Project 2 - 3D Visualization: write up

Yanchen Lu

@February 21, 2023

Concept Summary

This project analyzes the check-out and check-in data for books related to fiber and textile crafts. I looked into 4 keywords primarily, "Crochet", one of its sub-form "Amigurumi", "Macrame", and its sub-form "Friendship Bracelet". Then, I queried for book records with titles containing these keywords, and measured the length of time each book spent outside the Seattle Public Library. Additionally, I aggregated the number of check-outs and check-ins that occurred for each of the four categories of books on each day to produce frequency datasets.

The queried data contains records from 2005 until the first two weeks of February 2023. The year-month-day timestamp data maps each record to a point in the 3D space, with days of a month organized along the x-axis, and months and years organized along the z-axis, with the height of each curve (along -y) determined by the number of days each book spent outside the SPL. For the aggregated frequency data, I visualized the number of check-outs and check-ins on each day as spheres, with the size of each sphere corresponding to check-out and check-in counts.

The GUI allows the user to navigate the 3D space and includes toggles to show or hide components of the visualization, revealing patterns in the dataset.

MySQL Query + Data PreProcessing

I first explored check-out and check-in records of books with keywords “Crochet”, “Amigurumi”, “Macrame”, and “Friendship Bracelet”.

```
SELECT *
FROM (
    SELECT DISTINCT(barcode), itemNumber, title, cout, cin, deweyClass, bibNumber, collcode, callNumber
    FROM spl_2016.inraw
    WHERE title LIKE '%macrame%'
    OR title LIKE '%friendship bracelet%'
    OR title LIKE '%crochet%'
    OR title LIKE '%amigurumi%'
) AS macrame_books
ORDER BY itemNumber, barcode, cout;
```

query0_macrame_crochet_cout_cin

barcode	itemNumber	title	cout	cin	deweyClass	bibNumber	collcode	callNumber
10015744476	4387	Knit to fit a comprehensive guide to hand and machine knitting and crochet	2005-12-20 16:23:00	2006-01-04 11:08:00	746	25861	nanf	746 D912K2
10015744476	4387	Knit to fit a comprehensive guide to hand and machine knitting and crochet	2006-01-07 16:37:00	2006-01-14 16:03:00	746	25861	nanf	746 D912K2
10015744476	4387	Knit to fit a comprehensive guide to hand and machine knitting and crochet	2006-01-07 16:37:00	2006-06-02 17:16:00	746	25861	nanf	746 D912K2
10015744476	4387	Knit to fit a comprehensive guide to hand and machine knitting and crochet	2006-01-07 16:37:00	2006-06-05 12:49:00	746	25861	nanf	746 D912K2
10015744476	4387	Knit to fit a comprehensive guide to hand and machine knitting and crochet	2006-06-10 11:22:00	2006-06-10 11:23:00	746	25861	nanf	746 D912K2
10015744476	4387	Knit to fit a comprehensive guide to hand and machine knitting and crochet	2006-06-10 11:22:00	2006-06-13 10:07:00	746	25861	nanf	746 D912K2

Screenshot of dataset from the 1st query

In the second query, I counted the number of days between check-out and check-in dates for each record, as “time spent outside the SPL”, and added a column to denote the artform category.

```
SELECT *
FROM
  (SELECT
    itemNumber, title, cout, cin, deweyClass,
    DATEDIFF(cin, cout) AS time_outside_lib,
    (CASE
      WHEN (title LIKE '%macrame%') THEN 'macrame'
      WHEN (title LIKE '%friendship bracelet%') THEN 'friendship bracelet'
      WHEN (title LIKE '%crochet%') THEN 'crochet'
      WHEN (title LIKE '%amigurumi%') THEN 'amigurumi'
    END) AS artform
  FROM
    spl_2016.inraw
  WHERE
    title LIKE '%macrame%'
    OR title LIKE '%friendship bracelet%'
    OR title LIKE '%crochet%'
    OR title LIKE '%amigurumi%') AS textile_art
ORDER BY artform, itemNumber, cout;
```

itemNumber	title	cout	cin	deweyClass	time_outside_lib	artform
3030771	Kyuuto Japanese crafts Amigurumi	1970-01-01 00:00:00	2008-02-01 10:45:00	746.43404	13910	amigurumi
3030771	Kyuuto Japanese crafts Amigurumi	2008-02-01 13:41:00	2008-02-26 13:57:00	746.43404	25	amigurumi
3030771	Kyuuto Japanese crafts Amigurumi	2008-02-28 11:45:00	2008-03-06 15:30:00	746.43404	7	amigurumi
3030771	Kyuuto Japanese crafts Amigurumi	2008-03-10 10:12:00	2008-03-31 10:36:00	746.43404	21	amigurumi
3030771	Kyuuto Japanese crafts Amigurumi	2008-03-10 10:12:00	2008-04-01 12:11:00	746.43404	22	amigurumi
3030771	Kyuuto Japanese crafts Amigurumi	2008-04-01 15:19:00	2008-04-21 14:48:00	746.43404	20	amigurumi
3030771	Kyuuto Japanese crafts Amigurumi	2008-04-01 15:19:00	2008-04-22 23:09:00	746.43404	21	amigurumi

Screenshot of dataset from the 2nd query, added time_outside_lib and artform columns

There are only two irrelevant books in the resulting data from the query, “Dead men dont crochet” and “un friendship bracelet”. I removed them from the dataset.

I noticed that there are a lot of duplicate check-out dates that match to different check-in dates, even if it’s the same copy of a book. It might be due to readers renewing books repeatedly before finishing and

returning the book to the library. Therefore, I chose to keep the last check-in date for each check-out record of each book copy. The dataset is in `data_nodup_last.csv`.

data_nodup_last

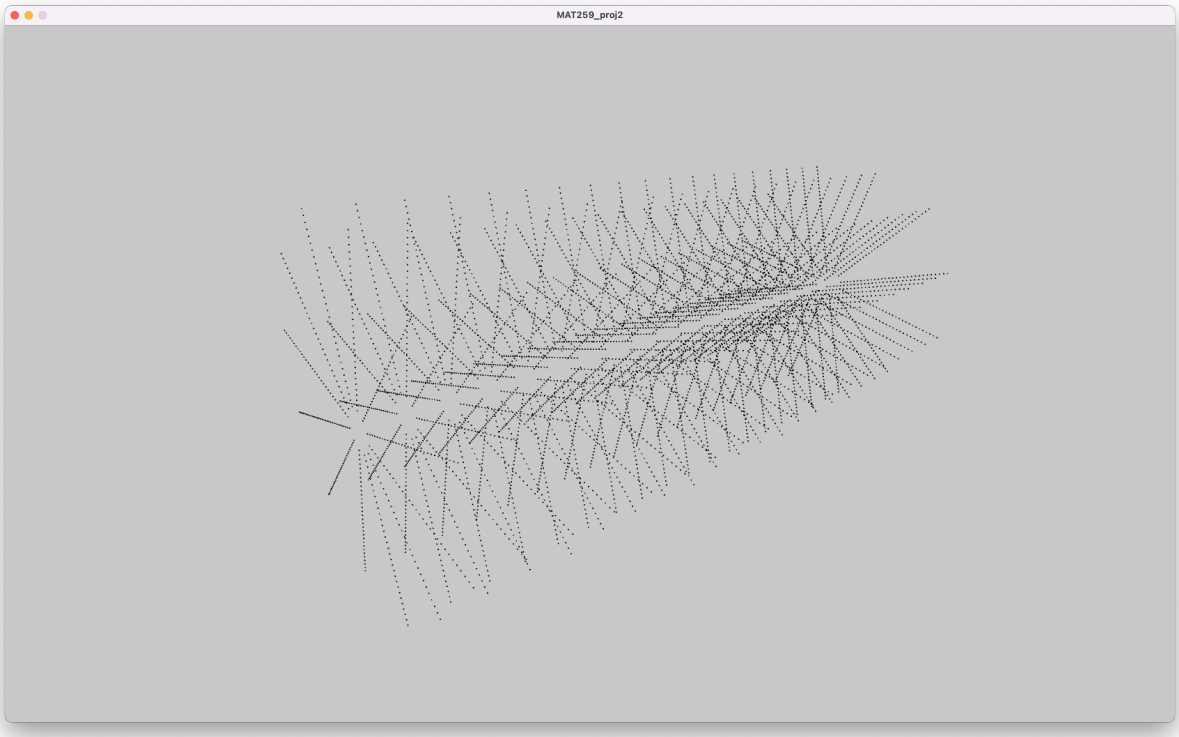
itemNumber	title	cout	cin	deweyClass	time_outside_lib	artform
3030771	Kyuuto Japanese crafts Amigurumi	2008-02-01 13:41:00	2008-02-26 13:57:00	746.43404	25	amigurumi
3030771	Kyuuto Japanese crafts Amigurumi	2008-02-28 11:45:00	2008-03-06 15:30:00	746.43404	7	amigurumi
3030771	Kyuuto Japanese crafts Amigurumi	2008-03-10 10:12:00	2008-04-01 12:11:00	746.43404	22	amigurumi
3030771	Kyuuto Japanese crafts Amigurumi	2008-04-01 15:19:00	2008-04-23 14:51:00	746.43404	22	amigurumi
3030771	Kyuuto Japanese crafts Amigurumi	2008-05-06 18:40:00	2008-06-11 16:27:00	746.43404	36	amigurumi
3030771	Kyuuto Japanese crafts Amigurumi	2008-06-15 16:40:00	2008-07-05 16:47:00	746.43404	20	amigurumi
3030771	Kyuuto Japanese crafts Amigurumi	2008-07-12 12:32:00	2008-07-28 15:35:00	746.43404	16	amigurumi

Dataset after removing duplicate check-out dates

Lastly, I aggregated check-out and check-in counts for each date within the dataset. This produced two more datasets, `aggregated_cout.csv` with check-out counts, and `aggregated_cin.csv` with check-in counts.

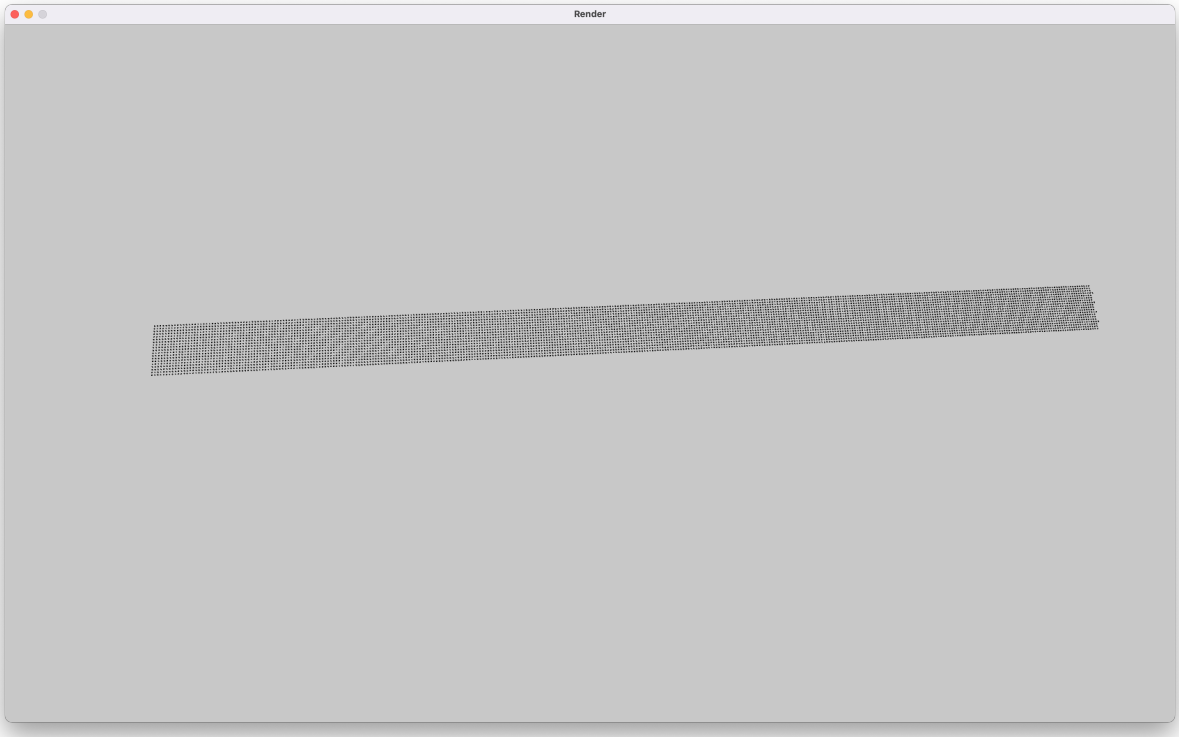
Ideation+Sketches

I started out with the idea of a calendar page. The design evolved from an opened book, to a cylinder consisted of 12 planes, then to a flat disc that represents each year. I sketched out the geometry in processing. It formed a spindle like structure. However, I didn't really like the shape.

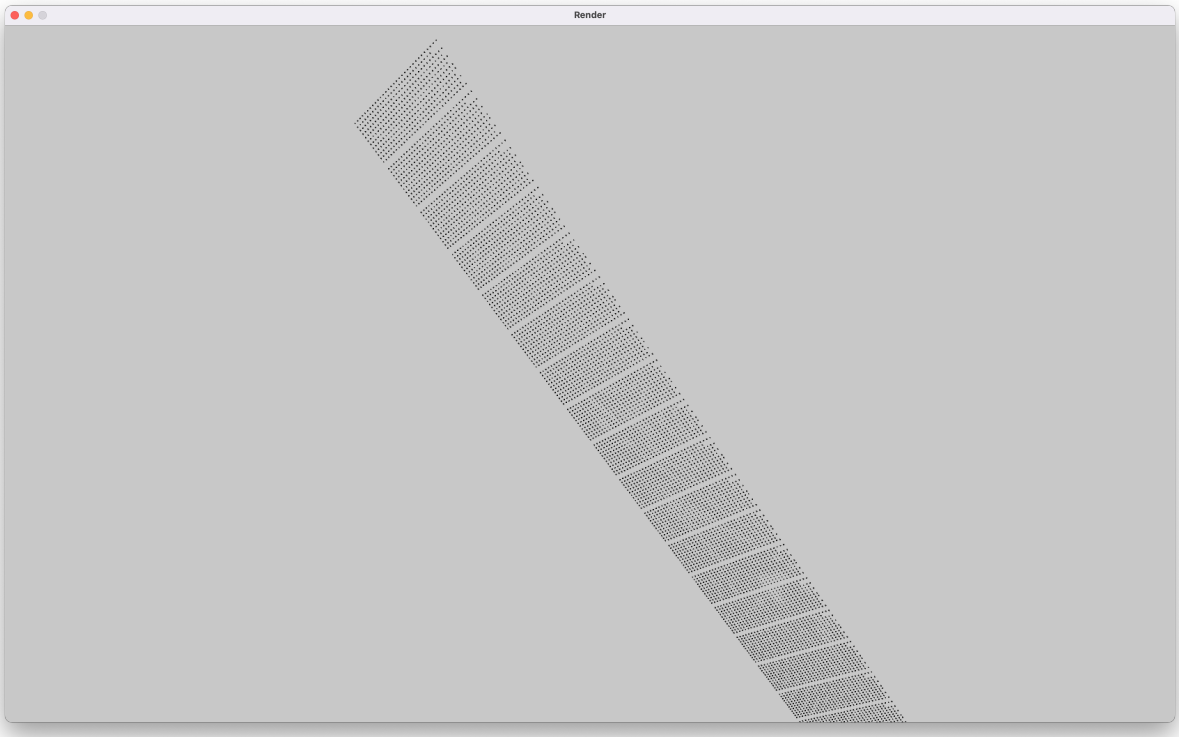


Spindle geometry

So, I decided to go back to my initial calendar idea. With each line presenting a year and the progression through the decades, it forms a 2D plane that maps each day from 2005/1/1 to 2023/2/14. The duration of a check-out → check-in record would determine the height of its curve.



Each year as a line, from 2005 to 2023

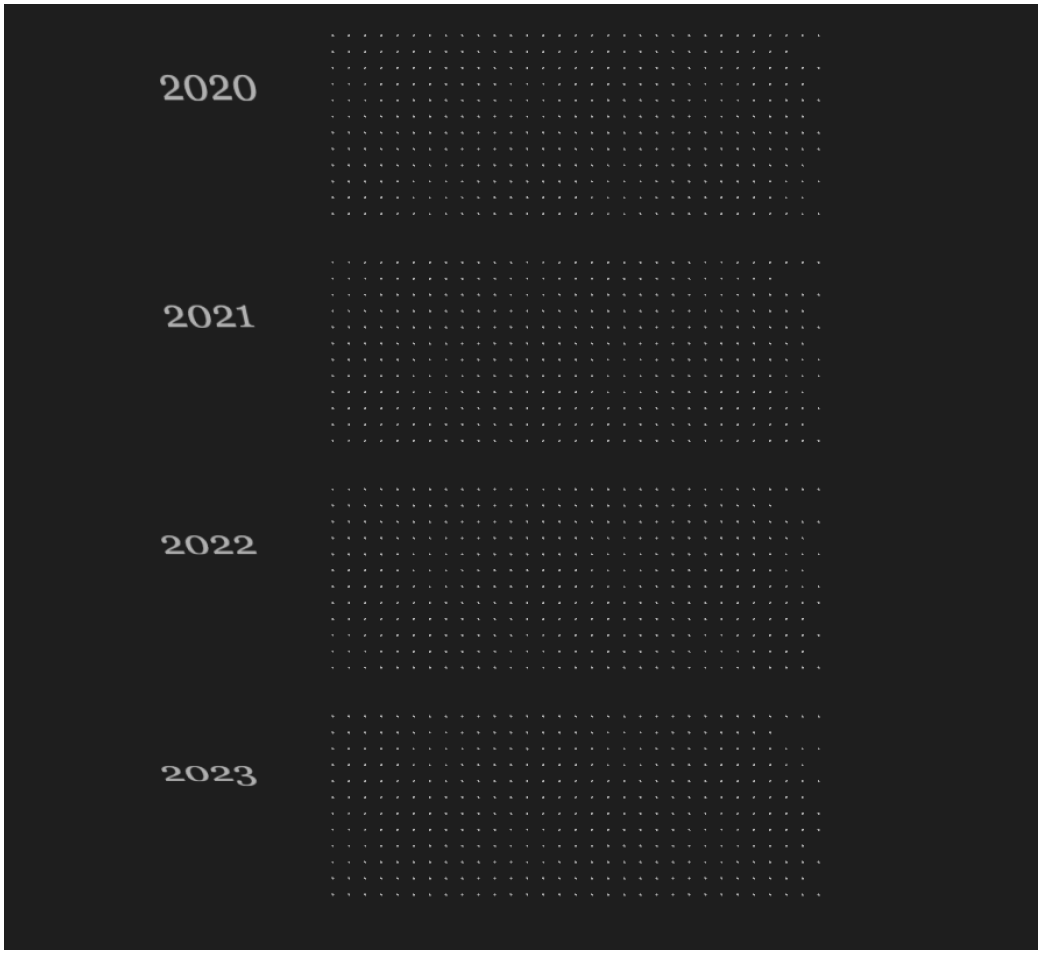


The geometry I ended up going with, each year as a rectangle

Since each horizontal line represents dates increasing to the right, though the right most dates may actually be closer in time with the left most dates on the neighboring line, the length the of the curve connecting those dates would be disproportionately long. Therefore, I thought using a “snake grid”, altering the direction or order of the dates on horizontal lines might facilitate a more accurate visual representation.

The Visualization

3D Space



The dates are mapped as the following:

Each point drawn in the space represents a day.

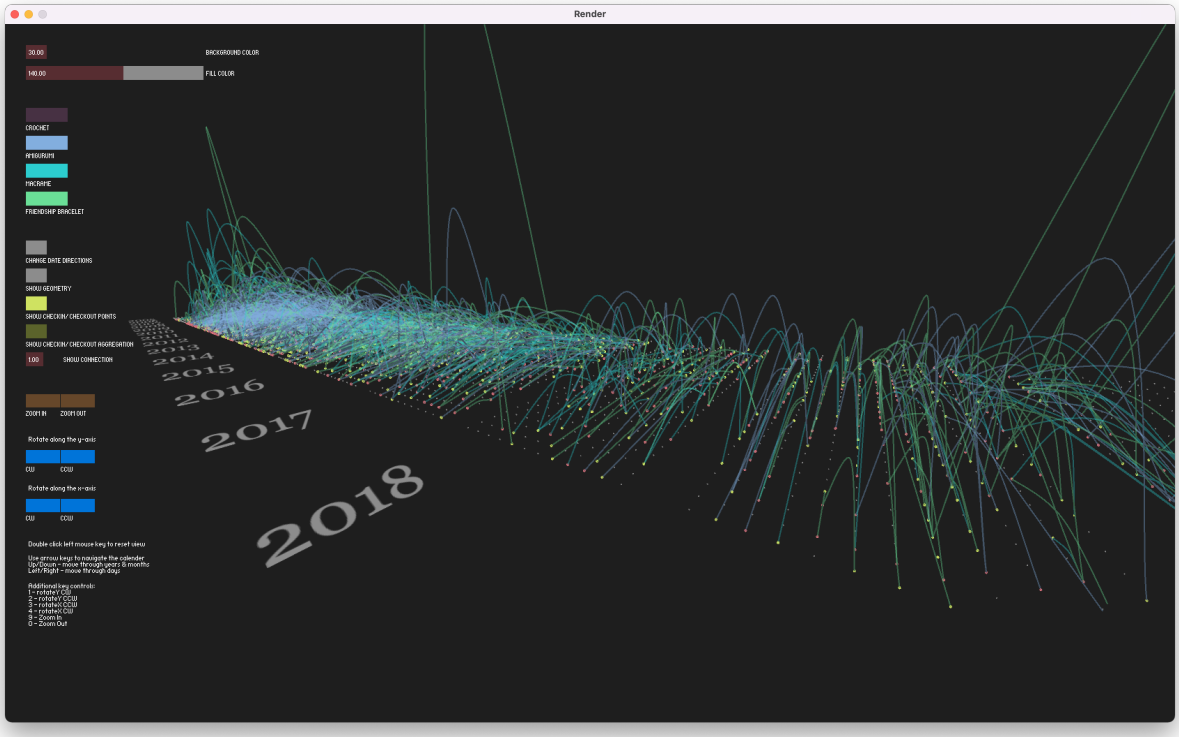
Each horizontal line along the positive x-axis represents an entire month. It shows the different number of days in each month and leap years.

The months, aka horizontal lines, are organized along the positive z-axis. The larger the z-value, the closer the line is to initial camera view position, closer in time (to current day).

Thus, each year is represented by a rectangle, also placed along the positive z-axis.

Curves

Check-outs are drawn as red points, and check-ins are drawn as yellow points. A curve is drawn from the check-out day to the check-in day. The taller the curve is, the longer the copy of the book is outside the SPL.



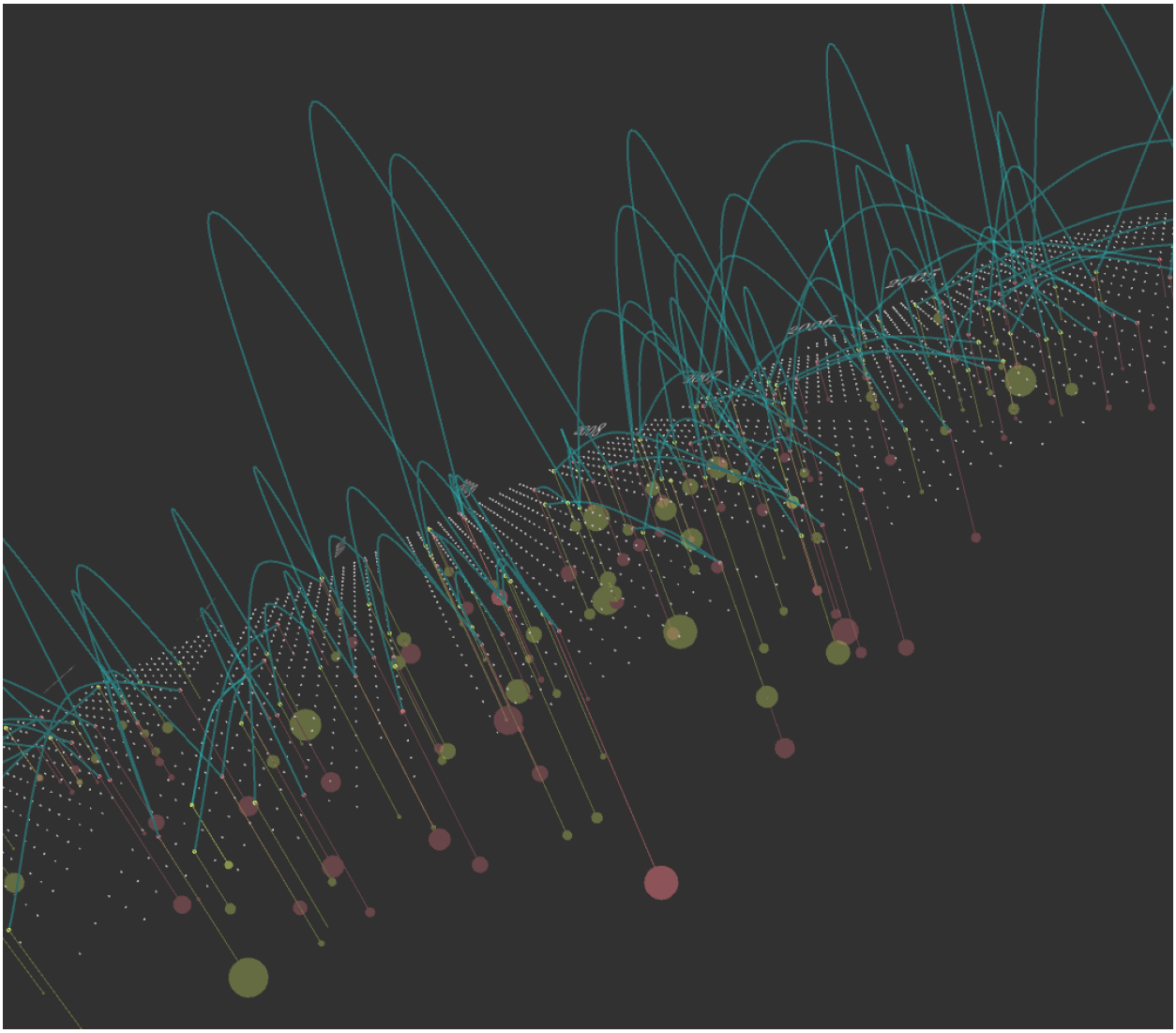
All check-out and check-in curves of categories “amigurumi”, “macrame”, and “friendship bracelet”

Spheres

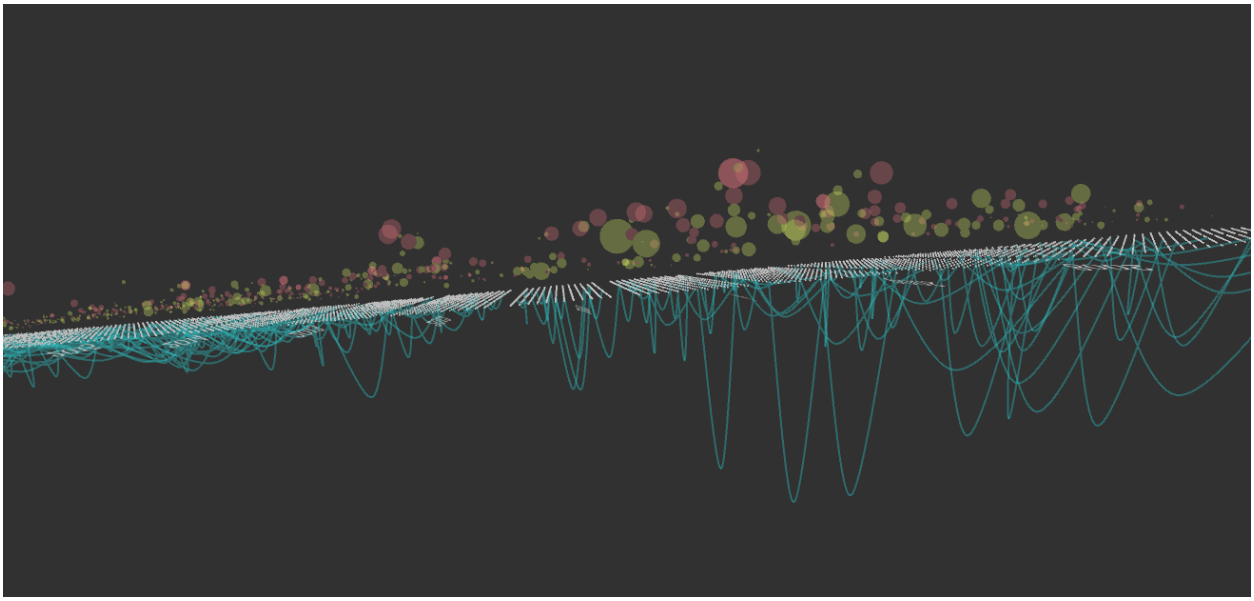
The aggregated counts of check-outs and check-ins records of each day are represented by spheres, with the diameters of the spheres increasing with the number of records. To prevent overlaps, the spheres’ y-positions also scale with the number of records.

On a single day, more check-outs and check-ins records means a larger sphere connected to that date in the 3D space, and the sphere is placed lower along the y-axis (larger y-value).

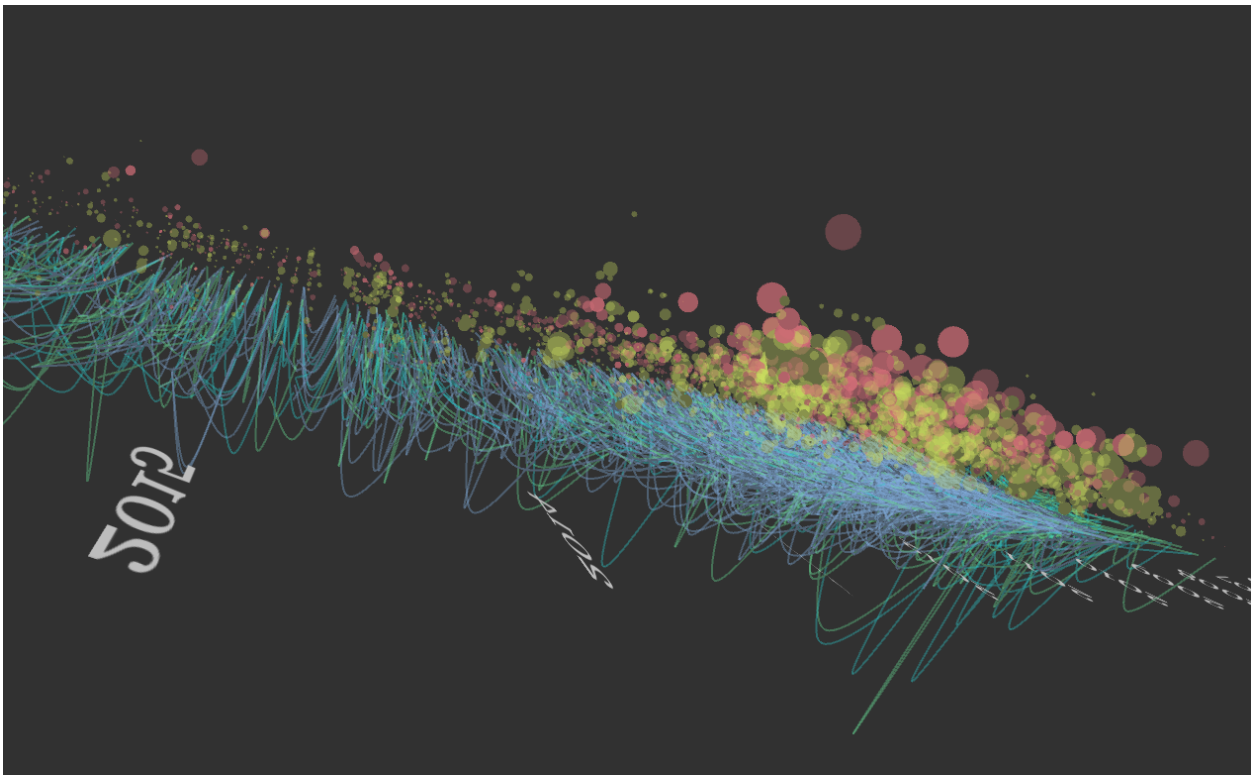
I thought it’s interesting to show not only the density of check-outs and check-ins, but also the scale of the number of check-out and check-in records.



Corresponding counts to each date represented by spheres



Spheres show without connecting lines

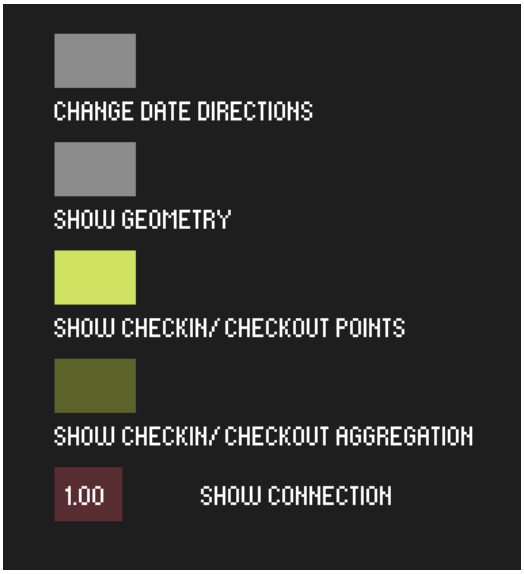


Aggregated counts in 3 categories

GUI

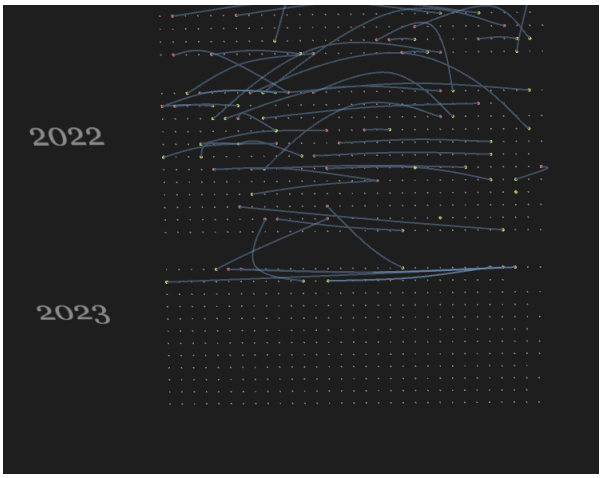
The background color, the fill color for the year labels and the geometry can be controlled with sliders.

The user can toggle to show or hide all 4 categories of records.

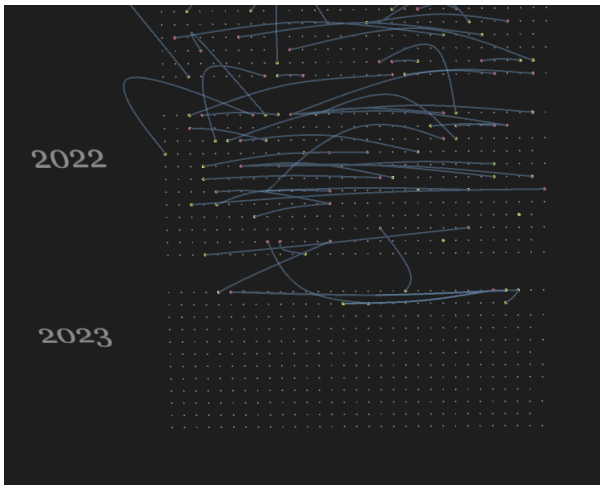


The user can toggle the geometry, the check-out/check-in points, the aggregated record counts spheres.

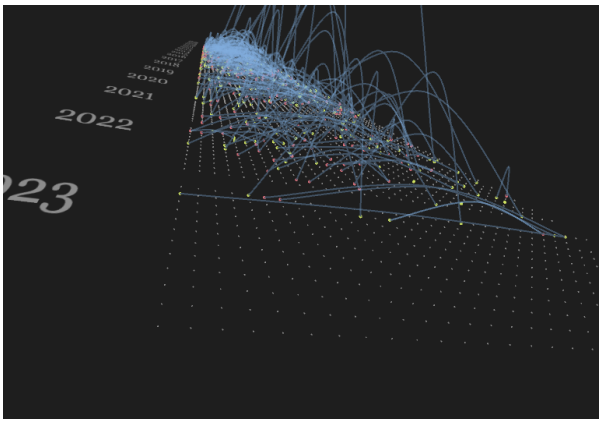
I've also added a toggle to alternate the date-direction to coordinate mapping. As mentioned in Ideation+Sketches, the alternating "snake grid" may show a more accurate visual representation, by placing closer dates in time closer in space.



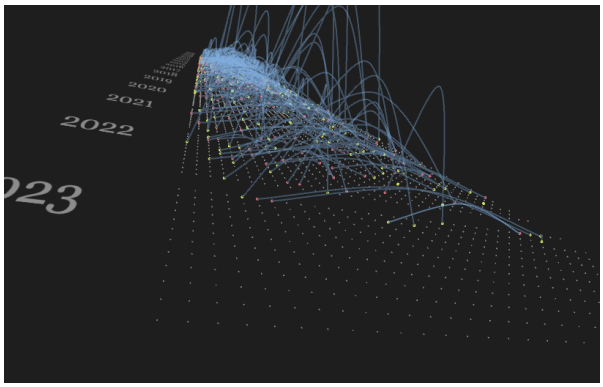
zigzag



alternating date-directions

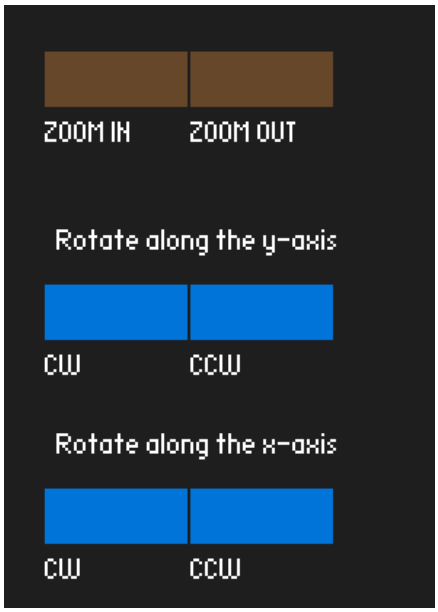


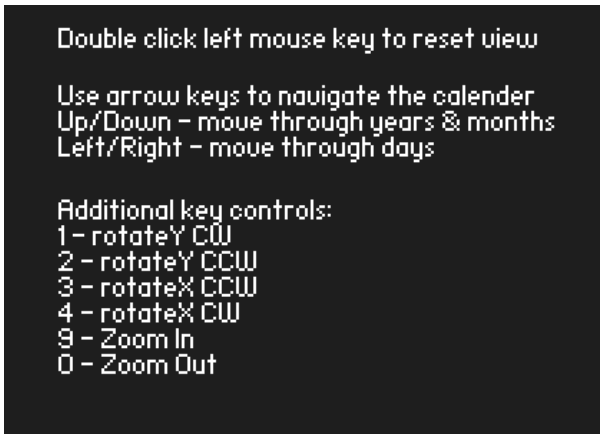
zigzag
notice how "short" one of the curve is, even though it's very wide, very long along the x-axis



alternating date-directions
the curves' height and width correlate better

The user can also zoom-in/ zoom-out, rotate along the x-axis and the y-axis, and navigate the dates with arrow keys.





Future Work

There are a few things I'd like to continue to work on:

1. Show the corresponding book title when the mouse is near a sphere or a curve
    a. To be able to click a sphere or curve to show check-out/check-in records of that book
2. Apply log-scale to transform the height of the curves - some areas are very dense and obscured shorter curves
3. The current strip spanning 19 year is really long. I'd like to find a better way to contain everything on the canvas/computer screen, e.g. map the date coordinates to a spherical shape to create a yarn ball shape.