

Alanna Bartolini

MAT259

Professor George Legrady

January 15, 2019

Project 1: MySQL Report on Charles Dickens

Context: Because I'm writing a dissertation chapter about Charles Dickens, this project seemed a good opportunity to investigate how the Seattle Public Library collects and circulates Dickens's works among a general reading public.

Questions:

1. What are the SPL's Dickens-related holdings? And;
2. What can we learn about the public's interest and engagement in of these items?

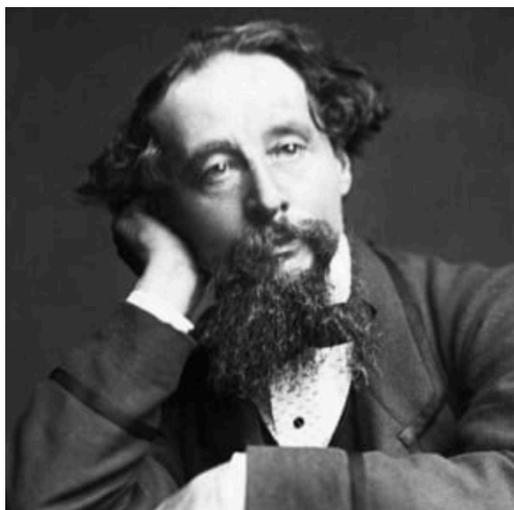


Figure 1: Charles Dickens

Difficulty:

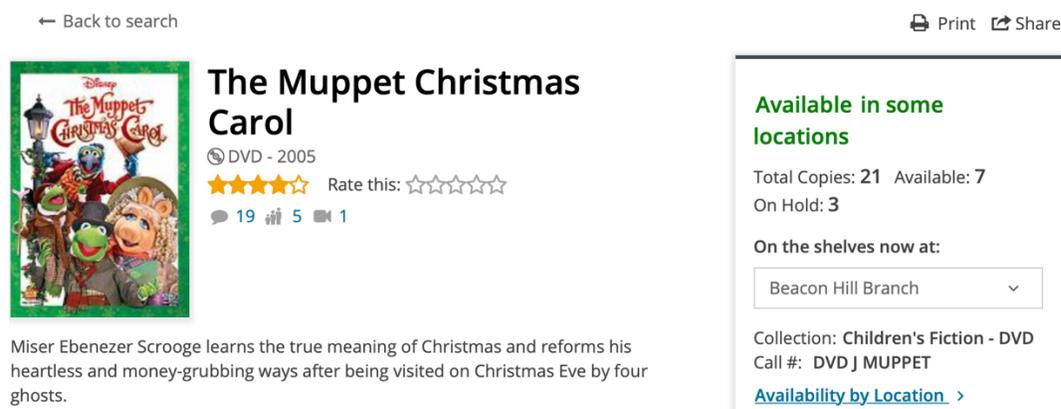
I immediately encountered an obstacle: there is no way to search the database by author's name. To a literature PhD, this seems unbelievable. The conspicuous absence of one of the most fundamentally important pieces of metadata led to the realization that, simply by virtue of participating in different disciplines, all of us will approach a database with certain biases that are invisible to us.

I thought to myself, "How can this be possible? It's a library. It's full of books. All books have authors. There must be some mistake." And yet, from the data science perspective, it is not true that all books have authors – some are anthologies compiled by editors, for example. How do we decide to enter that metadata into the "Author" column, then? The name of every author? That could be in the hundreds sometimes. The name of the editor, then? But they are not authors. Authorship becomes a slippery concept when you examine it closely, and it often lacks the neat, tidy edges that make for good metadata – and when some parts fit and others do not, it is no longer

possible to call it a “category” at all. It was also a misconception on my part to assume that books are the dominant media type in a library, and would therefore have proportionate dominance in the decisions about the metadata that describes – and indeed must describe – everything in the library’s holdings. Our own preconceptions shape our inquiries, how we get to the questions we ask, and how we look for them - and therefore, possibly, the results.

Solution:

I considered scrapping the question and simply focusing on a few of the more popular novels, but again “most popular” can only be proven by looking at the whole picture, so I went back to the SPL website and searched through their collection. The website says they have 544 items with Dickens credited as the author. However, as I scrolled through and saw “A Muppet Christmas Carol” listed as one of these items, it was clear that I’d need to narrow things down a little more. I noticed that the Call Numbers for the more literary items tended to be “FIC DICKENS” or “E DICKENS” depending on the collection, while the Call Numbers for the Muppet Christmas Carol were “DVD J MUPPET” so I felt safe that this distinction would help filter out the items I did not want.



← Back to search Print Share



The Muppet Christmas Carol
 DVD - 2005
 ★★★★★ Rate this: ☆☆☆☆☆
 19 5 1

Miser Ebenezer Scrooge learns the true meaning of Christmas and reforms his heartless and money-grubbing ways after being visited on Christmas Eve by four ghosts.

Available in some locations

Total Copies: 21 Available: 7
 On Hold: 3

On the shelves now at:

Beacon Hill Branch ▾

Collection: Children's Fiction - DVD
 Call #: DVD J MUPPET
[Availability by Location >](#)

Figure 2: Muppet Christmas Carol – Call Number DVD J MUPPET

But then I realized that this was true for all the film adaptations, which I was interested in looking at, like the excellent 2009 BBC adaptation of *Bleak House*:





Bleak House
 ★★★★★

 **DVD - 2009**
DVD BLEAK H
All copies in use [View details](#)
 Holds: 5 on 3 copies

Figure 3: DVD *Bleak House*

I also realized that if I wanted to include the BBC adaptation, I might actually have to include the Muppet version as well. A significant percentage of Dickens's work centers around children, and many of his works are, if not specifically directed at them, certainly suitable for children. Many read *Oliver Twist* or *David Copperfield* or *A Christmas Carol* in elementary school, and many K-12 students will read *Great Expectations* or *A Tale of Two Cities* as part of middle or high school. This means that there is no way to determine how much of the novels' circulation is happening among those populations. This also means that I cannot excise the film adaptations for children from the dataset of DVD materials, and subjectively draw some arbitrary line that I cannot really define between the ones for adults and the ones for kids.

I decided to put a pin in the DVD/CD/adapted materials issue and focus on getting the information I wanted about the items the library defines as having Dickens as the author. At first I only queried `Callnumber LIKE '%FIC DICKENS%'` but thought I was missing entries from other collections, so I went back to the library website to check and there were other ones: E DICKENS, which I knew about, but I was sure there were others I was missing. I finished by using `Callnumber LIKE '%DICKENS%'` instead. This worked well enough, and gave me 39 more findings.

Adjusted Query Results

Query	SELECT callNumber, itemType, title, COUNT(bibnumber) AS Counts FROM spl_2016.inraw WHERE CALLNUMBER like '%FIC DICKENS%' GROUP BY callNumber, itemtype, title Order BY counts DESC	SELECT callNumber, itemType, title, COUNT(bibnumber) AS Counts FROM spl_2016.inraw WHERE CALLNUMBER like '%DICKENS%' GROUP BY callNumber, itemtype, title Order BY counts DESC
Results	139	176

Problems:

It presented some problems though. The first is that it returned more than just the novels – it returned things like Dickens's short stories, collected short fictions, and novellas, which I was not interested in examining. It was also returning all the different Itemtypes – which I wanted – but it also included a DVD titled “Dickensian,” a modern term used as a catchall descriptor of poverty and poor working and living conditions, and not something I want in my dataset. I was getting a lot of different media types – and that's interesting because it may be interesting in future to come assess the proportion of Dickens DVD checkouts to Dickens novel checkouts. But first, the novels.

Final Solution:

Because I knew exactly which titles I wanted, and how to ask for those titles in book format, and could use the Callnumber specification as a workaround for my “Author” search, it seemed best to use the SumCase approach:

```

1 • SELECT year(cout),
2   sum( case when title = 'Great Expectations' then 1 else 0 end) as Great_Expectations,
3   sum( case when title LIKE '%Christmas Carol%' then 1 else 0 end) as Christmas_Carol,
4   sum( case when title LIKE '%Barnaby Rudge%' then 1 else 0 end) as Barnaby_Rudge,
5   sum( case when title = 'Oliver Twist' then 1 else 0 end) as Oliver_Twist,
6   sum( case when title LIKE '%Hard Times%' then 1 else 0 end) as Hard_Times,
7   sum( case when title = 'Bleak House' then 1 else 0 end) as Bleak_House,
8   sum( case when title LIKE '%Pickwick Papers%' then 1 else 0 end) as Pickwick_Papers,
9   sum( case when title = 'David Copperfield' then 1 else 0 end) as David_Copperfield,
10  sum( case when title LIKE '%Old Curiosity Shop%' then 1 else 0 end) as Old_Curiosity_Shop,
11  sum( case when title = 'Our Mutual Friend' then 1 else 0 end) as OMF,
12  sum( case when title LIKE '%Nicholas%' then 1 else 0 end) as Nicholas_Nickelby,
13  sum( case when title LIKE '%Little Dorrit%' then 1 else 0 end) as Little_Dorrit,
14  sum( case when title LIKE '%Dombey%' then 1 else 0 end) as Dombey_Son,
15  sum( case when title LIKE '%Martin Chuzzlewit%' then 1 else 0 end) as Martin_Chuzzlewit,
16  sum( case when title LIKE '%Tale of Two Cities%' then 1 else 0 end) as Tale_of_Two_Cities
17 FROM spl_2016.outraw WHERE (itemtype Like '%bk%') AND (callnumber LIKE '%dickens%') AND (year(cout) > 1970)
18 GROUP BY year(cout)

```

Figure 4: Sum Case query

I originally had more `title = 'actual title'` but three columns come back empty, so I had to modify the lines for *Little Dorrit*, *Nicholas Nickelby*, and *Dombey & Son*. This was an easy fix with *Little Dorrit*, and *Dombey & Son* gave me some trouble with the ampersand, but `'%Dombey%'` worked in the end. One odd thing I cannot explain, though, is that in the case of *Nicholas Nickelby*, `LIKE '%Nickelby%'` did not work, but `'%Nicholas%'` did.

The Results:

Dickens all novel COUT by year FULL RESULTS															
year(cout)	Great_Expectations	Christmas_Carol	Barnaby_Rudge	Oliver_Twist	Hard_Times	Bleak_House	Pickwick_Papers	David_Copperfield	Old_Curiosity_Shop	OMF	Nicholas_Nickelby	Little_Dorrit	Dombey_Son	Martin_Chuzzlewit	Tale_of_Two_Cities
2006	123	179	26	39	31	95	2	69	28	15	29	22	13	15	177
2007	123	171	25	50	24	55	0	72	38	14	24	11	15	16	192
2008	189	190	26	83	18	39	3	82	33	15	31	21	14	15	214
2009	177	265	33	132	26	35	5	84	50	21	33	125	18	27	249
2010	158	242	27	137	24	55	5	78	26	25	30	66	30	13	170
2011	133	204	25	151	27	73	5	54	16	15	21	44	6	18	166
2012	97	191	26	139	35	61	40	70	38	36	50	25	6	35	135
2013	105	196	15	177	25	54	28	59	36	37	26	9	8	20	142
2014	129	170	13	127	28	52	30	55	35	35	13	7	7	17	123
2015	92	164	12	105	12	47	35	41	16	22	17	17	9	13	139
2016	112	163	10	124	9	24	16	57	14	18	8	5	3	11	79
2017	77	135	10	69	8	25	2	58	10	14	8	8	2	11	91
2018	37	130	11	43	6	25	4	41	3	0	9	16	4	12	56
2019	47	107	22	44	5	23	12	40	11	8	22	10	0	10	78
2020	3	4	0	4	0	1	0	1	1	0	0	1	0	0	2

Figure 5: Full Results of Dickens Novel Checkouts 2006 – 2020

I then did some basic calculations to get total checkout counts per year, average yearly checkout per novel, and to rank them (end column) on a scale of 1 to 13, with 1 being the most popular year for Dickens's novels to be checked out.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	year(cout	Great_Exp	Christmas	Barnaby_Ru	Oliver_Twist	Hard_Times	Bleak_Hous	Pickwick_P	David_Copp	Old_Curiosit	OMF	Nicholas_Ni	Little_Dorrit	Dombey_So	Martin_Ch	Tale_of_Two		
2	2006	123	179	26	39	31	95	2	69	28	15	29	22	13	15	177	863	6
3	2007	123	171	25	50	24	55	0	72	38	14	24	11	15	16	192	830	8
4	2008	189	190	26	83	18	39	3	82	33	15	31	21	14	15	214	973	4
5	2009	177	265	33	132	26	35	5	84	50	21	33	125	18	27	249	1280	1
6	2010	158	242	27	137	24	55	5	78	26	25	30	66	30	13	170	1086	2
7	2011	133	204	25	151	27	73	5	54	16	15	21	44	6	18	166	958	5
8	2012	97	191	26	139	35	61	40	70	38	36	50	25	6	35	135	984	3
9	2013	105	196	15	177	25	54	28	59	36	37	26	9	8	20	142	937	5
10	2014	129	170	13	127	28	52	30	55	35	35	13	7	7	17	123	841	7
11	2015	92	164	12	105	12	47	35	41	16	22	17	17	9	13	139	741	9
12	2016	112	163	10	124	9	24	16	57	14	18	8	5	3	11	79	653	10
13	2017	77	135	10	69	8	25	2	58	10	14	8	8	2	11	91	528	11
14	2018	37	130	11	43	6	25	4	41	3	0	9	16	4	12	56	397	13
15	2019	47	107	22	44	5	23	12	40	11	8	22	10	10	10	78	439	12
16	average	106.8	167.4	18.7333333	94.9333333	18.5333333	44.2666667	12.4666667	57.4	23.6666667	18.3333333	22.9285714	27.5714286	9.64285714	16.6428571	143.642857	11510	
17																		
18																		
19	2020	3	4	0	4	0	1	0	1	1	0	0	1	0	0	2		
20																		
21																		

Figure 6: Full Results of Dickens Novel Checkouts 2006 – 2019 with Calculations

Because we are only two weeks into January 2020, I moved the row for that year out of my dataset because the low numbers would produce artificially low results for the average rate of checkout per title since 2006. I left it in the picture, however, because it is interesting to see what books are being checked out since the new year.

Findings and Interpretations:

Finding 1: In 2009 and 2010, there is a spike in the number of checkouts. Furthermore, the top 5 years are, in descending order: 2009, 2010, 2012, 2008, and 2011.

Interpretation: This is likely due to a renewal of interest caused by a proliferation of film adaptations in the period, which were popularized in mainstream media.

Bearing in mind that correlation does not prove causation, and further research would be needed to solidify these claims, I present some possible connections.

In 2009 and 2010, there appears to be a large spike in the popularity of *A Christmas Carol*. A cursory Google search reveals that the film adaptation starring Jim Carrey as Scrooge was released in late 2009.

In 2005, an excellent adaptation of *Bleak House* was released and got a lot of media attention for its high production value and all-star cast. Then, in the following year (2006), the number of novel checkouts for *Bleak House* is more than double the average for other years in a thirteen year period.

In late October of 2008, an award-winning BBC miniseries adaptation of *Little Dorrit* began to air. It seems odd that the book's checkouts are low in 2008, with only 21 (not significantly different from the prior two years, with 22 checkouts in 2006, and 11 checkouts in 2007). But, when one takes into account the fact that the show ran for fourteen weeks, and only finished airing in January of 2009, the huge increase to 125 checkouts makes a lot more sense. The numbers taper off at a steady increment until 2016, when there is a small but nevertheless noticeable spike again.

It could be explained, though, because the actress who played Dorrit in the BBC adaptation, Claire Foy, was highly critically acclaimed and received media attention. Then, in 2016, she also starred in the highly anticipated tv adaptation of Hilary Mantel’s ultra-successful novel, *Wolf Hall*, for which the actress also won numerous high-profile accolades. This leads to the hypothesis that there is the possibility of a causal link between the correlation; it is not impossible that some people who enjoy video adaptations of books might have seen the same person in lead roles in film versions of novels they wanted to read, and been motivated to check those novels out.

	year(cout)	Little_Dorrit	Tale_of_Two_Cities
	2006	22	177
	2007	11	192
	2008	21	214
	2009	125	249
	2010	66	170
	2011	44	166
	2012	25	135
	2013	9	142
	2014	7	123
	2015	17	139
	2016	5	79
	2017	8	91
	2018	16	56
	2019	10	78
	average	27.5714286	143.642857

Figure 7: Detail of data concerning *Little Dorrit* and *A Tale of Two Cities*

Finding 2: From 2015 onwards, we see a steep decline in the number of checkouts every year.

Interpretation: It is likely that, some time around 2015, the library rolled out the function that allows users to “borrow” e-books by downloading them onto their e-readers (Kindles, Kobos, etc.) for a certain period of time. This would account for a downturn in the rate of checkouts of physical books. However, until we can track the rates of checkouts for the digital items, we cannot

determine if there really is a significant change in the rate at which people are reading Dickens novels.

Finding 3: *A Tale of Two Cities* is the most checked out novel, aside from *A Christmas Carol* – and in an unusually consistent way.

Interpretation: The consistency of the numbers, as well as the drop-off that happens in 2016 onwards that then remains fairly constant also, leads me to believe this book is very likely part of a Seattle public school curriculum. It would also explain an outlier I noticed earlier in my research demonstrating that the SPL has 97 copies of the same bibnumber version of the book, and several more from other bibnumbers for the same title. It would be logical for a public institution like a library to help support the curricula of other local public institutions like schools. I looked up the Seattle Public School website to see if there was any information, but nothing specific came up.

Other concluding remarks: One interesting side finding was that the SPL has over 1400 community-generated lists related to Dickens. It would be interesting in future to do SPL inquiries having to do with these community-generated lists, and see if that is something the general reading public responds to. It would seem, for now, Dickens's popularity is alive and well in Seattle.