

# Trends in Seattle Library Usage

Taylorle Prinsen  
taylorle.prinsen@colorado.edu  
Colorado University at Boulder  
Boulder, Colorado

Sky Johnson  
sky.johnson@colorado.edu  
Colorado University at Boulder  
Boulder, Colorado

Peter Minwegen  
peter.minwegen@colorado.edu  
Colorado University at Boulder  
Boulder, Colorado

## ACM Reference Format:

Taylorle Prinsen, Sky Johnson, and Peter Minwegen. 2019. Trends in Seattle Library Usage. In *SIGGRAPH '19 Talks, July 28 - August 01, 2019, Los Angeles, CA*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3306307.3328180>

## 1 PROBLEM STATEMENT/MOTIVATION

The Seattle Public Library is a landmark of the city for its architectural style, and unique community gatherings. However we wanted to look further into its functions as an actual library.

It is currently unclear whether or not trends in rentals of library books (or other materials) in the Seattle area are impacted by the weather trends in the area. We intend to use a data set with weather statistics in Seattle as well as a data set with statistics related to the library check outs in the area. We would like to explore whether or not people check out books more or less when the weather is rainy, sunny, etc. We would also like to see if there is a correlation between winter and summer months and the topics of books that people check out. This could be very interesting to help us better understand how people tend to cope with varying weather conditions.

Many people suffer from Seasonal Affective Disorder (SAD) and are left facing varying side effects from weather impacting their mood and demeanor. We are intrigued by the ways that weather can impact people's moods and behaviors and are intrigued to begin to understand how this may impact people's daily habits, going to the library, etc. We are also intrigued to learn if these weather trends tend to influence the genre of books or music that people check out.

<https://www.mayoclinic.org/diseases-conditions/seasonal-affective-disorder/symptoms-causes/syc-20364651>

## 2 LITERATURE SURVEY

There has been very little research done with these data sets in the past. The only work we came across that used the Seattle Public Library data set only referenced data specific to that database - what time of day is most popular for rentals at the library, which type of rental is most popular at certain times, etc. At the end of this quick study the author came to the conclusion that 4pm, Fridays in January and July are the busiest days for activity within the library's checkout system dataset. However this is limited and doesn't ask much of why that is the case.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted by ACM, provided that the copies are not made for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*SIGGRAPH '19 Talks, July 28 - August 01, 2019, Los Angeles, CA, USA*

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6317-4/19/07.

<https://doi.org/10.1145/3306307.3328180>

2019-07-13 00:26. Page 1 of 1-3.



Figure 1: The Seattle Public Library

We intend to extend this research by combining it with weather data from the area to better understand how weather might impact rental trends.

Some of the trends that previously have been studied in the Seattle Public Library Data set include:

- Number of checkouts across the years (separated by books and DVD/CD)  
A steady number of book checkouts (around 4 million) observed, while CD/DVD checkouts decreased over time, peaking around 2009 - probably from the emergence of Netflix, Spotify, etc.

- Checkout temporal trends:
  - Time of Day  
4 p.m. is the most popular hour for checkouts - possibly because people get off work and then head to the library?
  - Day of the Week  
Saturday has remained the most popular day for all years in the data set (Friday and Sunday are the least popular days).
  - Month of the Year  
"The beginning of the year and the summer seem to be the most popular months for checkouts, with September and December the least popular." (CITE - from toward data science link)

<https://towardsdatascience.com/how-and-when-people-use-the-public-library-1b102f58fd8a>

### 3 PROPOSED WORK

Due to the exploratory nature of this project, we will have to revise and plan to explore more options of analysis as we continue to work on this. However, we do know some basic parts that we will have to work on at some point

#### 3.1 Data Pre-Processing

The nature of our data sets will involve a decent amount of effort within integrating. The checkout data doesn't inherently follow the same structure as the weather data, so we will have to figure out the best way to align those. Currently we plan to work with the checkout data initially to find basic statistics related to the libraries usage. (These will be nothing but high level observations and extrapolation). Then, we will compare those observations and directly add in weather data in a way that matches, that way we can begin to create more accurate observations that includes the weather data.

#### 3.2 Time Series Analysis

The next big challenge that we will have to approach is how we handle time series within the data set. Currently we plan to use some basic mathematical statistics APIs to quickly find and visualize time data on how check out patterns change.

### 4 DATA SETS

We will be using two (possibly more) data sets for our main research. We will use one data set that tells us about the checkout trends at the Seattle Library, and concatenate this with a data set that tells us the weather trends in the same months. By looking at these together, we hope to answer complex questions about trends in weather and library rentals.

#### 4.1 Seattle Checkouts by Title

This data set contains information for the Seattle Public Library for physical and digital data. There are 1,431,563 unique objects in the data set, each with data for 11 attributes:

- UsageClass (Nominal)  
Denotes if item is physical or digital
- CheckoutType (Nominal)  
Denotes the vendor tool used to check out the item.
- MaterialType (Nominal)  
Describes the type of item checked out (examples: book, song, movie, music, magazine)
- CheckoutYear (Numeric/Interval)  
The 4-digit year of checkout for this record.
- CheckoutMonth (Nominal)  
The month of checkout for this record.
- Checkouts (Numeric/Interval)  
A count of the number of times the title was checked out within the Checkout Month.
- Title (Nominal)  
The full title and subtitle of an individual item
- Creator (Nominal)  
The author or entity responsible for authoring the item.

- Subjects (Nominal)  
The subject of the item as it appears in the catalog.
- Publisher (Nominal)  
The publisher of the title.
- PublicationYear (Numeric/Interval)  
The year from the catalog record in which the item was published, printed, or copyrighted.

This data set begins in 2005. Link: <https://www.kaggle.com/city-of-seattle/seattle-checkouts-by-title>

#### 4.2 Seattle Weather Data

*4.2.1 Did It Rain in Seattle? (1948-2017).* "Besides coffee, grunge and technology companies, one of the things that Seattle is most famous for is how often it rains. This dataset contains complete records of daily rainfall patterns from January 1st, 1948 to December 12, 2017."

This data was collected at Seattle Tacoma International Airport, WA. <https://www.kaggle.com/ratatman/did-it-rain-in-seattle-19482017>  
<https://www.wunderground.com/hourly/us/wa/seattle/KSEA>  
<https://www.ndbc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND>

*4.2.2 Seattle Weather csv.* This data set contains 1463 unique objects. Each is a day included in the range from January 1 2012 to December 31 2015. This data was found on Github:

The attributes for the objects include:

- Date (Numeric/Continuous)  
The unique day, month, year for each entry
- Precipitation (Numeric/Ratio)  
Amount of precipitation.
- Temp<sub>Max</sub>(Numeric/Ratio)  
The maximum temperature that day.
- Temp<sub>Min</sub>(Numeric/Ratio)  
The minimum temperature that day.
- Wind (Numeric/Ratio)  
The wind speed in mph.
- Weather (Nominal)  
Drizzle, Rain, Sun, or Snow

This data was found on Github: <https://github.com/domoritz/maps/blob/master/data/weather.csv>

#### 4.3 Seattle Traffic Data

If we have enough time left on the project, we may also be able to include traffic data within our analysis. To see how weather can influence the effects of people's travels within the city, and if that has a direct influence on when people utilize the library.

The attributes for the objects include:

- ID (Nominal)  
Unique ID for entry
- stname (Nominal)  
Name of the street
- countlocation (Nominal)  
Specific street location
- year (Ordinal)  
year that data was taken.
- aawdt (Numeric/Ratio)  
Annual Average Weekday Traffic

<https://data.seattle.gov/resource/38vd-gyvtv.json>

## 5 EVALUATION METHODS

How we evaluate and approach the data will obviously change as the project continues.

We will start however, with the aim to look into discovering interesting correlations between data based on time series analysis techniques. One of the biggest challenges we will face is trying to differentiate between when there is a significant event that has a correlation to another pattern, vs random noise.

I think it will be appropriate to measure how successful our project is based on how many interesting correlations we can find, and how we can support them with statistical significance. However we won't be able to figure out this amount, or if it's possible at all, without more work on the data set.

## 6 TOOLS

For this project we will use the standard libraries with Python and Jupyter Notebooks. Specifically, we plan to import and handle our data using Panda's DataFrames so we can quickly sort through and evaluate large portions of the data at a time.

We can then use Numpy's tools and methods to create any confusion matrices and arrays, while also letting us handle the large data sets without too much trouble.

Finally for Time Series, we can make use of the statsmodels API library to quickly get started in looking for statistically interesting events within the time series data, and finally we will use Matplotlib to visualize.

## 7 MILESTONES

We plan to have the project completely by the Due date obviously, August 9th, 2019.

However we can set much more internal dates by then. Setting these milestones will allow us to also evaluate how our project is going so far, as well as help manage time.

Currently we plan to have a basic exploration and visualization of the data set done by the next submission, July 20th. This will include data integration, and analysis of trends within the library (basically leaving off where the last study left off).

The next section of the study will involve looking at how weather data and library data can influence one another, July 27th.

Finally, we will use the remainder of the time to address any additional questions or possible paths of exploration that might be interesting to pursue. This will include adding the Traffic Data to our analysis if it is possible.

This all aligns with our final due date less than a month out on August 9th.