

Usage Time for Each Dewey Class

Siming, Su

Concept Description

- I am interested to see the usage time of each Dewey Class in the library. By analyzing the estimated time people spend on each Dewey class, we could get a sense of which categories are easier to understand, which takes a long time to read, and so on. Based on this information, we could set up different deadlines of returning each Dewey Class. For example, Arts and Recreation usually take less time to borrow than others, the library can set up a shorter deadline of returning this category so that others may have more chance to borrow it. To dive deeper, I would like to make some visualizations to visualize how estimated usage time (day) is distributed among each Dewey class.

SQL Code with Explanation (part 1)

- Part1. `SELECT AVG(datediff(cin, cout)) as usageTime,
FLOOR(deweyClass/100) as class
FROM spl_2016.inraw inraw
WHERE YEAR(cout) >= 2010 AND YEAR(cout) <= 2020
AND deweyClass != ""
GROUP BY FLOOR(deweyClass/100)
ORDER BY usageTime DESC;`
- Explanation: By using this code, we get to see the average time people spend on books for each Dewey Class from 2010 to 2020.

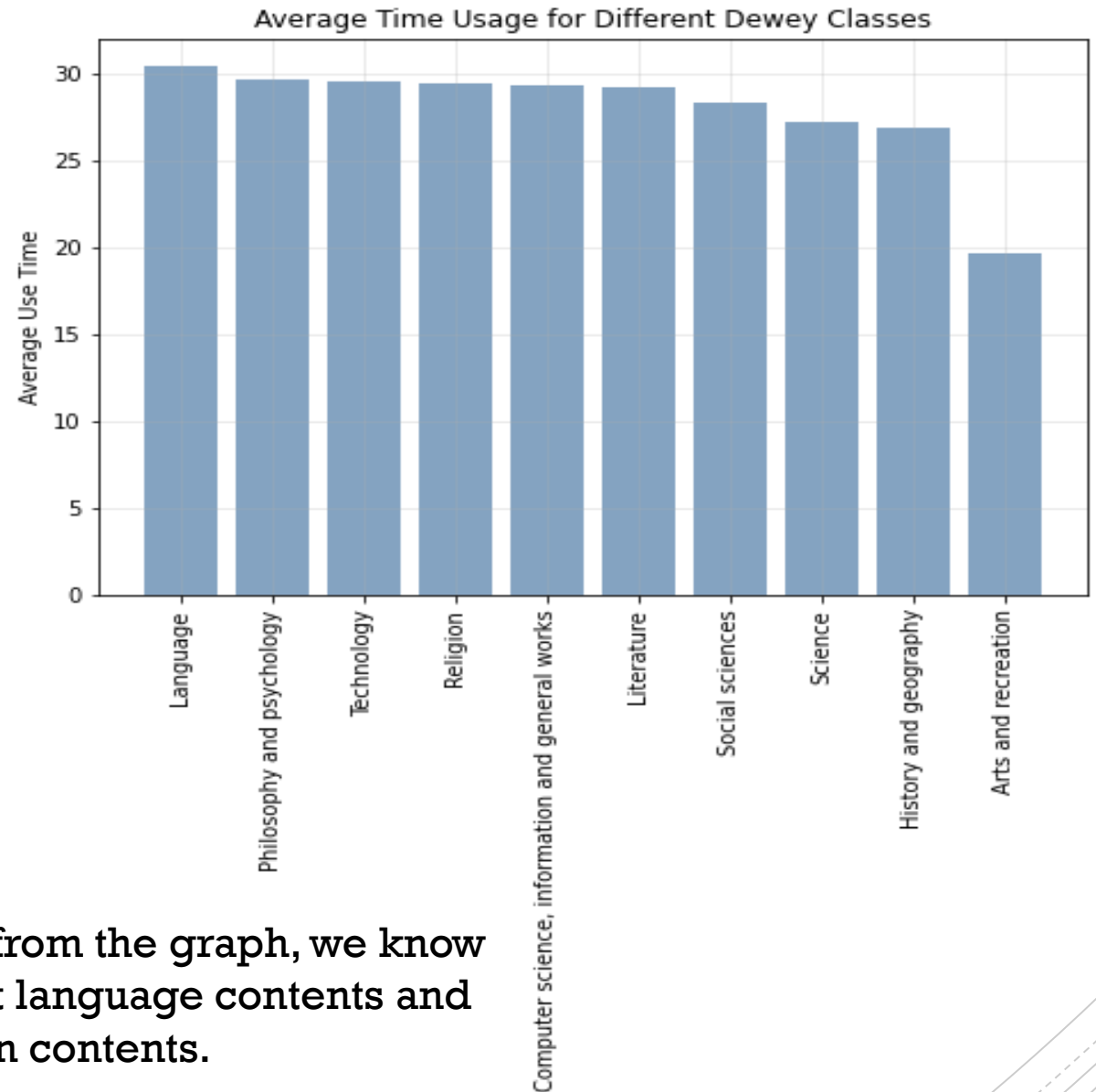
SQL Code with Explanation (part 2)

- **SELECT**
YEAR(cout) as eachYear,
AVG(TIMESTAMPDIFF(day, cout, cin)) as usageTime,
FLOOR(deweyClass/100) as class
FROM spl_2016.inraw
WHERE YEAR(cout) >= 2010 AND YEAR(cout) <= 2020
AND deweyClass != ""
GROUP BY YEAR(cout), FLOOR(deweyClass/100)
ORDER BY YEAR(cout) DESC,
AVG(TIMESTAMPDIFF(day, cout, cin)) DESC;
- **Explanation:** This code helps me to see the average spending time since checking out for each Dewey class for **every year** from 2016 to 2022

Data Results

			eachYear	usageTime	class			
usageTime	class			2020	111.7091	4		
				2020	91.4802	5		
				30.4305	4	2020	90.5173	1
				29.6174	1	2020	89.6343	6
						2020	86.5127	2
				29.5665	6	2020	83.2096	3
				29.4512	2	2020	81.3666	8
				29.262	0	2020	80.5719	9
						2020	72.8017	0
				29.1898	8	2020	66.8179	7
				28.3081	3	2019	28.2969	4
						2019	25.4688	1
				27.2226	5	2019	24.8855	6
				26.8214	9	2019	24.3547	8
				19.6426	7	2019	24.0789	2
						2019	23.8805	3
						2019	23.1563	5
						2019	22.2375	9
						2019	18.2427	0
						2019	17.6861	7

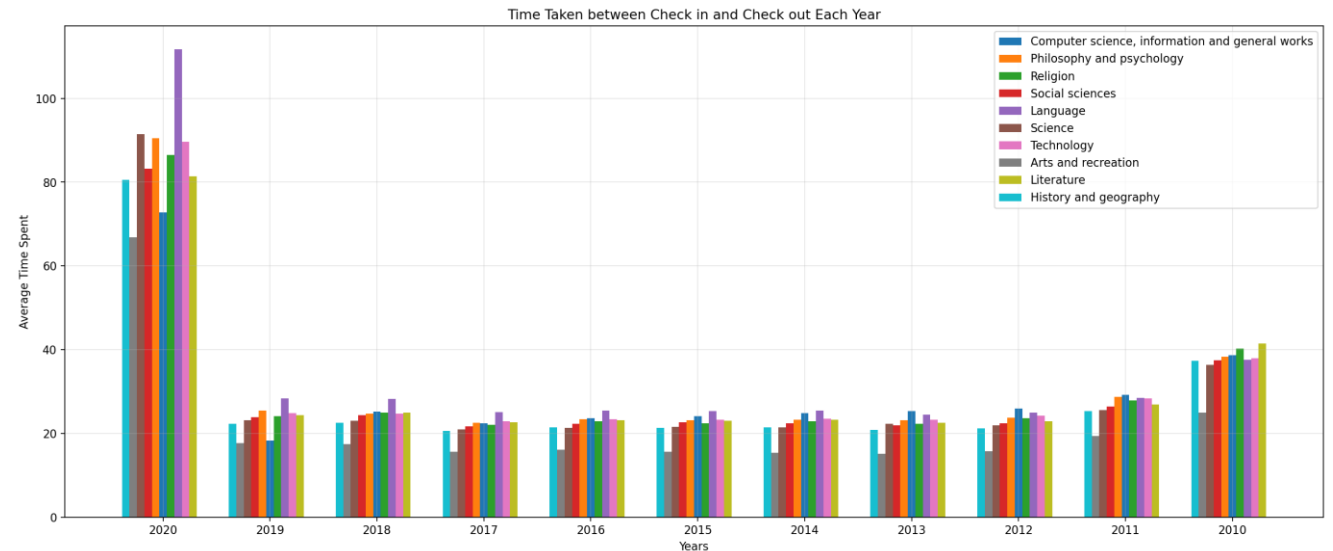
Data Visualization (by python)



This is quite interesting to analyze, because from the graph, we know that people are taking more time to interpret language contents and spend the fewest time on Arts and Recreation contents.

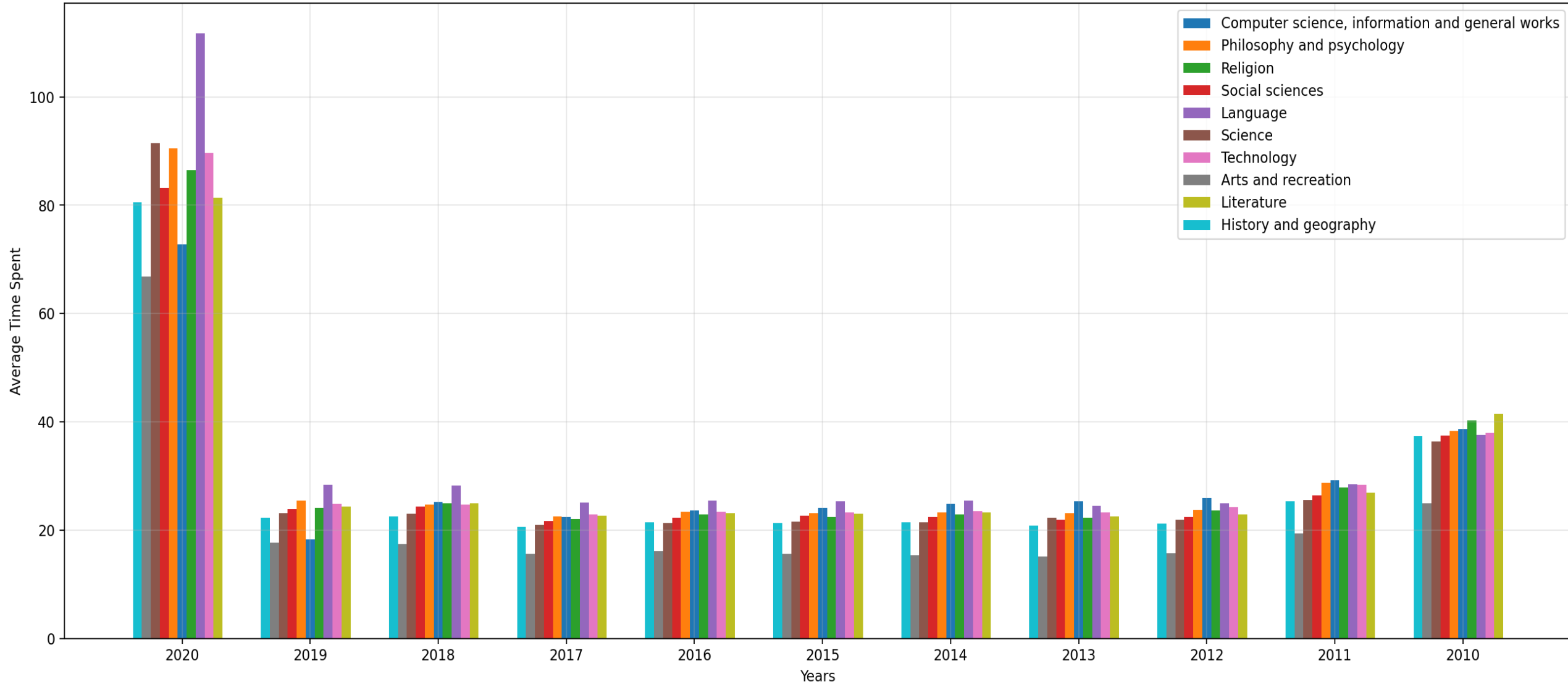
Bigger picture on next page

Data Visualization (by python)



Notice that due to Covid in 2020, we have a much larger time gap between checking in and checking out, it is probably because electronic copies have much longer return deadline than hard copies. Furthermore, notice that the structure of estimated time spent in each Dewey class is almost the same across years. That means, even though the total time spent in each year is different, but the ratio of spending on each Dewey class remains stable.

Time Taken between Check in and Check out Each Year



Conclusion

- By visualization above, we know that the ratio time spent in each Dewey Class is stable, so it is probably a good idea to shorten the deadline for borrowing the art and recreation books so that it could increase the flow of borrowing and returning in the library.
- There remains an issue why such a ratio structure presents, my theory is that it is probably because language books take a longer time to learn. For art books, people tend to just read faster to get the central idea. However, more investigations are needed.

Reference

- Data is courtesy of the Seattle Public Library database (2016)
- Floor function idea is taken from https://www.mat.ucsb.edu/~g.legrady/academic/courses/20w259/MySQL_student.pdf