

LexisNexis Appeals Data Review

Citations

Version 1.0, March 28, 2020

Tom Balmat, Duke University

To date, two primary data sets have been received from LexisNexis (LN): set A (3/13/2019) and set B (12/3/2019). Citations were received with data set A only. When asked whether additional or revised citations would be included with data set B, LN stated that no changes were made to the citations file as a result of the extract definition of data set B. The following summarizes the citation file and several comparisons made between citations related to cases appearing in data sets A and B.

List of Sections

1	Citation file contents	2
2	Evaluation of citing and cited LNI	4
3	Distribution of citations per citing case	5
4	Distribution of citations per cited case	6
5	Distribution of citations by court and year	8
6	Citations with LNI that do not appear in case header records	10
7	Duplication of citations by reporter	11
8	Distribution of the difference in number of citations by Shepard's treatment letter	12

1 Citation file contents

Citations were supplied in a comma separated (CSV) file with each record having the intended format:

- Column 1: Citing LNI
- Column 2: Cited LNI
- Column 3: Citation
- Column 4: Shepards treatment

Example records with valid format appear in table 1. 133 records have citation entries that contain one or more commas which, in a CSV context, cause additional columns to be inferred during import and generate an error. Example records with commas in the citation field appear in table 2. Some records have commas surrounded by double quotes, as in the final record of table 2. To correct erroneously formatted citation records prior to import into the project SQL database, double quotes were replaced with an exclamation point and each citation entry was delimited by double quotes. With this format, commas within fields are treated as text and not field delimiters, giving records a valid CSV format.

Table 1: Example citation records with valid format

Citing LNI	Cited LNI	Citation	Shepard's treatment
4M87-WC80-TVVRV-K21W-00000-00	43D0-X0T0-0004-S09B-00000-00	2001 U.S. Claims LEXIS 107	f
3S4X-CYF0-003B-P400-00000-00	3S4X-1YH0-008H-V23G-00000-00	970 F.2d 338	f
5BCM-RH61-F04K-P1BN-00000-00	4T8S-B740-TX4N-G0WY-00000-00	539 F.3d 327	d
52NM-DKG1-652R-11CY-00000-00	3S4X-FC80-0039-P0VX-00000-00	1985 U.S. App. LEXIS 22611	f
5CSF-P681-F04K-V3H3-00000-00	5C2D-M901-F04K-V0F4-00000-00	757 F.3d 885	e
5SNC-JKS1-F60C-X1RC-00000-00	3S4X-14C0-006F-M32J-00000-00	1996 U.S. App. LEXIS 22201	f
5CBY-6WH1-JCNH-X00V-00000-00	3S4X-6J80-003B-S169-00000-00	451 U.S. 527	f
3S4X-BR90-001B-K2MH-00000-00	3S4X-J2B0-003B-42RB-00000-00	93 L. Ed. 2d 800	f
55GJ-PW41-F04K-N11H-00000-00	4CM8-CD60-004C-2004-00000-00	124 S. Ct. 2333	f
5GN8-1YT1-F04K-N1D4-00000-00	5D6V-F471-F04K-P0HV-00000-00	768 F.3d 524	f
5MFN-XV51-F04K-S07G-00000-00	3S4X-BVX0-003B-4338-00000-00	1989 U.S. LEXIS 1568	d
5DW5-7CP1-F04K-Y12R-00000-00	3RRT-8RV0-003G-1194-00000-00	633 A.2d 834	f
3S5N-KMC0-0038-X48X-00000-00	3S4X-6WH0-008H-V230-00000-00	951 F.2d 1109	f
58HV-0FG1-F04K-X06K-00000-00	3YBB-JCF0-003B-C01R-00000-00	103 L. Ed. 2d 685	d
3S4W-YGN0-0039-M198-00000-00	3S4X-B0R0-003B-S3RS-00000-00	96 S. Ct. 1075	d
5FM2-CSS1-F04K-J00C-00000-00	3S4X-J640-00B1-D131-00000-00	3 Wage & Hour Cas. 2d (BNA) 1405	f
7YM7-9JJ1-2RHT-103S-00000-00	3T4K-SJD0-0038-X1YF-00000-00	73 Empl. Prac. Dec. (CCH) P45436	f
5SST-RX41-K054-G1RC-00000-00	3S4X-HT80-00B1-D3NM-00000-00	109 F.3d 578	f
5RVK-2P91-F528-G303-00000-00	3S4X-3FN0-003B-S3TN-00000-00	80 L. Ed. 2d 674	f
3S4W-TMB0-003B-G0FS-00000-00	3S4X-7590-003B-S174-00000-00	65 L. Ed. 2d 115	d

Table 2: Example citation records with comma in citation field

Citing LNI	Cited LNI	Citation	Shepard's treatment
3S4X-4310-0039-M2XJ-00000-00	3S4X-0X80-0039-X1TB-00000-00	Md. Code Ann., Health Occ. sec. 7-101 (h)	f
3TC7-NDR0-0038-X0BH-00000-00	3S4X-HDB0-003B-P4ST-00000-00	Wash. Cty., Md. Pub. Loc. L. 1930 sec. 572 (1941C526)(i)	f
3S4X-J2P0-003B-P104-00000-00	3S4X-H480-008H-V1HX-00000-00	Copyrights Class TX Minnesota Multiphasic Personality Schedule, The	f
3S4X-17P0-0039-M00R-00000-00	3S4X-2490-0039-M0J0-00000-00	Worc. Cty., Md. Pub. Loc. L. 1930 sec. 193	d
3S4X-2J60-001B-K469-00000-00	3S4X-3F10-003B-G4S4-00000-00	Kan. Stat. Ann. sec. 66-1,112g	f
3S4W-YSY0-0039-P47G-00000-00	3S4X-1RS0-003B-G2KJ-00000-00	Neb. Rev. Stat. sec. 81-8,110.13	d
3S4X-35G0-003B-52YN-00000-00	3S4X-JNV0-0039-P02C-00000-00	Balt. Cty., Md. Cty. Code 1978 2 (22-74)	f
3S4X-5RB0-0039-W25N-00000-00	3S4W-VGK0-0039-M08C-00000-00	Trademarks Bandido by Cervantes, El	d
3S4W-YSH0-006F-M3TY-00000-00	3S4X-2230-008H-V2NH-00000-00	Md. Code Ann., Health Occ. sec. 14-316 (c)	f
3S4X-77J0-003B-53H9-00000-00	3S4W-WHD0-0039-M55G-00000-00	Trademarks Dabs Creation, A	d
3S4X-0B30-0039-M0T3-00000-00	3S4X-15M0-0039-Y0SX-00000-00	Md. Code Ann., Fin. Inst. sec. 6-215 (d)	e
3S4X-5HG0-003B-50JP-00000-00	3S4V-S420-0054-8054-00000-00	Copyrights Class PA Jungle Killer, The	f
3S4X-5950-0039-W1N8-00000-00	3S4V-P6H0-003B-20JV-00000-00	Wicomico Cty., Md. Pub. Loc. L. 1959 sec. 23E	f
3S4X-1NT0-003B-P4CG-00000-00	3S4X-HDB0-003B-P4ST-00000-00	Wash. Cty., Md. Pub. Loc. L. 1930 sec. 572 (1941C526)(i)	f
3S4X-2BN0-003B-P2W7-00000-00	3S4X-5V90-003B-P34Y-00000-00	Balt. Cty., Md. Pub. Loc. L. 1948 sec. 399	d
3S4X-KJS0-0039-W0B6-00000-00	3S4W-XP90-0039-M4CC-00000-00	Pr. Geo. Cty., Md. Pub. Loc. L. 1953 sec. 1117	d
4401-R1M0-0038-X1R7-00000-00	3S4X-5V90-003B-P34Y-00000-00	Balt. Cty., Md. Pub. Loc. L. 1948 sec. 399	ca
3S4X-B4R0-003B-507G-00000-00	3S4X-HRM0-0039-P1Y4-00000-00	Harford Cty., Md. Pub. Loc. L. 1965 sec. 214	d
3S4X-40K0-003B-G262-00000-00	3S4X-0X80-0039-X1TB-00000-00	Md. Code Ann., Health Occ. sec. 7-101 (h)	f
3S4W-T9N0-0039-M0FN-00000-00	3S4W-WHD0-0039-M55G-00000-00	Trademarks Dabs Creation, A	f
3S4X-2NV0-006F-M138-00000-00	3S4X-8XW0-0039-P2SX-00000-00	Trademarks Better Jacket-Thru Science“, ”A	d

2 Evaluation of citing and cited LNI

LNIs are formatted strings that uniquely identify citing and cited cases. A valid LNI is composed of six dash-separated fields. Individual field formats are listed in table 3. An example, valid, LNI is 45CC-WT50-FM72-200M-00000-04.

Observations:

- All citing and cited LNIs are properly formatted as required by the patterns in table 3
- A total of 3,731,213 citations are recorded in the supplied citations file
- 325,307 unique citing LNIs are recorded. Of these:
 - 340,508 appear in case header records of data set A
 - 352,282 appear in case header records of data set B
- 327,330 unique LNIs are cited. Of these:
 - 176,493 appear in case header records of data set A
 - 179,333 appear in case header records of data set B

Table 3: Valid LNI format, as defined by LexisNexis

Field	Contents
1	Four positions, one character from {3, 4, 5, 7, 8} followed by three characters from set c*
2	Four positions, three characters from set c* followed by one character from {0, 1, 2} followed by three characters from set c*
3	Four positions, one character from {0, 2, 6, D, F, J, K, N, T, R, Y} followed by three characters from set c*
4	Four positions, one character from set c* followed by one character from {0, 1, 2, 3, 4, 5} followed by two characters from set c*
5	Five positions, each containing 0
6	Two positions, a 0 followed by one character from {0, 1, 2, 3, 4, 5}

* c = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, B, C, D, F, G, H, J, K, M, N, P, R, S, T, V, W, X, Y}

3 Distribution of citations per citing case

Figure 1 shows the distribution of the number of cited cases per citing case. There exist 352,207 unique citing LNIs in the citations file. Of these, 340,508 appear in case records of data set A, while 352,282 occur in case records of data set B. Citing cases used in figure 1 are limited to those with an LNI appearing in case header records. In the plot, symbol x indicates observation frequencies from data set A and symbol o indicates those from data set B.

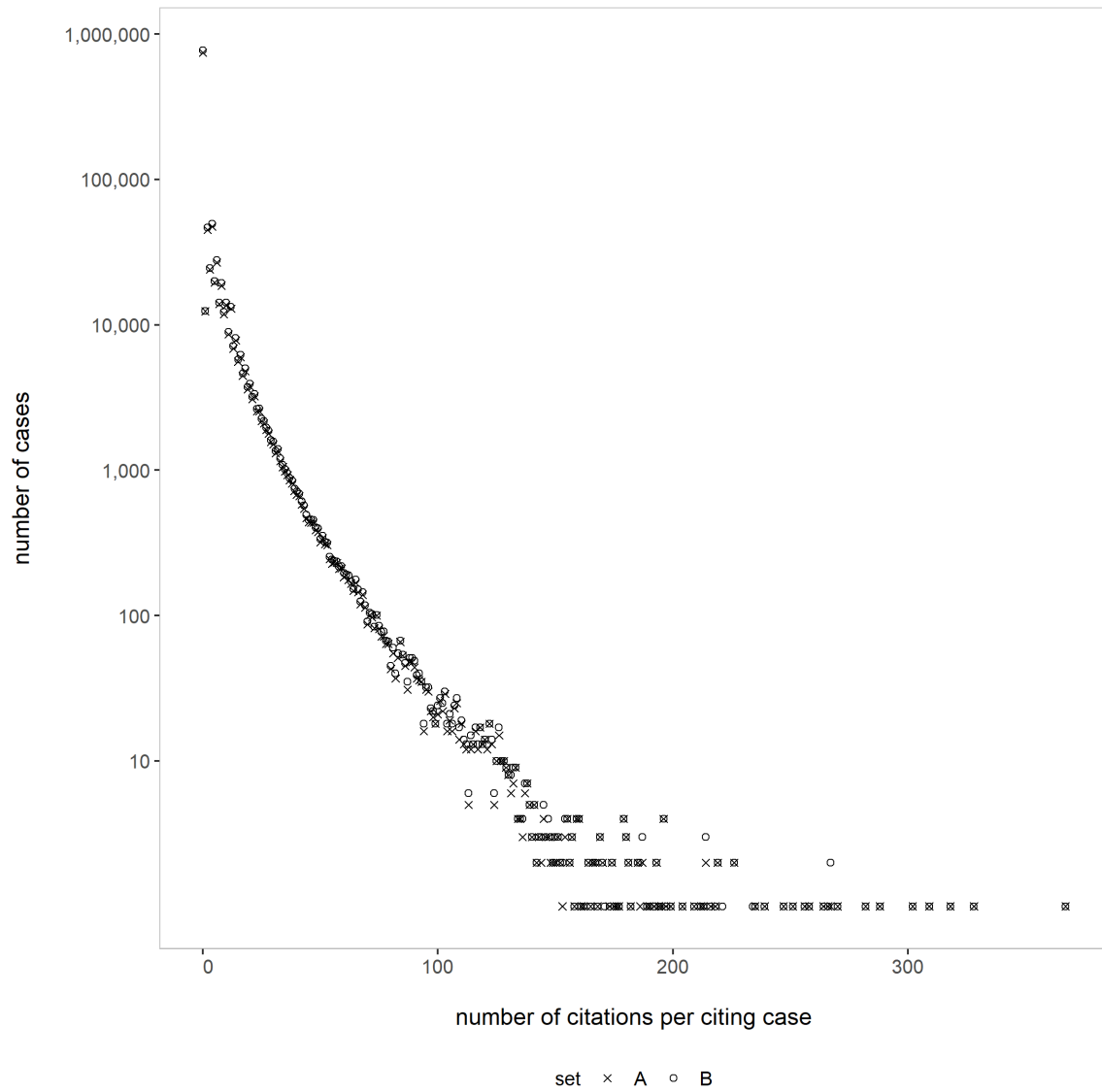


Figure 1: Distribution of citations per citing case

4 Distribution of citations per cited case

Figure 2 shows the distribution of the number of citing cases per cited case. There exist 327,330 unique cited LNIs in the citations file. Of these, 176,493 appear in case records of data set A, while 179,333 occur in case records of data set B. Cited cases used in figure 2 are limited to those with an LNI appearing in case header records. In the plot, symbol x indicates observation frequencies from data set A and symbol o indicates those from data set B. The extreme point on the x-axis, corresponding to a citation frequency of 5,364, is case 4349-83B0-0038-X066-00000-00 (John Hardy Rose v. Warden of Central Prison, Raleigh; 4th Circuit; 5/24/2001). Figure 3 excludes this case.

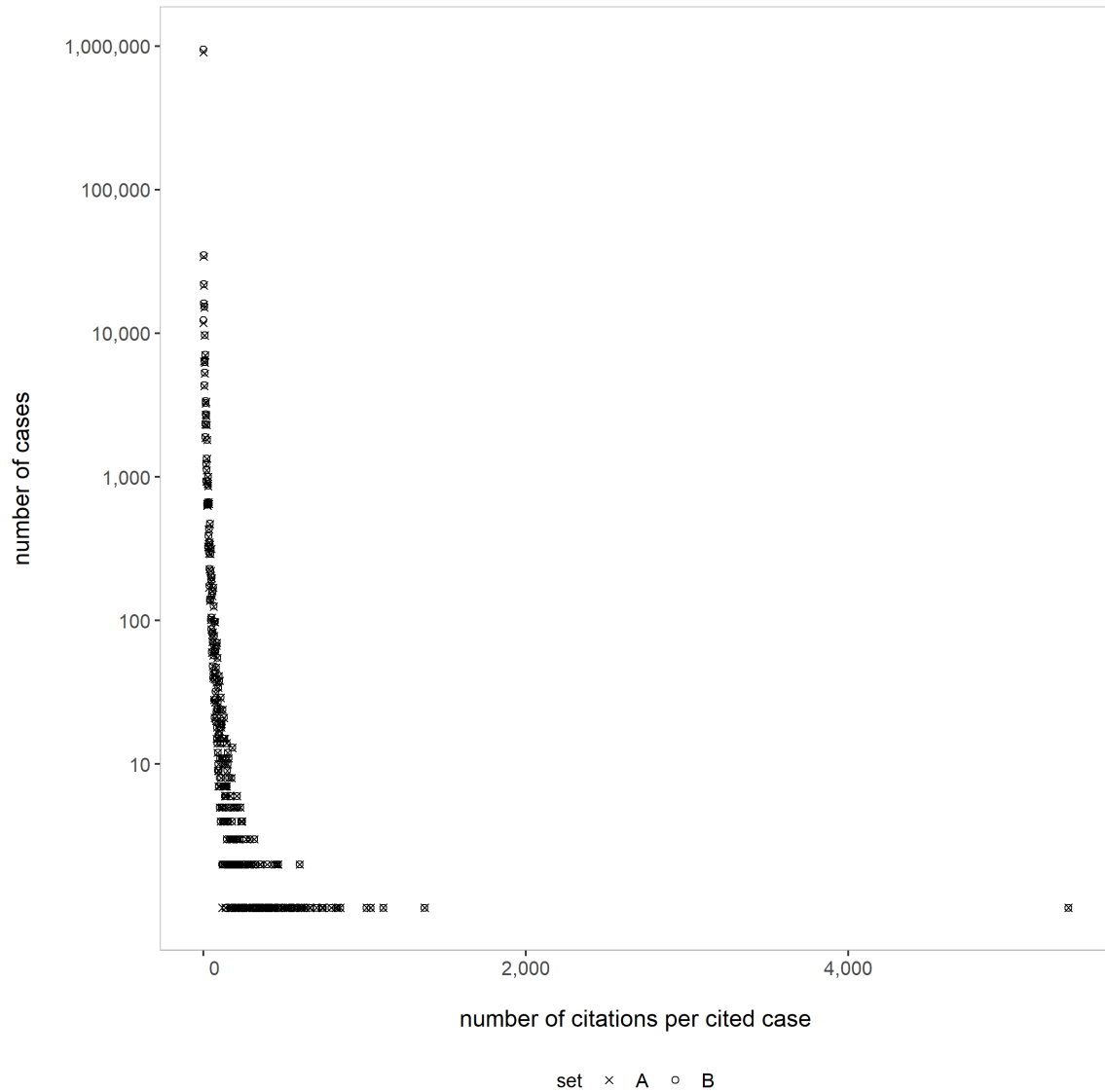


Figure 2: Distribution of citations per cited case

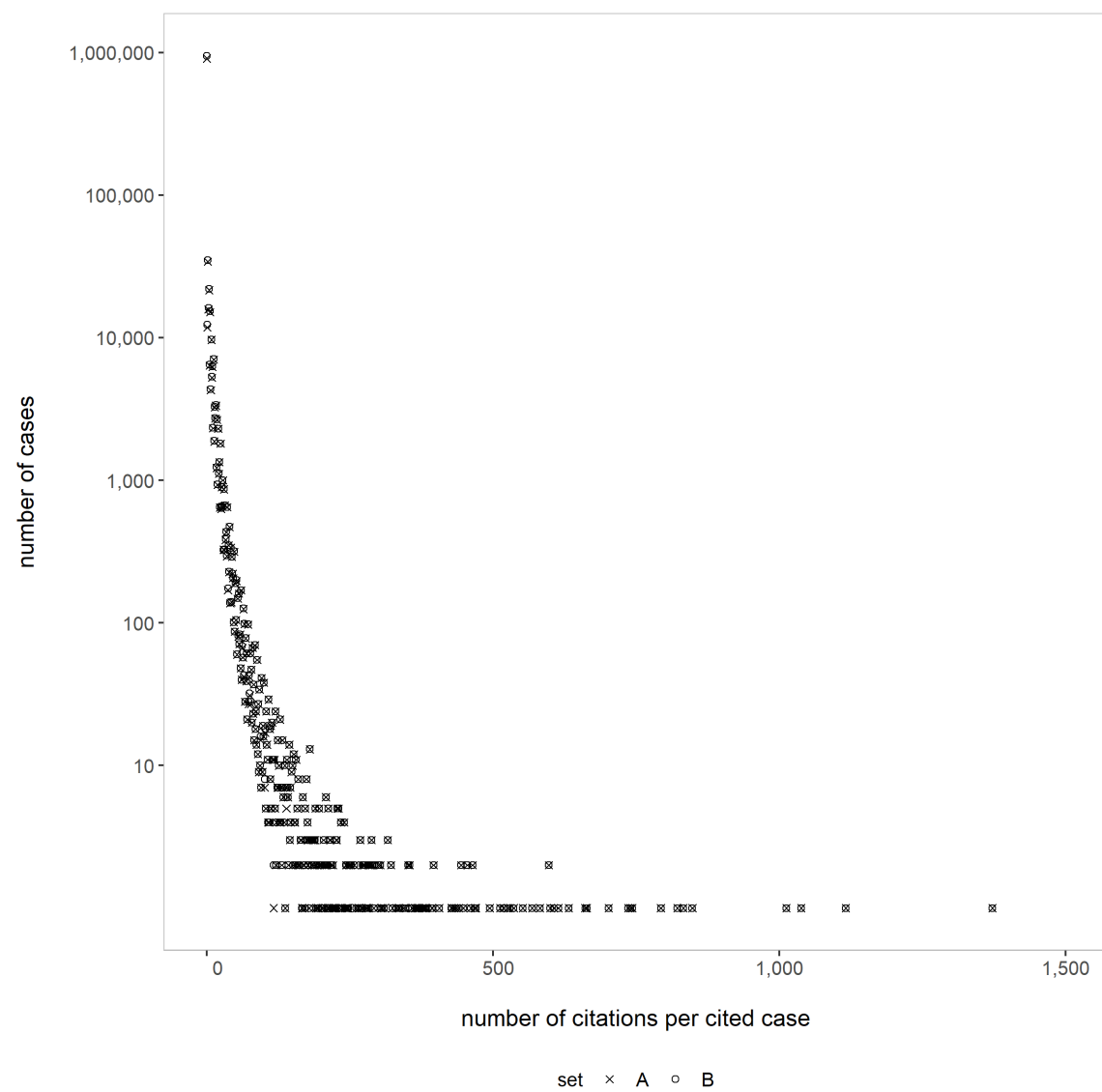


Figure 3: Distribution of citations per cited case, excluding Hardy v. Raleigh Central Prison

5 Distribution of citations by court and year

Figure 4 shows the distribution of citations by year. Citing cases are limited to those with an LNI that appears in case header records. Differences in late year citation frequency by data set may be due to the addition of cases in data set B. Figure 4 shows the distribution of citations by court and year.

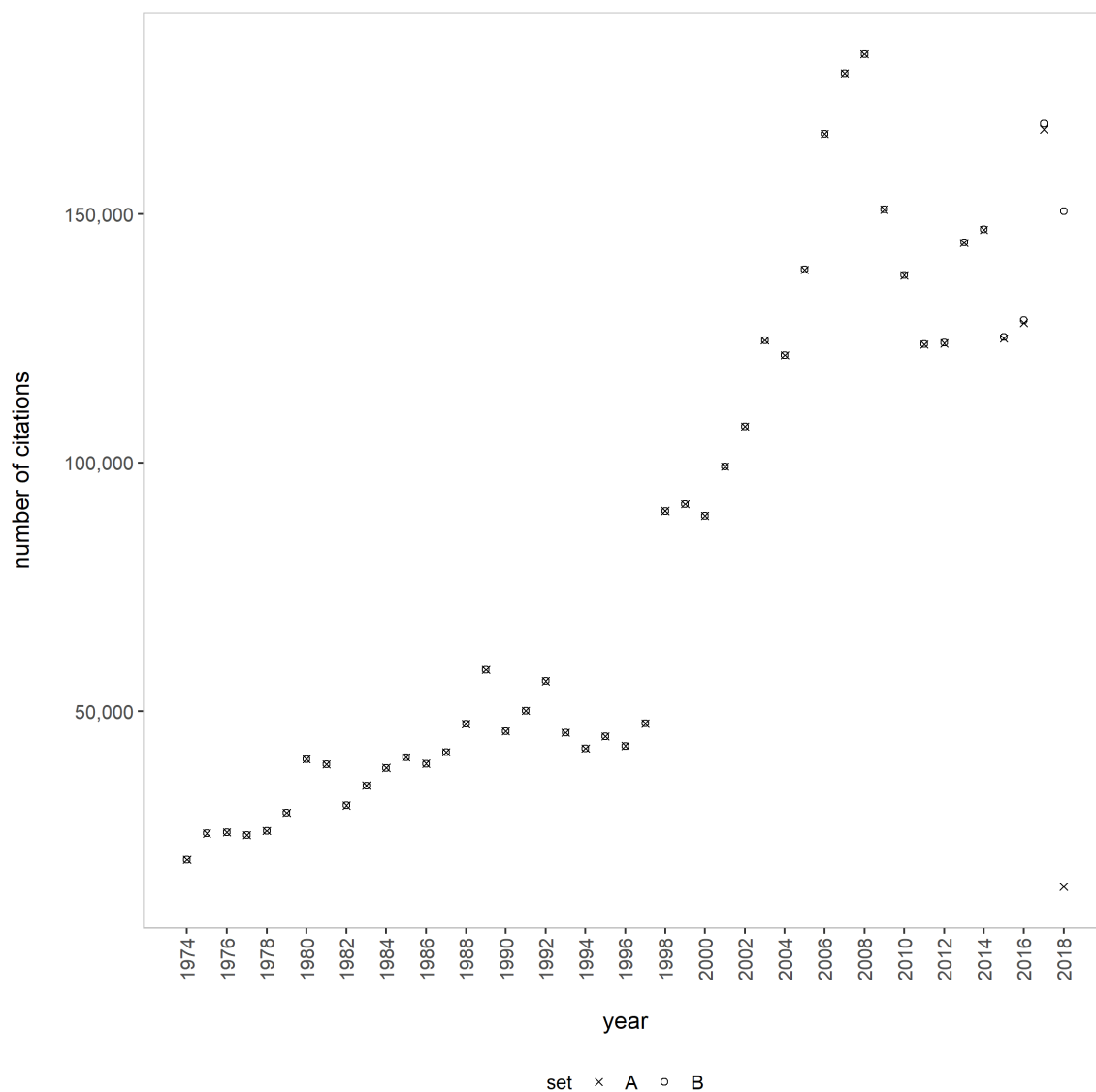


Figure 4: Distribution of citations by year

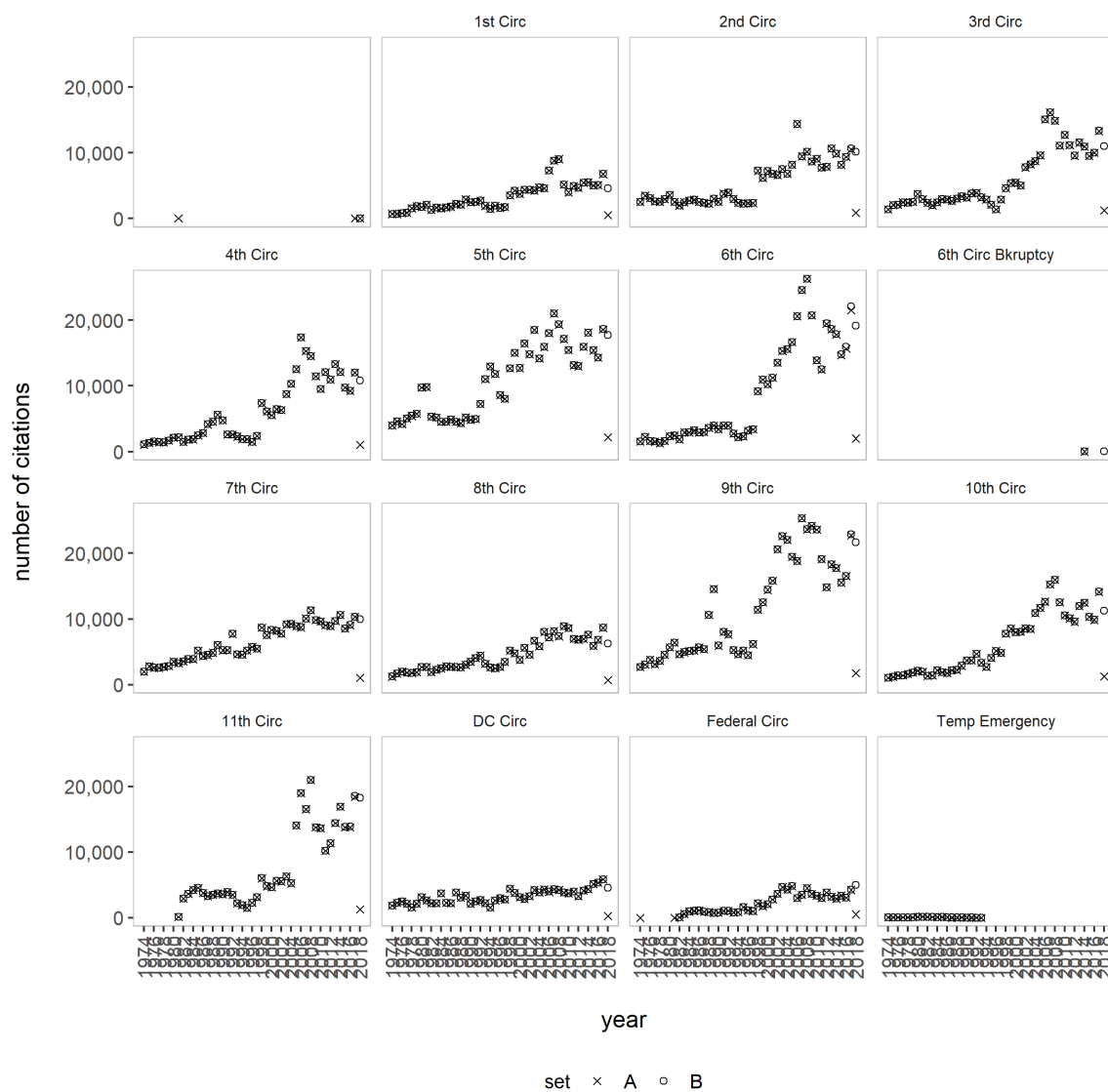


Figure 5: Distribution of citations by court and year

6 Citations with LNI that do not appear in case header records

11,799 and 25 citing LNIs in data sets A and B, respectively, do not appear in case header records of the respective data set. Although this is unexpected, it merely results in having citations that do not have a reference in the data. However, 147,202 and 147,995 cited LNIs in data sets A and B, respectively, do not appear in case records of the corresponding data set. In a relational database setting, these are termed "orphaned records" since there is a reference to a case by ID, but nothing is known about that case. Information beyond the fact that a citation exists, such as the court, date of decision, case type, outcome, or panel, then the existence of orphaned citations seems problematic.

No citations were received with data set B. When asked if we would receive revised citations, LN explained that changes to the scripts that generated case files for data set B did not affect citations. Cases were included in data set B that do not appear in data set A (particularly for years 1973 and 2015-2018). Is it likely that the additional cases are neither citing nor cited cases?

7 Duplication of citations by reporter

Many citations are duplicated, having multiple reporters. For instance, case 3RFG-6SY0-0038-X00H-00000-00 cites case 3S42-KX20-00B1-D24P-000000-00 on five separate records, one for each of reporters

- Ct. App. 2nd Dkt. No. 96-1712
- 97-2 U.S. Tax Cas. (CCH) P50608
- 80 A.F.T.R.2d (RIA) 97-6085
- 1997 U.S. App. LEXIS 21273
- 122 F.3d 112

Each of the above citations references Shepard's treatment letter f. Care must be taken when aggregating cases by citation to make sure that actual citations, and not reported citations, are counted. Additional information on reporters is available at <https://nyulaw.libguides.com/c.php?g=773843&p=5551618>. Following are excerpts from this site:

- Federal case law will be divided between three branches: U.S. Supreme Court, U.S. Courts of Appeal (Circuit Courts), and the U.S. District Courts.
- Case law is published in official and unofficial case law reporters, in chronological order. Supreme Court opinions are the only federal opinions published in official case reporters. West is the major publisher of unofficial case reporters.
- Case citations are the easiest way to retrieve a case (it is possible to find cases by party name, but may not be as exact). Case citations are structured by volume number, reporter abbreviation, and first page of the case.
- U.S Supreme Court Cases: Supreme Court opinions are published in 3 different case law reporters: United States Reports (the official reporter) - "U.S.", Supreme Court Reporter (West) - "S. Ct.", and Lawyers' Edition (Lexis) - "L. Ed.". Sample citation for Brown v. Board of Ed. of Topeka: 347 U.S. 483, 74 S. Ct. 686, 98 L. Ed. 873.
- Federal Circuit Courts of Appeal: Circuit Court opinions are published in unofficial reporters (West): Federal Reporter - "F.", Federal Reporter Second Series - "F.2d" and Federal Reporter Third Series - "F.3d".
- Federal District (Trial) Courts: The District Court opinions are published in unofficial case reporters (West): Federal Supplement - "F. Supp." and Federal Supplement Second Series - "F. Supp.2d"
- Earlier Federal Cases (dating before the Federal Reporter series) are available in Federal Case "F. Cas." bound volumes

8 Distribution of the difference in number of citations by Shepard's treatment letter

Figure 6 shows the difference, between data sets B and A, in citations by Shepard's treatment. Citing cases are limited to those that appear in case header records. Figure 7 plots the difference, between data sets B and A, in citation frequency by Shepard's letter and year. Letters are limited to those with a total difference greater than 1,000 between data sets. Red indicates a negative difference.

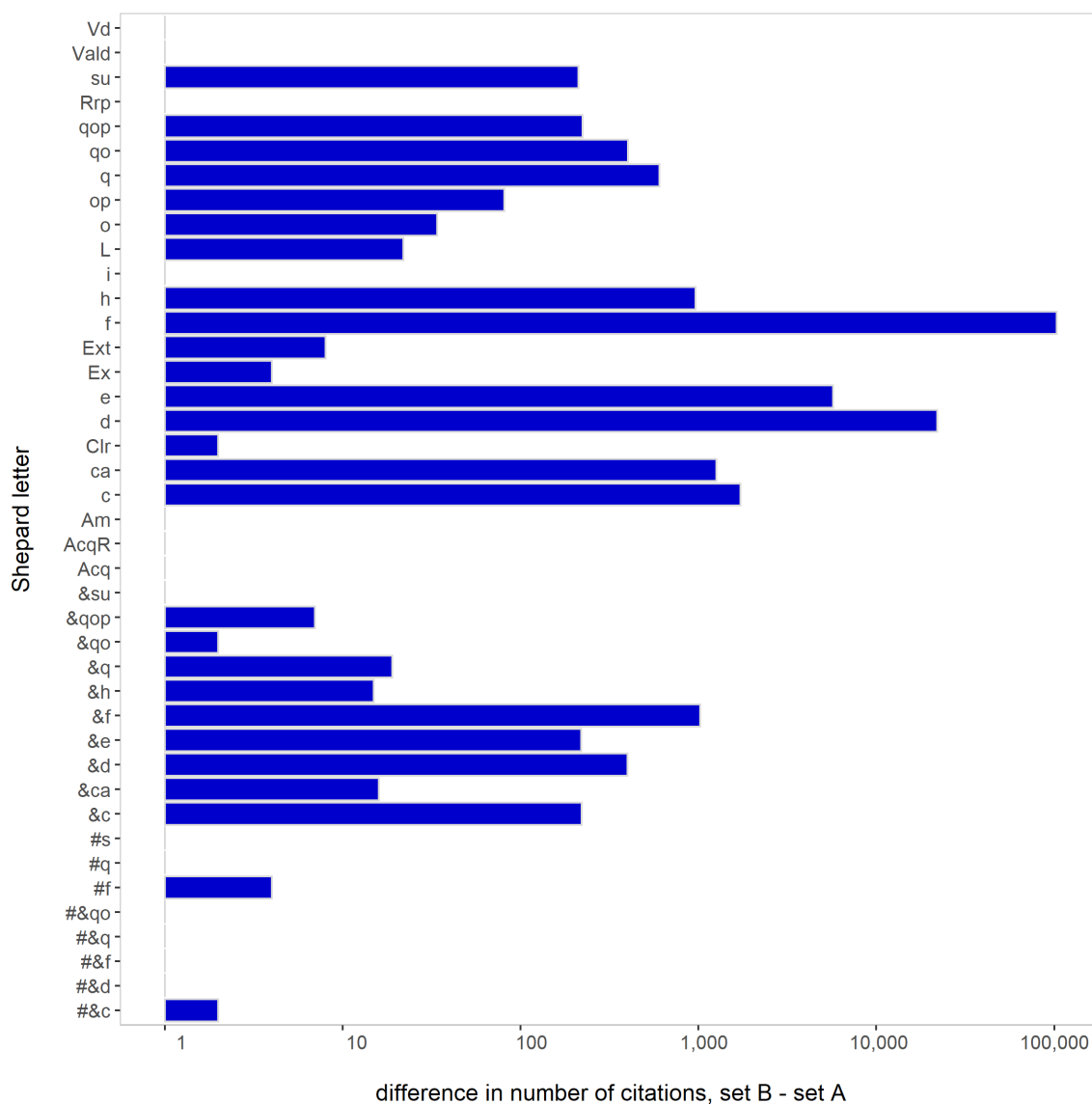


Figure 6: Difference, between data sets A and B, in citation frequency by Shepard's letter

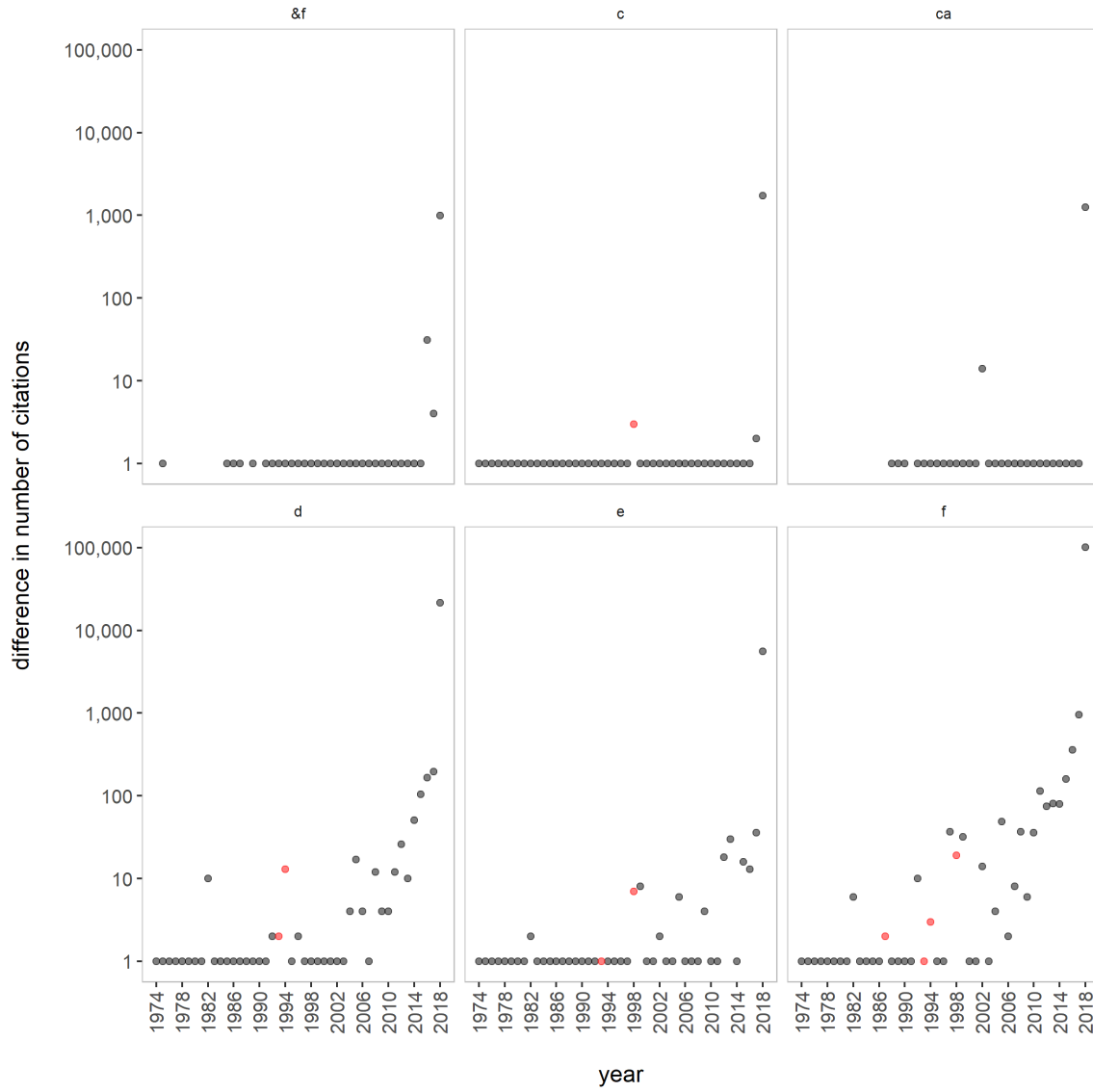


Figure 7: Difference, between data sets A and B, in citation frequency by Shepard's letter, by year. Letters with total difference greater than 1,000. Red indicates a negative difference.