

This spreadsheets_hist2430_readme.txt file was generated on 2020-08-21 by Milan Skobic
Date format: [W3C/ISO 8601 date standard](#) - YYYY-MM-DD

GENERAL INFORMATION

1. Title of Datasets: Spreadsheets for the class HIST 2430 - Digital Histories of Ethnic Boston - Fall 2020

2. Author Information

Contact Information - class instructor

Name: Simon Rabinovitch

Institution: Northeastern University, Boston campus

Address: 360 Huntington Ave, Boston, MA 02115

Email: s.rabinovitch@northeastern.edu

Contact Information - dataset initiator

Name: Milan Skobic

Institution: Northeastern University, Boston campus

Address: 360 Huntington Ave, Boston, MA 02115

Email: skobic.m@northeastern.edu

3. Date of data collection (single date, range, approximate date)

2020-06-15 - 2020-06-29

4. Geographic location of data collection

Boston, MA

SHARING/ACCESS INFORMATION

1. Licenses/restrictions placed on the data:

Datasets are for in-class use only; the produced visualizations are intended for public presentation online

2. Links to other publicly accessible locations of the data:

<http://jgsgb.org/pdfs/MassSynagogues.pdf>

3. Was data derived from another source?

Yes, from <http://jgsgb.org/pdfs/MassSynagogues.pdf>

4. Recommended citation for this dataset:

TBD

DATA AND FILES OVERVIEW

1. File List:

raw_dataset

example_dataset

copy_dataset

2. Are there multiple versions of the dataset?

No

3. raw_dataset short description

This dataset contains the reproduction of the dataset obtained from the source website administrators with their permission in the form of an excel spreadsheet, into a google spreadsheet from which it can be turned into a .csv file with no modifications.

4. example_dataset short description

This spreadsheet contains a sample of data from raw_dataset. This spreadsheet serves exclusively educational purposes. Data has been modified in order to exemplify data cleaning and data standardization issues and procedures that are presented in the course of the class.

5. copy_dataset short description

This spreadsheet presents a working folder in which the instructors and students engage in data standardization- New columns (e.g. geo_coordinates, religious types of congregations) and data are being introduced during the preparation and execution of the class.

6. Format of the file

Files are created in google spreadsheet, and will be exported as .csv files for further use.

7. Date that raw_dataset was created

2020-08-03

7a. Date that raw_dataset was updated and the nature of the updates

None

8. Date that example_dataset was created

2020-07-13

8a. Dates that example_dataset was updated and the nature of the updates

2020-08-03

Inputting first 4 columns and first 20 records (21 rows including heading)

Populating first column with settlement names
Introducing one entry (cell A20) with misspelled name for classroom demonstration purposes (controlled vocabularies)
Introducing one entry (cell C21) with added semicolon for classroom demonstration purposes (splitting text into columns)

9. Date that copy_dataset was created
2020-07-20

9a. Date that copy_dataset was updated and the nature of the updates
2020-08-03

Changing column names to be more machine readable
Changing "Year founded" column into "year_orig"; added "year_founded" and "year_incorp" for entering standardized dates
"year_orig" column is protected
Populating "settlement" column with settlement names
Unmerging the merged cells in first 20 records and copied the initial entry into each new cell

2020-08-13
Added additional protected index column in case rows get mishandled in the further processing

Added geodata where known for first 20 records
2020-09-17
Merged previously existing columns "latitude" and "longitude" into "geocoordinates", to be split in later work into the two. It is formatted "latitude,longitude".

2020-09-18
Standardized rows concluding with the row 48, and developed an initial guide to standardization of entry of geocoordinates in unclear cases

METHODOLOGICAL INFORMATION

1. Description of methods used for collection/generation of data:

After data was obtained, it was copied into the raw_dataset, a sample of it was copied into example_dataset and there modified for educational purposes, and the whole raw_dataset was copied and modified in the copy_dataset file.

2. Methods for processing the data:

3. Instrument- or software-specific information needed to interpret the data:

Google sheets; .csv compatible instruments and software

4. Standards and calibration information, if appropriate:

Guide developed for standardization of dates in the copy_dataset file:

- "c. year", change to year (for example, change "c. 1950" to "1950")
- "<" or "> year", change to year
- when there is a year range indicated or multiple years (e.g., 1896-97; 1893 or 1904), change to the earliest year indicated
- when the decade is indicated (e.g. 1980s), change to the middle of the decade (e.g., 1985)
- when there is a year with text (e.g., 1960s as Harvard Hillel Children's School) change to the year indicated and remove the text (e.g., 1965)
- empty cells, or cases where the date is unknown, change to "unknown"

Guide developed for standardization of geocoordinates entry from google maps

- when google maps search drops the pin on another address:
 - If the address is next to the searched address, enter the coordinates of the parcel next to it
 - If the address is significantly different in number or in name, enter "unknown"
- when google maps search drops the pin on the street
 - Determine the side of the street on which the address location should be located, and enter the coordinates of the point on which that side of the street is the closest to the dropped pin
- In both cases, enter the description of the situation with that address into the research journal

5. Describe any quality-assurance procedures performed on the data:

Team reading and revision of the dataset

6. People involved with sample collection, processing, analysis and/or submission:

Colleen Nugent, Sarah Connell, Laura Nelson, Simon Rabinovitch, Milan Skobic

DATA-SPECIFIC INFORMATION FOR: raw_dataset

1. Number of variables:

18

2. Number of cases/rows:

575 rows, heading + 574 records

3. Variable List:

<list variable name(s), description(s), unit(s) and value labels as appropriate for each>

City/Town, location of congregation, place in Massachusetts or no entry, no label
 Name of congregation, name of congregation, no standard unit, no label
 Contact Information, address, phone number and e-mail where known, same unit, no label
 Yearfounded, approximation of year, if two years in the entry then the second signifies the
 year of incorporation, year in numerals with additional characters if necessary (eg. "c."), no label
 Founded by (if other than localJews), no description, no standard unit, no label
 Past locations, previous locations of congregation (if applicable), name of the place or address, no standard unit, no label
 Any mergers, information on mergers with whom and when, also can include information on closures and other information, no standard unit, no label
 Cemetery
 Yahrzeit
 Member
 Officer History
 Death
 Marriage
 Bris
 BBM
 Pictures
 OtherRecords

4. Missing data codes:

<list code/symbol and definition>n

5. Specialized formats or other abbreviations used:

DATA-SPECIFIC INFORMATION FOR: example_dataset

1. Number of variables:

4

2. Number of cases/rows:

22 rows, 21 cases + heading

3. Variable List:

<list variable name(s), description(s), unit(s)and value labels as appropriate for each>

City/Town, location of congregation

Name of congregation, name of congregation

Contact Information, address, phone number and e-mail where known

Year founded, approximation of year, if two years in the entry then the second signifies the

4. Missing data codes:

<list code/symbol and definition>n

5. Specialized formats or other abbreviations used:

DATA-SPECIFIC INFORMATION FOR: copy_dataset

<repeat this section for each dataset, folder or file, as appropriate>

1. Number of variables:

23

2. Number of cases/rows:

22 rows, 21 cases + heading

3. Variable List:

<list variable name(s), description(s), unit(s) and value labels as appropriate for each>

Settlement, location of congregation, place in Massachusetts, no label
name_of_congregation, name of congregation, no standard unit, no label
address, contains address, phone number and e-mail where known, same unit, no label
year_orig, approximation of year, if two years in the entry then the second signifies the
year of incorporation, year in numerals with additional characters if necessary (eg. "c."), no label
year_founded, year of founding the congregation, year in numerals, no label
year_incorp, year of incorporation of the congregation (if applicable), year in numerals,
no label
founded_by, information of founders if they were not local Jews, no standard unit, no
label
past_locations, previous locations of congregation (if applicable), name of the place or
address, no standard unit, no label
any_mergers, information on mergers with whom and when, also can include information
on closures and other information, no standard unit, no label
cemetery
yahrzeit
member

officer
history
death
marriage
bris
bbm
pictures
other_records

1, index column for the preservation of original order of rows, number of row, no label
geocoordinates, geographical latitude comma longitude, degree, no label

4. Missing data codes:

<list code/symbol and definition>n

5. Specialized formats or other abbreviations used: