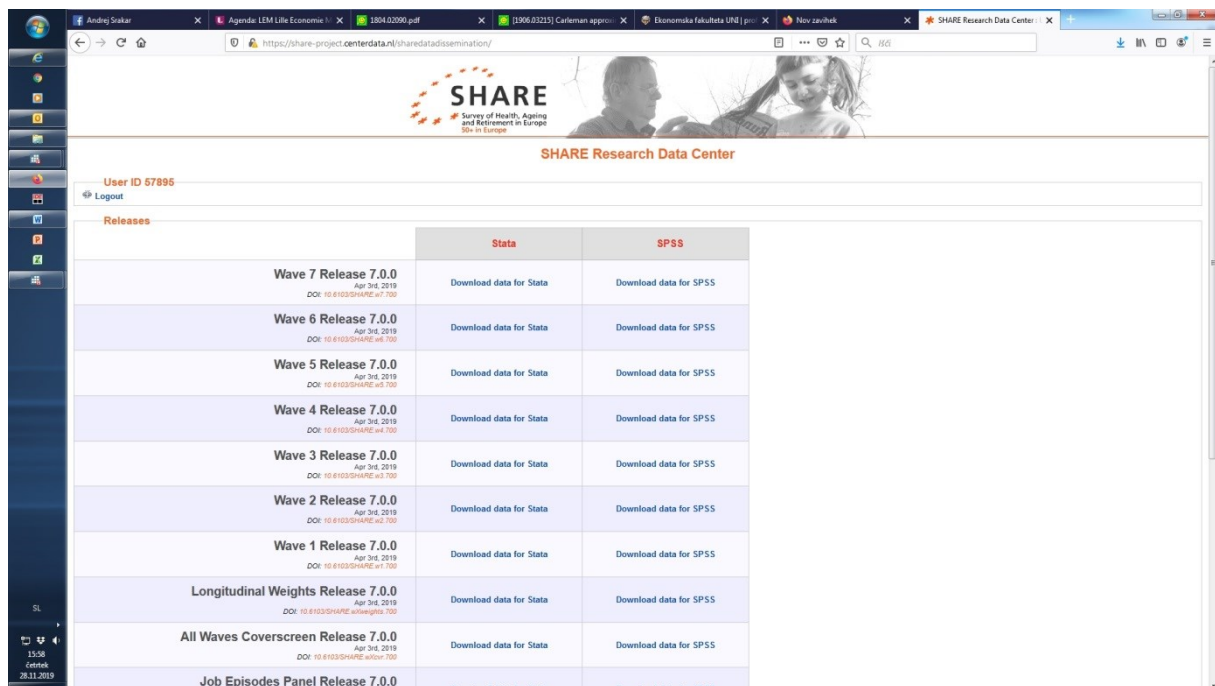


PREPROST UVOD V UKAZ readSHARE V PROGRAMU STATA: POMOČ K OBLIKOVANJU ŽELENE PODATKOVNE BAZE V SHARE

readSHARE je ukaz, ki so ga napisali poljski kolegi v projektu SHARE in omogoča karseda preprosto in hitro pripravo zelene podatkovne baze iz osnovnih podatkov SHARE, dostopnih na spletu vsem registriranim uporabnikom.

Kako torej najhitreje pripraviti lastno podatkovno bazo?

Najprej, registrirajte se v portal s podatki SHARE, morali bi videti nekaj podobnega, kot je na spodnji sliki.



The screenshot shows the SHARE Research Data Center website. At the top, there is a header with the SHARE logo and the text 'SHARE Research Data Center'. Below the header, there is a section for 'Releases' which contains a table of data releases. The table has three columns: 'Release', 'Stata', and 'SPSS'. Each row represents a different wave of data, from Wave 1 to Wave 7, plus Longitudinal Weights, All Waves Coverscreen, and Job Episodes Panel. Each row has a 'Download data for Stata' link and a 'Download data for SPSS' link.

Release	Stata	SPSS
Wave 7 Release 7.0.0 Apr 3rd, 2019 DOI: 10.6103/SHARE.v7.700	Download data for Stata	Download data for SPSS
Wave 6 Release 7.0.0 Apr 3rd, 2019 DOI: 10.6103/SHARE.v6.700	Download data for Stata	Download data for SPSS
Wave 5 Release 7.0.0 Apr 3rd, 2019 DOI: 10.6103/SHARE.v5.700	Download data for Stata	Download data for SPSS
Wave 4 Release 7.0.0 Apr 3rd, 2019 DOI: 10.6103/SHARE.v4.700	Download data for Stata	Download data for SPSS
Wave 3 Release 7.0.0 Apr 3rd, 2019 DOI: 10.6103/SHARE.v3.700	Download data for Stata	Download data for SPSS
Wave 2 Release 7.0.0 Apr 3rd, 2019 DOI: 10.6103/SHARE.v2.700	Download data for Stata	Download data for SPSS
Wave 1 Release 7.0.0 Apr 3rd, 2019 DOI: 10.6103/SHARE.v1.700	Download data for Stata	Download data for SPSS
Longitudinal Weights Release 7.0.0 Apr 3rd, 2019 DOI: 10.6103/SHARE.vw.weights.700	Download data for Stata	Download data for SPSS
All Waves Coverscreen Release 7.0.0 Apr 3rd, 2019 DOI: 10.6103/SHARE.vw.cover.700	Download data for Stata	Download data for SPSS
Job Episodes Panel Release 7.0.0	Download data for Stata	Download data for SPSS

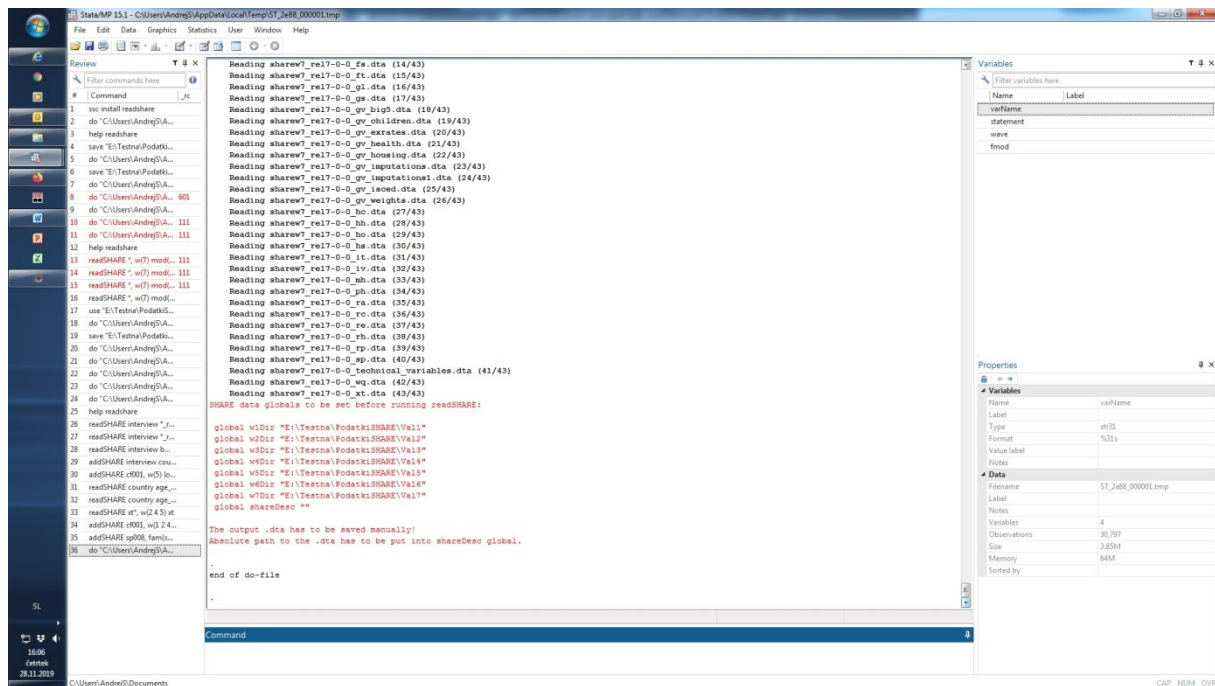
Naložite si podatke v programu Stata za vsakega od sedmih valov posebej, s klikom na ukaz "Download data for Stata". Shranite podatke za vsak val v svojo mapo na računalniku, ki ji dajte ime vala (Val1, Val2, Val3, itd.)

Najprej moramo shraniti posebno datoteko s podatki o vseh vključenih spremenljivkah. To storimo s spodnjim ukazom:

```
createSHAREDesc, dirs("E:\Testna\PodatkiSHARE\Val1"  
"E:\Testna\PodatkiSHARE\Val2" "E:\Testna\PodatkiSHARE\Val3"  
"E:\Testna\PodatkiSHARE\Val4" "E:\Testna\PodatkiSHARE\Val5"  
"E:\Testna\PodatkiSHARE\Val6" "E:\Testna\PodatkiSHARE\Val7")
```

Pri tem je denimo "E:\Testna\PodatkiSHARE\Val1" mesto, kjer se na računalniku nahajajo podatki prvega vala, podobno velja za vseh sedem valov v zgornjem ukazu. Sami morate seveda to spremeniti glede na mesto, kamor ste shranili podatke vsakega od valov.

Ko se izvede zgornji ukaz (kar lahko traja kakšno minuto, Stata na ekran izpiše kar nekaj izpisa/imen modulov v vmesnem času), bi morali videti sliko, podobno tej spodaj.



Zapis v rdečem ne pomeni napake! Pomeni le, da morate sedaj ročno shraniti rezultate, ki jih je proizvedel začetni ukaz. To naredite s spodnjim ukazom:

```
save "E:\Testna\PodatkiSHARE\SHAREDesc.dta"
```

(pri tem smo zopet uporabili mesto na računalniku, kamor sem sam shranil podatke (E:\Testna\PodatkiSHARE\), v vašem primeru spremenite zgornji ukaz in dodajte mesto, kamor ste vi shranili podatke, v tem primeru splošno/zbirno mapo, ki vsebuje vseh sedem valov in ne le enega).

Sedaj moramo določiti še zbirna mesta za vse podatke in datoteke, od koder bomo zajemali podatke. To naredimo s spodnjim naborom ukazov:

```
global w1Dir "E:\Testna\PodatkiSHARE\Val1"
global w2Dir "E:\Testna\PodatkiSHARE\Val2"
global w3Dir "E:\Testna\PodatkiSHARE\Val3"
global w4Dir "E:\Testna\PodatkiSHARE\Val4"
global w5Dir "E:\Testna\PodatkiSHARE\Val5"
global w6Dir "E:\Testna\PodatkiSHARE\Val6"
global w7Dir "E:\Testna\PodatkiSHARE\Val7"
global shareDesc "E:\Testna\PodatkiSHARE\SHAREDesc.dta"
```

Ponovno sem uporabil mesta, kamor sem sam shranjeval vse podatke, sami jih prilagodite vašemu računalniku, spreminjajte pa le del v narekovajih!

Sedaj je skoraj vse nared, da pričnemo povezovati podatke. Manjka le še en kamenček – če nameravam v podatkovno bazo vključiti tudi spremenljivke iz modula imputacij, je najbolje, da ga predhodno prilagodimo, saj vsebuje petkrat več opazovanj – vsaka manjkajoča vrednost dobi pet "implicatov", torej pet imputiranih vrednosti po metodi FCS. Da bi tudi ta modul lahko povezali z vsemi drugimi, manjšimi, je najlažje (in tudi metodološko korektno, kar svetuje literatura), če izberemo le prvi "implicat", torej le eno od petih vrednosti, v našem primeru prvo.

Da storimo to, izvedemo še spodnji nabor ukazov:

```
use "E:\Testna\PodatkiSHARE\Val7\sharew7_rel7-0-0_gv_imputations.dta", clear
drop if implicat!=1
save "E:\Testna\PodatkiSHARE\Val7\sharew7_rel7-0-0_gv_imputations1.dta"
```

Zopet velja enako kot prej – uporabil sem mesto, kamor sem sam shranjeval podatke, sami morate to spremeniti in prilagoditi svojemu računalniku, spreminjajte pa ponovno le del v narekovajih!

Sedaj je res vse pripravljeno za končno povezovanje.

V kolikor želimo denimo povezati vse module sedmega vala, lahko sicer izvedemo splošen readSHARE ukaz:

```
readSHARE *, w(7)
```

vendar pa bo zaradi nekaterih zelo posebnih modulov (zlasti: imputations, dropoff in xt/vprašalnik za umrle osebe) navadno veliko bolje, da se imensko specificira vse module, ki jih želimo povezati, to naredimo na primeru sedmega vala takole:

```
readSHARE *, w(7) mod(ac as br cc cf ch co cv_r dn dq ep ex fs ft gl
gs hc hh ho hs it iv mh ph ra rc re rh rp sp wq technical_variables
gv_children gv_health gv_housing gv_isced gv_big5 gv_weights
gv_imputations1)
```

Ukaz teče kar dolgo časa in zopet proizvede dolg izpis. Lahko se mu izognemo, če namesto zgornjega izvedemo spodnji ukaz, ki se razlikuje le v predponi/skrajšalnici qui, ki označuje, da program ne izpisuje sprotnega izpisa:

```
qui readSHARE *, w(7) mod(ac as br cc cf ch co cv_r dn dq ep ex fs
ft gl gs hc hh ho hs it iv mh ph ra rc re rh rp sp wq
technical_variables gv_children gv_health gv_housing gv_isced
gv_big5 gv_weights gv_imputations1)
```

Poglejmo si še praktični primer. Denimo, da želite napraviti podatkovno bazo spremenljivk, na področju zdravja. Zanimajo vas spremenljivke fizičnega zdravja in vedenjskih tveganj (kajenje, alkohol, itd.). Želite narediti panelno bazo, ki naj vsebuje vseh sedem valov. Hkrati želite vključiti tudi generirane spremenljivke iz zdravstvenega modula, torej modula gv_health, ter dodatno sociodemografsko spremenljivko zakonskega stana (dn014_ iz modula dn). Poleg tega želite vključiti še vse odzive anketarjev samih, shranjene v modulu iv.

Za začetek pogledamo, ali vseh sedem valov resnično vsebuje vse štiri zahtevane (celotne) module, torej ph (fizično zdravje), br (vedenjska tveganja), gv_health ter iv. Izkaže se, da izmed vseh sedmih valov, šest od njih (vsi razen tretjega, ki ne vsebuje nobenega od treh) vsebuje prve tri, vseh sedem valov pa vsebuje modul iv v takšni ali drugačni obliki.

Štiri zelene celotne module povežemo s spodnjim ukazom:

```
readSHARE *, w(1 2 3 4 5 6 7) mod(br ph gv_health iv)
```

Združena podatkovna baza vsebuje spremenljivko *wave*, ki je identifikator posameznega vala pri posamezniku. Baza je že pripravljena v panelni obliki (opazovanja so naložena ena na drugo), potrebno jo je le še urediti, kar naredimo s spodnjim ukazom:

```
sort mergeid wave
```

Potrebno kje le še dodati eno, dodatno spremenljivko, torej zakonski stan. Slednji je vključen v sociodemografski modul *dn*, in sicer v vseh valovih razen v tretjem. Izvedemo torej spodnji ukaz:

```
addSHARE dn014_, w(1 2 4 5 6 7)
```

Naša podatkovna baza je tako pripravljena.

Na podoben način ravnamo v drugih primerih, ob različnih modulih, ki jih želimo povezati, valovih, v katerih so bili vključeni, ter posameznih spremenljivkah.

Nekaj dodatnih informacij o delu z readSHARE se vam izpiše v Stati s spodnjim ukazom:

```
help readshare
```

V primeru nadaljnjih vprašanj nam prosim pišite, odgovorili vam bomo z veseljem. Prav tako z veseljem priskočimo na pomoč pri pripravi podatkovne baze – kot povedano na delavnici, je to ena naših osnovnih nalog.

Gradivo pripravil:

doc. dr. Andrej Srakar, Inštitut za ekonomska raziskovanja (IER) in Ekonomska fakulteta Univerze v Ljubljani

strokovni sodelavec SHARE Slovenija, zadolžen za podporo uporabnikom podatkov

srakara@ier.si