

# Introduction to Data and Data Modeling

---

HIST 2430 Digital Histories of Ethnic Boston  
Simon Rabinovitch  
Fall 2020



Northeastern University  
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point  
during the presentation!*

# What is data?

What kinds of data do you use every day?

How do your different daily activities rely on data? (For example, checking the weather, reading email.)

How much do you know about the structure and contents of the data you use?



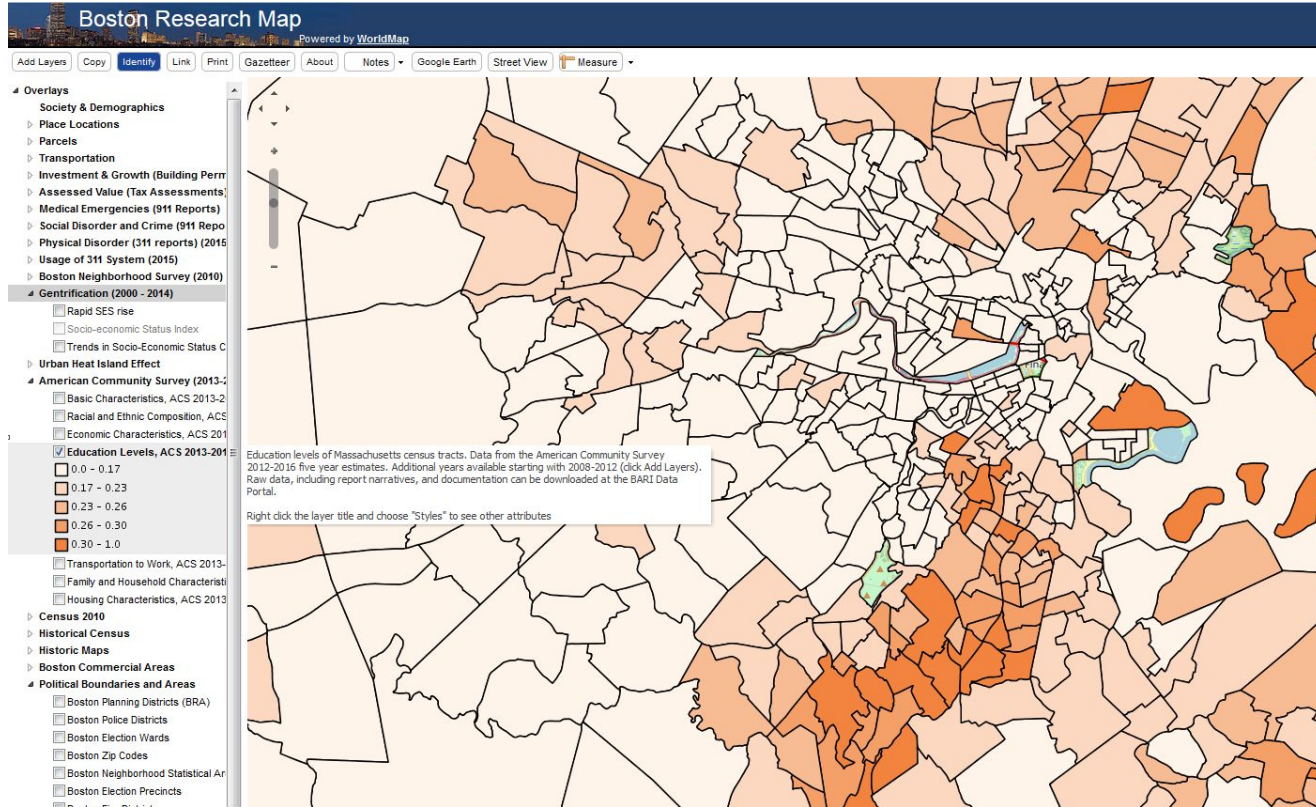
# Example: weather forecast

What data categories are present here?

What data would be needed to produce this app?



# At the end of data building comes...



# And what happens behind the image is...

adPer	SameSexCoupPer	GrandHeadPer	LessThanHS	HSGrad	SomeColl	Bach	Master	Prof	Doc	CommuteLess10	Commute1030
40046	9.473334e-01	0.000000e+00	7.319685e-02	0.172511	0.195735	0.336727	0.144816	3.564450e-02	4.137282e-02	0.448130	0.392466
12861	2.222250e-01	0.000000e+00	1.665624e-02	0.270387	0.186635	0.307122	0.183441	1.837072e-02	1.738933e-02	0.299253	0.393155
12864	2.222218e-01	0.000000e+00	1.665440e-02	0.270390	0.186628	0.307128	0.183443	1.836887e-02	1.738918e-02	0.299248	0.393155
32031	3.582095e-01	3.848857e-08	4.129928e-02	0.229930	0.289327	0.234570	0.141995	4.385138e-02	1.902550e-02	0.376687	0.404089
32031	3.582084e-01	0.000000e+00	4.129937e-02	0.229931	0.289328	0.234571	0.141996	4.385155e-02	1.902555e-02	0.376687	0.404090
23550	3.493735e-01	2.204291e-02	7.131495e-02	0.206305	0.293923	0.274097	0.121386	1.397819e-02	1.899565e-02	0.313884	0.370100
43237	6.229508e-01	9.871668e-03	2.652705e-02	0.153927	0.193717	0.405236	0.137871	3.246073e-02	5.026178e-02	0.249493	0.395538
36394	1.473668e-01	7.717905e-03	7.454486e-02	0.180729	0.280000	0.311998	0.113818	2.290944e-02	1.599978e-02	0.516147	0.353876
20039	2.679214e-04	0.000000e+00	2.158664e-02	0.298481	0.227921	0.254239	0.156811	1.975137e-02	2.121089e-02	0.356055	0.299790
35472	4.710172e-01	0.000000e+00	3.695200e-02	0.221877	0.219139	0.309641	0.151897	3.790324e-02	2.258889e-02	0.208973	0.577843
11216	1.815340e-01	2.288312e-05	3.863060e-02	0.231285	0.318652	0.230817	0.143632	2.545305e-02	1.153195e-02	0.198005	0.614922
31040	2.097580e-01	8.450970e-06	5.830264e-02	0.280979	0.297716	0.246405	0.092305	1.646212e-02	7.834582e-03	0.212224	0.528976
34333	1.058288e-03	1.440196e-02	4.150749e-02	0.258886	0.302047	0.239643	0.111758	3.959331e-02	6.565285e-03	0.201304	0.586980
39785	1.162788e-01	0.000000e+00	2.858148e-02	0.182075	0.292519	0.315102	0.135145	2.328861e-02	2.328861e-02	0.277368	0.522191



# Workshop Agenda

- What is data?
- Human vs. machine readable
- Delimiter separated files
- Metadata and data cleaning
- Introduction to our data-preparation activity

Slides, handouts, and data available at <https://bit.ly/diti-fall2020-rabinovitch2>



# Workshop Objectives

- Get a sense of how to effectively create and organize data
- Discuss how to plan data collection and organization for open use
- Learn the difference between Excel and CSV files
- Learn how CSV files can be used for the greatest effect
- Basic vocabulary - guidelines for further individual exploration of CSV and data organization in general
- Basic operations on CSV - toward cleaning data



# Key Terms

- Data - organized information
- Machine readable - set of properties that make possible for a machine to operate on a certain file
- Delimiter - special character, chosen to function as a border between fields in which data points are contained
- CSV - comma separated values - a format for presenting data that uses commas as delimiters
- Data cleaning - the process of standardizing data input





# Example dataset

example\_dataset ☆ 📄 🌐

File Edit View Insert Format Data Tools Add-ons Help Last edit was 3 minutes ago

100% \$ % .0 .00 123 Arial 10 B I U A

	city/town				
	A	B	C	D	E
1	city/town	Name of congregation	Contact Information	Year founded	
2	ACTON	Congregation Beth Elohim	133 Prospect St. (PO Box2218), 01720; 978-263-3061; cbeacton@tiac.org	1969	
3	ACTON	Chabad Center	148 Great Road, 01720;978-929-2513	1999	
4	AMHERST	Jewish Community of Amherst	742 Main St., 01002; 413-256-0160; info@J-C-A.org	1969	
5	ANDOVER	Congregation Beth Israel	501 S. Main St.(PO Box929), 01810; 978-474-0540;office@bethisraelmv.org	2004	
6	ANDOVER	Chabad Lubavitch of MerrimackValley	310 N. Main St., 01810;978-470-2288;rabbi@chabadmv.com	1992	
7	ANDOVER	Temple Emanuel	7 Haggetts Pond Rd.,01810; 978-470-1356;www.templemanuel.net	1920	
8	ANDOVER	Havurat Shalom	17 Olde Berry Rd.,01810; 978-470-2542		
9	ANDOVER	Tifereth Israel	510 S. Main St., 01810	c. 1990	
10	ARLINGTON	Hebrew Congregation of ArlingtonHeights	122 Sylvia St.	1906	
11	ASHLAND	Shaarei Shalom, Reform JewishCongregation of Ashland	PO Box 454, 01721; 508-231-4700; info@rjca.org	1995	
12	ATHOL	Agudas Achim	411 Pine Street	1907	
13	ATHOL	Temple Israel	107 Walnut St., 01331;978-249-9481;www.templeisraelathol.org	1950	
14	ATTEBORO	Agudas Achim Congregation	901 N. Main St., 02703;508-222-2243;office@agudasma.org	1910/1911	
15	AYER	Congregation Anshey Shalom	Cambridge St. &Norwood Ave.	1948	
16	BELMONT	Beth El Temple Center	2 Concord Av., 02178;617-484-6668;office.betc@verizon.net	1932/1936	
17	BEVERLY	Temple B'nai Abraham	200 East Lothrop St.,01915; 978-927-3211;tba.beverly@verizon.net	1897/ 1908	
18	BEVERLY	Congregation Ohav Shalom	3 Beckford St.	1903/1916	
19	BOSTON	Chabad House	491 CommonwealthAve., 02215; 617-424-1190;rsp@chabadboston.org	1979	
20	BSTON	Temple Israel	477 Longwood Ave,02215; 617-566-3960;info@tisrael.org	1854	
21	BOSTON	Congregation Bnai Jacob/ Zvhil-Mezbuz Beis Medrash	15 School St., 02108; 617-227-8200;Rebbe@rebbe.org;	1919	
22	BOSTON	Congregation Ohavi Sedeck		<1887	
23					



# Machine readability

What is machine readability and how is it different from human readability and digital accessibility?





*Feel free to ask questions at any point during the presentation!*

# Machine readability

- Purpose of machine readability
  - To enable the data to be processed with computer programs
  - Distinction between documents (static, “frozen”) and data (dynamic, operable)



# Example

Is this database machine readable, human readable, digitally accessible, or all of the above?

This example is the pdf document from which we made our dataset



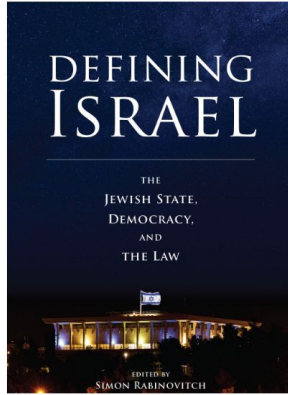
Massachusetts Synagogues and Their Records, Past and Present

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	City/Town	Name of congregation	Contact Information	Year founded	Founded by (if other than local Jews)	Past locations	Any mergers	Cemetery	Yahrzeit	Member	Officer	History	Death	Marriage	Bris	BBM	Pictures	Other Records
1	ACTON	Congregation Beth Elohim	133 Prospect St. (PO Box 2218), 01720; 978-263-3061; cbeacton@tiac.org	1969			1969--merged with Congregation Rodoph Shalom, Maynard	Section in town cemetery; new plots in Beit Olam	x							To mid-70s		
2		Chabad Center	148 Great Road, 01720; 978-929-2513	1999														
3	AMHERST	Jewish Community of Amherst	742 Main St., 01002; 413-256-0160; info@J-C-A.org	1969				122 x		X	X	website	X	X	maybe	X	x	
4	ANDOVER	Congregation Beth Israel	501 S. Main St.(PO Box 929), 01810; 978-474-0540; office@bethisraelmv.org	2004	Merger of Tifereth Israel, Andover, and Beth El, Lowell	Moving to 17 Boutwell Rd		11,68,72,73	X	Bulletins	acctning	exhibit in storage						
5																		





# NOW AVAILABLE: DEFINING ISRAEL: THE JEWISH STATE, DEMOCRACY, AND THE LAW



## Purchase This Book

November 2018

409pp

ISBN-13 9780878201624

ISBN-10 0878201629

Paperback \$35.95

The timeliness and exceptional comprehensiveness of this work cannot be overstated. With passage of Israel's nation-state law, this volume brings together the evolution of that law and a superbly definitive selection of the entire spectrum of views on the deeper underlying issues. It is required reading for anyone with an abiding interest in the issues of Jewish statehood and the direction of Israeli society and politics.

## SIMON RABINOVITCH

*Defining Israel: The Jewish State, Democracy, and the Law* is the first book in any language devoted to the controversial passage of Israel's nation-state law. Israel has no constitution, and though it calls itself the Jewish state there is no agreement among Israelis on how that fact should be reflected in the government's laws or by its courts. Since the 1990s a number of civil society groups and legislators have drafted constitutions and proposed Basic Laws with constitutional standing that would clarify what it means for Israel to be a "Jewish and democratic state."

Are these bills liberal or chauvinist? Are they a defense of the Knesset or an attack on the independence of the courts? Is their intention democratic or anti-democratic? The fight over the nation-state law—whether to have one and what should be in it—toppled the 19<sup>th</sup> Knesset's governing coalition and, even after its passage on July 29, 2018, remains a point of contention among Israel's lawmakers and increasingly the Israeli public. *Defining Israel* brings together influential scholars, journalists, and politicians, observers and participants, opponents and proponents, Jews and Arabs, all debating the merits and meaning of Israel's nation-state law. Together with translations of each draft law, the final law, and other key documents, the essays and sources in *Defining Israel* are essential to understand the ongoing debate over what it means for Israel to be a Jewish and democratic state.

[Listen to Simon Rabinovitch discuss the Nation-State Law with Jewish History Matters](#)



**Northeastern University**  
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point  
during the presentation!*

date,pages,ISBN-13,ISBN-10,binding-price  
November 2018,409pp,9780878201264,878201629,Paperback \$39.95



**Northeastern University**  
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point  
during the presentation!*

# Defining CSV (comma-separated values)

- Text-based, delimited data format
- Compatible with the principles of open data
- Doesn't have standard formatting
- Doesn't require any particular software to read or edit





# Putting the concepts to use

- Delimiter-separated values
  - TSV (tab-separated values), CSV (comma-separated values), PSV (pipe-separated values - “|”)
- Delimiter (aka field separator) - tab, comma, or pipe
- Each row is also sometimes referred to as a “record”
- Special characters - qualifier, escape character



# Who and what is data created for?

- Open and closed data (as a political issue)
- Issue of ownership over data and ownership over insights from it
  - Making data in/operable can reflect the political intentions over who can use it, and for what
  - Commodification of data such as selling data to target audiences: potential for abuse, lack of accountability, and the difference in power between the person on whom data is collected, and the data collector
- Privacy - identifiable information: sometimes stored personal data gets stolen; sometimes making connections between different data points enable identification even in a seemingly anonymized dataset, etc.



# Metadata

- What is metadata? Information on the dataset - who, what, where, when, how, and why
- Developing metadata as early as possible not only helps the users of your data, but helps you to plan for possible issues in creating and building the dataset
- Developing column naming conventions more compatible with machine readability: lower cases and underscores, rather than multiple words and characters that often have specific functions within programs (e.g. comma)
- Readme style of metadata writing, available [here](#)





# Anatomy of a tool

- This is a CSV exported and opened in a notepad
- Characters and commas - that's all there is to it

```
city/town,Name of congregation,Contact Information,Year founded
ACTON,Congregation Beth Elohim,"133 Prospect St. (PO Box2218), 01720; 978-263-3061; cbeacton@tiac.org",1969
ACTON,Chabad Center,"148 Great Road, 01720;978-929-2513",1999
AMHERST,Jewish Community of Amherst,"742 Main St., 01002; 413-256-0160; info@J-C-A.org",1969
ANDOVER,Congregation Beth Israel,"501 S. Main St.(PO Box929), 01810; 978-474-0540;office@bethisraelmv.org",2004
ANDOVER,Chabad Lubavitch of Merrimackvalley,"310 N. Main St., 01810;978-470-2288;rabbi@chabadmv.com",1992
ANDOVER,Temple Emanuel,"7 Haggetts Pond Rd.,01810; 978-470-1356;www.templemanuel.net",1920
ANDOVER,Havurat Shalom,"17 Olde Berry Rd.,01810; 978-470-2542",
ANDOVER,Tifereth Israel,"510 S. Main St., 01810",c. 1990
ARLINGTON,Hebrew Congregation of ArlingtonHeights,122 Sylvia St.,1906
ASHLAND,"Shaarei Shalom, Reform JewishCongregation of Ashland","PO Box 454, 01721; 508-231-4700; info@rjca.org",1995
ATHOL,Agudas Achim,411 Pine Street,1907
ATHOL,Temple Israel,"107 Walnut St., 01331;978-249-9481;www.templeisraelathol.org",1950
ATTLEBORO,Agudas Achim Congregation,"901 N. Main St., 02703;508-222-2243;office@agudasma.org",1910/1911
AYER,Congregation Anshey Shalom,Cambridge St. &Norwood Ave.,1948
BELMONT,Beth El Temple Center,"2 Concord Av., 02178;617-484-6668;office.betc@verizon.net",1932/1936
BEVERLY,Temple B'nai Abraham,"200 East Lothrop St.,01915; 978-927-3211;tba.beverly@verizon.net",1897/ 1908
BEVERLY,Congregation Ohav Shalom,3 Beckford St.,1903/1916
BOSTON,Chabad House,"491 CommonwealthAve., 02215; 617-424-1190;rsp@chabadboston.org",1979
BOSTON,Temple Israel,"477 Longwood Ave,02215; 617-566-3960;info@tisrael.org",1854
BOSTON,Congregation Bnai Jacob/ Zvihil-Mezbuz Beis Medrash,"15 School St., 02108; 617-227-8200;Rebbe@rebbe.org",1919
BOSTON,Congregation Ohavi Sedek,,<1887
```





# Demonstration of Tool - Google Sheets

CSV in Google  
Sheets - meaning of  
delimiters (,) and  
qualifiers ("")

```
city/town,Name of congregation,Contact Information,Year founded
ACTON,Congregation Beth Elohim,"133 Prospect St. (PO Box2218), 01720; 978-263-3061; cbeacton@tiac.org",1969
ACTON,Chabad Center,"148 Great Road, 01720;978-929-2513",1999
AMHERST,Jewish Community of Amherst,"742 Main St., 01002; 413-256-0160; info@J-C-A.org",1969
ANDOVER,Congregation Beth Israel,"501 S. Main St.(PO Box929), 01810; 978-474-0540;office@bethisraelmv.org",2004
ANDOVER,Chabad Lubavitch of MerrimackValley,"310 N. Main St., 01810;978-470-2288;rabbi@chabadmv.com",1992
ANDOVER,Temple Emanuel,"7 Haggetts Pond Rd.,01810; 978-470-1356;www.templemanuel.net",1920
ANDOVER,Havurat Shalom,"17 Olde Berry Rd.,01810; 978-470-2542",
ANDOVER,Tifereth Israel,"510 S. Main St., 01810",c. 1990
ARLINGTON,Hebrew Congregation of ArlingtonHeights,122 Sylvia St.,1906
ASHLAND,"Shaarei Shalom, Reform JewishCongregation of Ashland" PO Box 454, 01721; 508-231-4700; info@rjca.org",1995
ATHOL,Agudas Achim,411 Pine Street,1907
ATHOL,Tempel Israel,"107 Walnut St., 01331;978-249-9481;www.templeisraelathol.org",1950
ATTLEBORO,Agudas Achim Congregation,"301 N. Main St., 02703;508-222-2243;office@agudasma.org",1910/1911
AYER,Congregation Anshey Shalom,Cambridge St. &Norwood Ave.,1948
BELMONT,Beth El Temple Center,"2 Concord Av., 02178;617-484-6668;office.betc@verizon.net",1932/1936
BEVERLY,Tempel B'nai Abraham,"200 East Lothrop St.,01915; 978-927-3211;tba.beverly@verizon.net",1897/ 1908
BEVERLY,Congregation Ohav Shalom,3 Beckford St.,1903/1916
BOSTON,Chabad House,"491 CommonwealthAve., 02215; 617-424-1190;rsp@chabadboston.org",1979
BOSTON,Tempel Israel,"477 Longwood Ave,02215; 617-566-3960;info@tisrael.org",1854
BOSTON,Congregation Bnai Jacob/ Zvhil-Mezbuz Beis Medrash,"15 School St., 02108; 617-227-8200;Rebbe@rebbe.org",1919
BOSTON,Congregation Ohav Sedek,,<1887
```

	city/town	Name of congregation	Contact Information	Year founded
1	ACTON	Congregation Beth Elohim	133 Prospect St. (PO Box2218), 01720; 978-263-3061; cbeacton@tiac.org	1969
2	ACTON	Chabad Center	148 Great Road, 01720;978-929-2513	1999
3	AMHERST	Jewish Community of Amherst	742 Main St., 01002; 413-256-0160; info@J-C-A.org	1969
4	ANDOVER	Congregation Beth Israel	501 S. Main St.(PO Box929), 01810; 978-474-0540;office@bethisraelmv.org	2004
5	ANDOVER	Chabad Lubavitch of MerrimackValley	310 N. Main St., 01810;978-470-2288;rabbi@chabadmv.com	1992
6	ANDOVER	Temple Emanuel	7 Haggetts Pond Rd.,01810; 978-470-1356;www.templemanuel.net	1920
7	ANDOVER	Havurat Shalom	17 Olde Berry Rd.,01810; 978-470-2542	
8	ANDOVER	Tifereth Israel	510 S. Main St., 01810	c. 1990
9	ARLINGTON	Hebrew Congregation of ArlingtonHeights	122 Sylvia St.	1906
10	ASHLAND	Shaarei Shalom, Reform JewishCongregation of Ashland	PO Box 454, 01721; 508-231-4700; info@rjca.org	1995
11	ATHOL	Agudas Achim	411 Pine Street	1907
12	ATHOL	Tempel Israel	107 Walnut St., 01331;978-249-9481;www.templeisraelathol.org	1950
13	ATTLEBORO	Agudas Achim Congregation	301 N. Main St., 02703;508-222-2243;office@agudasma.org	1910/1911
14	AYER	Congregation Anshey Shalom	Cambridge St. &Norwood Ave.	1948
15	BELMONT	Beth El Temple Center	2 Concord Av., 02178;617-484-6668;office.betc@verizon.net	1932/1936
16	BEVERLY	Tempel B'nai Abraham	200 East Lothrop St.,01915; 978-927-3211;tba.beverly@verizon.net	1897/ 1908
17	BEVERLY	Congregation Ohav Shalom	3 Beckford St.	1903/1916
18	BOSTON	Chabad House	491 CommonwealthAve., 02215; 617-424-1190;rsp@chabadboston.org	1979
19	BOSTON	Tempel Israel	477 Longwood Ave,02215; 617-566-3960;info@tisrael.org	1854
20	BOSTON	Congregation Bnai Jacob/ Zvhil-Mezbuz Beis Medrash	15 School St., 02108; 617-227-8200;Rebbe@rebbe.org	1919
21	BOSTON	Congregation Ohav Sedek		<1887



# Demonstration of Tool - Google Sheets

Here, we have different types of information in the same column: street addresses, phone numbers, emails.

We won't be standardizing these columns in our own work on this dataset, but we'll practice splitting this information up as a way to discuss how delimiters work in CSV files.

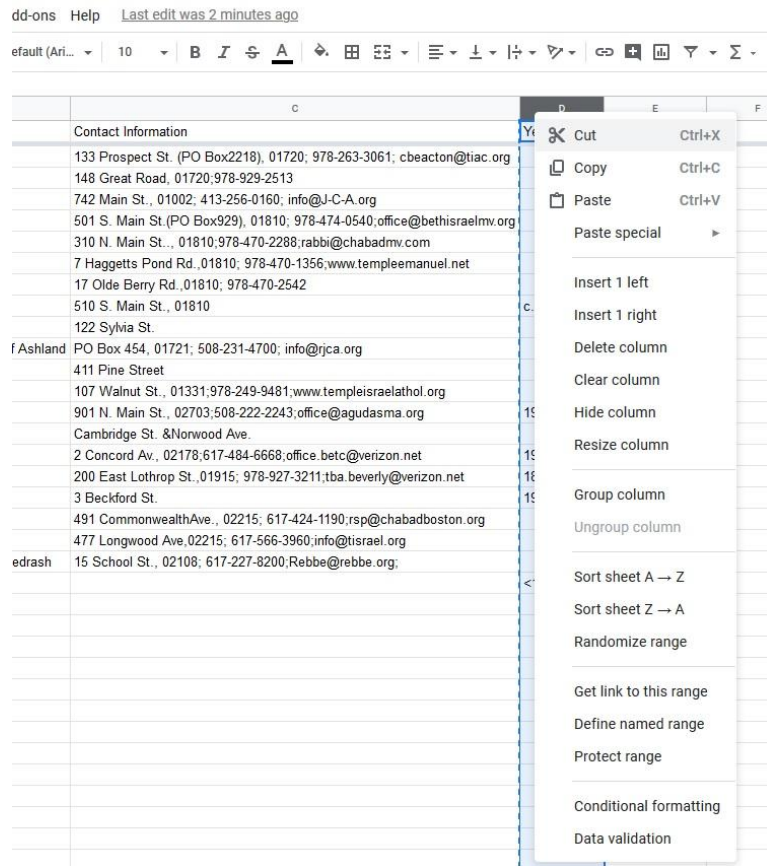
	C	D
	Contact Information	Year found
	133 Prospect St. (PO Box2218), 01720; 978-263-3061; cbeacton@tiac.org	1911
	148 Great Road, 01720;978-929-2513	1911
	742 Main St., 01002; 413-256-0160; info@J-C-A.org	1911
	501 S. Main St.(PO Box929), 01810; 978-474-0540;office@bethisraelmv.org	2011
	310 N. Main St., 01810;978-470-2288;rabbi@chabadmv.com	1911
	7 Haggetts Pond Rd.,01810; 978-470-1356;www.templemanuel.net	1911
	17 Olde Berry Rd.,01810; 978-470-2542	
	510 S. Main St., 01810	c. 1990
	122 Sylvia St.	1911
f Ashland	PO Box 454, 01721; 508-231-4700; info@rjca.org	1911
	411 Pine Street	1911
	107 Walnut St., 01331;978-249-9481;www.templeisraelathol.org	1911
	901 N. Main St., 02703;508-222-2243;office@agudasma.org	1910/1911
	Cambridge St. &Norwood Ave.	1911
	2 Concord Av., 02178;617-484-6668;office.betc@verizon.net	1932/1936
	200 East Lothrop St.,01915; 978-927-3211;tba.beverly@verizon.net	1897/ 190
	3 Beckford St.	1903/1916
	491 CommonwealthAve., 02215; 617-424-1190;rsp@chabadboston.org	1911
	477 Longwood Ave,02215; 617-566-3960;info@tisrael.org	1911
ledrash	15 School St., 02108; 617-227-8200;Rebbe@rebbe.org;	1911
		<1887



# Demonstration of Tool - Google Sheets

Since we have three different kinds of info, we have to prepare two more empty columns to project our results into.

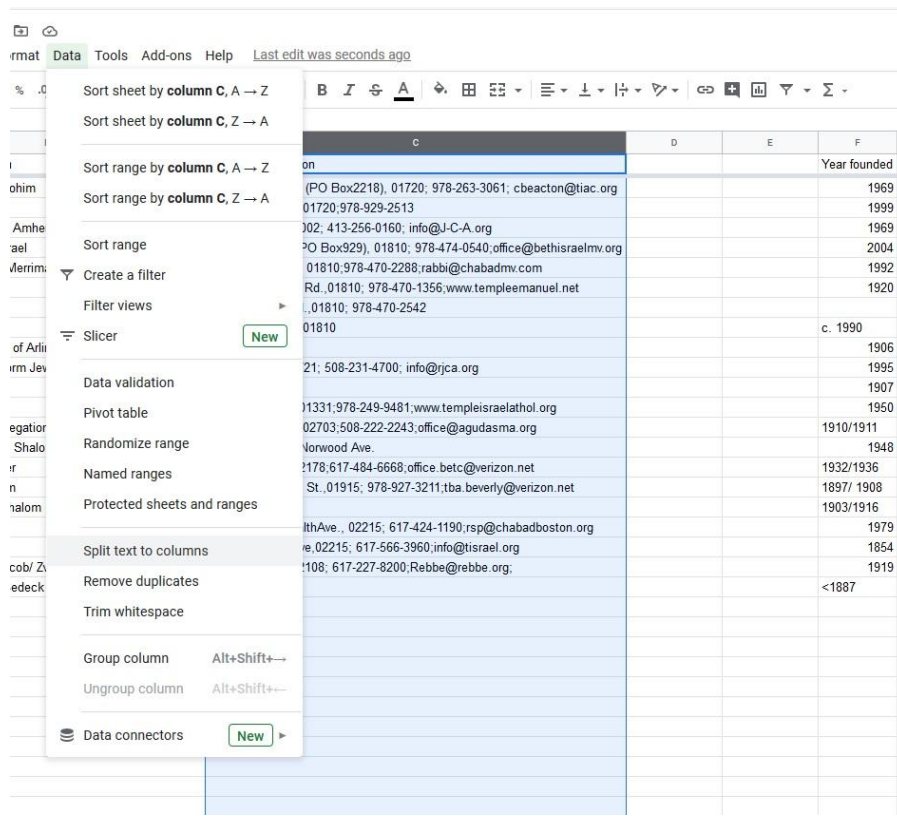
Otherwise, the new columns will be pasted across already existing columns





# Demonstration of Tool - Google Sheets

Go to "Data," select "Split text to columns"



The screenshot shows the Google Sheets interface with the 'Data' menu open. The 'Split text to columns' option is highlighted. The background spreadsheet contains data in columns C, D, E, and F. Column C contains contact information, and column F contains 'Year founded'.

C	D	E	F
on			Year founded
(PO Box2218), 01720, 978-263-3061; cbeacton@tiac.org			1969
01720, 978-929-2513			1999
002, 413-256-0160; info@J-C-A.org			1969
PO Box929), 01810, 978-474-0540; office@bethisraelmv.org			2004
01810, 978-470-2288; rabbi@chabadmn.com			1992
Rd, 01810, 978-470-1356; www.templemanuel.net			1920
01810, 978-470-2542			
01810			c. 1990
			1906
21, 508-231-4700; info@njca.org			1995
			1907
11331; 978-249-9481; www.templeisraelathol.org			1950
02703; 508-222-2243; office@agudasma.org			1910/1911
Norwood Ave.			1948
2178; 617-484-6668; office.betc@verizon.net			1932/1936
St., 01915, 978-927-3211; tba.beverly@verizon.net			1897/ 1908
1thAve., 02215; 617-424-1190; rsp@chabadboston.org			1903/1916
re, 02215; 617-566-3960; info@tisrael.org			1979
108; 617-227-8200; Rebbe@rebbe.org;			1854
			1919
			<1887



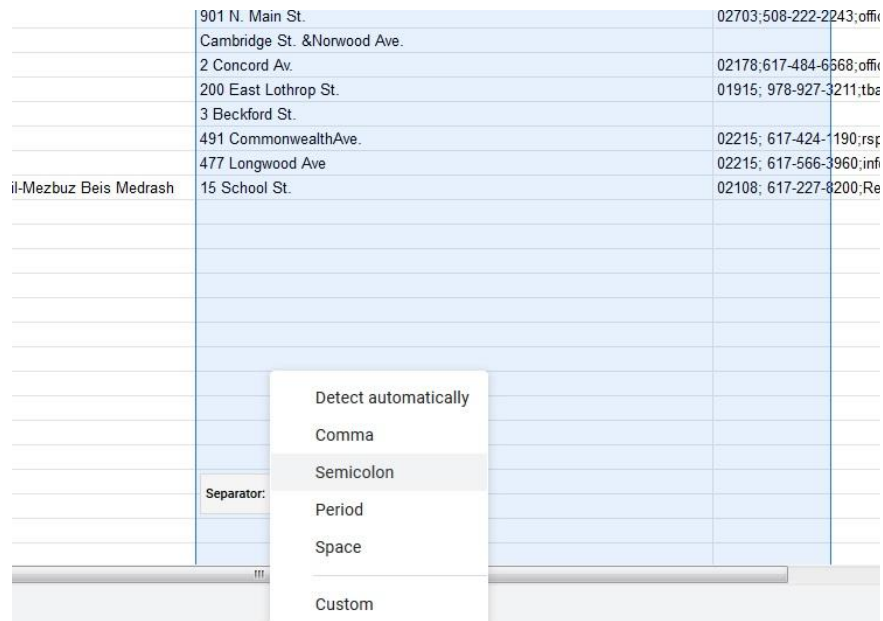
Google Sheets proceeds to separate the text based on commas, splitting the street addresses in the middle.

We need some other separator.

[illegible]

# Demonstration of Tool - Google Sheets

We click on that small window for choosing a "Separator" and select semicolon



Bummer - what happened? Why are we missing the “year founded” column?

*Feel free to ask questions at any point during the presentation!*



(Ari... 10 B I S A                    

	C	D	E	F
	Contact Information			Year founded
	133 Prospect St. (PO Box2218), 01720; 978-263-3061; cbeacton@tiac.org			1969
	148 Great Road, 01720;978-929-2513			1999
	742 Main St., 01002; 413-256-0160; info@J-C-A.org			1969
	501 S. Main St.(PO Box929), 01810; 978-474-0540;office@bethisraelmv.org			2004
	310 N. Main St., 01810;978-470-2288;rabbi@chabadmv.com			1992
	7 Haggetts Pond Rd.,01810; 978-470-1356;www.templemanuel.net			1920
	17 Olde Berry Rd.,01810; 978-470-2542			
	510 S. Main St., 01810			c. 1990
	122 Sylvia St.			1906
and	PO Box 454, 01721; 508-231-4700; info@rjca.org			1995
	411 Pine Street			1907
	107 Walnut St., 01331;978-249-9481;www.templeisraelathol.org			1950
	901 N. Main St., 02703;508-222-2243;office@agudasma.org			1910/1911
	Cambridge St. &Norwood Ave.			1948
	2 Concord Av., 02178;617-484-6668;office.betc@verizon.net			1932/1936
	200 East Lothrop St.,01915; 978-927-3211;tba.beverly@verizon.net			1897/ 1908
	3 Beckford St.			1903/1916
	491 CommonwealthAve., 02215; 617-424-1190;rsp@chabadboston.org			1979
	477 Longwood Ave,02215; 617-566-3960;info@tisrael.org			1854
1	15 School St., 02108; 617-227-8200;Rebbe@rebbe.org;			1919
				<1887



*Feel free to ask questions at any point during the presentation!*

# One tiny semicolon messes up everything...

668;office.betc@verizon.net			1932/1936
3-927-3211;tba.beverly@verizon.net			1897/ 1908
			1903/1916
617-424-1190;rsp@chabadboston.org			1979
666-3960;info@tisrael.org			1854
8200;Rebbe@rebbe.org;			1919
			<1887



efault (Ari... 10 B I S A                    

*Feel free to ask questions at any point during the presentation!*



## Additional Issues

- Quote character

new plots in Deerham, A., , , , , to include  
,"148 Great Road, 01720; 978-929-2513",  
Community of Amherst "742 Main St. 01

- Special characters

City/Town

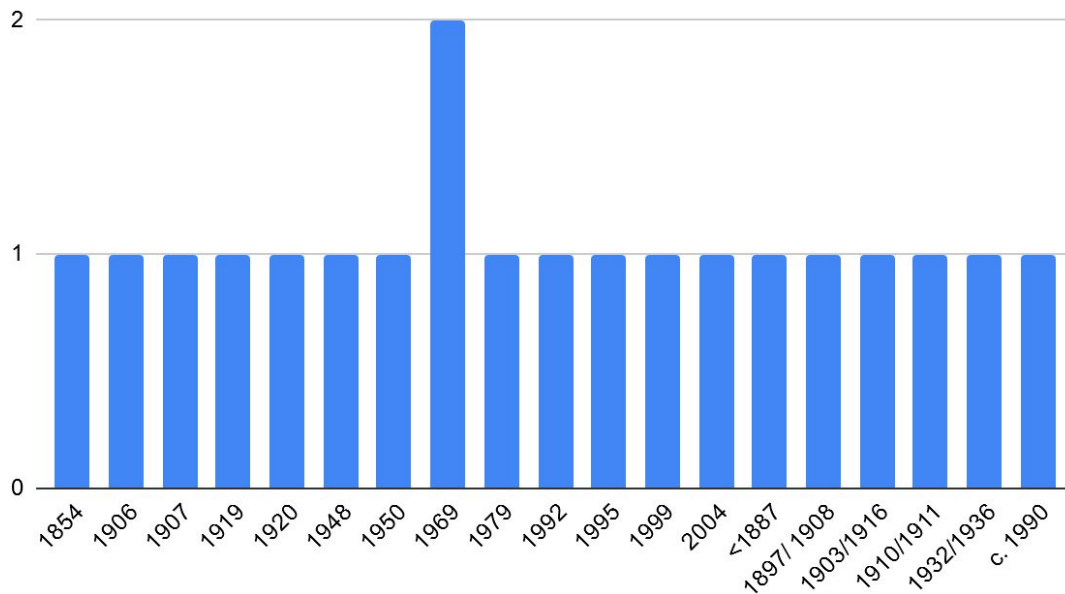
- Missing values

 $\dot{X}, , X$ 



# Common Pitfall #1: Data Types

Number of congregations founded per year



# Common Pitfall #1: Data Types

example\_dataset

File

Edit

View

Insert

Format

Data

Tools

Add-ons

Help

Last edit was 3 minutes ago

100%

\$

%

.0

.00

123

Arial

10

B

I

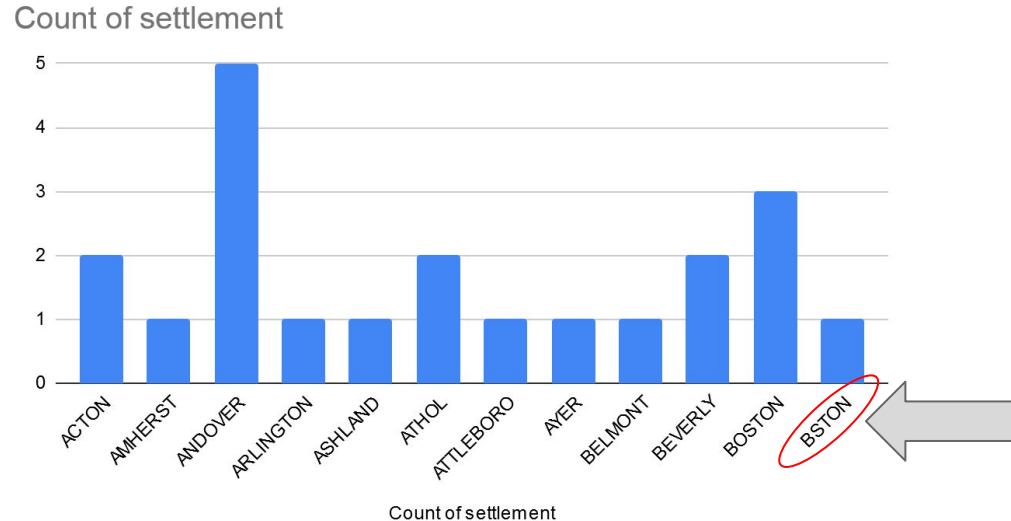
U

A

- This column requires standardization
- Standardization means humans making decisions on what is a standard form of the information



# Common Pitfall #2: Controlled Vocabulary



# Controlled Vocabularies

- Controlled vocabularies, a way to standardize input of categories: choose from a list prepared in advance

example\_dataset ☆ 📄 ☁

File Edit View Insert Format Data Tools Add-ons Help Last edit was 3 minutes ago

100% \$ % .0 .00 123 Arial 10 B I U

	A	B	C	D	E	F
1	city/town	Name of congregation	Contact Information	Year founded		
2	ACTON	Congregation Beth Elohim	133 Prospect St. (PO Box 2218), 01720; 978-263-3061; cbeacton@tiac.org	1969		
3	ACTON	Chabad Center	148 Great Road, 01720; 978-929-2513	1999		
4	AMHERST	Jewish Community of Amherst	742 Main St., 01002; 413-256-0160; info@J-C-A.org	1969		
5	ANDOVER	Congregation Beth Israel	501 S. Main St. (PO Box 929), 01810; 978-474-0540; office@bethisraelmv.org	2004		
6	ANDOVER	Chabad Lubavitch of Merrimack Valley	310 N. Main St., 01810; 978-470-2288; rabbi@chabadmv.com	1992		
7	ANDOVER	Temple Emanuel	7 Haggetts Pond Rd., 01810; 978-470-1356; www.templemanuel.net	1920		
8	ANDOVER	Havurat Shalom	17 Olde Berry Rd., 01810; 978-470-2542			
9	ANDOVER	Tiereth Israel	510 S. Main St., 01810	c. 1990		
10	ARLINGTON	Hebrew Congregation of Arlington Heights	122 Sylvia St.	1906		
11	ASHLAND	Sugar Shalom, Reform Jewish Congregation of Ashland	PO Box 454, 01721; 508-231-4700; info@rjca.org	1995		
12	ATHOL	Agudas Achim	411 Pine Street	1907		
13	ATHOL	Temple Israel	107 Walnut St., 01331; 978-249-9481; www.templeisraelathol.org	1950		
14	ATTLEBORO	Agudas Achim Congregation	901 N. Main St., 02703; 508-222-2243; office@agudasma.org	1910/1911		
15	AYER	Congregation Anshey Shalom	Cambridge St. & Norwood Ave.	1948		
16	BELMONT	Beth El Temple Center	2 Concord Av., 02178; 617-484-6668; office.betc@verizon.net	1932/1936		
17	BEVERLY	Temple B'nai Abraham	200 East Lothrop St., 01915; 978-927-3211; tba.beverly@verizon.net	1897/ 1908		
18	BEVERLY	Congregation Ohav Shalom	3 Beckford St.	1903/1916		
19	BOSTON	Chabad House	491 Commonwealth Ave., 02215; 617-424-1190; rsp@chabadboston.org	1979		
20	BSTON	Temple Israel	477 Longwood Ave, 02215; 617-566-3960; info@tisrael.org	1854		
21	BOSTON	Congregation Bnai Jacob/ Zvihil-Mezbuz Beis Medrash	15 School St., 02108; 617-227-8200; Rebbe@rebbe.org;	1919		
22	BOSTON	Congregation Ohavi Sedek		<1887		
23						




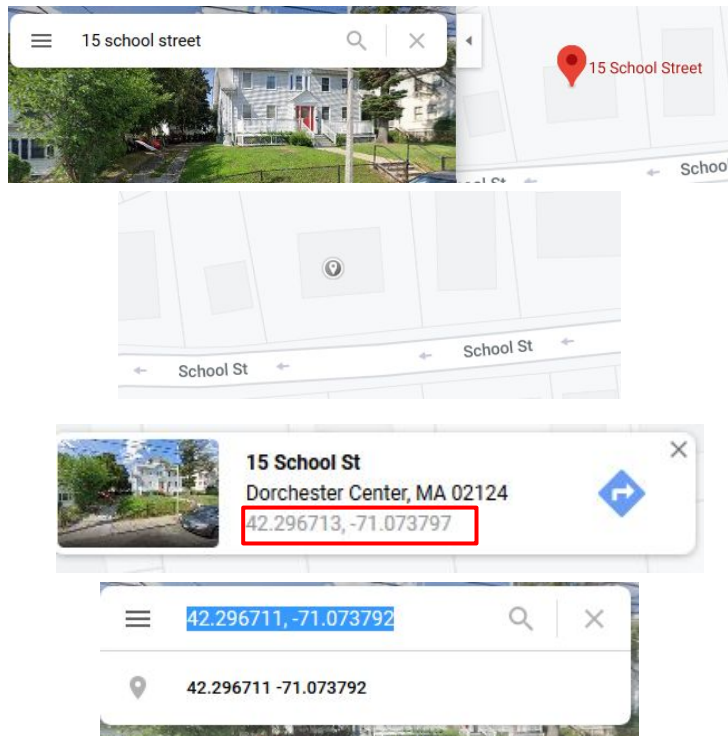
# Introducing the activity

- Everyone covers a certain number of rows
- Standardize the data in the year\_founded and year\_incorp columns
- Look up cases where the years when congregations were founded are unknown and add those to the 'year\_founded' column
- Identify any physical addresses that are missing (including PO boxes), with as much detail as is possible (use the coordinates for the whole settlement as a fallback)
- Add the coordinates of the addresses into the latitude and longitude columns. Use Google Maps to find the coordinates of the congregations.



# Finding geo-coordinates in Google

1. Type in the address and find it on the map
2. Remove the query (  ) and drop a pin on the same location (click or double click on the map)
3. Click on the geo-coordinates
4. Copy them from the search tab on the left side



# Your Turn!

Open the "copy\_dataset" spreadsheet and go to the first row you were assigned

Standardize that row and a few more rows: input the missing data as outlined in the assignment instructions - dates and geo coordinates

Note: later one person will use the “Split text to column” option to separate the geo coordinates into different columns. But for now we are copying the whole string into a single column “geocoordinates”



# In the next episodes...

Fun stuff is ahead. But the tediousness is a part of the process. This is also helpful in getting familiar with the dataset you will be working on, which is essential.





# In the next episodes...

This dataset will be useful in particular for two following assignments:

- 1) Making maps and visualizations in **Tableau**: the goal is to make big picture visualizations and try to identify trends or different divisions among the congregations
- 2) Making stories about different neighborhoods or congregations in **Knight Lab's StoryMap**: rather than big picture, this tool enables us to zoom in onto particular interesting cases that might help make sense of the broader picture identified with the help of Tableau



# Post-Exploration Group Discussion

What did you notice?

How does the world of data appear to you now?

What things were weird/tricky?

How might you use this way of thinking in the future?



# Thank you!

If you have any questions, contact DITI at [nulab.info@gmail.com](mailto:nulab.info@gmail.com)

**Developed and taught by**

**Colleen Nugent**

Digital Integration Teaching Initiative  
DITI Research Fellow

**Milan Skobic**

Digital Integration Teaching Initiative  
DITI Assistant Director

Slides, handouts, and data available at

<https://bit.ly/diti-fall2020-rabinovitch2>

Schedule an appointment with us! <https://calendly.com/diti-nu>



**Northeastern University**  
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point  
during the presentation!*