

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

```
* TITLE:                wb1_cleaning_final.do
```

```
* AUTHOR:                IDinsight
```

```
* LAST UPDATED:         July 28, 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:                 4,550 Observations & 200 Variables
```

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

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

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

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

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

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

```
    use "$db/wb1_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
```

```
*/
```

```
* ~~~~~*
* A. Multi-Select
* ~~~~~*
```

// These are the questions in which multiple answer choices could be selected

```
* Reasons for not working
  rename reason_nowork lab_nowork_reasons
  ren (reason_nowork_0 reason_nowork_1 reason_nowork_2
reason_nowork_3 ///
      reason_nowork_4 reason_nowork_5 reason_nowork_6
reason_nowork_99 ///
      reason_nowork__999 reason_nowork__888) ///
  (lab_nowork_none lab_nowork_cant_find lab_nowork_wages ///
   lab_nowork_shut lab_nowork_materials lab_nowork_covid_choice ///
   lab_nowork_fines lab_nowork_other lab_nowork_dk lab_nowork_ref)
  drop lab_nowork_other
  rename reason_nowork_oth lab_nowork_oth
```

```
* Reasons for not selling crops
  rename sell_start agr_nosell_reasons
  ren (sell_start_1 sell_start_2 sell_start_3 sell_start_4 ///
      sell_start_5 sell_start_6 sell_start_7 sell_start_8 ///
      sell_start_9 sell_start_10 sell_start_99 ///
      sell_start__999 sell_start__888) ///
  (agr_nosell_notready agr_nosell_seeds agr_nosell_destroyed ///
   agr_nosell_demand agr_nosell_markets agr_nosell_prices ///
   agr_nosell_lckdwn agr_nosell_labor agr_nosell_transport ///
   agr_nosell_machinery agr_nosell_other agr_nosell_dk
agr_nosell_ref)
  drop agr_nosell_other
  rename sell_start_other agr_nosell_oth
```

```
* Challenges in selling crops
  rename sell_diff agr_sellcrops_difficulties
  ren (sell_diff_0 sell_diff_1 sell_diff_2 sell_diff_3 sell_diff_4 sell_diff_5 ///
      sell_diff_6 sell_diff_7 sell_diff_99 sell_diff__999 sell_diff__888) ///
  (agr_selldiff_none agr_selldiff_labor agr_selldiff_transport
agr_selldiff_markets ///
      agr_selldiff_demand agr_selldiff_travel agr_selldiff_police
agr_selldiff_prices ///
      agr_selldiff_other agr_selldiff_dk agr_selldiff_ref)
  drop agr_selldiff_other
  rename sell_diff_other agr_selldiff_oth
```

```
* State of ration shops
  rename price_change rel_ration_prc_changes
  ren (price_change_0 price_change_1 price_change_2 price_change_3
price_change_4 ///
      price_change_5 price_change_6 price_change__999
price_change__888) ///
```

```

        (rel_ration_prc_closed rel_ration_prc_nogo rel_ration_prc_none
rel_ration_prc_free ///
        rel_ration_prc_cheaper rel_ration_prc_exp rel_ration_prc_unavail
rel_ration_prc_dk ///
        rel_ration_prc_ref)

```

```

* Health care services avoided
    rename service_avoided hea_services_avoided
    ren (service_avoided_1 service_avoided_2 service_avoided_3
service_avoided_4 ///
        service_avoided_5 service_avoided_6 service_avoided_7
service_avoided_99 ///
        service_avoided__999 service_avoided__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_other ///
    hea_avoid_dk hea_avoid_ref)

```

```

    drop hea_avoid_other
    ren service_avoided_oth hea_avoid_oth

```

```

* COVID symptoms known
    ren symptoms hea_symptoms
    ren (symptoms_1 symptoms_2 symptoms_3 symptoms_4 symptoms_5
symptoms_6 symptoms_7 ///
        symptoms_8 symptoms_9 symptoms_10 symptoms_11
symptoms_99 symptoms__999 symptoms__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_other
hea_symp_dk hea_symp_ref)
    rename symptoms_oth hea_symp_oth
    drop symptoms__99 hea_symp_other

```

```

* Prevention methods to protect against COVID
    ren prevention hea_prevention_methods
    ren (prevention_1 prevention_2 prevention_3 prevention_4 prevention_5
prevention_6 ///
        prevention_7 prevention_8 prevention_9 prevention_10
prevention_11 prevention_12 ///
        prevention_13 prevention_99 prevention__999 prevention__888) ///
    (hea_prevent_wash hea_prevent_san hea_prevent_cover
hea_prevent_face hea_prevent_dist ///
        hea_prevent_crowd hea_prevent_avoid_sick hea_prevent_contact
hea_prevent_touch ///
        hea_prevent_clean hea_prevent_mask hea_prevent_nospit
hea_prevent_home hea_prevent_other ///
        hea_prevent_dk hea_prevent_ref)

```

```
drop hea_prevent_other
rename prevention_oth hea_prevent_oth
```

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

```
* B. Standardize
```

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

```
// Here, we standardize a naming convention in which pre-fixes link to the
questionnaire module
```

```
* Module 0: Demographic/geographic information
```

```
rename state_id geo_state
rename district_id geo_dist
rename state_new geo_state_new
rename district_new1 geo_dist_new
rename village geo_village
rename block geo_block
rename gender demo_gender
rename age demo_age
rename edu demo_edu
rename caste demo_caste
rename caste_oth demo_caste_oth
rename religion demo_religion
rename religion_oth demo_religion_oth
rename self_help demo_selfhelp
rename cultivate_land demo_ag_hh
rename hh_size demo_hh_size
```

```
* Module 1: Migration
```

```
rename return_size mig_return_size
rename not_return mig_no_return_size
rename mostinc mig_most_inc
rename wage_avg_resp mig_avg_wage
rename inc_freq mig_wage_unit
rename wage_rec_resp mig_daily_wage
rename inc_freq1 mig_daily_wage_unit
```

```
* Module 3: Consumption
```

```
rename con_month con_feb
rename con_week con_wk
rename price_week_atta con_wk_atta
rename price_week_onions con_wk_onions
rename price_week_rice con_wk_rice
rename price_curr_atta con_curr_atta
rename price_curr_onions con_curr_onions
rename price_curr_rice con_curr_rice
```

```
* Module 2: Labor and Income
```

```
rename occup_march lab_march_occu
rename occup_march_oth lab_march_occu_oth
rename earning_march lab_march_wage
```

```
rename days_march lab_march_freq
rename occup_now lab_curr_occu
rename earning_now lab_curr_wage
rename days_now lab_curr_freq
```

\* Module 4: Agriculture

```
rename crop_grown agr_crop_grown
rename crop_grown_oth agr_crop_grown_oth
rename primary_crop agr_prim_crop
rename primary_crop_time agr_crop_planted
rename sell_act agr_crop_status
rename sell_over agr_crop_sold
rename price_holi agr_prc_holi
rename unit_qty_holi agr_prc_holi_unit
rename unit_qty_holi_oth agr_prc_holi_unit_oth
rename price_last_week agr_prc_curr
rename unit_qty_last_week agr_prc_curr_unit
rename unit_qty_last_week_oth agr_prc_curr_unit_oth
rename price_last_yr agr_prc_yr
rename unit_qty_yr agr_prc_yr_unit
rename unit_qty_yr_oth agr_prc_yr_unit_oth
rename sell_qty_ldwn agr_sell_lckdwn
rename sell_qty_prevyr agr_sell_prevyr
rename land_next_ssn agr_next_ssn_land
rename unit_land_next agr_next_ssn_unit
rename unit_land_next_oth agr_next_ssn_unit_oth
rename inputs_next_ssn agr_next_ssn_inputs
rename borrow_nxt_ssn agr_next_ssn_borrow
rename land_prev_yr agr_land_prevyr
rename unit_land_prev agr_land_prevyr_unit
rename unit_land_prev_oth agr_land_prevyr_unit_oth
rename inputs_prevyr_m agr_monsoon_inputs
rename borrow_prevyr_m agr_monsoon_borrow
rename land_curr_ssn agr_curr_land
rename unit_land_curr agr_curr_land_unit
rename unit_land_curr_oth agr_curr_land_unit_oth
rename inputs_curr_season agr_curr_inputs
rename borrow_curr_season agr_curr_borrow
rename land_prevyr_pcrop agr_prevyr_land_main
rename unit_land_pcrop agr_prevyr_land_main_unit
rename inputs_prevyr_pcrop agr_prevyr_main_inputs
rename borrow_prevyr_pcrop agr_prevyr_main_borrow
```

\* Module 5. Relief

```
rename bank_ac rel_bank
rename rec_govt_transfer rel_transfer_rec
rename check_bal rel_check_balance
rename amt_govt_transfer rel_transfer_amt
rename transfer_mem rel_transfer_mem
rename nrega_avl rel_nrega
```

\* Module 6. Health

```
rename health_fac_visit hea_fac_visit
rename health_fac_type hea_fac_type
rename health_fac_type_oth hea_fac_type_oth
rename service_taken hea_service
```

\* Weights

```
rename raked_wgt weight_raked
```

\*-----\*

\* 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.

\*/

\* Renaming

```
rename service_taken__999 service_taken_999
rename service_taken__888 service_taken_888
```

\* Generate parallel locals for the different option choices for health services

```
local services_long "Treatment Immunization AntenatalCare
FamilyPlanning ChildGrowth HealthCheck Education DontKnow Refuse"
local nums "1 2 3 4 5 6 7 999 888"
local services "treat imm anc fp child_grw health edu dk ref"
```

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

```
forval i=1/9 {
    local service: word `i' of `services'
    local service_long: word `i' of `services_long'
    local num: word `i' of `nums'

    rename service_taken_`num' hea_service_`service'
    replace hea_service_`service' = 0 if hea_service!=.
    replace hea_service_`service' = 1 if hea_service==`num'
```

```

label var hea_service_`service' "6.3 Health service - `service_long'"
}
/* This loop generates a corresponding binary variable in each
service bucket based on the response to "health service taken"
[which was coded as a single-select in the WB data]

```

The code also relabels the multi-select variables

```

*/

*-----*
* 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
*~~~~~*
// These are the questions in which multiple answer choices could be selected

* Reasons for not working
label var lab_nowork_reasons "2.7 Reasons for not finding work"
label var lab_nowork_none "2.7 Not working - Do not work outside
home"
label var lab_nowork_cant_find "2.7 Not working - Couldn't find work"
label var lab_nowork_wages "2.7 Not working - Wages were too low"
label var lab_nowork_shut "2.7 Not working - Workplace shut due to
lockdown"
label var lab_nowork_materials "2.7 Not working - Required material not
found"
label var lab_nowork_covid_choice "2.7 Not working - Workplace open,
but scared of COVID"
label var lab_nowork_fines "2.7 Not working - Workplace open, but
lockdown fines"
label var lab_nowork_dk "2.7 Not working - don't know"
label var lab_nowork_ref "2.7 Not working - refuse"
label var lab_nowork_oth "2.7 Not working - other"

* Reasons for not selling crops
label var agr_nosell_reasons "4.6 Not started selling crops"
label var agr_nosell_notready "4.6 Selling crops - not ready"
label var agr_nosell_seeds "4.6 Selling crops - keeping crops for seeds"
label var agr_nosell_destroyed "4.6 Selling crops - destroyed"
label var agr_nosell_demand "4.6 Selling crops - buyers aren't buying"
label var agr_nosell_markets "4.6 Selling crops - markets clsoed"
label var agr_nosell_prices "4.6 Selling crops - prices too low"
label var agr_nosell_lckdwn "4.6 Selling crops - not leaving house"
label var agr_nosell_labor "4.6 Selling crops - labor unavailable"
label var agr_nosell_transport "4.6 Selling crops - transport unavailable"

```

label var agr\_nosell\_machinery "4.6 Selling crops - machinery  
unavailable"

label var agr\_nosell\_dk "4.6 Selling crops - don't know"

label var agr\_nosell\_ref "4.6 Selling crops - refuse"

label var agr\_nosell\_oth "4.6 Selling crops - other"

\* Challenges in selling crops

label var agr\_sellcrops\_difficulties "4.13 Hurdles for selling harvested  
crops"

label var agr\_selldiff\_none "4.13 Selling difficulties - none"

label var agr\_selldiff\_labor "4.13 Selling difficulties - labor unavailable"

label var agr\_selldiff\_transport "4.13 Selling difficulties - transport  
unavailable"

label var agr\_selldiff\_markets "4.13 Selling difficulties - markets closed"

label var agr\_selldiff\_demand "4.13 Selling difficulties - buyers aren't  
buying"

label var agr\_selldiff\_travel "4.13 Selling difficulties - have to travel far"

label var agr\_selldiff\_police "4.13 Selling difficulties - police harassment"

label var agr\_selldiff\_prices "4.13 Selling difficulties - prices are too low"

label var agr\_selldiff\_dk "4.13 Selling difficulties - don't know"

label var agr\_selldiff\_ref "4.13 Selling difficulties - refuse"

label var agr\_selldiff\_oth "4.13 Selling difficulties - other"

\* State of ration shops

label var rel\_ration\_prc\_changes "5.6 Ration shop price change reasons"  
label var rel\_ration\_prc\_closed "5.6 Ration price change - ration shop  
closed"

label var rel\_ration\_prc\_nogo "5.6 Ration price change - don't go"

label var rel\_ration\_prc\_none "5.6 Ration price change - no changes"

label var rel\_ration\_prc\_free "5.6 Ration price change - at least one thing  
free"

label var rel\_ration\_prc\_cheaper "5.6 Ration price change - price of at  
least one thing reduced"

label var rel\_ration\_prc\_exp "5.6 Ration price change - price of at least  
one thing increased"

label var rel\_ration\_prc\_unavail "5.6 Ration price change - ration not  
available"

label var rel\_ration\_prc\_dk "5.6 Ration price change - don't know"

label var rel\_ration\_prc\_ref "5.6 Ration price change - refuse"

\* Health care services avoided

label var hea\_services\_avoided "6.5 Health services avoided"

label var hea\_avoid\_treat "6.5 Health service avoided - treatment"

label var hea\_avoid\_imm "6.5 Health service avoided - immunization"

label var hea\_avoid\_anc "6.5 Health service avoided - antenatal care"

label var hea\_avoid\_fp "6.5 Health service avoided - family planning"

label var hea\_avoid\_child\_grw "6.5 Health service avoided - growth  
monitoring of child"

label var hea\_avoid\_checkup "6.5 Health service avoided - health



checkup"

label var hea\_avoid\_edu "6.5 Health service avoided - health education"  
label var hea\_avoid\_oth "6.5 Health service avoided - other"  
label var hea\_avoid\_dk "6.5 Health service avoided - don't know"  
label var hea\_avoid\_ref "6.5 Health service avoided - refuse"

\* COVID symptoms known

label var hea\_symptoms "6.6 COVID symptoms known"  
label var hea\_symp\_fever "6.6 Know symptoms - fever"  
label var hea\_symp\_cough "6.6 Know symptoms - cough"  
label var hea\_symp\_tired "6.6 Know symptoms - tiredness"  
label var hea\_symp\_breath "6.6 Know symptoms - difficulty breathing"  
label var hea\_symp\_pain "6.6 Know symptoms - muscle pain/body aches"  
label var hea\_symp\_appetite "6.6 Known symptom - loss of appetite"  
label var hea\_symp\_throat "6.6 Known symptom - sore throat"  
label var hea\_symp\_diarrhea "6.6 Known symptom - diarrhea"  
label var hea\_symp\_nausea "6.6 Known symptom - nausea"  
label var hea\_symp\_congest "6.6 Known symptom - nasal and throat

congestion"

label var hea\_symp\_smell "6.6 Known symptom - loss of smell and taste"  
label var hea\_symp\_oth "6.6 Known symptom - other"  
label var hea\_symp\_dk "6.6 Known symptom - don't know"  
label var hea\_symp\_ref "6.6 Known symptom - refuse"

\* Prevention methods to protect against COVID

label var hea\_prevention\_methods "6.7 COVID prevention methods

known"

label var hea\_prevent\_wash "6.7 Know prevention - wash hands

frequently"

label var hea\_prevent\_san "6.7 Know prevention - use alcohol-based  
hand sanitizer"

label var hea\_prevent\_cover "6.7 Know prevention - cover nose and  
mouth with tissue"

label var hea\_prevent\_face "6.7 Know prevention - avoid touching face,  
eyes, nose, mouth"

label var hea\_prevent\_dist "6.7 Know prevention - be at least 1 meter  
away from everyone"

label var hea\_prevent\_crowd "6.7 Know prevention - avoid crowded  
places"

label var hea\_prevent\_avoid\_sick "6.7 Know prevention - avoid people  
who cough"

label var hea\_prevent\_contact "6.7 Know prevention - avoid contact with  
infected people"

label var hea\_prevent\_touch "6.7 Know prevention - avoid touching  
common surfaces"

label var hea\_prevent\_clean "6.7 Know prevention - clean common  
surfaces"

label var hea\_prevent\_mask "6.7 Know prevention - wear a mask"

label var hea\_prevent\_nospit "6.7 Know prevention - don't spit in public"

label var hea\_prevent\_home "6.7 Know prevention - stay at home"

```

label var hea_prevent_oth "6.7 Know prevention - other"
label var hea_prevent_dk "6.7 Know prevention - don't know"
label var hea_prevent_ref "6.7 Know prevention - refuse"

```

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

```
* B. Other
```

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

```
// These are other variables with missing labels
```

```
* Geography
```

```

label var geo_state_new "0.6 Current state"
label var geo_dist_new "0.15 Current district"
label define states 8 "Rajasthan" 9 "Uttar Pradesh" 10 "Bihar" 20
"Jharkhand" 23 "Madhya Pradesh" 28 "Uttar Pradesh" 2 "Arunachal Pradesh" ///
3 "Assam" 5 "Chhattisgarh" 6 "Goa" 7 "Gujarat"
38 "Haryana" 39 "Himachal Pradesh" 11 "Karnataka" ///
12 "Kerala" 14 "Maharashtra" 15 "Manipur" 16
"Meghalaya" 17 "Mizoram" 18 "Nagaland" ///
19 "Odisha" 37 "Punjab" 22 "Sikkim" 40 "Tamil
Nadu" 24 "Telangana" 25 "Tripura" 27 "Uttarakhand" ///
41 "West Bengal" 29 "Andaman and Nicobar
islands" 30 "Chandigarh" 31 "Daman and Diu" 32 "Delhi" ///
33 "Jammu and Kashmir" 34 "Ladakh" 35
"Lakshadweep" 36 "Puducherry"
label values geo_state_new states
replace geo_state_new = . if geo_state_new== -999 |
geo_state_new== -888

```

```
* Variables
```

```

label var rel_transfer_mem "5.5 Number of members receiving
government transfer"
note rel_transfer_mem: "5.5 The relief that you told us in the previous
question was received by your household is received by how many members?"
label define rel_nrega 1 "Yes, days have increased" 2 "Days have
reduced" 3 "Days remained the same" 4 "NREGA has stopped in the area"
label values rel_nrega rel_nrega
label var agr_prc_yr_unit_oth "4.12 Price Unit - other"

```

```
* Weights
```

```

label var strata_id "Strata ID"
label var weight_base "Base Weight"
label var weight_raked "Raked Weight"
label var weight_hh "Household Weight"
label var sampling_frame "Sampling Frame"

```

```
* Survey Specific
```

```

label var health_eligible "Eligible for health section"
label var fully_complete "Completeness of survey"
label define fully_complete 0 "Partially Complete" 1 "Fully Complete"
label values fully_complete fully_complete

```

\*-----\*

## \* 5. Final Cleaning

\*-----\*

### \* Drop unnecessary variables

```
drop district_new2 mod* sample primary_crop_final primary_crop_label  
crop_grown_label  
state_id_str district_id_str
```

### \* Order

```
order sampling_frame geo_state* geo_dist* geo_village geo_block  
hh_id_long_anon_str consent ag_dcision_maker* demo*, first  
order fully_complete strata_id weight_base weight_raked weight_hh, last
```

### \* Save

```
save "$db/wb1_cleaned_dataset.dta", replace
```