\*Data Import Stage for Indian Time-Use Data 2019, by Jihye Kim and Wendy Olsen

\* 2020, updated in 2023.

\*This file is related to our publication as shown, which you must cite! creative commons copyright.

\* Kim, Jihye, and Wendy Olsen (2023), Harmful Child Labour in India from a Time-Use \* Perspective, Development in Practice, 33:2, DOI

\* https://www.tandfonline.com/doi/full/10.1080/09614524.2022.2155620, Open access.

\*This first step involves taking data from India's Time-use data webpage and merging it.

\* This step comes before cleaning the data. We use stata as that is common in India.

\* Send any queries to wendy.olsen@manchester.ac.uk or jihye.kim@manchester.ac.uk

\* March 2023.

\*rename the online txt files as stata file in memory, save them as a C: drive file.

\*call these TUS106\_L01 etc.

\*These are sort of invisible in our system.

\*So you may use Stata's advised methods to download these files:

\*NOTE SPELLING https://mospi.gov.in/sites/default/files/tus/TUS106\_L02.TXT etc.

cd "C:\data\IATUR\IATUR\ITUS\"

clear

copy https://mospi.gov.in/sites/default/files/tus/TUS106\_L01.TXT "raw/level1.raw", replace

copy https://mospi.gov.in/sites/default/files/tus/TUS106\_L02.TXT "raw/level2.raw", replace

copy https://mospi.gov.in/sites/default/files/tus/TUS106\_L03.TXT "raw/level3.raw", replace

copy https://mospi.gov.in/sites/default/files/tus/TUS106\_L04.TXT "raw/level4.raw", replace

copy https://mospi.gov.in/sites/default/files/tus/TUS106\_L05.txt "raw/level5.raw", replace

dir raw/level\*.\*

\*Some people like to create a txt file as indicated by Indian Time-Use webpage. However,

\*the default infile type is .raw in Stata so we chose to rename these. Renaming often fails.

\*If you are a Windows user you will end up with level4.raw.txt if you are not careful.

\*NOTICE A flaw in the original filenames of Indian Time-Use webpage: the very small typing difference, where one file has surname .txt and another .TXT. In software this is very important. Hence we are showing it to you, see approximately lines 20-25 above,

\*where on line 25 you see small letters .txt but on lines 20-24 you have .TXT !.

\*you now have raw txt files, so use the dictionary to unpack and label them. See next step. You do not have to unzip these files.

\*File sizes expected are: level 1 19 Megabytes, level 2 71 Mb, level 3 19 Mb,

\* level 4 61 Mb, and level 5 1309 Mb. The last is so large that you must give

\* the machine plenty of time for autodownloading.

\*\*\*\*Import txt. file using dct. file\*\*\*\*

infix using "DCT\level1.dct", using (raw/level1.raw)

gen hhid= fsu + hh\_no

gen weight = mlt /100

tabstat weight, statistics( sum ) format(%14.0f)

sort fsu hh\_no sl\_no

gen state=substr(nss\_region,1,2)

label define state 1 "Jammu & Kashmir" 2 "Himachal Pradesh" 3 "Punjab" 4 "Chandigarh" 5 "Uttarakhand" 6 "Haryana" 7 "Delhi" 8 "Rajasthan" 9 "Uttar Pradesh" 10 "Bihar" 11 "Sikkim" 12 "Arunachal Pradesh" 13 "Nagaland" 14 "Manipur" 15 "Mizoram" 16 "Tripura" 17 "Meghalaya" 18 "Assam" 19 "West Bengal" 20 "Jharkhand" 21 "Odisha" 22 "Chhattisgarh" 23 "Madhya Pradesh" 24 "Gujarat" 25 "Daman & Diu" 26 "D & N Haveli" 27 "Maharashtra" 28 "Andhra Pradesh" 29 "Karnataka" 30 "Goa" 31 "Lakshadweep" 32 "Kerala" 33 "Tamil Nadu" 34 "Puducherry" 35 "A & N Islands" 36 "Telangana"

destring state, replace

label value state state

save "data\level1.dta", replace

clear

infix using "DCT\level2.dct", using (raw/level2.raw)

gen hhid= fsu + hh\_no

gen psid = fsu + hh\_no + sl\_no

sort fsu hh\_no sl\_no

save "data\level2.dta", replace

clear

infix using "DCT\level3.dct", using (raw/level3.raw)

gen hhid= fsu + hh\_no

sort fsu hh\_no

save "data\level3.dta", replace

clear

infix using "DCT\level4.dct", using (raw/level4.raw)

gen hhid= fsu + hh\_no

gen psid = fsu + hh\_no + sl\_no

sort fsu hh\_no sl\_no

save "data\level4.dta", replace

codebook

clear

infix using "DCT\level5.dct", using (raw/level5.raw)

\* see note below about Level 5 weight variable, which you may avoid.

gen hhid= fsu + hh\_no

gen psid = fsu + hh\_no + sl\_no

sort fsu hh\_no sl\_no

save "data\level5.dta", replace

codebook

\*Fist, combine level1 with level3, giving us files for household ;

\*2ndly then combine these by adding them to the level 2 data, which is about individuals.

\*thus, the file 'household data' is a roster plus household information,

\* having one line per individual! (sic)

\* If in doubt read the spreadsheet, Text Data Layout for Time Use Survey : Schedule-10.6, URL https://mospi.gov.in/sites/default/files/README\_TUS106.pdf, available

\* VIA https://mospi.gov.in/time-use-survey-0 and see also our article.

\*Thirdly, now add levels 5 and 4 to the combined data from level 1-2-3,

\*thus giving us a huge file of indiv data on time use per stint.

\*doing it this way, we do not create a distinct 'household' file as we would if using R software. If your machine hangs, you could get your data files from Smriti Rao and Vijayamba via the webpage of Foundation for Agrarian Studies (see \*@\* below).

clear

use "data\level1.dta", clear

merge 1:1 hhid using "data\level3.dta"

drop \_merge

save "data\household.dta", replace

clear

use "data\level2.dta", clear

merge m:1 hhid using "data\household.dta"

drop \_merge

merge 1:1 psid using "data\level4.dta"

drop \_merge

save "data\individual.dta", replace

clear

use "data\level5.dta", clear

merge m:1 psid using "data\individual.dta"

drop \_merge

\* there are 9,436,777 matched cases in this indiv file.

\*\*\*\*Generate relative weight\*\*\*\*

\*\*\*\* Be cautious. \*@\* Make sure your weight raw variable is from files 1, 2, 3 or 4, not from 5. Reference: S. Rao and Vijayamba R., online document on "Using the India TUS Unit Level Data: Some notes on method", URL https://fas.org.in/using-the-india-tus-unit-level-data-some-notes-on-method-and-some-stata-code/, accessed March 2022. This covers key issues in interpreting and using the documentation and data together.

sort hhid psid time\_from

egen meanWT=mean(weight)

gen fweight=weight/meanWT

gen fwt=round(fweight, 1)

recode fwt 0=1

\*\*\*\*Calculate daily hours\*\*\*\*

gen time\_from2 = substr(time\_from,1,2) + "." + substr(time\_from,4,2)

replace time\_from2=substr(time\_from,1,2)+ "." + "50" if substr(time\_from,4,2)== "30"

gen time\_to2 = substr(time\_to,1,2) + "." + substr(time\_to,4,2)

replace time\_to2=substr(time\_to,1,2)+ "." + "50" if substr(time\_to,4,2)== "30"

destring time\_to2 time\_from2, replace

gen time=time\_to2 - time\_from2

replace time=time+24 if time<0 & time!=.

drop if time==.

gen time\_7days=time\*7

tostring activity\_code, replace

gen activity\_1digit=substr(activity\_code,1,1)

label define activity\_1digit 1 "employment" 2 "production\_goods" 3 "unpaid hh domestic services" 4 "unpaid hh caregiving services" 5 "other unpaid work" 6 "learning" 7 "socializing" 8 "leisure" 9 "selfcare"

destring activity\_1digit, replace

label value activity\_1digit activity\_1digit

gen activity\_2digits=substr(activity\_code,1,2)

\*\*\*\*Constructing Variables\*\*\*\*

gen female=1 if gender==2

replace female=0 if female==.

save "data\timeuse.dta", replace

use "data\timeuse.dta", clear