gesis

Leibniz Institute
for the Social Sciences

PIAAC Data Analysis in Stata:
A practical guide

Video 2: Data Access and First Steps in Stata

GESIS – Leibniz Institute for the Social Sciences

Leibniz
Association

In a Nutshell

Goal:

- Practical guide for three existing tools to perform PIAAC data analysis in Stata

Target Group:

- Researchers with some experience in Stata, but little or no experience on analysis of PIAAC data in Stata

Structure:

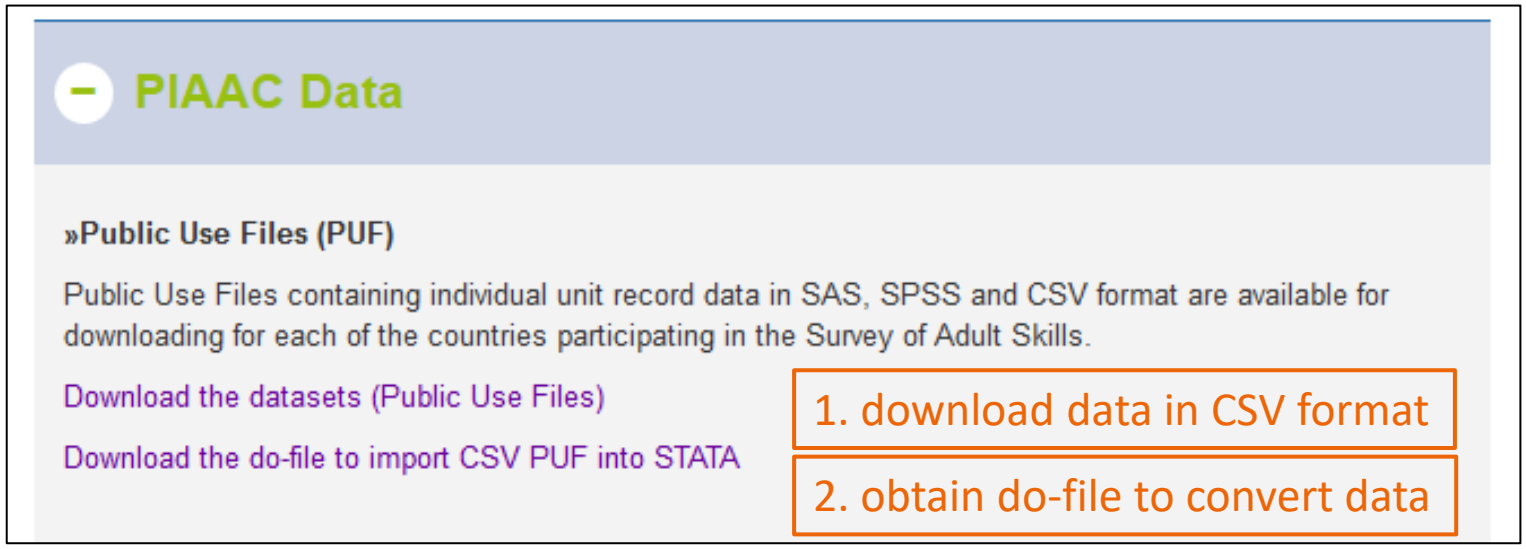
- General overview of how the three tools work, as well as several exemplary analyses with the PIAAC data

Video 2:
Data Access and First Steps in Stata

Data Access I

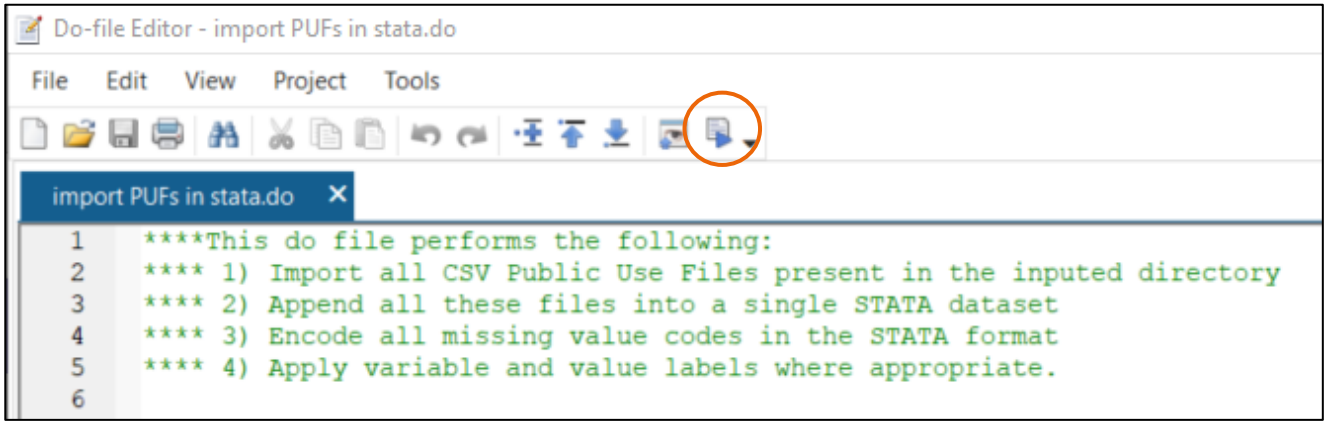
Public Use Files (PUFs) for most countries can be accessed via the OECD webpage:
http://www.oecd.org/skills/piaac/data/

1. Download data in CSV format
2. Download do-file to convert data into Stata format

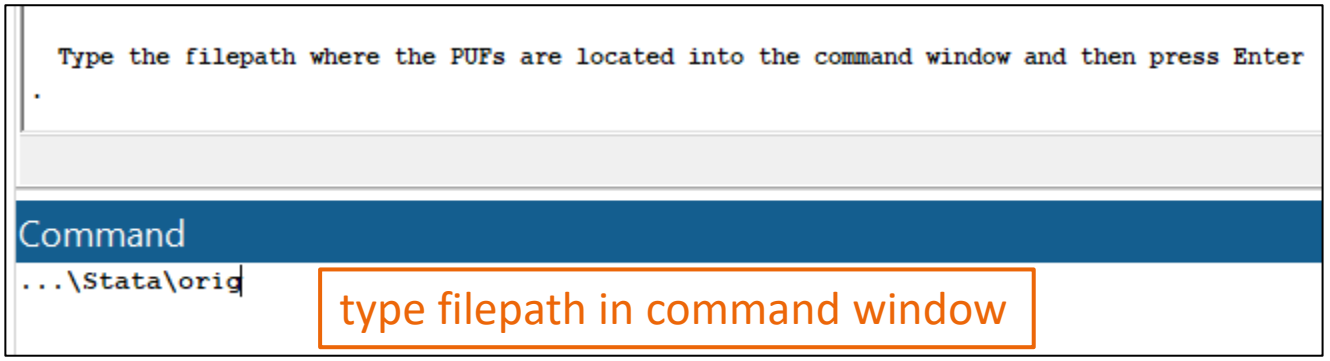


Data Access II

3. Run do-file

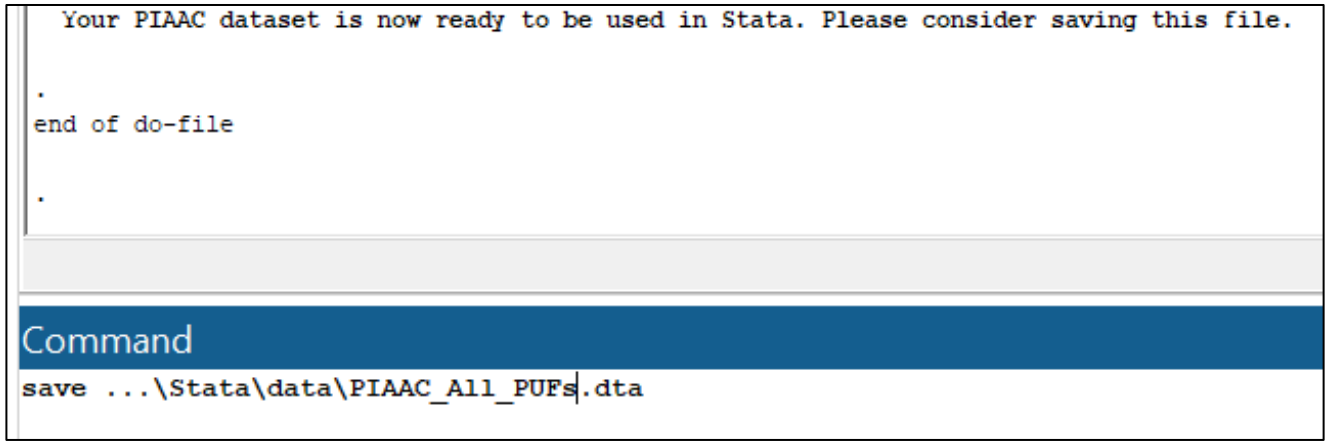


4. Stata will request filepath where CSV files are saved:



Data Access III

5. Save data file with all PUFs

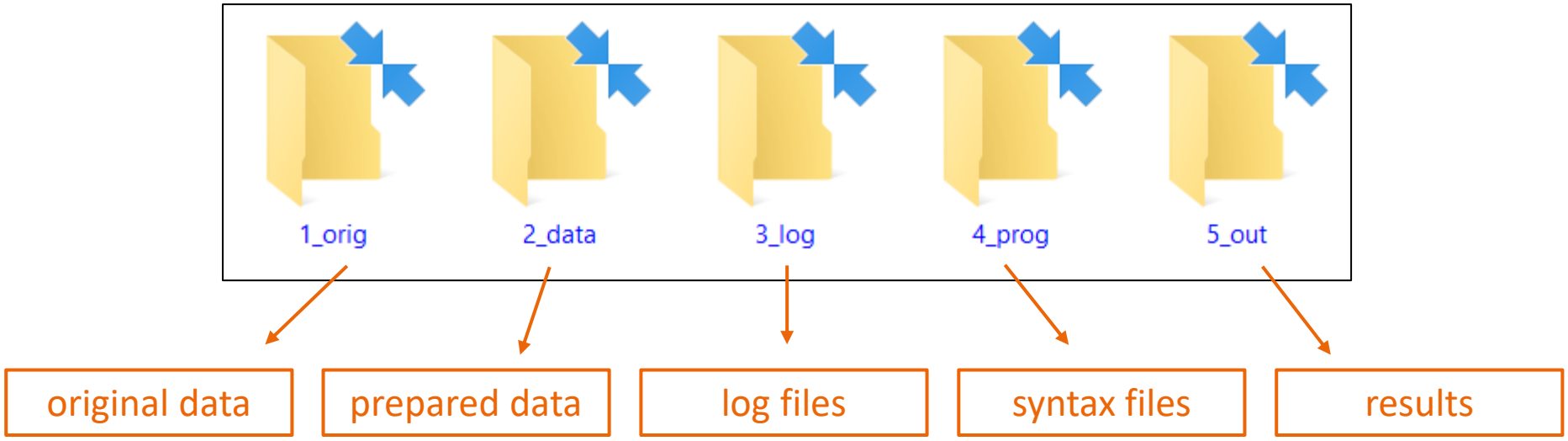


Note: Country-specific Scientific Use Files (SUFs) may contain further variables and may be obtained from national data providers,

e.g. for Germany: Research Data Center PIAAC, https://www.gesis.org/en/piaac/rdc

Data Management

- Suggested folder structure for data analysis:



Variables in PIAAC

1. Respondents' answers to background questionnaire (BQ)
   - e.g., job satisfaction: `d_q014`
2. Variables derived from respondents' answers to BQ
   - e.g., current employment status: `c_d05`
3. Respondents' answers to cognitive items
   - e.g., scored response on “orchestra ticket”: `e664001s`
4. Ten plausible values (PVs) for each skill domain
   - e.g., `pvlit1, pvlit2, …, pvlit10` for literacy
5. Interviewers' answers to observation module
   - e.g., assessment taking too long: `zz5`
6. Auxiliary variables on survey workflow and design
   - e.g., paper-based vs. computer-based assessment: `pbroute`

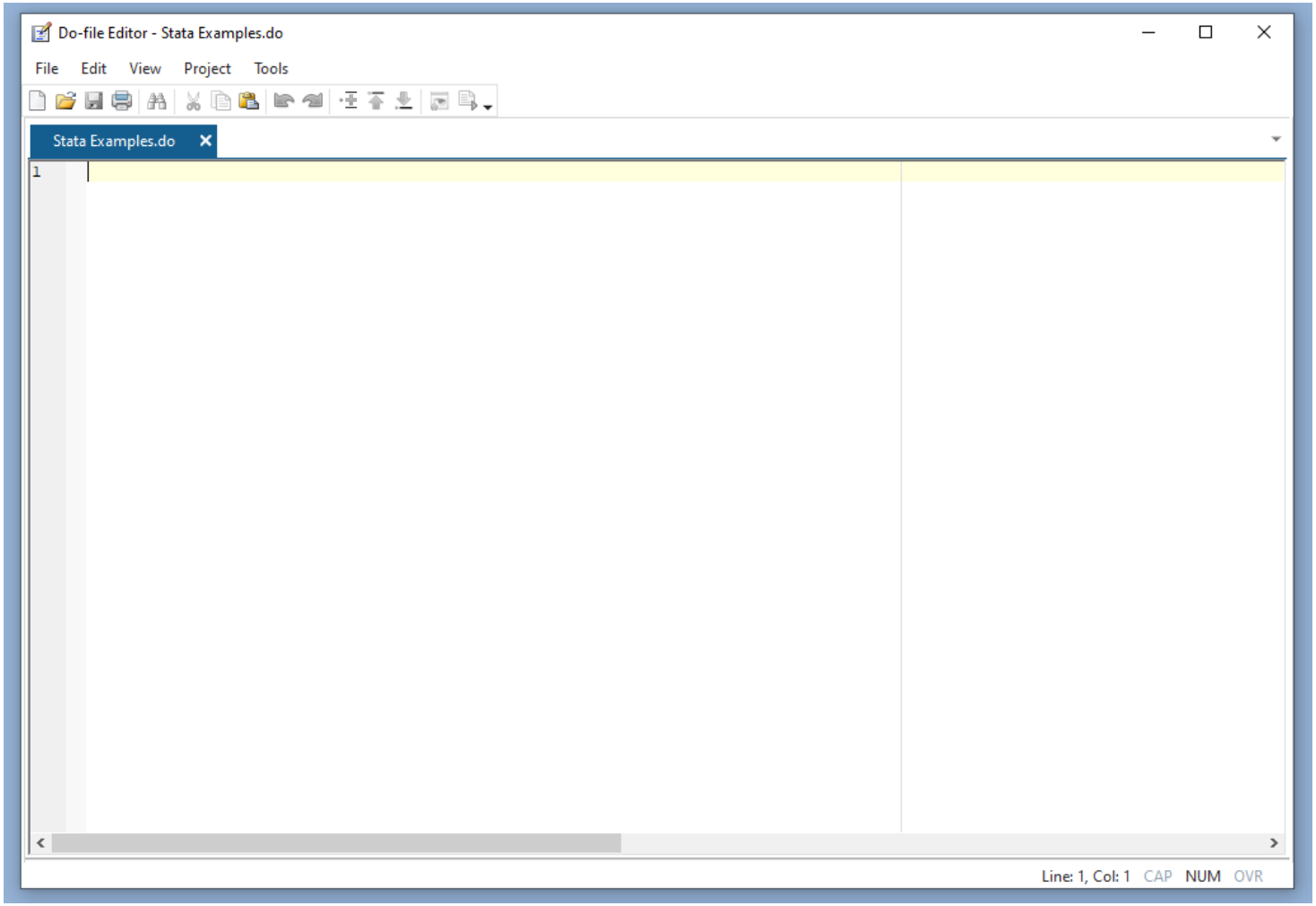
International codebook available at oecd.org/skills/piaac/publicdataandanalysis/

First Steps in Stata

- Tabulate and label country IDs
  - `cntryid`
- Tabulate plausible values for all skill domains:
  - `pvlit*`
  - `pvnum*`
  - `pvpsl*`
- Tabulate replicate weights:
  - `spfwt*`
- Save smaller data set only including Germany, Greece, Singapore, Sweden, and the US

  → we will use this data set for the following sample analyses

Let‘s go to Stata



gesis

Leibniz Institute
for the Social Sciences

Leibniz
Association