Finding and Analyzing Data on FEC.gov

Updated 3/6/2015; template links updated 8/12/2016

Christopher Schnaars, USA TODAY
cschnaars@usatoday.com, @chrisschnaars

The Federal Election Commission offers several flavors of campaign finance data. You can view an individual report in PDF or HTML format, look at charts and maps summarizing campaign finance data or search for specific contributions and expenditures. You also can download datasets ranging from the raw text data for an individual report to a summary file containing all individual contributions made during the two-year election cycle.

There's a lot to digest and a lot of pitfalls that can trip you up along the way. Your best bet is to read this guide and look over the URLs provided below, then spend some time poking around the FEC website to familiarize yourself with all the other cool stuff I left out. If you find something particularly interesting, or if you have any comments or suggestions for this guide, please email me: cschnaars@usatoday.com

If you need help, send me an email or post a message on the NICAR listserv. There are lots of people with NICAR – Derek Willis, Aaron Bycoffe, Jacob Fenton and Scott Klein, among others – who have taught me a lot about campaign finance data. Alternatively, you can reach Judith Ingram in the FEC's press office at JIngram@fec.gov or (202) 694-1220. She's extremely helpful and can put you in touch with the FEC's data folks.

A note about committees:

Throughout this document, I refer to committees. In this context, a committee is any entity required to file campaign finance reports with the FEC, which assigns each committee a "C-number" (the letter C followed by an 8-digit number). A committee could be a candidate's principal campaign committee, a legislator's leadership political action committee (PAC) or joint fundraising committee, a PAC for a corporation or union, or an independent expenditure only committee (sometimes referred to as a Super PAC). Entities not required to file reports with the FEC that detail contributors – most notably 501(c)(3) and 501(c)(4) non-profits – don't apply here, and you won't find contributor data for these groups on the FEC website other than, perhaps, contributions from those nonprofits to committees that file reports.

Candidates for political office also receive an ID number from the FEC. These IDs are similar to committee IDs but are prefixed with P, H or S (rather than C) depending on whether the person is a candidate for president, the House or the Senate. The FEC assigns multiple IDs to candidates who run for more than one office.

A note about the Senate: They suck (albeit a little less)

My favorite rant about campaign-finance data is the Senate. While this is the 21st century and virtually every Senate candidate surely uses a computer to track contributions and generate these reports, senators still allow themselves to file reports on paper, thereby obscuring the identities of their donors and making that data a lot more difficult to work with. The FEC, in turn, spends taxpayer dollars to pay contractors to input the once-computerized data from these paper reports. A limited number of these input fields show up in the weekly master data files (discussed below), but because of the lag between when the FEC gets the reports and how long it take a contractor to input the data, weeks or even months can pass before any of that data is available electronically.

In late 2014, the FEC made a slight improvement to this system: You now can download individual reports electronically once they have been keyed in. You're still stuck in terms of having to wait a while for this data to be keyed in, but individual reports contain a lot more data than the subset of fields housed in the weekly master files, most notably donor addresses and parsed donor names.

Of course, if all you want to do is eyeball a report, PDF images of Senate reports -- which often are several thousand pages -- can be viewed on the FEC website, but it usually takes a week after the filing deadline for them to show up. (A handful of senators opt to file their reports electronically. You can search for these “unofficial” reports here: http://docquery.fec.gov/senate)

Now I'm going to hop onto my soapbox for a minute: Too often, the end result of the long delay in the availability of Senate data is we ignore it because by the time it becomes available, it's old and we've moved on to other things. I'm as guilty of this as everyone else, and we shouldn't let them get away with it. Look at the data, even if it's old. Write a story about the data, even if it's only a brief. If a senator or Senate candidate you're covering doesn't file "unofficial" electronic reports, ask for them electronically. If he or she doesn't comply, ask why and write about that. I'm doubtful the Senate will ever surrender its paper-filing privileges, but it definitely won't happen without an incentive for them to do so, namely pressure from the media and an indifferent public.

1) Third party data

If you want to save yourself a lot of pain, you don’t need to see campaign finance reports the day they’re filed and don’t have issues crediting a third party, there is a lot of very good data available from the Sunlight Foundation (http://sunlightfoundation.com) and the Center for Responsive Politics (http://www.opensecrets.org) that you can download and use for free.

The Sunlight Foundation's Influence Explorer site (http://realtime.influenceexplorer.com/download-index), for example, boasts bulk FEC data files for Schedules A, B and E as well as data on a per PAC or per candidate basis. New data often is available within 12 hours of when a report is filed.

You’d be wise to check out these websites before reading any more of this guide and deposit it in the nearest recycling bin if Sunlight or CRP data meet your needs.

Note that just as with any other data, if you use data compiled by the Sunlight Foundation or CRP, you should make sure your story and any related graphics credit the organization for its work.

2) Weekly master files

If all you care about is contributions, and you don’t want to use data from a third party, the fastest, easiest way to get the data you need is to download the master contribution files compiled by the FEC. You can find them at: http://www.fec.gov/finance/disclosure/ftpdet.shtml

You’ll see there’s a link for seven files per election cycle. Each of these summary files is useful, but the four files I’m going to focus on now are highlighted below:

2015 - 2016 Data Files

Name	Data File	Total Records	Updated	Format Description
Committee Master File	cm16.zip	11455	06-MAR-2015	CM Data Dictionary
Candidate Master File	cn16.zip	3552	06-MAR-2015	CN Data Dictionary
Candidate Committee Linkage File	ccl16.zip	3959	06-MAR-2015	CCL Data Dictionary
Any Transaction from One Committee to Another	oth16.zip	4042	01-MAR-2015	OTH Data Dictionary
Contributions to Candidates (and other expenditures) from Committees	pas216.zip	1947	01-MAR-2015	PAS2 Data Dictionary
Contributions by Individuals	indiv16.zip	27540	01-MAR-2015	INDIV Data Dictionary
Operating Expenditures	oppexp16.zip	7473	01-MAR-2015	OPPEXP Data Dictionary

Links to data dictionaries are provided in the far-right column.

The *indiv* file supposedly contains all individual contributions of $200 or more to federal committees. However, contributions that appear to be from corporations, PACs or other organizations sometimes are included in campaign-finance reports on the line set aside for itemized individual contributions and therefore show up in these master files. (You can use the Entity Type field to identify these.) Contribution refunds, usually found on Schedule B, also are included.

Similarly, the *pas2* file contains contributions to candidates from committees, and the *oth* file contains contributions from one committee to another.

NEW! The FEC recently added a weekly master file (*oppexp*) listing committees' operating expenditures, which are found on Schedule B. Also, the FEC has ceased offering three files listing adds, changes and deletes for the individual contributions file each week. These files were intended to help people manage weekly changes without having to re-download the whole *indiv* file each week. This isn't much of a loss, however. I don't know anyone who used those files, particularly given the messy, error-prone process of trying to update your database each week. Despite the file's size, it's much easier to just reload fresh individual contributions data each week.

Besides links to the files, the chart above contains the record count for each file, the date the file was last updated and links to data dictionaries, where you can get the data type and a brief description of each field. The contributions and expenditures files are updated every Sunday. The candidates, committees and candidate-committee linkage files (which I don't discuss here) are updated daily.

The rest of this guide largely focuses on how to download and analyze individual reports. But if you find the weekly master files meet your needs, you most likely are better off using them rather than individual filings. Before you go, though, let's take a quick look at the pros and cons of each.

3) Using the weekly master files vs. individual reports

Here are some reasons you might want to use the weekly master files discussed above rather than download individual filings:

1. Fewer files: You need to download only three files to have all contributions of $200 or more for the election cycle.
2. Transaction types: The FEC adds a Transaction Type code to each row. This field can be used to characterize the type of contribution or disbursement on a much more granular level than just

using the line number field contained in the original filing. Presently, there are about 75 different transaction types, though you'll find fewer than half of them in the master files. The data dictionary for each master file lists the specific transaction types included in that file. A complete list of transaction type codes can be found at: http://www.fec.gov/finance/disclosure/metadata/DataDictionaryTransactionTypeCodes.shtml

3. Refunds included: An important gotcha when you use the original filings is you have to be careful to check Schedules A and B for contribution refunds. Refunded contributions are included in the master files for you. To find refunds, look for Transaction Type 22Y.
4. No amended reports: If a report has been amended, only contributions from the most recent reports are included in the master files. When you work with individual electronic filings, you always have to check to see if a report has been amended.
5. FEC vetting: In addition to the transaction type codes, the FEC looks for missing or incorrect candidate and committee IDs and adds this data to the master files.

Here are some advantages to using individual filings rather than the master files:

1. Timeliness: You can download these filings as soon as they're uploaded to the FEC website whereas the master files are generated only once a week (Sunday afternoons).
2. Address information: The master files contain a donor's city and state but omit street address data found in individual filings.
3. Names parsed: The individual filings include separate fields for first, middle and last names as well as prefix and suffix and a completely separate field for organization names. In the master files, names are concatenated into one field.
4. More forms and schedules: The master files include contributions, operating expenditures and refunds from Schedules A and B and some data from Schedules E (independent expenditures) and F (coordinated expenditures). All data found in another 12 schedules are excluded, though other than loans, you probably won't care about most of this missing data anyway. Additionally, while this guide focuses on Form 3 and all related schedules, there are other interesting forms you won't find in the master files, including committee statements of organization (Form 1), 24- and 48-hour contribution and expenditure notices, statements of candidacy (Form 2), lobbyist bundlers and the full text of miscellaneous correspondence sent to the FEC (Form 99).
5. More contributions: I'll discuss the technicalities of when a contribution has to be itemized later in this guide, but the important distinction here is that itemized contributions found in electronic filings are found in the master files only if they are for $200 or more.
6. More fields: The master files contain fewer fields than the individual filings. Whereas Schedule A offers nearly 50 fields, for example, fewer than two dozen are included in the master files (though other than street address and parsed names, all the important ones are there).
7. Smaller: If you only want specific reports or committees, you can download only what you need. The master files include all contributions for the entire election cycle, so the files are very large.

4) How to find a committee's reports

Now that I've shown you the quick and dirty way to get at campaign contributions, let's dive into working with individual filings. Again, if after reading the pros and cons of each type above you think the weekly master files will serve your needs, you can skip most of the rest of this guide (though in the next section, I'll point you to some additional files that can give you summary numbers for each report, such as total contributions, cash on hand and total disbursements).

To find an individual report, go to www.fec.gov, where you'll see a menu on the left side of your screen. Hover over (but don't click) Campaign Finance Disclosure Portal, then Searchable Systems, then click Candidate & Committee Viewer (screenshot below). Or you can skip all that and just click/type/bookmark: http://www.fec.gov/finance/disclosure/candcmte_info.shtml

On the search page that comes up, you can enter an ID or partial name for any candidate or committee. If you type "obama" (without the quotes) and click the Get Listing button, you'll see four entries for Barack Obama near the top of your search results: Three rows for his three candidate IDs (one each for President, House and Senate) and a listing for Obama for America, his principal campaign committee in the presidential race. Each candidate has exactly one principal campaign committee. This committee, rather than the candidate individually, is the fundraising arm that accepts campaign contributions.

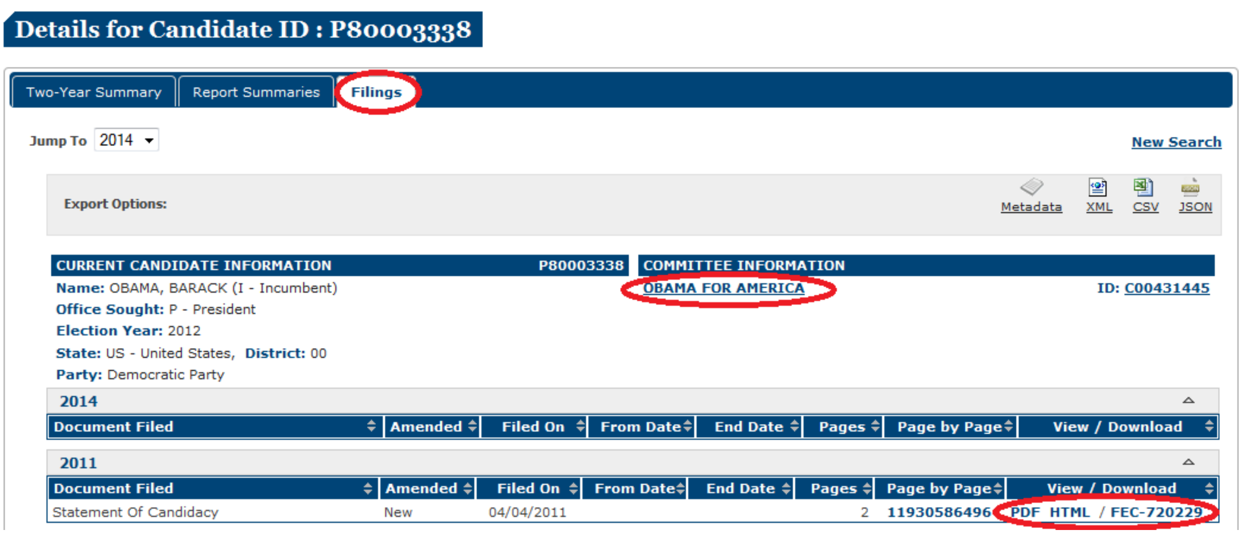
Navigating to the Candidate & Committee Viewer

Performing a search using the Candidate & Committee Viewer

If you click the link for P80003338, then select the Filings tab, you'll see only Statements of Candidacy (also known as Form 2). On the far-right of each row, you'll see a link to view that form in PDF or HTML format or to download it as text. Ignore the text link for now and click either the PDF or HTML link to see a printer-friendly version of the report.

There's not much to it, but you'll see Joe Biden listed as Obama's running mate. You also will see the campaign mailing address, the name of Obama's principal campaign committee (Obama for America) and a reference to a second authorized committee called Obama Victory Fund. This is a joint fundraising committee. It allows donors to write single big checks to one group which in turn disburses that money up to legal contribution limits to a host of other organizations, such as Obama's campaign committee, the Democratic National Committee and state Democratic parties.

Notice that Form 2 contains no financial information and there are no financial reports on this page. That's because you're looking at the candidate rather than his campaign committee. To see financial reports, you can either look under Committee Information (near the top-right corner of your screen) and click the committee's hyperlink or go back to your search results and find a link there.



The Filings tab for Barack Obama as a candidate for president, with links to his statement of candidacy and campaign committee.

Before we look at a financial report (Form 3), here's your first curveball: There are four different versions of Form 3. The version filed depends on the type of committee filing the report. Candidate committees for the House and Senate use Form 3; presidential campaign committees use Form F3P; PACs, Super PACs and other committees use Form F3X and lobbyists use Form F3L. I'll discuss these distinctions a little later. For now, all you need to know is that while they contain similar information (contributions, disbursements, loans and so on), there are subtle and some not-so-subtle differences. (Note that lobbyists are a different animal, and their reports aren't covered in this tipsheet.)

One nice consistency, though, is the different flavors of Form 3 share the same subforms. Just like a 1040 tax form, Form 3 has a bunch of optional "schedules," such as Schedule A (contributions), Schedule B (disbursements) and Schedule C (loans). The schedules go all the way up to L, but you'll rarely use anything but A and B. I'll come back to this later as well.

Now let's look at a report. Click the link for Obama for America, then click the Filings tab. Look for (but don't click) the July 2013 quarterly report. Do you notice anything? There actually are FOUR reports (screenshot below).

The first report, filed on 7/15/2013, is the original filing. The next row is a request for additional information (RFAI) from the FEC, and the third and fourth rows are amended reports. It's not that interesting, but if you click the PDF link for the RFAI, you'll see the FEC had questions concerning some contribution refunds and a debt. The committee then filed an amended report to address the FEC's questions, but it's likely the new report contains more than that: Look at the page counts for the original report and the amended report. The amended reports contain an additional 228 pages.

2013

Document Filed	Amended	Filed On	From Date	End Date	Pages	Page by Page	View / Download
Miscellaneous Report to FEC		01/02/2013			1	13960013520	PDF HTML / FEC-843659
Miscellaneous Report to FEC		01/07/2013			1	13960029195	PDF HTML / FEC-845086
Filing Frequency Change Notice		01/31/2013			1	13960642116	PDF HTML / FEC-853115
Acknowledgement of Filing Frequency Change		02/08/2013			2	13330022318	PDF
Statement Of Organization	Amend	03/01/2013			4	13961107507	PDF HTML / FEC-858885
Miscellaneous Report to FEC		04/12/2013			2	13961599729	PDF HTML / FEC-865465
Miscellaneous Report to FEC		05/20/2013			1	13940756991	PDF HTML / FEC-871938
Statement Of Organization	Amend	08/27/2013			4	13964573241	PDF HTML / FEC-886981
Miscellaneous Report to FEC		10/01/2013			1	13941676154	PDF HTML / FEC-889570
Miscellaneous Report to FEC		10/01/2013			1	13941676155	PDF HTML / FEC-889571
April Quarterly	New	04/15/2013	01/01/2013	03/31/2013	6612	13961631444	PDF HTML / FEC-866707
RFAI - April Quarterly		08/27/2013	01/01/2013	03/31/2013	15	13330037968	PDF
April Quarterly	Amend	12/13/2013	01/01/2013	03/31/2013	10310	13944033243	PDF HTML / FEC-897124
July Quarterly	New	07/15/2013	04/01/2013	06/30/2013	329	13941144551	PDF HTML / FEC-879003
RFAI - July Quarterly		08/27/2013	04/01/2013	06/30/2013	4	13330037989	PDF
July Quarterly	Amend	12/13/2013	04/01/2013	06/30/2013	557	13944032686	PDF HTML / FEC-897123
July Quarterly	Amend	03/28/2014	04/01/2013	06/30/2013	557	14960567272	PDF HTML / FEC-914130
October Quarterly	New	10/14/2013	07/01/2013	09/30/2013	620	13964684077	PDF HTML / FEC-891179
RFAI - October Quarterly		02/21/2014	07/01/2013	09/30/2013	3	14330044618	PDF
Year-End	New	01/31/2014	10/01/2013	12/31/2013	173	14940264637	PDF HTML / FEC-905573

Obama for American amended its July 2013 quarterly report twice after the FEC requested more information.

Amended reports are filed ALL THE TIME, and this often is done to correct mistakes or add additional information without any prompting from the FEC. So if you write a story using numbers pulled from one of these reports, it's a good idea to check the FEC website for amended reports before you publish. Sometimes amended reports are filed years after the initial filing. In this example, the most recent amendment was filed eight months after the original. Even the first amendment filed to address the FEC's concerns took five months.

5) How to read a report summary

Now click the PDF link for the amended report filed on 12/13/2013. Generally speaking, if you want to look only at summary numbers atop a report, the HTML link is better. Whereas the PDF link will load the entire report into your browser, the HTML version displays only the topline numbers for the report and provides links to specific line numbers within the various schedules. Note that Obama's post-general election report is more than 300,000 pages! The July 3013 report is small enough that you can grab the whole PDF, though.

SUMMARY

6.	CASH ON HAND AT BEGINNING OF REPORTING PERIOD	588173.98
7.	TOTAL RECEIPTS THIS PERIOD (From Line 22, Column A, Page 3)	2163181.33
8.	SUBTOTAL (Lines 6 and 7)	2751355.31
9.	TOTAL DISBURSEMENTS THIS PERIOD (From Line 30, Column A, Page 2)	2450850.11
10.	CASH ON HAND AT CLOSE OF THE REPORTING PERIOD (Subtract Line 9 from 8	300505.20
11.	DEBTS AND OBLIGATIONS OWED TO THE COMMITTEE (Itemize All on Schedule C-P or Schedule D-P)	0.00
12.	DEBTS AND OBLIGATIONS OWED BY THE COMMITTEE (Itemize All on Schedule C-P or Schedule D-P)	3454778.23
13.	EXPENDITURES SUBJECT TO LIMIITATION	0.00

You'll find Summary numbers near the top, but Cash On Hand is the only really useful number here. Dig deeper for better stuff.

At the top of the report, you'll see basic information, such as the committee's name and ID number, the treasurer's name and the period the report covers (in this case, 4/1/2013 through 6/30/2013). Reports generally cover a month, a quarter or six months.

The next page has a Summary section, but the only number of interest here is line 10: cash on hand at the close of the reporting period. If you're covering campaigns on deadline and looking at a recently filed report, cash on hand is a way to take a quick pulse of the committee's war chest going into the next period. Do they have millions of dollars to spend on media buys and voter outreach, or are they broke?

You might think total receipts and total disbursements are interesting numbers, but they contain a lot of extraneous stuff, such as candidate loans (for receipts) and contribution refunds (for disbursements). It's better to bypass these summary numbers and go a little deeper into the report.

The next gold nugget is Line 17, which is broken down into individual contributions (itemized and unitemized), as well as contributions from political party committees (such as the DNC and the RNC), other political committees (PACs) and the candidate. Line 18 shows transfers from other authorized committees, which typically are joint fundraising committees.

I. RECEIPTS	COLUMN A Total This Period	COLUMN B Election Cycle-to-Date
16. FEDERAL FUNDS (Itemize on Schedule A-P)	0.00	0.00
17. CONTRIBUTIONS (other than loans) FROM:		
(a) Individuals/Persons Other Than Political Committees		
(i) itemized	7349.00	526055.14
(ii) unitemized	8127.16	1236355.43
(iii) Total contributions	15476.16	1762410.57
(b) Political Party Committees	0.00	0.00
(c) Other Political Committees	0.00	0.00
(d) The Candidate	0.00	0.00
(e) TOTAL CONTRIBUTIONS (other than loans) (Add 17(a), 17(b), 17(c) and 17(d))	15476.16	1762410.57
18. TRANSFERS FROM OTHER AUTHORIZED COMMITTEES	0.00	8700000.00

Line 17 (or Line 11 for House, Senate and PACs) is where you can get details on individual and PAC contributions.

Notice that there are two columns of numbers. The first column shows amounts for the three-month reporting period. The second column shows numbers to-date for the election cycle.

Here’s another pitfall to be aware of: While candidate committees’ election cycles are two years or more, the clock resets for PACs and Super PACs – including the RNC and DNC – every January 1. This means that if you want to look at how much money a PAC received and spent during the 2014 election cycle, you might want to look at the PAC’s year-end report for 2013 AND 2014.

The difference between itemized and unitemized contributions can be confusing. Let’s demystify it now: Contributions must be itemized when the total amount given (the aggregate) by a particular donor EXCEEDS $200. Let’s use an extreme example: If John Smith makes four $50 contributions to Obama for America, all of that money can be counted as unitemized contributions. But John really wants to give all he can, so he digs into his couch cushions, finds $1 in change and mails it to Chicago. That $1 – and only that $1 – must be itemized while the original $200 can remain unitemized. When you start looking at the Schedule A records a little later in the guide, you’ll see that contribution records include two dollar figures: the amount of the specific contribution and the aggregate amount given during the election cycle. John's latest contribution would be listed as $1, with an aggregate of $201. (Occasionally, committees voluntarily itemize contribution aggregates of less than $200, but this is rare.)

Looking at the Line 17 numbers as a group can tell you a little about a candidate’s financial support. One calculation I often do is Line 17a2 (unitemized contributions) divided by Line 22 (total receipts) to show how much of a candidate’s money comes from small contributions, indicating the level of grassroots support. Comparing 17a1 with 17a2 can show you to what extent wealthier donors back the candidate. Line 17d shows whether the candidate is fueling his or her own campaign. Line 17c shows how much money the candidate got from special interests (PACs), though you should dive into the Schedule A data and look at the details to be sure.

Under disbursements (lines 23 to 30), you generally care most about operating expenditures on Line 23, but lines 24 and 25 will tell you if the committee is giving money to other committees and how much it's spending on fundraising. If the committee is in debt, you'll want to pay attention to Line 27.

The FEC also compiles master files summarizing the topline numbers for all electronically filed reports. You can find those summary files at http://www.fec.gov/finance/disclosure/ftpsum.shtml. One nice thing about these files is they are updated DAILY, so they're very current.

If you think you've got a handle on everything so far, let's mix things up a little.

6) Dealing with multiple versions of the same form
As I mentioned above, presidential candidates, House and Senate candidates, PACs and lobbyists each use a slightly different version of Form 3. For presidential candidates, contributions are on Line 17. House and Senate candidates and PACs (forms F3 and F3X) use line 11.

Make sense? Well, there's more. From time to time, the FEC changes the headers for one or more forms. The most recent header version is 8.1: the fourteenth version of headers released to date. When you consider all the different header iterations of all the different forms required by the FEC, there presently are 664 different sets of headers.

However, while this demonstrates a large variation in the data's format over time and a potential complication to be aware of, generally it's not a big deal and matters only when you're downloading data. If you're just looking at the PDF or HTML version of a single report, you can clearly see what each number means. Additionally, not all forms change with each iteration. The headers for all but three forms remained exactly the same when the FEC moved from v8.0 to v8.1 in January 2014.

I know I've been saying this a lot, but I'll come back to this point later. For now, I'm going to give you some resources you can use to see the different forms and header versions. Bear with me. It's important, and the reasons why will be clear very soon.

First, here are links to blank PDFs for Form 3, each of which includes all related Schedules. You should download or at least bookmark these:

- Form 3 (House/Senate): http://www.fec.gov/pdf/forms/fecfrm3.pdf
- Form 3P (President): http://www.fec.gov/pdf/forms/fecfrm3p.pdf
- Form 3X (PAC): http://www.fec.gov/pdf/forms/fecfrm3x.pdf
- Form 3L (Lobbyist): http://www.fec.gov/pdf/forms/fecfrm3l.pdf

Now let's get information on the various headers. If you go to http://www.fec.gov/finance/disclosure/ftpefile.shtml and scroll down, you'll see a chart listing each version of the headers and the dates during which each version was used.

Electronic Filing Format	Start Date	Stop Date
V1	06/20/1997	01/24/2001
V2	03/16/1998	01/31/2002
V3	03/02/1999	01/29/2004
V5.0	03/09/2003	02/09/2004
V5.1	12/22/2003	02/11/2005
V5.2	02/02/2005	02/13/2006
V5.3	02/06/2006	01/02/2008
V6.1	01/02/2008	06/08/2008
V6.2	06/09/2008	03/09/2009
V6.3	03/10/2009	05/13/2009
V6.4	05/14/2009	04/04/2011
V7.0	04/05/2011	09/25/2011
V8.0	09/26/2011	01/13/2014
V8.1	01/14/2014	current

Various header formats and the dates they were used.

Near the top of the webpage, you'll see a link for "Filing Format Documentation." Click this link or just go to http://www.fec.gov/elecfil/eFilingFormats.zip to download the headers. When you unzip this file, you'll see a file called: "e-filing headers all versions.xlsx" Open this file and go to the second tab (all versions). You'll see the headers for each version of every form. The other tabs are form-specific and don't contain any additional information.

	A	B	C	D	E	
1	v8.1	F1	1-FORM TYPE	2-FILER COMMITTEE ID NUMBER	3-CHANGE OF COMMITTEE NAME	4-COMM
2	v8.0	F1	1-FORM TYPE	2-FILER COMMITTEE ID NUMBER	3-CHANGE OF COMMITTEE NAME	4-COMM
3	v7.0	F1	1-FORM TYPE	2-FILER COMMITTEE ID NUMBER	3-CHANGE OF COMMITTEE NAME	4-COMM
4	v6.4	F1	1-FORM TYPE	2-FILER COMMITTEE ID NUMBER	3-CHANGE OF COMMITTEE NAME	4-COMM
5	v6.3	F1	1-FORM TYPE	2-FILER COMMITTEE ID NUMBER	3-CHANGE OF COMMITTEE NAME	4-COMM

A snip of the worksheet listing the headers for all versions of all FEC forms.

Open the FEC_v8x folder, then open: "FEC_Format_v8 1.xlsx" This file gives you a summary of all changes made since v8.0. As it turns out, changes were made only to Form 5, Form 5.7 and Schedule D. So all versions of Form 3 and all Schedules except D remained the same.

The various tabs in this Excel file also show the schema for each form and schedule, which you'll need if you want to import campaign-finance data into Access or some other database manager. Sadly, the documentation for each field is scant at best. Looking at the blank forms (URLs provided above) is a good tool to help you puzzle out what a particular field means.

If this seems a little overwhelming, don't worry. I'm showing you all the header versions and schema now so you'll have these resources handy later in the tutorial. For now, all you really need to keep in mind is that headers change over time, and now you have the tools you need to discern the differences when it matters.

7) How to inspect the detailed data in schedules A, B and so on

The text data for an individual filing can be confusing because that one file contains the summary numbers and all schedules. Each schedule can have a different number of fields as well as its own headers, none of which is in the file.

Let's look at an example. Download the amended post-general report for the Texas Democratic Party: http://docquery.fec.gov/showcsv/nicweb312910/908542.fec, then open the file in Excel or some other spreadsheet program. To make it easier, you can change the .fec extension to .csv for this exercise. If you open the file in Excel, you'll see something like this:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	HDR	FEC	8.1	Patton Te	3															
2	F3XN	C0009926	Texas Democratic Pa		4818 East Ben White		Austin	TX	78741	M2				20140101	20140131		Hinojosa	Gilberto		
3	SA11AI	C0009926	11ai-000391494			IND		Nettles	Scott				1701 Barn Swallow D		Austin	TX	78746			2
4	SA11AI	C0009926	11ai-000391565			IND		Whiteside	John				2519 Loving		Dallas	TX	75214			2
5	SA11AI	C0009926	11ai-000391845			IND		Book	Robert				7191 Kendallwood D		Dallas	TX	75240			2
6	SA11AI	C0009926	11ai-000391863			IND		Wells	Katheryn				9900 Oak Run Drive		Austin	TX	78758			2
7	SA11AI	C0009926	11ai-000391891			IND		Maloney J	Pat				239 E Commerce St		San Anton	TX	78205			2
8	SA11AI	C0009926	11ai-000391719			IND		Sims	John				P.O. Box 10236		Lubbock	TX	79408			2
9	SA11AI	C0009926	11ai-000391747			IND		Perez	Paul				11528 James Grant D		El Paso	TX	7.99E+08			2
10	SA11AI	C0009926	11ai-000391752			IND		Perez	Barbara				11528 James Grant D		El Paso	TX	7.99E+08			2
11	SA11AI	C0009926	11ai-000392603			IND		Nettles	Scott				1701 Barn Swallow D		Austin	TX	78746			2
12	SA11AI	C0009926	11ai-000391932			IND		Homer	Mark	S.			2735 Oak Creek		Paris	TX	75462			2
13	SA11AI	C0009926	11ai-000391946			IND		Neland	Linda				1500 Barton Creek Bl		Austin	TX	78735			2
14	SA11AI	C0009926	11ai-000391918			IND		Johnson	Luci	B.			114 W. 7Th St. Ste. 9		Austin	TX	78701			2
15	SA11AI	C0009926	11ai-000392647			IND		Murr	Priscilla				1710 Palma Plaza		Austin	TX	78703			2
16	SA11AI	C0009926	11ai-000391972			IND		Fearey	Patricia				1301 Spring St #24G		Seattle	WA	98104			2
17	SA11AI	C0009926	11ai-000392679			IND		Chapin	Jess				3900 Avenue C		Austin	TX	78751			2
18	SA11AI	C0009926	11ai-000392685			IND		Jones	Benjamin				2419 NW 61St Street		Seattle	WA	98107			2
19	SA11AI	C0009926	11ai-000392083			IND		Kostura	Judith	Ann			3214 Thousand Oaks		Austin	TX	78746			2

908542.fec from the Texas Democratic Party

At first glance, this doesn't look too bad. The first row is a header for the file. The only really useful piece of information is the third column, which tells you the version of headers used in the report (8.1). The second row is the longest: It contains all the topline numbers, and if you scroll all the way to the right, you'll see there are more than 120 columns.

The schedules data begins on the third row, and the key to deciphering each row lies in the first column. If you look at cell A3, you see: SA11AI. This means the row contains a Schedule A (SA) record that applies to line 11ai in the topline numbers. If you go to Form F3X used by all PACs, you'll see that itemized individual contributions are recorded on line 11ai.

I won't spend a lot of time on this file, but let's chop it up a little bit:

- Row 40 shows a $120 contribution from Cargas for Congress. In this file, the amount of the contribution is in Column U and the aggregate for the election cycle is in Column V.
- Rows 41 to 44 show more than $50,000 in transfers from Democratic groups.
- Rows 48 to 116 show that a bunch of groups – mostly state and local politicians – paid more than $75,000 for access to the group's voter and email lists. (This is noted in Column W.)

- Rows 117 to 137 itemize the group's operating expenses (Schedule B, Line 21b). As with Schedule A, some details concerning the transactions are in Column W.
- Rows 138 to 141 show contributions to federal candidates.
- Payroll is on lines 310 to 323 (Schedule B) and lines 406 to 422 (Schedule H4), with a Memo about payroll on line 480.

Those are the highlights. If you're curious about the rest, just use Column A to figure out what line number the row refers to, then go grab the appropriate headers.

A note about delimiters:
Revisiting the above example, if you go to the Candidate and Committee Viewer (http://www.fec.gov/finance/disclosure/candcmte_info.shtml), search for filings by the Texas Democratic Party (C00099267) and click the link for the February 2014 report (FEC-908542), you get a pop-up window asking whether you want a CSV (comma-separated values) or ASCII file. If you're going to view this data in Excel, pick CSV. But if you're going to import the data into a database manager other than Access, consider ASCII (specifically, ASCII 28).

As someone who has worked with data in various formats for more than a decade, I've always been mistrustful of CSV files. Because commas occur so frequently in data, you often have to surround your data fields with some kind of text qualifier. The FEC uses quotation marks as its qualifier. If a committee submitting an electronic report includes quotation marks in a field, however, it can cause all kinds of headaches for someone trying to parse the data.

The FEC's alternative to commas is the ASCII-28 delimiter. Unfortunately, I don't know of an easy way to import a file using this delimiter into Excel or Access, but if you're using a different database manager, ASCII-28 offers a couple advantages. First, you won't find ASCII-28 in any data field submitted to the FEC, so it parses quite smoothly. Second, because ASCII-28 doesn't show up in the data, it doesn't need text qualifiers (quotation marks) around every field, thereby reducing the size of the files. If you don't think this is a big deal, consider this: the CSV version of the Texas Democratic Party file you've been working with is **40 percent** larger than the ASCII-28 version. This size difference disappears once you convert the file to Excel format or import it into a database, but you still have to download all those files.

8) Viewing an electronic report in Excel
As you saw above, you can get a ton of information from the topline numbers of a campaign finance report. If you're writing a story on the filing deadline, when the reports can be submitted as late as midnight, this might be all you have time for. But your story is going to be just a numbers comparison with little or no context and – hopefully, at least – a few quotes. In short: it's going to suck.

If you can find as little as 10 minutes, you can dig into Schedules A and B to get some context and rich details that will give your story some depth and a diversion from just reporting a bunch of big numbers. These numbers also will help you pinpoint human sources for your story and provide ideas about what questions to ask

As you saw above, with a whole bunch of data from different forms with different headers cobbled together into one file, viewing Form 3 data in Excel is cumbersome at best. So let's deal with that problem once and for all by creating an Excel template.

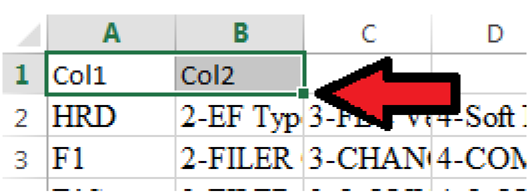
One caveat: Because even the latest version of Excel is limited to a little more than 1 million rows, this solution is not going to work if you are trying to look at a larger file with more than 1 million rows or if you are using a version of Excel older than Excel 2007 (where you're limited to 65,536 rows). In those cases, you're going to need to import the data into a database manager. This is covered later.

If you're lazy: I can't promise how long these links will be active, but here are pre-made templates for header versions 8.0 and 8.1. You can download these and skip ahead to the section on using the templates.

- bit.ly/Form3Template_v80
- bit.ly/Form3Template_v81

To create your template:

1. If you didn't already do so in Step 3, download the .zip file from the FEC containing all the header information (http://www.fec.gov/elecfil/eFilingFormats.zip), decompress it and open: "v8.1 e-filing headers .xlsx"
2. Make sure you're on the first tab: v8.1 headers.
3. Click cell B1.
4. Press Shift-Ctrl-End to highlight all your data except Column A.
5. Press Ctrl-C to copy this data.
6. Press Ctrl-N to create a new worksheet.
7. Click on cell A2. (You want to leave a blank row at the top.)
8. Type Ctrl-V to paste your copied data onto the new sheet.
9. We don't need Column B, so click the B heading to highlight all of Column B, right-click somewhere in this highlighted column and select Delete to delete the entire column.
10. Now let's create some generic headers. Go to A1 and just type a 1, Col1 or something similar, then go to B1 and type 2, Col2, etc.
11. Highlight A1 and B1, grab the fill handle (the little square in the bottom-right corner of the frame around the two highlighted cells) and drag it all the way to cell GX1. This will give you 206 generic column headings.

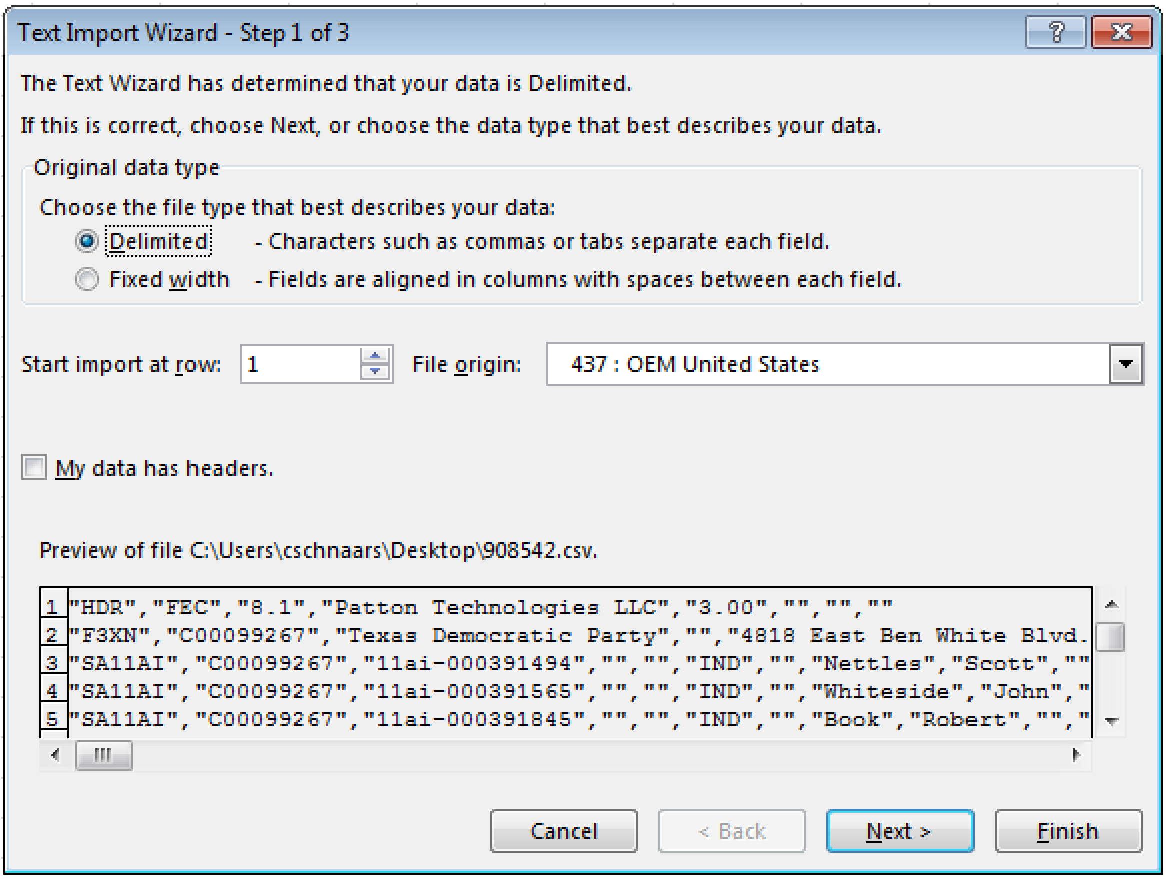


The Excel fill handle is a handy way to copy incremented data to other cells.

12. Go to File, Save As to save your template. Make sure you select the Excel Template file type and give it a name like Form3Template_v81.xltx.

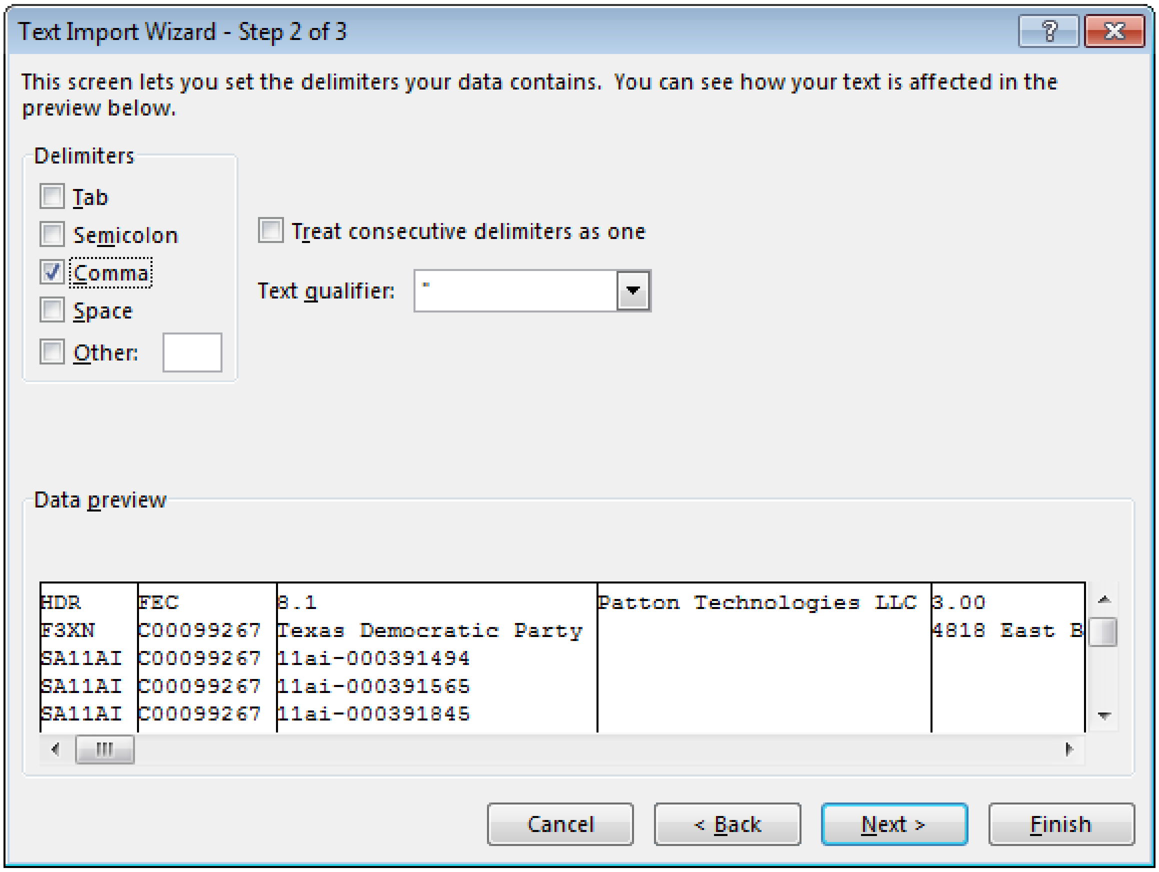
To use your template:

1. Download (but don't open) the report you want in CSV format.
2. Open your template.
3. To make sure you don't overwrite your template by accident, immediately go to File, Save As and save a copy of your template as a regular Excel Workbook. A good name is one that reflects the report you're going to view, such as: TexasDemPty_Feb2014Monthly_908542.xlsx
4. Click on the first blank cell in Column A, which should be A50.
5. Go to Data, Get External Data, From Text. Navigate to the report you downloaded.
6. Navigate to the electronic report you downloaded and double-click it to open the file.

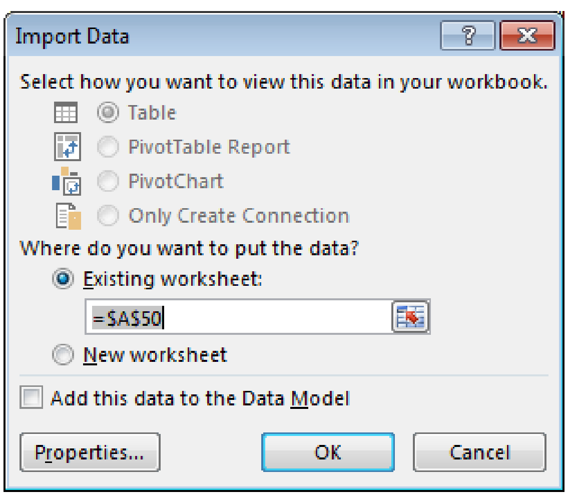


Step 1: Select Delimited data, no headers.

7. When the Text Import Wizard opens, make sure Delimited is selected and the "My data has headers" box is unchecked, then click Next.
8. In Step 2, uncheck the Tab delimiter, check the Comma delimiter and make sure the text qualifier is set to double quotation marks. Click Next.
9. In Step 3, leave the defaults and click Finish.

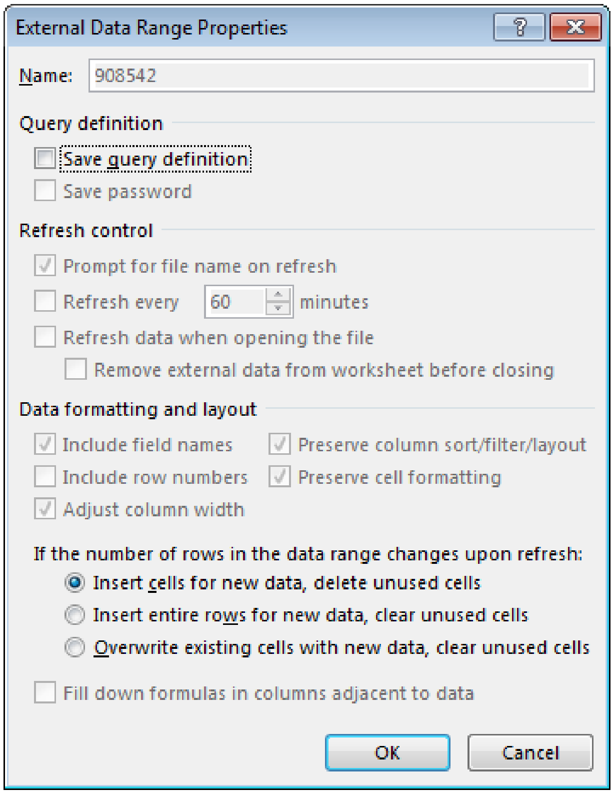


Step 2: Select comma delimiter and quotation marks as the text qualifier.



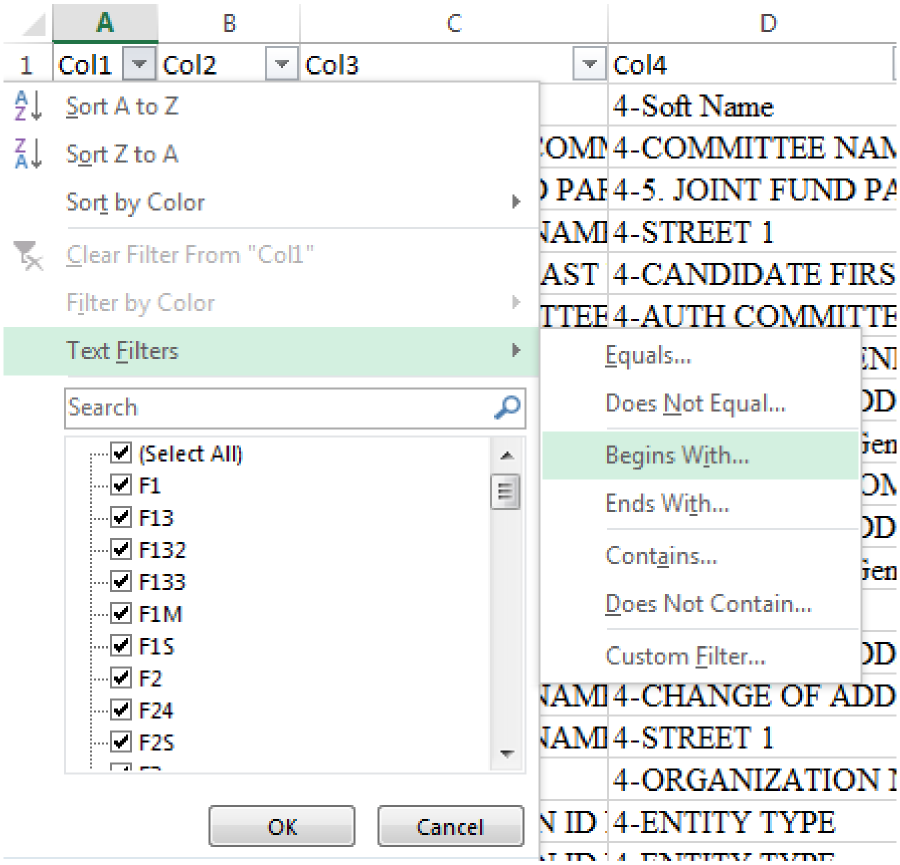
You might see an Import Data window after you click Finish.

10. If an Import Data window pops up, make sure “Existing worksheet” is selected and the cell address is the first blank cell (which should be “=A50”). If you want to disable the data connection between your Excel file and the source text (I would), click Properties, and uncheck “Save query Definition.” Click OK, then click OK again. Save your file.

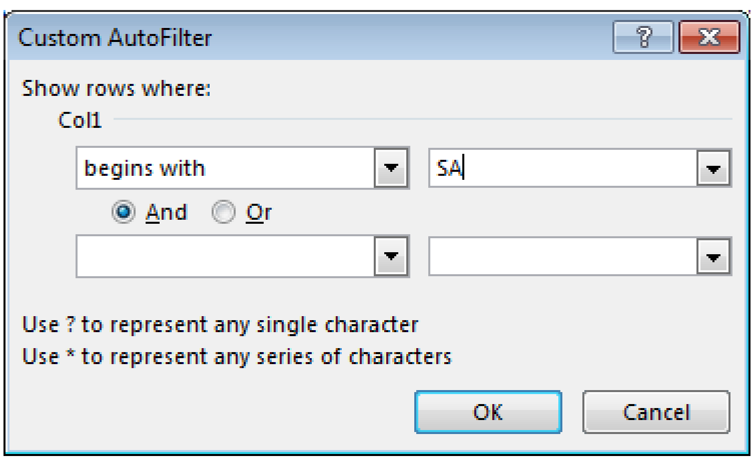


Use this window to break the connection between your Excel file and the text data.

11. Now you're ready to look at your data: Click a non-empty cell near the top of your spreadsheet, then go to Data, Filter. You'll see the generic column headings in the first row now have dropdown arrows.
12. Click the dropdown arrow for A1 and select Text Filters, then Begins With.
13. In the dialog that opens, enter "SA" (without the quotes), then click OK, and voila: You are now looking at just your Schedule A data, with a nice row of headers in the second row.
14. If you want to keep your headers visible at all times click the first cell of your first data row (below the headers). If you're using the Texas Democratic Party file, it should be A53. Then go to View, Freeze Panes.



Setting the text filter in your template.



Using our template to filter for Schedule A records.

This template is not ideal. You probably should avoid sorting, since you can scramble your data if you don't have everything selected, and this may be more likely to occur with some rows hidden by filtering. What's more, even with the filter on, you've got two rows of headers. Excel doesn't know the second row needs to stay at the top and will freely move it around your file if you're not careful.

If you want to keep the file long term, you could take the time to split this sheet up into separate tabs (or maybe just for Schedules A and B). To do that, turn off the Filter, highlight all your data and sort by

Column A to get your header row and all your data into one contiguous block. Then just copy and paste to a new tab.

Now that all your data is in one place, you can more easily examine it: Who are the big donors? Where are those donors from? (Are they out-of-state?) Are certain employers or industries providing a disproportionate amount of support? What PACs are giving to this group and vice versa? How much are they spending on their payroll? What other expenses do they have? Now that the data is in Excel, you can sort and filter to answer some of these questions and discover new questions.

If you like the template, you can create additional templates for older header versions (8.0 and so on), but you might not need them unless you want to go back and look at older reports. No reports filed after mid-January 2014 should use older headers, Schedule D is the only part of Form 3 that has changed since 8.0, and 8.0 has been in use since September 2011.

If you have any experience with Access, mySQL or other database managers, you're probably thinking about how much easier it would be to work with campaign finance data using one of those tools. Sadly, this is not an easy proposition. But before we discuss the pros and cons of putting campaign finance data into a database, let's look at another pitfall. Heck, let's look at two: Excluded rows and contribution refunds

9) Excluded rows

The Texas Democratic Party report you've been working with doesn't contain an example of what I'm about to cover. If you want to test the skills you've learned and make a template out of another report, download file 907289.fec, an amended 2013 year-end report for Minnesota Democratic-Farmer-Labor Party (C00025254). Otherwise, just read along and look at the screenshot.

You'll note that Column 43 in Schedule A is called MEMO CODE. This is a horrible name for this field, particularly since the next column is called MEMO. You might think MEMO CODE, which should always be blank or contain an X, just means there's something in MEMO, but that's not the case. MEMO CODE means this row is not included in the topline number referred to in the first column because the amount already is counted somewhere else.

This happens a lot with joint fundraising committees and other groups, such as ActBlue, that collect money from a bunch of donors, then write one big check to a campaign. In cases like this, the grand total is counted in the topline numbers (meaning the MEMO CODE field is blank). The excluded rows that identify the individual donors satisfy the disclosure requirements, and excluding those rows is a way to make sure no money is double-counted. However, if your goal is to identify individual donors and you omit any rows with an X in MEMO CODE, you're stripping out a lot of the data that identifies those donors and using a faceless aggregate.

In the case of the Minnesota Democratic-Farmer-Labor report, if you filter on Schedule A records with an X in MEMO CODE, then look at the MEMO field, you'll see notes pointing to Dollars for Democrats, Minnesota State Party Victory Fund and Minnesota Senate Victory 2014. You'll also note all of these rows point to Line 12, which is where money transferred from other committees is tallied.

If you look for SA12 records where the MEMO CODE is blank, you'll see a contribution from the Democratic National Committee and two contributions from Minnesota Senate Victory 2014. If you add these up, you get $25,275. The same amount is recorded on Line 12 at the top of the report.

If you add up the SA12 rows that have X in MEMO CODE (screenshot below), the total is more than $86,000, including about $57,000 in unitemized contributions from Dollars for Democrats and State Party Victory Fund. If you back these out, the sum you get is now within $4,000 of the amount on Line 12, but they still don't match. Maybe the group decided not to include contributions where the aggregate is less than $200. Backing these out gets the two numbers within about $2,000 of each other, but again: They don't match.

This kind of discrepancy can be immensely frustrating, particularly when you are trying to be exceedingly careful with an extremely complex dataset. Unfortunately, such discrepancies are the norm. If you really need the numbers to match, you'll need to call the organization. Otherwise, all you can do is get as close as you can and learn to be comfortable with being in the ballpark.

	A	H	I	O	P	U	V	AQ	AR
1	Col1	Col8	Col9	Col15	Col16	Col21	Col22	Col43	Col44
34	SA	8-CONTRIBU	9-CONTRI	15-CONTRIBUT	16-CONTRIBUTOR STATI	21-CONTRIE	22-CONTRIE	43-MEMO CODE	44-MEMO TEXT/DESCRIPTION
146	SA12	Anderson	Ronald	Deerwood	MN	$ 75.00	$ 75.00	X	* Dollars for Democrats
147	SA12	Batty	Ronald	Golden Valley	MN	$ 100.00	$ 100.00	X	* Dollars for Democrats
148	SA12	Bergland	Bob	Roseau	MN	$ 20.00	$ 85.00	X	* Dollars for Democrats
149	SA12	Bergland	Bob	Roseau	MN	$ 20.00	$ 85.00	X	* Dollars for Democrats
150	SA12	Bergland	Bob	Roseau	MN	$ 20.00	$ 85.00	X	* Dollars for Democrats
151	SA12	Bohne	Cleo	New Ulm	MN	$ 47.50	$ 47.50	X	* Minnesota State Party Victory Fund
152	SA12	Borrud	Aleta	Rochester	MN	$ 475.00	$ 2,725.00	X	* Minnesota State Party Victory Fund
153	SA12	Brandes	Judith	Elk River	MN	$ 100.00	$ 100.00	X	* Dollars for Democrats
154	SA12	Brooker	Charlotte	Maplewood	MN	$ 95.00	$ 595.00	X	* Minnesota State Party Victory Fund
155	SA12	Dahlberg	Megan	Roseville	MN	$ 475.00	$ 475.00	X	* Minnesota State Party Victory Fund
157	SA12			Washington	DC	$ 46,755.00	$ 188,189.02	X	* Dollars for Democrats
158	SA12	Eichelberger	Louise	Willmar	MN	$ 26.00	$ 356.00	X	* Dollars for Democrats
159	SA12	Eichelberger	Louise	Willmar	MN	$ 26.00	$ 356.00	X	* Dollars for Democrats
160	SA12	Eichelberger	Louise	Willmar	MN	$ 26.00	$ 356.00	X	* Dollars for Democrats
161	SA12	Engleson	Russel	Willmar	MN	$ 95.00	$ 485.00	X	* Minnesota State Party Victory Fund
162	SA12	Fairchild-Ehm	Audrey	Roseville	MN	$ 71.25	$ 313.75	X	* Minnesota State Party Victory Fund
163	SA12	Fairchild-Ehm	Audrey	Roseville	MN	$ 142.50	$ 313.75	X	* Minnesota State Party Victory Fund
164	SA12	Hall	Marion	Minneapolis	MN	$ 475.00	$ 475.00	X	* Minnesota State Party Victory Fund
165	SA12	Heaney	Timothy	Falcon Heights	MN	$ 95.00	$ 145.00	X	* Minnesota State Party Victory Fund

907289.fec: A sample of data rows excluded from Schedule A, Line 12.

10) Contribution refunds

If you're bleary-eyed from reading this far, consider this: I think the only time (knock, knock) I've had to run a correction related to an analysis of FEC data was when I missed a contribution refund, thereby overstating the amount of money a donor had given to Super PACs.

If you're going to report how much an individual or group gave to a committee, make sure you check for refunds. Refunds generally should be in Schedule B, but sometimes they show up as negative amounts in Schedule A. Keep in mind that weeks or months might pass between the contribution and the refund, so the two transactions might not be on the same report.

A good rule of thumb is to do name searches on Schedules A and B to look for all contributions and refunds. You also can check TEXT records to see if there are references to refunds (though these should be in addition to, rather than in lieu of, Schedule A or B records showing the refund). Also pay attention to the aggregate of the most recent contribution to make sure everything adds up.

11) Putting FEC data into a database
Now that you finally know how to find and download campaign finance reports from the FEC and inspect the guts using an Excel kludge, you might be thinking, "This would be so much better and easier if I just put the data into Access/SQL Server/mySQL/etc." Well, yes and no.

Unless you cover campaign finance nearly full time, it probably is not worth the effort. For one thing, if you don't take the time to build some kind of system to automate a lot of this data maintenance, you'll need to manually check the FEC website for new and amended filings. It's not feasible to do this for more than a handful of committees, and a manual process is error-prone and time-consuming at best.

To give you some perspective, the system I've pieced together during the last three years contains more than 600,000 files that go back to the 2000 election cycle. The database itself is more than 100 GB in size and hosted on a fairly nice SQL Server box. To keep it all up and running, I have an army of Python scripts to download, parse and load the data. (If you're curious about this, all of my code is available on github at https://github.com/cschnaars/FEC-Scraper-Toolbox; I recently added code to help you create the necessary database objects as well.)

If access to hardware is the only issue, you could scale back the volume of data by just building a system in support of the 2016 Election. You also could delete amended data and might even limit your database to specific committees, schedules and/or fields. If you do all these steps, you might even be able to get by with SQLite or Microsoft Access. An older version of my FEC database (https://github.com/cschnaars/FEC-Scraper) downloads reports only for the committees you specify.

But automation is the key here, and that means working knowledge of a scripting language such as Python or Ruby. You need a system in place to help you maintain and load all the data, and that system has to be good enough that, on the eve of publishing a story, you can say: "I'm confident my data is accurate, complete and up-to-date." If you can't do that, your system has a fatal flaw.

12) In conclusion
My goal with this tutorial has been to shed some like upon the murky nature of campaign finance data and help journalists competently and confidently work with it. The toll the economy has taken on newsrooms throughout the country for much of the past decade has, I believe, resulted in fewer reporters to keep an eye on money in politics. Given the vehicles that allow donors to pump unlimited millions of dollars into an election – largely through 527s, Super PACs and 501(c) nonprofits – helping to make this data more accessible to other reporters is critical. We still compete for stories, but I think collaborating to help everyone obtain clean, workable data is in the best interest of readers everywhere.

When I first took an interest in campaign finance, some very smart reporters at various news organizations helped me get started and have been an ongoing resource over the years. Without their help, I never would have gotten to the point where I could write the guide you have now. I'd like to thank some of them now: Derek Willis, *The New York Times*; Bob Biersack, *Center for Responsive Politics*; Aaron Bycoffe, *Huffington* Post; Paul Clark Jr., *Federal Election Commission*; Jacob Fenton, *Sunlight Foundation*; Scott Klein, *ProPublica*; and Fredreka Schouten, *USA TODAY*.