# Project 1: Draft

Evgeny Noi

January 15, 2020

## 1. Introduction

For my project with Seattle Public Library (SPL) I decided to visualize a journey of a book in space and time. Unfortunately, the provided dataset does not have any information on the readers or their places of residence, so I decided to employ a Gravity model prevalent in geography and spatial analysis as well as other probabilistic methods to assess the service areas of SPL branches and hypothesize about the trajectory of the book given random or pseudorandom processes.

## 2. Data Acquisition

### 2.1 SPL inventory

SPL publishes its data on City of Seattle Open Data Portal, where library inventory may be found and downloaded. The inventory file contains information on the resources owned by the library, including the bibNumber from the Library of Congress, as well as the ItemType and ItemCollection that could be used to geneate unique identifier to be joined to the spl_2016 database. Additionally, I found an ealier copy of the inventory on the Portal (2017) on kaggle.com. Logically, it was interesting to compare the prescription of books to branches from these two time frames: 2017 and 2020.

### 2.2 SPL branches

The information on branches is located on the official SPL website. Overall, there are 27 branches in Seattle. Their geographic coordinates can be downloaded from City of Seattle Open Data Portal.

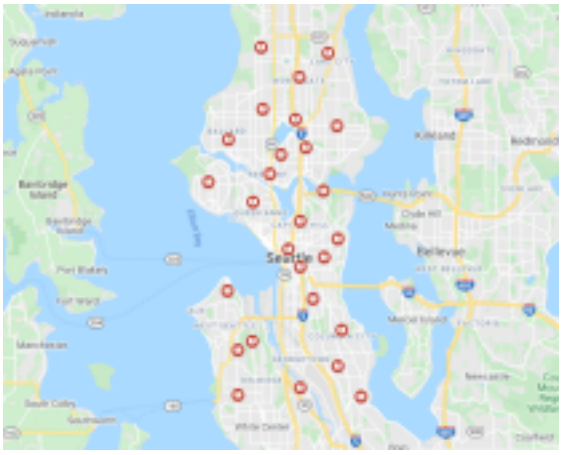


Figure 1: Branches of SPL library

### 2.3 SPL check-ins and check-outs

The SPL dataset is made available as part of the MAT259A course. MySQL Workbench was used to access data and run queries on the database.

## 3. Data Processing

There were four stages in data processing: 1. Downloading and pre-processing inventory data files 1. Inventory data from Kaggle (2017), denoted as $\mathcal{I}_{\infty}$, 2. Inventory data from Seattle Open Data Portal (2020), denoted as $\mathcal{I}_{\in}$, 2. Use 2017 inventory file as a baseline and keep only interesting records between the datasets. We can refer to this set as $\mathcal{I} \in \mathcal{I}_{\infty} \cap \mathcal{I}_{\in}$.

```
import pandas as pd
import numpy as np
print('Hello World')
df = pd.read_csv('/home/file.csv')
```

3. Creating unique identifier for merging inventory ($\mathcal{I}$) and check-ins/outs ($\mathcal{C}$)
    1. Because inventory files do not have a barcode or other unique identifiers for a join, I created a unique identifier for both inventory and check-ins/outs by concatenating *'bibNumber', 'CollCode' / 'ItemCollection', and 'ItemType'*. See *Figure 2* for the formula.
4. Running SQL query
    1. I copied the identifiers into a generated SQL query and ran it on the database to build the following query:

```
SELECT
  *
FROM spl_2016.inraw
WHERE cout>'2016-12-31' AND
  CONCAT(bibNumber,collcode,itemtype) in ('261cs9rarbk',
  -- < ... > here 10,900 identifiers were inserted
  '3304438canfacbk');
```

Q2 | =CONCAT("'", B2, J2, I2, "'", ",")

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | BibNum | Title | Author | ISBN | Publicatio | Publisher | Subjects | ItemType | ItemColle | FloatingI | ItemLoca | ReportDa | ItemCou | locbib | tycobib | |
| 2 | 0 | 261 | American | Holt, Alfred Hubbar | | 1969 | Gale Rese | Names G | arbk | cs9r | | cen | ### | 1 | cen2610 | 261cs9rarbk | '261cs9rarbk', |
| 3 | 1 | 813 | Central P | Johnston, Nancy | | [1968] | Sierra Clu | Central P | acbk | canf | | cen | ### | 1 | cen8130 | 813canfacbk | '813canfacbk', |
| 4 | 2 | 2524 | The Virgi | Dowdey, Clifford, 1 | | [1969] | Little, Bro | Carter Ro | acbk | canf | | cen | ### | 1 | cen2524 | 2524canfacbk | '2524canfacbk', |
| 5 | 3 | 3119 | The light | MacDonald, George | | [1969] | Farrar, St | Fairy tale | jcbk | ncfic | | wts | ### | 1 | wts3119 | 3119ncficjcbk | '3119ncficjcbk', |
| 6 | 4 | 4439 | An encyc | Franklyn, | 8E+07 | [1969] | Pergamo | Heraldry | arbk | cs9r | | cen | ### | 1 | cen4439 | 4439cs9rarbk | '4439cs9rarbk', |
| 7 | 5 | 4764 | Pearce pi | Pearce, Marvin J., 1 | | [1969] | | Pierce far | arbk | cs9r | | cen | ### | 1 | cen4764 | 4764cs9rarbk | '4764cs9rarbk', |
| 8 | 6 | 4974 | History o | Morrison, Leonard | | 1971 | Privately | Kimball f | arbk | cs9r | | cen | ### | 1 | cen4974 | 4974cs9rarbk | '4974cs9rarbk', |
| 9 | 7 | 6212 | Iggie's ho | Blume, Ju | 0134508 | [1970] | Bradbury | Race rela | jcbk | ncfic | | cap | ### | 1 | cap6212 | 6212ncficjcbk | '6212ncficjcbk', |
| 10 | 8 | 6229 | A financi | Myers, M | 2.31E+08 | 1970 | Columbia | Finance P | acbk | canf | | cen | ### | 1 | cen6229 | 6229canfacbk | '6229canfacbk', |
| 11 | 9 | 6504 | Quakeris | Carroll, Kenneth L. ( | | [1970] | Maryland | Society o | arbk | cs9r | | cen | ### | 1 | cen6504 | 6504cs9rarbk | '6504cs9rarbk', |

Figure 2: Generating unique identifiers for joining inventory and check-ins/outs

## 4. Results

The resulting table is illustrated in *Figure 2* and has the follwoing dimensions: XXX. This data will be further used to visualize the journey of a book in space-time using probabilistic models and p5.js.

| | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | id | itemNumber | bibNumber | cout | cin | collcode | itemtype | barcode | title | callNumber | deweyClass | subj |
| 2 | 83770886 | 767844 | 6212 | 2017-02-09 18:12:00 | 2017-03-11 16:34:00 | ncfic | jcbk | 10013772867 | Iggies house | J BLUME | | NULL |
| 3 | 84681966 | 767844 | 6212 | 2017-05-01 16:11:00 | 2017-05-21 16:10:00 | ncfic | jcbk | 10013772867 | Iggies house | J BLUME | | NULL |
| 4 | 83254746 | 1693272 | 33200 | 2017-01-04 14:26:00 | 2017-01-22 16:32:00 | canf | acbk | 100489012 | American slavery American freedom th | 975.5 M821A | 975.5 | NULL |
| 5 | 84686680 | 166431 | 46636 | 2017-05-01 15:31:00 | 2017-05-22 12:51:00 | nchol | jcbk | 10007090375 | Humbug rabbit | E BALIAN | | NULL |
| 6 | 83408679 | 570489 | 70465 | 2017-01-04 19:29:00 | 2017-02-02 19:03:00 | naaab | acbk | 10012382163 | Nigger an autobiography | B G86216 | | NULL |
| 7 | 84267864 | 570489 | 70465 | 2017-03-28 12:35:00 | 2017-04-19 11:24:00 | naaab | acbk | 10012382163 | Nigger an autobiography | B G86216 | | NULL |
| 8 | 86132367 | 570489 | 70465 | 2017-08-24 16:00:00 | 2017-09-12 15:35:00 | naaab | acbk | 10012382163 | Nigger an autobiography | B G86216 | | NULL |
| 9 | 86165829 | 570489 | 70465 | 2017-09-13 14:50:00 | 2017-09-15 08:07:00 | naaab | acbk | 10012382163 | Nigger an autobiography | B G86216 | | NULL |
| 10 | 86169666 | 1130295 | 75901 | 2017-06-30 14:37:00 | 2017-09-15 13:53:00 | canf | acbk | 10004684923 | Perspective for artists | 742 L272P | 742 | NULL |

Figure 3: The resulting table of check-ins/outs