IPyStata

Stata + Python + Jupyter Notebook

“The whole is greater than the sum of its parts.”
- Aristotle
Introduction
Workflow with Python and Stata

Generate data
- Web scraping
- I/O operations and IOT (Internet-of-things)

Process data
- NLP (Natural language processing)
- Ever have the urge to “just do it manually in Excel”? ➔ Use Pandas!
  Document all your actions in code ➔ replicability

Insight from data
- Statistical analysis (Stata)
- Visualize results / data (Stata / Python)

Tilburg University
Stata Conference 2016 | Ties de Kok
What is the Jupyter Notebook?

Open source, interactive data science and scientific computing across over 40 programming languages.

Jupyter Notebook

The Jupyter Notebook is a web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more.
Demonstration: Jupyter basics

Click here to play
What is IPyStata?

A wrapper to execute Stata code in a Jupyter Notebook environment

Seamlessly use both Python and Stata in one environment

- Interchangeable data structures using Pandas DataFrames

- Stata do-file / log-file on steroids
  - Input + Output in one place
    - (plots are supported)
  - Format / Headers / Comments
    - Markdown / Images / Latex equations
  - Run multiple Stata sessions at the same time from one notebook
  - Easy sharing of code + results
    - export ‘static’ version of the notebook
  - Works well with version control
    - GitHub renders notebooks!
Click here to play
Move data around between Stata and Python

Moving data around is very easy using the `-d` (`--data`) or `-o` (`--output`) arguments

```
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
```

Click here to play
Click here to play
Click here to play
Use IPyStata to make offline materials

• Jupyter Notebooks can easily be converted into offline HTML documents

• These HTML documents can be easily edited for teaching and demonstration purposes
  ➔ Converting to PDF works well
Estimate HEALY variable

This model assumes that the non-discretionary accrual part depends on the accruals of last year.

\[ NDA_t = TA_{t-1} \]

```
In [10]:
gen HEALY=WTA-lag_WTA
```

```
In [11]:
histogram HEALY
```

```
(bin=45, start=-1.0332508, width=.04592226)
```

```
Out[11]:
```

![Histogram of HEALY](image)
Jupyter Notebook + Version Control

• Officially supported by GitHub

Code, results, comments, etc. all under version control!

Click here to play
How to start using IPyStata?

• Check out my “getting started” guide on the IPyStata GitHub:

https://git.io/vKrff

• IPyStata GitHub page:

https://github.com/TiesdeKok/ipystata

• Official Jupyter website:

http://www.jupyter.org/

Questions? Contact me!
- GitHub/TiesdeKok
- t.c.j.dekok@uvt.nl

1) Install the Anaconda distribution
2) Install IPyStata
3) Configure IPyStata