



UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

April 30, 2014

2014 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT
MEMORANDUM SERIES #ACS14-RER-18

MEMORANDUM FOR ACS Research and Evaluation Advisory Group

From: James B. Treat (**signed on 05/12/2014**)
Chief, American Community Survey Office

Prepared by: Samantha M. Fish
American Community Survey Office

Subject: Evaluation of the 2013 Automated Response Cleanup for
American Community Survey Internet Open Response Data

Attached is the final American Community Survey (ACS) Research and Evaluation report, "Evaluation of the 2013 Automated Response Cleanup for American Community Survey Internet Open Response Data". We conducted this evaluation to assess the Internet open responses that the Automated Response Cleanup (ARC) changed during the first four months of 2013, which helped ACS managers determine if the survey should enhance the current ARC program. If you have any questions about this report, please contact Samantha Fish at 301-763-7542.

Attachment

cc:
ACS Research and Evaluation Workgroup
Gail Denby (ACSO)
Dameka Reese (ACSO)
Colleen Hughes (POP)
Rachel Marks (POP)
Nicole Scanniello (SEHSD)

APRIL 24, 2014

Evaluation of the 2013 Automated Response Cleanup for American Community Survey Internet Open Response Data

FINAL REPORT

Table of Contents

INTRODUCTION	2
BACKGROUND	2
RESEARCH QUESTIONS	6
METHODOLOGY	6
RESULTS	9
1. HOW OFTEN ARE INTERNET OPEN RESPONSES CHANGED BY THE ARC?	9
2. WHAT ARE THE MAIN REASONS THAT THE ARC CHANGED RESPONSES?	10
3. WHAT ARE SOME EXAMPLES OF RESPONSES ARC CHANGED THAT WE CATEGORIZE AS “SPECIAL CHARACTER ONLY” AND “OTHER”? HOW DID ARC HANDLE THEM?	15
CONCLUSIONS	19
NEXT STEPS	20
APPENDIX A	21
APPENDIX B	24
APPENDIX C	29
APPENDIX D	35

INTRODUCTION

Beginning in 2013, the American Community Survey (ACS) collected data from respondents via the Internet as part of its self-response data collection. The ACS designed its Internet instrument to work with any electronic or mobile device, which came with the tradeoff that the instrument could not limit the types of characters respondents entered into open item survey questions.¹ To limit incoming characters that ACS sends to downstream data processing, management created the Automated Response Cleanup (ARC) program to blank responses with insufficient data and perform other cleanup that the keyers of the paper questionnaire would normally handle at the National Processing Center. ARC also removes invalid characters from and properly formats Internet open responses.

The programming of ARC was subject to the same testing standards used in the development of the 2013 Internet instrument. Teams from various divisions within the Census Bureau planned, created, and reviewed the ARC program. The testing confirmed that the ARC was programmed according to its specifications; however, it did not confirm that the cleanup rules addressed all possible inputs or that that interaction of individuals rules produced output that the analysts and researchers had intended.

Team members expected that ARC would change entries for very specific reasons (like monetary items containing commas or decimals), but also counted on the program to blank any inputs that the ARC did specifically address in its cleanup rules. Members agreed that they would look at the data coming out of ARC after it was in production to make any enhancements to the program.

This report assesses the total responses that ARC changed during the first four months of 2013 to help determine if the ACS should consider enhancing its current rules. We quantify the input patterns of the responses that the ARC program specifically addressed (like commas or decimals in monetary field) in addition to responses that the program blanked because they were either uninformative/insufficient or the program did not have specific instructions to salvage the data otherwise.

BACKGROUND

What is the American Community Survey?

The ACS is a continuous survey that measures population and housing characteristics of large and small areas across the United States. The ACS eliminated the need for the decennial census long form and instead provides detailed demographic, social, economic, and housing data every year based on a sample. The survey produces data that determine how federal and state governments distribute their funds each year. The survey also gives communities current information, which many use for planning investments and services.

The Census Bureau has collected ACS data since 2000, although it started out as a test survey called the Census 2000 Supplemental Survey. Since then, the ACS has expanded its annual sample to

¹ The ACS accepts data from respondents that do not have Java enabled on their Internet browsers. Since respondents must have Java on their devices in order to limit the incoming characters, the Internet instrument could not limit incoming characters and post-collection processing needed a way to cleanup open responses from the Internet mode.

roughly 3.54 million addresses. This sample size ensures the adequate geographic coverage needed in order to produce quality estimates for the smallest geographic level, block groups.

To see the questions asked in the 2013 ACS, please visit the webpage, “Questions on the form and why we ask”, www.census.gov/acs/www/about_the_survey/questions_and_why_we_ask/.

Data Collection in the ACS

The ACS divides its annual sample into 12 monthly panels for data collection. The panels consist of three sequential phases of data collection, each lasting about one month, and occur in the following order: Self-Response, Telephone, and Personal Visit.² In the January 2013 panel, for example, the ACS collected self-responses in the calendar month of January 2013 with telephone responses collected in February 2013 and personal visit interviews collected in March 2013, as shown in Table 1. A new panel begins each calendar month so that in any month of the year all three phases run simultaneously.

Table 1. Phases of Data Collection by Panel

CALENDAR MONTH	PANEL			
	January 2013	February 2013	March 2013	April 2013
January 2013	Self-Response			
February 2013	Telephone Self-Response (cont'd)	Self-Response		
March 2013	Personal Visit Self-Response (cont'd)	Telephone Self-Response (cont'd)	Self-Response	
April 2013		Personal Visit Self-Response (cont'd)	Telephone Self-Response (cont'd)	Self-Response
May 2013			Personal Visit Self-Response (cont'd)	Telephone Self-Response (cont'd)

Prior to 2013, self-response data collection included the just the mail (i.e. paper questionnaire) and telephone questionnaire assistance (TQA) operations. Beginning in 2013, the ACS added an Internet response option to its self-response materials. Instead of sending a paper questionnaire in its first mailing, the Census Bureau now mails instructions for completing the ACS online to all sample addresses with a mailable address. After about two weeks, if a sample address does not respond online then the Census Bureau sends the address a paper questionnaire. At any time respondents have the option to reply by TQA.

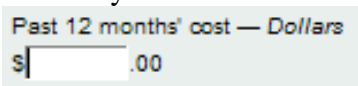
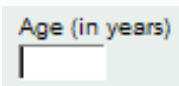
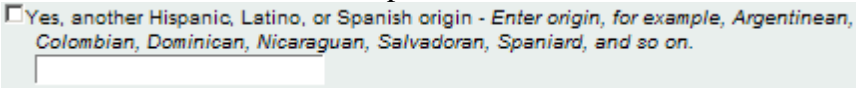
² Late self-response returns are accepted throughout all data collection modes.

For more detail on the data collection operations or topics not covered in this report, please consult the ACS Design and Methodology Handbook. The following web address houses the current version: www.census.gov/acs/www/methodology/methodology_main.³

Open Response Questions

ACS Internet questions require mostly checkbox responses; however, some questions require “fill-in-the-blank” responses. We call questions like this open response items. The 2013 ACS Internet response option contains 83 open response questions: 22 items require a monetary response, 26 items require an “other numeric” response (i.e. non-monetary), and 35 items require an alphabetic response. Figure 1 shows examples of the Internet instrument’s answer boxes for these kinds of questions.

Figure 2. Internet Open Response Box Examples

<p>Monetary:</p> 	<p>Other Numeric (Non-Monetary):</p> 
<p>Alphabetic:</p> 	

Source: 2013 American Community Survey Internet Instrument

Answer boxes requiring an alphabetic and other numeric response are plain blank boxes that allow for a certain number of characters or digits depending upon the item. Monetary answer boxes are similar in that they accept a fixed number of digits but they also include a dollar sign in front of the answer box and a “.00” at the end. The ACS made this distinction in an attempt to indicate to the respondent that they should enter a whole dollar amount for monetary items. Additionally, the ACS positions all open response fields on a green background, as on the paper questionnaire to draw the respondent’s attention to the answer box.

³ As of the date of this publication, the ACS had not yet updated the handbook with details on the Internet response option. Please see the background section of the ACS R&E Assessment of Multiple Responses and Internet Returns that Identify as Vacant Units ([Fish, 2013](#)) until it is updated.

Internet Data Processing

The ARC cleans and formats most Internet data for open response items before sending them on to subsequent processing. Only the alphabetic responses go to coding, which is an operation with human intervention for responses that cannot be auto-coded. The clerks review alphabetic responses and give them a code. The other numeric and monetary Internet write-ins, however, do not receive content checks by a person and pass through ARC for cleanup.

In general, the ARC program removes special characters that subsequent processing operations cannot handle, rounds other numeric and monetary responses that contain fractions, blanks entries with insufficient information, and formats the data as necessary. For example, a response to an income item of “1.2,345” has insufficient information because we do not know what the respondent intended. One of the purposes of this report is to highlight these situations, so we can understand how respondents interact with the Internet open response fields and enhance the ARC program as necessary.

Basis of ARC Specification

The ACS based the ARC program off the 2012 mail mode’s keying rules for open response items. In mail questionnaire processing, keyers at the National Processing Center electronically view and key the handwritten responses from questionnaires into a database. Keyers are also responsible for “cleaning up” open responses. For example, if a respondent writes “\$70,000” for his or her income on the paper form, the keyers must enter “70000”, without any special characters or punctuation. ARC intends to imitate this process automatically by applying the NPC keyer rules to the incoming Internet data. Likewise, a response of “\$70,000” to an income item in the Internet instrument should become “70000” after ARC processing.

Researchers and subject-matter representatives worked with the American Community Survey Office to come up with an ARC specification that closely resembled the 2012 NPC keying instructions. They formed an ARC Team, comprised of several representatives from each of these areas, to carry out the task. The ARC Team proposed a specification, which they sent out to the subject-matter experts to review within their individual areas of expertise. The analysts provided feedback on the draft and developed additional rules for their items of interest. Although there were a few general guidelines for Internet cleanup across multiple items, the ARC Team developed the specification primarily using an item-by-item approach. This caused some rules to be inconsistent across similar types of items. The team delivered its original specification and completed testing of the ARC program by December 2012.

Changes to the Original ARC Program

Since drafting this report in early 2013, the ACS identified several issues with the ARC specification and/or program. Many issues resulted from rule interactions, and the ARC program was restructured with this in mind. For example, the original ARC program changed open responses to monetary items that contained all alphanumerics (like “Don’t know” for an income item) by changing any letter “o” to zero

and then checking for all alphabetic characters instead of checking for all alphabetic characters first and then checking for the letter “o” used as a zero. The order in which ARC executes these rules can make a difference (e.g. “Don’t know” became “D0n’tKn0” instead of blank). The updates made to ARC in summer 2013 addressed issues like these as well as any programming errors that the December 2012 testing did not catch.

Because these errors were more than just enhancements, managers implemented a list of changes to the ARC program on August 19, 2013. These changes, although minor on paper, required the ACS to re-structuring how it wrote the ARC program. The ARC Team got together to approve the new changes and verify testing output of the rewritten program. These modifications affect all ACS Internet data collected after August 19, 2013. The ACS did not correct the data it collected prior to this date due to the operational difficulties that doing so would involve. Please see [Appendix B](#) for the list of specific changes made to the ARC program.

In the ARC’s next update, the ACS should consider additional enhancements that members of the ARC team suggested after looking at the output data. Although this report looks at responses with inputs patterns of “special character only” or “other”, team members also found room for improvement in how ARC handles inputs involving other patterns (see [Table 3](#) for the full list we used). In the future, the ARC Team will need regularly consider enhancements to ensure that the ACS properly cleans up and salvages Internet open response data. The need to enhance ARC will exist as long as the devices people use to complete the Internet survey change or the ACS finds a different way to limit the incoming characters that respondents enter into open response fields in the Internet instrument.

RESEARCH QUESTIONS

This report answers the following questions regarding the original ARC program:

1. How often are Internet open responses changed by the ARC?
2. What are the main reasons that the ARC changed responses?
3. What are some examples of responses ARC changed that we categorize as “special character only” and “other”? How did ARC handle them?

METHODOLOGY

This study analyzes all Internet open responses collected from December 17, 2012 to April 30, 2013 that came from sample addresses in the January through April 2013 panels. We look at Internet responses regardless of their Internet outcome code (i.e. respondent’s progress through the Internet instrument). We analyze every non-blank, Internet open response as each existed before and after ARC processing.

We categorize open response items based on whether they primarily require a monetary, other numeric, or alphabetic response. Further dividing the items into sub-categories help us to evaluate

whether ARC rules are consistent among similar items. The following list shows the sub-categorizations we use for this report:

Monetary

- Annual Income
- Annual Expense
- Monthly Expense
- Other Value

Other Numeric

- Quantity
- Year
- Hour/Minute
- Zip Code
- Education
- Month
- Telephone number

Alphabetic

- Origin
- Address & Location
- First/Last Names & Middle Initial
- Miscellaneous

For a full list of the 83 Internet open response items we study, please see [Appendix A](#). This appendix also lists the item’s abbreviated names (along side their full descriptions), which we use throughout the remainder of the report.

When the ARC program changes an open response item, it does not flag the reason(s) why. For this evaluation, we created a program that identified patterns among Internet responses in order to infer why the ARC made each modification it did. For example, we concluded that the ARC likely changed the response of “3,000” to interest income to a response of “3000” due to the comma. The input patterns we identified are mutually exclusive and listed in [Table 3](#) by the type of response required for the open response field. The input patterns we identified vary by item type; patterns not applicable are shaded gray. We estimate how often the ARC modified responses for a particular reason by summing the instances of these patterns for each item.

We excluded two types of ARC modifications from [Table 3](#). The first is “zero filling”, which is when the program fills the input with zeroes in front of the numeric response (e.g. “3” changed to “003”). The ACS zero-fills the mail mode data, and this is more of a formatting issue than a change to the response, so we exclude it from our summary of ARC changes. The second modification we exclude in our report is blank entries that the ARC program incorrectly changed to a non-blank response. For example, some blank responses were changed to “.” dot. We had not anticipated these changes, and ACS managers promptly corrected this error for future panels. However, due to the excessive resources required to fix the error for early panels, the ACS left the first few months of data untouched.

We report all statistics unweighted to represent the actual number of Internet responses changed by the ARC per item. Our unweighted figures do not account for ACS sample selection, and thus they do not necessarily portray the characteristics of all Internet mode respondents. Unweighted rates act as a guideline for the proportions of changed responses the ACS can expect in production for the near future.

Table 3. Input Patterns Tabulated for ARC Evaluation

Pattern	Monetary	Other Numeric	Alphabetic
Commas only	Commas before any set of three consecutive digits and/or at the end of a response with no other alphabetic or symbolic characters (e.g. “30,000” or “3,”).		One or more commas and no other non-alphabetic characters (e.g. “Spanish, French, and Russian”).
Decimals only	Decimal with zero, one, or two digits following it and all other characters are numeric (e.g. “450.13”, “450.1”, or “450.”)		
Question mark only	One or more question marks as the only characters or with other numeric digits (e.g. “30?” or “???”).		One or more question marks and no other non-alphabetic characters (e.g. “???” or “sometimes?”).
Leading Dollar Sign only	First character is a dollar sign and all the remaining characters are numeric (e.g. “\$1”).		
Ampersand only			Contains one or more ampersands with no other special characters (e.g. “French & Spanish”).
Forward slash only			Contains one or more forward slashes with all alphabetic (e.g. “French/Irish”).
Main Characters only	Contains a leading dollar sign with any combination of commas and decimals along with all numerics (e.g. “\$150,00” or “\$150,000.”).		Contains any combination of commas, question marks, ampersands, or forward slashes along with all alphabetic (e.g. “Half French, Half Irish”).
All Alphabetic	Contains alphabetic characters only (e.g. “same”).		
Postfix only	Contains numeric characters followed by one or more alphabetic characters (e.g. “2m”).		
Leading/Trailing Colon/Semi-colon		Begins or ends with a colon or semi-colon and contains only other numeric (e.g. “2;”).	
All Zeros		Contains only zeros (e.g. “00000”).	
2-digit Year only		For year fields, contains only 2 digits.	
Leading Zeros only		An all numeric value padded with zeros, which ARC removes.	
Needing Extra Digit only	All numeric entry that ends in a comma followed by 2 digits (e.g. “150,00”).		
Special Character only	Contains a non-alphanumeric character not mentioned for monetary variables or has a comma, decimal, question mark, or dollar sign used in a manner other than described.	Contains a non-alphanumeric character not mentioned for numeric variables or has a decimal, question mark, colon, or semi-colon used in a manner other than described.	Contains a non-alphanumeric character not mentioned for character variables or has a question mark, comma, ampersand, or forward slash used in a manner other than described.
Other	Entry was changed by the ARC for any other reason or combination of reasons.		

The graph shows the range in rates of item responses changed by ARC grouped according to the type of response each requires. The other numeric items had a small portion of their responses changed; ARC changed only 0.0 percent to 1.1 percent of other numeric items' responses. However, on average, ARC changed monetary and all-alphabetic items more often. The program modified 4.1 percent to 36.6 percent of the monetary items' responses and about 0.0 percent to 32.8 percent of the all-alphabetic items' responses.

We note that the following items had more than 20 percent of their responses modified by ARC:

- Property value (VAL, 36.6 percent);
- Hispanic origin write-in (HISW, 32.8 percent);
- Social Security amount (SS, 26.8 percent);
- Retirement income amount (RET, 25.5 percent);
- Total income estimate (TIEST, 23.7 percent);
- Total income (TI, 22.9 percent); and
- Wages amount (WAG, 21.5 percent).

We were surprised to see that some items had more than 20 or 30 percent of responses changed by ARC because preliminary results indicated that ARC would change very few item responses (please refer to [Appendix C](#) for the rates of change for all items in by the figure). Note that these rates represent only how often the ARC changed the response, not necessarily how often the changed resulted in a blank. Based on the first month of 2013 data collection, the percentage of open responses changed to blank was minimal— less than 0.05 percent per item (see [Appendix D](#) for more information).

As mentioned in the background section, we noticed that the ARC specification and/or program resulted in some unintended consequences and the ACS modified the program accordingly in the summer of 2013. ACS managers implemented the ARC Team's proposal in order to correctly handle all Internet inputs, but testing results from the new program indicate that those changes would not substantially increase or decrease these rates. For more detail on the corrections ACS made to the original ARC program, please refer to the [“Changes to the Original ARC Program”](#) subsection.

2. What are the main reasons that the ARC changed responses?

Not only did we expect that the ARC would modify very few cases, but we also anticipated that the modifications the ARC made to monetary items would be mostly due to clean up of commas, decimals, and leading dollar signs. This section looks at the reasons that the ARC changed Internet open responses.

Figures 5A through 5C show how often we saw certain input patterns in the open responses that the ARC changed for monetary, other numeric, and alphabetic items, respectively. The color shading indicates the distribution of input patterns among each items' changed responses. We group similar items together and sort each grouping by ascending proportion of changed item responses having “other” patterns (or by the “other special character” input pattern for alphabetic items since there were

no responses with the “other” pattern). We do this because the “other” pattern includes mostly responses that ARC changed if there was insufficient information or the program did not have a specific rule regarding how to handle them. By looking at the responses that fall into this category (and the “other special character” category), which we do in the next results section, we observe many ways to enhance the ARC program. We exclude a few survey items from the figures that the ARC did not modify at all.⁴

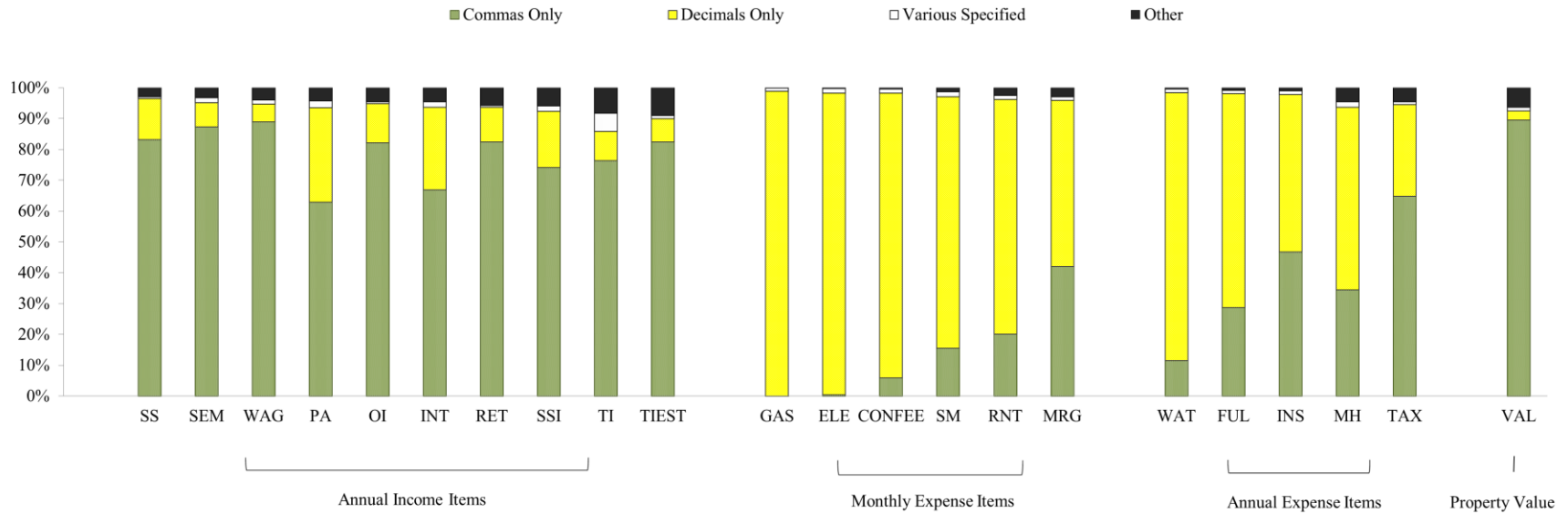
In order to make the graphs easy to read, we group together several input patterns that account for a small proportion of changed responses. [Table 3](#) of the methodology section details the full list of patterns we tallied. In all, we looked for sixteen different patterns, but ultimately recoded some of the patterns occurring less frequently into one category called “various specified”. This category groups together the input patterns not explicitly mentioned in the graph.

Every bar in the graph represents an item’s total changed responses, and that total varies from item to item. For example, ARC changed 77 responses for the zip code one year ago item (MGW6), while it changed 75,691 responses for the property value item (VAL) during the same period. Despite the large difference, both items’ changed responses are represented by one bar in the following figures because the purpose of the figure is to see the main input patterns of changed responses. To understand the frequencies associated with each input pattern, please see [Appendix C](#) for a detailed tabulation.

Figures 5A through 5C use abbreviated item names. Please see [Appendix A](#) for the full item descriptions.

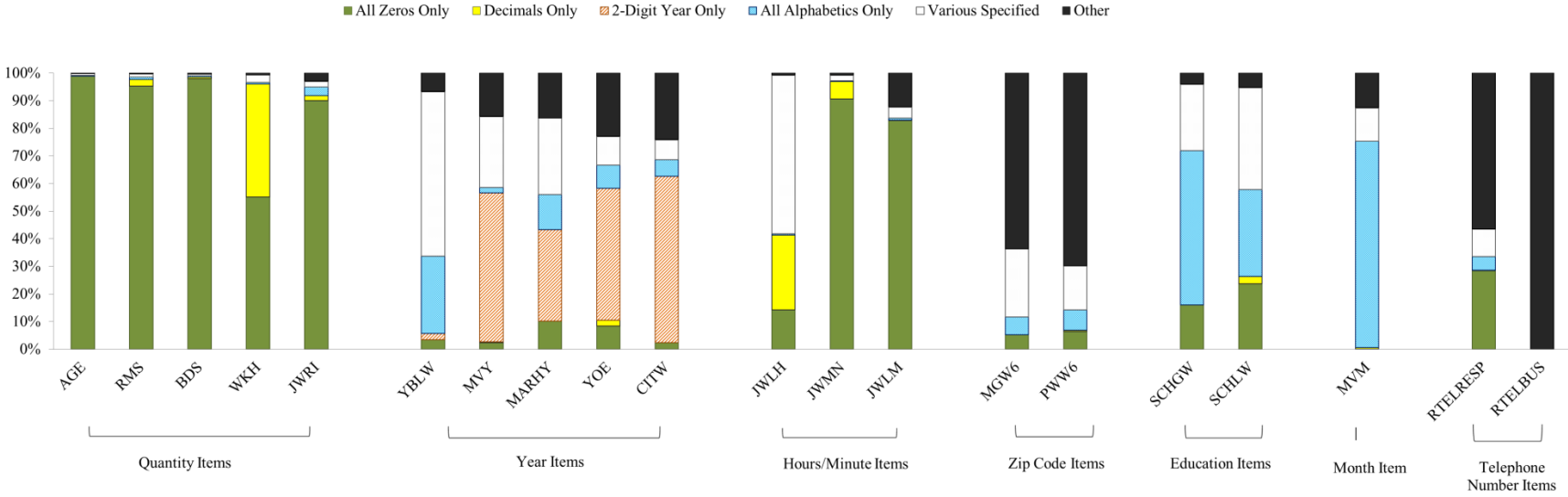
⁴ Date of birth year (DBY), last year respondent accessed the Internet instrument (LASTYR), date of birth month (DBM), last month respondent accessed the Internet instrument (LASTMO), date of birth day (DBD), Last day respondent accessed the Internet instrument (LASTDA), respondent middle initial (RMIBUS), state lived in one year ago (MGW5), and place of work city (PBW2).

Figure 5A. Estimated Reasons for ARC Response Modifications— Monetary Items



Source: American Community Survey Internet Responses Collected from December 17, 2012 to April 30, 2013

Figure 5B. Estimated Reasons for ARC Response Modifications— Other Numeric Items



Source: American Community Survey Internet Responses Collected from December 17, 2012 to April 30, 2013

Figure 5C. Estimated Reasons for ARC Response Modifications— All-Alphabetic Items



Source: American Community Survey Internet Responses Collected from December 17, 2012 to April 30, 2013

Figure 5A confirmed our intuition that ARC would change monetary responses mostly due to commas and decimals in the entry. In addition, on average, the income items and the property value item (VAL) had nominally larger proportions of responses changed by ARC due to commas only compared to the expense items. This is because the income items and VAL generally have longer responses lengths than the expense items. For example, total income (TI) usually involves five or more digits (e.g. “50,000”), whereas monthly electricity bill usually involves three or fewer digits (e.g. “100”).

Figure 5B shows that, for the other numeric items, the patterns of changed responses varied by groupings of similar items. For example, the ARC seemed to changed most quantity items due to the responses containing “all zeroes only”, while the year items seemed mostly changed due to “2-digit year” or “all alphabetic” inputs. Still, a couple other item groups and individual items had a large proportion of changed responses involving “various specified” or “other” input patterns. Unlike the monetary items, the numeric items differ by units of measure (i.e. they are not all measured in dollars), so we see greater variability in their response patterns.

From Figure 5C, we can see that most alphabetic changed responses contained “commas only”, a “forward slash only” or a “special character only”. The origin items were the only character items that contained a notable proportion of “forward slash only” responses. The address and location items, name and middle initial, and miscellaneous items prominently involved “commas only” and a “special character only”. This means that many responses contained a symbolic character other than a comma, decimal, ampersand, etc., which are symbols that the ARC has specific rules to handle. We investigate these types of responses in the next results section.

To determine how many responses the ARC changed for a certain reason, you should refer to [Appendix C](#). This research question merely looks at the main reasons why open responses were changed for each item, but does indicate the total items were changed per reason. For example, we learned from the last research question that ARC changed responses to the property value (VAL) item 36.6 percent of the time. When considering this rate of modification along with main reasons ARC changed VAL responses, you can make the following conclusion: Although ARC changed 36.6 percent of property value responses, the program changed most responses due to “commas only”. This input pattern accounted for 32.8 percentage points of the total 36.6 percent, as shown in the appendix.

Similar to the property value item, you will find several items with high rates of ARC changes that have one or two main input patterns easily handled by the ARC. However, there are a few items with low rates of ARC changes with a sizable proportion of “other” or “special character only” responses. For example, ARC modified 0.3 percent of the place of work zip code (PWW6) responses, yet we identified most of these responses as “other” (0.2 percentage points of the 0.3 percent total). These “other” or “special character only” inputs add up overall, although they may not account for a large portion of one item’s changed responses.

3. What are some examples of responses ARC changed that we categorize as “special character only” and “other”? How did ARC handle them?

In this section, we specifically look at examples of changed responses characterized as “special character only” or “other”. The designation “special character only” is one of the many input patterns

encapsulated by the “various specified” grouping from Figures 5A through 5C. We specifically tabulated changed responses with the special character only and other patterns in the tables of [Appendix C](#). For context, please refer back as needed.

The “special character only” pattern distinguishes changed responses with a character other than those dealt with by the remaining input patterns. For example, we did not specifically look for the “less than” or “greater than” symbols in changed responses for any items. Therefore, we categorize a response such as “>50” to the AGE item as an input with a special character only. If the response contains more than two special characters, then it does not meet the special character *only* designation and we would categorize it as “other”.

The “other” category distinguishes changed responses that contained multiple special characters or was not one of the specified input patterns. For example, an amount response to the total income item of “\$50,000+” involves three special characters: dollar sign, comma, and the plus symbol. We would categorize this response as “other” since no specified input patterns for monetary items deal with the plus sign. Although the pattern called “main characters” designates responses with a dollar sign, comma, and/or decimal combination, it does not include combinations with the plus symbol. Additionally, since “\$50,000+” includes more than one special character, we do not categorize it as “special character only”.

Originally, ACS management thought that reviewing changed responses with these input patterns would catch any mistakes or unsatisfactory rules that existed in ARC. However, we quickly realized that this was a faulty assumption. We found that only some of the example inputs categorized as “special character only” or “other” were problematic or needed to be put on our list of recommended enhancements. For additional details on the issues we found, please see the [“Changes to the Original ARC”](#) subsection.

The next few subsections describe examples of changed responses characterized by the “special character only” and “other” input patterns in early 2013. We report on monetary, other numeric and alphabetic items separately, but concentrate on the monetary and other numeric items. The ACS sends alphabetic responses that cannot be auto-coded to coders at the Bureau’s National Processing Center. Unlike the alphabetic items, the Internet monetary and other numeric fields receive no human intervention. Thus, they require more attention to confirm that ARC processes them correctly because the ACS has no method in place to catch miscellaneous content issues after ARC processing.

Monetary Items

Most monetary items had similar responses in the “special character only” and “other” categories. For example, responses to total income items and monthly expense items used similar symbols to indicate approximations or made similar typing errors. Responses to all monetary items included things like too many digits, too few digits, or symbols used to indicate estimates.

Below we list examples of changed responses with special character only and other patterns and show how the original ARC program transformed them. The ones we list here seemed to be the most frequent, although other examples do exist. To be clear, the arrows after the example inputs show exactly how the original ARC program handled the response during the first four months in 2013. The

labels to the right of the example inputs give one possible interpretation of what the respondent may have meant:

Semi-colon as Comma	740;000 → blank
Text and Numbers	loss12000 → blank
Apostrophe as Comma	44'000 → blank
Comma Typo	20,.000 → blank
Monthly Amounts	135/m → blank
Plus symbol	15+ → blank
Lowercase L as #1	15o,000 → blank
Too Many Zeroes	2,3000 → 23000
Range of Values	75-80 → 75
Question Mark	30,000 ? → blank
Approximations	17k appro → blank
Combination of Symbols	+3000. → blank
Plus and minus sign	3.000+/- → blank
Altilda as Approximation	~3000 → blank
Number One Typo	!80,000 → blank
Number Two Typo	@50,000 → blank
Multiplication	1500x12 → blank
Zero, Not a Cent More	.00 → blank

The corrections made to the original ARC program alleviated some, but not all, of the instances where the ACS lost data. For example, a response such as “30,000 ?” is still being blanked by the updated ARC program instead of keeping the numeric portion of “30,000”. We consider this issue as a desired enhancement for the subject matter analysts and programmers to review. The ACS will address identified enhancements in the next scheduled round of ARC updates.

Other Numeric Items

In the previous results section, we saw that input patterns for other numeric items varied by sub-type. Here we look at just examples of changed responses with a “special character only” or “other” by item sub-type and indicate how the original ARC program treated each. We saw these input patterns most often in the data, although other patterns do exist. The labels before each example indicate one possible interpretation of the respondent’s intention and the information following each arrow after the example input indicates the exact information that the original ARC program output:

Quantity Items:	Plus symbol	20+ → blank
	Less than/Great than Symbols	>50 → blank
	Range of Values	4-8 → blank
	Stray Trailing Character	71' → blank
	Number One Typo	!8 → blank
	Lowercase L as #1	l8 → blank
	All Alphabetic Response	non → . (dot)
	Hourly Amount	3hr → blank

Year Items:	Two-Digit Year	76 → blank
	Apostrophe with 2-Digit Year	'85 → blank
	Anniversary Date (mm/dd)	4/19 → blank
	Month and Year (mm/yy)	6 87 → blank
	Stray Leading/Trailing Symbol	'1960 → blank
	Missing Digit	198- → blank
	Number One Typo	!992 → blank
Hours/Minutes:	Leading ; , or .	.6 → blank
	Trailing ; , or .	6. → blank
	Plus symbol	20+ → blank
	Range of Values	1-2 → 1
	Hour Indicators	1 h → blank
Zip Codes:	Too few digits	4952 → blank
	Leading/Trailing Char.)4999 → blank
	Symbols Used as Numbers	337o7 → blank
	Other	T2E6P → blank
Education:	Kindergarten	K → blank
	Leading/Trailing Char.	1, → blank
Month Item:	Number and Letter	6u → blank
	Letter and Number	j7 → blank
Telephone Number:	Letters Converted to Numbers	480xxxxxxx → 4809999999

As with many of the monetary items, the updated ARC program still blanks many instances where the respondents provided some but not perfectly formatted or complete data (e.g. '85 for year items). We have added these issues to our list of recommended ARC enhancements for the analysts and programmers to consider. However, responses like “)4999” to zip code, which could be a typo of “04999”, are more difficult to accurately decipher and may be more of an edit than a clean-up. The ACS management, subject matter analysts, and programmers will have to discuss how to handle these more difficult responses. Ideally, we would like flag these responses as having some data, but current data processing methods make it challenging to do so at this time.

Alphabetic Items

The proportions of alphabetic changed responses with a “special character only” varied greatly by item, and no items’ changed responses involved the “other” pattern. Thus, we look at changed responses with a special character only. Because coders at the National Processing Center ultimately view alphabetic responses that fail auto-coding, ACS managers are less concerned about the changes that ARC makes to these responses.

Here we look specifically at three alphabetic items with nominally large levels of changed responses involving a special character: other Hispanic origin (HISW), name of employer (INW2) and place of work street address (PWW1). The ARC program changed more than 20,000 responses for these items in the first four months of 2013 due to a special character.

The HISW responses that the ARC changed due to the special character only pattern mostly involved responses with the ñ character. ARC incorrectly translated this character originally, but the ACS quickly caught this mistake and modified the ARC program to correctly translate this character in early 2013. The INW2 and PWW1 changed responses characterized by the special character only pattern were similar in that they included various punctuation marks. Most of these changed responses contained an apostrophe in combination with commas or periods. For example, one respondent wrote for name of employer, “Lowe’s Auto Parts, Inc.”. When ARC modified the response, it replaced the special characters and any commas/periods with a space. This example became “Lowe s Auto Parts Inc”, which was then sent to the coders if it failed auto-coding.

CONCLUSIONS

We found that the automated cleanup of Internet open responses changed a small percentage of “other numeric” responses, but the rates of changed responses for the monetary and alphabetic items varied greatly. Although many monetary and alphabetic items had rates less than ten percent, seven items had 20 percent or more of their item responses changed by ARC: property value, Hispanic origin, Social Security amount, retirement income amount, total income estimate, total income, and wages amount.

Looking at items’ changed responses by input pattern, however, our results indicate that the ARC program correctly changed most open responses. Most monetary responses were changed due to “commas only” or “decimals only” in the entry (e.g. “3,000” becomes “3000”). The alphabetic items, which we are less concerned about because they are sent to coding, were mostly changed due to “commas only”, “forward slash only” or a “special character only”. When ARC encountered these symbols, it stripped the character and inserted a space. Lastly, the other numeric items changed for various reasons that differed by the measuring unit for each item. For example, the year items changed most due to “2-digit year” or “all alphabetic” inputs, whereas the quantity items changed mostly due to responses containing “all zeroes only”.

In the final results section of this report, we listed examples of responses involving “special character only” or “other” input patterns. These are two of the sixteen input patterns that we used to tabulate changed responses. Changed responses in these two categories were instances where respondents provided some, but not perfectly formatted or complete, information. For the monetary and other numeric items, many of these changed items included typos or estimated amounts (e.g. “3,000 ?”). For the character items, these responses involved special characters such as “ñ” or a combination of punctuation/symbols.

After seeing the preliminary results from this report, we identified several issues with the original ARC specification and program that mostly stemmed from the interaction of rules in the computer program or from ambiguities in the specification. These changes were implemented on August 19, 2013 to correct and re-structure the ARC program. We documented the specific changes made to the original

ARC in [Appendix B](#), but, in making these changes, the ACS re-structured the actual coding of the computer program to allow for easily adding enhancements in the future. These changes corrected instances where the original ARC was incorrectly blanking some responses and provided the ability to easily add future enhancements. This will help integrate the ARC processing into the Census Bureau's new corporate software system for managing data from multiple surveys.

NEXT STEPS

Based on the results of this report, the ACS is enhancing the ARC program to account for data loss or mis-recording responses, as shown by some of the examples in this report. We held off on making enhancements until the ARC Team had the time to list and decide on them. ACS managers are testing and approving enhancements that will salvage much of the data we currently lose.

Additionally, the ACS should continue to test if changes made to the instrument would encourage respondents to provide open responses in the format required. There may be ways to do this, without burdening them.

APPENDIX A. Internet Open Response Items and Descriptions from the 2013 ACS

MONETARY

Annual Income Items	Internet Variable Name
Interest amount	INT
Other income amount	OI
Public assistance amount	PA
Retirement income amount	RET
Self-employment amount	SEM
Social security amount	SS
Supplemental Security Income (SSI) amount	SSI
Total income	TI
Estimated total income	TIEST
Wages amount	WAG
Other Annual Items	
Other type of fuel payment amount	FUL
Insurance	INS
Mobile home expenses	MH
Taxes	TAX
Water payment amount	WAT
Monthly Items	
Condo fee amount	CONFEE
Electricity payment amount	ELE
Gas payment amount	GAS
Mortgage payment amount	MRG
Monthly rent	RNT
Second mortgage payment amount	SM
Other Values	
Property value	VAL

OTHER NUMERIC (NON-MONETARY)

Quantity	Internet Variable Name
Age	AGE
Bedrooms	BDS
Number of riders	JWRI
Total rooms	RMS
Number of hours worked	WKH
Year	
Citizenship	CITW
Date of birth year	DBY
Last year respondent accessed	LASTYR
Year of marriage	MARHY
Year moved in	MVY
Year built	YBLW
Year of entry	YOE
Month	
Date of birth month	DBM
Last month respondent accessed	LASTMO
Month moved in	MVM
Day of Month	
Date of birth day	DBD
Last day respondent accessed	LASTDA
Hours/Minutes	
Time left for work hour	JWLH
Time left for work minute	JWLM
Minutes to work	JWMN
Zip Codes	
ZIP code one year ago	MGW6
Place of work ZIP code	PWW6
School Grade Level	
Grade attending	SCHGW
Education attainment	SCHLW
Telephone Number	
Respondent telephone number	RTELBUS
Respondent telephone number	RTELRESP

ALPHABETIC

Names	Internet Variable Name
Rosterer Person First Name	FNAME
Rosterer Person Last Name	LNAME
Respondent first name	RFNBUS
Respondent first name	RFNRESP
Respondent last name	RLNBUS
Respondent last name	RLNRESP
Initials	
Rosterer Person Middle Initial	MI
Respondent middle initial	RMIBUS
Respondent middle initial	RMIRESP
Origins	
Ancestry	ANCW
Other Hispanic Origin	HISW
American Indian and Alaska Native	RCW1
Specify some other race	RCW2
Other Asian	RCW3
Other Pacific Islander	RCW4
Street	
Street address one year ago	MGW2
Place of work street address	PWW1
City	
City one year ago	MGW3
Place of work city	PWW2
County	
County one year ago	MGW4
Place of work county	PWW4
State	
State one year ago	MGW5
Place of birth state	PBW2
Place of work state	PWW5
Country	
Country one year ago	MGW1
Place of birth - country	PBW3
Other Write-Ins	
Other type of computer	COMPOTHW
Field of degree	FODW
Specify other health insurance	HINSW
Name of employer	INW2
Type of industry	INW3
Specify other language	LANW
Type of work	OCW1
Major duties	OCW2
Other type of Internet service	OTHSVCEW

APPENDIX B. List of Automated Response Cleanup (ARC) Changes in Summer 2013

Purpose: This document lists the changes that the ARC Team implemented in August 2013 to the original program/specification and the issues that it saved for later review/enhancement. The column "Action Still Needed?" indicates whether the issue is outstanding. Any issues that the group saved for program enhancements will be addressed by the subject matter analysts in our next set of changes scheduled for ARC.

Variable(s)	Issue with Original ARC	Updated ARC (as of 8/19/2013)	Action Still Needed?
AGE	A handful of inputs like "non" were incorrectly transformed to a dot instead of a blank.	These kinds of responses are now blanked, in accordance with the original specification.	No. This was already approved by the ARC Team.
AGE	Inputs like "4mo" were being kept as "04m" by original ARC but the original specification said to blank.	New ARC now keeps numeric entries with trailing "mo".	No. This was already approved by the ARC Team.
COMPOTHW, HINSW, OTHSVCEW	All write-ins with a "-" dash were being blanked. It was unclear from the spec if this was supposed to be a valid character.	The "-" character is now part of valid character set for these items.	No. This was already approved by the ARC Team.
Amount fields (CONFEE, ELE, FUL, GAS, INS, INT, MRG, OI, RET, RNT, SEM, SM, SS, SSI, TAX, TI, TIEST, VAL, WAG, WAT)	The original ARC was keeping a partial entry for responses that the specification said to blank (e.g. "20,.000" was being kept as 20000).	New ARC now blanks these responses according to the original specification.	No. This was already approved by the ARC Team.

Variable(s)	Issue with Original ARC	Updated ARC (as of 8/19/2013)	Action Still Needed?
Amount fields (CONFEE, ELE, FUL, GAS, INS, INT, MRG, OI, RET, RNT, SEM, SM, SS, SSI, TAX, TI, TIEST, VAL, WAG, WAT)	In some cases, the respondent is trying to communicate "about this amount" and original ARC is probably keeping good information (e.g. "100+" is recorded as "100"). However, this is not what the specification intended to do.	New ARC blanks the field.	Maybe. The ARC Team should discuss developing enhancements for responses like these.
Amount fields (CONFEE, ELE, FUL, GAS, INS, INT, MRG, OI, RET, RNT, SEM, SM, SS, SSI, TAX, TI, TIEST, VAL, WAG, WAT)	In other cases, the meaning of the entry is changed by the original ARC program by retaining the number (a negative number is made positive, a range is read wrong, etc). For example, "75-80" is changed to "75". This is not what the specification intended to do.	New ARC blanks the field.	Yes. The ARC Team should discuss developing enhancements for responses like these.
Amount fields (CONFEE, ELE, FUL, GAS, INS, INT, MRG, OI, RET, RNT, SEM, SM, SS, SSI, TAX, TI, TIEST, VAL, WAG, WAT)	A few differences were trivial: the input "00" two zeroes is changed to "0" one zero in the original program. To keep formatting consistent, it should be kept as "00" two zeroes.	New ARC keeps "00" as "00".	No. This was already approved by the ARC Team.
FNAME, LNAME, MI, RFNRESP, RLNRESP, RMIRESP	The original ARC was blanking some entries if there was a dash, invalid character, or a number.	New ARC strips the invalid characters and keeps the valid characters, which includes keeping dashes and numbers, as per the original specification.	No. This was already approved by the ARC Team.
INTCALC	The original spec has a very simplified rule set for INTCALC that does not match TI/TIEST. Should we put INTCALC through the same transformations as TI?	We learned that INTCALC should not appear in ARC Diff. It is an instrument calculated variable that should be free of commas, dollar signs, etc. New ARC makes no change to INTCALC's rule set.	No. This was already approved by the ARC Team.

Variable(s)	Issue with Original ARC	Updated ARC (as of 8/19/2013)	Action Still Needed?
JWLH	The original specification said to blank entries with a number and leading/trailing "." dot. However, the original ARC program was keeping the number (e.g. "6." kept as "6"). Same issue as JWLM.	New ARC now allows leading and trailing ":" colon, ";" semi-colon, "." dot for numeric input	No. This was already approved by the ARC Team.
JWLM	The original specification said to blank entries with a number and leading/trailing "." dot. However, the original ARC program was keeping the number (e.g. ".6" kept as "6"). Same issue as JWLH.	New ARC now allows leading and trailing ":" colon, ";" semi-colon, and "." dot for numeric input.	No. This was already approved by the ARC Team.
JWMN	The original ARC blanked responses of the format #<space>#. Should this be? The original specification was unclear.	New ARC blanks digit-space-digit responses.	No. This was already approved by the ARC Team.
JWMN	Should ARC translate "hr"? E.g. "6hr" is blanked by original ARC.	New ARC also blanks and does not perform this translation.	Yes. The ARC Team will discuss this enhancement.
LANW	The original specification does not say to do translations, but the analysts indicated that they would like to do a translation of "&"-->"and" and "+"-->"plus" for LANW.	New ARC implements the "&"-->"and" and "+"-->"plus" translations.	No. This was already approved by the ARC Team.
LANW, PWW, MGW, COMPOTH, INTSVCE, RCW	These variables all have translations built into the original ARC program for "&" and "+", but how they are translated varies.	New ARC does not address the consistency of translations. Is this an issue?	Yes. The ARC Team should discuss developing enhancements for responses like these.

Variable(s)	Issue with Original ARC	Updated ARC (as of 8/19/2013)	Action Still Needed?
MARHY, MVY	Original ARC was not blanking responses in the form of #<space>##. The original specification indicated these should be blanked. These are responses where respondents seem to provide the month and two digit year of the date in question.	New ARC blanks these responses (e.g. "6 46").	No. This was already approved by the ARC Team.
MVM	The original program and specification did not match for entries of "j". Original ARC was making these "55" instead of "00".	New ARC makes "j" into "00".	No. This was already approved by the ARC Team.
RCW1, RCW2, RCW3	The analysts and the original specification indicated that the "/" forward slash is invalid, but Ali recalled a decision to make "/" valid only if it is a fraction. The original program was keeping "/".	New ARC now replaces all forward slashes "/" with a space, even fractions.	No. This was already approved by the ARC Team.
WKH	Invalid responses such as "non" were not being blanked by the original program.	New ARC now blanks responses that do not meet any of our specified rules for WKH.	No. This was already approved by the ARC Team.
WKH	In cases where the input contains foreign characters, the original ARC program does not blank the response.	New ARC now blanks responses with foreign characters.	Yes. The ARC Team should discuss developing enhancements for responses like these.

Variable(s)	Issue with Original ARC	Updated ARC (as of 8/19/2013)	Action Still Needed?
WKH	The original ARC blanked responses with trailing "hr" (such as "4hr"). However, the original specification accounted for only "h". It was unclear whether analysts meant to include "hr" as well.	New ARC now does the "hr" hour conversion (e.g. "3hr" becomes "003").	No. This was already approved by the ARC Team.

APPENDIX C. Estimated Reasons for ARC Modifications

Table C1. Frequency and Input Patterns of Internet Responses that ARC Changed— Monetary Items

		Annual Income Items									
		INT	OI	PA	RET	SEM	SS	SSI	TI	TIEST	WAG
Total Open Responses		94,956	28,798	6,623	58,192	31,457	91,364	9,920	25,401	43,103	355,499
		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Responses ARC Changed		13,566	4,253	556	14,841	4,990	24,460	1,688	5,805	10,223	76,331
		14.3%	14.8%	8.4%	25.5%	15.9%	26.8%	17.0%	22.9%	23.7%	21.5%
Input Pattern of Changed Responses	Commas Only	9,060	3,491	349	12,231	4,358	20,346	1,250	4,428	8,418	67,895
		9.5%	12.1%	5.3%	21.0%	13.9%	22.3%	12.6%	17.4%	19.5%	19.1%
	Decimals Only	3,637	544	171	1,658	385	3,228	308	549	776	4,388
		3.8%	1.9%	2.6%	2.9%	1.2%	3.5%	3.1%	2.2%	1.8%	1.2%
	Question Marks Only	54	3	5	6	17	20	8	70	7	111
		0.1%	0.0%	0.1%	0.0%	0.1%	0.0%	0.1%	0.3%	0.0%	0.0%
	Leading Dollar Sign Only	13	1	0	2	2	10	1	2	2	75
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Main Characters Only	17	0	1	7	8	4	0	4	11	116
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
All Alphabets Only	100	7	2	31	34	35	3	186	9	376	
	0.1%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.7%	0.0%	0.1%	
Postfix Only	5	0	0	1	0	1	0	4	0	4	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
Needing Extra Digit Only	24	9	2	26	7	45	9	27	64	217	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.2%	0.1%	
Special Character Only	37	4	2	19	17	37	9	56	14	116	
	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%	0.2%	0.0%	0.0%	
Other	619	194	24	860	162	734	100	479	922	3,033	
	0.7%	0.7%	0.4%	1.5%	0.5%	0.8%	1.0%	1.9%	2.1%	0.9%	

Source: American Community Survey Internet Responses from December 17, 2012 to April 30, 2013

Note: We round percentages to the nearest 0.1 percent. See [Table 3](#) for input pattern definitions.

(continued on next page)

Table C1 [Continued]. Frequency and Input Patterns of Internet Responses that ARC Changed— Monetary Items

		Monthly Housing Variables					Annual Housing Variables					Other Housing Variables		
		CONFEE	ELE	GAS	MRG	RNT	SM	FUL	INS	MH	TAX	WAT	VAL	
Total Open Responses		11,741	260,494	149,151	136,270	61,028	37,910	24,157	187,408	7,479	199,331	191,768	206,729	
		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	
Responses ARC Changed		624	21,636	10,044	8,733	2,480	1,882	1,382	12,382	634	23,106	9,808	75,691	
		5.3%	8.3%	6.7%	6.4%	4.1%	5.0%	5.7%	6.6%	8.5%	11.6%	5.1%	36.6%	
Input Pattern of Changed Responses	Commas Only	36 0.3%	71 0.0%	23 0.0%	3,664 2.7%	497 0.8%	290 0.8%	395 1.6%	5,779 3.1%	218 2.9%	14,980 7.5%	1,127 0.6%	67,759 32.8%	
	Decimals Only	577 4.9%	21,175 8.1%	9,896 6.6%	4,704 3.5%	1,889 3.1%	1,536 4.1%	961 4.0%	6,331 3.4%	376 5.0%	6,850 3.4%	8,518 4.4%	2,256 1.1%	
	Question Marks Only	0 0.0%	10 0.0%	13 0.0%	10 0.0%	1 0.0%	0 0.0%	2 0.0%	49 0.0%	4 0.1%	44 0.0%	18 0.0%	59 0.0%	
	Leading Dollar Sign Only	3 0.0%	174 0.1%	34 0.0%	21 0.0%	11 0.0%	8 0.0%	5 0.0%	17 0.0%	1 0.0%	33 0.0%	48 0.0%	28 0.0%	
	Main Characters Only	1 0.0%	64 0.0%	11 0.0%	8 0.0%	5 0.0%	0 0.0%	1 0.0%	1 0.0%	0 0.0%	23 0.0%	4 0.0%	141 0.1%	
	All Alphabets Only	1 0.0%	16 0.0%	12 0.0%	35 0.0%	5 0.0%	5 0.0%	2 0.0%	25 0.0%	2 0.0%	25 0.0%	11 0.0%	38 0.0%	
	Postfix Only	0 0.0%	4 0.0%	1 0.0%	1 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.0%	4 0.0%
	Needing Extra Digit Only	1 0.0%	40 0.0%	29 0.0%	19 0.0%	6 0.0%	13 0.0%	2 0.0%	40 0.0%	3 0.0%	66 0.0%	18 0.0%	550 0.3%	
	Special Character Only	2 0.0%	23 0.0%	7 0.0%	14 0.0%	2 0.0%	4 0.0%	2 0.0%	17 0.0%	1 0.0%	24 0.0%	13 0.0%	50 0.0%	
	Other	3 0.0%	59 0.0%	18 0.0%	257 0.2%	64 0.1%	26 0.1%	12 0.1%	121 0.1%	29 0.4%	1,061 0.5%	49 0.0%	4,806 2.3%	

Source: American Community Survey Internet Responses from December 17, 2012 to April 30, 2013

Note: We round percentages to the nearest 0.1 percent. See [Table 3](#) for input pattern definitions.

Table C2. Frequency and Input Patterns of Internet Responses that ARC Changed— Other Numeric Items

		Quantity						Year						Month			
		AGE	BDS	IDREF	JWRI	RMS	WKH	CITW	DBY	LASTYR	MARHY	MVY	YBLW	YOE	DBM	LASTMO	MVM
Total Open Responses		708,843	278,347	284,014	287,432	280,499	372,336	31,885	700,991	284,014	412,291	276,457	48,748	69,740	700,547	284,014	276,008
		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Responses ARC Changed		7,973	2,488	0	505	660	2,763	83	0	0	148	833	89	48	0	0	835
		1.1%	0.9%	0.0%	0.2%	0.2%	0.7%	0.3%	0.0%	0.0%	0.0%	0.3%	0.2%	0.1%	0.0%	0.0%	0.3%
Input Patterns of Changed Responses	All Zeros Only	7,887	2,438	0	455	629	1,525	2	0	0	15	20	3	4	0	0	0
		1.1%	0.9%	0.0%	0.2%	0.2%	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Decimals Only	2	18	0	9	16	1,129	0	0	0	0	2	0	1	0	0	4
		0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Question Marks Only	18	2	0	2	4	12	4	0	0	14	24	39	2	0	0	61
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%
	All Alphabets Only	21	9	0	16	6	19	5	0	0	19	16	25	4	0	0	625
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.2%
	Postfix Only	4	0	0	1	0	5	0	0	0	0	8	0	0	0	0	16
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Leading/Trailing Colon/Semi-Colon Only	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
2-Digit Year Only	0	0	0	0	0	0	50	0	0	49	450	2	23	0	0	0	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	
Leading Zeros Only	0	0	0	0	0	0	1	0	0	3	172	7	1	0	0	0	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	
Special Character Only	34	12	0	7	3	55	1	0	0	23	10	7	2	0	0	24	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
Other	7	8	0	15	2	17	20	0	0	24	131	6	11	0	0	105	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	

Source: American Community Survey Internet Responses from December 17, 2012 to April 30, 2013

Note: We round percentages to the nearest 0.1 percent. See [Table 3](#) for input pattern definitions.

(continued on next page)

Table C2 [Continued]. Frequency and Input Patterns of Internet Responses that ARC Changed— Other Numeric Items

		Day of Month		Hours/ Minutes			Zip Codes		Education		Telephone number	
		DBD	LASTDA	JWLH	JWLM	JWMN	MGW6	PWW6	SCHGW	SCHLW	RTELBUS	RTELRSP
Total Open Responses		697,962	284,014	301,782	283,744	307,470	42,525	294,703	102,350	104,533	384	281,748
		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Responses ARC Changed		0	0	1,933	1,699	1,793	77	857	25	38	1	1,143
		0.0%	0.0%	0.6%	0.7%	0.6%	0.2%	0.3%	0.0%	0.0%	0.3%	0.4%
Input Patterns of Changed Responses	All Zeros Only	0	0	275	1,408	1,627	4	55	4	9	0	325
		0.0%	0.0%	0.1%	0.5%	0.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
	Decimals Only	0	0	523	3	111	0	3	0	1	0	2
		0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Question Marks Only	0	0	7	5	10	7	51	0	2	0	1
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	All Alphabets Only	0	0	11	13	7	5	64	14	12	0	56
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Postfix Only	0	0	7	0	11	0	3	2	6	0	11
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Leading/Trailing Colon/Semi-Colon Only	0	0	1,085	58	4	0	0	0	0	0	0	
	0.0%	0.0%	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
2-Digit Year Only	0	0	0	0	0	0	0	0	0	0	0	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
Leading Zeros Only	0	0	0	0	0	6	46	0	0	0	11	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
Special Character Only	0	0	11	4	10	6	37	4	6	0	91	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
Other	0	0	14	208	13	49	598	1	2	1	646	
	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%	0.2%	0.0%	0.0%	0.3%	0.2%	

Source: American Community Survey Internet Responses from December 17, 2012 to April 30, 2013

Note: We round percentages to the nearest 0.1 percent. See [Table 3](#) for input pattern definitions.

Table C3 Frequency and Input Patterns of Internet Responses that ARC Changed— All-Alphabetic Items

		Origins						Street		City		County		State			Country	
		HISW	RCW1	RCW2	RCW3	RCW4	ANCW	MGW2	PWW1	MGW3	PWW2	MGW4	PWW4	MGW5	PBW2	PWW5	MGW1	PBW3
Total Open Responses		621,869	20,103	8,858	24,781	6,618	674	42,757	311,927	44,488	319,852	34,734	307,741	44,064	591,257	311,308	3,613	71,510
		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Responses ARC Changed		204,247	1,669	463	1,405	150	34	8,308	60,987	1,040	19,494	631	4,450	0	0	3,356	101	2,766
		32.8%	8.3%	5.2%	5.7%	2.3%	5.0%	19.4%	19.6%	2.3%	6.1%	1.8%	1.5%	0.0%	0.0%	1.1%	2.8%	3.9%
Input Pattern for Changed Responses	Commas Only	140,850	351	0	0	0	0	21	3,428	535	13,313	9	183	0	0	185	55	1,925
		22.6%	1.8%	0.0%	0.0%	0.0%	0.0%	0.1%	1.1%	1.2%	4.2%	0.0%	0.1%	0.0%	0.0%	0.1%	1.5%	2.7%
	Forward Slash Only	34,523	540	212	909	103	22	0	0	0	0	17	322	0	0	51	2	0
		5.6%	2.7%	2.4%	3.7%	1.6%	3.3%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.1%	0.0%
	Question Marks Only	846	23	25	31	2	1	26	41	3	5	7	130	0	0	3	0	1
		0.1%	0.1%	0.3%	0.1%	0.0%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Ampersands Only	3,831	0	0	0	0	0	3	778	0	15	0	51	0	0	6	0	60
	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	
Main Characters Only	2,062	67	46	89	4	1	0	99	0	5	1	13	0	0	0	0	1	
	0.3%	0.3%	0.5%	0.4%	0.1%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
Special Character Only	22,135	688	180	376	41	10	8,258	56,641	502	6,156	597	3,751	0	0	3,111	44	779	
	3.6%	3.4%	2.0%	1.5%	0.6%	1.5%	19.3%	18.2%	1.1%	1.9%	1.7%	1.2%	0.0%	0.0%	1.0%	1.2%	1.1%	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	

Source: American Community Survey Internet Responses from December 17, 2012 to April 30, 2013

Note: We round percentages to the nearest 0.1 percent. See [Table 3](#) for input pattern definitions.

(continued on next page)

Table C3 [Continued]. Frequency and Input Patterns of Internet Responses that ARC Changed— All-Alphabetic Items

		Names						Initials			Other Write-Ins									
		FNAME	LNAME	RFNBUS	RFNRESP	RLNBUS	RLNRESP	MI	RMIBUS	RMIRESP	COMPOTH W	FODW	HINSW	INW2	INW3	LANW	OCW1	OCW2	OTHSVC EW	
Total Open Responses		714,331	710,484	381	283,096	377	281,933	574,714	249	233,988	13,036	207,325	24,982	405,632	413,183	79,962	413,416	403,116	3,828	
		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	
Responses ARC Changed		1,374	7,200	1	378	5	2,115	369	0	252	248	23,435	4,031	59,310	10,287	4,278	15,451	46,685	396	
		0.2%	1.0%	0.3%	0.1%	1.3%	0.8%	0.1%	0.0%	0.1%	1.9%	11.3%	16.1%	14.6%	2.5%	5.4%	3.7%	11.6%	10.3%	
Input Pattern for Changed Responses	Commas Only	48	1,423	0	7	0	347	0	0	0	0	9,219	363	11,199	4,285	2,917	9,394	35,882	0	
		0.0%	0.2%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	4.5%	1.5%	2.8%	1.0%	3.7%	2.3%	8.9%	0.0%	
	Forward Slash Only	23	52	0	3	0	4	4	0	2	0	4,296	0	0	0	0	0	0	0	0
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Question Marks Only	7	15	0	3	0	3	15	0	3	3	9	7	81	38	2	29	61	18	
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	
	Ampersands Only	0	0	0	0	0	0	0	0	0	41	2,292	0	0	0	160	0	0	136	
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	1.1%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	3.6%		
Main Characters Only	2	1	0	9	0	1	0	0	0	2	380	16	1,575	306	15	312	1,700	7		
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.1%	0.4%	0.1%	0.0%	0.1%	0.4%	0.2%		
Special Character Only	1,294	5,709	1	356	5	1,760	350	0	247	202	7,239	3,645	46,455	5,658	1,184	5,716	9,042	235		
	0.2%	0.8%	0.3%	0.1%	1.3%	0.6%	0.1%	0.0%	0.1%	1.6%	3.5%	14.6%	11.5%	1.4%	1.5%	1.4%	2.2%	6.1%		
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		

Source: American Community Survey Internet Responses from December 17, 2012 to April 30, 2013

Note: We round percentages to the nearest 0.1 percent. See [Table 3](#) for input pattern definitions.

APPENDIX D. Rates of Blanked Open Item Responses by the ARC

Item	Total Changed Responses by ARC	Responses Changed to Blank	Percent Blanked
AGE	112,135	9	0.0%
ANCW	105,019	89	0.1%
BDS	44,714	4	0.0%
CITW	5,409	10	0.2%
COMPOTHW	2,062	0	0.0%
CONFEE	1,923	0	0.0%
DBD	110,631	0	0.0%
DBM	110,995	0	0.0%
DBY	111,112	0	0.0%
ELE	41,848	6	0.0%
FNAME	112,608	10	0.0%
FODW	34,750	1	0.0%
FUL	3,971	0	0.0%
GAS	23,909	4	0.0%
HINSW	4,056	0	0.0%
HISW	3,124	0	0.0%
IDREF	45,334	0	0.0%
INS	30,202	9	0.0%
INT	15,337	30	0.2%
INW2	69,017	11	0.0%
INW3	70,303	7	0.0%
JWLH	50,223	167	0.3%
JWLM	47,380	14	0.0%
JWMN	51,132	6	0.0%
JWRI	47,780	3	0.0%
LANW	13,803	0	0.0%
LASTDA	45,334	0	0.0%
LASTMO	45,334	0	0.0%
LASTYR	45,334	0	0.0%
LNAME	112,157	6	0.0%
MARHY	66,850	6	0.0%
MGW1	708	0	0.0%
MGW2	7,519	3	0.0%
MGW3	7,852	1	0.0%
MGW4	6,102	3	0.0%
MGW5	7,796	0	0.0%
MGW6	7,515	12	0.2%
MH	1,141	1	0.1%
MI	91,929	54	0.1%
MRG	22,189	9	0.0%
MVM	44,195	16	0.0%
MVY	44,266	103	0.2%
OCW1	70,355	8	0.0%
OCW2	68,690	10	0.0%

Item	Total Changed Responses by ARC	Responses Changed to Blank	Percent Blanked
OI	4,938	5	0.1%
OTHSVCEW	598	1	0.2%
PA	1,097	0	0.0%
PBW2	99,346	0	0.0%
PBW3	12,238	1	0.0%
PWW1	52,194	7	0.0%
PWW2	53,303	1	0.0%
PWW4	51,310	14	0.0%
PWW5	51,982	0	0.0%
PWW6	49,265	135	0.3%
RCW1	1,409	6	0.4%
RCW2	3,660	0	0.0%
RCW3	1,098	0	0.0%
RCW4	93	0	0.0%
RET	9,271	12	0.1%
RFNBUS	78	0	0.0%
RFNRESP	45,174	2	0.0%
RLNBUS	78	0	0.0%
RLNRESP	45,024	2	0.0%
RMIBUS	53	0	0.0%
RMIRESP	37,568	42	0.1%
RMS	45,027	6	0.0%
RNT	10,123	0	0.0%
RTELBUS	79	0	0.0%
RTELRESP	44,983	152	0.3%
SCHGW	18,373	3	0.0%
SCHLW	18,544	5	0.0%
SEM	5,217	8	0.2%
SM	6,170	2	0.0%
SS	14,492	17	0.1%
SSI	1,687	0	0.0%
TAX	31,962	21	0.1%
TI	4,175	40	1.0%
TIEST	7,084	6	0.1%
VAL	32,994	40	0.1%
WAG	60,977	102	0.2%
WAT	30,826	6	0.0%
WKH	63,388	9	0.0%
YBLW	7,871	7	0.1%
YOE	11,987	3	0.0%

Note: We round percentages to the nearly tenth of one percent.

Source: January 2013 American Community Survey panel data collected on or before Jan 30, 2013