

Matthias Wagner
MAT 259, Winter 2015
Prof. George Legrady

Project 1: Data Query

Motivation:

While doing research on different categories of the Dewey Decimal System, I found a small list of sections, which specifically deal with women and their relation to a given field.

(<http://ndscs.nodak.libguides.com/content.php?pid=50434&sid=681502>) I chose to investigate on the significance by analyzing the percentage of specifically female works as part of their larger parent branch, and how this changes over time.

To this end, I list the total number of checked out items per month of both the female and general categories, and the female category percentage. The list extends from January 2006 to December 2013.

Query:

```
SELECT
    MONTH(cout),
    YEAR(cout),
    SUM(CASE
        WHEN
            # categories that relate specifically to women
            (deweyclass = 323.34
                OR deweyclass = 346.0134
                OR deweyclass = 809.933522
                OR deweyclass = 324.3
                OR deweyclass = 305.4
                OR deweyclass = 324.623
                OR deweyclass = 355.4082
                OR deweyclass = 323.34
                # civil and human rights
                # legal status
                # literature - criticism
                # political organizations
                # social groups/aspects
                # suffrage
                # in combat
                # rights
            )
        THEN
            1
        ELSE 0
    END) AS Women,
    SUM(CASE
        WHEN
            # general parent categories
            (deweyclass >= 323 AND deweyClass < 324
                OR deweyclass >= 346 AND deweyClass < 347
                OR deweyclass >= 809 AND deweyClass < 810
                OR deweyclass >= 324 AND deweyClass < 325
                OR deweyclass >= 305 AND deweyClass < 306
                OR deweyclass >= 355 AND deweyClass < 355
                # civil rights
                # legal status
                # literature - criticism
                # the political process
                # social groups
                # military science
            )
        THEN
            1
        ELSE 0
    END) AS General,
    # calculate percentage
    (SUM(CASE
        WHEN
            (deweyclass = 323.34
                OR deweyclass = 346.0134
                OR deweyclass = 809.933522
                OR deweyclass = 324.3
                OR deweyclass = 305.4
                OR deweyclass = 324.623
```

```

                                OR deweyclass = 355.4082
                                OR deweyclass = 323.34
                            )
                        THEN
                            1
                        ELSE 0
                    END) /
SUM(CASE
    WHEN
        (deweyclass >= 323 AND deweyClass < 324
        OR deweyclass >= 346 AND deweyClass < 347
        OR deweyclass >= 809 AND deweyClass < 810
        OR deweyclass >= 324 AND deweyClass < 325
        OR deweyclass >= 305 AND deweyClass < 306
        OR deweyclass >= 355 AND deweyClass < 355
        )
        THEN
            1
        ELSE 0
    END)) * 100 AS '%'
FROM
    spl2.inraw
WHERE
    DATE(cout) > '2005-12-31' AND DATE(cout) < '2014-01-01'
    AND ( (deweyclass >= 304 AND deweyclass < 356)
    OR (deweyClass >= 809 AND deweyclass < 810) )

GROUP BY YEAR(cout), MONTH(cout)
ORDER BY YEAR(cout), MONTH(cout)

```

Analysis:

Initially, I wanted to include *itemtypes* in my query to investigate a possible correlation between female areas of interest and the type of medium. This proved difficult due to the fact that apart from a few VHS tapes, books were the only item in the women categories. I decided not to distinguish between types of media. Also, as discussed in class, date-related data, though with entries going back as far as 1970, is very inconsistent until 2006, which I subsequently chose as the starting point.

A few things jump out when looking at the output of the final query. The percentages of women items make up between 0.3 – 3 % of their respective larger dewey classifications. This value declines quite steadily over the 8 year period. For 2012 and 2013, all but 2 months fall below 1%. For the same period, the total items checked out in the general categories is also generally lower than in prior years. But whereas the general total never seems to go below about ¼ of it's highest value (max: 4892, 01/2009, min: 1195, 12/2013), the women total fluctuates much stronger (max: 105, 03/2006, min: 6, 04/2013). Within the given range, there do not seem to be any discernible patterns, as far as yearly repeating cycles go. While the total number of all checked out items in this list declines, items in women categories seem to do so more dramatically.