

MAT259A Project 1

Zijian Wan

Description

In Dewey decimal classification, there are 10 topics (see Figure 1), which are subdivided into 100 subclasses for more detailed classification. It would be interesting to see the overall trend of the number of checkouts of each topic in the past 10 years (from 2012 to 2021). More interestingly, we might be able to identify anomalies, if there are any. The data was obtained from the database with MySQL and processed with Python for visualization and further analysis.

M259 Visualizing Information

George Legrady 2020 Winter

Dewey Decimal Classification

Ten topics each subdivided into 100 subclasses:

000-099 - Generalities
100-199 - Philosophy & Psychology
200-299 - Religion
300-399 - Social Science
400-499 - Language
500-599 - Natural Science & Mathematics
600-699 - Technology & Applied Sciences
700-799 - Arts
800-899 - Literature
900-999 - Geography & History

Figure 1. Dewey classification categories

MySQL query

The dataset was obtained from the database with the following MySQL query.

```
1. SELECT
2. YEAR(cout) AS years,
3. MONTH(cout) AS months,
4. COUNT(IF(deweyClass >= 0 AND deweyClass < 99,1,NULL)) AS "Generalities",
5. COUNT(IF(deweyClass >= 100 AND deweyClass < 199,1,NULL)) AS "Philosophy & Psychology",
6. COUNT(IF(deweyClass >= 200 AND deweyClass < 299,1,NULL)) AS "Religion",
7. COUNT(IF(deweyClass >= 300 AND deweyClass < 399,1,NULL)) AS "Social Science",
8. COUNT(IF(deweyClass >= 400 AND deweyClass < 499,1,NULL)) AS "Language",
9. COUNT(IF(deweyClass >= 500 AND deweyClass < 599,1,NULL)) AS "Natural Science & Mathematics",
10. COUNT(IF(deweyClass >= 600 AND deweyClass < 699,1,NULL)) AS "Technology & Applied Sciences",
11. COUNT(IF(deweyClass >= 700 AND deweyClass < 799,1,NULL)) AS "Arts",
12. COUNT(IF(deweyClass >= 800 AND deweyClass < 899,1,NULL)) AS "Literature",
13. COUNT(IF(deweyClass >= 900 AND deweyClass < 999,1,NULL)) AS "Geography & History"
14. FROM
15. spl_2016.outraw
16. WHERE
```

```
17. YEAR(cout) BETWEEN 2012 AND 2021
18. GROUP BY months , years
19. ORDER BY years , months;
20.
```

Further analysis

The obtained dataset was visualized using the Matplotlib package in Python (with codes provided below).

```
1. import copy
2. import pandas as pd
3. import seaborn as sns
4. import matplotlib.pyplot as plt
5. import matplotlib.ticker as mtick
6.
7.
8. # read data
9. cout_df = pd.read_csv("CheckOuts_Monthly_by_Topic_2012-2021.csv")
10. cats = cout_df.columns.to_list()[2:]
11.
12.
13. # year-month str
14. for rcdi in range(len(cout_df)):
15.     cout_df.loc[rcdi, 'ym'] = (str(cout_df.loc[rcdi, 'years']) + '-' +
16.                               str(cout_df.loc[rcdi, 'months']))
17. del rcdi
18.
19.
20. # set plot theme for aesthetics
21. sns.set_theme()
22. sns.set_context("talk")
23.
24.
25. # plot data by category
26. plt.figure(figsize=(30,20))
27. for cati in cats:
28.     plt.plot(cout_df['ym'], cout_df[cati], label=cati)
29. del cati
30. plt.xticks(fontsize=20, rotation=45)
31. plt.yticks(fontsize=20)
32. plt.xlabel('Time', fontsize=25)
33. plt.ylabel('Number of checkouts', fontsize=25)
34. plt.legend(loc='best', fontsize=20)
35. plt.title('Monthly checkouts by Dewey classification categories', fontsize=40)
36. ax = plt.gca()
37. ax.set_xticks(ax.get_xticks()[::6])
38. plt.show()
39.
```

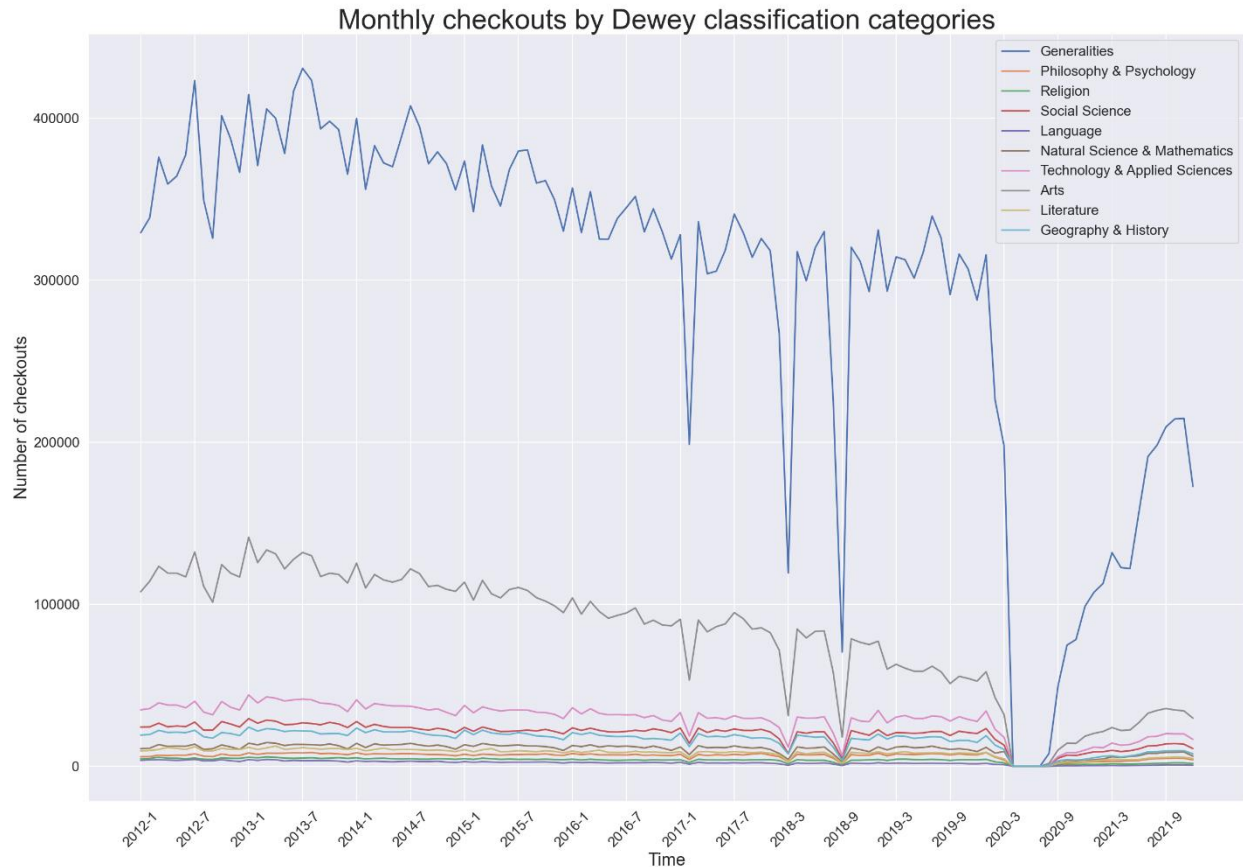


Figure 2. Monthly checkouts by Dewey classification categories

As shown in Figure 2, the monthly checkout number of generalities is far greater than that of others, leading to difficulty of comparative evaluation. Therefore, I tried to plot the checkout change instead. To calculate changes, I took the first record available for each category, i.e., the monthly checkout number of January 2012, as the baseline, and calculated the changes accordingly. The monthly checkout changes are visualized in Figure 3 (with Python codes provided below).

```

1. # compute change percentages of checkouts
2. # using the first record (2012-01) as the baseline
3. cout_change_df = copy.deepcopy(cout_df)
4. for cati in cats:
5.     for rcdi in range(len(cout_change_df)):
6.         if rcdi == 0: # the 1st record
7.             cout_change_df.loc[rcdi, cati] = 0
8.         else:
9.             cout_change_df.loc[rcdi, cati] = (
10.                 (cout_change_df.loc[rcdi, cati] - cout_df.loc[0, cati])
11.                 / cout_df.loc[0, cati])
12. del cati, rcdi
13.
14. # plot change percentages
15. plt.figure(figsize=(30,20))
16. for cati in cats:
17.     plt.plot(cout_change_df['ym'], cout_change_df[cat], label=cati)
18. del cati
19. plt.xticks(fontsize=20, rotation=45)
20. plt.yticks(fontsize=20)
21. plt.xlabel('Time', fontsize=25)

```

```

22. plt.ylabel('Change of checkouts (%)', fontsize=25)
23. plt.legend(loc='best', fontsize=20)
24. plt.title('Monthly checkout changes by Dewey classification categories',
25.           fontsize=40)
26. ax = plt.gca()
27. ax.set_xticks(ax.get_xticks()[::6])
28. ax.yaxis.set_major_formatter(mtick.PercentFormatter())
29. plt.show()
30.

```

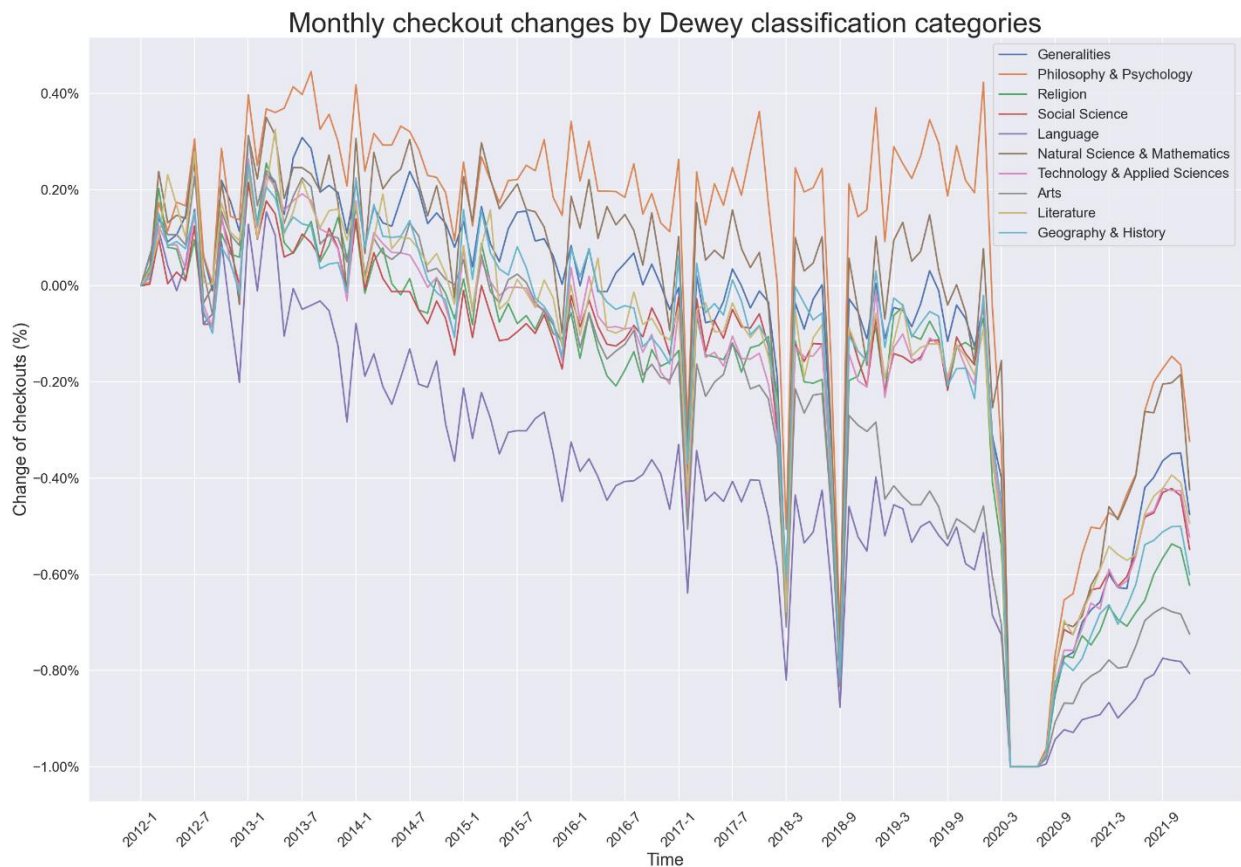


Figure 3. Monthly checkout changes by Dewey classification categories

As one can tell from Figure 3, there have been 4 major drops in monthly checkout numbers in the past 10 years. The biggest one happened in March 2020, due to COVID, which is a reason that we are all aware of. However, there are 3 other major drops, which are in February 2017, March 2018, and September 2018, respectively. The reasons behind those major drops are beyond my knowledge, which would be interesting if we can find out. Figure 3 demonstrates vividly how catastrophic and overwhelming the COVID pandemic is, as even after some extent of recovery, the current checkouts are not comparable to those in the past 10 years. In addition, the overall checkouts follow a downward trend in general (except one topic, philosophy and psychology). The reasons behind that could be twofold. People nowadays might be too busy to spare some dedicated time to reading. On the other hand, it might simply because that we can now read on all sorts of platforms, such as smart phones and tablets, thus leading to decreased demand for physical printed books.