# DHS reproductive calendar processing workflow

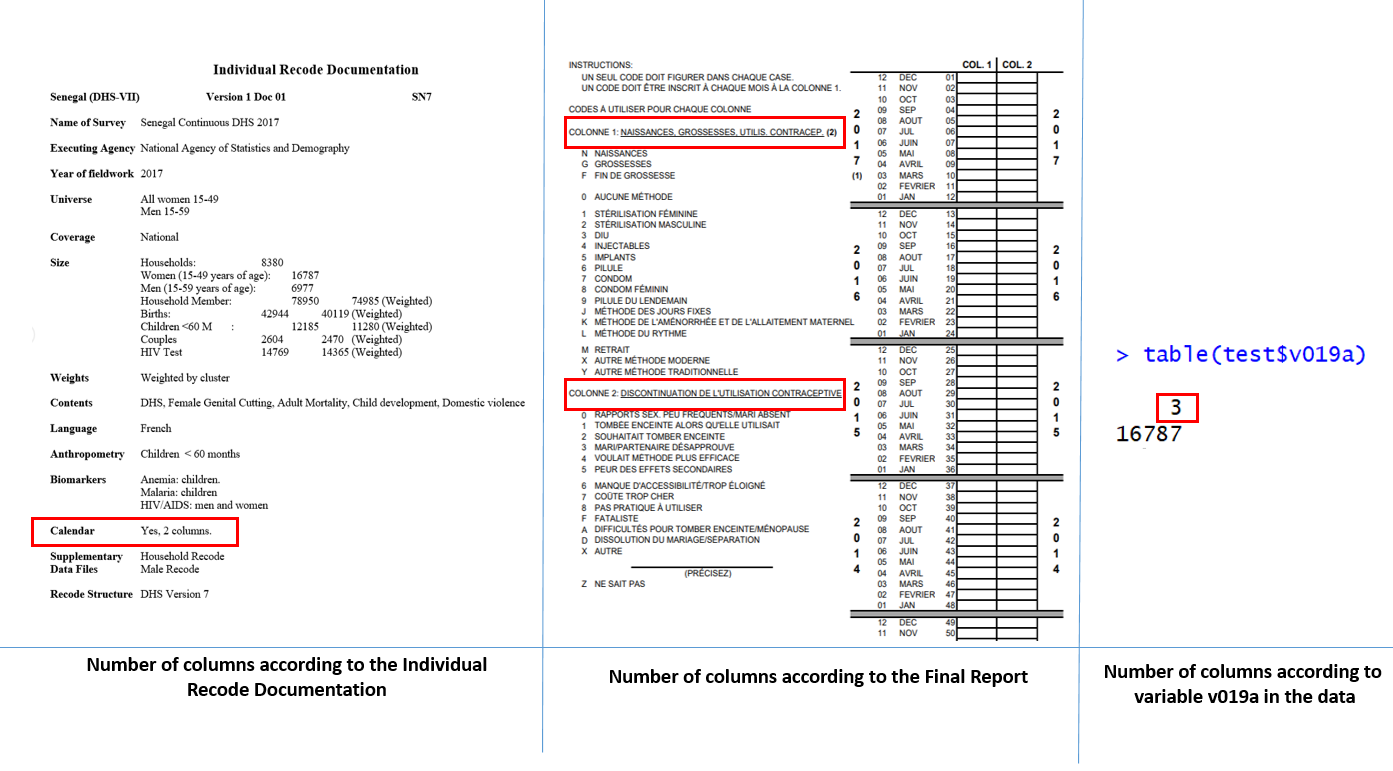
The purpose of this document is to provide a step-by-step description of transforming reproductive health calendar data from DHS. In short, the goal is to transform two or more (depending on the samples) 80-character long string variables on reproductive events to women-months, with every row representing a month in a reproductive history of a woman. In addition to transforming the calendar variables, we also create additional variables describing reproductive health events and contraceptive use. This document explains data processing and is accompanied by the R script. An overview of the DHS reproductive calendar is available here: <https://www.dhsprogram.com/data/calendar-tutorial/>.

There are nine possible calendar variables, although most samples only have the first two listed below. The variables are:

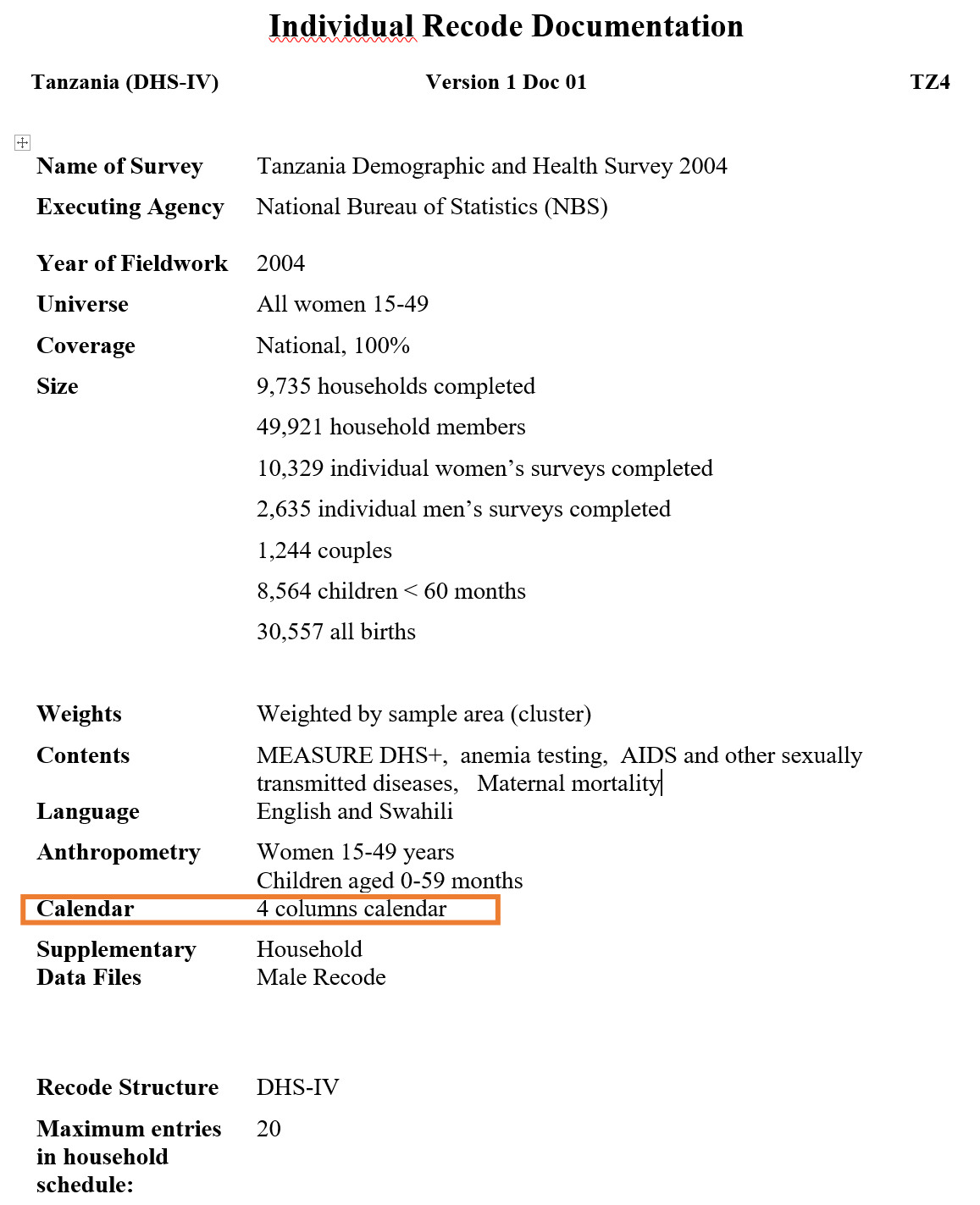
* Vcal\_1 – births, pregnancies, and contraceptive use
* Vcal\_2 – reasons for discontinuation of contraceptive use
* Vcal\_3 – marital status
* Vcal\_4 – residence
* Vcal\_5 – source of contraceptive
* Vcal\_6 – vcal9 – country-specific

# Processing logic

1. Set up your R environment – the first few lines in the script are dedicated to that. Specify a working directory and create the following folders inside it: **input**, **output**, and **scripts**.
   1. Copy the Functions script to the **scripts** directory, and save other sample-specific scripts in that folder. That script will load functions needed to process the data.
   2. Store IR data files and files with column names (what those are is described in more details below) to you **input** folder.
   3. Inside the **output** directory, create a directory called “**test\_examples**” to store test examples of processed data for manual checks.
2. Next, check if a country has calendar variables. There are multiple ways to do that. One way is to download an Individual Data recode dataset from the DHS website. Inside the data folder, there will be an Individual Recode Documentation file (in MS Word format) that documents whether the sample has calendar data – see below. According to this information, there are four calendar variables in the Tanzania 2004 sample. If it is not obvious right away how many calendar columns there are, to find out it’s best to check the data manually (variable v019a) and check the IR documentation as the number of columns reported in these sources sometimes differ – see figure below. For example, the figure below indicates there are two calendar columns in Senegal 2017 DHS, but variable v019a indicates there are three. If not sure about the number of columns, a third source to check is the sample final report – just scroll to the questionnaire section and find the calendar section questionnaires (it is usually located toward the very end of the final report).

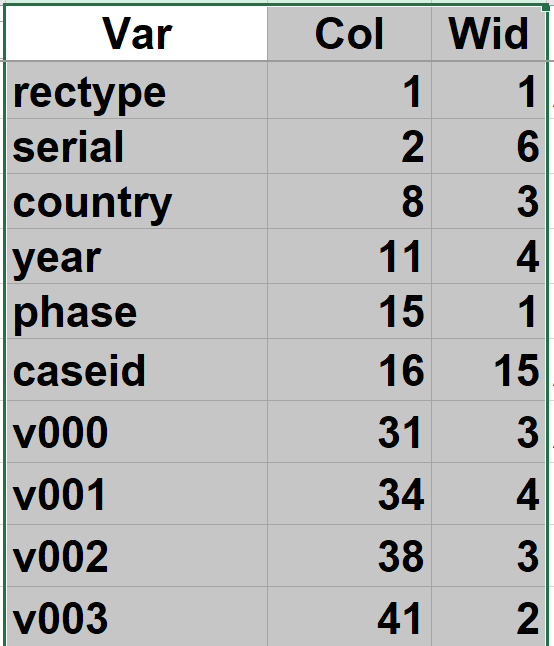


Number of calendar columns according to different sources

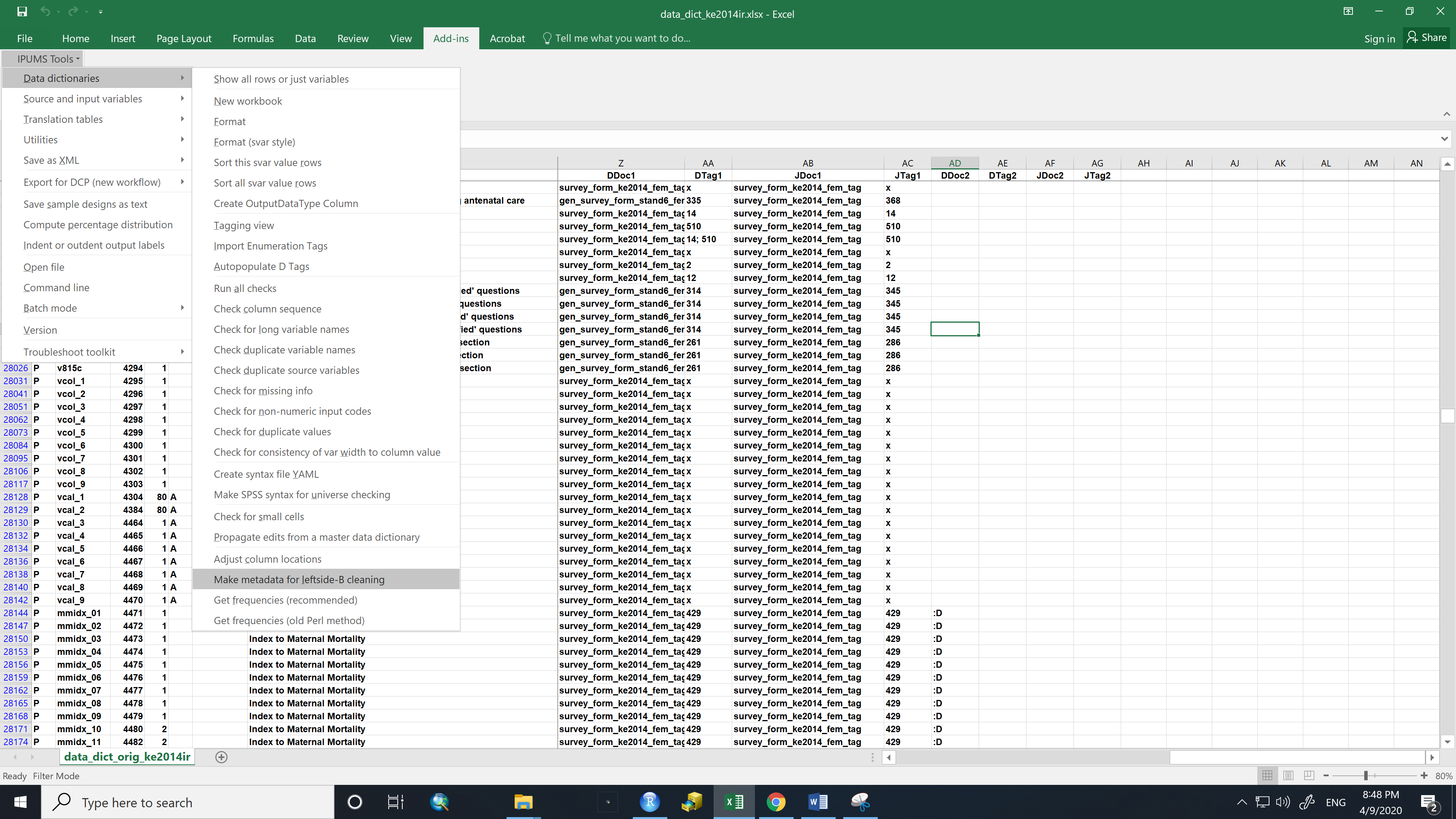


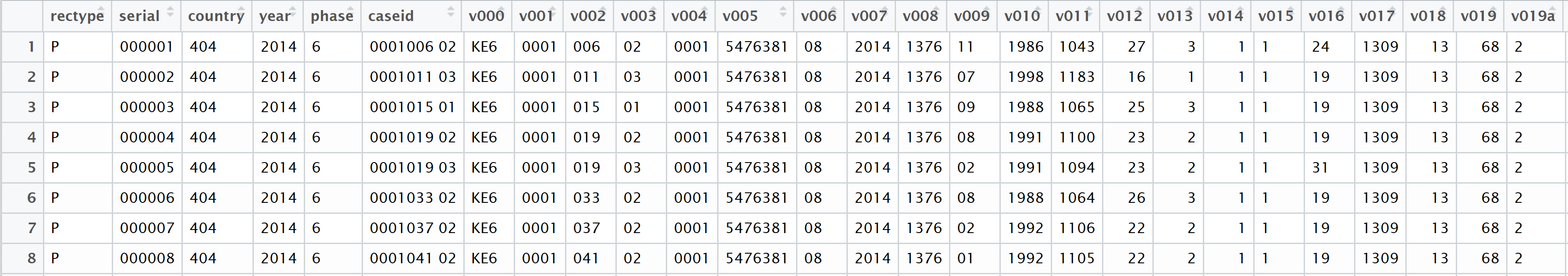
1. Calendar variables are stored in the individual recode (IR). The IR includes up to 20 births in the birth history, and up to 6 children under age 5, for whom pregnancy and postnatal care as well as immunization and health data were collected.

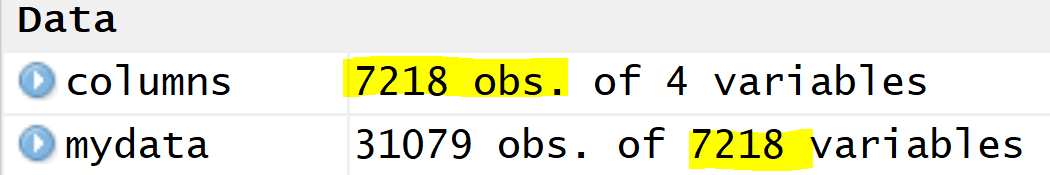
As input for data processing, we will individual recode data in .dat format from country-specific data directories on the IPUMS-DHS hard drive (e.g., DHS:\country\kenya\2014\data\ke2014ir.dat. To read in a file of this format to R, we need to get information on variables’ names (headers) and columns’ start and end position. To obtain that information, open a sample-specific data dictionary, copy and paste the following three columns (Var, Col, Wid) into a separate excel file, and save it as a .csv file. You can name the file “columns\_kenya2014.csv”, for example. This file will be read into R to get columns’ names for the IR file. Store the columns dataset inside the input folder.



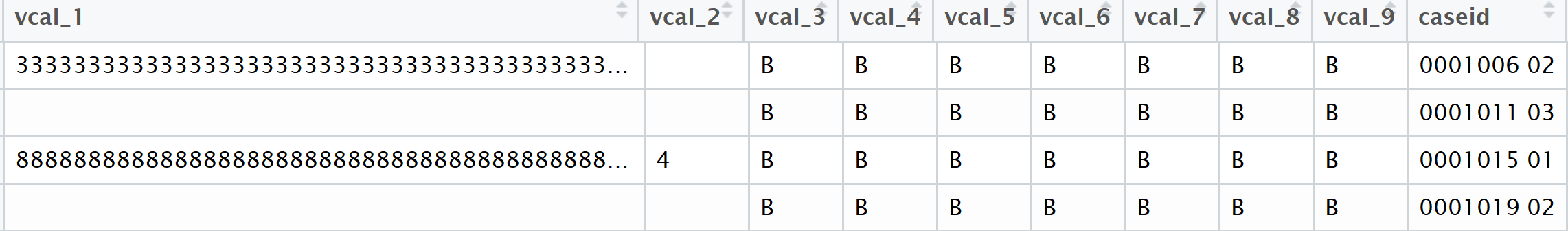
Note: Even though the “Make medata for leftside-B cleaning” (pictured below) tool creates a list of variables’ names and their positions, it does not include all the variables from the data dictionary, so we can’t use the output of that tool for column headers.



1. Read data into R – see comments in the code. There might be some warnings in the process, but they likely have to do with data types (integer, strings), so disregard them. The data should resemble something like this. 
2. Make sure the number of observations in the columns file is the same as the number of columns in the newly created dataframe – that means the data was read in correctly.



1. Next, we limit the IR data to the variables needed for processing the calendar variables. If you look at the data in R, you can see there are values in vcal\_1 and vcal\_2, whereas calendar variables vcal\_3 – vcal\_9 are populated with “B’s”, meaning they are empty. DHS creates nine calendar columns for all countries, but most samples only have vcal\_1 and vcal\_2. You can use variable v019a to check how many calendar variables with actual data there are in a sample, but also do a quick visual check. The code includes a command to check how many calendar columns there are.



We limit the data to the following variables necessary for next steps:

* 1. V006 – month of interview (Gregorian calendar for most countries)
  2. V007 – year of interview (Gregorian calendar for most countries)
  3. V008 – Century Month Code (CMC) of the interview

In Kenya 2014 DHS v008 takes on values 1373 through 1378.

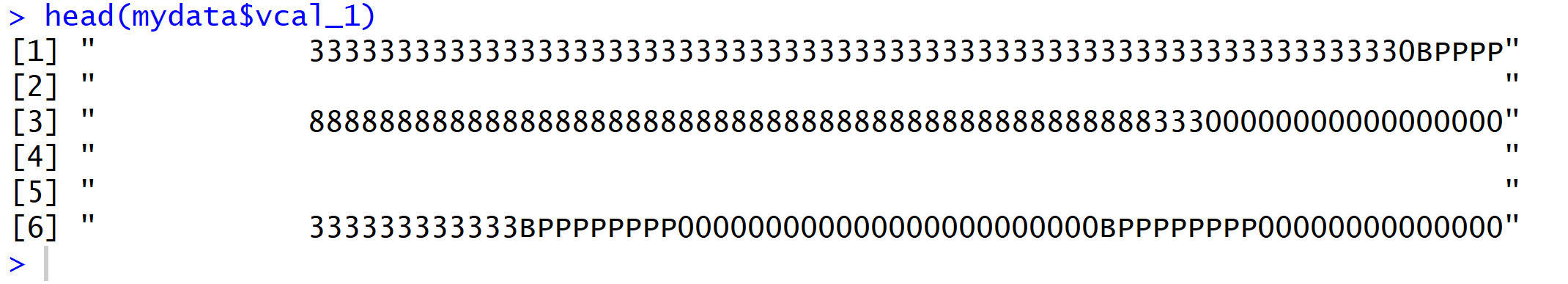
* 1. V017 - Century month code for the first month of the calendar. This variable is the same for all women and is the century month code of January of the first year of the calendar.

For Kenya 2014 it is CMC 1309 (January 2009).

CMC 1309 to Gregorian year = int(1900+1309/12)=2009

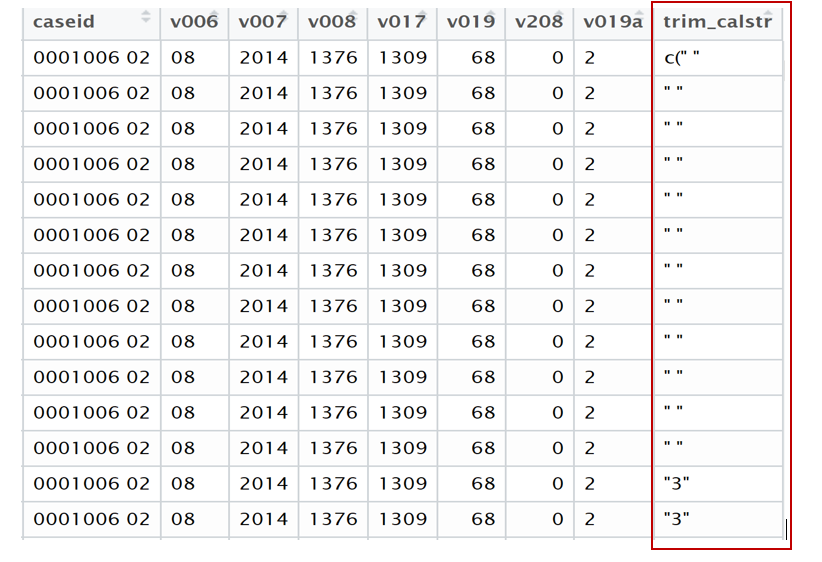
* 1. V019 – length of calendar (in months), unique for every woman. In Kenya 2014, v019 varies from 65 to 70 months. The lengths of calendars depend on when (in what months) a woman was interviewed. This variable is coded 0 for incomplete interviews.
  2. V019a – number of calendar columns with data. Again, there are 9 calendar columns possible, and even if they don’t contain any data, they are still included in the IR as empty columns.

1. Compute the length of the string in vcal\_1. This will show how many characters there are in the string, including spaces. (You will see later why keeping these spaces for now is important.) In the next step, we will transform the dataset in such a way that every calendar variable will be split into the number of rows equal to the length of that variable. Since the length of vcal\_1 is set to 80 for all women, the next step will create 80 rows for every woman.
2. Separate the first calendar variable, vcal\_1, to rows. Take a look at vcal\_1 first by printing out the first few cases:

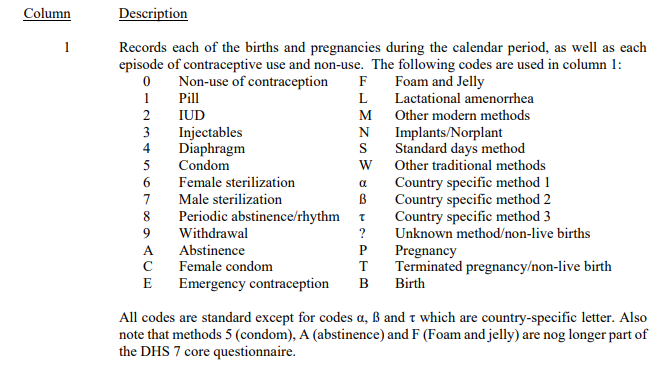


As can be seen, there are some trailing spaces. Don’t trim the trailing spaces yet even though they don’t contain any information. We will need to split other calendar variables and join the results. For the join to work correctly, the variables need to have the same length, and spaces are included in overall string length count, so keep them for now.

1. After splitting the string into rows, women-months records are stored in the column “trim\_calstr”. As you can see there are some characters that need to be removed from that column – such as the ‘c’, “ “, parentheses – these are “by-products” of splitting the string. Remove these characters by creating a new column “event” – it will represent a reproductive health event or contraceptive method relevant to a woman in a given month.

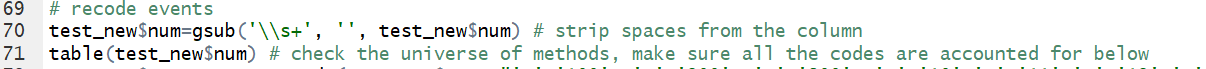


1. Compute Century Month Codes (CMC) for every month of reproductive calendar for every unique caseid (woman). Check manually the start and end CMC for a few women. Make sure the earliest CMC for all women is the same as the value of variable v017 (CMC of the start of the interview).
2. Recode codes from the original vcal\_1 variable. Right now reproductive events and contraceptive methods in vcal\_1 are represented by a combination of alphanumeric codes as shown below. Specific codes can be found in DHS Recode Manuals. In a DHS Recode Manual, search for “REC82” – this section usually represents calendar information. We will base the recoding on the codes for vcal\_1 from DHS Recode Manual VII. If there are new codes in a sample, document them in the **Calendar variables description** document and consult that document to assign the codes for other samples. Also check out **Country\_specific\_codes.txt** – a file supplied by Trevor Croft that documents country-specific codes for many samples.



We will recode the codes listed above to numeric characters. This means leaving codes 0-9 as is, and then assigning sequential numeric codes to the contraceptive methods coded with letters that follow method with code 9. This means assigning code 10 for A, 11 for C, 12 for E, and so forth (see more on this in the Calendar variable description document). Assign codes 100, 200, and 300 for birth, pregnancy, and terminations, respectively. We need all codes to be numeric to create other variables from them later on. For unknown methods/missing data, replace “?” with 99.

**An important note:** During initial IPUMS-DHS processing, sometimes ‘?’ in the original DHS data are replaced by ‘Q’. The current instructions are to only change ? to Q if there are no Qs in the data that represent actual methods of contraception/reasons for discontinuation. As such, when recoding the data, pay attention to what codes you see after running these lines of code, for variable vcal\_1 **and** vcal\_2:



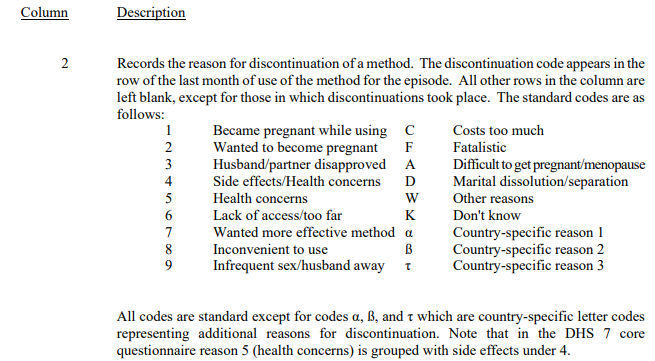
If there is a Q but no ‘?’, recode the Q to 99. If there’s just a ‘?’, recode to 99. If there are both ‘?’ and Q, Q likely represents a method/reason for discontinuation and will need to be assigned a numerical code. Check the Final Report to identify if Q stands for an actual method.

Before recoding, it is useful to check what codes for vcal\_1 a sample contains. Sometimes there are codes for contraceptive methods that are not represented in the corresponding DHS Recode Manual nor in the Individual Recode documentation. If unable to find out the meaning of the code, assign the next sequential number to it. For example, if there was code “Q” for Tanzania representing a contraceptive method, we would assign number 19 to that code.

1. Next, create the following variables based on the transformed calendar variable “event”:

* Preg - binary variable indicating whether a woman was pregnant during that CMC\_month (1-yes, 0-no)
* Birth - binary variable indicating whether a woman delivered a live birth during that CMC\_month (1-yes, 0-no)
* Term - binary variable indicating whether a woman terminated a pregnancy/delivered a non-live birth during that CMC\_month (1-yes, 0-no)
* Contr - binary variable indicating whether a woman used any contraceptive method during that CMC\_month (1-yes, 0-no)
* Total\_preg - total number of months a woman was pregnant during the entire length of her reproductive health calendar
* Count\_birth - total number of live births a woman had during the entire length of her reproductive health calendar
* Count\_term - total number of terminations/non-live births a woman had during the entire length of her reproductive health calendar
* Contr\_duration - total number of months a woman used any contraceptive methods during the length of her reproductive calendar
* Switch - binary variable indicating whether in any given month a woman was using a different contraceptive method compared to the previous month (1-yes, 0-no). This variable describes cases when women switched to a different contraceptive method *or* started using a contraceptive method prior to not using any methods. Comparison to a previous months when a woman was pregnant, gave birth, or termination does not make sense, so months after pregnancy, birth, and/or terminations will be filled with code 99 (NA).
* Switch\_new - binary variable indicating whether in a given month a woman was using a different contraceptive method compared to the month before (1-yes, 0-no). This variable describes instances of switching *from one method to another*, if a woman reported using a method in the previous month.
* duration\_1 … duration\_N - duration of use of specific contraceptive methods, where N is the numeric code for the last contraceptive method as recoded in variable Event. For example, duration\_2 refers to the total number of months a woman used an IUD (code 2) during her reproductive calendar. Duration variables are created for all the methods present in a sample.

1. Create a variable “prfirst” denoting if a woman was pregnant in the first month of her calendar. This binary variable denotes if a woman was pregnant in the first (earliest) month of her calendar (1-yes, 0-no). If a pregnancy had begun before the start of a woman’s calendar, it’s not possible to accurately determine the total duration of such pregnancy. Variable prfirst can be used to easily identify such cases. A variable denoting pregnancy in the last month of the calendar will be created at the end because we need to delete all the spaces (“ “) to create it, but we need these spaces in the meantime to join all the calendar variables.
2. We have finished processing the first variable, vcal\_1. The resulting dataframe represents women-months for every woman in the DHS sample. There is some clean-up to be done, but we will do that at the end.
3. Processing for vcal\_2 (reasons for discontinuation of contraceptive use) largely resembles that for vcal\_1. The same principle stands – we need to transform an 80-character string into women-months, remove unnecessary characters, compute CMC codes for every woman-month, recode the reasons to numeric codes, and create a few variables. To recode the original DHS code, leave codes 0-9 as is and assign sequential numbers to the alphabetic characters following the numeric codes. Again, consult the Calendar variable description document for the codes for specific reasons and document any new codes there. Below are the codes from DHS Recode Manual VII for vcal\_2.



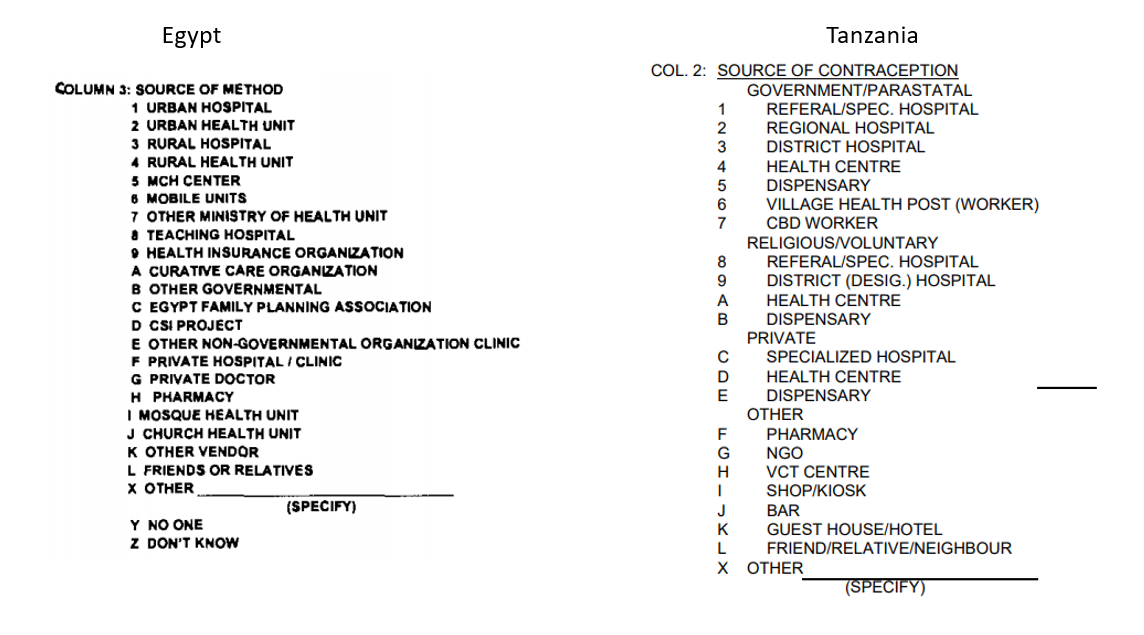
If there are reasons for discontinuation not represented in the DHS recode, assign the first sequential country-specific code to it. For example, if Tanzania 2004 contained code “Y” that represented a reason, we would assign the next sequential number – 16 – to that reason, since it is currently not represented in the codes. Remember that Q are most likely to represent ‘?’ and thus need to be recoded to 99 (more on that in the vcal\_1 section).

1. The result of transforming vcal\_2 are these variables:

* Reason - reason for discontinuation of a contraceptive method. The variable is recorded for the last CMC\_month of use of a contraceptive method. Codes for that variable are standard and can be found in a DHS Recode Manual. Just like with variable vcal\_1 and event, if there are codes in vcal\_2 that are not represented in the DHS Recode, check the sample’s Final Report to identify the meaning of the code and write it down in the Calendar Variables description document. Assign the next available number to a new code.
* Disc\_event - binary variable indicating whether a woman stopped using a contraceptive method during that CMC\_month (1-yes, 0-no)
* Disc\_total - total number of times a woman stopped using a contraceptive method throughout the entire length of her reproductive health calendar.

1. Join datasets produced by splitting columns vcal\_1 and vcal\_2 into rows.
2. Process variable vcal\_3 (marital/union status) and create a binary variable “married” indicating whether a woman was married/in a union in a given month (1-yes, 0-no).
3. Process variable vcal\_5 (source of contraception) and create variable “contr\_source” storing the codes. The source of contraception is recorded for the first month of use of a method. Variable vcal\_5 has both numeric and alphabetic codes. These codes are not standard (i.e., they are country-specific and vary by survey). An example of codes for vcal\_5 for Egypt and Tanzania is below. A description of the codes can be found in sample-specific Final Reports, usually toward the end of the document. For example, take a look at the final page of the Egypt 2000 Final Report: <https://www.dhsprogram.com/pubs/pdf/FR117/FR117.pdf>.

Assign numeric codes for these codes (see R script for that).



1. Join all the calendar variables.
2. You can skip this step if a country does not have any other calendar variables. Some countries (such as Egypt, for example) have multiple calendar variables. Egypt 2000 DHS has the following variables:

* Vcal\_1: Pregnancies & contraception
* Vcal\_2: Reason for discontinuation
* Vcal\_3: Marriage & unions
* Vcal\_5: Source of method
* Vcal\_6: Postpartum amenorrhea (months)
* Vcal\_7: Breastfeeding (months)
* Vcal\_8: Original Pregnancies & contraception (includes type of termination). This variable as the same as vcal\_1, except that it describes a specific type of termination. So, if an event is coded as T (termination of pregnancy) in vcal\_1, in vcal\_8 it will represent either A (abortion), M (miscarriage), or S (stillbirth).

As such, for Egypt, we will also create the following variables:

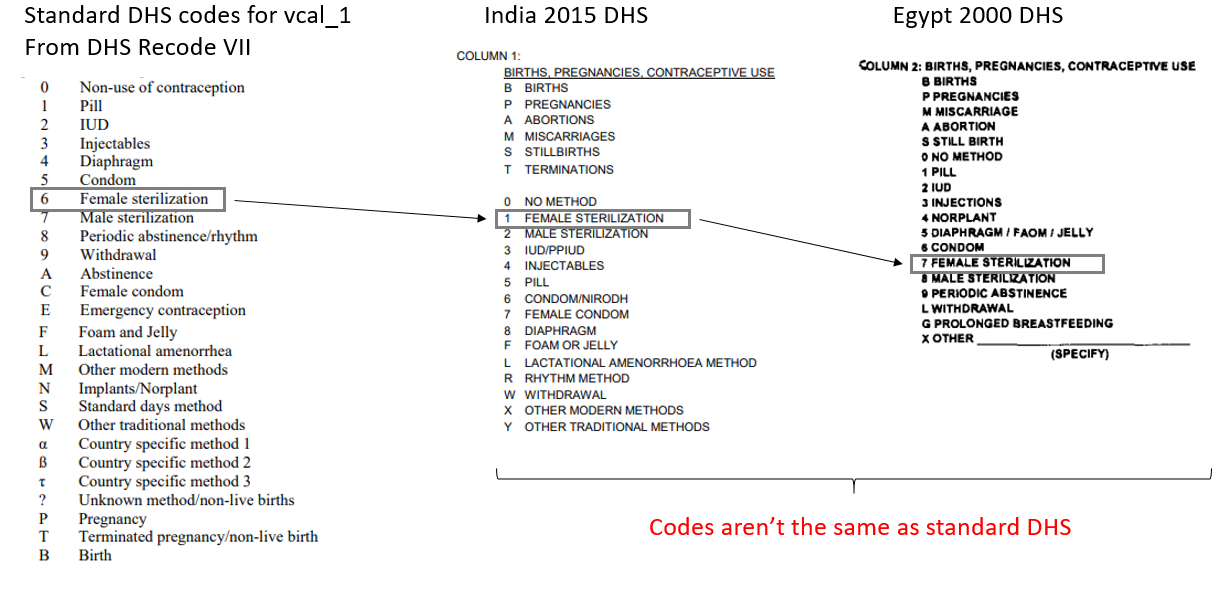
* ppam – indicates whether a woman experienced postpartum amenorrhea in that month (1-yes, 0-no) (based on vcal\_6)
* breastf - indicates whether a woman breastfed in that month (1-yes, 0-no, 2 - never breastfed) (based on vcal\_7)

variables based on vcal\_8:

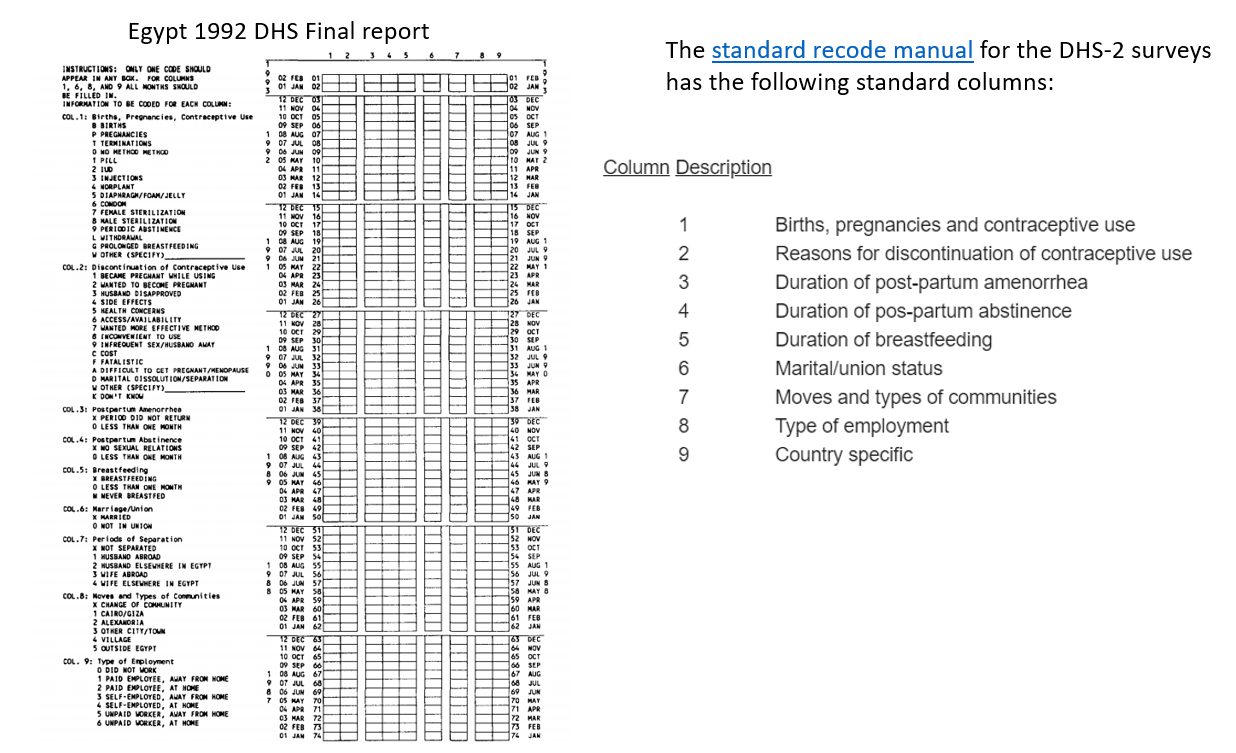
* abort - indicates whether a woman had an abortion in that month (1-yes, 0-no)
* misc - indicates whether a woman had a miscarriage in that month (1-yes, 0-no)
* sbirth - indicates whether a woman had a stillbirth in that month (1-yes, 0-no)
* count\_abort – the number of abortions a woman had during her calendar period
* count\_misc - the number of miscarriages a woman had during her calendar period
* count\_sbirth - the number of stillbirths a woman had during her calendar period

A number of samples have a calendar variable describing the type of pregnancy termination from vcal\_1. The position of that variable in the calendar recode changes from sample to sample: in Egypt 2000 it is vcal\_8, Egypt 2004 – vcal\_6, India 2015 – vcal\_7. In this variable, code T (pregnancy termination) from vcal\_1 is replaced by either A (abortion), M (miscarriage), or S (stillbirth), while the numeric codes for methods of contraception are not always standard, meaning that the codes are not always the same as in vcal\_1. The meaning of the numeric codes in this calendar variable can be found in the DHS Final Report and Recode manuals, if the codes are standard – see an image below. As such, those numeric codes can be assigned regular sequential numbers, because variable vcal\_1 already contains standardized methods codes.

**Important note**: The fact that the codes in the Final Reports differ from the standard codes in DHS Recode Manual does not apply to variable vcal\_1 because the codes for vcal\_1 have been standardized by DHS.



Refer to corresponding DHS Recode Manuals to identify the meaning of the calendar variables. The questionnaire in the report shows the structure used at the time of data collection, but not the structure of the data in the recode file. The data structure used in the recode file is designed to provide greater comparability, by using columns and codes that fit (as best as possible) to a standard system. As such, refer to the **survey-specific phase of DHS Recode Manuals**. See example below:



Some samples may have unique calendar variables not found in other samples. However, the processing logic for all the calendar variables is the same (described above). By far not all countries have country-specific calendar variables, and most countries have variables vcal\_1 and vcal\_2 only.

For example, samples for India 2005 and 2015 have a variable ultrasound:

ultrasound – a binary variable describing whether a woman had an ultrasound during her pregnancy (1-yes, 0-no). This variable is recorded in the last month of a woman’s pregnancy (i.e. the month of birth).

Vcal\_4 is another variable, available for older samples, such as Zimbabwe 1994 and Jordan 1997. It describes a woman’s residence, and we can construct the following variable from it:

Egypt 1992 DHS has variables vcal\_1 – vcal\_9, but because this is an older survey, the meaning of the variables after vcal\_2 is not the same as in more recent surveys, and can be found in DHS Recode Manual II, see page 70: (<https://dhsprogram.com/pubs/pdf/DHSG4/Recode2DHS.pdf>).

Egypt 1992 has the following variables: Vcal\_1, vcal\_2, vcal\_3 (postpartum amenorrhea), vcal\_4 (postpartum abstinence), vcal\_5 (breastfeeding), vcal\_6 (marital status), vcal\_7 (moves and communities), vcal\_8 (types of employment), vcal\_9 (periods of separation). Morocco 1992 and Jordan 1990 have the same vcal\_1 – vcal\_8 variables.

Move – categorical variable denoting whether a woman moved communities or the type of community she resided in a given month. Codes for these variables are country-specific and can be found in respective DHS Final Reports and corresponding DHS Recode. For example, Jordan 1997 has the following codes for variable move: (0-change of community, 1-Amman, 2-another city, 3-countryside/village, 4-outside Jordan). Zimbabwe 1994 has the following codes: (0-change of community, 1-main town, 2-other urban area, 3-rural area).

1. Final edits – fill columns with zeroes, 99 (=NA), etc.
   1. remove trailing spaces from vcal\_1 so that the table reflects the actual calendar length for every woman
   2. compute the difference between the length of calendar for every woman created by our processing and the original DHS length -> difference should be 0. This is just an internal check, and that variable can be deleted later.
   3. Trim spaces and assign code 0 to rows that have no data – such as assigning 0 to variable disc\_total for women who have no information in vcal\_2, for example.
   4. Create a unique id for every woman-month, based on “caseid” and “CMC\_month” (caseid\_CMC)
   5. Compute cumulative duration of every pregnancy in month (cml\_preg)
   6. Create variable “preg\_flag” if a pregnancy lasts > 9 months.
   7. Create variable “trunc” if a pregnancy is truncated at the end of the calendar. This binary variable indicates if a pregnancy is truncated at the end of the calendar (1-yes, 0-no). If a woman is pregnant during the last month of her calendar, we can’t find out the outcome of that pregnancy and include it in the data. This can lead to discrepancies between the number of months a woman has been pregnant (total\_preg) relative to the number of pregnancy outcomes (birth/termination) captured in the calendar.
2. In some samples the number of reported calendar columns in variable v019a or Individual Recode Documentation does not represent the actual number of columns. For Tanzania 2004, variable v019a indicates five calendar columns, whereas the Individual Recode Documentation lists four variables. As can be seen from the actual data, there are four columns.
3. It is useful to manually check the final recode for a few women. You can subset the final dataset to the first 10-50 women, save it, open it in Excel, and just look over it. It’s useful to freeze the top row to easily navigate across columns. Below is an example of a reproductive history for a few women for the Tanzania 2004 sample. Note that the calendar data are stored in reverse chronological order -- more recent events appear first. This is how the data are originally stored in the DHS. If you prefer for earlier events to appear first, you can sort the data by the CMC\_month variable. The reverse order essentially means that we start reading a woman’s history from the bottom up.

If we look at woman whose id=2, we can see she was using the pill during the first month of the calendar (CMC 1189). She used the pill for a total of 9 months before discontinuing the pill in month 9 (CMC 1197) of her calendar due to side effects. Since she started using a new method, injections, in month 10 (CMC 1198), and was using a different method (the pill) in CMC 1197, the woman switched from using one contraceptive method to another as indicated by codes in variables “switch” and “switch\_new” for CMC 1198. In CMC 1232 (sequential month 44 of her calendar) she stopped using the injections because of side effects (look at variable disc\_event), which amounted to two episodes of discontinuation of a contraceptive method over the course of her calendar (look at variable disc\_total). In CMC 1244, she became pregnant and gave a live birth in CMC 1252, for a pregnancy that lasted for a total of 9 months (look at variable cml\_prg). In CMC 1257 (sequential month 69 of her calendar) she started using the pill again and used it for two more months before her calendar period ended. Overall, she used the pill for 11 months (duration\_1) and injections for 35 months (duration\_3). When she started the pill again in month CMC 1257, she had not been using any contraception in the previous month, so variable “switch\_new” is set to 99, indicating that she did not switch from one contraceptive method to another without a gap in contraceptive use. Variable “switch” is coded as 1 because the woman did take up a contraceptive method (as opposed to not using any method a month before). The woman reported being in a marriage/union from CMC 1204 onwards in her calendar.

Let’s look at some other interesting cases. If you scroll down to woman with id=3, you can see that her first pregnancy lasted more than 9 months (10 months in that case), which is noted by variable “flag\_preg” and is also reflected in variable “cml\_prg” that describes the cumulative duration of pregnancy. Woman with id=9 was pregnant in the last month of her calendar, so her pregnancy is considered truncated at the end of the calendar. Woman with id=7 was pregnant in the first month of her calendar as indicated by column “prfirst”.

# Note about country-specific calendar systems (i.e., not Gregorian calendar)

Ethiopia used a different calendar that is about 92 months (~7.7 years) behind the Gregorian calendar used in Western countries. The original DHS columns v006, v007, v008, and v017 are in Ethiopian calendar. Country-specific script for Ethiopia has code for converting the dates to the Gregorian calendar. The script creates variables CMC\_g, yy\_greg, mo\_greg, and v017\_g to convert variables v006, v007, v008, and v017 to Gregorian calendar, respectively. Check our R code for Ethiopia or Nepal to see how the conversion was done.

Formula to convert the dates was obtained from the DHS User Forum: <https://userforum.dhsprogram.com/index.php?t=msg&goto=66&>.

Nepal also uses a different calendar system, called Nepali Patro. According to this website, “Nepali calendar is approximately 56 years and 8½ months ahead of the Gregorian calendar. There are 12 months, but unlike the Gregorian calendar, it has varied dates on every month, as they are not easy pre-determined and changes from year to year.” Afghanistan uses a Persian calendar, and months’ start and end dates in the Persian calendar do not align with the months in the Gregorian calendar. Because information on the day of the month is necessary to convert the Nepali/Agfhani to the Gregorian calendar, it is not possible to convert the month and year of interview with 100% certainty. However, it is possible to approximate Gregorian CMC by subtracting/adding a fixed number of months to the CMC reported in the local calendar systems.

See these notes on the DHS user forum:

<https://userforum.dhsprogram.com/index.php?t=msg&goto=18414&S=Google>

<https://userforum.dhsprogram.com/index.php?t=msg&goto=11126&S=Google>

# Link data from the Individual Recode to Calendar Recode on the server

As a final step, we need to link variables from the IR file to the transformed calendar data. Because joining the IR file to the women-months takes up quite some space and time, your local computer likely won’t be able to handle this. As such, we will do this in R on the MPC server.

1. Access R studio via <https://rstudio.gp1.pop.umn.edu/>, log in with your X500 password.
2. Copy the contents of your local package library to your instance on the server by using the code below:

> lib\_location <- file.path(

"~/R",

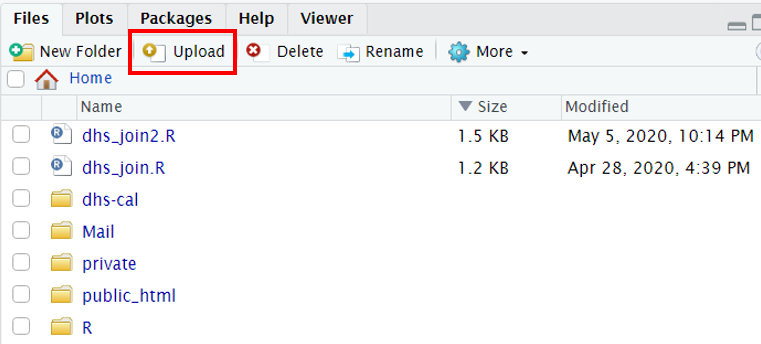
paste0(R.version$platform,"-library"),

paste0(R.version$major, ".", strsplit(R.version$minor, "\\.")[[1]][1])

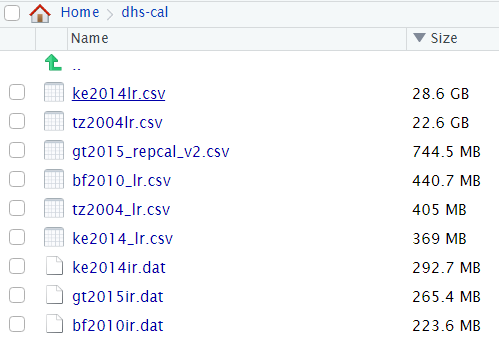
)

> Sys.setenv(R\_LIBS\_USER = lib\_location)

1. Use the Upload button to load the script (ir-to-calendar-join.R) to your home directory on the server:

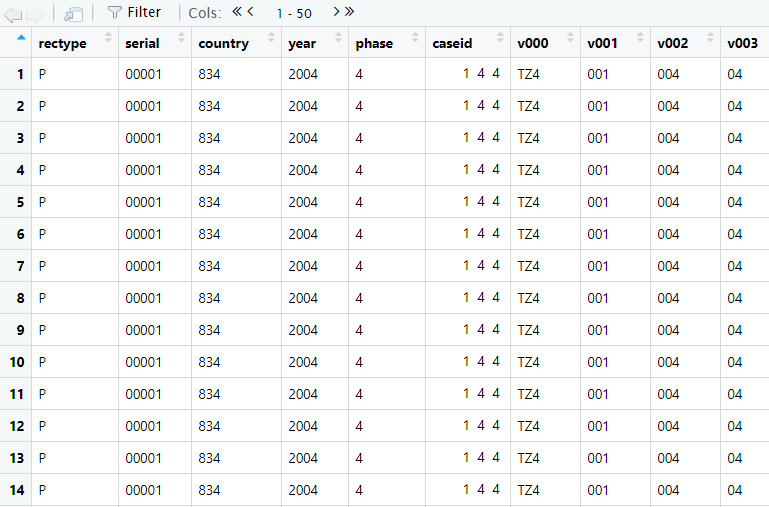


1. Once the script is loaded, create a folder (such as dhs-cal as in the image above). Use the Upload button again to upload the following:
   1. original IPUMS-DHS IR file, such as tz2004ir.dat
   2. columns names/width from the data dictionary (columns\_tz2004.csv)
   3. transformed calendar data, such as tz2004\_lr.csv
2. After the data is loaded, the folder will resemble this:

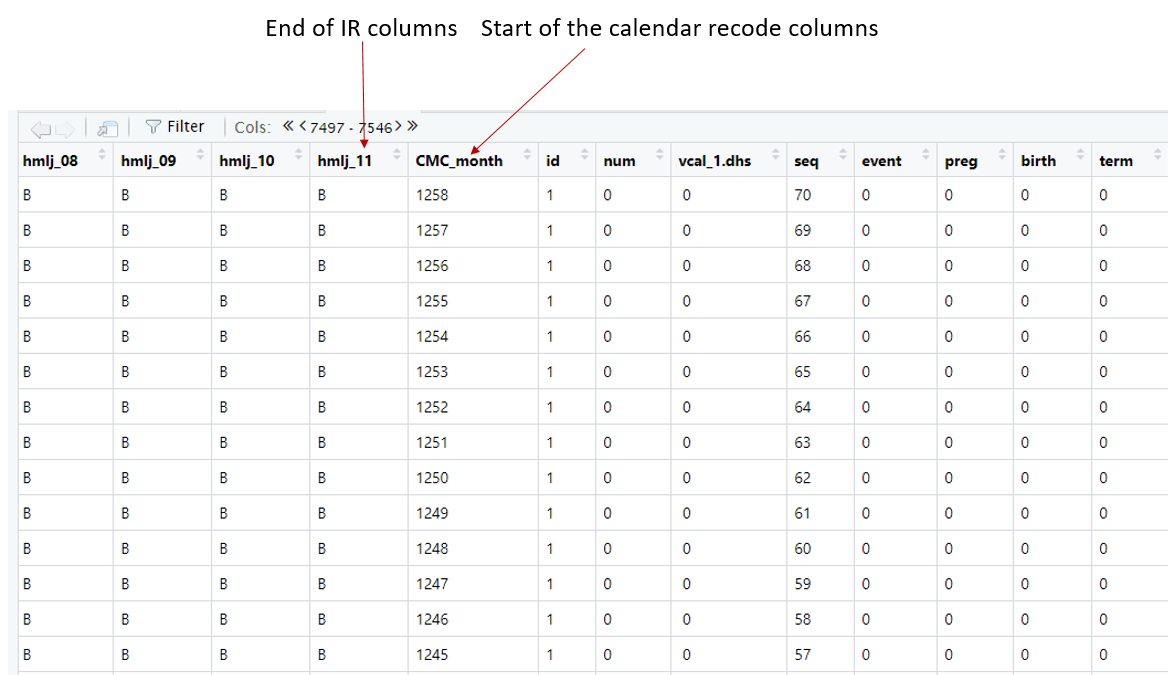


1. Run the script to join the data. Note that the join (appr. Line 24 in the code) might take quite a while to run, especially for large samples.
2. After the joining all the columns from the IR file, we need to sort the columns in order to ensure the original IR columns come first, and the newly created calendar columns appear after them. To do that, we will need to obtain the indices of certain columns’ names and rearrange them. The results will look similar to this:

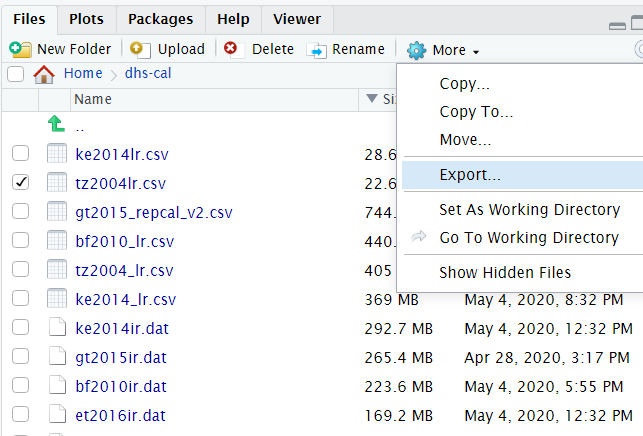
The first few columns:



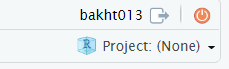
End of the IR columns and calendar columns:



1. Save output as tz2004lr. Notice I named the first iteration of the calendar data as tz2004\_lr.csv, and after joining that dataset with the original IR on the server, I named it tz2004lr.csv (removed the underscore sign).
2. You can export the file to your computer by selecting the file and clicking Export:



1. Sometimes R on the server acts up if you want to use it without previously having closed out your session. As such, at the end of your session click on the orange button to terminate your session:



# Other notes

1. Create a new script to process every sample. Once you figure out what calendar variables a sample has, copy a script for a sample with the same collection of calendar variables, save it as a script for the current sample and modify accordingly. Having a script for every sample makes it easy to quickly modify the data if needed.
2. To figure out what character was recorded to what number, use the Calendar variables description document and/or a sample-specific R script. Within the script, look at the following sections, for example:

For vcal\_1:

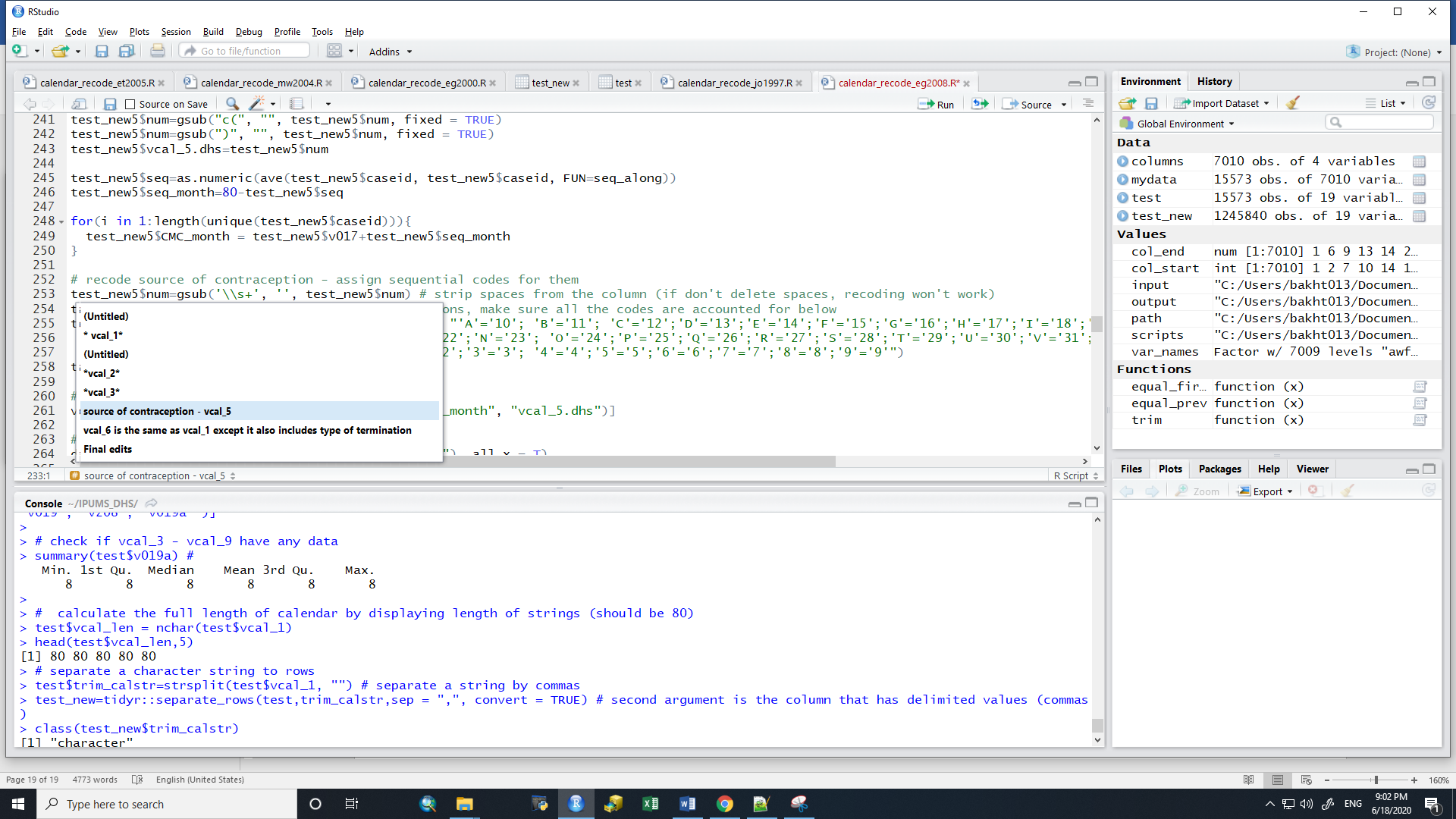


For vcal\_2:



1. Use the orange # button in R to navigate between sections of code and quickly figure out what calendar variables are represented in the code.

To create a heading for a code section, use one # sign before and four # signs after the section’s name: # section’s name ####



1. If need to modify a previously processed calendar sample and you want to do it in R, use the following code to open the previously saved output:

setwd("C:/Users/bakht013/Documents/IPUMS\_DHS") # set working directory

path="C:/Users/bakht013/Documents/IPUMS\_DHS" # set a file path

output=file.path(path,"output") # this is where previously created output is

data=read.csv(file=file.path("output", "eg2000\_lr.csv")) # read in the data

1. To read in original DHS IR file, use the following code:

library(foreign) # package to read in the data

data = read.dta("C:/Users/bakht013/Downloads/JOIR73DT/JOIR73FL.DTA", convert.factors = FALSE)

1. I processed India 2015 via R on the server because my desktop R couldn’t handle it. If someone needs to work with a large sample like this, it’s best to upload not the ir2015\_ir.dat file directly, for example, but upload its zipped version (it is located in the country data folder). You can see R script for India 2015 for more detail; you can read in a zipped file into R without unzipping it first, to save time/space. Once the file has been unzipped, it is automatically saved in the home directory on the server. Likewise, after all the processing is complete, save output as a zipped file.