

How the public data platform can disrupt the research community



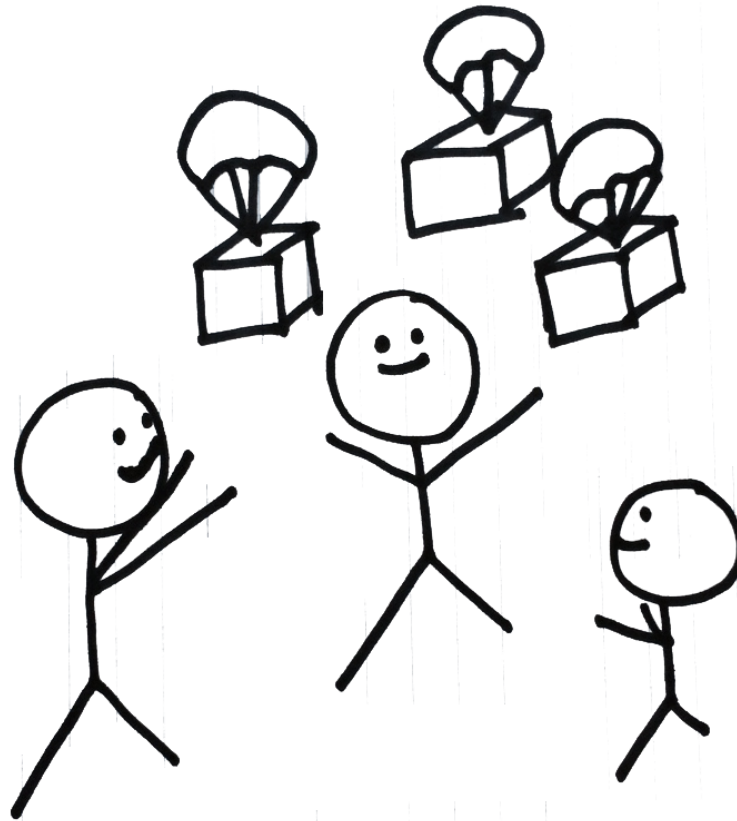
Consumer Financial
Protection Bureau

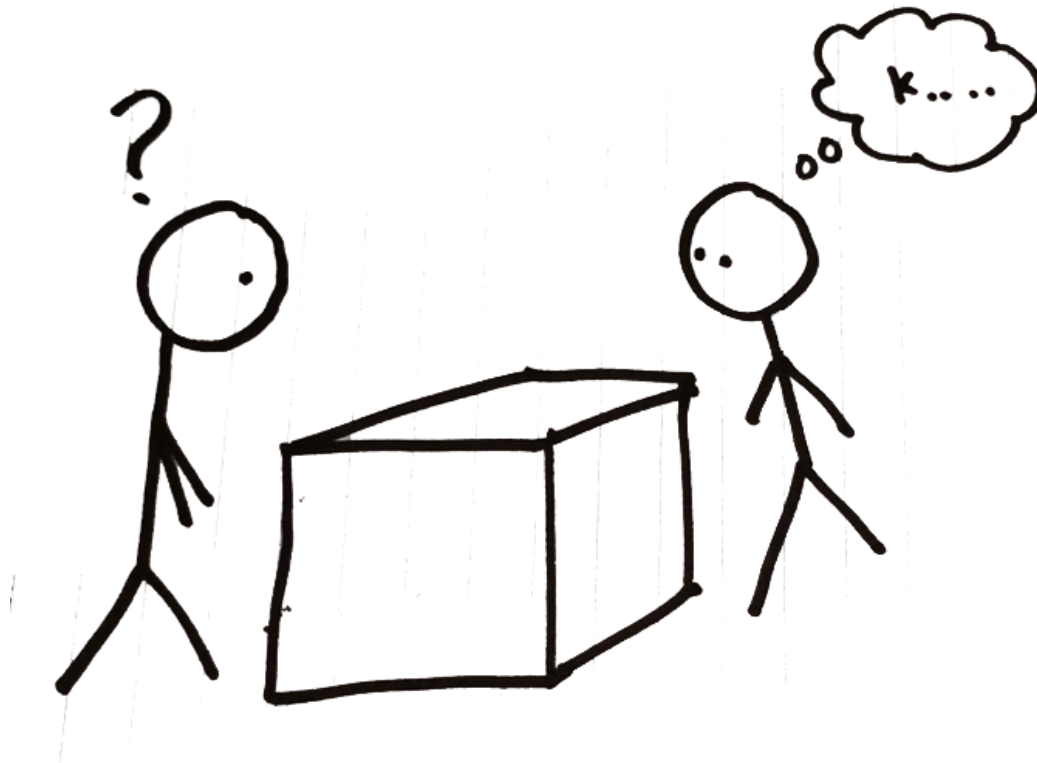
"This is why I can't be angry"



Consumer Financial
Protection Bureau

Increasing access to data





A typical workflow now

1. Make the dataset

1. ~~Make the dataset~~
2. Go to the organization or university's website

HMDA LAR & TS Raw Data Windows Application

This software includes the loan-level data, documentation, and a tool to query and export the data. This software is compatible with Windows only.

Download the software:

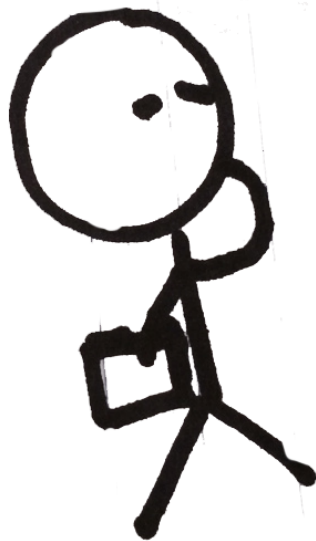
- ▶ [2012 \(621 MB\)](#)
- ▶ [2011 \(490 MB\)](#)
- ▶ [2010 \(535 MB\)](#)
- ▶ [2009 \(625 MB\)](#)
- ▶ [2008 \(595 MB\)](#)
- ▶ [2007 \(860 MB\)](#)

*“*I* don’t even use the FFIEC website to get HMDA data because it’s not formatted correctly.”*



Consumer Financial
Protection Bureau

1. ~~Make the dataset~~
2. ~~Go to the organization or university's website~~
3. **Hope your boss gives it to you**



Consumer Financial
Protection Bureau

1. ~~Make the dataset~~
2. ~~Go to the organization or university's website~~
3. ~~Hope your boss gives it to you~~
4. **Hope the researcher lets you have it**

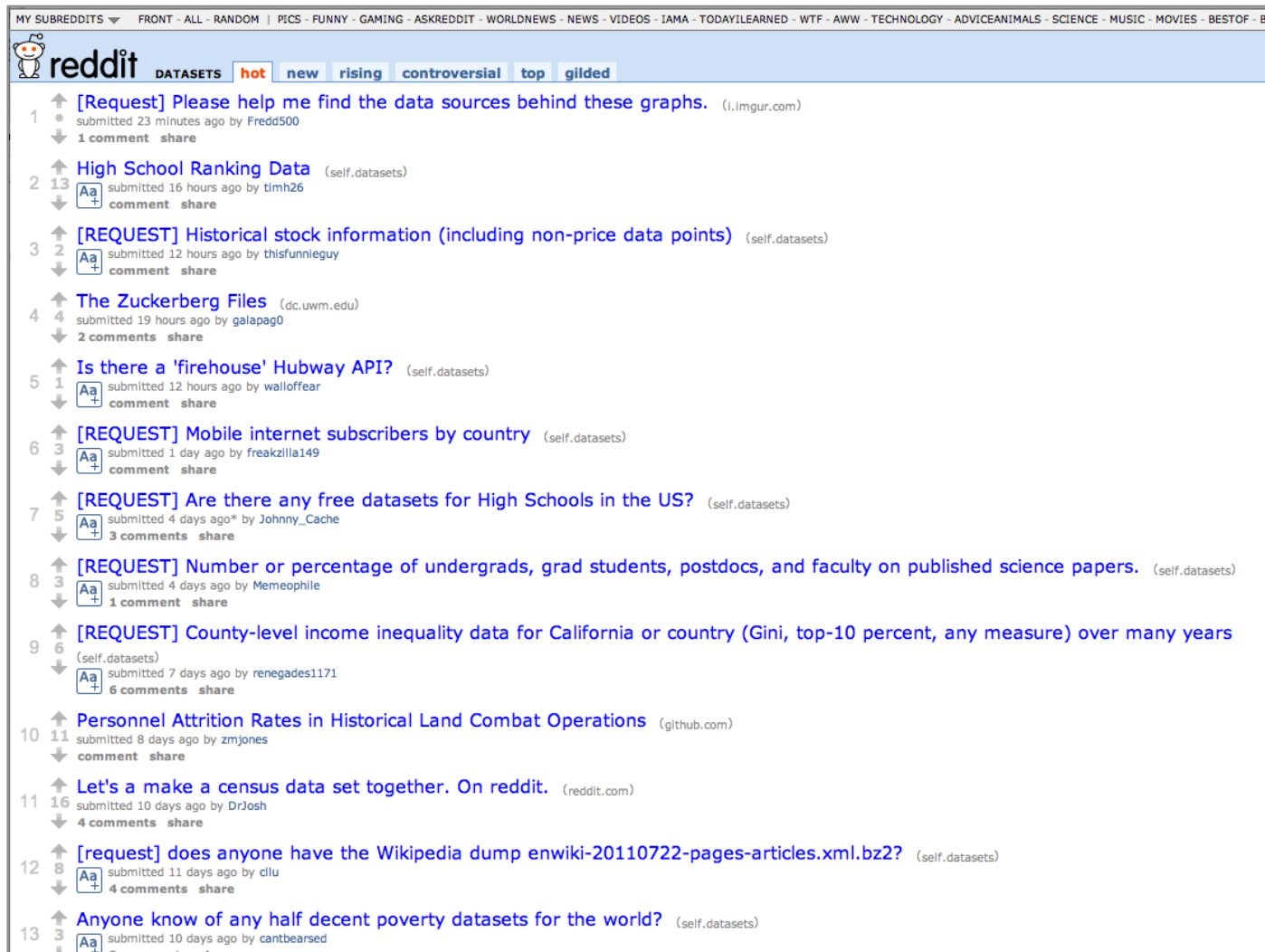
“I contact the original principal investigator (or their secretary), explain what I want it for...and if they think my reason for wanting it is good enough, they email it to me, probably as a .zip file.”



Consumer Financial
Protection Bureau

1. ~~Make the dataset~~
2. ~~Go to the organization or university's website~~
3. ~~Hope your boss gives it to you~~
4. ~~Hope the researcher lets you have it~~
5. **Ask around for it on listservs, mailing lists**

6. Reddit (seriously)



The screenshot shows the Reddit interface for the 'DATASETS' subreddit. The top navigation bar includes links for 'MY SUBREDDITS', 'FRONT', 'ALL', 'RANDOM', 'PICS', 'FUNNY', 'GAMING', 'ASKREDDIT', 'WORLDNEWS', 'NEWS', 'VIDEOS', 'IAMA', 'TODAYILEARNED', 'WTF', 'AWW', 'TECHNOLOGY', 'ADVICEANIMALS', 'SCIENCE', 'MUSIC', 'MOVIES', 'BESTOF', and 'B'. Below the navigation bar, the subreddit name 'reddit' is displayed, followed by tabs for 'DATASETS', 'hot', 'new', 'rising', 'controversial', 'top', and 'gilded'. The main content area lists 13 posts, each with a title, submission details, and interaction options (comment, share). The posts are as follows:

1. [Request] Please help me find the data sources behind these graphs. (i.imgur.com) submitted 23 minutes ago by Fredd500. 1 comment, share.
2. High School Ranking Data (self.datasets) submitted 16 hours ago by timh26. comment, share.
3. [REQUEST] Historical stock information (including non-price data points) (self.datasets) submitted 12 hours ago by thisfunniguy. comment, share.
4. The Zuckerberg Files (dc.uwm.edu) submitted 19 hours ago by galapag0. 2 comments, share.
5. Is there a 'firehouse' Hubway API? (self.datasets) submitted 12 hours ago by walloffear. comment, share.
6. [REQUEST] Mobile internet subscribers by country (self.datasets) submitted 1 day ago by freakzilla149. comment, share.
7. [REQUEST] Are there any free datasets for High Schools in the US? (self.datasets) submitted 4 days ago by Johnny_Cache. 3 comments, share.
8. [REQUEST] Number or percentage of undergrads, grad students, postdocs, and faculty on published science papers. (self.datasets) submitted 4 days ago by Memeophile. 1 comment, share.
9. [REQUEST] County-level income inequality data for California or country (Gini, top-10 percent, any measure) over many years (self.datasets) submitted 7 days ago by renegades1171. 6 comments, share.
10. Personnel Attrition Rates in Historical Land Combat Operations (github.com) submitted 8 days ago by zmjones. comment, share.
11. Let's a make a census data set together. On reddit. (reddit.com) submitted 10 days ago by DrJosh. 4 comments, share.
12. [request] does anyone have the Wikipedia dump enwiki-20110722-pages-articles.xml.bz2? (self.datasets) submitted 11 days ago by clu. 4 comments, share.
13. Anyone know of any half decent poverty datasets for the world? (self.datasets) submitted 10 days ago by cantbearsed.

Alright! I got it! Now what do I do?

Is the data complete?

Is the data correct?

Is the data in a format I can use?

Alright! I got it! Now what do I do?

2011HMDAInstitutionRecords.txt														
2011	00-3324993	1	11-3324993	STERLING NATL MTGE CO., INC	98 CUTTERMILL RD	GREAT NECK	NY	11021-3006	STERLING NATIONAL BANK 650 5TH AVE	NEW YORK				
2011	NY 10019-6108	1	STERLING NATL MTGE CO., INC	GREAT NECK	NY 964388	1	787	N						
2011	0000000047	1	35-0704860	FIRST FINANCIAL BANK NA	1401 S 3RD ST	TERRE HAUTE	IN	47802	FIRST FINANCIAL CORPORATION	ONE FIRST FINANCIAL PLAZA				
2011	TERRE HAUTE IN	47807	FIRST FNCL BK NA	TERRE HAUTE	IN	2363730	0	3	2950	N				
2011	0000000056	1	31-0308605	FIRST FINANCIAL BANK, N.A.	300 HIGH STREET	ATTN: MICKEY MULVANEY			HAMILTON	OH	45011			
2011	FIRST FNCL BK NA	1	HAMILTON	OH	6237352	0	3	2470	N					
2011	0000000086	1	31-0294798	FIRST NATIONAL BANK OF GERMANTOWN	17 NORTH MAIN STREET	GERMANTOWN	OH	45327-1349					FIRST NB	
2011	GERMANTOWN OH	1	68341	0	3	17	N							
2011	0000000182	1	48-1175601	EXCHANGE NATIONAL BANK _TRUST	600 COMMERCIAL	ATCHISON	KS	66002					EXCHANGE NATIONAL BANK	
2011	& TRUST ATCHISON	KS	181149	0	4	80	N							
2011	0000000249	1	25-1255405	FIRST NATIONAL BANK OF PA	166 MAIN ST	GREENVILLE	PA	16125					FIRST NB OF PA	
2011	GREENVILLE PA	8759397	0	1	9814	N								
2011	0000000324	1	23-0916095	FIRST NATIONAL BANK AND TRUST	40 SOUTH STATE ST	NEWTOWN	PA	18940					FIRST NATIONAL BANK AND	
2011	TRUST NEWTOWN PA	730382	0	1	194	N								
2011	0000000325	1	24-0558097	FNB BANK, N.A.	354 MILL STREET	DANVILLE	PA	17821	FULTON FINANCIAL CORPORATION	ONE PENN SQUARE	LANCASTER	PA	17604	
2011	FNB BK NA DANVILLE	PA	388866	0	1	518	N							
2011	0000000347	1	36-1078900	THE 1ST NATIONAL BANK OF LACON	111 S BROAD ST	PO BOX 308	LACON	IL	61540				FIRST NB LACON	
2011	LACON IL	56794	0	3	25	N								
2011	0000000497	1	06-0343370	THE FIRST NATIONAL BANK OF SUF	30 BRIDGE STREET, P.O. BOX 96	SUFFIELD	CT	06078					FIRST NB	
2011	SUFFIELD CT	1	299890	0	1	252	N							
2011	0000000644	1	24-0614415	THE HONESDALE NATIONAL BANK	P O BOX 350	HONESDALE	PA	18431-0350					HONESDALE NB	
2011	HONESDALE PA	521374	1	839	0	1	N							
2011	0000000661	1	23-0534545	DNB FIRST, NA	4 BRANDYWINE AVE.	DOWNINGTOWN	PA	19335					DNB FIRST NA	DOWNINGTOWN
2011	PA	602659	0	1	136	N								
2011	0000000676	1	04-2832782	NATIONAL GRAND BANK	91 PLEASANT STREET	MARBLEHEAD	MA	01945					NATL GRAND BK OF	
2011	MARBLEHEAD MARBLEHEAD	MA	248617	0	1	131	N							
2011	0000000866	1	04-1620797	THE MILFORD NATIONAL BANK	300 EAST MAIN ST	MILFORD	MA	01757	MNB BANCORP.	300 EAST MAIN ST	MILFORD	MA	01757	
2011	MILFORD NB&TC MILFORD MA	276105	0	1	96	N								
2011	0000000929	1	14-0913619	NATIONAL UNION BANK OF KINDERHOOK	ONE HUDSON STREET	KINDERHOOK	NY	12106	NATIONAL UNION BANK OF KINDERHOOK	1 HUDSON ST				
2011	KINDERHOOK NY	12106	1	248057	0	1	81	N						
2011	0000000976	1	14-0984850	PUTNAM COUNTY NATIONAL BANK	P.O. BOX 10	CARMEL	NY	10512-0010					PUTNAM COUNTY NB OF	
2011	CARMEL NY	164609	0	14	N									
2011	0000001027	1	15-0372430	THE LYONS NATIONAL BANK	35 WILLIAM STREET	LYONS	NY	14489					LYONS NB	LYONS
2011	NY	513234	0	1	805	N								
2011	0000001083	1	15-0328960	FIRST NATIONAL BANK OF GROTON	161 MAIN STREET	GROTON	NY	13073					FIRST NB OF GROTON	GROTON
2011	NY	127731	0	1	190	N								
2011	0000001090	1	15-0405885	ALLIANCE BANK NA	120 MADISON ST	SYRACUSE	NY	13202					ALLIANCE BK NA	SYRACUSE
2011	NY	1449065	0	1	833	N								
2011	0000001154	1	36-1588745	FIRST NATIONAL BANK OF OTTAWA	701 LASALLE STREET	OTTAWA	IL	61350	FIRST OTTAWA BANCSHARES	701 LASALLE STREET	OTTAWA			
2011	IL	61350	FIRST NB OF OTTAWA	OTTAWA	IL	285690	0	3	210	N				
2011	0000001194	1	04-1782774	ROCKPORT NATIONAL BANK	P O BOX 90	16 MAIN STREET	ROCKPORT	MA	01966				ROCKPORT NB	
2011	ROCKPORT MA	166351	0	1	92	N								
2011	0000001195	1	03-0143997	NATIONAL BANK OF MIDDLEBURY	30 MAIN STREET, P O BOX 189	MIDDLEBURY	VT	05753-0189						
2011	NATIONAL BK OF MIDDLEBURY	MIDDLEBURY	VT	273219	0	1	458	N						
2011	0000001253	1	14-0486153	BALLSTON SPA NATIONAL BANK	87 FRONT ST.	BALLSTON	SPA	NY	12020	BALLSTON SPA NATIONAL BANK	87 FRONT ST.	BALLSTON	SPA	
2011	NY	12020	BALLSTON SPA NB	BALLSTON SPA	NY	385348	0	1	317	N				
2011	0000001354	1	15-0395735	NBT BANK NA	52 SOUTH BROAD STREET	NORWICH	NY	13815	NBT BANCORP.	INC52 SOUTH BROAD STREET	NORWICH	NY	13815	NBT BK NA
2011	FRIENDSVILLE PA	5302543	0	1	4946	N								
2011	0000001368	1	03-0288082	COMMUNITY NATIONAL BANK	4811 US ROUTE 5, P O BOX 259	DERBY	VT	05829					COMMUNITY NB	
2011	DERBY VT	545104	0	1	723	N								
2011	0000001398	1	14-0912110	NATIONAL BANK OF COXSACKIE	3-7 REED STREET	COXSACKIE	NY	12051					NATIONAL BK OF	
2011	COXSACKIE COXSACKIE	NY	193512	0	1	147	N							
2011	0000001467	1	43-0263025	LANDMARK BANK N.A.	801 EAST BROADWAY	COLUMBIA	MO	65201					LANDMARK BK NA	COLUMBIA
2011	MO	1488307	0	4	1748	N								
2011	0000001493	1	61-0290000	FIRST SOUTHERN NATIONAL BANK	99 LANCASTER STREET	STANFORD	KY	40484					FIRST SOUTHERN	
2011	NB LANCASTER KY	655223	0	3	636	N								
2011	0000001750	1	85-0067165	FIRST NATIONAL BANK SANTA FE	P.O. BOX 609	SANTA FE	NM	87504-0609					FIRST NB OF	
2011	SANTA FE SANTA FE	NM	707367	0	4	268	N							
2011	0000001770	1	43-0189015	BOONE COUNTY NATIONAL BANK	720 E BROADWAY	COLUMBIA	MO	65201	CENTRAL BANCAMPANY	238 MADISON ST	JEFFERSON CITY	MO		
2011	65101 BOONE CTY NB OF COLUMBIA	COLUMBIA	MO	1233182	0	4	1465	N						
2011	0000001794	1	41-0519725	FIRST NATIONAL BANK MINNESOTA	500 LONG STREET	MANKATO	MN	56001					FIRST NB MN	SAINT PETER
2011	MN	186057	0	3	44	N								
2011	0000001950	1	71-0058845	FIRST NATIONAL BANK	P.O. BOX 7	FORT SMITH	AR	72902	FIRST BANK CORPP.	P.O. BOX 7	FORT SMITH	AR	72902	FIRST NB
2011	OF FORT SMITH FORT SMITH	AR	1087148	0	2	159	N							
2011	0000001997	1	31-0244820	NATIONAL BANK & TRUST COMPANY	PO BOX 711	WILMINGTON	OH	45177	NBT FINANCIAL GROUP, INC	PO BOX 711	WILMINGTON			
2011	OH	45177	NATIONAL B&TC	WILMINGTON	OH	689900	0	3	432	N				
2011	0000002036	1	31-4332540	VINTON COUNTY NATIONAL BANK	112 W MAIN ST	PO BOX 460	MCARTHUR	OH	45651				VINTON	
2011	COUNTY NATIONAL BANK	MC ARTHUR	OH	700538	0	3	553	N						
2011	0000002068	1	46-0208038	FIRST DAKOTA NATIONAL BANK	225 CEDAR ST	YANKTON	SD	57078-0156	FIRST DAKOTA NATIONAL BANK	225 CEDAR ST	YANKTON	SD		
2011	57078 FIRST DAKOTA NB	YANKTON SD	787788	0	4	705	N							



Consumer Financial
Protection Bureau

Alright! I got it! Now what do I do?

The screenshot displays the SAS software interface. The main window shows a data table with columns for various bank records. The 'Log - (Untitled)' window is open, showing the execution of a SAS program. The 'Hello World.sas' code editor is also visible, containing the following code:

```
data _null_;  
  do i=1 to 5;  
    put i "Hello World!";  
  end;  
run;
```

The log window shows the output of the program, including the text "NOTE: DATA statement used (Total process time):" and the execution time for the data statement.

The data table contains the following information:

Bank Name	Address	City	State	Zip	Other Info
STERLING NATL MTGE CO., INC	98 CUTTERMILL RD	GREAT NECK	NY	11021-3006	STERLING NATIONAL BANK 650 5TH AVE NEW YORK
STERLING NATL MTGE CO., INC	964388 1	787	N		
FIRST FINANCIAL BANK NA	1401 S 3RD ST	TERRE HAUTE	IN	47802	FIRST FINANCIAL CORPORATION ONE FIRST FINANCIAL PLAZA
FIRST FNCL BK NA	31-030860	HAMILTON			
FIRST FNCL BK NA	31-029479	HAMILTON			
GERMANTOWN OH	68341 0				
GERMANTOWN OH	48-117560				
ATCHISON KS	101149 0				
ATCHISON KS	25-125540				
GREENVILLE PA	8759397 0				
GREENVILLE PA	23-091609				
TRUST NEWTOWN PA	730382 0				
TRUST NEWTOWN PA	24-055809				
FNB BK NA DANVILLE	PA 38				
FNB BK NA DANVILLE	36-107890				
LACON IL 56794	0 3 25				
LACON IL 56794	06-034337				
SUFFIELD CT	209090 0				
SUFFIELD CT	24-061441				
HONESDALE PA	521374				
HONESDALE PA	23-053454				
PA 602659	0 1 136 N				
PA 602659	04-283278				
MARBLEHEAD MA	24				
MARBLEHEAD MA	04-162079				
MILFORD NB&TC MILFORD MA	276105 0				
MILFORD NB&TC MILFORD MA	14-091361				
KINDERHOOK NY	12106				
KINDERHOOK NY	14-098485				
CARMEL NY	164609 0				
CARMEL NY	15-037243				
NY 513234	0 1 805 N				
NY 513234	15-032896				
NY 127731	0 1 190 N				
NY 127731	15-040588				
NY 1449065	0 1 833 N				
NY 1449065	36-158874				
IL 61350	FIRST NB OF OTTAWA 07				
IL 61350	04-178277				
ROCKPORT MA	166351 0				
ROCKPORT MA	03-014399				
NATIONAL BK OF MIDDLEBURY	MIDDLEBURY				
NATIONAL BK OF MIDDLEBURY	14-048615				
NY 12020	BALLSTON SPA NB				
NY 12020	15-039573				
FRIENDSVILLE PA	5302543 0				
FRIENDSVILLE PA	03-028808				
DERBY VT	545104 0				
DERBY VT	14-091211				
COXSACKIE COXSACKIE	193512 0				
COXSACKIE COXSACKIE	43-0263025				
MO 1488307	0 4 1748 N				
MO 1488307	61-0290000				
NB LANCASTER KY	655223 0				
NB LANCASTER KY	85-0067165				
SANTA FE SANTA FE	NM 707367				
SANTA FE SANTA FE	43-0189015				
BOONE CTY NB OF COLUMBIA	COLUMBIA				
BOONE CTY NB OF COLUMBIA	41-0519725				
MINN 186057	0 3 44 N				
MINN 186057	71-0058845				
OF FORT SMITH FORT SMITH	AR 1087148 0				
OF FORT SMITH FORT SMITH	31-0244820				
OH 45177	NATIONAL B&TC WILMINGTON				
OH 45177	689900 0				
OH 45177	31-4332540				
COUNTY NATIONAL BANK MC ARTHUR	OH 700538 0				
COUNTY NATIONAL BANK MC ARTHUR	46-0208038				
FIRST DAKOTA NATIONAL BANK	225 CEDAR ST				



Consumer Financial
Protection Bureau

Alright! I got it! Now what do I do?

2011 00-3324993 1 11-3324993 STERLING NATL MTGE CO., INC 98 CUTTERMILL RD GREAT NECK NY 11021-3006 STERLING NATIONAL BANK 650 5TH AVE NEW YORK

2011 NY 10019-6108 STERLING NATL MTGE CO., INC GREAT NECK NY 964388 1 787 1401 S 3RD ST TERRE HAUTE TN 47802 FIRST FINANCIAL CORPORATION ONE FIRST FINANCIAL PLAZA

2011 000000047 1 35-0704860 FIRST FINANCIAL BANK NA

2011 SAS 1

2011 File Edit View Tools Run Solutions Window Help

2011 Explorer

2011 Contents of 'SAS Environment'

2011 Libraries File Shortcuts

2011 Favorite Folders My Computer

2011 Log - (Untitled)

```
9 data _null_;
10 do i=1 to 5;
11 put i "Hello World!";
12 end;
13 run;
```

1 Hello World!
2 Hello World!
3 Hello World!
4 Hello World!
5 Hello World!

NOTE: DATA statement used (Total process time):
real time 0.12 seconds
cpu time 0.01 seconds

2011 Hello World.sas

```
data _null_;
do i=1 to 5;
put i "Hello World!";
end;
run;
```

2011 SAS/IML Studio

2011 File Edit View Program Graph Analysis Tools Window Help

2011 SAS/IML Studio

```
declare DataObject dobj;
dobj = DataObject.CreateFromFile( "usnews.sas7bdat" );

/* Change observation colors: BLACK & RED */
run ColorCode( dobj, "PubPriv", (0 0 0, 255 0 0), 2 );

/* Create a scatter plot and change attributes */
declare ScatterPlot plot;
plot = ScatterPlot.Create( dobj, "Tuition", "DollarPerStud" );
plot.ShowAxisReferenceLines( YAXIS );
plot.SetAxisLabelVar( "College" );
/* add ymx line */
plot.DrawUseDataCoordinates();
plot.DrawLine( 0, 0, 25000, 25000 );

/* Create another plot that is linked to the first */
declare Histogram hist;
hist = Histogram.Create( dobj, "SAT" );
hist.SetWindowPosition(0, 50, 100, 50);
```

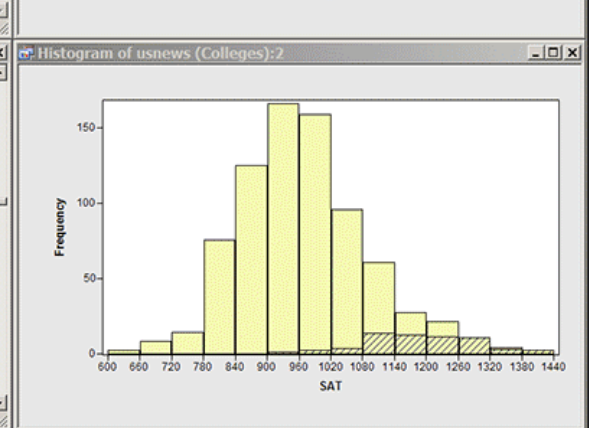
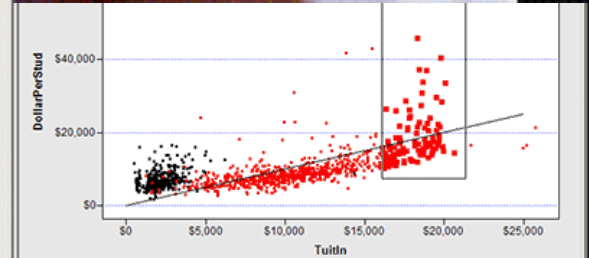
2011 usnews (Colleges):3

College	State	PubPriv	SATM
Harding University	AR	2	5
Grand Canyon University	AZ	2	4
Prescott College	AZ	2	4
Azusa Pacific University	CA	2	4
Biola University	CA	2	5
California Baptist College	CA	2	4
California Institute of Technology	CA	2	7
California Lutheran University	CA	2	4
United States International University	CA	2	4
Chapman University	CA	2	5
Claremont McKenna College	CA	2	6
Harvey Mudd College	CA	2	7
Pitzer College	CA	2	7
Pomona College	CA	2	7
Scotch College	CA	2	4
College of Notre Dame	CA	2	4
Holy Names College	CA	2	4
Dominican College of San Rafael	CA	2	4
La Sierra University	CA	2	4
University of La Verne	CA	2	4
Menlo College	CA	2	4

2011 Histogram of usnews (Colleges):2

2011 Frequency

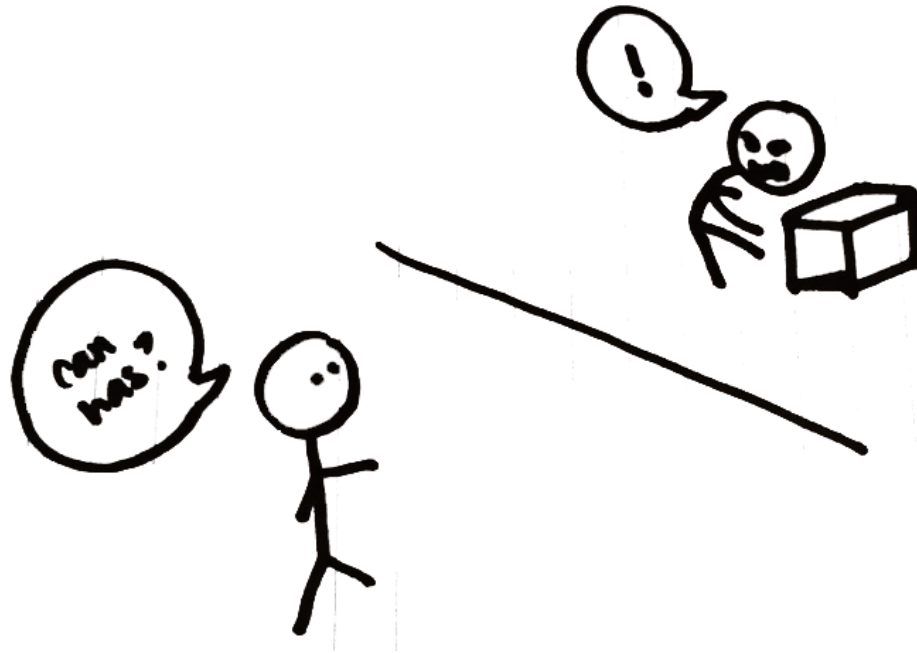
2011 SAT



Consumer Financial
Protection Bureau

This process is a pain in the ass.

People are territorial



Software is expensive and superfluous

SAS® Analytics Pro

SAS Analytics Pro provides all of the functionality needed to access, manipulate, analyze and present information. For more information, click [here](#).

How many user licenses? **Update**

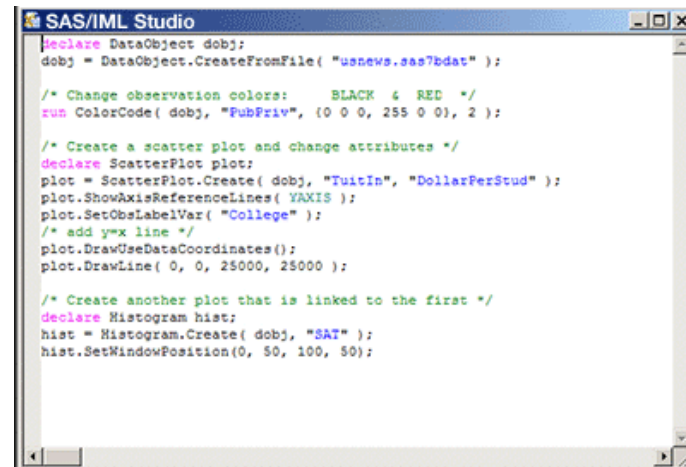
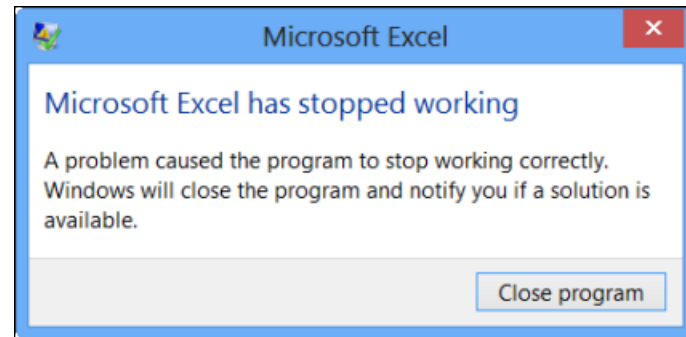
\$8,700

First-year license fee** for Desktop Version of Windows

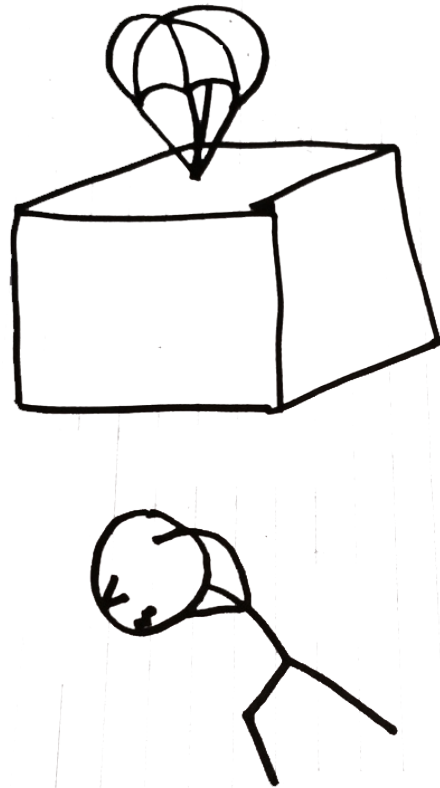
- * Includes Technical Support
- * Includes Documentation
- * Includes Software Upgrades

Continue You can add other products to your order in the next step

****ANNUAL LICENSE:** SAS software licenses are renewable on an annual basis. Annual renewal fees are approximately 28 percent of the then-current first-year license fee. The fee shown does not include applicable taxes.



Data is getting bigger



Sounds like a design problem to me!

A better workflow

MARIO
021650

● × 13

WORLD
1-2

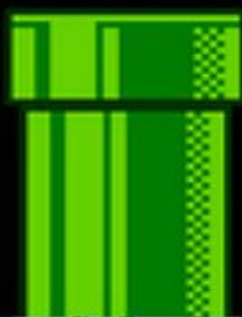
TIME
327

WELCOME TO WARP ZONE!

4

3

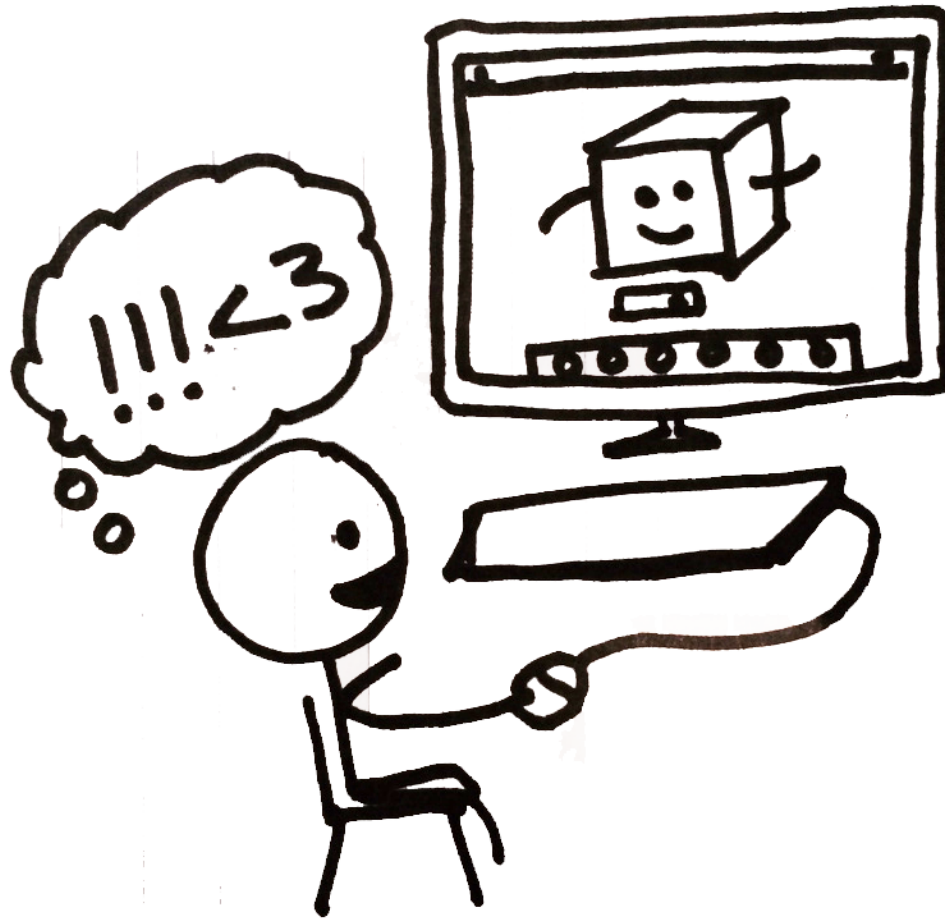
2



SUPER MARIO BROS - NINTENDO - 1985

**We can eliminate the majority of the cruff
some researchers go through *just* to start
doing their work.**

1. ~~Make the dataset~~
2. ~~Go to the organization or university's website~~
3. ~~Hope your boss gives it to you~~
4. ~~Hope the researcher lets you have it~~
5. ~~Ask around for it on listservs, mailing lists~~
6. ~~Reddit~~



Filter the data

Filter the data to just what you're interested in, and then download the results or aggregate them into a summary table.



Most people search for **loan originations** for **first-lien, owner-occupied, 1-4 family** and **manufactured housing** in **2012**. To save you time, we have pre-filtered the data to these results.

[Download the data now](#) or change the filters below. [\[Clear all fields\]](#)

Select year(s) of data:

2012 x



LOCATION

State, metro area, county, and census tract of the property



APPLICANT

Demographic information for applicants and co-applicants



LENDER

Lender ID and federal agency



PROPERTY

Property type and occupancy



LOAN DESCRIPTION

Loan purpose, type, amount, and more



LOAN ACTION

Status of the application, secondary purchaser, and more

[Clear all fields](#) | [Reset to default](#)

Want more insight? [Create a summary table >](#)



Consumer Financial
Protection Bureau



There are **8,401,970** HMDA records from **2012** with the previously selected filters.

Not quite what you're looking for? You can also go back and filter your results differently. [Modify your filters >](#)

Save & share your work

Save your filters, or share them with a link:

<https://github.cfpb.gov/pages/data-platform/p>

Share



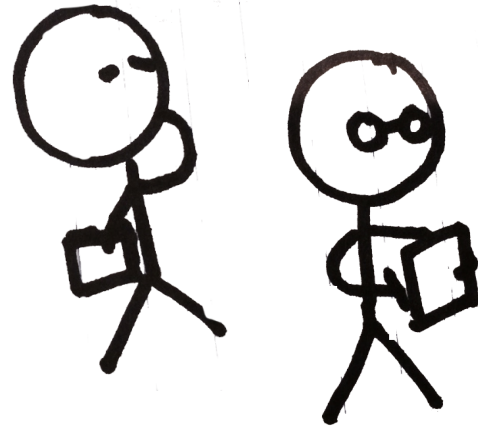
Consumer Financial
Protection Bureau



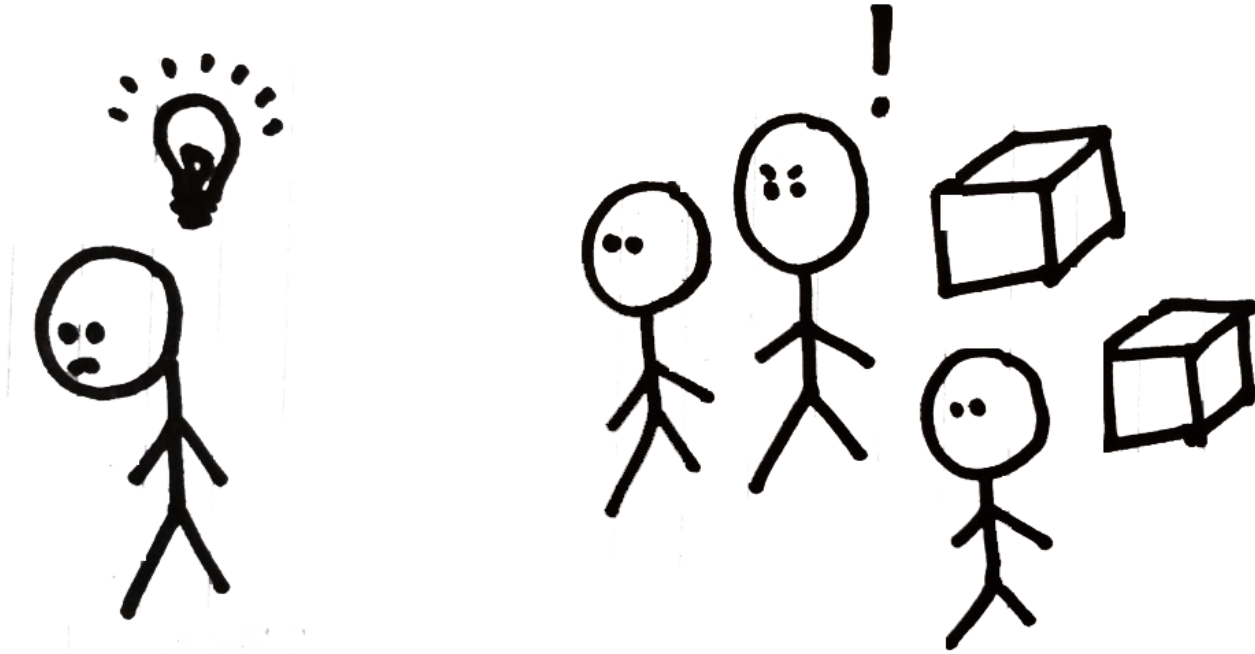
Time



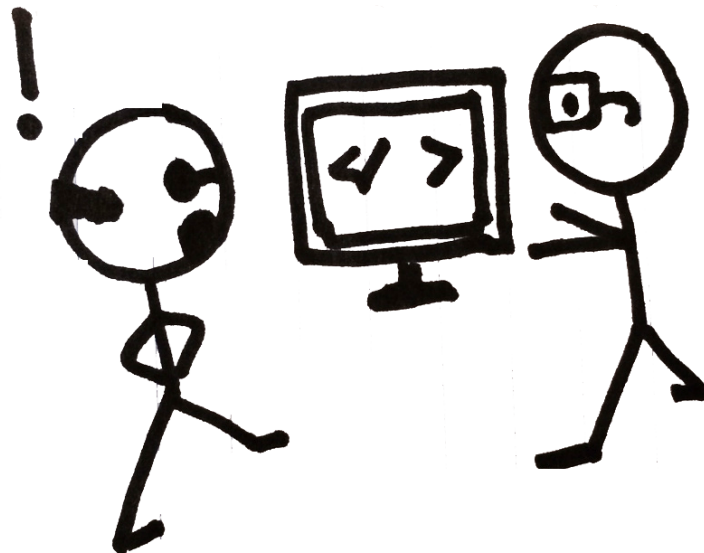
Money



People



Empathy and user experience



Good design solves problems.