**Workflow Documentation Group 6**

Christabel Bonfoh, Benjamin Winsemius, Vanessa Terlouw

**Project Overview**

The purpose of this research is to examine whether national political events in the United States of America or local socio-political and/or economic developments influence book borrowing patterns at the Seattle Public Library, indicating how such events may affect public engagement with political topics. Our research question is: *Do national political events such as Obama’s re-election in 2012 and Trump’s election in 2016 shape borrowing patterns of books on their policies at the Seattle Public Library (SPL), compared to the influence of local socio-economic events?*

Seattle is situated within Washington State, a region consistently classified as a “blue state” due to its long-standing support for Democratic candidates and progressive policies. The city itself is known for its liberal political orientation, with residents typically advocating for labor rights, environmental sustainability, social equity, and access to public services. This political culture can shape community behaviors in notable ways, particularly regarding information-seeking practices. As a result, analyzing library borrowing patterns in Seattle may offer valuable insight into how a politically engaged population seeks out information and responds to socio-political events.As stated by the Journal of Radical Librarianship, “Historically, libraries have been highly successful in promoting themselves as having critical importance in a democratic society” (Sierpe, E., 2017). Similarly, Arseneau and Dodd explain that “[l]ibrarians are traditionally strong advocates for open access because it encourages the freedom of information,” further noting that “library workers create systems that ensure members of their community can freely access the information they need for learning, growth, and empowerment regardless of technology, format, or delivery methods” (as cited in Arseneau & Dodd, 2025, p. 6). In other words, libraries function as a democratic civic infrastructure, providing open access to knowledge that supports public education, informed decision-making, and civic engagement, particularly in a city like Seattle where residents are actively interested in social, economic, and political issues.

Through the examination of SPL book checkout data across national and local contexts, we aim to research if and how libraries may function as mirrors of civic awareness. When significant political or social developments occur, shifts in book borrowing may reveal how communities seek to understand the world around them. This approach moves beyond digital engagement (e.g. social media activity) and rather explores a more reflective form or information seeking - borrowing a book from a local library. At the national level, we focus on the 2012 and 2016 presidential elections, while at the local level, we consider the following events: the 2012 May Day protests and the 2016 Seattle Homeless Needs Assessment.

Regarding the first of these local events, *May Day* serves as a global day of labor solidarity and protest, with origins in the United States tracing back to the late 19th century, most notably *the Haymarket Affair of 1886* in Chicago, which significantly shaped the day’s legacy. On that day, [unarmed strikers] at the McCormick Reaper Works factory clashed with police, resulting in “the deaths of six workers [which] became a call for direct action, [where] a public rally was called for the following day to be held in Haymarket Square” (The Haymarket Affair, 1886 | Gilder Lehrman Institute of American History, n.d.), and consequently, tensions escalated between labor activists and anarchists on one side and authorities on the other. In the aftermath, May 1 was adopted nationally as a day to honor labor movements and advocate for workers’ rights. Similarly, the 2012 *May Day protests in Seattle* echoed this historical spirit, as various groups, including the Occupy movement, organized demonstrations addressing economic inequality, labor rights, and corporate influence. These events were characterized by both peaceful marches and instances of civil rebellion, including property damage and confrontations with law enforcement, which in turn attracted significant media coverage and public discussion (The Guardian, 2012; The Seattle Times, 2016).

In 2016, Seattle faced a pronounced homelessness crisis that exposed significant social and economic challenges within the city. As reported by The Guardian, “The past decade has seen a sharp rise in the number of homeless people – around 3,000 sleep rough, with thousands more in shelters and transitional housing” (Rock, 2016). The article further notes that after 47 homeless individuals died on the streets in the previous year, Mayor Ed Murray declared a “state of emergency,” calling for increased state and federal support and urging the city to unite in addressing the crisis (Rock, 2016). This heightened visibility of homelessness not only drew public attention to the issue but also stimulated civic discourse, encouraging residents to seek information about housing policy, social services, and economic inequality. Complementing this, the City of Seattle’s 2016 Homeless Needs Assessment highlighted the structural causes of homelessness, including the lack of affordable housing, limited access to mental health services, and systemic economic inequality (City of Seattle, 2016).

On a national level we focus on the presidential election of 2012 and 2016. In 2012, President Barack Obama was re-elected, running a campaign building upon his progress in his first term. The main focus was about restrengthening the economy after it hit the biggest economic crisis since the Great Depression, the Great Recession in 2008. The Great Recession hit many middle class families and Obama sought to provide financial stability for them through his campaign. With the help of the Dodd-Frank Act in 2010, which made legislation in the US financial system safer, Obama sought to change investments. He wanted to improve education with training for jobs where people were needed and lowering tuition with pell grants. He also wanted to improve healthcare by further strengthening the Affordable Care Act, also known as Obamacare, that helps millions of people with health insurance. Furthermore, creating new sources of energy in America was a campaign point. These investments would help families financially and help regrow the economy. With the ending of the Iraq war and the killing of Osama Bin Laden in 2011 in Obama’s first term, some campaign points of 2012 were about reducing investments in wars and putting them towards education, new energy sources and tax cuts. Obama believed that taking money that was saved for the two wars the United States were battling would help rebuild America (PBS NewsHour, 2020). With new directions of investments Obama strived to make American families and businesses progress, thereby standing behind his campaign slogan: “Forwards”.

The 2016 presidential election was a historic election that brought many discussion points. For the first time in American history a woman was the presidential nominee of a major US political party: Hillary Clinton (NBC News, 2016). The Clinton family was well known to the public as Bill Clinton was president from 1993 till 2001. Criticism about Bill’s presidency, family scandals and the fact that Hillary Clinton was a woman were brought up in the debates to convince people that she would not have been a good president. In his campaign points Donald Trump emphasizes the phrase “Make America Great Again”. He criticizes the economy and culture that America has found itself in. The main focus of his campaign was on immigration. Illegal immigrants would have been taking jobs from American civilians, bringing drugs and bringing crime into the land. His solution includes building a wall at the border of Mexico and mass deportation of illegal immigrants (CBS News, 2015). For the economy, Trump proposed tariffs on countries and giving out tax cuts that benefit the wealthy. Trump also had a strong stance on foreign affairs that included backing out of foreign trade deals like NAFTA and TPP, improving relations with Russia and renegotiating the Iran deal. Donald Trump ended up winning the presidential election despite being behind in favorability where Clinton took the lead of 47% while Trump stood on 42% (CNN, 2016).

Such visible civic engagement frequently stimulates broader public interest in underlying social and political issues, prompting communities to actively seek information on labor movements, social inequality, and economic policy. Consequently, libraries, such as the Seattle Public Library, provide critical access to curated collections, government reports, and historical analyses, enabling researchers and the general public to contextualize and interpret complex social challenges. By tracking library engagement and checkout patterns during periods of heightened public attention, such as the 2012 May Day protests or the 2016 homelessness crisis, researchers can gain insight into community information-seeking behaviors and civic participation. In this way, libraries function not only as repositories of information but also as essential sites for reflective learning, democratic engagement, and the dissemination of knowledge in response to pressing societal issues (City of Seattle, 2016; Rock, 2016).

Accordingly, we hypothesize that at the Seattle Public Library (2010–2017), national election events (Obama 2012, Trump 2016) will produce short-lived, event-month spikes in checkouts of books tied to presidential policy agendas, whereas local socio-economic events (e.g., May Day labor protests, 2016 homelessness assessment) generate more targeted spikes in topic-specific borrowing (housing, labor, social justice) within±1–2 months.

**Data Acquisition**

For the project we analyzed the checkout of books in the Seattle Public Library covering the years 2010-2017, this was provided as “Seattle\_Book\_Checkouts\_2010\_2017.csv”. The dataset consists of 11 columns containing bibliographic and temporal fields: [‘UsageClass’: Digital or Physical, ‘CheckoutType’: Type of checkout (renewal or in-person checkout), ‘MaterialType’: Format of item, ‘CheckoutYear: Calendar year of checkout', ‘CheckoutMonth’: Calendar month of checkout, ‘Checkouts’: Number of times book was borrowed that month, ‘Title’: Book title, ‘Creator’: Author, ‘Subjects’: Thematic subject labels , ‘Publisher’: Publish entity, ‘PublicationYear’]. The “Who” and “What” we analyze are individual checkout transactions linked to a certain book related to the relevant subject. This is done by categorizing the monthly topic totals by subject category, built from matched book titles. The dataset is anonymized, meaning it can under no circumstances be used to retrace an identification of a user. The data provided goes from January 2010 to December 2017. The original Seattle Library Dataset contains books from 2005 till present and is public for everyone to use.

**Methods (Concepts and Models explained):**

*1: Interrupted time series (ITS)*

Our research question is about event-timed attention, do two election events and two local events show a brief spike in specific topic borrowing? ITS is a short window regression around a given event that tests an instant jump in the event month and any trend change after the event, this method relates well to our research question. It holds two constants that could help reduce overfitting. 1) Checking on existing trends that were already underway, 2) By adding Fourier seasonality terms to regular seasonality (quiet vs busy months). To calculate a discrete change in monthly checkouts related to a certain topic and whether interest stayed afterward. Three event terms form the hard part of the ITS equation.

1) Pre-event trend (Beta1):

The pre-event trend measures if checkouts were already rising or lowering before the event, thus separating pre event-trends from the actual event that needs studying.

This is necessary because it captures the drift and we don’t mistake it for an event effect.

2) Level jump at the event (Beta2):

This checks if there is a jump of checkouts in the precise month that an event occurred, over and above the already predicted pre-trend and seasonality. It is calculated by asking, what should have happened if there had been no event, based of only the pre- event time and seasonality this outputs numbers. Now it checks the same but with looking at the event (Beta2 going from 0 to1), this could give a peak in checkouts suggesting causation or a decrease.

3) Post-event slope change (Beta3)

After the event we allow the last slope change, beta3 tells us whether attention persists, falls or stays flat after the event has happened.

We report for Beta2 and Beta3 point estimates in checkouts, 95% CI, p-value, and a one-line interpretation (“short-lived jump”, “no detectable change”, “post-event decay”, etc.). Also includes a plot (observed vs fitted) with a vertical event line.

*2: Fourier Seasonality*

The last two variables in our equation for ITS are the Fourier seasonal terms, Fourier his terms capture the natural yearly rhythm of checkouts in the library taking into account the busy and quieter months of the year. Fourier seasonality does exactly this by adding just two smooth terms sin(2πm/12) and cos(2πm/12); these trace the annual up and down pattern. This method reduces overfitting and causes the model to decrease its independent variables that are free to vary when estimating a parameter. Our Fourier terms do not need to be statically significant in addressing seasonal variety, they simply serve as the least invasive way to guard against seasonal confounding variables. Resulting in that the estimated event jump (beta\_2 ) and post-event trend change (beta\_3) are not artifacts of routine seasonal fluctuations or other noisy month-to-month variation, but reflect changes associated with the event itself.

*3: Ordinary Least Square Regression (OLS)*

OLS is a statistical method to calculate our coefficients (the pre-event trend, the immediate jump at the event and the post-event trend change) in the ITS model. The coefficients function as the definers of the relationship between our variables (time, event, seasonality) and the outcome (monthly checkouts). Monthly checkouts are presented as dots on a graph, OLS works by drawing a prediction line through those dots. Some dots may lay further from any given prediction line than others. The task of OLS is to find the line where the sum of the squared values of these distances or errors is as small as possible. Squaring is crucial to penalize errors harder creating a more sensitive fitted line, also it helps in ensuring that positive and negative values don’t cancel each other out.

After the model finds its best-fitting line, it looks at how much unexplained “noise” is left and uses that to judge how trustworthy each coefficient is. In statistics, noise is the random, unpredictable variation in your data that is not part of the relevant underlying pattern your model is trying to capture. The standard error reflects the typical wobble around an estimate: when it’s small, the estimate is precise; when it’s large, the estimate is unreliable, because the data are noisy or the sample is small. The t-statistic is just the coefficient divided by its standard error, so it tells you how big the effect is relative to the uncertainty. A large t means the signal stands out from the noise, a small t means it doesn’t. The p-value then translates that t into a probability: how likely it would be to see an effect at least this big purely by chance if, in truth, there were no effect at all. A small p-value suggests the effect is real rather than a fluke, a large p-value says the evidence is too weak to call.To judge the model as a whole, we look at how much variation it explains compared to a naive baseline that simply predicts the same average every month.. If the explained share is high, the regressors capture meaningful structure. If it is low, most variation remains unaccounted for. A complementary omnibus test, the F-statistic with its p-value, asks whether the full set of regressors improves predictions beyond a flat mean. A small p-value here indicates the model provides a real explanatory signal as a group, whereas a large p-value indicates the full specification is not clearly better than a constant.

*4: Keyword Categorization*

To measure event-timed attention, we turn raw circulation into four topic series: Obama, Trump, workers\_social\_justice, and Homelessness. We do this with a simple, transparent step: match curated keyword lists to each record’s librarian-assigned Subjects field. Subjects are comma-separated descriptors and already encode context. For every checkout, we split the Subjects string, normalize it, and ask whether any subject term contains a keyword from one of our lists. If yes, that checkout is assigned to the corresponding topic.

The keyword lists are built to capture the real discourse around each event. For the national cases (Obama/Trump), the lists go beyond names and include signature policies (“Affordable Care Act,” “border wall”), slogans and movements (“hope and change,” “MAGA”), and relevant issues tied to each administration (“Paris Agreement,” “January 6,” “fake news,” “racial justice”). For the local cases, we look for terms to the issues at hand: workers’ rights, unions, wages, and inequality for workers\_social\_justice; and for Homelessness, not only “homeless/unhoused” but also shelter terms, housing-crisis language, and co-occurring issues such as poverty and substance use that routinely appear in subject metadata for homelessness titles.

A natural concern is that some terms look broad (e.g., “inequality,” “drug abuse”). In our setting this is acceptable because matching happens against Subjects, not titles. In practice, this two-step filter first librarian categorization, then our keyword screen gives high coverage without pulling in unrelated material. It is fast, reproducible, and provides a clean bridge from concrete events to topic-specific borrowing, which we then analyze with ITS for immediate jumps and post-event trend changes. After automated classification, we manually review high frequency matches, top-contributing titles in spike months, and a random sample of edge cases to catch inconsistencies (e.g., ambiguous acronyms, homonyms). Any mislabels are corrected and the keyword lists adjusted accordingly.

**Workflow**

Firstly we began by setting up a Jupiter notebook using python. Our dataset was imported using Pandas, this allows us to analyze big data and to restructure big messy datasets. We also used matplotlib for the visualization of our charts. For our statistical analysis statsmodels.api was used. Using Python made it possible to rerun and end to end and made our decisions easily retraceable.

*1: Data acquisition & integrity checks*

We loaded our dataset with pandas.read\_csv, read\_csv is a function in Pandas it is used to read data from CSV files into Pandas DataFrame. First we needed to check the integrity of our data. We checked column types, missing values and uniqueness tests using [df.info](http://df.info)(), .isna().sum() and small random samples. To standardize our dataset we lowercased our columns, this should avoid mismatches, syntax errors and hardships during grouping. Missing values that were not relevant for our field of interest e.g. Creator and Publisher were tolerated, these classes were not essential for topic classification. However missing titles and subjects would prevent categorization into topics affecting the dependent variable, topic -based checkouts.

*2: Temporal Variable Creation*

Next we created a reliable time index that would be used as our x-axis. To achieve this a date column was made by combining CheckoutYear and CheckoutMonth into a single timestamp. This was relevant because it enabled us to link the data to specific event windows e.g. political elections.

*3: Text Normalization and tokenization*

The subject columns required the most thoughtful preprocessing. The records of our books were most of the time consisting of several subjects consisting of one comma-separated string . This caused a book to only have one topic, obviously the first topic in the string of subjects. Normalizing our Subject column was relevant because this field often has words starting with capital letters, punctuation and white spaces. This ensures that semantically identical strings are treated the same way e.g. Homelessness and homelessness. This consistency improves both measurement validity (higher recall and precision) and reproducibility, ensuring the same raw input consistently yields the same classifications for auditing and replication.

*5: Thematic Categorization*

Based on related literature custom keyword lists were made to assign books into thematic categories (Homelessness, Workers & Social justice, Obama, Trump). The selection of keywords was guided by recurring terms and phrases identified in (academic) written works, academic discussions, media discourse, and previous studies addressing these topics. In addition, through exploratory analysis of book titles and summaries, we tried to ensure that the chosen keywords accurately captured the nuances of each theme.

After sorting through the Subjects field (lowercasing, trimming, splitting on commas), we used a simple substring match: if any subject term contained a keyword from a theme’s list, that record was assigned to that theme Subjectcategroy. This rule-based approach is transparent and fast to check, lets us iterate keywords easily, and keeps per-topic counts understandable. We note a limitation: straightforward substring matching can miss synonyms and may include a few borderline cases. We address this by curating the lists carefully and reviewing edge cases during analysis.

*6: Aggregation*

The totals of the monthly checkouts were calculated per SubjectCategory using the built-in group function on Date and SubjectCategory. This meant we had transformed raw data to a time series suitable for more clear analysis. This allows us to easily visualize and model. The number of titles on a topic shifts over time (e.g., very few Trump books in 2012, many more by 2017), which can inflate total checkouts even if per-title interest is unchanged. We checked catalog continuity: month-by-month and within every ±12-month ITS window, across all subjects, there is no spike in newly added, topic-related books around the events. In other words, the observed checkout jumps are not driven by a sudden influx of relevant titles; they reflect genuine changes in reader attention.

*7: Event Definition and Windows*

We highlight event months on national level (Obama 2012 and Trump 2016 election) and local socio-economic events such as the May 1 protests in 2012 asking attention for worker rights and the Seattle survey about homelessness in 2016. We analyze short windows on each event of 6 months post and 6 months pre event. Most of the time longer window will result in more significant outcome, but we choose to keep close to our research question that focuses on a brief event-timed attention. We only keep ± 12 months as sensitivity checks to measure stability.

*7: Modeling*

Using a ±6-month window around the event, we estimate a segmented ITS (Interrupted time series) with centered time, an indicator for the event month, and a post-event slope term, plus two Fourier terms to control for monthly seasonality. A sine and a cosine of the month. Think of them as two smooth waves that together reproduce the typical yearly up-and-down pattern using only two parameters. In a 13-month window this is much safer than adding 11 separate month dummies (which would overfit and drain degrees of freedom). These Fourier terms absorb the broad seasonal swing without chasing random noise, so the estimated event jump and after-event trend are clearer and more reliable.The event month coefficient (post\_event) captures the instant level change attributable to the event, while (time\_after\_event) tests for a trend change afterward. We report ordinary least squares (OLS). We use OLS for the ITS because it directly estimates the event timed level and slope changes in a short window. For our results 4 plots will be outputted, these plots are time series aggregated to checkouts per month. These plots will be visualized in our repository on GitHub, under the name *plots\_local\_and\_national\_events.*

**TIMELINE / WORK PACKAGES**

| **WEEK** | **TASKS** | **RESPONSIBLE** | **MEETINGS** |
| --- | --- | --- | --- |
| 2 | 1. Data familiarization (going through the dataset) | Task 1: All members | meeting 1: Friday 12-09-25 (Online) |
| 3 | **!! PROJECT PROPOSAL DEADLINE FRIDAY 19-09-25 !!**    1. Coming up with a research question + thesis statement    2. Finding at least 3 sources (articles/ books / journals)    3. Working on the Project Proposal    4. Filtering dataset for the relevant categories (race, gender, religion, politics) | Task 1-3: All members    Task 4: Benjamin | meeting 1: Friday 19-09-25 (Online) |
| 4 | 1. Fine-tune research objectives by revising the RQ + thesis statement based on the professor’s feedback 2 local and 2 national events    2. Document potential ethical concerns, biases, and limitations of the dataset    3. Curate keywords lists for new topics (Worker rights, Homelessness, Obama, Trump) + review ambiguousness    4. Search for new sources on libraries + the American presedential elections of 2012 and 2016 and organize the findings    5. Set up a new Jypter Notebook + load the CSV file + validate dataset integrity    6. Combine year/month into timestamps, make columns (e.g. Subject column), python coding to remove punctuation and whitespace etc. + preprare data for ITS analysis | Task 1: Christabel    Task 2: Christabel + Vanessa    Task 3: All members    Task 4: Vanessa    Task 5 + 6: Benjamin | meeting 1: Thursday 25-09-25 (In class)    meeting 2: Friday 26-09-25 (Online) |
| 5 | 1. Write paragraphs for **the project overview** (the “why?” for the project + background information on local / national events) in the Workflow Document    2. ITS Modeling + Fourier Seasonality implement ITS regression with event indicators, calculate Bea coefficients, incorporate Fourier terms to acount for the seasonality aspect    3. Evaluate the short-term spikes and interpret early results    4. Draft report sections on findings | Task 1: Christabel + Vanessa    Task 2: Benjamin    Task 3: Benjamin + Christabel    Task 4: Vanessa + Christabel | meeting 1: Thursday 02-10-25 (In class)      meeting 2: Friday 03-10-25 (Online) |
| 6 | 1. Generate charts, document workflow reproducibility, finalize the code in Jupyter Notebook    2. Team review where we reflect on lessons learned, the workflow, potential improvements etc.      3. Making the PowerPoint  + making sure everything is uploaded to GitHub | Task 1: Benjamin    Task 2+3:  All members | meeting 1: Thursday 09-10-25 (In class)      meeting 2: Friday 10-10-25 (Online) |
| 7 | **!! PRESENTATION THURSDAY 16-10-25 !!** | All members | - |

**Challenges and Solutions.**

*1: Lexical Ambiguity*

Our dataset uses selected keywords manually added from literature and news. As synonyms, polysemy and homographs are represented in other texts or lists, they were also presented in our data. For example union as a labor movement, a political union or as in our dataset thousands of books related to the Soviet Union. In our data a key example was the word ‘Union’, representing the relation to social workers justice. As we did with all the related books to the specific subjects we checked the uniqueness of the titles and if they were not false positives. In the case of the social workers we saw a really high number of checkouts related to homelessness, that raised questions. When looking in the list of titles, we saw that of the roughly 1500 books 1200 were not related to social workers rights but to books on the Soviet Union. Although the Soviet Union and workers rights do share a special relation, for our project based in a Seattle environment we did not feel the need to add these books. Because ambiguity cannot be fully eliminated, we treat the categories as proxies for interest (not perfect counts), document the dictionary and exclusions, and interpret coefficients with that uncertainty in mind. Also our code categorized the keywords on strings, this means that for example the word ‘union’ also relates to ‘reunion’. For most of the cases this could be neglected because these unwanted relations were scarce. The only notable was the abbreviation of the ‘Affordable care act’ in our keyword lists represented as ‘ACA’. When looking at our data, words related to Obama were a lot more present than words related to Trump. After looking manually into the related titles and subjects to Obama we found that of the 2849 books related to Obama only 520 books really were related to Obama. This was because ‘ACA’ also represented the wildly present subject category ‘Juvenile vacation’, thus ‘ACA’ equaling the ‘aca’ in vacation.

*2: Seasonality*

Adding seasonality came from a mistake we made. As I was working on decreasing the p-value of social\_workers\_justice we looked up ways to reduce overfitting. It only wasn’t the problem with overfitting, the problem lay in checking that outputted the regressions of 2012-05 (the wrong month). Instead 2012-04 was the relevant month for the protests for social workers rights. Confused, because a spike was seen in the graph at the 2012-04 we searched ways to reduce the p-value. In this confusion we already forgot that the tested values from the first regression came from the 5th month of 2012. Only after adding Fourier terms, we intuitively checked regression for the 4th month resulting in way better results. After this mistake came to light, we checked regression with and without Fourier, result: no difference. Two reasons we still view Fourier as a *good* choice: (1) it’s a lightweight guard against calendar effects in a short window just two parameters instead of 11 month dummies, so it reduces overfitting risk without eating degrees of freedom; and (2) showing that results are robust to including/excluding Fourier actually strengthens credibility, because the event effect doesn’t hinge on a modeling tweak.

*3: Pulse*

Our original ITS treated the event as a one-month switch (post\_event: 0 → 1 at the event month). That’s appropriate if interest spikes only in that single month. But some interventions create multi-week salience (e.g. the homelessness survey fielded over several weeks). To reflect that, we used a pulse: a dummy that is 1 not only in the event month but also in the immediately following month (or however long the event realistically lasts). Conceptually, the pulse says, “treat these two months as the active treatment period,” and only after it ends do we let a new slope begin (time\_after\_pulse).

Statistically, pulsing helps when the true effect is spread over more than one month. A single-month dummy dilutes the signal half shows up in month 0, half in month +1 so each coefficient looks smaller and noisier, and p-values inflate. The pulse pools that signal into one parameter that aligns with the real duration, which raises precision (less attenuation) and often lowers the p-value. In our homelessness example, interest visibly covered roughly two months; modeling it as a two-month pulse captured that sustained bump more cleanly than a one-month step and produced a more decisive test. We still report both specifications for transparency, but the pulse is the better match when the event’s impact plausibly spans multiple weeks.

**Ethical Considerations**

To start, since the dataset reflects the borrowing patterns of library users rather than the entire Seattle population, it is important to acknowledge the limitations this introduces in terms of representativeness. Library visitors may differ from non-users in terms of, to name a few, differences in education, heritage background, access to technology, or political views, which can shape the types of books they might borrow. As such, we threaded carefully and discussed that the findings should not be interpreted as a factual, comprehensive reflection of community-wide attitudes, but rather as an indication of patterns among individuals who actively participate in public library services. Another ethical consideration involves the algorithmic and interpretative bias during data processing and analysis. As stated before, the team developed keywords used for the code, meaning that our own perspectives and biases inevitably influenced the classification process. Consequently, the results reflect not only the dataset itself but also our interpretation embedded in the coding decisions.

**Results**

Public events move books. Across our four cases, both national elections and local shocks produced statistically significant increases in topic-specific borrowing at the Seattle Public Library. The size and duration of those effects differed, with the 2016 presidential election standing out as the clear outlier.

Two elections, two footprints. Obama’s 2012 re-election shows an event-month jump of approximately +131 checkouts on Obama topics (p = .002), followed by a sharp decay of about 38 per month (p = .002). This reads as a flash of attention that faded quickly back to trend. By contrast, Trump’s 2016 victory produces an event-month surge of roughly +497 Trump-topic checkouts (p = .034) with no detectable post-event decline (slope p = .196), indicating that attention reset to a higher plateau and remained there over subsequent months. The 2016 Trump election was unique among the events studied for its ability to not only generate a massive spike in attention but also to sustain that interest over time. Local events, while capable of producing large spikes, tended to capture public attention for only a brief period. This shows that a highly consequential and unexpected national political event appears to have the greatest power to create a lasting shift in public reading habits.

Local shocks are sharp but short. Around the May Day 2012 mobilization, workers/social-justice borrowing rose by about +52 in the event month (p = .009) with no sustained change before or after. During the 2016 homelessness assessment, modeled as a two-month pulse, related titles show a short-lived level increase of approximately +417 during the survey window (p = .044) before returning to baseline. These patterns align with our hypothesis that high-stakes local issues channel attention into specific topics, even if the effect is transient.

What explains the Trump exception is the composition of demand. In 2017, over 12% of all Trump-topic checkouts came from a small cluster of books, led by Devil’s Bargain and The Dangerous Case of Donald Trump, and these works were predominantly critical or analytical in tone. The data therefore suggest that the post-2016 elevation was driven less by admiration than by interpretation and debate. In short, predictable national events tend to produce brief, election-timed spikes, whereas disruptive national shocks can generate large and durable shifts, and local socio-economic events trigger intense but short-horizon attention within the directly relevant topics.

**Documentation and Sustainability.**

Our repository can be found under the name “[Intro-To-DH-SA-Group-6-Seattle-Public-Library-s-Open-Checkout-Data-2010-2017-](https://github.com/VanessaTerl/Intro-To-DH-SA-Group-6-Seattle-Public-Library-s-Open-Checkout-Data-2010-2017-)”

We classify national events (e.g. Obama 2012, Trump 2016) under the National category and local socio-economic events (e.g. May Day labor protests, 2016 homelessness assessment) under the Local category. All borrowing time series, models, and figures are produced separately for these two groups so national shocks are not mixed with local dynamics.

Changes & updates. If we add or revise events, we (1) place them in the correct group (National or Local), (2) re-run the same pipeline (classification -> ITS -> figures), and (3) record what changed in a short changelog note (date, event added/edited, reason). This keeps results comparable over time while making updates transparent.

**Reflection**

Reflecting on our workflow, the process proved both effective and educational. Using Python allowed us to build a transparent and reproducible pipeline, which made it possible to trace analytical steps and refine our decisions throughout the project. This structure strengthened our confidence in the results while also revealing areas for improvement. In terms of the latter, we have learned quite an important lesson concerning lexical ambiguity, such as the misclassification of “*Union*,” which highlighted the limitations of rule-based keyword matching. Future projects could therefore benefit from incorporating more advanced natural language processing methods, including lemmatization and semantic similarity modeling, to enhance the accuracy of topic classification. Additionally, accounting for seasonality and event timing through Fourier terms showed us the importance of aligning statistical methods with real-world patterns. It reminded us that models should not only fit the data mathematically but also make sense conceptually, guaranteeing that observed data truly reflects event-related changes rather than regular seasonal trends. Another valuable insight emerged when we adjusted our event modeling by introducing the pulse variable, which more accurately captured attention patterns that extended beyond a single month. This adjustment improved both the precision and realism of our results.

Accordingly, the project emphasized the importance of ethical reflection, as the dataset represents only library users rather than the entire Seattle population. Our findings should therefore be interpreted as patterns of engagement within a specific group rather than as a universal reflection of public behavior. Overall, we enjoyed working on a project in the field of the Digital Humanities and appreciate the opportunity to combine computational methods with social and cultural inquiry. This interdisciplinary approach allowed us to explore how data-driven techniques can shed light on patterns of public engagement, while also deepening our understanding of the broader societal context behind them.

**Literature / Works Cited**

Arseneau, R., & Dodd, H. (n.d.). *View of Vol. 7 No. 1 (2025): Emerging Library & Information Perspectives*. <https://ojs.lib.uwo.ca/index.php/elip/issue/view/2030/820>

Bush, E. (2016, April 28). *May Day in Seattle has featured violence after peaceful marches*. The Seattle Times.<https://www.seattletimes.com/seattle-news/may-day-in-seattle-4-years-of-violence-after-peaceful-marches/>

*Clinton becomes first woman presidential nominee of major U.S. party*. (2016, July 27). [Video]. NBC News. <https://www.nbcnews.com/storyline/2016-conventions/hillary-clinton-becomes-first-female-nominee-major-u-s-political-n617406>

City of Seattle. (2016).  *City of Seattle 2016 Homeless Needs Assessment.* Applied Survey Research. <https://www.seattle.gov/documents/Departments/HumanServices/CDBG/CityOfSeattle2016-HomelessNeedsAssessment.pdf>

Devereaux, R. (2012, May 2). *Occupy May Day protests across US as activists and unions link up*. The Guardian; The Guardian. <https://www.theguardian.com/world/2012/may/02/occupy-may-day-protests>

*Election 2016: Transcript - Donald Trump announces presidential candidacy*. (2015, June 16). CBS News. <https://www.cbsnews.com/news/transcript-donald-trump-announces-his-presidential-candidacy/>

PBS NewsHour. (2020, September 26). *Obama vs. Romney: The first 2012 presidential debate* [Video]. YouTube. <https://www.youtube.com/watch?v=KfaBRyCKRhk>

*Post-debate, Clinton takes the lead.* (2016, October 4). CNN Politics. <https://edition.cnn.com/2016/10/03/politics/hillary-clinton-donald-trump-presidential-polls/index.html>

Rock, L. (2016, January 31). *Sleeping rough in Seattle: homeless crisis exposes dark side of hi-tech city* (L. Rock, Ed.). The Guardian. <https://www.theguardian.com/us-news/2016/jan/31/sleeping-rough-in-seattle-homeless-crisis-exposes-dark-side-of-affluent-microsoft-city>

Sierpe, E. & Southern Connecticut State University. (2017). The Election of Donald Trump to the Presidency and the Crisis of Liberalism in Librarianship: The Need to Reconsider the Social Function of the Library and its Role in Critical Information Literacy and Political Education in Response to the Rise of Alt-right Fascism in the United States. In *Journal of Radical Librarianship* (Vols. 3–3, pp. 65–75). <https://journal.radicallibrarianship.org/index.php/journal/article/view/23/35>

*The Haymarket Affair, 1886 | Gilder Lehrman Institute of American History*. (n.d.). <https://www.gilderlehrman.org/history-resources/spotlight-primary-source/haymarket-affair-1886>