cchsflow: An open science approach to transform & combine population health datasets

Kitty Chen, Warsame Yusuf, Carol Bennett, Yulric Sequeira, Douglas Manuel

December 6th, 2022

Agenda

- Background of Canadian Community Health Survey (CCHS)
- Issues relating to CCHS cycles
- *cchsflow*: development and current status
- *recodeflow*: development and current status

Canadian Community Health Survey (CCHS)

- CCHS is a population-based cross-sectional survey of Canadians that has been administered by Statistics Canada every two years since 2001.
  - Information related to health status, health care utilization and health determinants for the Canadian population.
- CCHS is one of the largest and most robust ongoing population health surveys worldwide with approximately 130,000 respondents per cycle.
- Available as public use microdata file (PUMF) from 2001-2018.

Issue

- Data cleaning, including transforming variables into harmonized or common variables, is typically the most time-consuming part of data analyses.
- With the CCHS, data cleaning and harmonization issues arise when combining CCHS surveys.
  - The names of variables change across cycles.
  - The categories change across cycles.

Categorical age:

- 2001 CCHS: DHHAGAGE, 15 categories.
- 2005 CCHS: DHHEGAGE, 16 categories.

Variable Name	DHHAGAGE	**Length**	2	**Position**	16 - 17
Question Name					
Concept	Age - (G)				
Question					
Universe	All respondents				
Note	Based on DHHA_AGE.				

Content	Code	Sample	Population
12 TO 14 YEARS	1	6,476	1,186,119
15 TO 19 YEARS	2	11,081	2,131,999
20 TO 24 YEARS	3	7,584	2,112,568
25 TO 29 YEARS	4	8,742	2,006,021
30 TO 34 YEARS	5	10,281	2,158,989
35 TO 39 YEARS	6	12,447	2,587,642
40 TO 44 YEARS	7	12,886	2,707,970
45 TO 49 YEARS	8	11,388	2,369,433
50 TO 54 YEARS	9	10,255	2,051,946
55 TO 59 YEARS	10	8,355	1,585,225
60 TO 64 YEARS	11	7,152	1,244,611
65 TO 69 YEARS	12	6,842	1,151,556
70 TO 74 YEARS	13	6,360	1,003,709
75 TO 79 YEARS	14	5,237	740,459
80 YEARS OR OLDER	15	5,794	749,088
Total		130,880	25,787,334

Variable Name	DHHEGAGE	**Length**	2	**Position**	26 - 27
Question Name	ANC_AGE				
Concept	Age - (G)				
Question	What is your age?				
Universe	All respondents				
Note	Derived from DHHE_DOB, DHHE_MOB and DHHE_YOB during interview and confirmed with respondent.				

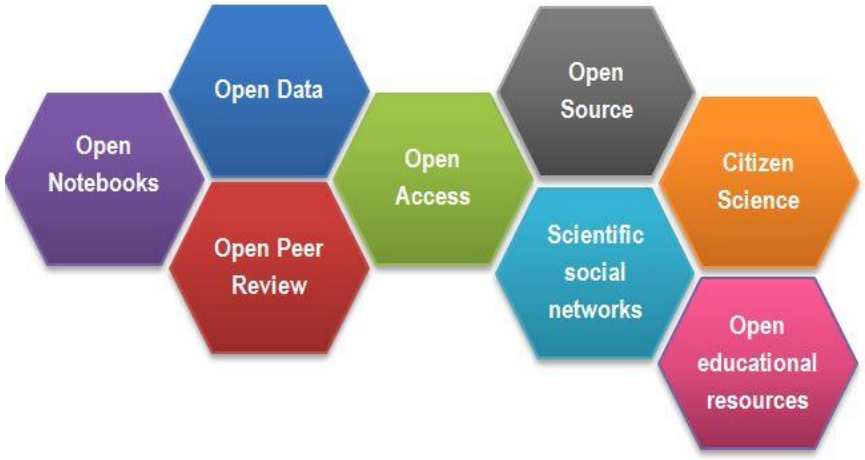
Content	Code	Sample	Population
12 TO 14 YEARS	1	6,172	1,235,378
15 TO 17 YEARS	2	6,145	1,304,374
18 TO 19 YEARS	3	3,989	813,009
20 TO 24 YEARS	4	7,740	2,239,016
25 TO 29 YEARS	5	9,227	2,103,064
30 TO 34 YEARS	6	10,252	2,079,630
35 TO 39 YEARS	7	10,058	2,282,458
40 TO 44 YEARS	8	11,172	2,803,793
45 TO 49 YEARS	9	9,143	2,532,177
50 TO 54 YEARS	10	10,296	2,251,247
55 TO 59 YEARS	11	10,645	1,973,801
60 TO 64 YEARS	12	9,268	1,580,369
65 TO 69 YEARS	13	7,846	1,213,621
70 TO 74 YEARS	14	7,124	1,030,975
75 TO 79 YEARS	15	5,961	807,900
80 YEARS OR OLDER	16	7,183	875,354
Total		132,221	27,126,165

Traditional method: hard-coding data across cycles

- Writing scripts that recode individual variables.
- Problems:
  - Labour intensive.
  - Error prone.
  - No standardized method of recoding variables.

```
  **VISIBLE MINORITY;
ethnicity_a= "Unknown";
if SDCDGCGT= 1                                    then ethnicity_a="White";
       else if SDCDGCGT = 2                       then ethnicity_a="Non-White";

** HOUSEHOLD INCOME (provincial-level);
            Income_pr_5="Unknown";
            Income_ca_5="Unknown";
         If INCDVRPR in (1, 2) then Income_pr_5= "Q1";
         If INCDVRPR in (3, 4) then Income_pr_5= "Q2";
         If INCDVRPR in (5, 6) then Income_pr_5= "Q3";
         If INCDVRPR in (7, 8) then Income_pr_5= "Q4";
         If INCDVRPR in (9, 10) then Income_pr_5= "Q5";
         If INCDVRCA in (1, 2) then Income_ca_5= "Q1";
         If INCDVRCA in (3, 4) then Income_ca_5= "Q2";
         If INCDVRCA in (5, 6) then Income_ca_5= "Q3";
         If INCDVRCA in (7, 8) then Income_ca_5= "Q4";
         If INCDVRCA in (9, 10) then Income_ca_5= "Q5";



** EDUCATION;
               Education_Cat= "Unknown";
               If EHG2DVR3= 1        then Education_Cat= "LessThanSecondary";
          else If EHG2DVR3= 2        then Education_Cat= "SecondaryGraduate";
          else if EHG2DVR3 in (3, 4) then Education_Cat= "MoreThanSecondary";

    /* Education - 5 cat*/
          Education_d = "Unknown";
          If EHG2DVR3 = 1 then Education_d =  "LessThanSecondary";
          else If EHG2DVR3 in (2, 3, 4)then Education_d = "SecondaryGraduate";
```

Open science

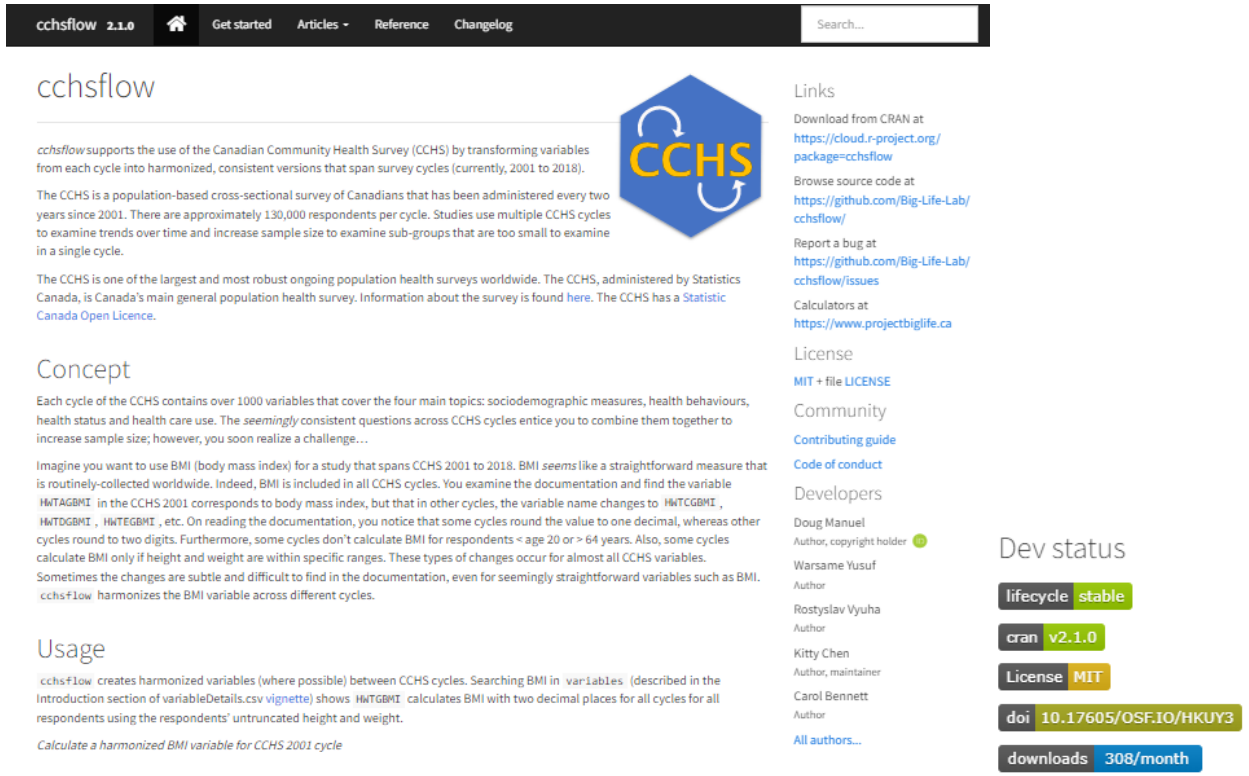


https://www.fosteropenscience.eu/content/what-open-science-introduction

- Defined as “transparent and accessible knowledge that is shared and developed through collaborative networks”. (Vicente-Saez, Martinez-Fuentes, 2018)
- Includes:
  - Sharing of data (e.g. the CCHS PUMF).
  - Use of open source languages (e.g. R, Python, Julia).
  - Sharing of code.

New approach: *cchsflow*

- R package developed in 2019 that transforms and harmonizes CCHS variables across survey cycles.
- Contains a repository of 330+ variables from 2001 to 2018 for PUMF and share files.
- Available for installation through the Comprehensive R Archive Network (CRAN).



https://big-life-lab.github.io/cchsflow/

Using *cchsflow*

- Can be used to transform individual variables or entire survey cycles.
- Provides metadata for each variable.
- Transformed datasets can then be combined to create a harmonized dataset across many years.

```
> cchs2001_BMI <- rec_with_table(cchs2001_p, variables = "HWTGBMI", notes = TRUE)
No variable_details detected.
             Loading cchsflow variable_details
Using the passed data variable name as database_name
NOTE for HWTGBMI: CCHS 2001 restricts BMI to ages 20-64. Consider using using HWTGBMI_der for
 the most concistent BMI variable across all CCHS cycles. See documentation for BMI_fun() in
 derived variables for more details, or type ?BMI_fun in the console.
NOTE for HWTGBMI: CCHS 2001 and 2003 codes not applicable and missing variables as 999.6 and
 999.7-999.9 respectively, while CCHS 2005 onwards codes not applicable and missing variables
 as 999.96 and 999.7-999.99 respectively
NOTE for HWTGBMI: Don't know (999.7) and refusal (999.8) not included in 2001 CCHS
```

Example 1: Transforming BMI in the 2001 survey cycle

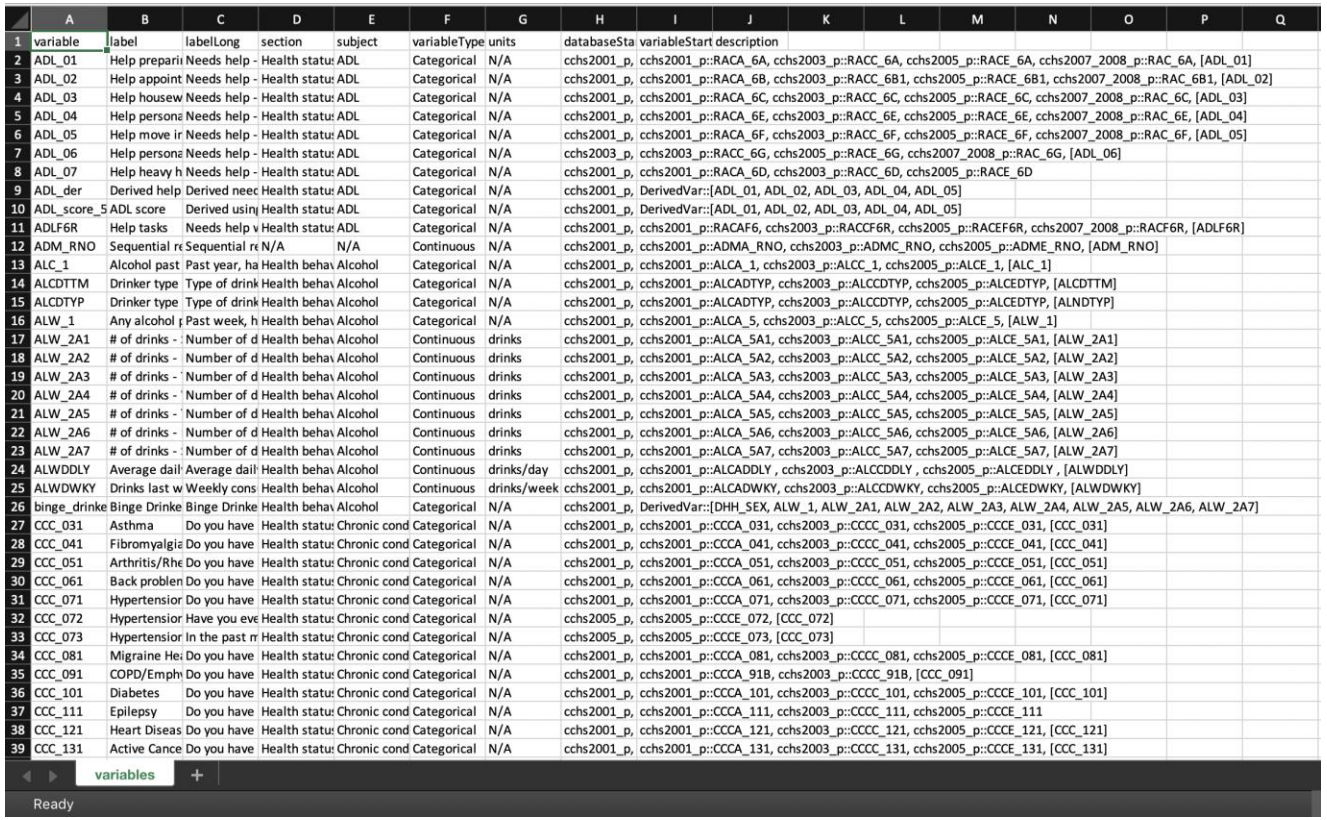
```
> cchs2001_full <- rec_with_table(cchs2001_p, notes = TRUE)
No variable_details detected.
             Loading cchsflow variable_details
No variables detected.
             Loading cchsflow variables
Using the passed data variable name as database_name
NOTE for ADL_02: In the 2001 CCHS, respondents were asked, "Because of any condition or healt
h problem, do you need the help of another person in shopping forgroceries or other necessiti
es?"
NOTE for ALCDTTM: In CCHS cycles 2001, 2003, and 2005, ALCDTTM was derived from ALCDTYP in wh
ich former and never drinkers were combined into "No drink in the last 12 months"
NOTE for ALCDTYP: Don't know (7) and refusal (8) not included in 2001 CCHS
NOTE for ALWDDLY: 2007-08, 09-10, 2010, 2012 cycles are categorical
NOTE for ALWDWKY: shown as categorical variable in CCHS 2014 cycle
```

Example 2: Transforming the entire 2001 survey cycle

Contents of the *cchsflow* package

- Specification worksheets:
  - *variables.csv* – specifies all the variables available in the package.
  - *variable_details.csv* – details about each variable (which cycles it is found, category structure etc.).
- Processing functions:
  - *rec_with_table()* – recoding of variables & survey cycles.
  - *set_data_labels()* – adding labels to transformed variables and survey cycles.
  - *merge_rec_data()* – merging and labelling transformed survey cycles.
- Derived variable functions:
  - Custom functions to generate derived variables.
- Sample data:
  - Subsets of 200 respondents for each CCHS cycle from 2001 to 2018 (PUMF).
  - Subsets of 200 respondents for each CCHS cycle for 2009 – 2012 (share files)

variables.csv



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	variable	label	labelLong	section	subject	variableType	units	databaseSta	variableStart	description							
2	ADL_01	Help preparii	Needs help -	Health statu	ADL	Categorical	N/A	cchs2001_p,	cchs2001_p::RACA_6A, cchs2003_p::RACC_6A, cchs2005_p::RACE_6A, cchs2007_2008_p::RAC_6A, [ADL_01]								
3	ADL_02	Help appoint	Needs help -	Health statu	ADL	Categorical	N/A	cchs2001_p,	cchs2001_p::RACA_6B, cchs2003_p::RACC_6B1, cchs2005_p::RACE_6B1, cchs2007_2008_p::RAC_6B1, [ADL_02]								
4	ADL_03	Help housew	Needs help -	Health statu	ADL	Categorical	N/A	cchs2001_p,	cchs2001_p::RACA_6C, cchs2003_p::RACC_6C, cchs2005_p::RACE_6C, cchs2007_2008_p::RAC_6C, [ADL_03]								
5	ADL_04	Help persona	Needs help -	Health statu	ADL	Categorical	N/A	cchs2001_p,	cchs2001_p::RACA_6E, cchs2003_p::RACC_6E, cchs2005_p::RACE_6E, cchs2007_2008_p::RAC_6E, [ADL_04]								
6	ADL_05	Help move ir	Needs help -	Health statu	ADL	Categorical	N/A	cchs2001_p,	cchs2001_p::RACA_6F, cchs2003_p::RACC_6F, cchs2005_p::RACE_6F, cchs2007_2008_p::RAC_6F, [ADL_05]								
7	ADL_06	Help persona	Needs help -	Health statu	ADL	Categorical	N/A	cchs2003_p,	cchs2003_p::RACC_6G, cchs2005_p::RACE_6G, cchs2007_2008_p::RAC_6G, [ADL_06]								
8	ADL_07	Help heavy h	Needs help -	Health statu	ADL	Categorical	N/A	cchs2001_p,	cchs2001_p::RACA_6D, cchs2003_p::RACC_6D, cchs2005_p::RACE_6D								
9	ADL_der	Derived help	Derived neec	Health statu	ADL	Categorical	N/A	cchs2001_p,	DerivedVar::[ADL_01, ADL_02, ADL_03, ADL_04, ADL_05]								
10	ADL_score_5	ADL score	Derived usin	Health statu	ADL	Categorical	N/A	cchs2001_p,	DerivedVar::[ADL_01, ADL_02, ADL_03, ADL_04, ADL_05]								
11	ADLF6R	Help tasks	Needs help v	Health statu	ADL	Categorical	N/A	cchs2001_p,	cchs2001_p::RACAF6, cchs2003_p::RACCF6R, cchs2005_p::RACEF6R, cchs2007_2008_p::RACF6R, [ADLF6R]								
12	ADM_RNO	Sequential re	Sequential re	N/A	N/A	Continuous	N/A	cchs2001_p,	cchs2001_p::ADMA_RNO, cchs2003_p::ADMC_RNO, cchs2005_p::ADME_RNO, [ADM_RNO]								
13	ALC_1	Alcohol past	Past year, ha	Health beha	Alcohol	Categorical	N/A	cchs2001_p,	cchs2001_p::ALCA_1, cchs2003_p::ALCC_1, cchs2005_p::ALCE_1, [ALC_1]								
14	ALCDTTM	Drinker type	Type of drink	Health beha	Alcohol	Categorical	N/A	cchs2001_p,	cchs2001_p::ALCADTYP, cchs2003_p::ALCCDTYP, cchs2005_p::ALCEDTYP, [ALCDTTM]								
15	ALCDTYP	Drinker type	Type of drink	Health beha	Alcohol	Categorical	N/A	cchs2001_p,	cchs2001_p::ALCADTYP, cchs2003_p::ALCCDTYP, cchs2005_p::ALCEDTYP, [ALNDTYP]								
16	ALW_1	Any alcohol p	Past week, h	Health beha	Alcohol	Categorical	N/A	cchs2001_p,	cchs2001_p::ALCA_5, cchs2003_p::ALCC_5, cchs2005_p::ALCE_5, [ALW_1]								
17	ALW_2A1	# of drinks -	Number of d	Health beha	Alcohol	Continuous	drinks	cchs2001_p,	cchs2001_p::ALCA_5A1, cchs2003_p::ALCC_5A1, cchs2005_p::ALCE_5A1, [ALW_2A1]								
18	ALW_2A2	# of drinks -	Number of d	Health beha	Alcohol	Continuous	drinks	cchs2001_p,	cchs2001_p::ALCA_5A2, cchs2003_p::ALCC_5A2, cchs2005_p::ALCE_5A2, [ALW_2A2]								
19	ALW_2A3	# of drinks -	Number of d	Health beha	Alcohol	Continuous	drinks	cchs2001_p,	cchs2001_p::ALCA_5A3, cchs2003_p::ALCC_5A3, cchs2005_p::ALCE_5A3, [ALW_2A3]								
20	ALW_2A4	# of drinks -	Number of d	Health beha	Alcohol	Continuous	drinks	cchs2001_p,	cchs2001_p::ALCA_5A4, cchs2003_p::ALCC_5A4, cchs2005_p::ALCE_5A4, [ALW_2A4]								
21	ALW_2A5	# of drinks -	Number of d	Health beha	Alcohol	Continuous	drinks	cchs2001_p,	cchs2001_p::ALCA_5A5, cchs2003_p::ALCC_5A5, cchs2005_p::ALCE_5A5, [ALW_2A5]								
22	ALW_2A6	# of drinks -	Number of d	Health beha	Alcohol	Continuous	drinks	cchs2001_p,	cchs2001_p::ALCA_5A6, cchs2003_p::ALCC_5A6, cchs2005_p::ALCE_5A6, [ALW_2A6]								
23	ALW_2A7	# of drinks -	Number of d	Health beha	Alcohol	Continuous	drinks	cchs2001_p,	cchs2001_p::ALCA_5A7, cchs2003_p::ALCC_5A7, cchs2005_p::ALCE_5A7, [ALW_2A7]								
24	ALWDDLY	Average dail	Average dail	Health beha	Alcohol	Continuous	drinks/day	cchs2001_p,	cchs2001_p::ALCADDLY , cchs2003_p::ALCCDDLY , cchs2005_p::ALCEDDLY , [ALWDDLY]								
25	ALWDWKY	Drinks last w	Weekly cons	Health beha	Alcohol	Continuous	drinks/week	cchs2001_p,	cchs2001_p::ALCADWKY, cchs2003_p::ALCCDWKY, cchs2005_p::ALCEDWKY, [ALWDWKY]								
26	binge_drinke	Binge Drinke	Binge Drinke	Health beha	Alcohol	Categorical	N/A	cchs2001_p,	DerivedVar::[DHH_SEX, ALW_1, ALW_2A1, ALW_2A2, ALW_2A3, ALW_2A4, ALW_2A5, ALW_2A6, ALW_2A7]								
27	CCC_031	Asthma	Do you have	Health statu	Chronic cond	Categorical	N/A	cchs2001_p,	cchs2001_p::CCCA_031, cchs2003_p::CCCC_031, cchs2005_p::CCCE_031, [CCC_031]								
28	CCC_041	Fibromyalgia	Do you have	Health statu	Chronic cond	Categorical	N/A	cchs2001_p,	cchs2001_p::CCCA_041, cchs2003_p::CCCC_041, cchs2005_p::CCCE_041, [CCC_041]								
29	CCC_051	Arthritis/Rhe	Do you have	Health statu	Chronic cond	Categorical	N/A	cchs2001_p,	cchs2001_p::CCCA_051, cchs2003_p::CCCC_051, cchs2005_p::CCCE_051, [CCC_051]								
30	CCC_061	Back problen	Do you have	Health statu	Chronic cond	Categorical	N/A	cchs2001_p,	cchs2001_p::CCCA_061, cchs2003_p::CCCC_061, cchs2005_p::CCCE_061, [CCC_061]								
31	CCC_071	Hypertensior	Do you have	Health statu	Chronic cond	Categorical	N/A	cchs2001_p,	cchs2001_p::CCCA_071, cchs2003_p::CCCC_071, cchs2005_p::CCCE_071, [CCC_071]								
32	CCC_072	Hypertensior	Have you eve	Health statu	Chronic cond	Categorical	N/A	cchs2005_p,	cchs2005_p::CCCE_072, [CCC_072]								
33	CCC_073	Hypertensior	In the past m	Health statu	Chronic cond	Categorical	N/A	cchs2005_p,	cchs2005_p::CCCE_073, [CCC_073]								
34	CCC_081	Migraine Hea	Do you have	Health statu	Chronic cond	Categorical	N/A	cchs2001_p,	cchs2001_p::CCCA_081, cchs2003_p::CCCC_081, cchs2005_p::CCCE_081, [CCC_081]								
35	CCC_091	COPD/Emph	Do you have	Health statu	Chronic cond	Categorical	N/A	cchs2001_p,	cchs2001_p::CCCA_91B, cchs2003_p::CCCC_91B, [CCC_091]								
36	CCC_101	Diabetes	Do you have	Health statu	Chronic cond	Categorical	N/A	cchs2001_p,	cchs2001_p::CCCA_101, cchs2003_p::CCCC_101, cchs2005_p::CCCE_101, [CCC_101]								
37	CCC_111	Epilepsy	Do you have	Health statu	Chronic cond	Categorical	N/A	cchs2001_p,	cchs2001_p::CCCA_111, cchs2003_p::CCCC_111, cchs2005_p::CCCE_111								
38	CCC_121	Heart Diseas	Do you have	Health statu	Chronic cond	Categorical	N/A	cchs2001_p,	cchs2001_p::CCCA_121, cchs2003_p::CCCC_121, cchs2005_p::CCCE_121, [CCC_121]								
39	CCC_131	Active Cance	Do you have	Health statu	Chronic cond	Categorical	N/A	cchs2001_p,	cchs2001_p::CCCA_131, cchs2003_p::CCCC_131, cchs2005_p::CCCE_131, [CCC_131]								

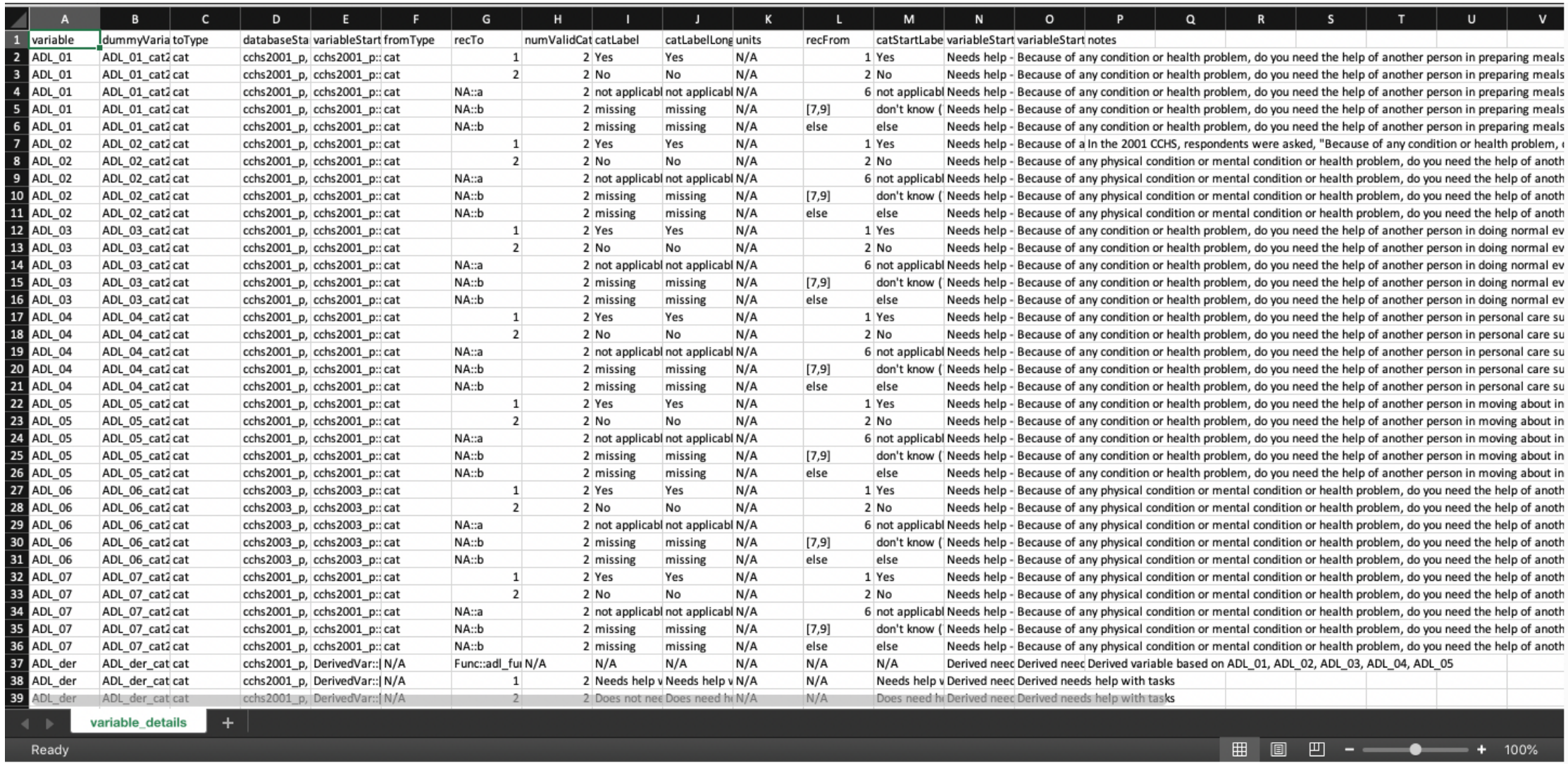
variables +

Ready

- Specifies all the variables available in *cchsflow*.
- Provides metadata of each variable.
  - Variable labels.
  - Variable type.
  - Subject and section.
  - Units (if applicable).
  - Description of variable.

variable_details.csv

- Outlines the structure of each variable.
- Guides the transformation process by:
  - Identifying the survey cycles the variable is available.
  - Specifying the original and final variable names.
  - Specifying the original and final variable types (categorical or continuous).
  - Specifying the original and final category structure (original and final ranges for continuous variables).
- Provides additional metadata of each variable.
  - Provides labels of categories.
  - Notes are provided to identify potential issues when combining between survey cycles.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	variable	dummyVaria	toType	databaseSta	variableStart	fromType	recTo	numValidCat	catLabel	catLabelLong	units	recFrom	catStartLabe	variableStart	variableStart	notes						
2	ADL_01	ADL_01_cat2	cat	cchs2001_p,	cchs2001_p::	cat	1	2	Yes	Yes	N/A	1	Yes	Needs help -	Because of any condition or health problem, do you need the help of another person in preparing meals							
3	ADL_01	ADL_01_cat2	cat	cchs2001_p,	cchs2001_p::	cat	2	2	No	No	N/A	2	No	Needs help -	Because of any condition or health problem, do you need the help of another person in preparing meals							
4	ADL_01	ADL_01_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::a	2	not applicabl	not applicabl	N/A	6	not applicabl	Needs help -	Because of any condition or health problem, do you need the help of another person in preparing meals							
5	ADL_01	ADL_01_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::b	2	missing	missing	N/A	[7,9]	don't know (	Needs help -	Because of any condition or health problem, do you need the help of another person in preparing meals							
6	ADL_01	ADL_01_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::b	2	missing	missing	N/A	else	else	Needs help -	Because of any condition or health problem, do you need the help of another person in preparing meals							
7	ADL_02	ADL_02_cat2	cat	cchs2001_p,	cchs2001_p::	cat	1	2	Yes	Yes	N/A	1	Yes	Needs help -	Because of a	In the 2001 CCHS, respondents were asked, "Because of any condition or health problem,						
8	ADL_02	ADL_02_cat2	cat	cchs2001_p,	cchs2001_p::	cat	2	2	No	No	N/A	2	No	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
9	ADL_02	ADL_02_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::a	2	not applicabl	not applicabl	N/A	6	not applicabl	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
10	ADL_02	ADL_02_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::b	2	missing	missing	N/A	[7,9]	don't know (	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
11	ADL_02	ADL_02_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::b	2	missing	missing	N/A	else	else	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
12	ADL_03	ADL_03_cat2	cat	cchs2001_p,	cchs2001_p::	cat	1	2	Yes	Yes	N/A	1	Yes	Needs help -	Because of any condition or health problem, do you need the help of another person in doing normal ev							
13	ADL_03	ADL_03_cat2	cat	cchs2001_p,	cchs2001_p::	cat	2	2	No	No	N/A	2	No	Needs help -	Because of any condition or health problem, do you need the help of another person in doing normal ev							
14	ADL_03	ADL_03_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::a	2	not applicabl	not applicabl	N/A	6	not applicabl	Needs help -	Because of any condition or health problem, do you need the help of another person in doing normal ev							
15	ADL_03	ADL_03_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::b	2	missing	missing	N/A	[7,9]	don't know (	Needs help -	Because of any condition or health problem, do you need the help of another person in doing normal ev							
16	ADL_03	ADL_03_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::b	2	missing	missing	N/A	else	else	Needs help -	Because of any condition or health problem, do you need the help of another person in doing normal ev							
17	ADL_04	ADL_04_cat2	cat	cchs2001_p,	cchs2001_p::	cat	1	2	Yes	Yes	N/A	1	Yes	Needs help -	Because of any condition or health problem, do you need the help of another person in personal care su							
18	ADL_04	ADL_04_cat2	cat	cchs2001_p,	cchs2001_p::	cat	2	2	No	No	N/A	2	No	Needs help -	Because of any condition or health problem, do you need the help of another person in personal care su							
19	ADL_04	ADL_04_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::a	2	not applicabl	not applicabl	N/A	6	not applicabl	Needs help -	Because of any condition or health problem, do you need the help of another person in personal care su							
20	ADL_04	ADL_04_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::b	2	missing	missing	N/A	[7,9]	don't know (	Needs help -	Because of any condition or health problem, do you need the help of another person in personal care su							
21	ADL_04	ADL_04_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::b	2	missing	missing	N/A	else	else	Needs help -	Because of any condition or health problem, do you need the help of another person in personal care su							
22	ADL_05	ADL_05_cat2	cat	cchs2001_p,	cchs2001_p::	cat	1	2	Yes	Yes	N/A	1	Yes	Needs help -	Because of any condition or health problem, do you need the help of another person in moving about in							
23	ADL_05	ADL_05_cat2	cat	cchs2001_p,	cchs2001_p::	cat	2	2	No	No	N/A	2	No	Needs help -	Because of any condition or health problem, do you need the help of another person in moving about in							
24	ADL_05	ADL_05_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::a	2	not applicabl	not applicabl	N/A	6	not applicabl	Needs help -	Because of any condition or health problem, do you need the help of another person in moving about in							
25	ADL_05	ADL_05_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::b	2	missing	missing	N/A	[7,9]	don't know (	Needs help -	Because of any condition or health problem, do you need the help of another person in moving about in							
26	ADL_05	ADL_05_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::b	2	missing	missing	N/A	else	else	Needs help -	Because of any condition or health problem, do you need the help of another person in moving about in							
27	ADL_06	ADL_06_cat2	cat	cchs2003_p,	cchs2003_p::	cat	1	2	Yes	Yes	N/A	1	Yes	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
28	ADL_06	ADL_06_cat2	cat	cchs2003_p,	cchs2003_p::	cat	2	2	No	No	N/A	2	No	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
29	ADL_06	ADL_06_cat2	cat	cchs2003_p,	cchs2003_p::	cat	NA::a	2	not applicabl	not applicabl	N/A	6	not applicabl	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
30	ADL_06	ADL_06_cat2	cat	cchs2003_p,	cchs2003_p::	cat	NA::b	2	missing	missing	N/A	[7,9]	don't know (	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
31	ADL_06	ADL_06_cat2	cat	cchs2003_p,	cchs2003_p::	cat	NA::b	2	missing	missing	N/A	else	else	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
32	ADL_07	ADL_07_cat2	cat	cchs2001_p,	cchs2001_p::	cat	1	2	Yes	Yes	N/A	1	Yes	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
33	ADL_07	ADL_07_cat2	cat	cchs2001_p,	cchs2001_p::	cat	2	2	No	No	N/A	2	No	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
34	ADL_07	ADL_07_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::a	2	not applicabl	not applicabl	N/A	6	not applicabl	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
35	ADL_07	ADL_07_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::b	2	missing	missing	N/A	[7,9]	don't know (	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
36	ADL_07	ADL_07_cat2	cat	cchs2001_p,	cchs2001_p::	cat	NA::b	2	missing	missing	N/A	else	else	Needs help -	Because of any physical condition or mental condition or health problem, do you need the help of anoth							
37	ADL_der	ADL_der_cat	cat	cchs2001_p,	DerivedVar::	N/A	Func::adl_fu	N/A	N/A	N/A	N/A	N/A	N/A	Derived neec	Derived neec	Derived variable based on ADL_01, ADL_02, ADL_03, ADL_04, ADL_05						
38	ADL_der	ADL_der_cat	cat	cchs2001_p,	DerivedVar::	N/A	1	2	Needs help v	Needs help v	N/A	N/A	Needs help v	Derived neec	Derived needs help with tasks							
39	ADL_der	ADL_der_cat	cat	cchs2001_p,	DerivedVar::	N/A	2	2	Does not nee	Does need h	N/A	N/A	Does need h	Derived neec	Derived needs help with tasks							

variable_details

Ready

100%

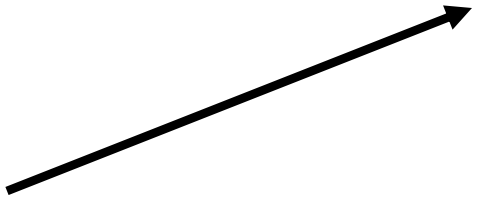
rec_with_table()

- Function used to transform variables in a given survey cycle.
- Uses information from *variables.csv* and *variable_details.csv* to transform variables.
- Can be used to transform individual variables or an entire survey cycle using all available variables in package.

```
sex2001 <- rec_with_table(cchs2001_p, "DHH_SEX",
  log = TRUE,
  var_labels = c(DHH_SEX = "SEX")
)
#> No variable_details detected.
#>               Loading cchsflow variable_details
#> Using the passed data variable name as database_name
#> The variable DHHA_SEX was recoded into DHH_SEX for the database cchs2001_p the following recodes
#>   value_to  From rows_recoded
#> 1        1     1           79
#> 2        2     2          121
#> 3    NA::a     6            0
#> 4    NA::b [7,9]            0
```

```
transformed2001 <- rec_with_table(data = cchs2001_p, notes = FALSE)
#> No variable_details detected.
#>               Loading cchsflow variable_details
#> No variables detected.
#>               Loading cchsflow variables
#> Using the passed data variable name as database_name
```

rec_with_table(data, variables, variable_details, notes)



Specify which dataset to be recoded.

Specify which variables to be recoded. Function will recode all variables in package variables sheet if not specified.

Specify a variable details sheet. Function will use package details sheet if not specified.

Specify if notes should be printed to console during recode.

Other processing functions

- *set_data_labels()*
  - Adds labels to transformed data.
  - Uses metadata specified in *variables.csv* and *variable_details.csv* to label variables.

- *merge_rec_data()*
  - Binds transformed data together into one dataset.
  - Adds labels to final transformed dataset.

```
transformed2001 <- rec_with_table(data = cchs2001_p, notes = FALSE)
#> No variable_details detected.
#>               Loading cchsflow variable_details
#> No variables detected.
#>               Loading cchsflow variables
#> Using the passed data variable name as database_name

transformed2012 <- rec_with_table(data = cchs2012_p, notes = FALSE)
#> No variable_details detected.
#>               Loading cchsflow variable_details
#> No variables detected.
#>               Loading cchsflow variables
#> Using the passed data variable name as database_name

combined_cchs <- merge_rec_data(transformed2001, transformed2012)
```

	ADL_01 Help preparing meals	ADL_02 Help appointments/errands	ADL_03 Help housework	ADL_04 Help personal care	ADL_05 Help move inside house	ADL_07 Help heavy household chores
1	2	2	2	2	2	2
2	2	1	1	2	2	1
3	2	2	2	2	2	2
4	2	2	2	2	2	1
5	2	2	2	2	2	2
6	2	1	1	2	2	1
7	2	2	2	2	2	2
8	2	2	2	2	2	2
9	2	2	2	2	2	2
10	2	2	2	2	2	2
11	2	2	2	2	2	2
12	2	2	2	2	2	1
13	2	2	2	2	2	2
14	2	2	2	2	2	2
15	2	2	2	2	2	2
16	2	2	2	2	2	2
17	2	2	2	2	2	2
18	2	2	2	2	2	2
19	2	2	2	2	2	2
20	2	2	2	2	2	2
21	2	2	2	2	2	2
22	1	2	1	1	2	1
23	2	2	2	2	2	2
24	2	2	2	2	2	2
25	1	1	1	2	2	1
26	2	2	1	2	2	1
27	2	2	2	2	2	2
28	2	2	2	2	2	2

Derived variables

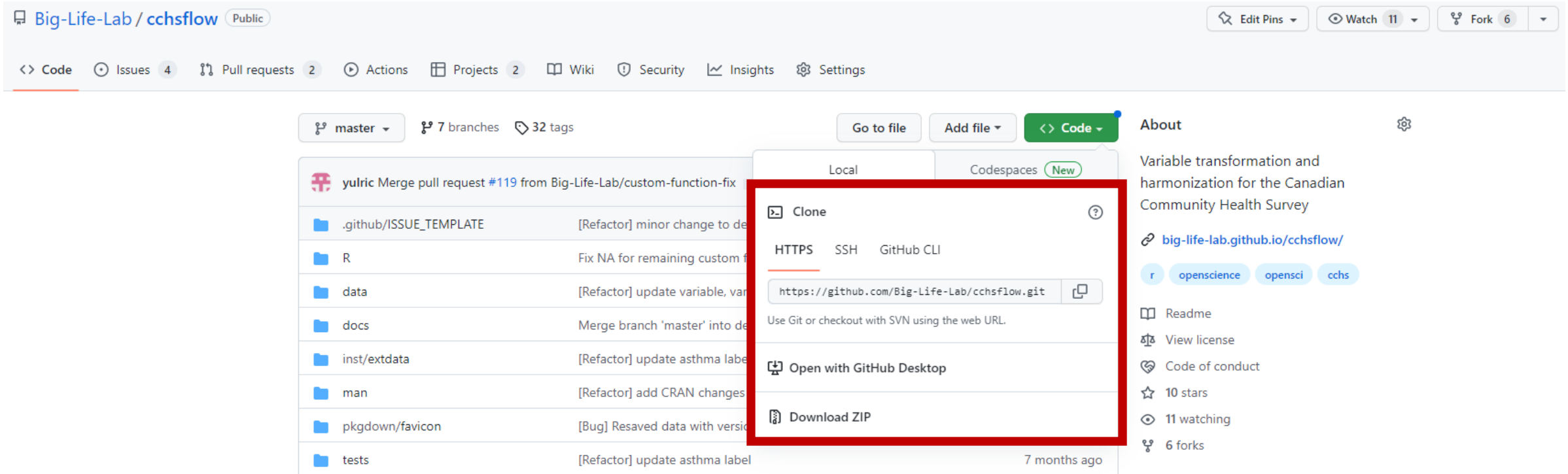
- Along with transforming existing CCHS variables, *cchsflow* can be used to create derived variables.
- Simple derived variables can be created using *variable_details.csv,* complex variables require custom functions.
- Can be based on transformed CCHS variables, other derived variables, or both.

```
adjusted_bmi_fun <-
  function(DHH_SEX, HWTGHTM, HWTGWTK) {
    # BMI adjusted for male
    if_else2(
      (!is.na(HWTGHTM)) & (!is.na(HWTGWTK)) & DHH_SEX==1,
      -1.07575 + 1.07592*(HWTGWTK / (HWTGHTM * HWTGHTM)),
      # BMI adjusted for female
      if_else2(
        (!is.na(HWTGHTM)) & (!is.na(HWTGWTK)) & DHH_SEX==2,
        -0.12374 + 1.05129*(HWTGWTK / (HWTGHTM * HWTGHTM)),
        tagged_na("b")
      )
    )
  }
```

Custom function for adjusted BMI.

Contributing to *cchsflow*

- Package is publicly available on GitHub: https://github.com/Big-Life-Lab/cchsflow
- There are two ways users can contribute to the package:
  - Cloning the repository to their computer and submitting changes via pull requests.
  - Instructions on how to add variables can be found here: https://big-life-lab.github.io/cchsflow/articles/how_to_add_variables.html
  - Using the issues section where users can request variables to be added or submit bug reports to help improve the package.


Big-Life-Lab / cchsflow
Public
Edit Pins
Watch 11
Fork 6
Code
Issues 4
Pull requests 2
Actions
Projects 2
Wiki
Security
Insights
Settings
master
7 branches
32 tags
Go to file
Add file
Code
yulric Merge pull request #119 from Big-Life-Lab/custom-function-fix
.github/ISSUE_TEMPLATE
[Refactor] minor change to de
R
Fix NA for remaining custom f
data
[Refactor] update variable, var
docs
Merge branch 'master' into de
inst/extdata
[Refactor] update asthma labe
man
[Refactor] add CRAN changes
pkgdown/favicon
[Bug] Resaved data with versi
tests
[Refactor] update asthma label
7 months ago
Local
Codespaces New
Clone
HTTPS
SSH
GitHub CLI
https://github.com/Big-Life-Lab/cchsflow.git
Use Git or checkout with SVN using the web URL.
Open with GitHub Desktop
Download ZIP
About
Variable transformation and harmonization for the Canadian Community Health Survey
big-life-lab.github.io/cchsflow/
r
openscience
opensci
cchs
Readme
View license
Code of conduct
10 stars
11 watching
6 forks

Big-Life-Lab / cchsflow Public

Edit Pins | Watch 11 | Fork 6

Code | Issues 4 | Pull requests 2 | Actions | Projects 2 | Wiki | Security | Insights | Settings

Bug report
Create a report to help us improve — Get started

Variable request
Suggest a variable for cchsflow — Get started

Big-Life-Lab / cchsflow Public

Edit Pins | Watch 11 | Fork 6

Code | Issues 4 | Pull requests 2 | Actions | Projects 2 | Wiki | Security | Insights | Settings

Label issues and pull requests for new contributors — Dismiss

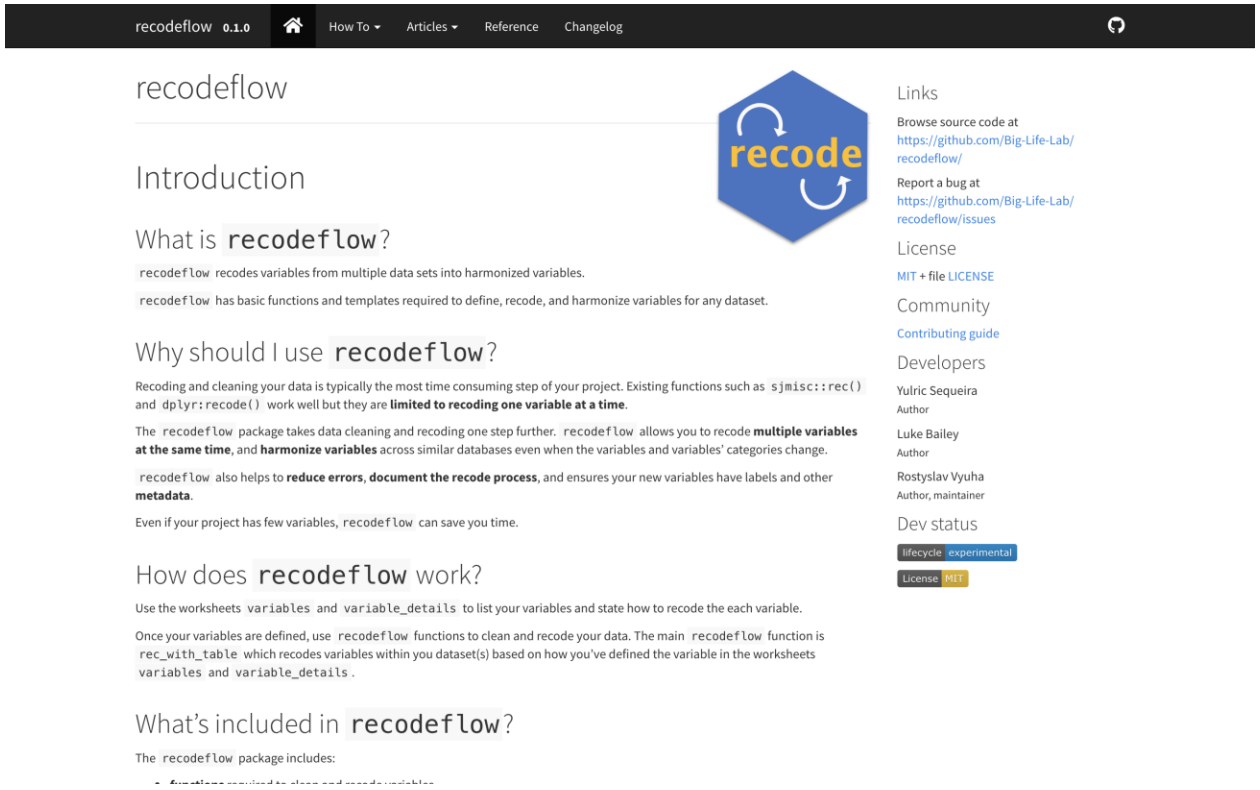
Now, GitHub will help potential first-time contributors discover issues labeled with good first issue

Filters | is:pr is:open | Labels 10 | Milestones 6 | New pull request

2 Open | 74 Closed — Author | Label | Projects | Milestones | Reviews | Assignee | Sort

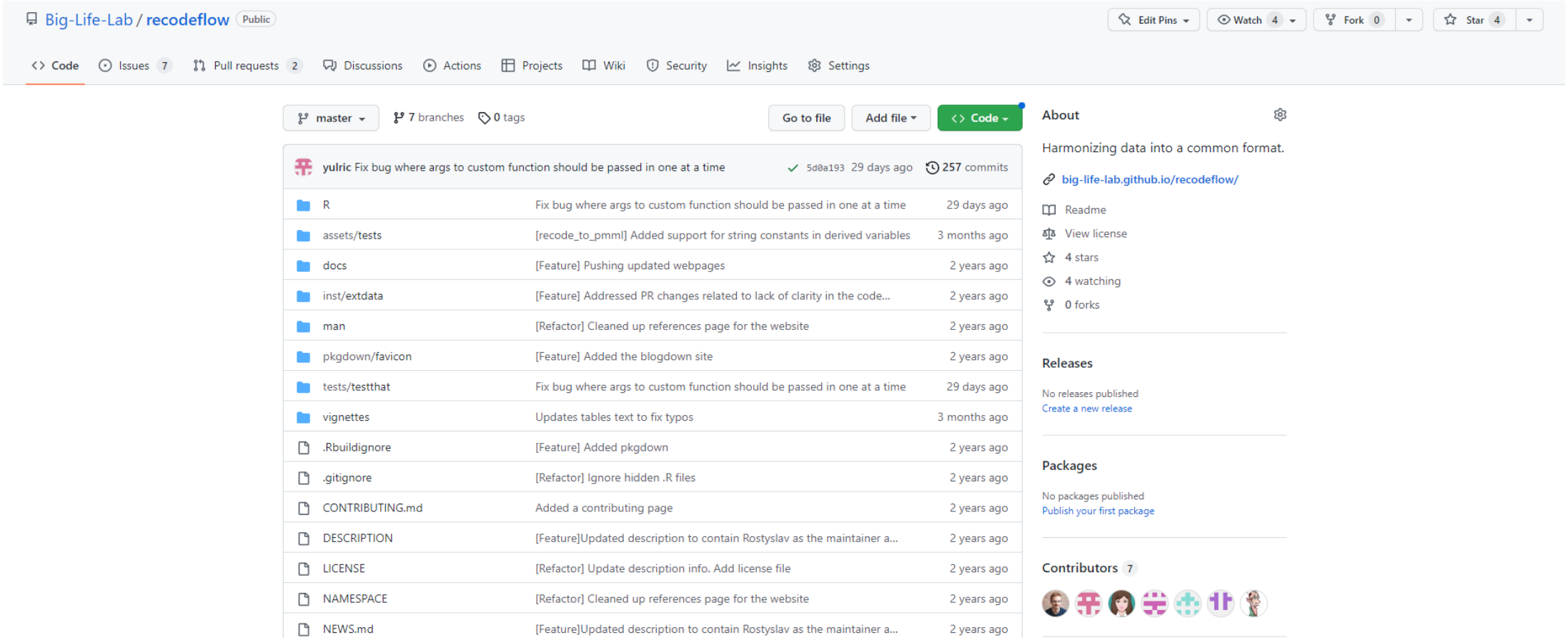
- **Sedentary activity** — 1 | 7
  #114 opened on Feb 16 by kittychenn V2.1
- **Spec Review - DO NOT MERGE** — 13
  #71 opened on Sep 28, 2020 by davidschramm • Changes requested

Transforming non-CCHS data: *recodeflow*



https://big-life-lab.github.io/recodeflow/

- While *cchsflow* can be a powerful tool in transforming variables, it is restricted to data from the CCHS.
- *recodeflow* has been developed to transform and harmonize variables from other surveys and datasets.
  - Uses the same transformation principles as *cchsflow.*
- Available for installation on CRAN.



Users can contribute to the package on the GitHub repository: https://github.com/Big-Life-Lab/recodeflow/

Summary

- *cchsflow* is an open-source package that transforms and harmonizes variables across numerous CCHS survey cycles.
- Instead of spending time recoding and transforming variables, Canadian health researchers can use *cchsflow's* existing library of variables to conduct longitudinal analyses.
- Specification worksheets are used to guide the transformation process and provide metadata of transformed variables.
- The use of GitHub allows users to contribute to the package, allowing them to add variables that may be of value to other researchers.
- *recodeflow* can be used to transform variables from other surveys and datasets using the same transformation principles.

References

- Vicente-Saez R, Martinez-Fuentes C. Open Science now: A systematic literature review for an integrated definition. Journal of Business Research 2018. https://doi.org/10.1016/j.jbusres.2017.12.043.

Questions?