522	0
2021	15352	0	6532	0
2020	6229	0	4581	0
2019	22205	0	15343	0
2018	17260	0	13653	0
2017	23424	0	22272	0
2016	24297	0	27735	0
2015	27000	0	26609	0
2014	30467	0	37097	0
2013	33409	0	45863	0
2012	29995	4	48259	2
2011	33289	18	59656	50
2010	36029	82	70969	241
2009	37986	258	78409	773
2008	39136	404	67005	1811
2007	29691	529	45225	2594
2006	30993	726	47693	4052

Data Visualization

Checkout Numbers for Various Item Types in Chinese Related Book(2006-2022)
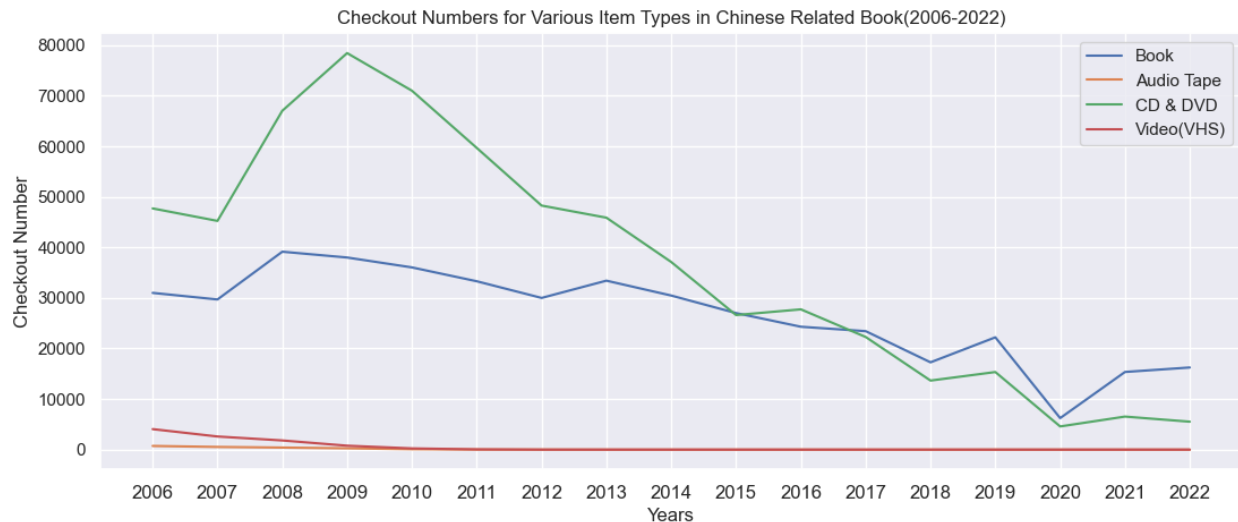
Analysis

The above table presents the checkout numbers for various types of Chinese-related media, including books, audio tapes, CDs & DVDs, and videos (VHS) at the Seattle Public Library from 2006 to 2022. It is evident that audio tapes and videos are not popular items in the library, which could be due to a lack of resources or interest among patrons. In contrast, books and CDs & DVDs have consistently increased in popularity over the years, with the highest number of checkouts recorded in 2022 for books and in 2016 for CDs & DVDs. However, the numbers for CDs & DVDs show a slight decrease in 2017-2022. It's worth mentioning that the sudden decline in 2020 for all item types is likely due to the impact of the COVID-19 pandemic.

Possible explanations for these trends include an increase in interest or population of Chinese culture in Seattle, changes in technology and media consumption habits, or increased availability and promotion of Chinese-related media at the library. The library's collection development policy and budget allocation may also play a role. Additionally, the decrease in the popularity of CDs & DVDs in recent years can be attributed to the growing popularity of streaming services and digital downloads.

Query 3:

```
select
YEAR(cout) AS years,
count(IF(deweyClass >= 000 and deweyClass < 100, 1, NULL)) AS 'Computer
science, information and general works',
count(IF(deweyClass >= 100 and deweyClass < 200, 1, NULL)) AS 'Philosophy
and psychology',
count(IF(deweyClass >= 200 and deweyClass < 300, 1, NULL)) AS 'Religion',
count(IF(deweyClass >= 300 and deweyClass < 400, 1, NULL)) AS 'Social
sciences',
count(IF(deweyClass >= 400 and deweyClass < 500, 1, NULL)) AS 'Language',
count(IF(deweyClass >= 500 and deweyClass < 600, 1, NULL)) AS 'Technology',
count(IF(deweyClass >= 600 and deweyClass < 700, 1, NULL)) AS
'Associations, organizations and museums',
count(IF(deweyClass >= 700 and deweyClass < 800, 1, NULL)) AS 'Arts and
recreation',
count(IF(deweyClass >= 800 and deweyClass < 900, 1, NULL)) AS 'Literature',
count(IF(deweyClass >= 900 and deweyClass < 1000, 1, NULL)) AS 'History and
geography'
from
(SELECT  distinct id,barcode,itemNumber,itemtype, title, cout, deweyClass
FROM spl_2016.outraw
inner JOIN spl_2016.subject ON outraw.bibNumber = subject.bibNumber
where subject.subject like '%chinese' or subject.subject like '%china%') as
A
where YEAR(cout) not in (2023) and deweyClass != ''
GROUP BY YEAR(cout)
ORDER BY YEAR(cout) DESC;
```

when querying for Dewey Decimal System classification of Chinese culture, the query uses a filter of "deweyClass >= 000 and deweyClass < 001" to retrieve relevant data. However, this filter may lead to inaccurate information because, in some cases, the deweyClass column may be empty. In SQL, the CHAR data type is used to store fixed-length character strings, and when a CHAR column is empty, it will still take up the same amount of storage space as if it were filled with a value. In this case, the value stored in the column will be a series of spaces. When using the CAST or CONVERT functions to convert an empty CHAR value to an integer, it will be considered as 0, which can cause the query to return incorrect results. To avoid this problem, the query needs to handle empty or null values in the deweyClass column properly.

Data Visualization

Dewey Classification Checkout Numbers in Chinese Related Books(2006-2022)
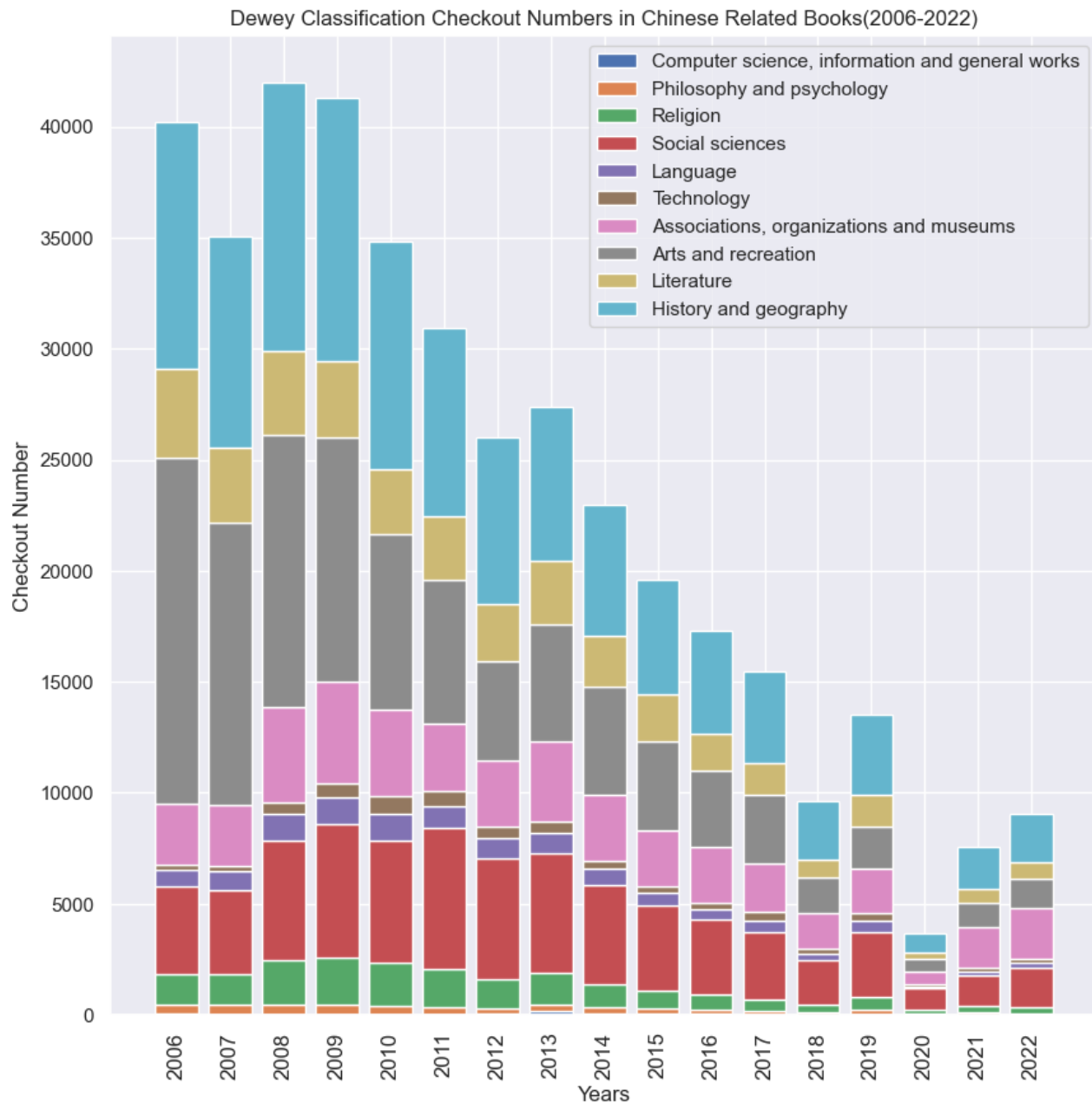
This plot presents the Dewey Classification checkout numbers for Chinese-related books at the Seattle Public Library from 2006 to 2022. The Dewey Classification categories include Computer science, information and general works, Philosophy and Psychology, Religion, Social sciences, Language, Technology, Associations, organizations and museums, Arts and recreation, Literature, and History and geography. The numbers are separated by years and Dewey Classification categories. The Social sciences and Literature categories have had the highest number of checkouts. The number of checkouts for the Literature category is consistently higher than for other types. Meanwhile, the number of checkouts for Technology and Associations, organizations, and museum classes has been increasing.

Likely, the trend of high checkouts in the Social sciences and Literature categories is due to the rich culture of China, which is reflected in the country's literature and social science studies. The literature of China is diverse and includes a wide range of genres, from poetry and fiction to non-fiction. Additionally, Chinese culture has a long history and rich tradition reflected in the country's social science studies. The increasing popularity of the Technology and Associations, organizations, and museums categories could be due to the growing interest in Chinese technology and society and the increased availability and promotion of Chinese-related media at the library. Additionally, as China is becoming more and more influential globally, the interest in learning about China's culture, society, and technology is increasing, which reflects in the increasing popularity of these Dewey Classification categories.

Query 4:

```sql
select
YEAR(cout) AS years,
count(IF(deweyClass = 0000, 1, NULL)) AS 'Not in Dewey Class',
count(IF(deweyClass > 000 and deweyClass < 100, 1, NULL)) AS 'Computer
science, information and general works',
count(IF(deweyClass >= 100 and deweyClass < 200, 1, NULL)) AS 'Philosophy
and psychology',
count(IF(deweyClass >= 200 and deweyClass < 300, 1, NULL)) AS 'Religion',
count(IF(deweyClass >= 300 and deweyClass < 400, 1, NULL)) AS 'Social
sciences',
count(IF(deweyClass >= 400 and deweyClass < 500, 1, NULL)) AS 'Language',
count(IF(deweyClass >= 500 and deweyClass < 600, 1, NULL)) AS 'Technology',
count(IF(deweyClass >= 600 and deweyClass < 700, 1, NULL)) AS
'Associations, organizations and museums',
count(IF(deweyClass >= 700 and deweyClass < 800, 1, NULL)) AS 'Arts and
recreation',
count(IF(deweyClass >= 800 and deweyClass < 900, 1, NULL)) AS 'Literature',
count(IF(deweyClass >= 900 and deweyClass < 1000, 1, NULL)) AS 'History and
geography'
from
(SELECT  distinct id,barcode,itemNumber,itemtype, title, cout, deweyClass
FROM spl_2016.outraw
inner JOIN spl_2016.subject ON outraw.bibNumber = subject.bibNumber
where subject.subject like '%chinese' or subject.subject like '%china%') as
A
where YEAR(cout) in (2006,2022)
GROUP BY YEAR(cout);
```

To understand the difference between books that are and are not classified using the Dewey Decimal System, I have created a new query that separates books by whether or not they have a Dewey Classification. To illustrate this comparison, I have chosen to focus on 2006 and 2022 as examples.

Data Visualization:

Comparing Dewey Classification Checkout Numbers in Chinese Related Books for 2006 and 2022
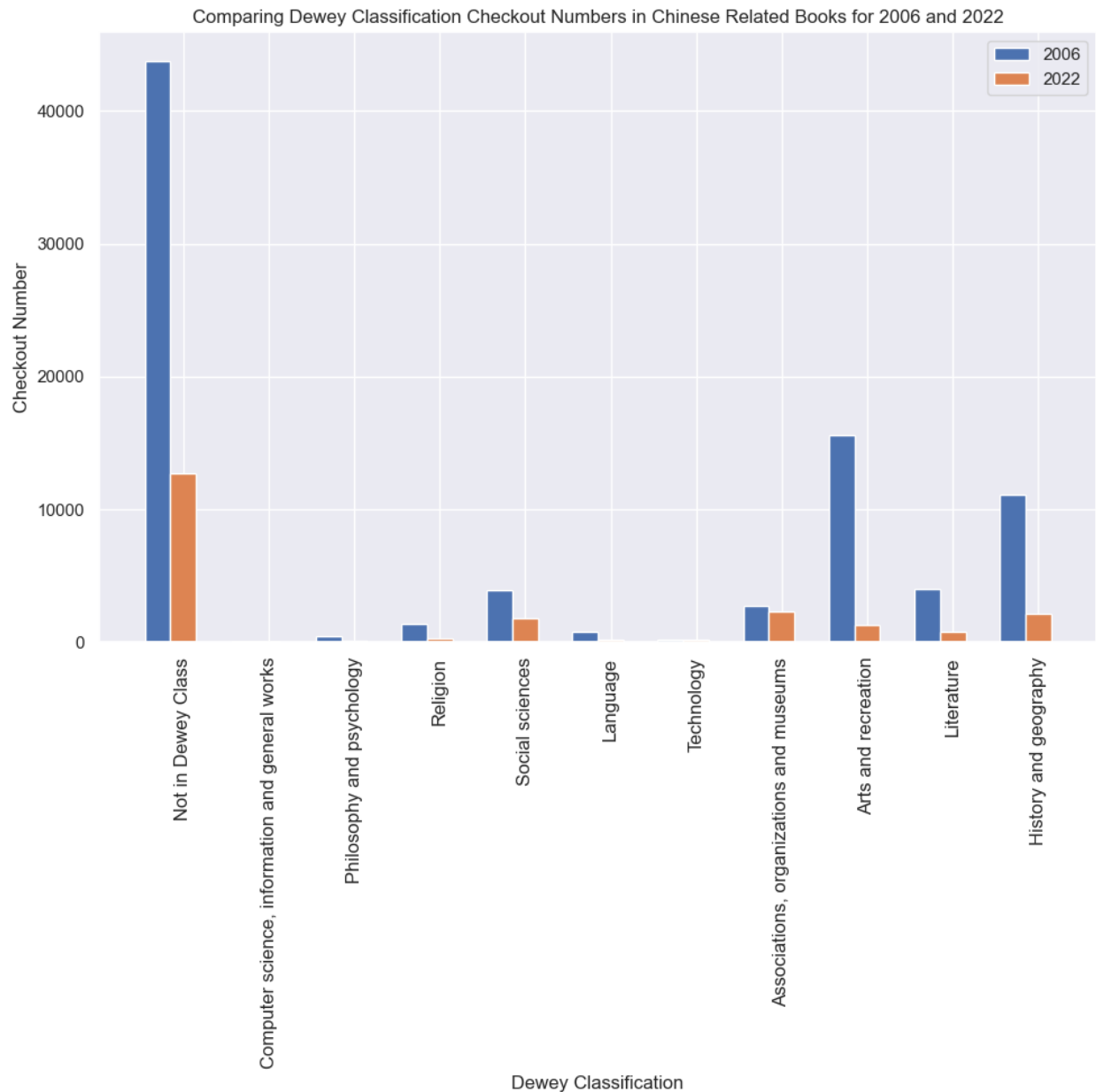
Books not in Dewey's class have higher checkout numbers than books in Dewey's class. This could be due to several reasons. First, Dewey classification is optional for all books,

and some books may not be classified using the Dewey Decimal System. This means that a significant portion of the library's collection may not be included in the Dewey Classification checkout numbers.

Second, Dewey Classification is imperfect and may only be suitable for some types of books and media, especially new and emerging formats such as e-books, audiobooks, and digital media. As a result, some patrons may prefer to browse the collection by author, title, or subject outside Dewey's class. Also, other factors, such as collection development policy and the popularity of certain authors, titles, or subjects, influence the checkout numbers of books outside Dewey's class.