UnderStandingAmericaStudy

DATA GUIDE



USC Dornsife Center for Economic and Social Research

01/30/2019

WELCOME

Thank you for your interest in the UnderStanding America Study (UAS) at the USC Dornsife Center for Economic and Social Research! It is our aim to make the data that we collect available to an as broad audience as possible including (but not limited to) researchers, policy makers, media representatives.

In this guide you will find a short introduction about the UAS. This is followed by detailed information about how you can access UAS data and the types of data that we provide. Should you have any questions or comments after reading the guide, please just feel free to contact us:

BY EMAIL

uas-l@mymaillists.usc.edu

BY PHONE/FAX

213.821.1819 (phone)

213.821.2716 (fax)

BY MAIL

UnderStanding America Study

University of Southern California

PO Box 77902

Los Angeles, CA 90007

WHAT IS THE UNDERSTANDING AMERICA STUDY?

The UnderStanding America Study (UAS) is a probability based panel of American households recruited for their demographic diversity. UAS respondents answer researchers' queries once to twice a month via an online interface that is technologically powerful yet friendly for the respondents and quick to deliver results. At this time the UAS only includes adults aged 18 and over. It is not possible to volunteer, as people around the country are randomly selected using postal codes. The majority of the panel members have their own online access, while those who didn't at the time they joined the panel have been provided, by USC, a tablet and/or an Internet subscription.

The surveys – commissioned by (among others) USC clients, international universities and other government researchers – make inquiries into decision making, social and economic policy and health. The UAS also is used for methodological experiments. Researchers around the world sketch out the surveys; then the UAS team at the Dornsife Center for Economic and Social Research designs them and ensures their rigor through testing before sending them to the panel respondents. More information about the UAS and its respondents can be found <u>here</u>.

HOW TO FIND YOUR WAY AROUND THE UAS DATA PAGES

Access to data collected in the UAS is centralized through the <u>Understanding America Study Data</u> <u>Pages</u> (located at <u>https://uasdata.usc.edu</u>). On these pages you will find generic information about the UAS as well as documentation on a variety of topics. You can also view individual surveys and the questions those surveys asked, search for a particular topic, look at overall survey response rates and timings or download data files.

Navigation on the UAS Data Pages is provided through the main menu located at the top of the page:



The site is divided into the following main sub sections:

- About: glean overall information about the panel, the data we provide, the technology we use, the team behind the UnderStanding America Study, and delve deeper with access to UAS resources (such as this data guide).
- Methodology: gain insight into the UnderStanding America Study sample and its recruitment and attrition, information on the weighing of UAS data to match against the US population, and the standard variables we collect and disseminate.
- Surveys: learn about the surveys conducted in the UAS and the possibility to view quick aggregate information about the data or to download data sets
- Special datasets: explore specific topics such the US Election 2016 surveys, the US Midterm Election 2018 surveys, the UAS comprehensive file and the UAS Visualization Toolkit.
- Publications: see what has been written using UAS data. If you have a paper that is missing from the list, please contact us at <u>uas-l@mymaillists.usc.edu</u>
- Contact us: explore our services and pricing, get quotes, find answers to commonly asked questions and visit ours newsletter archive.
- Search: search the UAS Data Pages.

• Login/register: login to the UAS Data Pages or register for an account.

In what follows, we will focus foremost on the Data section of the UAS Data Pages. Note that registration is required to obtain UAS data. The registration process consists of two simple steps:

- Fill out the registration form at <u>https://uasdata.usc.edu/page/Register</u>. We ask some basic information such as your name, address as well as for your affiliation. This information is used solely for processing your registration request and maintaining your account. It is not shared with anyone.
- Download, print, sign, scan and return the UnderStanding America Study Data User agreement.
- Wait for your registration confirmation. Once your request has been processed we will provide you with user credentials with which you will be able to access all the functionality of the UAS Data Pages.

EXPLORING THE DATA

The data that has been collected in the UAS can be explored in several ways:

- o Interactively through a drill down mechanism to look at individual questions
- o By downloading a variety of data sets
- Through the search function (potentially in combination with the question cart)

In addition, a variety of so-called meta-data (i.e. data about the data) is available.

DRILL DOWN MECHANISM

Firstly, the UAS Data Pages provide a drill down mechanism to navigate all surveys and individual questions. The main entrance point for this mechanism can be accessed by clicking on 'Data' in the main menu. This will open up a listing of UAS surveys (small subset shown):



Each survey is listed with its name, a short description, whether the data is under embargo (i.e. accessible for download or not), the number of observations, the survey's field dates, and its status (in the field or completed). The My Household survey is special in the sense that it is always in the field and its number of observations is not explicitly listed.

To drill down to a specific survey, simply click on its name. For example, opening up the UAS 1 survey page provides you with the following screen (note: this is the screen shown when you are logged on and the data is not under embargo). Unregistered users do not have access to the data and paradata sections):

FINANCIAL LITERACY; PERSONALITY; UNDERSTANDING PROBABILITIES; NUMERACY		UAS 1 - RESPONSE OVERVIEW	#
Solver Torica. Obginerie Ab	incles / Financial Electracy / Fsychology	Size of selected sample	7171
SECTION	DESCRIPTION	Completed the survey	6870
base	Base questions	Started but did not complete the survey	109
demographics	Demographics	Did not start the survey	192
lipkus	Lipkus	Response rate	95.80%
admc	ADMC Consistency		
bigfive	Big Five	TIMING DISTRIBUTION	
financialliteracy	Financial Literacy	LIAS1 - Timing distribution	n =
closing	Closing questions	Source: UAS	_
METADATA Survey codebook Cognitive reflection vs. ca UAS 1 Scoring Big Five Personality Scorie DATA DATA Non response data PARADATA INFORMATION Timings (CSV)	alculation in decision making ng Chart Data Data Data Data Platform data (STATA)	1000 500 500 7 ⁻⁵ 50 1 ⁻¹⁰ 1	ris rin
	Platform data (CSV)		

As can be seen from the figure there is a wealth of information available about each individual UAS survey. The survey page is divided in multiple parts. First, towards the top of the screen the survey's title, a short description and a list of topics associated with the survey appear:

FINANCIAL LITERACY; PERSONALITY; UNDERSTANDING PROBABILITIES; NUMERACY SURVEY TOPICS: Cognitive Abilities / Financial Literacy / Psychology

Just below that is the list of the survey's sections (in this context a section is no more than a grouping of logically related questions, e.g. demographics, financial literacy, and etceteras).

SECTION	DESCRIPTION
base	Base questions
demographics	Demographics
lipkus	Lipkus
admc	ADMC Consistency
bigfive	Big Five
financialliteracy	Financial Literacy
closing	Closing questions

There are two sections that are present in every UAS survey: the base section and the demographics section. The base section contains the respondent identifier, the time each respondent began and ended the survey, and the survey language (English or Spanish). The demographics section groups together all the standard variables collected for every respondent through the My Household survey. The records for these variables are added to the survey data at the time a respondent starts a survey, i.e. they reflect the respondent demographics at that point in time. Detailed information about these basic demographic variables can be found online here as well as in the UAS Standard Variables guide (available for download in the Documentation and Papers section of the UAS Data Pages).

Zooming in on an individual section shows an overview of the available variables:

LIPKUS

VARIABLE	QUESTION TEXT	LABEL	ACTIONS
lip001	Imagine that we roll a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up as an even number?	number of times dice even	芹
lip002	In the BIG BUCKS LOTTERY, the chances of winning a \$10.00 prize are 1%. What is your best guess about how many people would win a \$10.00 prize if 1,000 people each buy a single ticket from BIG BUCKS?	number of people winning lottery	Ä
lipoog	In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?	percent of tickets win car	Ä
lip004	If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?	time to drink one barrel	Ϋ́
lip005	Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class?	how many students in class	Ä
lip006	A man buys a pig for \$60, sells it for \$70, buys it back for \$80, and sells it finally for \$90. How much has he made?	how much made	Ä
lip007	Simon decided to invest \$8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has:	stocks ahead	Ä

Each variable is listed with its name, question text and label. More details can be seen by clicking on a variable name, such as answer options, whether respondents were allowed to leave the question unanswered (empty allowed) and whether the variable is part of an array (e.g. asking the gender of multiple household members where all variables are named 'hhmembergender' followed by an identifier to distinguish between members).

If you have registered and are logged on to the UAS Data Pages you will also see a tabular and graphical representation. These are based on data sets that are generated nightly (i.e. they are not real time). For "select one" and "select all that apply" variables the table and graph are built by utilizing their answer options. For numeric variables appropriate brackets are automatically determined. For variables containing textual answers no tabular or graphical representation is provided.



GENERAL INFORMATION

Question text:	Which of the following statements describes the main function of the stock market?
Answer type:	Radio buttons
Answer options:	1 The stock market helps to predict stock earnings 2 The stock market results in an increase in the price of stocks 3 The stock market brings people who want to buy stocks together with those who want to sell stocks 4 None of the above 5 I don't know
Label:	stock market
Empty allowed:	One-time warning
Error allowed:	Not allowed
Multiple instances:	No



stock market

RESPONDENTS DATA

A second option to get access to UAS data is by downloading available data sets. The download section of a survey is located below the section listing (assuming you are logged on and the data is not under embargo):

METADATA	
🚣 Survey codebook	
La Cognitive reflection vs. cal	culation in decision making
LAS 1 Scoring	
🛓 Big Five Personality Scoring	g Chart
DATA	
🖽 Data	Data
🖶 Non response data	Non response data
DADADATA INFORMATION	
진 Timings (CSV)	🖽 Platform data (STATA)
	Platform data (CSV)

Respondent data sets (located under the 'Data' header) contain the data collected in the survey. On the site this data is provided either in STATA or CSV format. Should you require a different format, then please just let us know. Note that STATA files are provided in STATA 13 format. If you are using STATA 12 or earlier version, you can download the user-written command use13 (ssc install use13) to load the files.

All survey data files come with a set of <u>preloaded demographics</u> for each respondent. These variables represent the basic respondents' characteristics at the time they started the survey (as collected in the <u>My Household survey</u>). Note that for some variables their underlying raw representations differ from those in the downloaded data. Details on the mapping from the cleaned up data variables to the raw data can be found <u>here</u>.

Weights are available for completed surveys and are part of the data file. Weights for a survey that is ongoing are by request (<u>uas-weights-l@mymaillists.usc.edu</u>). The following weight variables are included (detailed information on the UAS weighing procedures can be found <u>here</u>):

- **Base_weight**: Relative base weights correcting for the over-representation of Native Americans in the survey sample. They average to one and sum to the UAS survey sample size.
- **Rel_weight**: Relative post-stratification weights which ensure representativeness of the survey sample with respect to key selected variables (raking factors). They include the correction for the over-representation of Native Americans. They average to one and sum to the UAS survey sample size.
- Imputation_flag: A binary variable indicating whether any of the variables used within the weighting procedure has been imputed.

Survey data files provide "clean" data, that is, answers given to questions that are not applicable anymore at survey completion (for example because a respondent went back in the survey and skipped over a previously answered question) are treated as if the questions were never asked. In the data files all questions that were asked, but not answered by the respondent are marked with ".e". All questions never seen by the respondent (or any dirty data) are marked with ".a". The latter may mean that a respondent did not view the question because s/he skipped over it; or alternatively that s/he never reached that question in the survey due to a survey break off.

In terms of the formatting of the survey data sets (as well as non-respondents data sets discussed shortly) the following applies:

- In STATA all questions come with short descriptions.
- 'Please select one' questions come with value labels for each answer option. In STATA these labels will include the labels "Not asked" and "Not answered" for ".a" and ".e", and will show in tabulations such as "tab q1, missing".
- For 'select all that apply' questions a binary variable is created for each answer option indicating whether the option was selected or not. A summary variable is also provided in the format "1-3-2" reflecting which options were selected and in which order. For example, if a question asked about favorite animals with options cat, dog, and horse, then if a respondent selected horse and then cat, the binary variables for horse and cat will be set to yes, while the overall variable would have a string value of 3-1. If no answer was given, all binary variables and the summary will be marked with ".e".

- Questions that are asked multiple times are often implemented as so-called array questions. Supposing the name of such question was Q1 and it was asked in 6 different instances, your data set would contain the variables Q1_1_ to Q1_6_. To illustrate, if a survey asked the names of all children, then child_1_ would contain the name of the first child the respondent names and so on.
- If a survey uses randomization, please be sure to consult the codebook. It will indicate whether you will need to de-randomize your data or not; and if so, in what way. Should you have further questions please just contact us.

NONRESPONDENTS DATA

In addition to the survey data files we also provide a non-respondents data set for each UAS survey. This data contains the basic demographics of respondents (as collected in the <u>My</u> <u>Household survey</u>) who did not start the survey (in STATA or CSV format). Like survey data sets, non-response data files are cleaned to match the definitions of the standard variables. The demographics used for each non-respondent are the most recent available with respect to the time the panel member was invited for the survey. It should be noted that in a few instances only a respondent identifier is available in a non-respondents data set. This is the case when the respondent was invited to the survey, but had not taken the My Household survey prior to being invited.

TIMINGS DATA

To learn more about the amount of time respondents spent on each UAS survey a few options are available.

First, in the respondent data set the start and end time of each interview is marked. These can be used as rough approximation of how long respondents took to complete the survey. Note that if all the end time related variables are empty, this means that the respondent never completed the survey. This method has some limitations in the sense that it does not account for any prolonged periods of time that a respondent discontinued the survey before returning to it. To obtain more detailed and accurate timing information, we recommend using the timings data file. This file contains the following for each survey:

- Listing of the time spent per respondent for the interview as a whole (NOTE: any time spent on a screen longer than 5 minutes is excluded from this listing):
 - Uasid of the respondent.
 - Total time spent in seconds.
 - Total time spent in minutes.
 - Number of screens viewed by the respondent.
 - Average time spent per screen in seconds.
- Listing of the time spent per question by all respondents combined (NOTE: any time spent on a screen longer than 5 minutes is excluded from this listing):
 - o Total number of times.
 - Average time spent in seconds.
 - Average time spent in minutes.
- Listing of each period of time spent per question by each individual respondent (NOTE: any time spent on a screen longer than 5 minutes is included in this listing. As such, the number of times a question is listed here may slightly differ from the total number of times a question is reported to have been shown in the listing of the time spent per question by all respondents combined):
 - o Uasid of the respondent.
 - Time spent in seconds.

PLATFORM INFORMATION

Another source of data available for analysis is the platform information file. Platform information files contain information about the type of device, operating system and browser used by survey respondents. The file comes with the following variables:

- Uasid of the respondent.
- Device used by the respondent, divided into desktop/laptop, mobile or tablet.
- Operating system used by the respondent.
- Browser used by the respondent

Please note that these are based on the user agent strings as reported by respondent's browsers. Such browser-reported user agent strings are known to be limited in their accuracy, and as such, the provided platform information should be treated as indicative in nature (rather than as absolute). More information on browser user agent strings and their limitations can be found <u>here</u>.

META INFORMATION

In addition to the previous types of data provided for download, a survey page also give access to a variety of online data as well as meta-data (that is, information about the survey instrument):

- Response rate informing about the gross sample size, the number of members who started the survey, the number of members who completed the survey, and the number of members who did not started the survey. The response rate is generated in real time and slight discrepancies may exist with the survey data set (which is generated nightly).
- Graphical illustration of the time spent by respondents on the survey.
- Survey information including survey author(s), sample selection, respondent compensation, field dates, and average time spent in minute.
- Codebook (PDF) describing the survey questions and skip patterns. If a codebook is not available yet, a link to view the routing in the browser is available instead. This link can be found under 'Metadata' in the data section of a survey page just below the section listing:

METADATA

Survey codebook

- 🛓 UAS 1 Scoring
- 🛓 Big Five Personality Scoring Chart

In addition to the above, additional documentation may be provided. If this is the case, this will be located underneath the 'Supporting files' header. Lastly, the following are available upon request:

- o Item non response summaries.
- Custom extracts from the survey logs; for example, how often respondents went back and forth for a particular question.

SEARCH FUNCTION AND QUESTION CART

A third way in which you can explore the UAS data is through the search function. Located in the far right of the top menu, the keyword based search allows to find, for instance, variables by name:

UNDERSTA	NDING AMERICA STUDY » SEARCH
/ :	SEARCH
CRITERIA	
Look for	
Look in	All surveys ×
	☑ Papers
	Veb site
SEAR	СН

The search can be combined with the data cart. This mechanism allows to add selected variables to be added to a cart:

UNDERSTAND	VERSTANDING AMERICA STUDY » SEARCH RESULTS SEARCH AGAIN VEY VARIABLES urvey Variable Question text Label			
SEARC	H AGAIN			
SURVEY VARIA	BLES			
Survey	Variable	Question text	Label	Actions
My Household	gender	What is your gender?	Gender - Male	Ϋ́

To add variables to the cart just click on the 'Cart' image next to a variable. To then access your cart, click 'Data cart' in the 'Data' menu. In the data cart screen you can then select to download data for variables in the cart:

HOME » MY DATA CART						
O O CURRENT CART LOAD CART						
UNSAVEL	UNSAVED CART					
	Variable	Survey	Question	Label	Data	
Ä	gender	My Household	What is your gender?	Gender - Male	\oslash	
EMF	PTY CART	SAVE	DOWNLOAD A	S CSV		

The resulting data set will contain the individual identifier and the variables in the cart. Note that in order to be able to export variables with the same name from different surveys through the question cart, outputted variables are prefixed with the survey name. For example, 'D001' in the screenshot above would be outputted as 'uas1_d001'.

COMBINING DATASETS

One of the benefits of using data from the UnderStanding America Study is that you can combine data from multiple data sets to obtain a rich contextual background for your research.

Merging multiple UAS data sets into one is done based on the "uasid" identifier variable, which is present in every data set. To merge two data sets, you should follow these steps (for STATA):

- Open the first data set, e.g. UAS1.dta and rename the variables to ensure uniqueness, e.g. by prefixing them with "rename * prefix_=". Make sure that the "uasid" variable remains named "uasid".
- Do the same for the second data set, for example UAS24.dta.
- Merge the two data sets with "merge 1:1 (uasid) using "uas24.dta".
- o (optional and depending on your analysis purposes) Keep the observations that are common to the two data sets using "keep if _merge == 3"
- Drop the merge variable using "drop _merge".

Should you require any assistance, please do not hesitate to contact us. We are more than happy to answer any questions you might have.

RESPONDENT LEVEL VERSUS HOUSEHOLD LEVEL ANALYSIS

The UAS is a panel of households, and as such, a sub set of its respondents is part of the same household. Because of this, UAS data can be considered for:

- Respondent level analysis: in this type of analysis the sample unit of interest is a single respondent. For example, your outcome of interest is individual health. Clearly, since respondents belonging to the same households may share or influence each other preferences and habits, appropriate econometric techniques should be used to take into account potential within-household interdependencies.
- Household level analysis: in this type of analysis the sample unit of interest is a household.
 For example, your outcomes of interest are income and expenditure at the household level. Since surveys are administered to individual respondents, not to households, household-level data may be gathered from multiple respondents within the household. In this cases, particular attention should be paid to treat potential discrepancies between information reported by different household members.

The presence of multiple respondents within a household also makes it possible to study intra-household decisions and how the relative bargaining power of household members influences resource allocation within the household.

The relevant variables to identify respondents within a household are:

- primary_respondent: indicates whether the respondent was the first person within the household to become a member or added subsequently. A household in this regard is broadly defined as anyone living together with the primary respondent. That is, a household comprises individuals who live together, e.g. as part of a family relationship (like a spouse/child/parent) or in context of some other relationship (like a roommate or tenant). This variable will have a value of 1 for all primary respondents including those who are the only member of their household to also be an UAS respondent.
- **uasmembers**: is the number of other household members who are also UAS panel members at the time of the survey. Since individuals can answer the same survey at different points in time (which can be relatively far apart if the survey is kept in the field for a prolonged time), it may be possible that, within the same data set, the primary respondent of a household has a value of "0", whereas the second UAS household

respondent has a value of "1". Therefore uasmembers should be interpreted as the number of household and UAS panel members at the time the respondent answers the survey. Note: in the My Household survey uasmembers is set to unknown (".u") for respondents who last participated in the My Household survey prior to January 21, 2015.

- **uashhid**: is the household identifier of the respondent. Every member is assigned a household identifier, stored in the variable uashhid. For the primary respondent this identifier is his or her 'uasid'. All other eligible members of the primary respondent's household (everyone who is 18 or older in the household) who become UAS respondents receive the uasid of the primary respondent as their household identifier. The identifier uashhid remains constant over time for all respondents. Thus it is always possible to find the original UAS household of an UAS panel member (even after they, for example, have moved out to form another household).
- o **survhhid**: uniquely identifies the household a UAS panel member belongs to in a given survey. For instance, if the primary respondent and his/her spouse are both UAS members at the time of a given survey, they both receive the same survhhid identifier for that survey. If they subsequently split, they receive two different survhhid in subsequent surveys. They, however, always share the same uashhid. The identifier 'survhhid' is set to missing (".n") if no other household members are UAS panel members at the time of the survey. Since individuals can answer the same survey at different points in time (which can be relatively far apart if the survey is kept in the field for a prolonged time), it may be possible that, within the same data set, household members have different survhhid reflecting different household compositions at the time they answered the survey. For instance, suppose that the primary respondent and his/her spouse are both UAS members. If the primary respondent answers the survey when he/she is living with the spouse, but the spouse answers the survey when the couple has split, they receive different survhhid. Hence, the variable survhhid identifies household membership of UAS panel members, at the time the respondent answers the survey. Note: in the My Household survey survhhid is set to unknown (".u") for respondents who last participated in the My Household survey prior to January 21, 2015.

Now, how can the above variables be utilized for the two types of analysis we sketched?

Starting with respondent level analysis, we noted that it may be the case that the fact that respondents are part of the same household in some way affects analysis results. In order to assess

whether this is the case, an option would be to control for this by including a 'multi-household flag' as a control variable. Constructing such flag variable can be done by looking at for example the "survhhid" variable. Any respondent who has a value for "survhhid" was part of a household containing at least one other UAS respondent at the time the survey was started. Another option is to control for number of household members in which case "uasmembers" can be utilized as a control variable. Note: using the "uashhid" variable as a control variable is discouraged for respondent level analysis as it remains attached to a respondent even if at some point in time there are no other UAS respondents in their household.

Household level analysis can also leverage the described variables. For example, suppose you are using data on household income. To find out which respondents were in the same household at the time of the survey, you can use the aforementioned "survhhid" variable or the "uashhid" variable. Any respondents with the same "survhhid" were in the same household at the time of the survey. You can then decide on how to treat the multiple reports. One option would be to only consider primary respondents, that is, the respondents who were the first member of their household to sign up. Another could be to use demographic information to decide.

Now, it may be possible that a second respondent signed up with the panel just after the first respondent had completed the survey. As a result, the first and second respondent would not have the same "survhhid", but their answers would be describing e.g. household income around the same time. In such situations discrepancies may be valid if the second respondent themselves just moved in and also added income to the household. However, if the person was part of the household already and merely became an UAS member, then such discrepancies might be indicative of errors. To locate such potential cases you can use the "uashhid" to determine which respondents are part of the same UAS household (independent of when they became UAS members).

CREDITING THE UAS

The UAS data provided through the UAS Data Pages is free of charge. All we ask in return is that you cite usage of our data. For this purpose we recommend the following text:

"The project described in this paper relies on data from survey(s) administered by the UnderStanding America Study, which is maintained by the Center for Economic and Social Research (CESR) at the University of Southern California. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of USC or UAS."

Also, we would love to hear from you if you presented or published work using UAS data, so we may include it in our list of papers on the UAS Data Pages web site.