MAT 259 –Winter 2016

Project 1: Data Query

Jameson Woodfin
January 15, 2016
Visualizing Data
Prof. Legrady

Mathematicians' Interests

Concept:

Math is very widely despised subject. Even people who work in scientific fields that require the use of math often cringe at the thought of having to evaluate a nasty looking integral. What type of masochist would subject themselves to reading a book on mathematics when there are so many other topics in the library?
I plan to find out.
Let's study the behavior of people who checkout books on mathematics and see what else they checkout. Are they shut-ins that only focus on pure math? Are they more open and interest themselves in a wide range of subjects? Let's explore.

MySQL Query:

I wrote two queries: the first one helps me describe the behavior of people checking out math books, the other to help me compare them to "normal" people.

```
SELECT
   (CASE WHEN (deweyClass>=0 AND deweyClass <100) THEN '000'
   WHEN (deweyClass>=100 AND deweyClass <200) THEN '100'
   WHEN (deweyClass>=200 AND deweyClass <300) THEN '200'
   WHEN (deweyClass>=300 AND deweyClass <400) THEN '300'
   WHEN (deweyClass>=400 AND deweyClass <500) THEN '400'
   WHEN (deweyClass>=510 AND deweyClass <520) THEN '510 (Math)'
   WHEN (deweyClass>=500 AND deweyClass <510) OR (deweyClass>=520 AND deweyClass <600)
THEN '500 not 510'
   WHEN (deweyClass>=600 AND deweyClass <700) THEN '600'
   WHEN (deweyClass>=700 AND deweyClass <800) THEN '700'
   WHEN (deweyClass>=800 AND deweyClass <900) THEN '800'
   WHEN (deweyClass>=900 AND deweyClass <1000) THEN '900'
   END) AS dClassSec,
   COUNT(deweyClass) AS freq
FROM
         (SELECT /* This is the INNER JOINed table with times from the table below */
    _rawXmlDataCheckOuts.deweyClass, _rawXmlDataCheckOuts.checkOut
FROM
   spl3._rawXmlDataCheckOuts
      INNER JOIN /*Find checkouts that happened at the same time as a math book checkout. */
   (SELECT /* This table contains math books */
       _rawXmlDataCheckOuts.deweyClass, _rawXmlDataCheckOuts.checkOut
```

```
   FROM
      spl3._rawXmlDataCheckOuts
   WHERE
      deweyClass >= 510 AND deweyClass < 520
   /*LIMIT 100*/) AS mathBookCheckOut ON _rawXmlDataCheckOuts.checkOut  =
mathBookCheckOut.checkOut
/*LIMIT 100*/) AS mathBookTable
WHERE
         length(deweyClass)>0
GROUP BY dClassSec
ORDER BY dClassSec
;


SELECT
   (CASE WHEN (deweyClass>=0 AND deweyClass <100) THEN '000'
   WHEN (deweyClass>=100 AND deweyClass <200) THEN '100'
   WHEN (deweyClass>=200 AND deweyClass <300) THEN '200'
   WHEN (deweyClass>=300 AND deweyClass <400) THEN '300'
   WHEN (deweyClass>=400 AND deweyClass <500) THEN '400'
   WHEN (deweyClass>=510 AND deweyClass <520) THEN '510 (Math)'
   WHEN (deweyClass>=500 AND deweyClass <510) OR (deweyClass>=520 AND deweyClass <600)
THEN '500 not 510'
   WHEN (deweyClass>=600 AND deweyClass <700) THEN '600'
   WHEN (deweyClass>=700 AND deweyClass <800) THEN '700'
   WHEN (deweyClass>=800 AND deweyClass <900) THEN '800'
   WHEN (deweyClass>=900 AND deweyClass <1000) THEN '900'
   END) AS dClassSec,
   COUNT(deweyClass) AS freq
FROM
         spl3._rawXmlDataCheckOuts
WHERE
         length(deweyClass)>0
GROUP BY dClassSec
ORDER BY dClassSec
;
```

Query Explanation:

The first query should really be split up into three separate queries that create tables but the mat259 user didn't have the permissions to create new or even temporary tables so I had to nest queries to get the results I wanted. I'll explain them in the order in which I wrote them:

The innermost nested SELECT statement looks through the raw checkout table and returns rows containing checkouts with Dewey class between 510 and 520, this being the Dewey class interval for books on mathematics. With this I knew the times at which math books were checked out.

The second SELECT statement picks the Dewey class and checkout times out the raw checkout table and inner joins that with the table explained above with respect to checkout time. This query returns a table containing a list of checkouts that occurred at the same exact time that a math book was checked out. I gave this table the alias "mathBookTable".

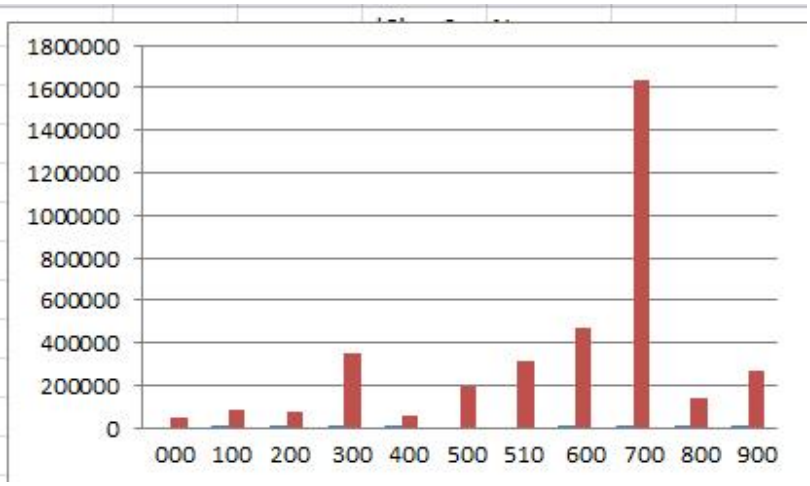
The outermost SELECT statement (the first one) takes the "mathBookTable" table and counts the number of checkouts within a certain Dewey class interval. I chose to group Dewey classes by intervals of 100. I also ignored entries with no Dewey class (i.e. fiction books) and I looked at the 510 (i.e. math) Dewey class specifically. With the results from this query we can see the frequency in which people who checkout math books checkout books in other Dewey classes.

The second query was my "control group" query. It does the exact same thing at the first query except it acts on the raw checkouts table instead of the "mathBookTable" so that I can get an idea of "normal" interests. With these results I can compare with the results from the mathematician's interests and see how they compare to "normal" people.

Results:

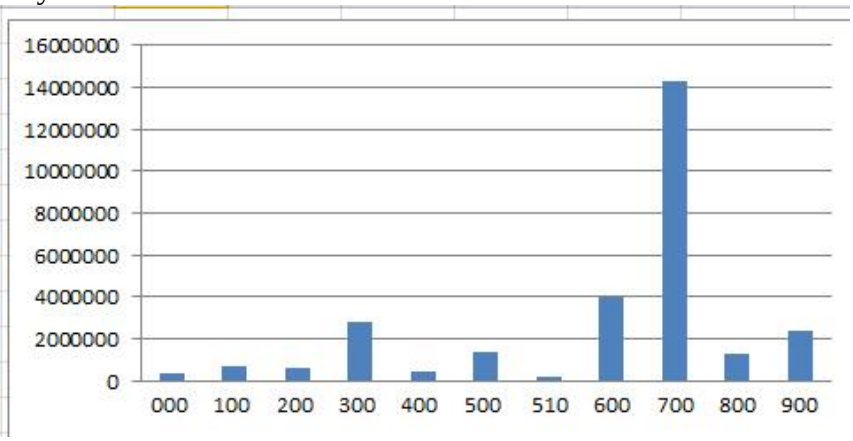
From the first query:

dClassSec	freq
0	44586
100	84767
200	72644
300	350101
400	60560
500 not 510	197691
510 (Math)	316435
600	470526
700	1635850
800	142118
900	272054



From the second query:

dClassSec	freq
0	361222
100	709076
200	630399
300	2813926
400	422155
500 not 510	1341161
510 (Math)	152451
600	3955707
700	14321687
800	1264278
900	2384498



Processing Time:

The processing time for the first query was 118.202 seconds.
The processing time for the second query was 0.031 seconds.

Result Analysis:

A first glance at the results might make you think that mathematicians have a particular interest in the arts (700), technology (600), social sciences (300), and history and geography (900). But, when you compare these to our control group results you see that interest in these subjects is "normal". The two graphs look very much the same proportionately, except for one Dewey class, the mathematics Dewey class.

The results from the second query confirm our initial intuition about the majority of people not taking much interest in math. Looking at the first query results and comparing the two we find that really the only difference between mathematicians and "normal" people is just that they like math. Otherwise, they take interest in the same topics as people who hate math. In other words, mathematicians are "normal" people.