LexisNexis Appeals Data Review (Data Received March 13, 2019)

Version 1.1   April 10, 2019

Tom Balmat

1 Data Receipt

U.S. District and Circuit Court appeals data were received by Duke from LexisNexis on March 13, 2019. Table 1 lists files and corresponding numbers of records received. Tables 2 through 5 give field definitions for the supplied files.

Table 1: LexisNexis Appeals Data Set

File	Contents	Records
DukePartnn.csv (eleven parts)	Primary case data	1,083,602
Judges.csv	Judge IDs and names	60,096
LNI-LNI-Citations.csv	Citation with Shepard's treament	3,731,205
Shepard's-Treatment-Phrases.xlsx	Shepard's treatment letters and definitions	176

Table 2: Primary Case Data

Field	Description
LNI	Unique case identifier
Date	Decision date? (some are empty)
CourtShortName	
CourtLongName	
Title	Case title, LN assigned
ShortCaseName	
FullCaseName	
ThisCaseRefs	Not used
CaseTypes	Criminal, Civil, empty (pipe delimited)
Pubstatus	Reported, Unreported, empty
JudgeEntityList	Panel judge IDs (pipe delimited)
Judges	Not used
OpinionBy	Judge name(s)*
ConcurBy	Judge name(s)*
DissentBy	Judge name(s)*
PerCuriam	per curiam indicator
Outcome	Average 175 characters
LegalTopics	Topic IDs (pipe delimited)
InternalTopicIDs	LN Topic IDs (pipe delimited, one for each legal topic per case)
TreatedCites	Not used

* No IDs supplied

Table 3: Judges

Field	Description
ID	Not used
JudgeID	Unique LN assigned identifier
JudgeName	First, M, Last

Table 4: Citations (LNI-LNI-Treatments)

Field	Description
CaseLNI	LNI of citing case
LNICited	LNI of case cited
NormCite	Normalized citation
ShepardLetter	Shepard's treatment ID

Table 5: Shepard's Treatment Definitions

Field	Description
Letter	Treatment identifier
CurrentCategory	Treatment, Statute, etc.
Subcategory	Agency, General, etc.
CaseStatute	Case, Statute
TreatmentPhrase	Brief description
Description	Full description

2 Data Import

Supplied data were imported into a MySQL database. To implement a hierarchical structure, fields with delimited values (JudgeEntityList, LegalTopics, InternalTopicIDs) were stored in separate tables along with referring case LNI. Common values such as court names and Shepard's treatment definitions were stored in separate tables and referred to by ID as needed. Table 6 lists database table names and associated record counts.

Table 6: Appeals Database Tables

Table	Contents	Records
CaseHeader	Primary case data	1,083,600*
CaseLegalTopics	Legal Topic IDs by LNI	4,699,274
CaseLNTopics	LN Topic IDs by LNI	4,699,274
CaseType	Case type by LNI	937,701
Citation	LNI-LNI reference with norm cite and Shepard's Treatment letter	3,731,205
Court	Duke assigned ID with official short/long names	18
Judge	LN assigned ID with FML name	60,077**
Citations	Citation with Shepard's treatment	3,731,205
Opinion	Opinion, concurring, dissenting author IDs	0
Panel	LNI with Judge IDs	2,305,731
ShepardTreatment	Shepard's treatment letters and definitions	176

* Two records omitted (this is explained in section 3)
** Twenty-two records omitted (this is explained in section 3)

3 Initial Data Validation

The following initial data integrity checks were made. Emphasis is on columnar consistency of format and existence of reference records for all IDs appearing in referring records. Items in red require attention.

- Tested for duplicate CaseHeader LNI: none found

- Tested for duplicate LNI, CaseType: none found

- Is the CaseHeader field labeled "Date" the decision date?

- Missing dates in CaseHeader: 29 records (no decision rendered?)

- The following cases are missing a court identifier:
  3S4X-BYM0-001B-K3HW-00000-00
  5RK7-J161-DYFH-X0W8-00000-00
  5NMN-RCM1-DYTB-X1V5-00000-00
  5NMN-RDC1-JCB9-61DX-00000-00
  3S4W-WSK0-0039-M0TG-00000-00

5NMN-RGW1-DYTB-X1VV-00000-00
5NMN-RCY1-DYTB-X1VF-00000-00
5NYK-PXX1-DYTB-X4XM-00000-00
5NMN-RDF1-JCB9-61F0-00000-00
5NMN-RCT1-DYTB-X1V9-00000-00
5RJ8-3DP1-F04K-S2GK-00000-00
3S12-VYB0-0038-X3BN-00000-00
5NMN-RDR1-JCB9-61F4-00000-00
5NYK-PY01-DYTB-X4XR-00000-00
3S4X-6M00-008H-V0GV-00000-00
5NMN-RDN1-JCB9-61F1-00000-00
5NMN-RJ41-JCB9-61GJ-00000-00
3S4X-5JR0-003B-50T8-00000-00
5NMN-RFF1-DYTB-X1VK-00000-00
5NMN-RD71-JCB9-61DG-00000-00
3RV5-KSB0-003B-50M4-00000-00
3S4X-0650-0039-W4RT-00000-00
5NMN-RH41-JCB9-61G4-00000-00
3S4W-T920-0039-M08N-00000-00
5NMN-RD91-JCB9-61DP-00000-00
3S4X-C0G0-001T-D46N-00000-00
5NMN-RCT1-DYTB-X1VC-00000-00
5NMN-RJ11-JCB9-61GD-00000-00
3S4X-1PS0-001B-K1XV-00000-00
5NMN-RD81-JCB9-61DM-00000-00
5NRR-WCV1-F04K-W0B6-00000-00
4F7Y-YXX0-TVRV-K31B-00000-00
3RHM-PKM0-0038-X3K4-00000-00
4KND-FMG0-TVRV-K3DG-00000-00
5NYK-PXX1-DYTB-X4XP-00000-00
4G24-T8Y0-TVRV-K2NF-00000-00
3S4X-6M00-008H-V0GW-00000-00
3S4X-CSF0-003B-5370-00000-00
3RV5-KP00-008H-V07P-00000-00
5NYK-PX51-DYTB-X4WP-00000-00
5NMN-RH11-JCB9-61FX-00000-00

- 212,937 cases have no panel

- The following CaseHeader records have opinion text in the OpinionBy field:
  3T3Y-TS60-0038-X1MJ-00000-00: rec 31214 of part01 file
  3T8C-CSW0-0038-X3XN-00000-00: rec 19963 of Part08 file

- The following judge records are missing a name:
  urn:entity:jud-500000312
  urn:entity:jud-500000557
  urn:entity:jud-500000563
  urn:entity:jud-500000660
  urn:entity:jud-500000860
  urn:entity:jud-500000946
  urn:entity:jud-500002100
  urn:entity:jud-500002286
  urn:entity:jud-500002313

urn:entity:jud-500002359
urn:entity:jud-500002376
urn:entity:jud-500002486
urn:entity:jud-500002575
urn:entity:jud-500002595
urn:entity:jud-500002615
urn:entity:jud-500002937
urn:entity:jud-500003017

- Tested for judges appearing in Panel that do not appear in the supplied list of judges: none found

- Duplicate Shepard treatment letters (with different treatment phrases): "Acq" and "Noacq" (each are repeated once)

- Several Shepard's treatment records contained the text "Obsolete" in the TreatmentPhrase field. They were omitted.

- Some normalized cites in the LNI-LNI citation records contain commas. This column should be quoted in the source file.

- The following Shepard's treatment letters appeared in LNI-LNI Citation records, but did not appear in the Shepard's treatment file (affected citations have been omitted, quantities appear after each code):
  &qab, 4
  &qabp, 33
  *c, 15
  *d, 23
  *e, 27
  *f, 47
  *o, 24
  *q, 31
  ab, 128
  abp, 366
  cf, 30
  Ltd, 4
  mi, 2
  pd, 27
  pg, 12
  qab, 1699
  qabp, 2019
  qr, 681
  qv, 514
  qw, 37

- 215,546 cases do not appear in as either a citing or cited case

- Tested for duplicate citations (identical LNI-LNI, norm cite, and Shepard's treatment): none found

- 638,060 cases have no legal topics

- Each supplied case has an identical number of legal and internal topics. Are both of these needed?

- Legal and internal topic codes were supplied for each case, but the corresponding definitions were not supplied

4 Data Properties

Following are properties and distributions observed in key data elements.

Table 7: Distribution of Cases by Year and Type

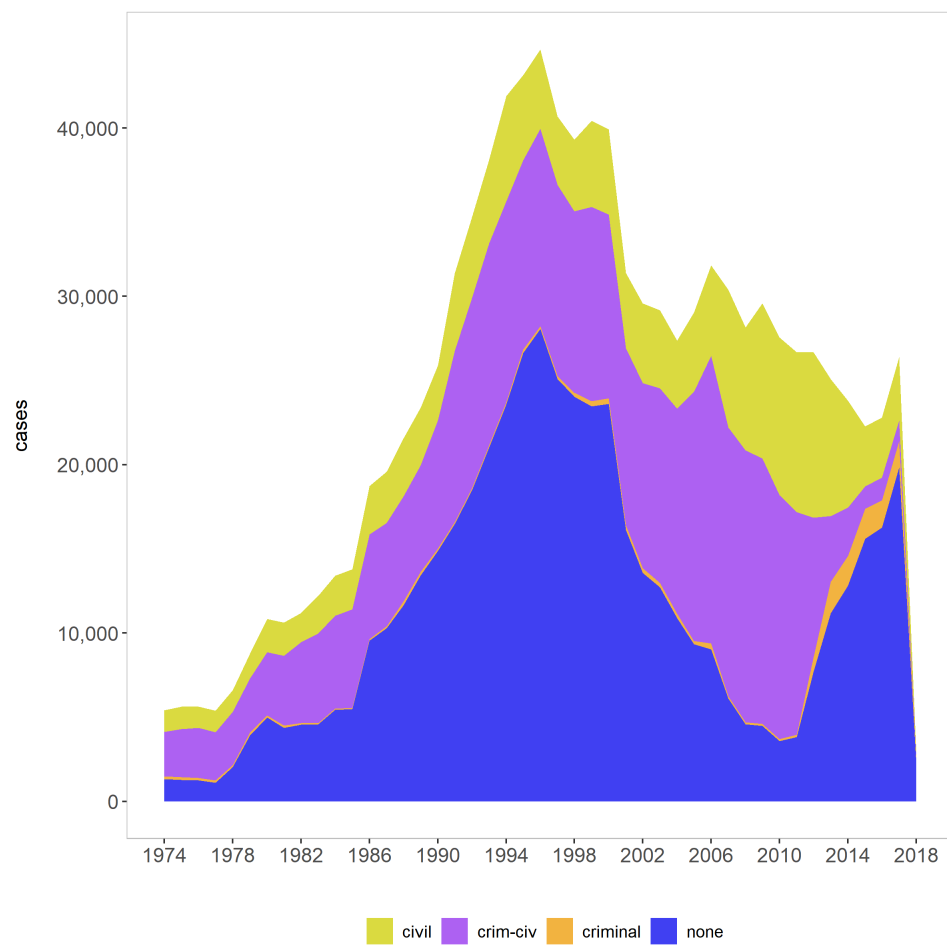| Year	Total Cases	Civil	Criminal	Civil-Criminal	Unidentified
0000	29	1	0	4	24
1974	5,406	1,271	167	2,638	1,330
1975	5,621	1,321	159	2,867	1,274
1976	5,618	1,261	146	2,973	1,238
1977	5,398	1,287	144	2,863	1,104
1978	6,604	1,292	128	3,159	2,025
1979	8,772	1,484	149	3,218	3,921
1980	10,829	1,960	125	3,764	4,980
1981	10,626	1,967	123	4,162	4,374
1982	11,201	1,746	90	4,807	4,558
1983	12,208	2,231	64	5,323	4,590
1984	13,411	2,381	71	5,533	5,426
1985	13,784	2,372	66	5,860	5,486
1986	18,736	2,874	88	6,213	9,561
1987	19,594	3,040	136	6,139	10,279
1988	21,560	3,449	239	6,235	11,637
1989	23,392	3,427	199	6,328	13,438
1990	25,901	3,275	148	7,599	14,879
1991	31,364	4,605	125	10,160	16,474
1992	34,692	4,809	133	11,281	18,469
1993	38,117	4,956	131	12,011	21,019
1994	41,913	6,256	129	11,934	23,594
1995	43,150	5,067	212	11,248	26,623
1996	44,659	4,700	170	11,738	28,051
1997	40,697	4,080	205	11,332	25,080
1998	39,323	4,279	232	10,774	24,038
1999	40,441	5,124	309	11,533	23,475
2000	39,921	5,061	318	10,925	23,617
2001	31,387	4,480	283	10,503	16,121
2002	29,578	4,734	280	10,991	13,573
2003	29,185	4,638	261	11,571	12,715
2004	27,379	4,054	251	12,189	10,885
2005	29,025	4,691	205	14,801	9,328
2006	31,847	5,380	356	17,079	9,032
2007	30,391	8,177	144	15,963	6,107
2008	28,153	7,310	118	16,153	4,572
2009	29,571	9,189	126	15,776	4,480
2010	27,567	9,384	122	14,483	3,578
2011	26,683	9,498	162	13,212	3,811
2012	26,679	9,822	858	8,293	7,706
2013	25,081	8,125	1,887	3,899	11,170
2014	23,799	6,340	1,764	2,881	12,814
2015	22,267	3,551	1,777	1,347	15,592
2016	22,790	3,563	1,625	1,338	16,264
2017	26,396	3,785	1,556	1,221	19,834
2018	2,855	378	133	135	2,209
	1,083,600	192,675	16,114	364,456	510,355

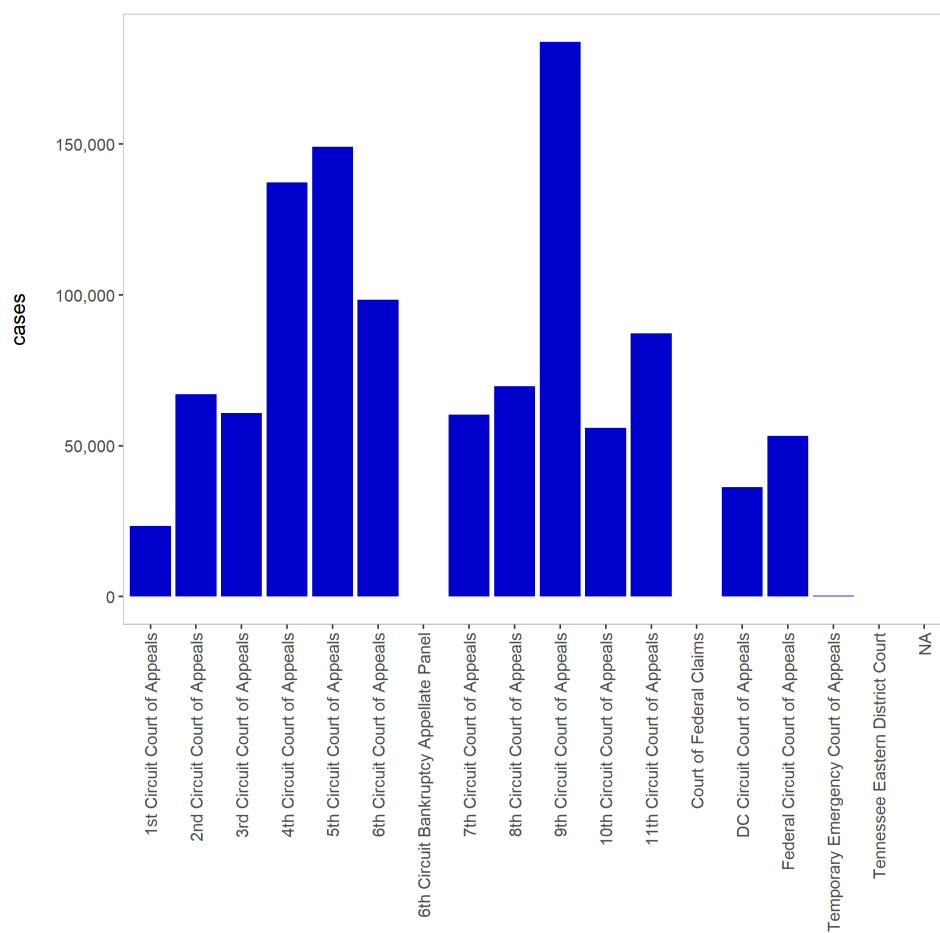Figure 1: Stacked distribution of cases by year and type
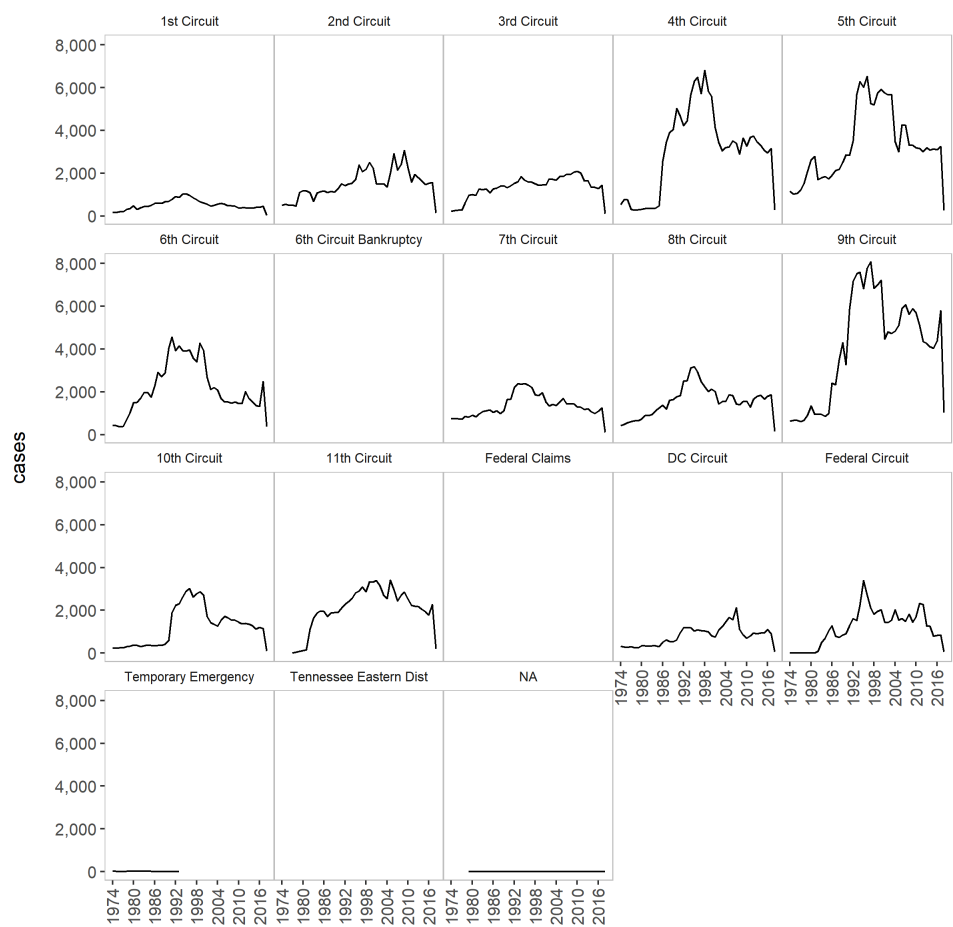
Figure 2: Distribution of cases by court
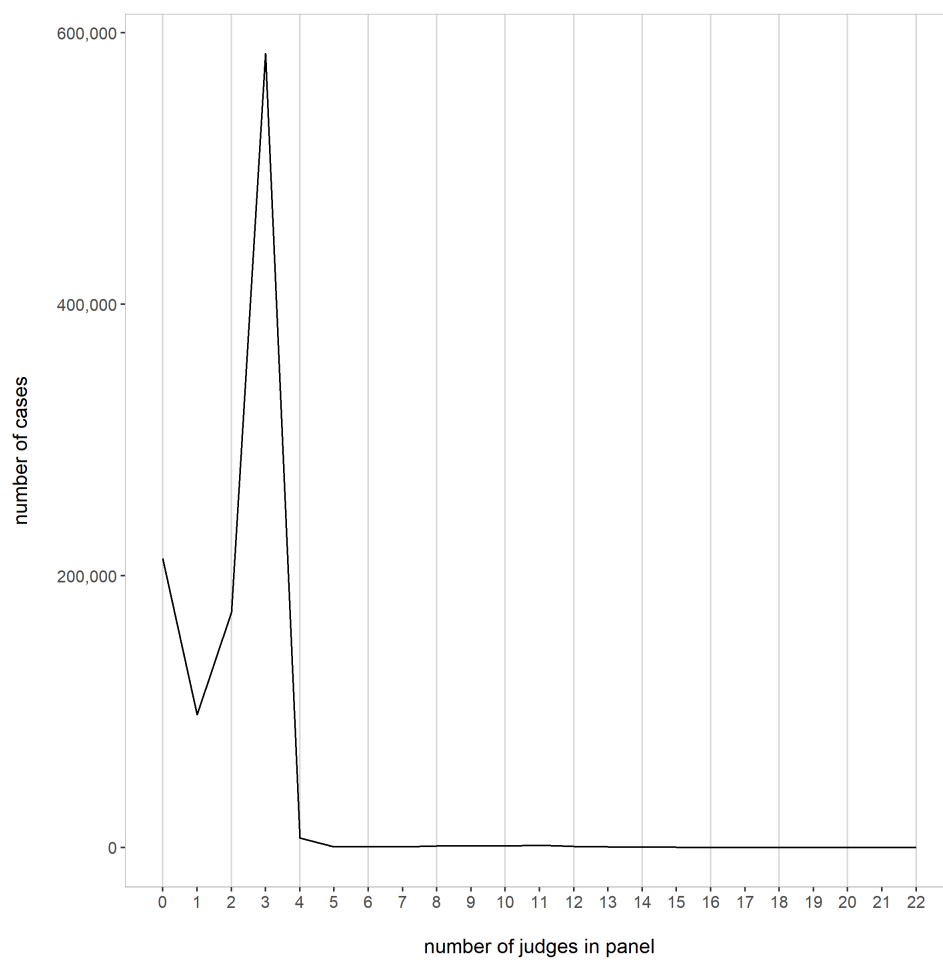
Figure 3: Cases by court and year

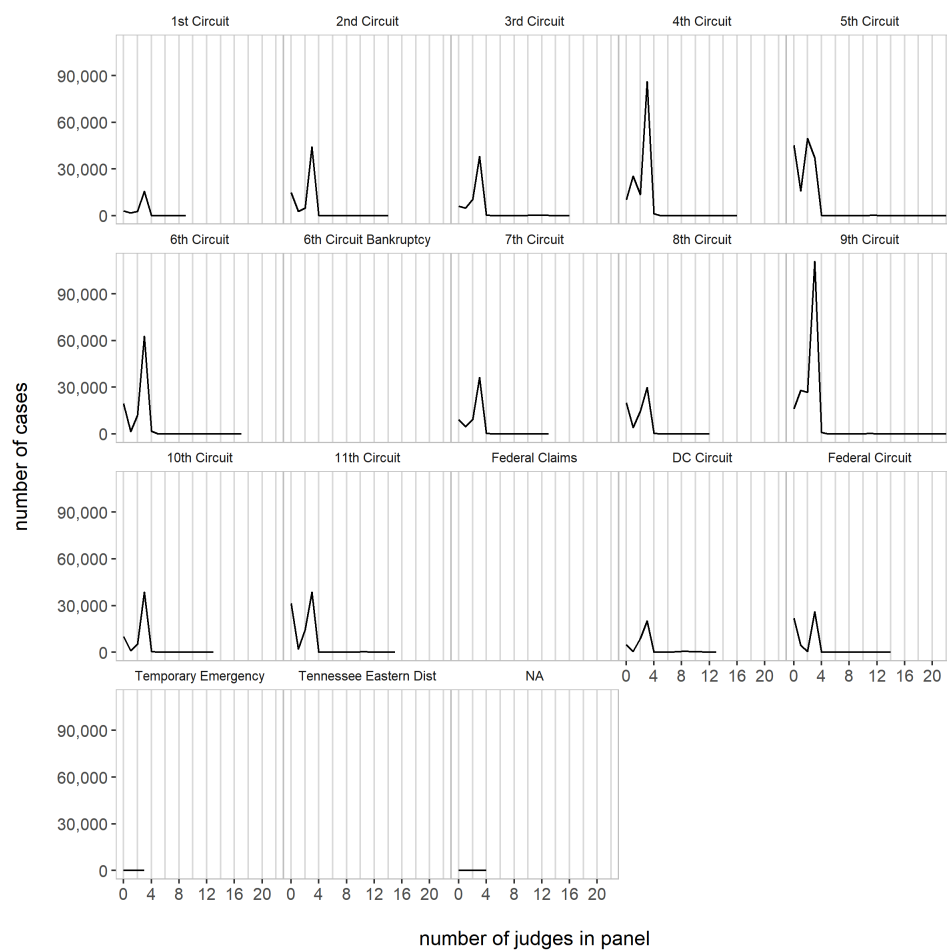Figure 4: Distribution of cases by panel size

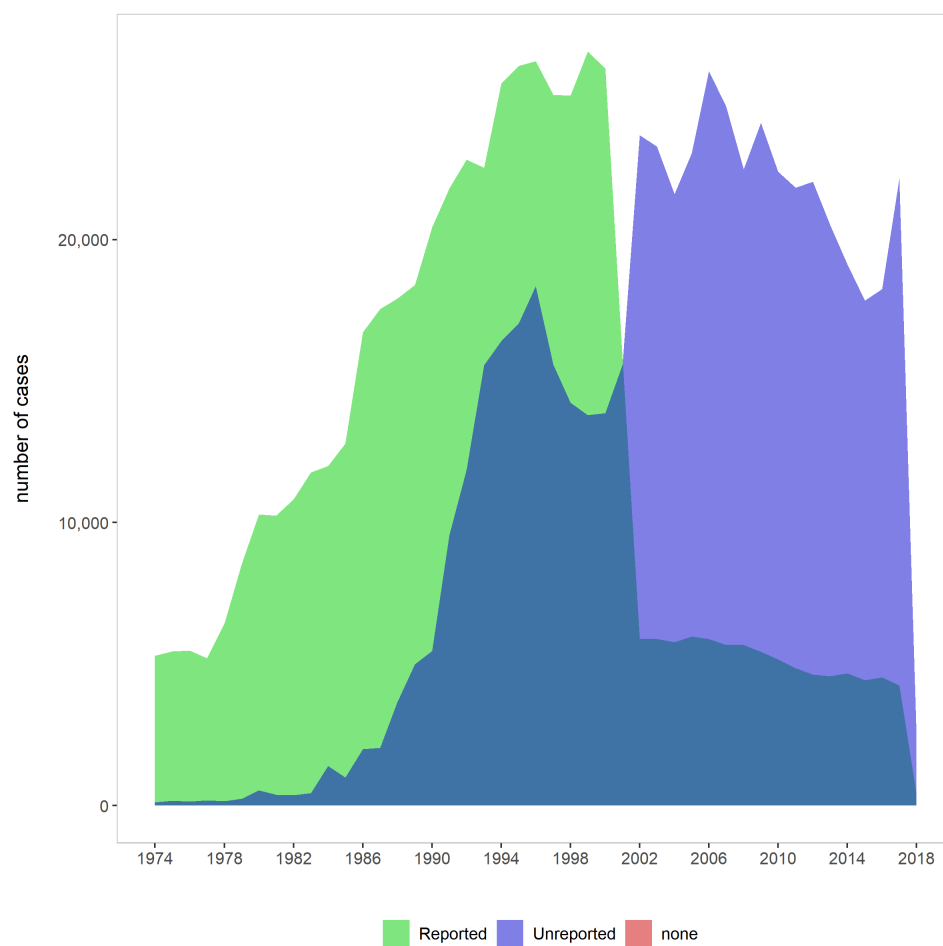Figure 5: Cases by panel size and court

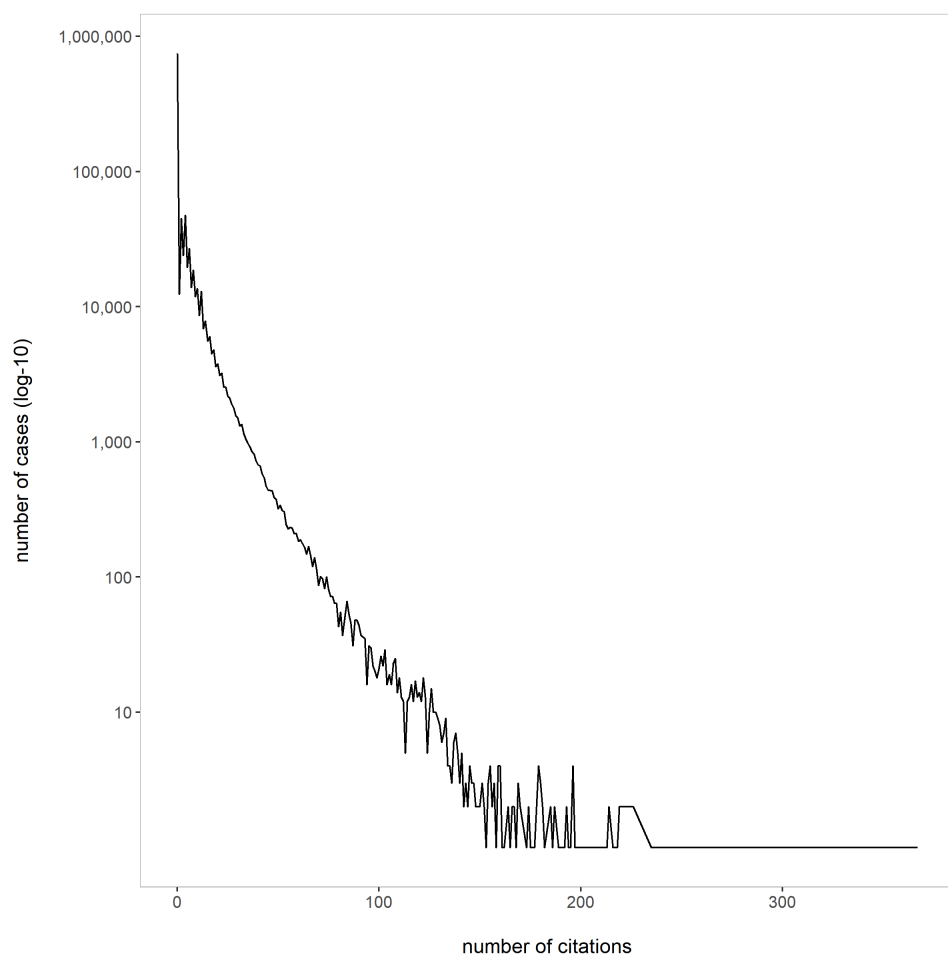Figure 6: Distribution of cases by year and PubStat

Figure 7: Distribution of cases by number of cases cited (note the log-10 Y axis scale)
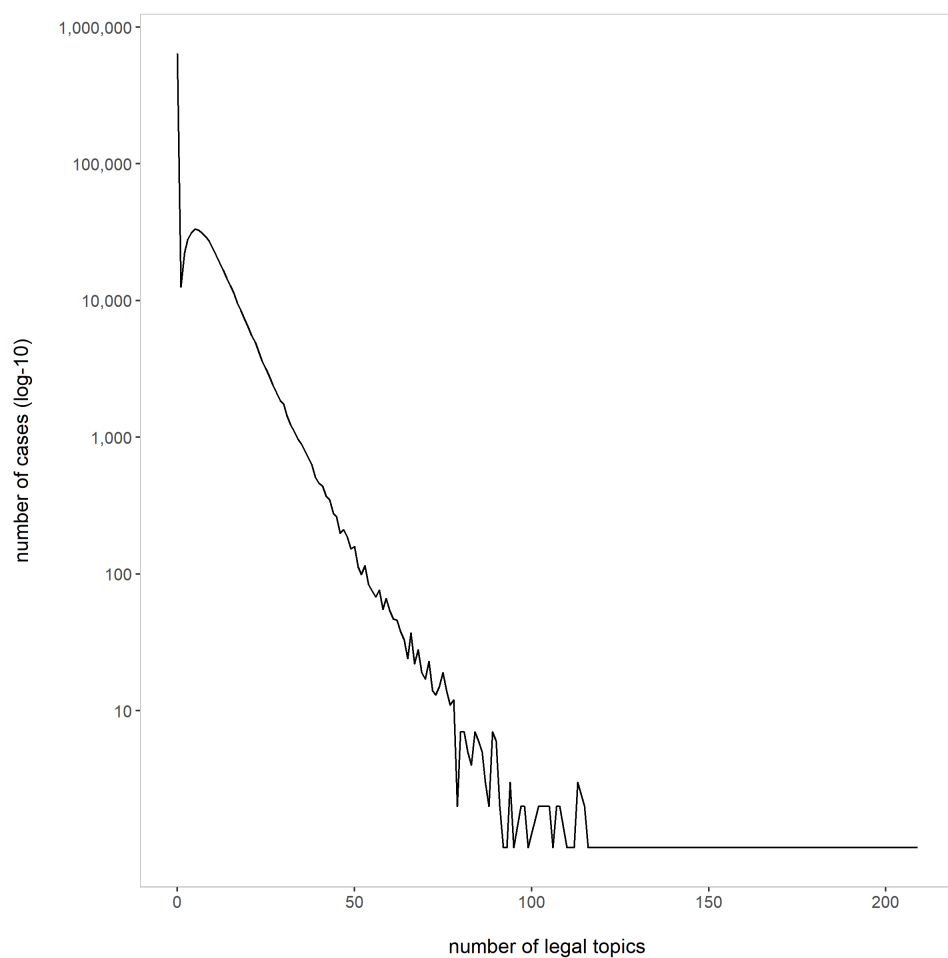
Figure 8: Distribution of cases by number of legal topics (note the log-10 Y axis scale)
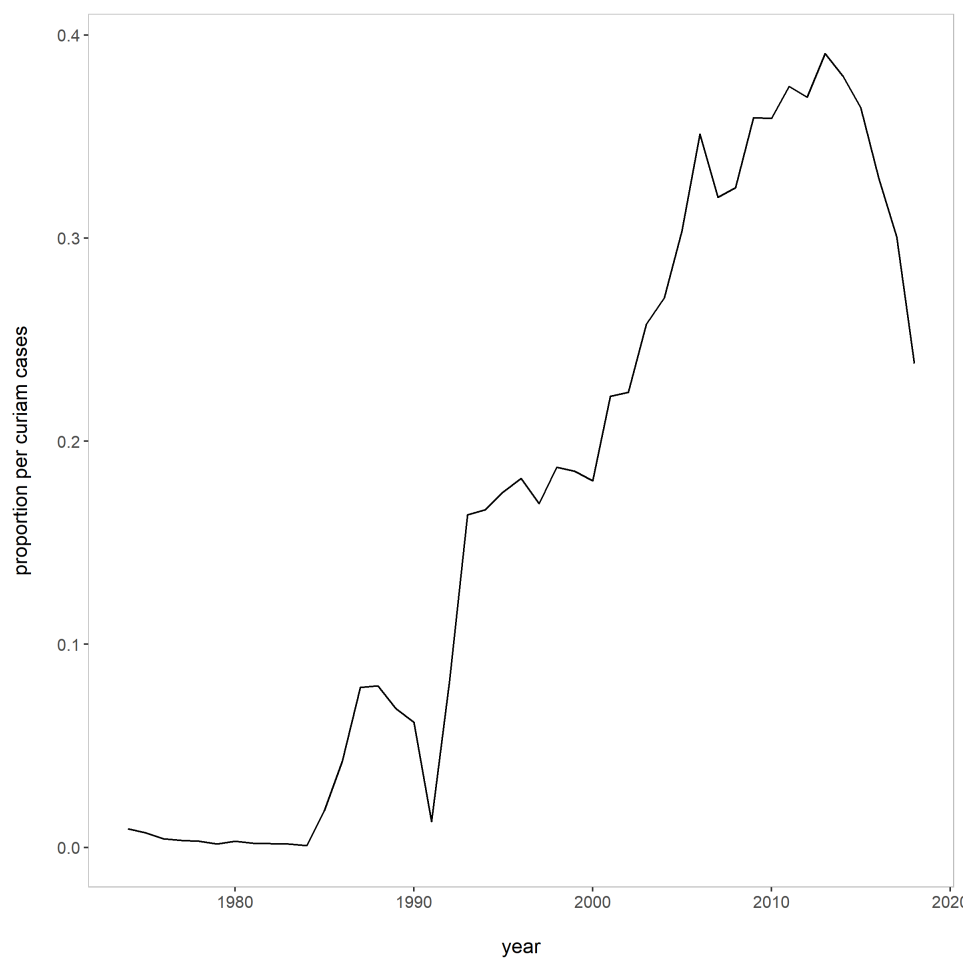
Figure 9: Per curiam cases by year

Table 8 lists the results of the logistic regression model $\mathrm{logit}(per\ curiam) = \beta_0 + year + courtID + caseType$. Near-zero $\Pr(>|z|)$ values indicate significance and it is seen that all but one year, most courts, and all case types appear to be significant predictors. Note that one level of year, court, and case type are omitted and used as reference levels. The gradual increase in $year$ coefficient estimate is consistent with the pattern observed in figure 9.

Table 8: Logistic Regression of P(per curiam) $\sim$ Year+Court+CaseType

Parameter	Estimate	Std. Error	z value	Pr($>$|z|)
(Intercept)	-9.7647	0.1488	-65.63	0.0000
year1975	-0.3383	0.2136	-1.58	0.1131
year1976	-0.9097	0.2507	-3.63	0.0003
year1977	-0.9332	0.2719	-3.43	0.0006
year1978	-0.9683	0.2667	-3.63	0.0003
year1979	-1.4726	0.3045	-4.84	0.0000
year1980	-0.8455	0.2287	-3.70	0.0002
year1981	-1.2824	0.2575	-4.98	0.0000
year1982	-1.1235	0.2666	-4.21	0.0000
year1983	-1.2183	0.2663	-4.58	0.0000
year1984	-1.7554	0.3220	-5.45	0.0000
year1985	1.2899	0.1579	8.17	0.0000
year1986	1.9435	0.1486	13.08	0.0000
year1987	2.5296	0.1465	17.27	0.0000
year1988	2.5463	0.1462	17.41	0.0000
year1989	2.4066	0.1463	16.44	0.0000
year1990	2.1985	0.1463	15.03	0.0000
year1991	0.6916	0.1525	4.53	0.0000
year1992	2.9417	0.1453	20.25	0.0000
year1993	3.8797	0.1446	26.83	0.0000
year1994	3.8036	0.1445	26.32	0.0000
year1995	3.8641	0.1445	26.75	0.0000
year1996	3.9498	0.1444	27.35	0.0000
year1997	3.9040	0.1446	27.01	0.0000
year1998	3.8870	0.1445	26.90	0.0000
year1999	3.9405	0.1445	27.27	0.0000
year2000	3.8955	0.1445	26.95	0.0000
year2001	4.1185	0.1446	28.48	0.0000
year2002	4.1916	0.1447	28.97	0.0000
year2003	4.5153	0.1447	31.21	0.0000
year2004	4.8384	0.1448	33.41	0.0000
year2005	5.1655	0.1447	35.69	0.0000
year2006	5.5407	0.1447	38.29	0.0000
year2007	5.2196	0.1448	36.05	0.0000
year2008	5.3502	0.1448	36.94	0.0000
year2009	5.5373	0.1448	38.24	0.0000
year2010	5.4776	0.1449	37.81	0.0000
year2011	5.4679	0.1449	37.74	0.0000
year2012	5.4043	0.1448	37.32	0.0000
year2013	5.5694	0.1449	38.44	0.0000
year2014	5.5054	0.1449	37.99	0.0000
year2015	5.4536	0.1450	37.60	0.0000
year2016	5.2863	0.1450	36.45	0.0000
year2017	5.1837	0.1449	35.78	0.0000

year2018	5.1792	0.1564	33.11	0.0000
CourtID2	2.0785	0.0451	46.06	0.0000
CourtID3	4.5660	0.0435	104.89	0.0000
CourtID4	7.0710	0.0394	179.27	0.0000
CourtID5	1.3930	0.0551	25.26	0.0000
CourtID6	6.4526	0.0393	164.06	0.0000
CourtID7	1.2996	0.0603	21.55	0.0000
CourtID8	3.8456	0.0405	95.03	0.0000
CourtID9	5.3527	0.0396	135.17	0.0000
CourtID10	-0.5147	0.0986	-5.22	0.0000
CourtID11	5.2001	0.0399	130.31	0.0000
CourtID12	3.8997	0.0409	95.40	0.0000
CourtID13	3.6247	0.0411	88.21	0.0000
CourtID14	4.0677	0.5951	6.83	0.0000
CourtID15	-4.4019	44.0205	-0.10	0.9203
CourtID16	-6.1776	196.9677	-0.03	0.9750
CourtID17	-6.2095	196.9677	-0.03	0.9749
CourtID19	-5.8559	196.9677	-0.03	0.9763
CaseTypeCriminal	-0.1968	0.0248	-7.95	0.0000
CaseTypeCriminalCivil	-0.6331	0.0106	-59.59	0.0000
CaseTypeNone	-1.1292	0.0100	-112.36	0.0000