Greater Pittsburgh Jewish Community Study, 2017
Cohen Center for Modern Jewish Studies, Brandeis University
Documentation of Public Use Data Set
For more information contact Matthew Boxer, mboxer@brandeis.edu
Version March 9, 2018

Introduction

This document describes the Public Use Data Set for the 2017 Greater Pittsburgh 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:
http://www.brandeis.edu/ssri/communitystudies/pittsburghreport.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. The syntax used to create complex variables will be included in the report appendix.

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_county: Respondents supplied the ZIP code of their primary residence in the Greater Pittsburgh area. To protect their identities, this variable groups ZIP codes into the City of Pittsburgh, the remainder of Allegheny county, and combined Beaver, Butler, Washington, and Westmoreland counties.

x_region: Respondents supplied the ZIP code of their primary residence in the Greater Pittsburgh area. To protect their identities, this variable groups ZIP codes into constructed regions that make up the Greater Pittsburgh area (detailed explanations of these regions can be found in Chapter 2 of the report).

x_respedu: Open-ended responses based off the "other" response were recoded to an existing education level whenever possible.

x_resprelig, x_hhadrelig2-6: Open-ended responses based off the "other religion" response were recoded to an existing religion whenever possible.

x_respjewtype, x_hhadjewtype2-6: 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 they are 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_hhadjewish2-6: Jewish adults are JBR, JNR, or JMR.

x_respdenom, x_hhaddenom2-6: Open-ended responses given as the "other denomination" response option were recoded to an existing or a newly grouped denomination wherever possible.

x_hhadwhere2-6: Open-ended responses based off the "other" response were recoded to an existing education level whenever possible.

x_hhchrelrsd1-9: 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_hhchjew1-9: 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_pkjewish1-5: The variables incorporate those who responded "no" to pknow1-5.

x_rlsyntype1-4: 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_isrpartcorrect: This constructed variable combines open-ended and quantitative responses to isrpart, whether the respondent knows the region exists, or not, and can correct identify it, or not.

x_ x_pwaid_ssdissi x_pwaid_energy x_pwaid_welfare x_pwaid_unemp 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_hhjewish: Jewish households contain at least one JBR, JNR, or JMR adult. There are 38 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_hhadjewct, x_hhchjewct, x_hhjewct: The number of Jewish adults, Jewish minor children, and total Jews in the household.

x_respmartype: This notes if the respondent is inmarried, intermarried, or unmarried.

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_pittjengage: This is the pattern of engagement of the respondent (see Chapter 3 of the report).

x_locyears3cat: The respondent's length of residence in Greater Pittsburgh.

x_synlocal: If the household belongs to a local congregation or no congregation; those belonging to a non-local synagogue are coded as missing.

x_hhchild, x_parresp: If there is a minor child in the household, and if the respondent is a parent of a minor child.

x_respage4cat: The age categories of the respondent used in the report.

Weighting
Two 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). 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).

The weight variables are also segmented by whether they refer to the primary sample (x_wtprimhh and x_wtprimresp) or the full sample (x_wtfullhh and x_wtfullresp). 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 to missing (. in Stata). Non-Jewish households that screened into the survey (x_hhjewish = 0) because they include an adult categorized as Jewish Background or Jewish Affinity also have weights set to missing. Weighting instructions are designed for use with Stata.

Constructed variables for weighting

x_draw: If the respondent comes from the primary or supplementary frames. See methodological appendix (Appendix A) for further details.

x_strata: 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.

<u>Primary Weights</u>

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_strata) 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_strata) 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_strata) vce(linearized) singleunit(missing)
svy, subpop(x_respjewish): tab jlfriend

<u>Full Weights</u>

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_strata) vce(linearized) singleunit(missing)
svyset _n [pweight= x_wtfullresp], strata(x_strata) 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_strata) 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_strata) vce(linearized) singleunit(missing)

svy, subpop(x_respjewish): tab jlfriend locown, col