

Carlos Baez
MAT 259
W2015
Legrady

Project 1: Data Query

Transportation scientist are often limited in the data that they can use in their analyses. For instance, in order to forecast the aggregate demand for travel to a destination, the best source of information is usually direct information of a population followed by travel surveys and historical travel data.

Direct information is mostly not an option due to the expenses associated with the operation and also, any such operation would raise questions about privacy. Surveys, although not as pricey or invasive as direct information, nonetheless, are beyond the financial means of most researches given the costs of survey operations. Historical data is often the most inexpensive of the three, however, the use of this information is limited due to its historical nature (it takes time to process data after it's collected) and also, often many sources of data are held tightly by the organizations that collect data. Thus, in order to develop estimates of demand for travel, transportation scientist often look for sources of data.

Here we are attempting to measure the amount of books about Texas (in general) that are related to the topics of Geography and History. Then we look at the months in which the books are checked out in hopes of seeing a seasonal pattern for planning of travel.

Query:

```
SELECT
    SUM(CASE WHEN MONTH(cout) = '01' THEN 1 ELSE 0 END) as Jan,
    SUM(CASE WHEN MONTH(cout) = '02' THEN 1 ELSE 0 END) as Feb,
    SUM(CASE WHEN MONTH(cout) = '03' THEN 1 ELSE 0 END) as Mar,
    SUM(CASE WHEN MONTH(cout) = '04' THEN 1 ELSE 0 END) as Apr,
    SUM(CASE WHEN MONTH(cout) = '05' THEN 1 ELSE 0 END) as May,
    SUM(CASE WHEN MONTH(cout) = '06' THEN 1 ELSE 0 END) as Jun,
    SUM(CASE WHEN MONTH(cout) = '07' THEN 1 ELSE 0 END) as Jul,
    SUM(CASE WHEN MONTH(cout) = '08' THEN 1 ELSE 0 END) as Aug,
    SUM(CASE WHEN MONTH(cout) = '09' THEN 1 ELSE 0 END) as Sep,
    SUM(CASE WHEN MONTH(cout) = '10' THEN 1 ELSE 0 END) as Oct,
    SUM(CASE WHEN MONTH(cout) = '11' THEN 1 ELSE 0 END) as Nov,
    SUM(CASE WHEN MONTH(cout) = '12' THEN 1 ELSE 0 END) as Deb
FROM
    spl2.inraw
WHERE
    (title LIKE "%Texas%") AND
    (((deweyClass >= 910) AND (deweyClass <920)) OR
    ((deweyClass >= 970) AND (deweyClass< 980)))
```

Analysis

There seems to be a couple of peak times Jan (263), Mar(258), Aug(176) with the colder periods of the year having a higher count than the warmer periods of the year. To further verify this model, we would need to isolate the History category and attempt to tease out the potential effects of students writing reports. Also, it would be interesting to compare multiple states with different climates to test whether

the travel planning pattern is stable or whether it varies depending on the weather (in terms of that at both places or of the origin/destination alone).