Working with Data on Political Behaviour: The CSES

Hannah Schwarz
GESIS Leibniz Institute for the Social Sciences, Cologne
hannah.schwarz@gesis.org

Presentation at CESSDA Workshop, 06 November 2017

Analysis Aim for Illustration

- Let's look at:
  - Individual-level relationship: Efficacy and turnout
  - Contextual effect: Electoral system proportionality on turnout
  - Inspired by Banducci & Karp (2008)

Advocates of proportional representation (PR) often cite its potential for increasing citizen involvement in politics as one of PR's fundamental advantages over plurality or first-past-the-post systems. The assumption is that plurality electoral systems distort the translation of votes into seats, discouraging and alienating small party supporters and other political minorities. In contrast, PR systems are believed to provide greater opportunities for representation which are assumed to instil greater efficacy and increase participation. We examine this theory linking institutions to electoral participation across a diverse set of countries using data from the Comparative Study of Electoral Systems. Using a multi-level approach we find evidence consistent with the expectations about the negative influence of disproportional systems on political minorities…

CSES
COMPARATIVE STUDY of ELECTORAL SYSTEMS

Data Download

- **www.cses.org**

- **Go to Data Center, sign up**
- **Choose Module 2**
- **Download codebook files (3) & stata file (→ unzip)**

- *Website hint: For an overview of included election studies and variables in all Modules, tables are provided under Quick Links:*
  http://www.cses.org/electionstudies.htm
  http://www.cses.org/vartable.htm

CSES
COMPARATIVE STUDY of ELECTORAL SYSTEMS

Navigating the Documentation

- CSES Codebook files
- 3 parts in Module 2
  - Introduction
  - Variables Description
  - Appendices (Parties and Leaders, Primary Electoral Districts)

Quick exercise:

What is the last district-level variable of the dataset?

- Keyword search
  - sections)))
  - subsections >>>
  - variables (e.g. B1010_1)

CSES
COMPARATIVE STUDY of ELECTORAL SYSTEMS

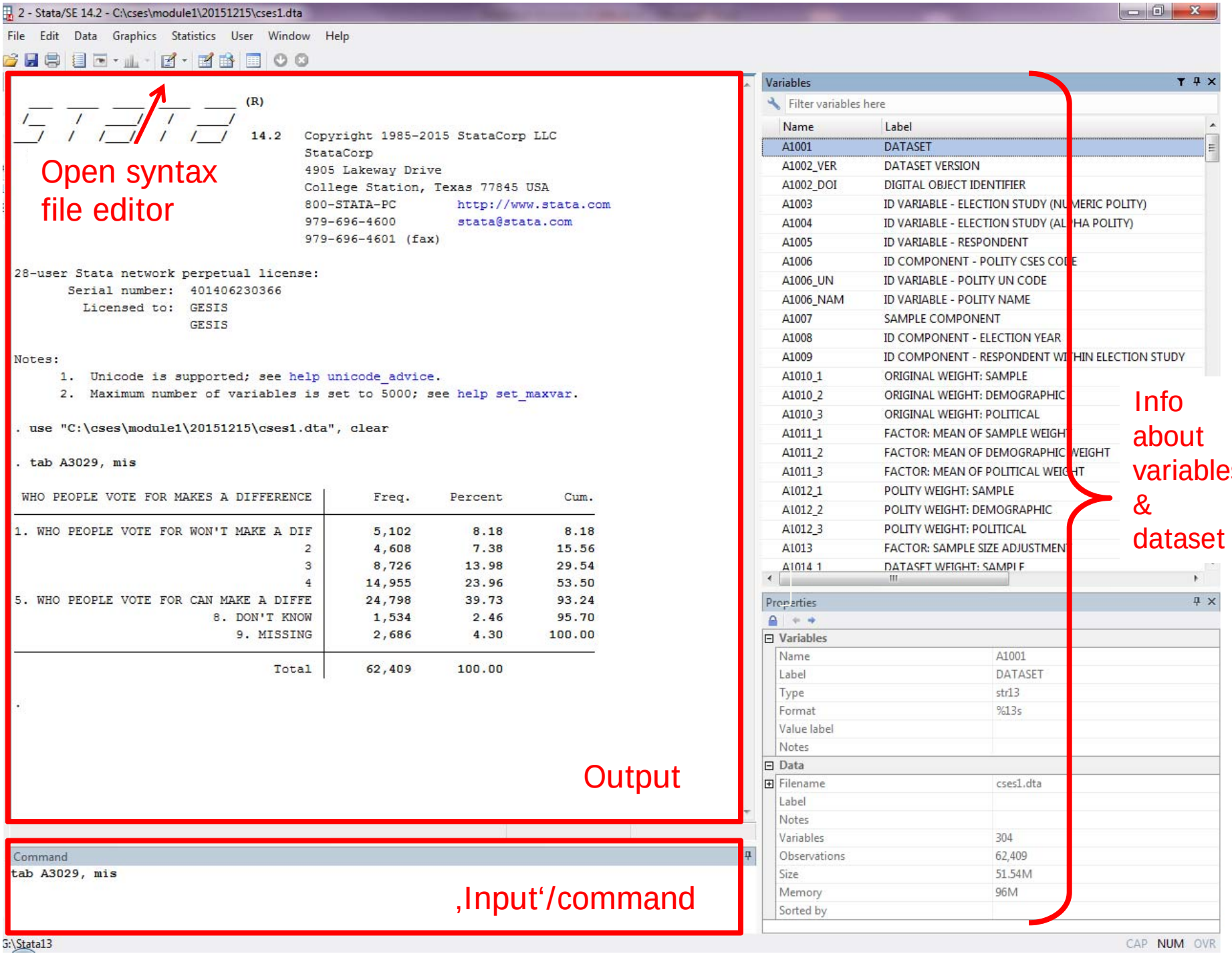
Navigating the Documentation ctd

- Important structure in variables section:
  - Variable name
  - Question text
  - Answer codes
  - Variable notes
  - Election Study Notes

Quick exercise:

How many answer codes does variable turnout (B3004_1/B3004_2) have?

- [SEE ELECTION STUDY NOTES] – election study specific codes!


2 - Stata/SE 14.2 - C:\cses\module1\20151215\cses1.dta
File Edit Data Graphics Statistics User Window Help
Open syntax file editor
14.2 Copyright 1985-2015 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)
28-user Stata network perpetual license:
Serial number: 401406230366
Licensed to: GESIS
GESIS
Notes:
1. Unicode is supported; see help unicode_advice.
2. Maximum number of variables is set to 5000; see help set_maxvar.
. use "C:\cses\module1\20151215\cses1.dta", clear
. tab A3029, mis
WHO PEOPLE VOTE FOR MAKES A DIFFERENCE | Freq. Percent Cum.
1. WHO PEOPLE VOTE FOR WON'T MAKE A DIF | 5,102 8.18 8.18
2 | 4,608 7.38 15.56
3 | 8,726 13.98 29.54
4 | 14,955 23.96 53.50
5. WHO PEOPLE VOTE FOR CAN MAKE A DIFFE | 24,798 39.73 93.24
8. DON'T KNOW | 1,534 2.46 95.70
9. MISSING | 2,686 4.30 100.00
Total | 62,409 100.00
Output
Command
tab A3029, mis
‚Input'/command
Variables
Filter variables here
Name Label
A1001 DATASET
A1002_VER DATASET VERSION
A1002_DOI DIGITAL OBJECT IDENTIFIER
A1003 ID VARIABLE - ELECTION STUDY (NUMERIC POLITY)
A1004 ID VARIABLE - ELECTION STUDY (ALPHA POLITY)
A1005 ID VARIABLE - RESPONDENT
A1006 ID COMPONENT - POLITY CSES CODE
A1006_UN ID VARIABLE - POLITY UN CODE
A1006_NAM ID VARIABLE - POLITY NAME
A1007 SAMPLE COMPONENT
A1008 ID COMPONENT - ELECTION YEAR
A1009 ID COMPONENT - RESPONDENT WITHIN ELECTION STUDY
A1010_1 ORIGINAL WEIGHT: SAMPLE
A1010_2 ORIGINAL WEIGHT: DEMOGRAPHIC
A1010_3 ORIGINAL WEIGHT: POLITICAL
A1011_1 FACTOR: MEAN OF SAMPLE WEIGHT
A1011_2 FACTOR: MEAN OF DEMOGRAPHIC WEIGHT
A1011_3 FACTOR: MEAN OF POLITICAL WEIGHT
A1012_1 POLITY WEIGHT: SAMPLE
A1012_2 POLITY WEIGHT: DEMOGRAPHIC
A1012_3 POLITY WEIGHT: POLITICAL
A1013 FACTOR: SAMPLE SIZE ADJUSTMENT
A1014_1 DATASET WEIGHT: SAMPLE
Info about variables & dataset
Properties
Variables
Name A1001
Label DATASET
Type str13
Format %13s
Value label
Notes
Data
Filename cses1.dta
Label
Notes
Variables 304
Observations 62,409
Size 51.54M
Memory 96M
Sorted by
G:\Stata13
CAP NUM OVR

CSES Weights

- "[…] "visualizing" the role of case-specific weights in survey data analysis is to consider the weight as the number (or share) of the population elements that is represented by the sample observation." (Heeringa et al., 2010, p.35)
- CSES has Sample, Demographic & Political Weights
- Sample Weights correct for unequal probabilities of inclusion in the sample → most important weights
- Other weights can be helpful for certain analyses.
- Use of weights often debated. Individual decision, depends on analysis.

CSES
COMPARATIVE STUDY of ELECTORAL SYSTEMS

CSES Weights

- Before working with weights:
  - Check documentation
  - Look at distributions and effects of weights
- Codebook: Weight Variables B1010_1 - B1010_3

Quick exercise:

What do the Sample Weights correct for in the case of New Zealand (2002)?

- For more details go to Design Report
- We don’t advise combining different CSES weights

CSES
COMPARATIVE STUDY of ELECTORAL SYSTEMS

Exercise 1

- Open the Stata syntax file ‘CSES_CESSDA_Training.do’. Use the commands in the first sections on recoding and descriptive statistics to either …

  a) … run the presented commands or similar ones while sticking with the variables/research question of the example, or …

  b) … adjust the commands or devise other syntax to derive descriptive statistics about variables you are interested in.

CSES
COMPARATIVE STUDY of ELECTORAL SYSTEMS

Nested Data

- Individual observations are nested in clusters sharing certain attributes and are therefore likely to be correlated
- E.g. pupils in a class (share teacher etc.)
- E.g. Individuals in a country (share all kinds of country-level attributes)
- Solution 1: Dummy variables for clusters (e.g. countries)
- Solution 2: Multi-level models

CSES
COMPARATIVE STUDY of ELECTORAL SYSTEMS

Multi-level Models

- *Statistically*, accounting in model for nested data structure important – else risk to underestimate Standard Errors

- *Theoretically*, modelling how variance is distributed across levels can be enlightening:
  - ICC (intraclass-correlation) shows how much variance of the outcome variable resides at context level (fraction of the total variance that is due to variation between countries)
  - Effects of context level predictors
  - Investigate whether individual-level effects vary between contexts

- CSES data structure suited for multilevel analysis (common: effects of country context on behavior)

CSES
COMPARATIVE STUDY of ELECTORAL SYSTEMS

Exercise 2

- Use the commands in the section on correlation and regression to either …

  a) … run the presented commands or similar ones while sticking with the variables/research question of the example, or …

  b) … adjust the commands or devise other syntax to run correlations and regression analyses with variables you are interested in.

Take Home Messages

- Analyzing a cross-national dataset is different than analyzing a single-country dataset.
- Use specific modelling techniques (e.g. multi-level models) to account for nested data structure.
  - CSES bibliography has examples of use/methods
- Use documentation to get acquainted with specificities of different election studies and country backgrounds.
- Find out more about local contexts.
  - Bibliography provides more sources
  - Ask colleagues from the region
  - Collaborators are known and can be contacted

CSES
COMPARATIVE STUDY of ELECTORAL SYSTEMS

Further Reading

Goldstein, H. (1995). *Multilevel Statistical Models*, 2nd edition. London: Edward Arnold.

Heeringa et al. (2010). *Applied Survey Data Analysis*. Chapman & Hall/CRC.

Hox, J. (2002). *Multilevel Analysis. Techniques and Applications*, 2nd edition. London: Lawrence Erlbaum.

Kreft, I. and Leeuw, J. de (2000). *Introducing Multilevel Modelling*. London: Sage

Rabe-Hesketh, Sophia and Anders Skrondal (2012). *Multilevel and Longitudinal Modeling Using Stata*. 3rd edition. Volume I: Continuous Responses. College Station, Texas: Stata Press.

Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical Linear Models. Applications and Data Analysis Methods*, 2nd edition. Sage: Thousand Oaks.

Snijders, Tom A. B. and Roel J. Bosker (1999). *Multilevel Analysis*. London: Sage.

CSES
COMPARATIVE STUDY of ELECTORAL SYSTEMS

This presentation is property of the CSES Secretariat. It should be cited as follows:

CSES (2017). Working with Data on Political Behaviour:
The CSES. Presentation at CESSDA Workshop, 06 November 2017.