

Which statistical software to use?

A reflection and review upon the use of STATA, R and Python for teaching in the social sciences

Thomas de Graaff¹*

Abstract

Keywords

Teaching—Empirical research—STATA —R—Python

¹Department of Spatial Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

*Corresponding author: ✉ t.de.graaff@vu.nl; 📧 thomasdegraaff.nl

Introduction: the empirical workflow

Econometrics is much easier without the data.

Marno Verbeek

In economics

A typical empirical workflow in the social sciences looks as follows:

Generate data Data is read from an external source (file or online database) or is simulated.

Transform data This is usually the most time demanding phase¹ and includes (amongst many other things) manipulating missing data, merging data and relabeling data

Analyse data This phase includes as well generatin

Present results

Statistical software packages

STATA

R

Python

Statistics in education

Concluding remarks

¹There is an old rule that says that 80% of your research time goes in transforming data, while 20% is only spent on analysing the data