Advanced Research Tools for Economics and Business Administration

Thomas de Graaff

January 15, 2016

Why this workshop?

- In the *social sciences* few attention to what tools to use (and why they make sense)
- LaTeXis used very much in the scientific world and can *work* together with
  - Stata/R
  - Markdown/HTML
  - Reference managers
- Why *I* want to give this workshop
  - intrinsic interest
  - my goal: pre-conferences workshops / courses

What I want (and don't want) with this workshop

- Give a general introduction of why some tools work together
  - Why LaTeX
  - Why reference managers
- Give an introduction to LaTeX
  - First the basics
  - Next workshop: some advanced stuff
- What *I* do not want
  - Tell you what applications to use (**you** need to decide and make a **well-informed** decision)

Background

- TeXhas been devised by Donald E. Knuth in the late 70's
- LaTeXis a set of macro's around TeX and devised in the 80's
- LaTeXis a *typesetting program*, not a *Word processor*
  - It is actually some code that needs to be compiled
  - Code is typed in by an editor
- So, huge differences between
  - Word processor: Open Office, Word
  - Typesetter: LaTeX, Adobe's InDesign (in general XML)
  - Editors:
    - Specific editors: TexStudio, TexShop, RStudio
    - General editors: Sublime, TextMate, Notepad++, Vim, Emacs

Disadvantages

- Not WYSIWYG
- You nead to learn (quite) some commands
  - Learning curve, but
  - hurray for cheat sheets and Google
- Difficult to cooperate with people that went to the *dark side*
- *Basic* LaTeXhas *difficulties* with incorporating new fonts (Hoefler, minion pro)
  - XeTeX
  - For the purists: LaTeX does it right (LaTeXvs Word)

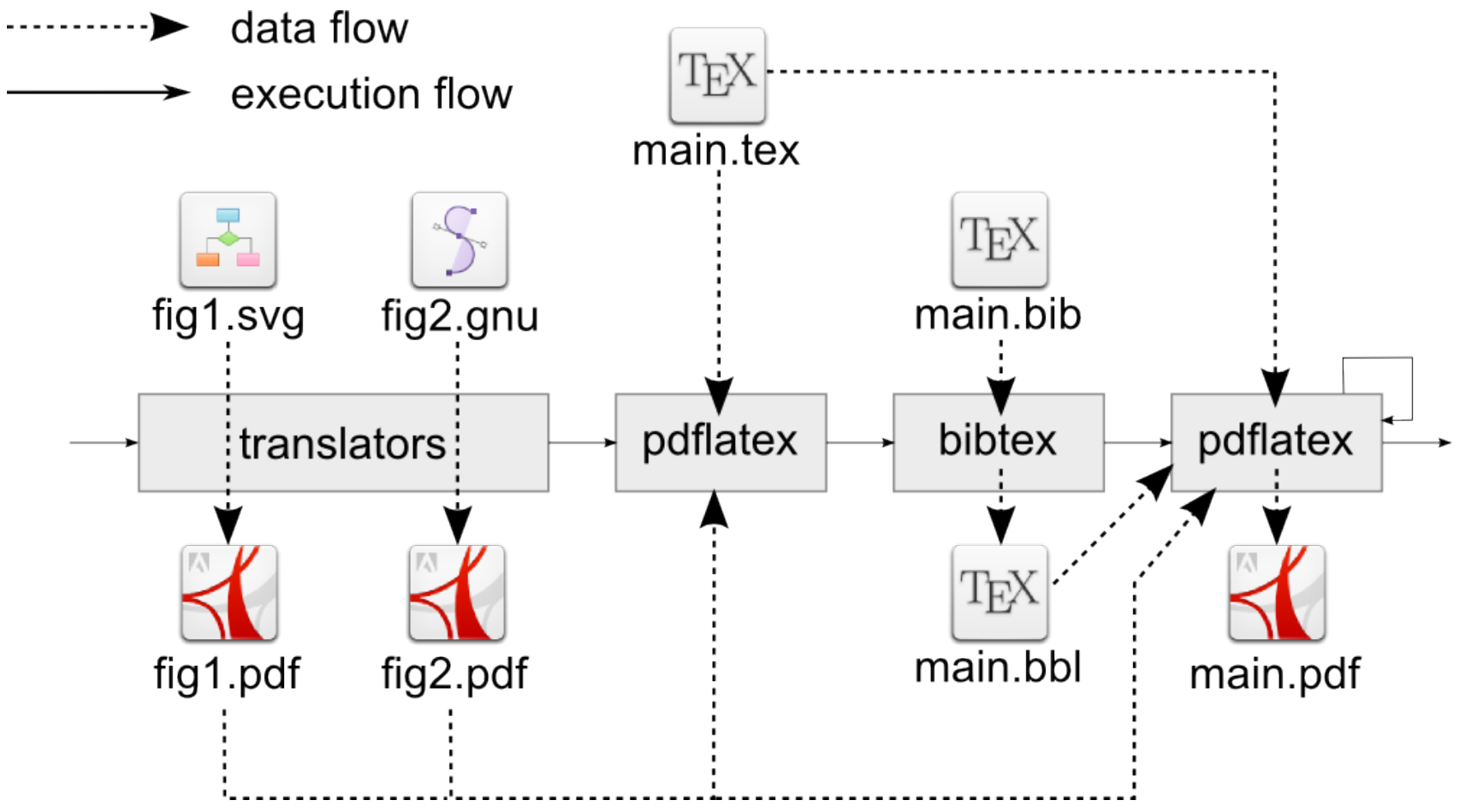
Advantages

- Free (as in beer) and ubiquitous
- WYSIWYM
- Consistent lay-out throughout the whole document (including tables, appendices, formulas, source code, etc)
- Internal references are a breeze (references, tables of. . . , indices)
- Forced to structure documents
- Macros, thus scriptable
- Large community, thus a package for almost everything (books, articles, presentation, posters, exams, musicscores)
- Superior typography & output
- Large publishers (i.e., Elsevier and Springer) have LaTeX templates for their articles

How does it work in practice?

- You edit a `.tex` file without thinking about how it looks
  - distraction free writing (yeah right)
- You then compile it
  - LaTeXis unforgiving: if there is an error, usually it does not compile
  - Typically, errors are missing brackets or parentheses.
- Typically, source `.tex` file is compiled into `.pdf`

A process diagram



Why bother about workflow or tools?

- Good scientific practice: *document how you have achieved your results*; this ensures
  - Reproducibility
  - Transparency
  - Modularity
  - Portability (across systems and users)
  - Efficiency
  - Self-sanity

When should I adopt new tools/workflow?

- The sooner the better (you really have time now)
- But think twice about which one (switching is costly; not in terms of beer but in terms of time)
- Start one step at a time (starting with LaTeXis a pretty neat idea)

A journey of a thousand miles begins with a single step

Lao-tzu

In general

In science consensus is irrelevant. What is relevant is reproducible results. The greatest scientists in history are great precisely because they broke with the consensus (Michael Crichton)

In data science

- Typically, a publication is not at the heart of research
  - Code
  - Data

The data and code used to make a finding are available and they are sufficient for an independent researcher to recreate the finding (Peng, 2011)

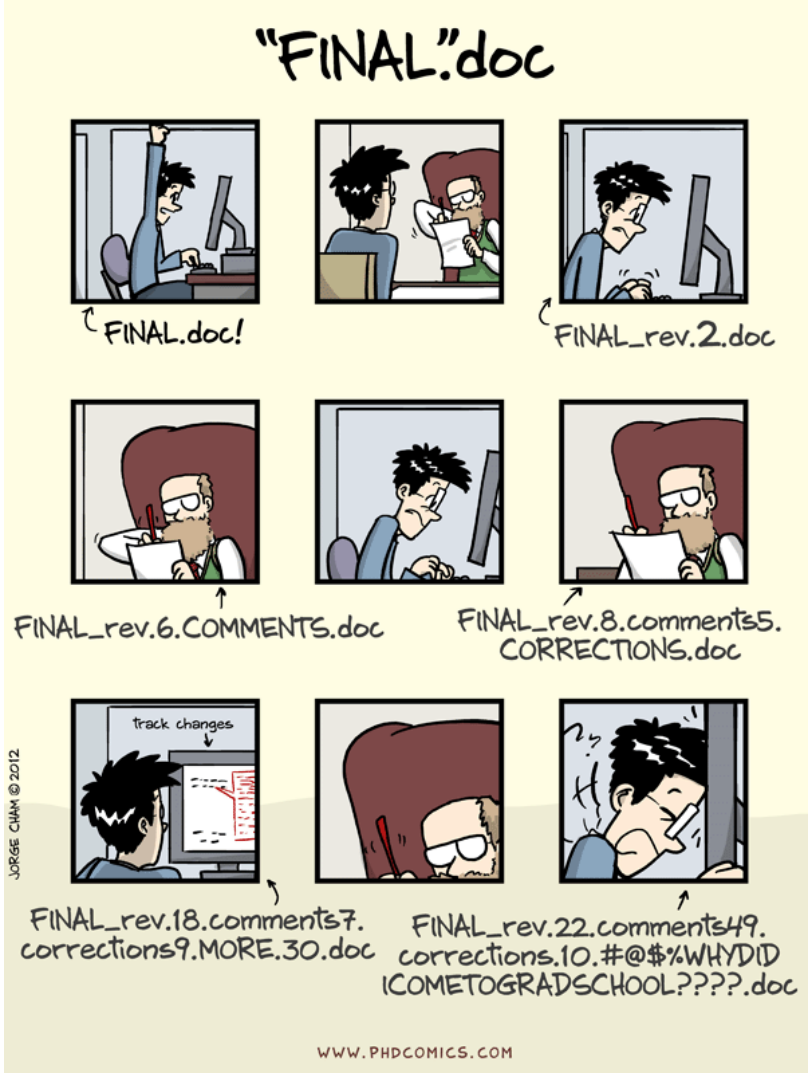
Code, documentation and output

1. Synonyms
2. All based on `.txt` files
3. Encompasses almost anything
   - data itself (`.csv`, `.txt`)
   - set of commands for data cleaning and statistical analysis (`.do`, `.R`)
   - database with references (`.bib`)
   - text for articles, presentations or websites (`.tex`, `.html`)
4. Only output is displayed/interpreted differently (e.g., in a browser or pdf viewer)

Folder structure of your new project (theses, paper, research)

- Think *a priori* about project set-up
  - Seperate analysis, data and output files
- Be careful with source data!
  - Seperate source and derived data files
  - Typically
    - you get/collect data
    - transform data
    - analyse data
  - Keep track of all these stages!

Why is version control systems such a neat idea



Why reference managers?

This is a life saver!

Use one!

Several applications out there:

- In this case Mendeley (free but not open source)
- Make sure it exports to `.bib` files
- Search for references (google scholar, jstor, etc.)
- Mendeley can import `.pdf`'s

Installation

1. First, install LaTeX
   - in default location
2. Install, TeXstudio
   - Which can then automatically find LaTeX

Basic set-up of a LaTeX file

```
\documentclass[]{article}
%opening
\title{}
\author{}

\begin{document}

\maketitle

\begin{abstract}
\end{abstract}

\section{}

\end{document}
```

Assignment 1

Create an abstract, title, authors, date, table of contents and create a section and some subsections:

- Give a date: `\date{}`
- Create subsections: `\section{}`, `\subsection{}`, `\subsubsection{}`, `\chapter{}`
- Insert a table of contents: `\tableofcontents{}`

Further text control

- itemization

```
\begin{itemize}
\item bla bla bla
\end{itemize}
```

- enumeration

```
\begin{enumerate}
\item bla bla bla
\end{enumerate}
```

- bold: `\textbf{}`
- emphasize: `\textit{}` or `\emph{}`

Inserting equations

- Inline: \$e=mc\^2\$ will be $e = mc^2$ or

  ```
  \begin{equation}
  e=mc^2
  \end{equation}
  ```

  will render in

$$e = mc^2 \tag{1}$$

- Equations can be as complex (cool) as you want
- Cheat sheet mathematics:

Assignment 2:

Produce the well-known univariate regression formula:

$$y_i = \alpha + \beta x_i + \epsilon_i$$

Referencing

- Internally:
  - `\label{}`, `\ref{}`
- footnotes (different symbol in title)
  - `\footnote{}`
- literature:
  - `cite{}`
- Bibliography:
  - `\bibliography{database1}`
  - for style: `\bibliographystyle{}`

Finally

- Several spaces of new lines are treated as one space or new line
- Some characters can not be used directly but with \ in front:
  - not: `# $ % ^ & _ { } ~ \`
  - but: `\# \$ \% \^ \& \_ \{ \} \~ \\`
- Commands always start with \
- Comments start with %

What are we going to do next time?

- Use of packages
- Figures
- Tables
- Automating do file outputs
- Slides