**Grater Washington, DC Jewish Community Study, 2017**

**Cohen Center for Modern Jewish Studies, Brandeis University**

**Documentation of Public Use Data Set**

*For more information contact Janet Krasner Aronson,* [*jaronson@brandeis.edu*](mailto:jaronson@brandeis.edu)

*Version March 23, 2018*

**Introduction**

This document describes the Public Use Data Set for the 2017 Greater Washington, DC Jewish Community Study. It explains the constructed variables in the dataset and the procedures for statistical weighting.

Variables that begin with the prefix “x\_” were not part of the original data, but were constructed from other variables, some of which are not included in the public data set (this is noted in the descriptions where applicable). All other variables are documented in the codebook (Appendix D of the report) and are taken directly from the survey.

You can find the report and technical appendices at: https://www.brandeis.edu/ssri/communitystudies/dcreport.html

**About the Public Use Data Set**

The Public Use Data Set contains all raw data provided by respondents except for data that might be used to identify individual respondents. Primarily, these removed variables were open-ended responses and household ZIP codes.

Descriptions of some variable naming conventions follow:

1. All variables with a “resp” in the name refer to the respondent (e.g., respage is the age of the respondent and x\_respmartype is the constructed variable noting the marital status of the respondent).
2. Variables with a “hhad” prefix refer to non-respondent adults in the household; these variables range from 2-9 because the respondent is considered as the first adult (e.g., hhadage3 is the age of the third adult in the household).
3. Variables with a “hhch” prefix refer to the children in the household; these range from 1-6 (e.g., hhchgrd1 is the grade-level of the first child in the household).

*Potential Issues with Data Interpretation*

CMJS as a rule maintained the integrity of the data as collected. As such, two potential issues warrant caution. First, due to programming conventions, both refusals and “don’t know” responses are in many cases coded together with the code “999.” Responses skipped through survey logic and individually skipped items are coded the same way, as system missing (. in Stata). Second, there are cases where respondents answered a question and its follow-up before seemingly backtracking and changing the original response, which would otherwise render the follow-up response invalid. Such overwritten responses remain in the data set.

It is up to the analyst’s own interpretation as to how to account for these characteristics of the data set.

**Definitions of Constructed Variables (x\_ prefix)**

Variables were constructed from raw data for two purposes:

1. Standardized coding of open-ended data
2. Analytical variables created through complex combinations of multiple raw variables.

x\_surveystatus: Whether the respondent screened into the survey, screened out, or completed the recent mover survey (see below for more information on the movers).

x\_hhadct, x\_hhchct, x\_hhsize: The number of individuals living in households was recoded to match the number of people for whom information was provided; for example, when a respondent said she lived with four other adults but only provided information on three of them, x\_hhadct was recoded as three.

x\_citycounty, x\_region, x\_state: Respondents supplied the ZIP code of their primary residence in the Greater Washington, DC area. To protect their identities, this variable groups ZIP codes into municipal boundaries, as well as constructed regions that make up the Greater Washington, DC area (detailed explanations of these regions can be found in Chapter 3 of the report).

x\_resprelig, x\_hhadrelig2-9: Open-ended responses based off the “other religion” response were recoded to an existing religion whenever possible.

x\_respjewtype, x\_hhadjewtype2-9: These variables denote the “type of Jew” corresponding to the respondent or the household adults. Jews by Religion (JBR) are those who say their religion is Jewish and have a Jewish background (i.e., a Jewish parent, was raised Jewish, or converted). Jews of No Religion (JNR) are atheists and agnostics who consider themselves Jewish aside from religion and have a Jewish background, or say their religion is both Jewish and atheist/agnostic, and have a Jewish background. Jews of Multiple Religions (JMR) either say they have two religions, one of which is Judaism, and have a Jewish background, or have another religion but consider themselves Jewish aside from religion, and have a Jewish background. Jewish Background (JB) say they have a religion other than Judaism, do not consider themselves Jewish aside from religion, but have a Jewish background. Jewish Affinity (JA) consider themselves Jewish either by religion or not by religion, but do not have a Jewish background. Non-Jews (NJ) have a religion other than Judaism, do not consider themselves Jewish in any way, and do not have a Jewish background. Messianic Jews with a Jewish background were categorized as JB, and those without a Jewish background were categorized as NJ.

x\_respjewish; x\_hhadjewish2-9: Jewish adults are JBR, JNR, or JMR.

x\_respdenom, x\_hhaddenom2-9: Open-ended responses given as the “other denomination” response option were recoded to an existing or a newly grouped denomination wherever possible.

x\_hhchrelrsd1-6: In the original survey, after identifying the religion in which the first child is being raised, respondents were asked if all children are being raised in the same religion (see variable hhchrelsame). These variables fill in the responses for children 2-6 who have the same religion as child 1.

x\_hhchjew1-6: Children are counted as Jews if they are being raised Jewish by religion, culturally Jewish, or Jewish and another religion (corresponding to the variables x\_hhchrelrsd1-6).

x\_yaactivity\_adas: Due to a programming error, not all eligible respondents eligible were asked yaactivity\_adas; open-ended responses based on responses to of yaactivity\_oth were recoded to this variable.

x\_rlsyntype1-5: A recoding of the synagogue names (up to 5) that respondents belong to. The actual names were removed from the data set to protect identities. Traditional, brick-and-mortar synagogues are separated by denomination, while independent *minyanim*, Chabad, and non-local or unknown congregations are listed separately, regardless of denomination.

x\_rlholhigh: Those who did not attend services (see rlsynsvc) were coded as having not attended High Holy Day services.

x\_wcharanyamt: Categorized amount donated by respondents to charity in the past year, including no donations.

x\_wwbaid, x\_wbsave3month x\_wbrent x\_wbsave400 x\_wbjewlife: Those indicating a standard of living as “Prosperous” or “Living very comfortably” (see wbstan) were not asked these questions, and in those cases, they were recoded as not being in economic need.

x\_hhadjewct, x\_hhchjewct, x\_hhjewct: The number of Jewish adults, Jewish minor children, and total Jews in the household.

x\_hhjewish: Jewish households contain at least one JBR, JNR, or JMR adult. There are 62 households that screened into the survey but do not contain any Jewish adults or children. These households include at least one person of Jewish Background or Jewish Affinity.

x\_hhmartype: This notes if the household contains an inmarried or intermarried couple, or no couple (whether or not the respondent is part of the married couple).

x\_respmartype: This notes if the respondent is inmarried, intermarried, or unmarried.

x\_dcjengage: This is the pattern of engagement of the respondent (see Chapter 4 of the report).

x\_respagecat: Categories are based on the age and parental status of the respondent for respondents between the ages of 22 and 64. All those 65 and older are grouped together, regardless of parent status. The respondents younger than 22 are excluded from the variable.

x\_hhagecat: Categories for reporting are based on the combined ages of all adults in the household. If all adults are ages 22-39, households are classified as “young adults 22-39” households. If all adults are ages 65 and over, households are classified as “seniors 65+” households. If at least one adult is age 40-64, the household is classified as “adults 40-64” household. The few households that do not fit any of these categories are coded as missing (e.g. a household with a 20 and 70 year old would be missing).

**Weighting**

Three sets of weights are available for this dataset. One set is at the household level (x\_wtprimhh and x\_wtfullhh), and one is at the respondent level (x\_wtprimresp and x\_wtfullresp). A third set is used for recent mover variables only (x\_wtprimmove and x\_wtfullmove). Household-level weights should be used to calculate characteristics of the household, population counts, and anything involving children. Respondent-level weights should be used to calculate characteristics of respondents (e.g., behaviors and attitudes). Young adult weights are respondent-level weights designed for estimates of young adult specific variables. Mover weights are household-level weights designed for estimates of those who recently moved away from Metro DC (all starting with “mv”).

The weight variables are also segmented by whether they refer to the primary sample (x\_wtprimhh, x\_wtprimresp, and x\_wtprimmove) or the full sample (x\_wtfullhh, x\_wtfullresp, and x\_wtfullmove). Primary weights are the only ones appropriate for generating counts or characteristics of the overall population. The full-sample weights are appropriate only for generating characteristics of subpopulations.

Notes: The public-use dataset includes all screener data, but households that screened out of the survey have all their weights set at 0. Non-Jewish households that screened into the survey (x\_hhjewish = 0) because they include an adult categorized as Jewish Background or Jewish Affinity have household-level weights (x\_wtprimhh, x\_wtfullhh). Weighting instructions are designed for use with Stata.

Constructed variables for weighting

x\_draw: If the respondent comes from the primary, supplementary, or RDD frames. Data from the RDD frame is used as part of the supplement.

x\_stratafull: The strata identified is used for weighting.

x\_wtprimhh: The primary-sample household weight.

x\_wtfullhh: The full-sample household weight.

x\_wtprimresp: The primary-sample respondent/individual weight.

x\_wtfullresp: The full-sample respondent/individual weight.

x\_wtprimmove: The primary-sample weight for mover-only variables (starting with “mv”).

x\_wtfullmove: The full-sample weight for mover-only variables (starting with “mv”).

Primary Weights

*Use ONLY the primary sample for generating population counts*

For household estimates—estimations on the number of households—use wtprimhh to estimate the percentage of households. For estimations on the number of people—i.e., counts—use the household weights with totals of count variables—e.g., x\_hhadct, x\_hhchct.

svyset \_n [pweight= x\_wtprimhh], strata(x\_stratafull) vce(linearized) singleunit(missing)

Use the respondent weights for questions asked only of respondents. Use wtprimresp to estimate percentage of adults (including respondent and non-respondent adults):

svyset \_n [pweight=x\_wtprimresp], strata(x\_ stratafull) vce(linearized) singleunit(missing)

Primary weights should be used for generating characteristics of the population as a whole, including counts and subgroups.

*Example: Household Characteristics*

The number or proportion of households that are synagogue members:

svyset \_n [pweight= x\_wtprimhh], strata(x\_ stratafull) vce(linearized) singleunit(missing)

svy, subpop(x\_hhjewish): tab rlsynany, count

svy, subpop(x\_hhjewish): tab rlsynany

*Example: Counts of Adults*

svyset \_n [pweight= x\_wtprimhh], strata(x\_stratafull) vce(linearized) singleunit(missing)

svy, subpop(x\_hhjewish): total x\_hhadct

*Example*: *Respondent or Individual Characteristics*

The proportion of close Jewish friends:

svyset \_n [pweight= x\_wtresp], strata(x\_stratafull) vce(linearized) singleunit(missing)

svy, subpop(x\_respjewish): tab jlfriend

Full Weights

*DO NOT use the full weights to generate population counts.* Full weights should be used for characteristics of subgroups (e.g., percentage of synagogue-member households that belong to other Jewish organizations)

svyset \_n [pweight= x\_wtfullhh], strata(x\_stratafull) vce(linearized) singleunit(missing)

svyset \_n [pweight= x\_wtfullresp], strata(x\_stratafull) vce(linearized) singleunit(missing)

*Example: Household Subpopulations*

The proportion of synagogue-member households that also belong to Jewish organizations:

svyset \_n [pweight= x\_wtfullhh], strata(x\_stratafull) vce(linearized) singleunit(missing)

svy, subpop(x\_hhjewish): tab orgmem rlsynany, col

*Example: Respondent or Individual Characteristics*

The proportion of close Jewish friends held by those who rent or own their homes:

svyset \_n [pweight= x\_wtfullresp], strata(x\_stratafull) vce(linearized) singleunit(missing)

svy, subpop(x\_respjewish): tab jlfriend locown, col

Mover Weights

Those who recently moved away from Metro DC were asked a special set of survey questions (see Chapter 3). These variables are:

* mvwhen\_
* month
* mvwhen\_year
* mvwhen\_month\_dkr
* mvregion
* mvregionyr
* mvplans
* mvback
* mvbase
* mvfam
* mvrespgen
* mvrespage
* mvresprelig
* mvrespconsider
* mvrespparents
* mvresprelrsd
* mvrespconvert
* mvadct
* mvadjct
* mvchct
* mvchjct
* mvjob
* mvgov
* mvwork
* mvspouse
* mvjobsp
* mvgovsp
* mvworksp
* mvsyn

x\_wtprimmove should be used in the same manner as primary weights, and x\_wtfullmove should be used in the same manner as full weights.