How Couples Meet and Stay Together (HCMST) 2017, 2020 and 2022

User's Guide

Draft March 12, 2024

Michael J. Rosenfeld*

* Michael J. Rosenfeld, Professor of Sociology, Stanford University. Web: www.stanford.edu/~mrosenfe. Email: mrosenfe@stanford.edu. Data gathered through the support of the National Science Foundation grant SES 2030593, and Stanford's United Parcel Service endowment. Data and documentation home: https://data.stanford.edu/hcmst2017

	HCMST 2017	HCMST 2020	HCMST 2022
Weight	**w1_weight_combo** (analytic weight) **w1_weight_combo_freqwt** (frequency weight)	**w2_combo_weight** (analytic weight) **w2_attrition_adj_weights** (analytic weight, adjusted for attrition from w1)	**w3_combo_weight** (analytic weight) **w3_attrition_adj_weight** (analytic weight adjusted for attrition from w1)
Sampling frame, theoretical	English literate adults in the US age 18+	English literate adults in the US age 21+	English literate adults in the US age 23+
Sampling frame, actual	Members of Ipsos KnowledgePanel, age 18+	Members of Ipsos KnowledgePanel, age 21+, who were subjects of HCMST 2017 and who were still in the KnowledgePanel in 2020	Members of Ipsos KnowledgePanel, age 23+, who were subjects of HCMST 2017 and who remained in the KnowledgePanel in 2022
Fielded	July, 2017	September-October 2020 **w2_month_of_survey**	March-April 2022 **w3_month_interview**
Sample size	3,510	2,107 variable: **w2_surveyed**	1,722 variable: **w3_surveyed**
Response Rate	4,033/6,753=60% (what Ipsos refers to as the screener rate) or 3,510/6,753= 52% (response rate) See also DiSogra and Callegaro (2008) for response rate to initial KnowledgePanel screener years ago.	2,107/2,431=87%	1,722/2,073=83%
Technical information	**caseid_new** (unique subject identifier) **w1_duration** (survey duration in minutes; not to be confused with relationship duration)	**caseid_new** (unique subject identifier) **w2_duration** (survey duration in minutes; not to be confused with relationship duration)	**caseid_new** (unique subject identifier) **w3_duration** (survey duration in minutes; not to be confused with relationship duration)
Survey Sections	Three Sections: 1) partnered adults 2) previously partnered adults 3) never partnered adults Variable: **w1_section**	Six sections, based on relationship status coming out of wave 1: 1) Still married 2) Marriage ended 3) Still in relationship (unmarried at last wave) 4) Relationship (unmarried at last wave) ended 5) Was unpartnered, now partnered 6) Was unpartnered still unpartnered Variable: **w2_section**	Six sections, based on relationship status from last contact (w1 or w2): 1) Still married 2) Marriage ended 3) Still in relationship (unmarried at last wave) 4) Relationship (unmarried at last wave) ended 5) Was unpartnered now partnered 6) Was unpartnered still unpartnered Variable: **w3_section**
Basic demography of study subjects. Note: pp* variables were supplied by Ipsos as part of their regular	**w1_ppage** (age) **w1_ppeduc** (education) **w1_ppgender** (gender)	**w2_ppage** (age) **w2_ppeduc** (education) **w2_ppgender** (gender)	**w3_ppage** (age) **w3_ppeduc** (education) **w3_ppgender** (gender)

background surveys. Note also: this is NOT a complete list! See codebooks, instruments, and data for more.	**w1_ppethm** (race/ethnicity) **w1_ppincimp** (HH income) **w1_ppreg9** (US region) **w1_ppmsacat** (metro residence) **w1_ppwork** (employment status) **w1_PPT01, w1_PPT25, w1_PPT612, w1_PPT1317, w1_PPT18OV** (# of children in the HH of different ages) **w1_partyid7** (political party affiliation) **w1_ppp20071** (born-again status) **w1_ppp20072** (religious service attendance) **w1_q14** (mother's education) **w1_q15a1** (country where grew up) **w1_q17** (times married) **w1_q16** (number of relatives seen each month) **w1_attraction** (gender of attraction) **w1_identity_all_modified** see also **p17_pppa_lgb** (sexual identity) **w1_outness_all** (LGB outness) **w1_otherdate_all** (did subject meet any new partners in past 12 months [and if partnered, any new partners besides main partner])? **w1_how_many_all** (how many new partners in past 12 months) political movement affiliations from **p17_ppp10206** (tea party affiliation) to **p17_pppa1705** (Black Lives Matter)	**w2_ppethm** (race/ethnicity) **w2_ppincimp** (HH income) **w2_ppreg9** (US region) **w2_ppmsacat** (metro residence) **w2_ppwork** (employment status) **w2_PPT01, w2_PPT25, w2_PPT612, w2_PPT1317, w2_PPT18OV** (# of children in the HH of different ages) **p20_ppp20071** (born-again status) **p20_ppp20072** (religious service attendance) **w2_q17** (times married for subjects who did not answer the Q in wave 1) **w2_relatives** (number of relatives seen each month) **p19_pppa_lgb** (sexual identity) **w2_otherdate_combo** (has subject gone on any dates if single) **w2_how_many** (how many dates for subjects who have been single) **w2_flirt_other** (how often flirt online with others beside partner, or how often flirt with anyone if unpartnered) **p18_ppp10206** (tea party affiliation) to **p18_pppa1705** (Black Lives Matter) and **p19_ppp10206** (tea party affiliation) to **p19_pppa1705** (Black Lives Matter) and **p20_ppp10206** (tea party affiliation) to **p20_pppa1705** (Black Lives Matter)	**w3_ppethm** (race/ethnicity) **w3_ppincimp** (HH income) **w3_ppreg9** (US region) **w3_ppmsacat** (metro residence) **w3_ppwork** (employment status) **w3_PPT01, w3_PPT25, w3_PPT612, w3_PPT1317, w3_PPT18OV** (# of children in the HH of different ages) **w3_relatives** (number of relatives seen each month) **w3_otherdate** (has subject gone on any dates if single) **w3_how_many** (how many dates for subjects who have been unpartnered) **w3_past_partners_gender_1** (number of lifetime female sexual partners) **w3_past_partners_gender_2** (number of lifetime male sexual partners) **w3_past_partners_gender_3** (number of lifetime trans or nonbinary sexual partners)
Basic demography of partners	**w1_q4** (partner gender) **w1_q9** (partner age) **w1_q6a** (partner Hispanicity) **w1_q6b** (partner race) **w1_q10** (partner education)	new partners recorded in w2 only: **w2_Q4** (partner gender) **w2_Q9** (partner age) **w2_Q6a** (partner Hispanicity) **w2_Q6b** (partner race)	new partners recorded in w3 only: **w3_Q4** (partner gender) **w3_Q9** (partner age) **w3_Q6A** (partner Hispanicity) **w3_Q6B** (partner race)

	w1_q11 (partner's mother's education) **w1_q12** (partner's political party ID)	**w2_Q10** (partner education) **w2_Q11** (partner's mother's education) **w2_Q12** (partner's political party ID)	**w3_Q10** (partner education) **w3_Q11** (partner's mother's education) **w3_Q12** (partner's political party ID)
Basic demography of couples (subject and partner). For most timing variables, month is also available. Note: this is not a complete list!	**w1_same_sex_couple** (same-sex ID) **w1_q19** (cohabiting w partner) **w1_q20** (ever lived w partner) **w1_married** **w1_sex_frequency** month as well as year is available for almost all timing variables: **w1_q21a_year** (year first met partner) **w1_q21b_year** (year relationship began) **w1_q21c_year** (year first lived w partner) **w1_q21d_year** (year married partner) **w1_q21e_year** (year broke up w partner) **w1_q21f_year** (year partner died) **w1_q23** (who earned more) **w1_q24*** (how couples met coded from text answers) **w1_q32** (how met closed-ended) **w1_q34** (relationship quality) **w1_who_breakup_combo** (who wanted the breakup) **w1_partnership_status** (relationship status at w1) **w1_relationship_end_nonmar** and **w1_relationship_end_mar** (how past relationships ended) **w1_relate_duration_in2017_years** (relationship duration in years for current relationships)	information for all couples intact at w2: **w2_partner_type** **w2_same_sex_couple** **w2_cohab** **w2_married** **w2_sex_frequency** **w2_rel_qual_combo** (relationship quality) **w2_fight** (how often do you fight w partner) **w2_flirt** (how often flirt online w partner) Information on new couples recorded in w2 (month as well as year is available for almost all timing variables): **w2_new_relationship** (is subject recording a new relationship in w2?) **w2_q21a_year** (year first met) **w2_q21b_year** (year relationship began) **w2_q21c_year** (year first lived w partner) **w2_q21d_year** (year married partner) **w2_q21e_year** (year broke up w partner) **w2_q24*** (how couples met coded from text answers) **w2_Q32** (how met closed-ended) **w2_p_monogamy** (expectations of partner's commitment to monogamy) information on breakups recorded in w2:	information for all couples intact at w3: **w3_partner_type** **w3_same_sex_couple** **w3_live_w_partner** **w3_married** **w3_sex_frequency** **w3_rel_qual** (relationship quality) **w3_fight** (how often do you fight w partner) **w3_flirt** (how often flirt online w partner) Information on new couples recorded in w3 (month as well as year is available for almost all timing variables): **w3_new_relationship** (is subject recording a new relationship in w3?) **w3_Q21A_year** (year first met) **w3_Q21B_year** (year relationship began) **w3_Q21C_year** (year first lived w partner) **w3_Q21D_year** (year married partner) **w3_Q21E_year** (year broke up w partner) **w3_Q23** (earnings comparison with old and new partners) **w3_q24*** (how couples met coded from text answers) **w3_Q32** (how met closed-ended) **w3_p_monogamy** (expectations of partner's commitment to monogamy)

		w2_relationship_end (how relationships ended) **w2_relationship_end_yr** **w2_partner_passaway_yr** **w2_who_breakup_combo** (who wanted the breakup) Other couple information: **w2_Q23_old** (earnings comparison with partner from w1) **w2_Q23_new** (earnings comparison with new partner from w2) **w2_move_in_yr** (timing of first cohabitation for previously identified relationships) **w2_relationship_duration** (relationship duration in years for relationships current at wave 2)	**w3_monogamy** (subject's commitment to monogamy with partner) information on breakups recorded in w3: **w3_relationship_end_combo** (how relationships ended) **w3_breakup year** **w3_partner_passaway_year** **w3_who_breakup** (who wanted the breakup) Other couple information: **w3_partner_source** (for couples in w3, which wave was the partner introduced in- i.e. which wave holds the demographic information about partner?) **w3_breakup_source** (for couples whose breakup was recorded in w3, which wave holds the information about partner?) **w3_move_in_yr** (timing of first cohabitation for previously identified relationships) **w3_when_married_yr** (timing of marriage w previously identified partner) **w3_relationship_duration_yrs** (relationship duration in years for relationships current at wave 3)
COVID-related variables	N/A	**w2_coronavirus_effect_combo** (is relationship better or worse during the pandemic?) **w2cov*** (a series of variables coded from open-ended answers about how the pandemic affected subjects' primary relationships) **w2_pandemic_income_combo** (has income gone up or down during the pandemic?)	**w3_coronavirus_effect_combo** (is relationship better or worse during the pandemic?) **w3cov*** (a series of variables coded from open-ended answers about how the pandemic affected subjects' primary relationships) **w2_pandemic_income** (has income gone up or down during the pandemic?)

		w2_app_use_combo (for single subjects: more or less dating app use during the pandemic) **w2_how_many_corona_combo** (for single subjects: texting with more or fewer potential partners during the pandemic?) **w2_corona_longing_combo** (for single subjects: longing for a relationship more or less than before the pandemic) **w2_corona_effect_dating_combo** (did the pandemic make it easier or harder to meet people?) **w2_shelter_combo_months** (how long sheltering in place)	**w3_app_use** (for single subjects: more or less dating app use during the pandemic) **w3_how_many_corona** (for single subjects: texting with more or fewer potential partners during the pandemic?) **w2_corona_longing** (for single subjects: longing for a relationship more or less than before the pandemic) **w3_coronavirus_effect_dating** (did the pandemic make it easier or harder to meet people?) **w3_shelter_combo_months** (how long sheltering in place) **w3_WFH** (increase in work from home during the pandemic?) **w3_subject_had_COVID** (has subject been sick from COVID-19) **w3_corona_test** (has subject tested positive for COVID-19) **w3_corona_vaccine** (has subject ben vaccinated) **w3_partner_had_COVID** (did partner have COVID) **w3_partcorona_test** (has partner tested positive for COVID) **w3_partner_corona_vaccine** (has partner been vaccinated) **w3_covid_masking** (response to public health restrictions **w3_government_response** (should Gov't mandate vaccination) **w3_COVID_agreement** (do subject and partner agree on approach to the pandemic?)
Open-ended variables coded, codes included in the first release of public data. See rubrics in the documents for how questions were coded.	*q24, how couples meet	*q24, how couples meet	*q24, how couples meet

Open-ended variables coded, where codes will be released to the public after the first release of public data.	*q35, relationship quality *why broke up	*q35, relationship quality *why broke up *COVID effect on relationship	*q35, relationship quality *why broke up *Preference for marriage versus domestic partnership for subjects in same-sex unions *Reasons for nonmonogamy *COVID effect on relationship *COVID effect on dating

Further notes on the HCMST 2017-2022 public data

* In the first public data release of HCMST 2017, "HCMST_2017_public_data_v1_stata.dta", many variables had a "w6_" prefix because HCMST 2017 was fielded at the same time as wave 6 of the HCMST 2009 panel. The "w6_" prefix was misleading because in fact HCMST 2017 was wave 1 of a fresh panel. We have corrected that unfortunate naming decision here: HCMST 2017 variables have the proper "w1_" prefix, HCMST 2020 is properly "w2_" and HCMST 2022 is properly "w3_". There are still a few value labels in the data that refer to "w6," you should interpret those to mean w1, i.e. 2017.

* In this dataset w1=2017; w2=2020; w3=2022.

* Self-identified LGB adults were oversampled in HCMST 2017, and therefore remain oversampled in subsequent waves. The weights correct for this oversample.

* When the user's guide refers to "partnered" individuals, this includes married and nonmarital relationships.

* The user's guide above lists only a small subset of all available variables. For instance, for timing variables month is almost always available as well as year. All variables are listed in the codebook.

* What the "small" version of the data includes: The "small" version of the data includes all answers from each of the 3 survey waves that can be shared with the public at this time. We want data users to avoid problems that can arise when the same survey question yields 7 different variables (6 specific to each survey section, and one summary variable that combines the other 6) in the same wave. The "small" version of the HCMST dataset includes only the summary variables that combine answers across survey sections within each survey wave. The small public data version includes a variable for section (**w1_section, w2_section, w3_section**) so users can figure out which section of the survey each respondent was in for each wave. The same question can have slightly variable wording depending on which survey section the question was in, that is depending on whether subjects were married or in unmarried partnerships, or were single.

* In order to access the exact wording of questions and of answer options, refer to the survey instruments from each wave. And note: survey question wording can vary slightly across survey sections, depending on whether the subject is currently married, in an unmarried partnership, or single. Why don't the variables in the dataset include all the information for exact question wording? There are several reasons. First, most of the variables in the public dataset are combinations (across survey sections) of several variables from the HCMST surveys that might have slightly different wording. Second, STATA has an 80 character limit for variable labels, which is where the text for survey questions would ordinarily go. Third, translation of the dataset from STATA to other formats is not perfect where value and variable labels are concerned.

* Analytic weights have a mean of 1, and account for LGB oversample and are weighted (by Ipsos) to correspond to Current Population Survey values for gender, age, income, region, race, and ethnicity. Wave 2 and wave 3 have attrition-adjusted weights which adjust the analytic weights for attrition since wave 1. Wave 1 has a frequency weight which has a mean value of 69,410. Applying the frequency weight yields a full 2017 US population of adults.

* Background variables from the KnowledgePanel generally have w*_pp* or pp17, pp18, pp19, pp20 (pp20 indicating background variable as of 2020) in their variable names. These variables come from annual KnowledgePanel surveys, and reflect the most recent values prior to each HCMST wave. Question wording comes from separate KnowledgePanel background surveys.

* The w*_xvar* variables carry information from one wave to the next, so that subjects ended up in the right section of the next survey (based on relationship status coming from the previous survey).

* On file formats: The HCMST data were created with Stata software, and translated versions were created with Stat/Transfer to SPSS (.SAV) and R (.Rdata) formats. Refer to the survey instruments for exact survey question wording.

Stata truncates value labels at 80 characters so value labels in the data and in the codebook may not reflect exact question wording if question was long.

* For R users: If you are going to use the R version of the data, you will need to do some work to bring the value labels (which are in the R dataset) into use. We have had success with the “labeled” package, the “to_labeled” function, and tabular output from the tidyverse packages, but your mileage may vary.

* As codes and rubrics are developed for the various open-text answers in the HCMST surveys, the codes and rubrics will be added to the public data. The original text answers will be cleaned of individually identifiable information (such as names and places) and the edited text answers will be deposited with ICPSR and will be available separately as restricted data (requiring a letter from the researcher’s IRB) to access, with **caseid_new** to allow merging back to the rest of the data.