

Mat 259, winter 2015

## Assignment 1: SQL Query

Keehong Youn

14 Jan 2015

Loan period, sorted by Pages and Dewey number.

Q1: Does longer books get checked out longer?

Q2: Is there any specific type of subject that gets checked out longer?

It was found that spl2 does not have proper pages data. So pages data was took from spl1.extra table, **assuming that bib number of spl1 and spl2 is same for each item.**

**select**

```
    floor(convert(pages, unsigned integer) / 50) * 50 as pages,
    floor(deweyClass / 100) * 100 as dewey,
    sum(case
        when cout is not null and cin is not null
            and timestampdiff(day, date(cout), date(cin)) >= 1
            and timestampdiff(day, date(cout), date(cin)) <= 365
            then timestampdiff(day, date(cout), date(cin))
        else 0
    end) as totaltime,
    sum(case
        when cout is not null and cin is not null
            and timestampdiff(day, date(cout), date(cin)) >= 1
            and timestampdiff(day, date(cout), date(cin)) <= 365
            then 1
        else 0
    end) as totalnum
```

**from**

```
spl1.extras,
spl2.inraw
```

**where**

```
extras.bib = inraw.bibNumber
and pages > 100
```

**group by**

```
    floor(convert(pages, unsigned integer) / 50) * 50,
    floor(deweyClass / 100) * 100
```

**limit** 1000;

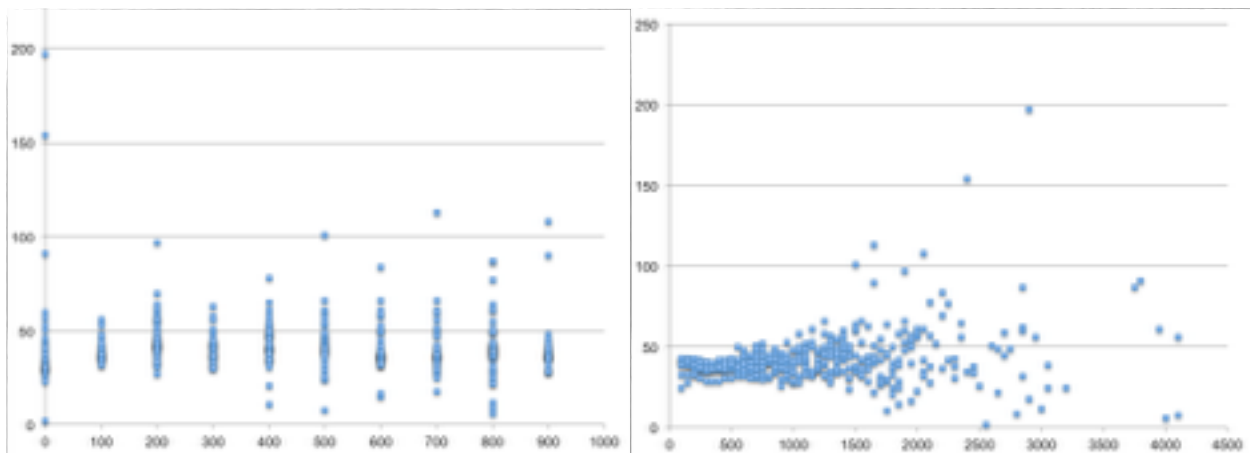
As seen in the query above pages were ranged by 50 and Dewey number by 100. And those went into 1st and 2nd column. **For 3rd and 4th column the sum of**

**total loan time and counted number of loan was calculated so that later the average loan time for each pair of total pages and Dewey number could be obtained (with MS Excel).**

pages	dewey	totaltime	totalnum	avgloantime
100	0	30738675	1252493	25
100	100	1499824	35274	43
100	200	2278314	54053	42
100	300	4952796	118063	42

First four columns are the originally generated data from SQL query. Fifth column is the one calculated after. Further processing will be using 1st, 2nd, and 5th column.

It took quite long (more than 10 min) to get all the data because the query barely excludes any item except ones that don't have Dewey number. After the query average time was calculated in MS Excel. Below is a brief overview on the result.



Dewey Number vs Loan Time

Total Pages vs Loan Time

When first thinking about this query the motivation was that the result might help running a library. (ex: They could buy more copies on items that tend to be loaned for longer time) However, though a precise statistic analysis could be done, it seems that there aren't very strong characteristic in the result. **This might be due to the possible behavior of the users that: They anyway return their items after maximum loan period whether or not they have finished the item.** However, it is still considerable that despite the maximum loan period for books are 21 days, average loan time tends to be larger than that.