Kitty Currier
MAT 259
13 March 2014

Project 4: Exploratory data visualization to investigate bar code anomalies

Project Description:

For the final project, I return to the subject of my first project: bar code anomalies. The *barcode* field only appears in two tables of the *spl2* database: *inraw* and *outraw*. Tables derived from these (e.g. *activity*, *callnum*, *collection*) include the item number but not the bar code as columns, suggesting that the bar code is redundant information. In the first project I posed queries to investigate whether, in fact, the bar code and item number are both unique to individual items and found that this is not the case: some item numbers are associated with more than one bar code and vice versa.

While the majority of *itemNumbers* are associated with only one bar code, a handful (26,900, or <1%) have more than one bar code over their checkout history at the SPL:

5 barcodes	4 barcodes	3 barcodes	2 barcodes	1 barcode
1	25	719	26156	3453793

My final project is a visualization to investigate the characteristics of these anomalous items. Are there temporal patterns corresponding to the date(s) at which items acquire new bar codes? Does this depend on item type or item location (Central library branch or other branch)? To investigate this I will visualize items' patterns of bar code behavior over time, highlighting the time at which each item acquires a new barcode.

Query:

```
SELECT itemNumber, numBarcodes, barcode, o AS cout, collcode, itemtype
FROM
  (SELECT itemNumber, COUNT(DISTINCT barcode) AS numBarcodes, barcode,
      collcode, itemtype
    FROM inraw
      WHERE YEAR(cout) >= 2006 && YEAR(cout) <= 2013
    GROUP BY itemNumber) AS a1,
  activity
      WHERE activity.item = a1.itemNumber && numBarcodes >=3 && YEAR(o) >=
      2006 && YEAR(o) <=2013
```

Explanation:

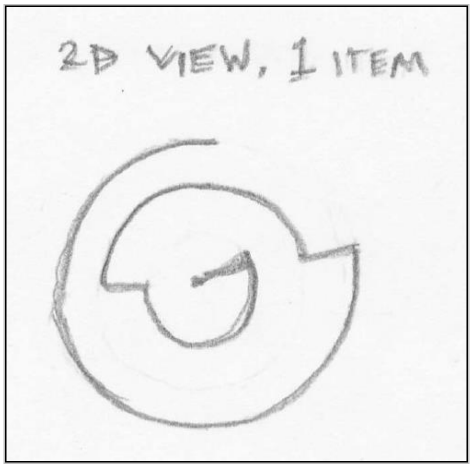
The data is limited to the years 2006–2013. A sub-query on **inraw** is first executed to count the number of bar codes (*numBarcodes*) associated with each item number and to grab values for the *title*, *collcode* and *itemtype* fields. Then, for each item number, **activity** is queried to return the check-out history of those items associated with more than one bar code. File is included as **query_output.csv**.

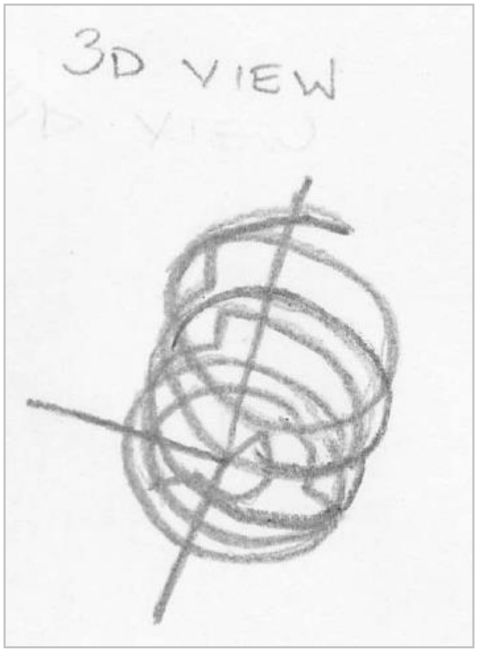
I manipulated the data further in Excel to produce a CSV file (**DAYS360_3_or_more.csv**) with the following fields for each item number:

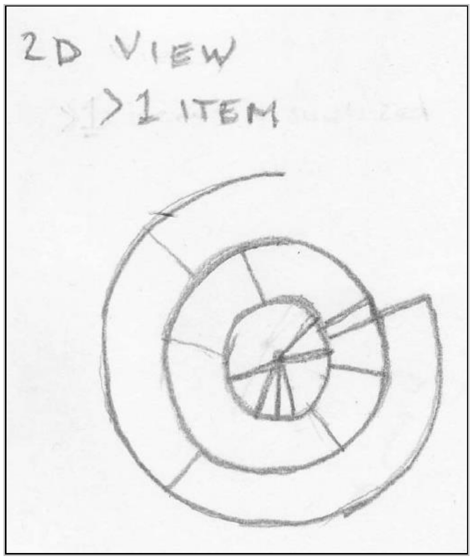
- *days360_start* – number of days, assuming a 360-day year, from 1/1/2006. This represents the item's first check-out during the sampling period (2006 – 2013).

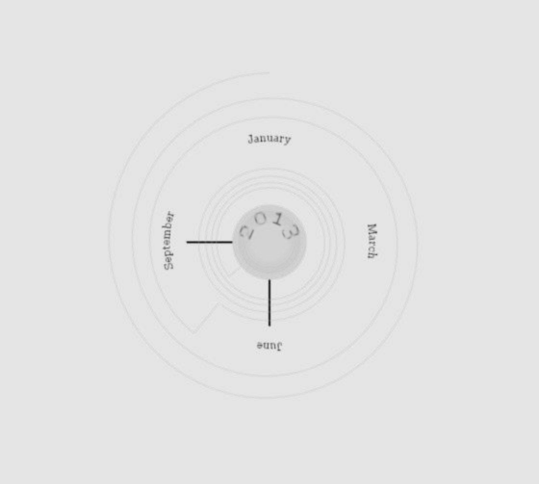
- *days360_chA* – date (in number of days as described above) of the item's first change to a new bar code.
- *days360_chB* – date of the item's second change to a new bar code.
- *days360_chC* – date of the item's third change to a new bar code (items that did not have a third change are coded -999).
- *days360_chD* – date of the item's fourth change to a new bar code (items that did not have a fourth change are coded -999).
- *isLocatedCentral* – if the item's home location is the Central library branch, 1, else 0.
- *type* – cd, dvd, bk, kit, or other.

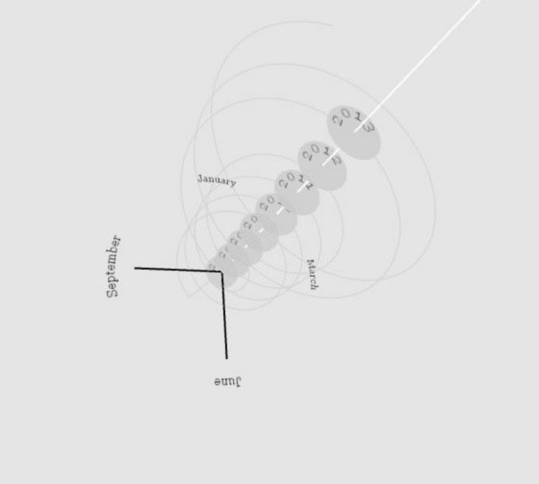
Doodles

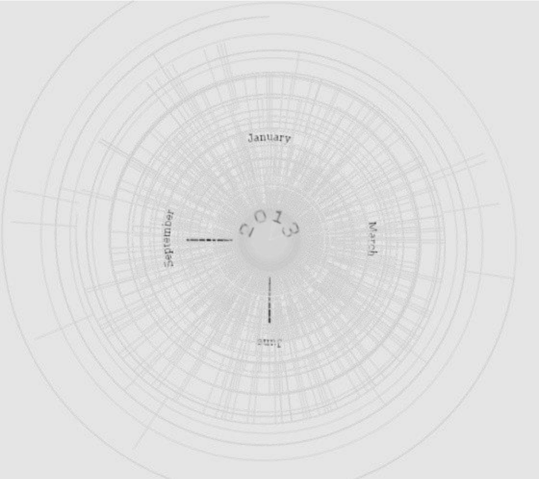


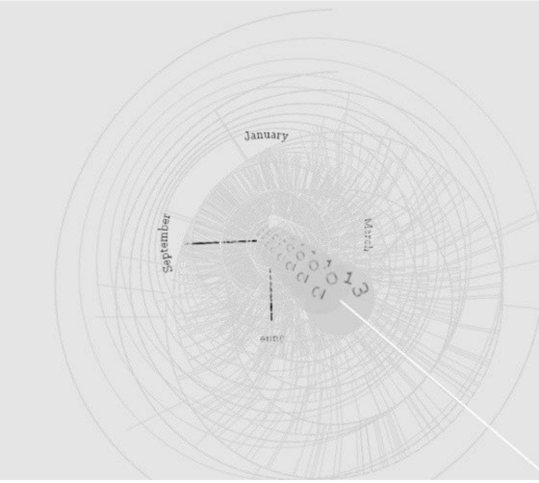


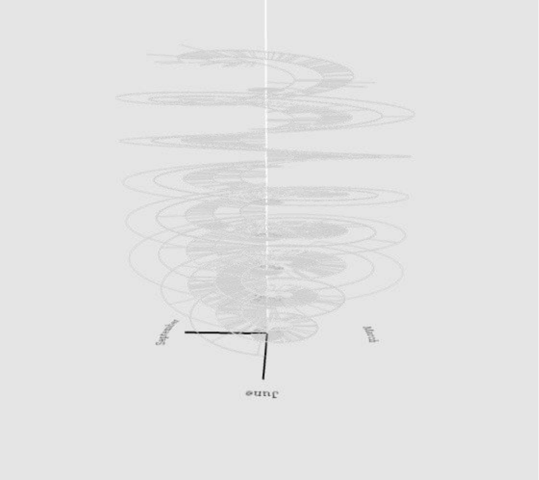












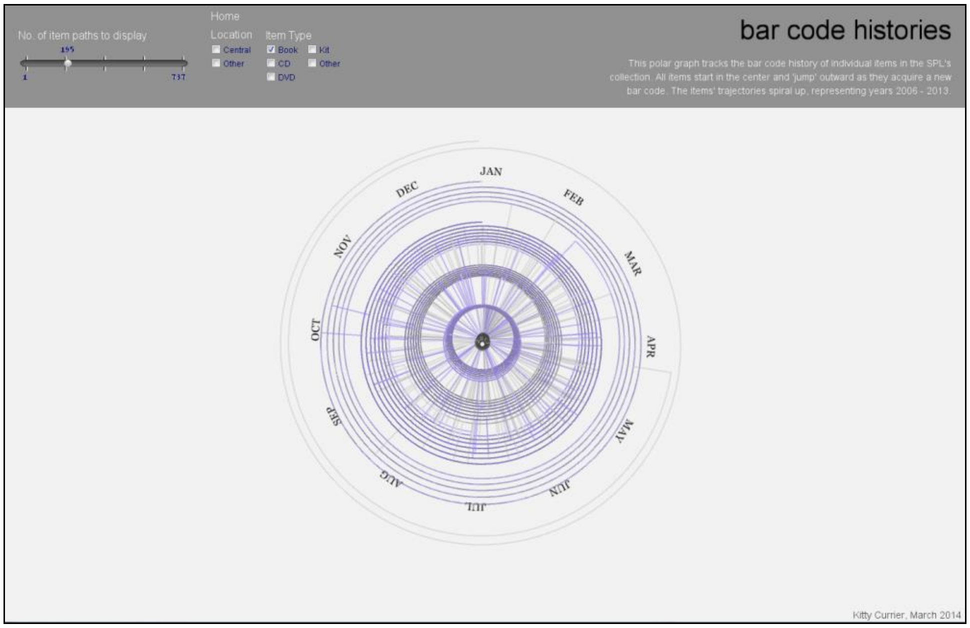
The general form of the visualization is a spiral, with each revolution representing one year. I find that visualizing more than 200 items takes a toll on the display performance, decreasing the frame rate so that motion looks jerky. I'll probably limit the number of items visualized to 200 or fewer, sampled from the pool of items that are associated with three or more bar codes.

Project timeline:

- Tuesday, 4 March: query
- Thursday, 6 March: query results
- Tuesday, 11 March: preliminary visualization
- Thursday, 13 March: present final project

Process

I first tried keeping the colors to a minimum and only selected one to highlight the item type or item's home location, based on user input:



bar code histories

This polar graph tracks the bar code history of individual items in the SPL's collection. All items start in the center and 'jump' outward as they acquire a new bar code. The items' trajectories spiral up, representing years 2006 - 2013.

No. of item paths to display: 195 (1 – 737)

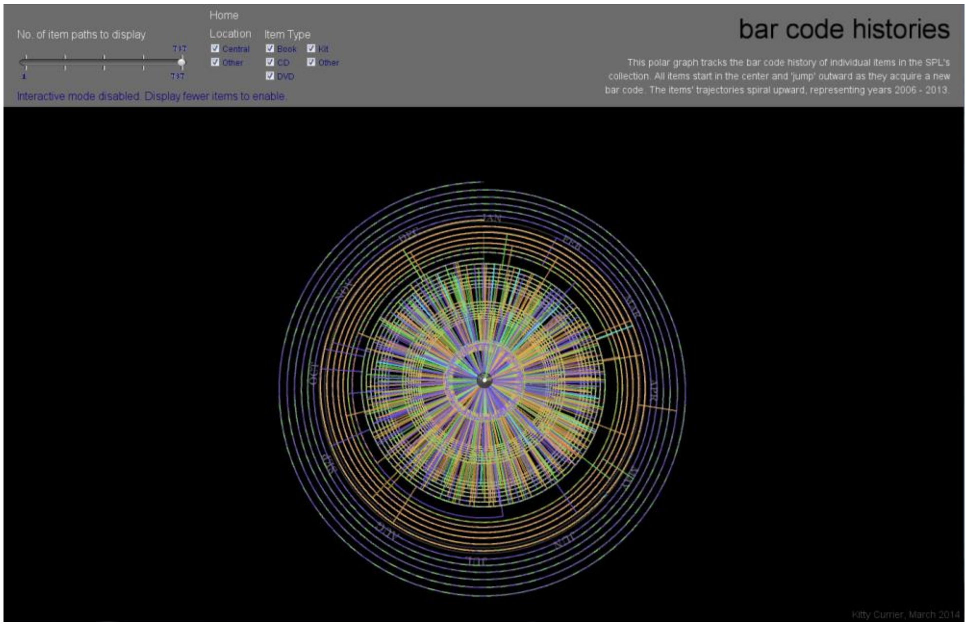
Home Location: □ Central □ Other

Item Type: ☑ Book □ Kit □ CD □ Other □ DVD

Kitty Currier, March 2014

Results & Analysis

Then I expanded and tried symbolizing different types with different colors:



bar code histories

This polar graph tracks the bar code history of individual items in the SPL's collection. All items start in the center and 'jump' outward as they acquire a new bar code. The items' trajectories spiral upward, representing years 2006 - 2013.

No. of item paths to display: 737 (1 – 737)

Home Location: ☑ Central ☑ Other

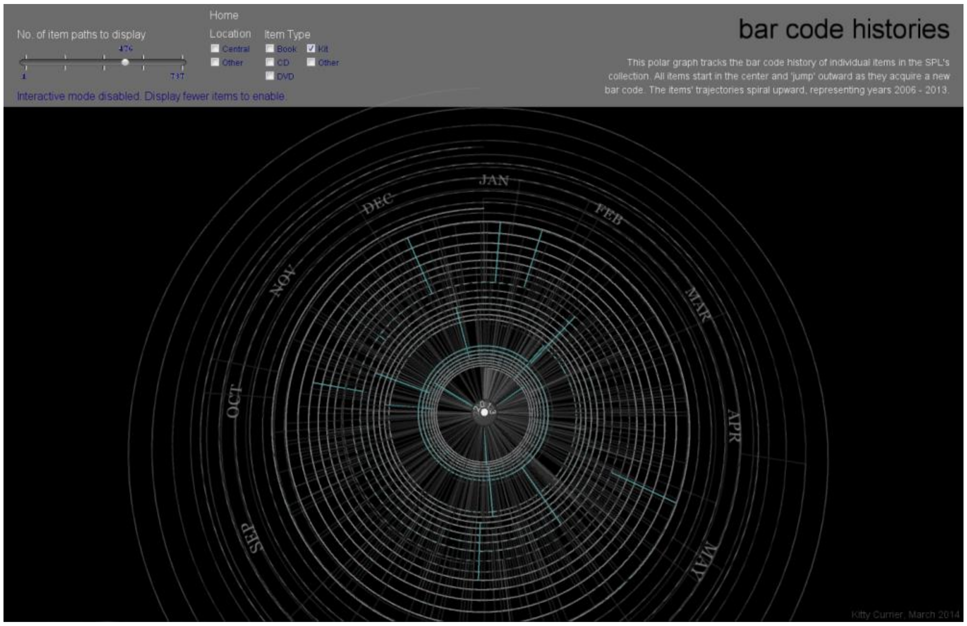
Item Type: ☑ Book ☑ Kit ☑ CD ☑ Other ☑ DVD

Interactive mode disabled. Display fewer items to enable.

Kitty Currier, March 2014

;

Kitty Currier
MAT 259
13 March 2014



Home

No. of item paths to display

476

1 737

Location

□ Central
□ Other

Item Type

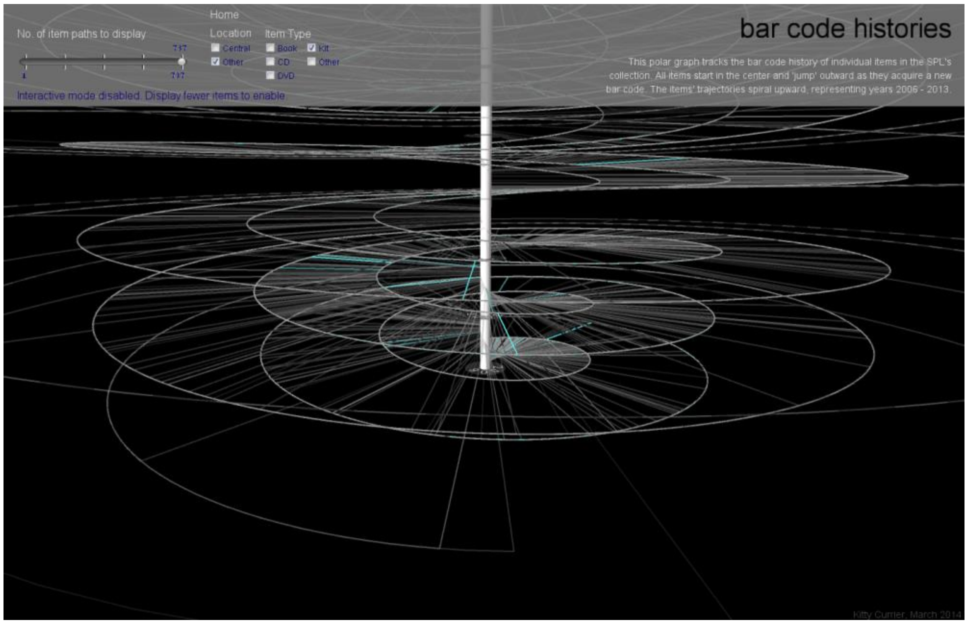
□ Book
□ CD
□ DVD
☑ Kit
□ Other

Interactive mode disabled. Display fewer items to enable.

bar code histories

This polar graph tracks the bar code history of individual items in the SPL's collection. All items start in the center and 'jump' outward as they acquire a new bar code. The items' trajectories spiral upward, representing years 2006 - 2013.

Kitty Currier, March 2014



Home

No. of item paths to display

737

1 737

Location

□ Central
☑ Other

Item Type

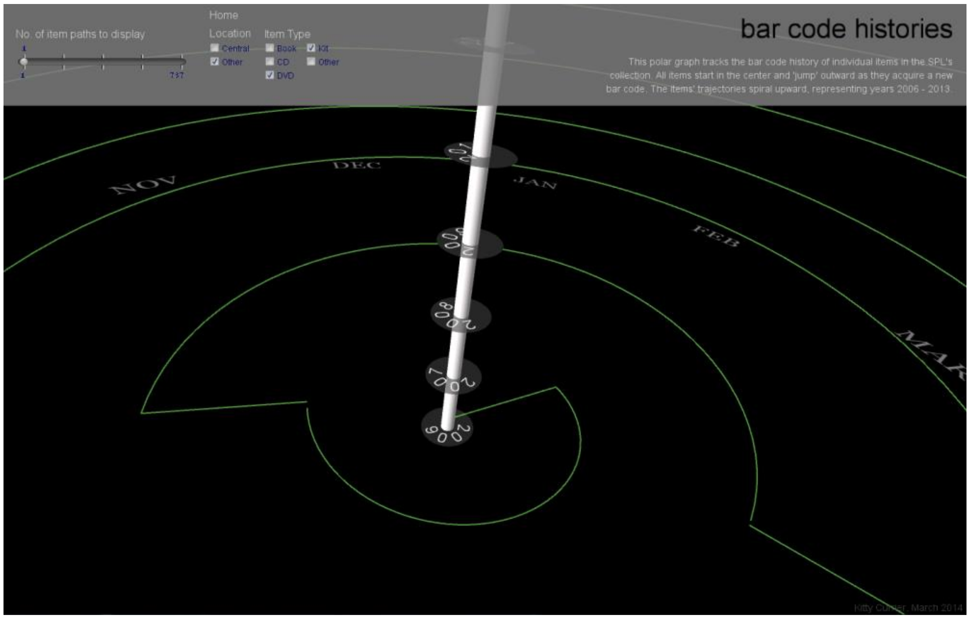
□ Book
□ CD
□ DVD
☑ Kit
□ Other

Interactive mode disabled. Display fewer items to enable.

bar code histories

This polar graph tracks the bar code history of individual items in the SPL's collection. All items start in the center and 'jump' outward as they acquire a new bar code. The items' trajectories spiral upward, representing years 2006 - 2013.

Kitty Currier, March 2014



Home

No. of item paths to display: 1 (1 – 737)

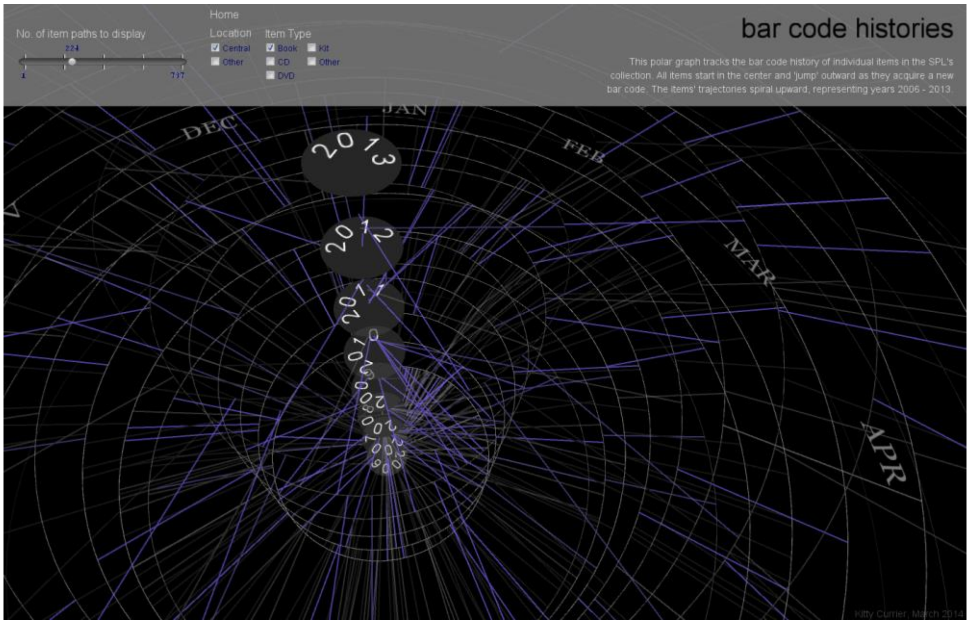
Location: ☐ Central ☑ Other

Item Type: ☐ Book ☑ Kit ☐ CD ☐ Other ☑ DVD

bar code histories

This polar graph tracks the bar code history of individual items in the SPL's collection. All items start in the center and 'jump' outward as they acquire a new bar code. The items' trajectories spiral upward, representing years 2006 - 2013.

Finally, I removed the inner cylindrical axis:



Home

No. of item paths to display: 224 (1 – 737)

Location: ☑ Central ☐ Other

Item Type: ☑ Book ☐ Kit ☐ CD ☐ Other ☐ DVD

bar code histories

This polar graph tracks the bar code history of individual items in the SPL's collection. All items start in the center and 'jump' outward as they acquire a new bar code. The items' trajectories spiral upward, representing years 2006 - 2013.

The visualization is not very useful, except to reveal a flaw in my data coding. An unusually large number of "spokes" appear at the beginning of the series in 2006, but this is actually an edge effect of sorts. The spokes connecting the radial axis to the first circular level simply represent the first time during the year 2006 that each item is checked out—jumping from "no bar code" to "first bar code". Subsequent radial levels represent the time an item changes from one bar code to another, an event with much less probability and much more randomness.

I preferred the gray background of my early sketch, but when I introduced more colors, they did not show up well on the gray background. All of the lines are translucent, which reduces their contrast with a gray background but less so with a black background.

Controls

A user can check boxes to visualize the trajectories (lines) by item type, represented by different colors, or the home location (central branch or other branch) of an item, represented by a change in brightness and opacity (brighter, more opaque lines for the option checked).