Data Science

Writing and Writing Tools

KEA – Data Science Elective PBA – Fall 2019 – Henrik Strøm

September 5, 2019

Table of Contents

Why All The Trouble?

"If I have seen further it is by standing on the shoulders of Giants."
– Sir Isaac Newton

- No need to reinvent the wheel.
- We use references to underpin our claims
- Our research is no stronger than the foundation it is built upon.

Finding Reference

Good sources for finding references:

- `https://scholar.google.com` – you can make a profile here and make bookmarks etc.
- Libraries
- University Libraries
- Beware of predatory journals

Demo

Demo: Google Scholar

Why Use a Reference Manager?

- It's another source of finding references
- Keep track of references over (long) time
- Integrate into writing tools
- Everything in one place
- I use Mendeley, but there are many other tools out there

Demo

Demo: Mendeley

Why LaTeX?

- Separate content from style (just like HTML/CSS)
- Focus on content
- Easily change template
- Everything is text, so it works well with Git
- Easy to integrate "live" data files
- You can write your own packages
- LaTeXis DRY
- Free and Open Source
- Used extensively in scientific communities

Short Intro to LaTeX

As simple as it gets:

```
\documentclass{article}

\begin{document}
First document. This is a simple example, with no extra parameters or
packages included.
\end{document}
```

Other document classes: book, report, beamer etc.
See https://tex.stackexchange.com/questions/782/what-are-the-available-documentclass-types-and-their-uses

Useful Tips

Support for international characters:

```
\usepackage[utf8]{inputenc}
```

Support for syntax-highlighted source code:

```
\usepackage{minted}
\definecolor{mintedbackground}{rgb}{0.95,0.95,0.95}
\setminted{bgcolor=mintedbackground,fontsize=\small,baselinestretch=1}
%\usemintedstyle{tango}
\inputminted{latex}{src/02-03.tex}
```

More Useful Tips

Setting up citation style:

```
\usepackage[backend=bibtex,style=authoryear,natbib=true]{biblatex}
\addbibresource{data-science.bib}
\usepackage[autostyle=true]{csquotes}
```

Custom packages:

```
\ProvidesPackage{preamble-defaults}[2019/01/27 Defaults Data Science Slides]
```

Then use this package:

```
\usepackage{preamble-defaults}
```

Demo

Demo: LaTeX

Writing a Good Paper

- Short and to-the-point
- Well referenced
- Figures and tables have captions, and are referenced from the text
- Use of cross-references
- Clearly stating purpose (e.g. research question) and concluding on that
- Outlining the state-of-the-art
- Clearly show methods behind the work
- Clear separation of analysis and findings

Exercise – Individual

Set up a LaTeXproject in Overleaf (`https://overleaf.com`) with a linked-up Mendeley bibliography.

The paper must include:

- At least one figure
- At least one external source file with syntax-highlighting
- At least one table
- At least three sections, seven subsections
- At least three bibliography references
- At least three internal cross-references
- At least one custom package
- Try change the template (extract all changes)
- Try use columns
- Get creative!