

EXHIBIT D



PUBLIC.RESOURCE.ORG ~ *A Nonprofit Corporation*

Open Source America's Operating System

"It's Not Just A Good Idea—It's The Law!"

October 24, 2008

The Honorable Lee H. Rosenthal
Chair, Committee on Rules of Practice and Procedure
Judicial Conference of the United States
Washington, D.C. 20544

Dear Judge Rosenthal:

Pursuant to my previous letter of October 3, 2008, I am pleased to present to you the audit results for the 32 district courts for which we have been able to examine data. Of the 2,282 suspect files we reported to you in the preliminary findings, we were able to eliminate 613 of the documents as "false positives," leaving 1,669 documents with verified Social Security numbers and other issues.

However, just as our primitive scanning tool yielded false positives, we believe that there are probably a large number of false negatives in the 2,706,431 PDF files we examined. Indeed, often when our tool reported a Social Security number violation, when we looked around the document we also picked up many other Social Security numbers, birth dates, driver license numbers, Alien IDs, and bank account numbers.

We thus consider our work to date to be preliminary, both in thoroughness and scope. In subsequent stages, we hope to be able to use more comprehensive tools to perform a more thorough scan. In terms of scope, of the 32 district courts for which we have data, we do not have the full collection of cases for many of the districts and we thus hope to expand our work to provide more definitive results over the full database.

You will find attached a DVD that has all 1,669 problem files, with both the original document and a redacted version. You may use the "audited.html" summary file to compare the two versions of these documents. Our detailed results are also attached as Appendix A. We have also provided you with this data in spreadsheet format should you wish to perform further statistical analysis.

In Table 1, you will find summary statistics for each of the 32 district courts. We provide the total number of PDF files, HTML files, size of the collection in gigabytes, and the total number of pages. In addition, we list the total number of documents with Social Security numbers and calculate a Privacy Problem Index, which is the number of problem documents found per gigabyte of data. The Privacy Problem Index takes into account the fact that larger district courts will be expected to have a larger number of problems. As an aid to interpretation, we have taken the liberty of assigning letter grades on a curve. Needless to say, the curve might change if we had complete data for all the districts.

On a personal note, in the course of redacting the 1,669 documents in the attached DVD, I was quite struck by how damaging these privacy violations can be. A few of the horror stories I encountered that have kept me up nights include:

- In the District of Massachusetts, a 54–page list filed in June 2008 contains the names, birth dates, Social Security numbers, and medical problems of 353 patients of a doctor.
- In the District of the District of Columbia, an attorney who was not paid in what he considered to be a timely fashion by the District of Columbia schools decided to raise his rate to \$405/hour and bill the schools for the difference. To support his claim, he listed page after page of the names, home addresses, birth dates, and psychological issues for countless minors he saw.
- In the District of Alabama, lawyers seem to feel a need to sign briefs with their Social Security numbers, and the court consistently exposes the Social Security numbers and birth dates of police officers, state employees, and even court administrators.
- In the Central District of Illinois, litigants involving pension funds representing labor unions frequently attach the unredacted list of all union members and their Social Security numbers.
- In a huge number of IRS actions, unredacted tax returns are filed, including a large number where the redaction was performed incorrectly by simply placing a black box on top of the taxpayer ID, leaving the numbers untouched underneath the graphic.

To assist in the analysis and interpretation of this information, we have prepared a set of summary figures:

- In Figure 1, 30 of the districts (excluding Oregon and Northern Mariana Islands) are plotted with the Privacy Problem Index on one axis and the size of the archive on the other. As you can see, the District of Delaware has a very large archive, but a low rate of problems. On the other hand, the Central District of Illinois has a large archive but a very high problem rate.
- In Figure 2, a map is presented that shows color–coding for letter grades by geographic distribution. Needless to say, we were able to only provide this information for 31 of the 94 districts so the map is incomplete.
- In Figure 3, we plot the number of incidents over time, demonstrating that this problem is still ongoing.
- In Figures 4 and 5, we analyze two individual district courts to show that the distribution of privacy issues varies quite a bit by Judge. We realize different judges have different case loads and different kinds of cases, so further analysis of this data would be needed before drawing any conclusions.

On a technical note, a large number of the PDF files we encountered were not valid PDF files. A variety of “tricks” such as redistilling the files were performed to enable us to open them in Adobe Acrobat Professional, the tool we use for redaction. However, the fact that we had issues opening the files means that a large number of users will have similar issues. We would thus recommend a scan for PDF validity be performed. The open source Ghostscript package, in particular the PDF2PS utility, is quite useful for validating PDF content. In addition, there are commercial packages such as Apago’s PDF Appraiser for validation and automatic correction of PDF/A compatibility.

The biggest obstacles we have encountered have not been technical, but appear to be administrative barriers imposed to restrict public access. Although the PACER data we have been examining has been online for several years, neither the computer staff of the PACER system nor the commercial retailers such as LexisNexis or Westlaw have taken any steps to either report or redact this sensitive personal information. Indeed, in the case of the commercial sector, the Social Security numbers are considered to be a feature not a bug, enabling the vendors to sell sophisticated personal data mining products.

We bring this point up not to criticize either our commercial cousins or the well-respected staff who operate the computer systems for the courts. They have different objectives and time pressures. However, public interest groups and the public in general, when given access to these public records, are able to provide the kind of feedback that leads to the correction of these privacy issues. As Justice Brandeis said, "sunlight is said to be the best of disinfectants; electric light the most efficient policeman." If we want to be serious about personal privacy, we can only do so if we are also serious about public access.

Public access is a fundamental, enabling characteristic of our judicial system. As the Massachusetts Supreme Court so eloquently put the matter in *Nash v. Lathrop*, 6 N.E. 559 (1886), "every citizen is presumed to know the law thus declared, and it needs no argument to show that justice requires that all should have free access to the opinions, and that it is against sound public policy to prevent this, or to suppress and keep from the earliest knowledge of the public the statutes or the decisions and opinions of the justices."

Unfortunately, it appears that public access is an afterthought on the PACER system. Despite \$60 million/year in revenue with direct expenses of only \$11m (a very healthy 59% gross margin), and an "unobligated balance" of \$146.6 million in the Judiciary Information Technology Fund, the only way members of the public may access PACER is to petition a judge or pay \$0.08/page, a rate that quickly leads to large bills and is a prohibitive barrier for most people. Only this year did the PACER system decide to run a public trial in 17 libraries to "discover if a segment of the public desires access to information contained in the PACER system." That trial was run with no written or oral guidelines on appropriate use, and was then abruptly cancelled.

Public access goes to the heart of the role of the judiciary in our modern society and is an important matter of public policy for the Judicial Conference to consider. As you have seen, when public access is provided, the result is that we are able to take seriously issues such as the protection of privacy, the accountability of our system of justice to the people, and the right of citizens to know the law.

Please let me know if I can provide further information on this report and I look forward to providing you with a full audit of all the data for all 94 district courts.

Respectfully yours,

Carl Malamud
Public.Resource.Org

cc: Mr. Peter G. McCabe, Esq.
The Honorable Mr. Duff

Table 1
Summary of Privacy Issues Found

District	# PDF Files Examined	# HTML Files	Total # Pages	# Gbytes Examined	# Docs With SSNs	Privacy Problem Index	
akd	54,434	15,976	293,045	10.0	58	5.80	D-
almd	139,471	26,719	956,991	39.0	171	4.38	D
azd	67,181	95,927	532,604	16.0	33	2.06	C
cand	192,607	38,105	1,486,510	55.0	30	0.55	A-
casd	90,346	57,666	620,522	21.0	14	0.67	B+
cod	87,134	66,779	609,624	18.0	26	1.44	C+
cofc	85,686	79,216	613,040	27.0	228	8.44	F
ctd	78,158	59,159	546,083	16.0	16	1.00	B-
dcd	220,340	50,460	1,423,178	69.0	145	2.10	C
ded	183,618	81,669	1,225,609	68.0	40	0.59	A-
flsd	36,632	3,081	231,298	5.6	10	1.79	C
gud	32,889	14,223	164,855	6.7	11	1.64	C
hid	47,448	45,977	324,971	12.0	4	0.33	A-
ilcd	193,049	95,843	1,530,197	52.0	299	5.75	D-
ilnd	159,562	25,589	828,186	31.0	26	0.84	B
laed	6,590	12,365	41,097	1.1	2	1.82	C
mad	217,701	32,139	1,643,126	54.0	124	2.30	C-
mdd	62,038	166,907	553,404	15.0	68	4.53	D-
mnd	43,896	26,118	299,590	8.4	6	0.71	B+
njd	141,041	139,365	1,018,050	46.0	48	1.04	B
nmid	8,556	5,760	59,064	1.7	34	20.00	F
nysd	238,404	59,001	2,154,572	73.0	82	1.12	B-
ohsd	19,220	11,511	152,009	4.8	29	6.04	D-
ord	0	112,762	112,762	1.2	0	na	S
paed	20,901	7,714	200,630	3.4	8	2.35	C-
pamd	13,423	6,174	118,659	4.2	3	0.71	B+
pawd	18,482	24,706	173,301	5.3	21	3.96	D
prd	19,776	44,713	106,216	2.6	12	4.62	D
rid	90,187	60,890	622,755	20.0	104	5.20	D
txsd	35,279	9,878	239,544	11.0	0	0.00	A+
vaed	2,367	267,111	281,857	3.9	12	3.08	D+
vtd	100,015	69,718	692,811	34.0	5	0.15	A
Total	2,706,431	1,813,221	19,856,160	735.9	1,669		

Figure 1
Rate of Incidents v. Size of Archive

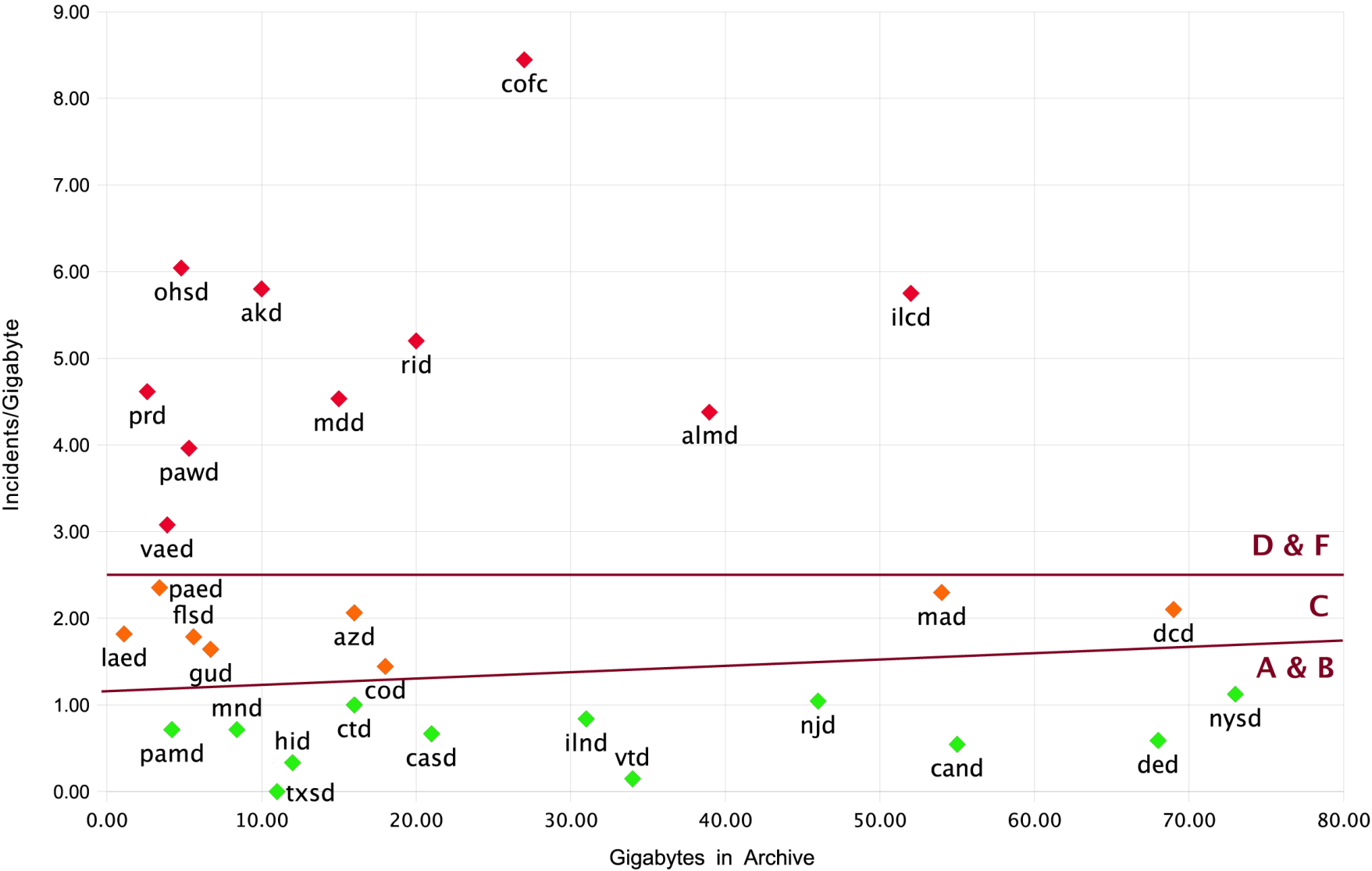


Figure 2
Distribution of Privacy Violations By Geographic Area

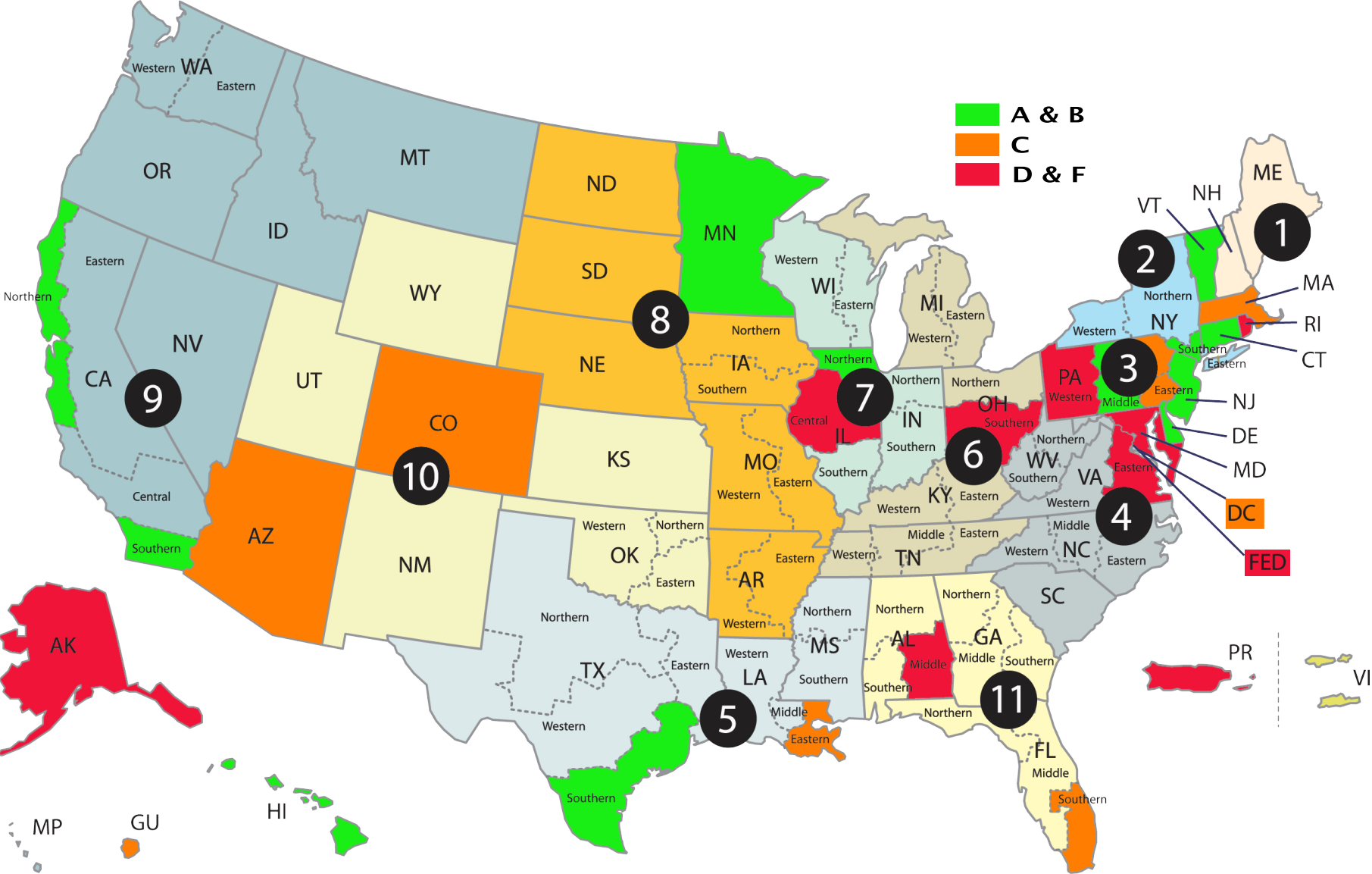


Figure 3
Number of Documents With Social Security Numbers Found By Month Of Filing

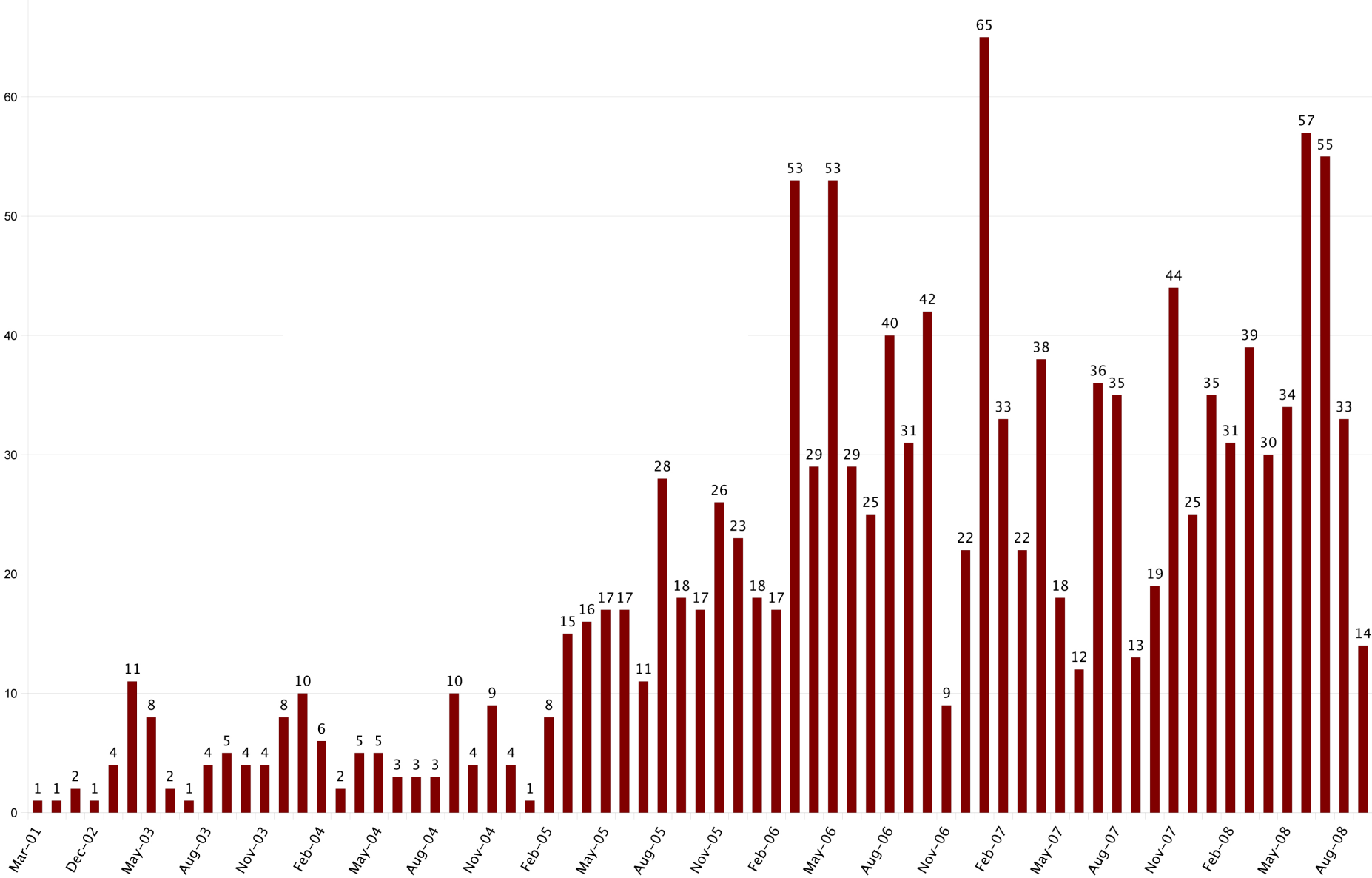
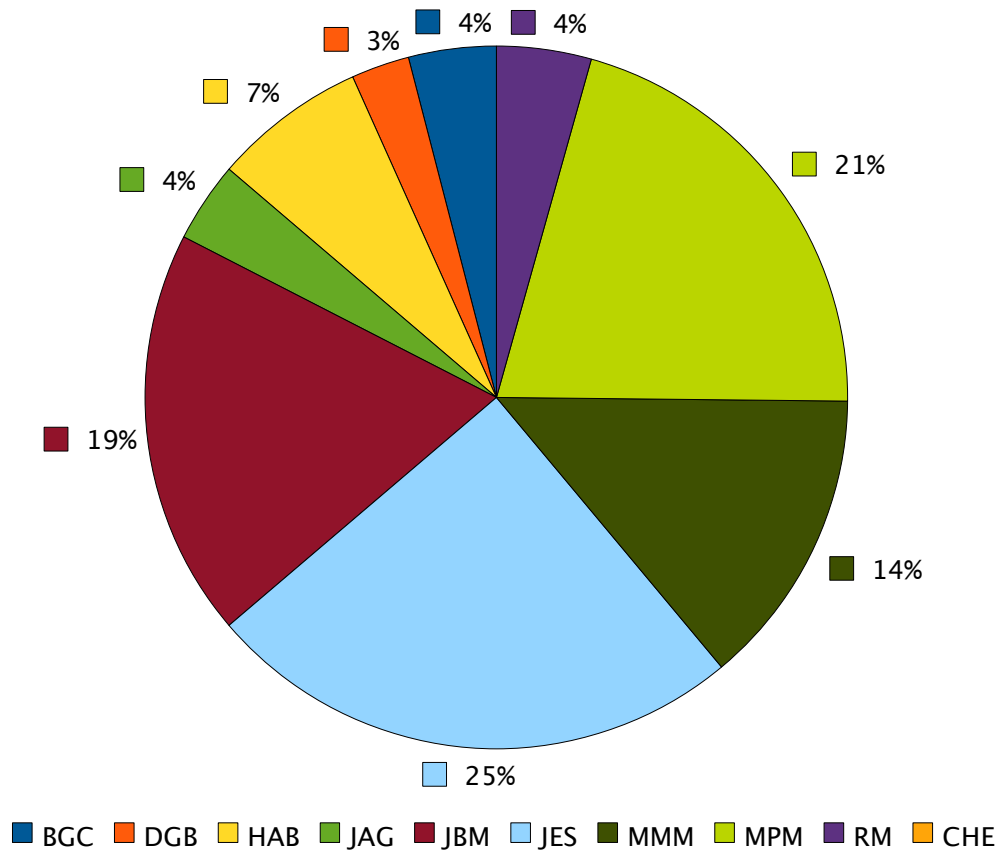
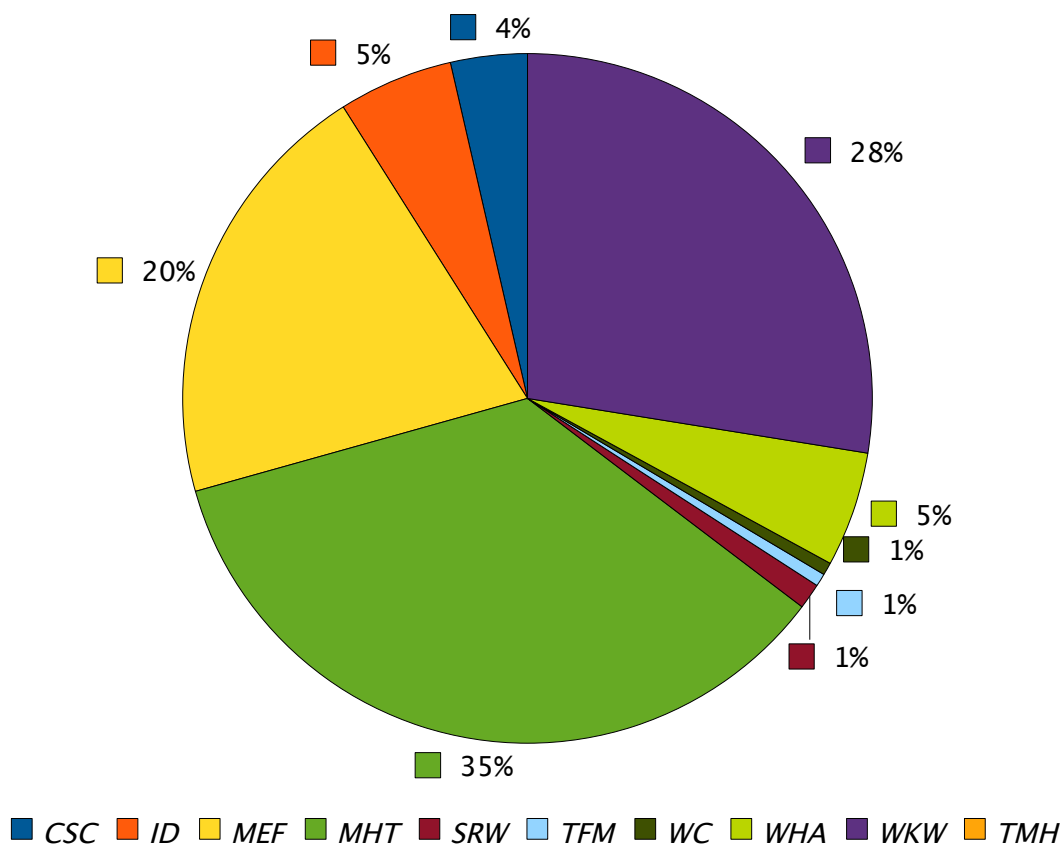


Figure 4
Number of Documents With Social Security Numbers
 District Court of the Middle District of Alabama



Initials	Judge	# of Docs	Percent
BGC	The Hon. Byron G. Cudmore	12	4%
DGB	The Hon. David G. Bernthal	8	3%
HAB	The Hon. Harold A. Baker	21	7%
JAG	The Hon. John A. Gorman	11	4%
JBM	The Hon. Joe Billy McDade	56	19%
JES	The Hon. Jeanne E. Scott	74	25%
MMM	The Hon. Michael M. Mihm	41	14%
MPM	The Hon. Michael P. McCuskey	62	21%
RM	The Hon. Richard Mills	13	4%
CHE	The Hon. Charles H. Evans	0	0%

Figure 5
Number of Documents With Social Security Numbers
 District Court of the Middle District of Alabama



Initials	Judge	# of Docs	Percent
CSC	The Hon. Charles S. Coody	6	4%
ID	The Hon. Ira DeMent	9	5%
MEF	The Hon. Mark E. Fuller	34	20%
MHT	The Hon. Myron H. Thompson	59	35%
SRW	The Hon. Susan Russ Walker	2	1%
TFM	The Hon. Terry F. Moorner	1	1%
WC	The Hon. Wallace Capel, Jr.	1	1%
WHA	The Hon. W. Harold Albritton	9	5%
WKW	The Hon. W. Keith Watkins	46	28%
TMH	The Hon. Truman M. Hobbes	0	0%

Pages 10-84 of this document, which consisted of an Appendix listing redacted personal identifying information, have been removed from the Exhibit.