

```
/*=====
=====*
```

```
* TITLE:                wb2_cleaning_final.do
```

```
* AUTHOR:                IDinsight
```

```
* LAST UPDATED:         September 11, 2020
```

```
* DESCRIPTION:           This .do file cleans the raw dataset to output
                          the cleaned dataset.
```

```
* SSC PROGRAMS:          None
```

```
* TABLE OF CONTENTS:
```

1. Set-Up
2. Renaming
3. Health Section Reconciliation
4. Labeling
5. Final Cleaning

```
* ABOUT:                 5,005 Observations
```

```
*=====
=====*/
```

```
*-----*
```

```
* 1. Set up the environment
```

```
*-----*
```

```
* Load anonymized raw data
  // Load from directory
```

```
    *global db `"" // Add your directory
```

```
    *use "$db/wb2_raw_dataset.dta", clear
```

```
/* NOTE: This raw dataset is unanonymized. All data is raw apart from the
weights variables, which were merged in separately.
```

```
*/
```

```
*-----*
```

```
* 2. Renaming
```

```
*-----*
```

```
/* NOTE: This section renames variables from the raw variable names in the
survey form to a more accessible and standardized convention
```

```
*/
```

* Multi-select questions

* agr_fert_diffs
ren (agr_fert_diffs_1 agr_fert_diffs_2 agr_fert_diffs_3 agr_fert_diffs_4
agr_fert_diffs_0 ///
agr_fert_diffs_99 agr_fert_diffs__888) (agr_fert_diffs_shut
agr_fert_diffs_nostock agr_fert_diffs_aadhar ///
agr_fert_diffs_charge agr_fert_diffs_none agr_fert_diffs_oth
agr_fert_diffs_ref)

* con_limit_lckdwn
ren (con_limit_lckdwn_1 con_limit_lckdwn_2 con_limit_lckdwn_3
con_limit_lckdwn_4 con_limit_lckdwn_0 ///
con_limit_lckdwn__888) (con_limit_lckdwn_reduce con_limit_lckdwn_out
con_limit_lckdwn_hungry ///
con_limit_lckdwn_noeat con_limit_lckdwn_none con_limit_lckdwn_ref)

* con_limit_wk
ren (con_limit_wk_1 con_limit_wk_2 con_limit_wk_3 con_limit_wk_4
con_limit_wk_0 ///
con_limit_wk__888) (con_limit_wk_reduce con_limit_wk_out
con_limit_wk_hungry ///
con_limit_wk_noeat con_limit_wk_none con_limit_wk_ref)

* lab_nowork
ren (lab_nowork_0 lab_nowork_1 lab_nowork_2 lab_nowork_3
lab_nowork_4 lab_nowork_5 ///
lab_nowork_6 lab_nowork_99 lab_nowork__888) (lab_nowork_none
lab_nowork_cant_find ///
lab_nowork_wages lab_nowork_materials lab_nowork_covid_choice
lab_nowork_fines ///
lab_nowork_sick lab_nowork_oth lab_nowork_ref)

* rel_withdraw_try
ren (rel_withdraw_try_1 rel_withdraw_try_2 rel_withdraw_try_3
rel_withdraw_try_4 ///
rel_withdraw_try_5 rel_withdraw_try__888) (rel_withdraw_notry
rel_withdraw_atm ///
rel_withdraw_biz rel_withdraw_shg rel_withdraw_unable
rel_withdraw_ref)

* rel_no_withdraw
ren (rel_no_withdraw_1 rel_no_withdraw_2 rel_no_withdraw_3
rel_no_withdraw_4 rel_no_withdraw_5 ///
rel_no_withdraw_6 rel_no_withdraw_7 rel_no_withdraw_99
rel_no_withdraw__888) (rel_nowithdraw_crowd ///
rel_nowithdraw_health rel_nowithdraw_far rel_nowithdraw_nocash
rel_nowithdraw_closed ///
rel_nowithdraw_unavail rel_nowithdraw_fail rel_nowithdraw_oth
rel_nowithdraw_ref)

```

* rel_shg_rec
  ren (rel_shg_rec_1 rel_shg_rec_2 rel_shg_rec_3 rel_shg_rec_4
rel_shg_rec_5 ///
      rel_shg_rec_6 rel_shg_rec_7 rel_shg_rec_8 rel_shg_rec_9
rel_shg_rec_10 rel_shg_rec_11 ///
      rel_shg_rec__888) (rel_shg_ppe rel_shg_san rel_shg_info rel_shg_food
rel_shg_meds ///
      rel_shg_econ_loan rel_shg_con_loan rel_shg_grains rel_shg_norec
rel_shg_none rel_shg_dk ///
      rel_shg_ref)

```

```

* rel_pds_rec
  ren (rel_pds_rec_1 rel_pds_rec_2 rel_pds_rec_3 rel_pds_rec_99
rel_pds_rec_0 ///
      rel_pds_rec__888) (rel_pds_rec_wheat rel_pds_rec_rice
rel_pds_rec_pulses rel_pds_rec_oth ///
      rel_pds_rec_none rel_pds_rec_ref)

```

```

* hea_avoid_serv
  ren (hea_avoid_serv_1 hea_avoid_serv_2 hea_avoid_serv_3
hea_avoid_serv_4 ///
      hea_avoid_serv_5 hea_avoid_serv_6 hea_avoid_serv_7 ///
      hea_avoid_serv__888) (hea_avoid_treat hea_avoid_imm hea_avoid_anc
hea_avoid_fp ///
      hea_avoid_child_grw hea_avoid_checkup hea_avoid_edu ///
      hea_avoid_ref)

```

```

drop hea_avoid_serv_99
rename hea_avoid_serv_oth hea_avoid_oth

```

```

* hea_symp
  ren (hea_symp_1 hea_symp_2 hea_symp_3 hea_symp_4 hea_symp_5
hea_symp_6 ///
      hea_symp_7 hea_symp_8 hea_symp_9 hea_symp_10 hea_symp_11
hea_symp__888) (hea_symp_fever ///
      hea_symp_cough hea_symp_tired hea_symp_breath hea_symp_pain
hea_symp_appetite ///
      hea_symp_throat hea_symp_diarrhea hea_symp_nausea
hea_symp_congest hea_symp_smell ///
      hea_symp_ref)

```

```

drop hea_symp_99

```

```

* hea_prev
  ren (hea_prev_1 hea_prev_2 hea_prev_3 hea_prev_4 hea_prev_5
hea_prev_6 ///
      hea_prev_7 hea_prev_8 hea_prev_9 hea_prev_10 hea_prev_11
hea_prev_12 hea_prev_13 ///
      hea_prev__888) (hea_prev_wash hea_prev_san hea_prev_cover

```

```

hea_prev_face ///
    hea_prev_dist hea_prev_crowd hea_prev_sick hea_prev_contact
hea_prev_touch ///
    hea_prev_clean hea_prev_mask hea_prev_nospit hea_prev_home
hea_prev_ref)

```

```

    drop hea_prev_99

```

```

* Drop half complete variables from multi-select questions
    drop *_777

```

```

*-----*

```

```

* 3. Health Section Reconciliation

```

```

*-----*

```

```

/* NOTE: In the CIFF survey, the question about health services
    taken (6.3) was coded as a multi-select question, whereas
    it was a single-select question in the WB survey.

```

The multi-select question type meant that the data listed binaries for each option selected. The following code generates a corresponding binary for each service type of the health service single-select variable (hea_service) from the WB survey.

At the end of running the code in this section, there will not be a singular variable (hea_service) to describe the health services taken. Instead, there will be a list of 7 option choices, with binaries representing each choice.

```

*/

```

```

* Variable generation

```

```

    local service_stub treat imm anc fp child_grw  checkup edu ref
    foreach stub of local service_stub {
        // Generate a new binary var for the different option choices for health services
        gen hea_service_`stub' =.
    }

```

```

    replace hea_service_ref=1 if hea_service== -888
    // Replace the value for ref

```

```

* Code single select [WB] responses into the multi-select format

```

```

    local services "treat imm anc fp child_grw  checkup edu"
    local nums "1 2 3 4 5 6 7"
    // Generate locals for the parallel loop

```

```

    forval i=1/7 {
        local service: word `i' of `services'
        local num: word `i' of `nums'
    }

```

```

replace hea_service_`service' = 0 if hea_service!=. | hea_service_ref==1
replace hea_service_`service' = 1 if hea_service==`num'
}

```

```

*-----*

```

* 4. Labeling

```

*-----*

```

```

/* NOTE: For the variables in which labels are missing, add in labels so that
the dataset is easier to understand
*/

```

```

*~~~~~*

```

* A. Multi-Select

```

*~~~~~*

```

```

* agr_fert_diffs
    label var agr_fert_diffs_shut "1.13: Shops were shut"
    label var agr_fert_diffs_nostock "1.13: Shops did not have sufficient
fertilizer stocks"
    label var agr_fert_diffs_aadhar "1.13: Aadhar authentication failed"
    label var agr_fert_diffs_charge "1.13: Retailers were overcharging"
    label var agr_fert_diffs_none "1.13: None of the above"
    label var agr_fert_diffs_oth "1.13: Other"
    label var agr_fert_diffs_ref "1.13: Refuse"

* con_limit_lckdwn
    label var con_limit_lckdwn_reduce "3.3 Reduced portion size or meals"
    label var con_limit_lckdwn_out "3.3 Ran out of food"
    label var con_limit_lckdwn_hungry "3.3 Someone was hungry and didn't
eat"
    label var con_limit_lckdwn_noeat "3.3 Someone went without eating"
    label var con_limit_lckdwn_none "3.3 None of the above"
    label var con_limit_lckdwn_ref "3.3 Refuse"

* con_limit_wk
    label var con_limit_wk_reduce "3.4 Reduced portion size or meals"
    label var con_limit_wk_out "3.4 Ran out of food"
    label var con_limit_wk_hungry "3.4 Someone was hungry and didn't eat"
    label var con_limit_wk_noeat "3.4 Someone went without eating"
    label var con_limit_wk_none "3.4 None of the above"
    label var con_limit_wk_ref "3.4 Refuse"

* lab_nowork
    label var lab_nowork_none "4.7 Do not work outside the home"
    label var lab_nowork_cant_find "4.7 Couldn't find work"
    label var lab_nowork_wages "4.7 Wages were too low"
    label var lab_nowork_materials "4.7 The required material could not be
found"
    label var lab_nowork_covid_choice "4.7 Workplace open, but hose to stay
home due to covid"

```

label var lab_nowork_fines "4.7 Workplace open, but lockdown fines"
label var lab_nowork_sick "4.7 Sick/fell ill"
label var lab_nowork_oth "4.7 Other"
label var lab_nowork_ref "4.7 Refuse"

* rel_withdraw_try
label var rel_withdraw_notry "5.5 Did not try to withdraw"
label var rel_withdraw_atm "5.5 Withdrawn using ATM/Bank branch"
label var rel_withdraw_biz "5.5 Withdrawn using business
correspondents"
label var rel_withdraw_shg "5.5 Withdrawn with the help of Bank Sakhis"
label var rel_withdraw_unable "5.5 Tried, but unable to withdraw"
label var rel_withdraw_ref "5.5 Refuse"

* rel_no_withdraw
label var rel_nowithdraw_crowd "5.6 ATM/Bank was crowded"
label var rel_nowithdraw_health "5.6 It is not safe for my health to step
out"
label var rel_nowithdraw_far "5.6 ATM/Bank is too far away"
label var rel_nowithdraw_closed "5.6 ATM/Bank was closed"
label var rel_nowithdraw_nocash "5.6 Cash is unavailable at the
ATM/Bank"
label var rel_nowithdraw_unavail "5.6 Banking correspondent was not
working/unavailable"
label var rel_nowithdraw_fail "5.6 Transaction failure"
label var rel_nowithdraw_oth "5.6 Other"
label var rel_nowithdraw_ref "5.6 Refuse"

* rel_shg_rec
label var rel_shg_ppe "5.10 Protective face masks/gloves"
label var rel_shg_san "5.10 Sanitizers/soaps"
label var rel_shg_info "5.10 Information about health and hygiene"
label var rel_shg_food "5.10 Food supplies"
label var rel_shg_meds "5.10 Medicine"
label var rel_shg_econ_loan "5.10 SHG loans for economic activities"
label var rel_shg_con_loan "5.10 SHG loans for consumption activities"
label var rel_shg_grains "5.10 Procured grains with SHG help"
label var rel_shg_norec "5.10 Did not receive"
label var rel_shg_none "5.10 No SHGs in village"
label var rel_shg_dk "5.10 Don't know about SHGs"
label var rel_shg_ref "5.10 Refuse"

* rel_pds_rec
label var rel_pds_rec_wheat "5.12 Wheat"
label var rel_pds_rec_rice "5.12 Rice"
label var rel_pds_rec_pulses "5.12 Pulses"
label var rel_pds_rec_oth "5.12 Others"
label var rel_pds_rec_none "5.12 Did not receive anything"
label var rel_pds_rec_ref "5.12 Refuse"

* hea_symp

label var hea_symp_fever "6.6 Fever"
label var hea_symp_cough "6.6 Cough"
label var hea_symp_tired "6.6 Tiredness"
label var hea_symp_breath "6.6 Difficulty breathing"
label var hea_symp_pain "6.6 Muscle pain/body aches"
label var hea_symp_appetite "6.6 Loss of appetite"
label var hea_symp_throat "6.6 Sore throat"
label var hea_symp_diarrhea "6.6 Diarrhea"
label var hea_symp_nausea "6.6 Nausea"
label var hea_symp_congest "6.6 Nasal and throat congestion"
label var hea_symp_smell "6.6 Loss of smell and taste"
label var hea_symp_ref "6.6 Refuse"
label var hea_symp_oth "6.6 Other"

* hea_prev

label var hea_prev_wash "6.7 Wash hands frequently"
label var hea_prev_san "6.7 Use alcohol-based hand sanitizer"
label var hea_prev_cover "6.7 Cover nose and mouth with
handkerchief/tissue/elbow"
label var hea_prev_face "6.7 Avoid touching your face, eyes, nose, or
mouth"
label var hea_prev_dist "6.7 Be at least one meter away from everyone"
label var hea_prev_crowd "6.7 Avoid crowded places"
label var hea_prev_sick "6.7 Stay away from people who sneeze/cough"
label var hea_prev_contact "6.7 Avoid physical contact with infected
individuals"
label var hea_prev_touch "6.7 Avoid touching common surfaces"
label var hea_prev_clean "6.7 Keep cleaning common surfaces"
label var hea_prev_mask "6.7 Wear a mask"
label var hea_prev_nospit "6.7 Don't spit in public"
label var hea_prev_home "6.7 Stay at home"
label var hea_prev_ref "6.7 Refuse"
label var hea_prev_oth "6.7 Other"

* hea_avoid

label var hea_avoid_treat "6.5 Medical treatment for self or household
member"
label var hea_avoid_imm "6.5 Immunization"
label var hea_avoid_anc "6.5 Antenatal Care"
label var hea_avoid_fp "6.5 Family planning"
label var hea_avoid_child_grw "6.5 Growth monitoring of child"
label var hea_avoid_checkup "6.5 Health check-up"
label var hea_avoid_edu "6.5 Health education"
label var hea_avoid_ref "6.5 Refuse"
label var hea_avoid_oth "6.5 Other"

* hea_service

label var hea_service_treat "6.3 Medical treatment for self or household
member"

```

label var hea_service_imm "6.3 Immunization"
label var hea_service_anc "6.3 Antenatal Care"
label var hea_service_fp "6.3 Family planning"
label var hea_service_child_grw "6.3 Growth monitoring of child"
label var hea_service_checkup "6.3 Health check-up"
label var hea_service_edu "6.3 Health education"
label var hea_service_ref "6.3 Refuse"
label var hea_service_oth "6.3 Other"

```

```
* ~~~~~*
```

```
* B. Misc Vars
```

```
* ~~~~~*
```

```
* Demo vars merged in from NRLP sample
```

```

label var demo_fuel_kerosene "Fuel source: Kerosene"
label var demo_fuel_firewood "Fuel source: Firewood"
label var demo_fuel_cropres "Fuel source: Crop residue"
label var demo_fuel_coal "Fuel source: Coal"
label var demo_fuel_dung "Fuel source: Dung"
label var demo_fuel_electricity "Fuel source: Electricity"
label var demo_fuel_dk "Fuel source - don't know"
label var demo_fuel_ref "Fuel source - ref"
label var demo_resp_gen "Gender of respondent who picked up"
label var demo_age "Respondent age"

```

```
* Section-level completions
```

```

label var agr_complete "Agriculture section fully-completed"
label var con_complete "Consumption section fully-completed"
label var mig_complete "Migration section fully-completed"
label var lab_complete "Labor & Income section fully-completed"
label var hea_complete "Health section fully-completed"
label var rel_complete "Relief section fully-completed"
label var hea_complete "Health section fully-completed"
destring fully_complete, replace
label define fully_complete 0 "Partially Complete" 1 "Fully Complete"
label var fully_complete "Completeness of survey"

```

```
* Other renaming
```

```

label var sampling_frame "Sampling frame"
label var demo_gen_ag "Gender of agriculture section respondent"
label var demo_gen_nonag "Gender of non-agriculture sections

```

```
respondent"
```

```
* ~~~~~*
```

```
* 5. Final Cleaning
```

```
* ~~~~~*
```

```
* Sampling frame
```

```
replace sampling_frame = "CIFF" if state_id==8 | state_id==20
```

```
* Respondent gender - agriculture
```



```
replace demo_gen_ag = "Male" if (demo_resp_gen==1 &  
demo_ag_dec_mkr==1) | (demo_resp_gen==0 & demo_ag_dec_mkr==1 &  
demo_agdm_consent==1)
```

* Order

```
order sampling_frame state district psu* hh_id_long_anon_str consent  
demo_ag_hh demo*, first  
order fully_complete, last
```

* Save

```
save "$db/wb2_cleaned_dataset.dta", replace
```