zeppelin universität

zwischen
Wirtschaft Kultur Politik

LaTeX and Reproducible Research

Beating the Monsters of Text-Setting and Repetitions

David Zimmermann, MSc
Email: david.zimmermann@zu.de | Blog: https://datashenanigan.wordpress.com/

05 October 2016

Institute of Entrepreneurship and Finance | Zeppelin Universität

Outline

Introduction

Motivation

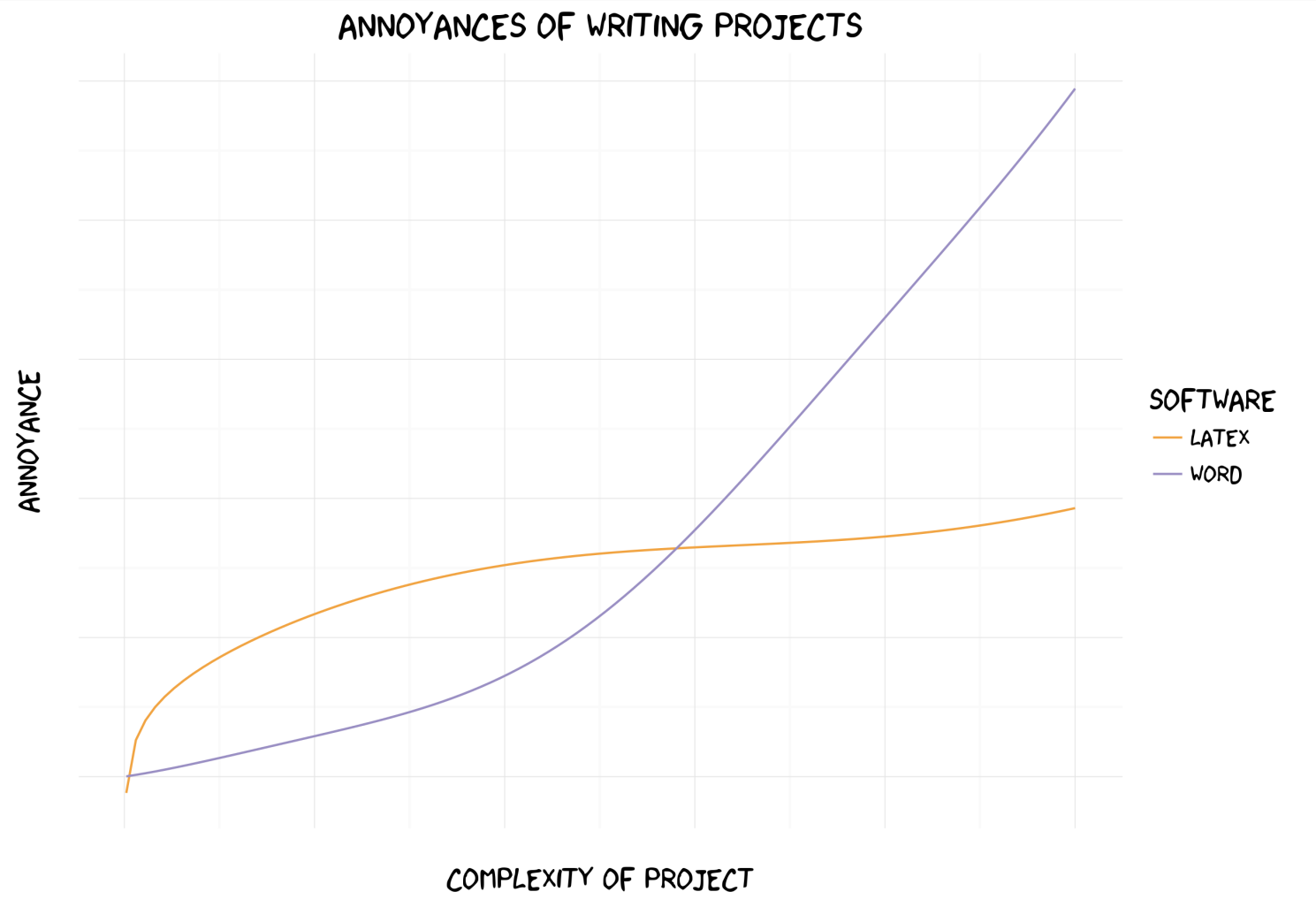


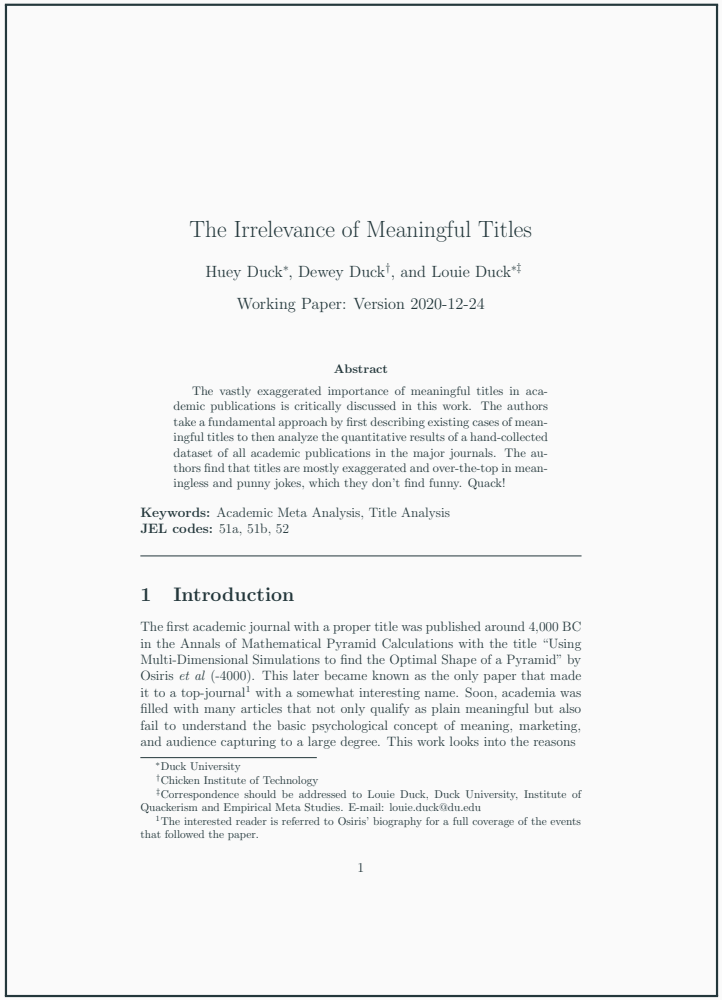
Figure 1: A completely realistic representation of project work

What is our Goal?

What is our goal?

- Tell a compelling story where the reader can concentrate on the content and is not annoyed by formatting
- Publication-ready output
  - Text, presentation, book, poster, ...
  - Table of Contents/List of Figures/List of Tables
  - Figures & Tables
  - (Complex) Math
  - Citations & References
- Spend resources on writing and minimize resources spent on formatting/setting

What is a Desired Result?



The Irrelevance of Meaningful Titles

Huey Duck*, Dewey Duck†, and Louie Duck*‡

Working Paper: Version 2020-12-24

Abstract

The vastly exaggerated importance of meaningful titles in academic publications is critically discussed in this work. The authors take a fundamental approach by first describing existing cases of meaningful titles to then analyze the quantitative results of a hand-collected dataset of all academic publications in the major journals. The authors find that titles are mostly exaggerated and over-the-top in meaningless and punny jokes, which they don't find funny. Quack!

Keywords: Academic Meta Analysis, Title Analysis
JEL codes: 51a, 51b, 52

1 Introduction

The first academic journal with a proper title was published around 4,000 BC in the Annals of Mathematical Pyramid Calculations with the title "Using Multi-Dimensional Simulations to find the Optimal Shape of a Pyramid" by Osiris *et al* (-4000). This later became known as the only paper that made it to a top-journal[1] with a somewhat interesting name. Soon, academia was filled with many articles that not only qualify as plain meaningful but also fail to understand the basic psychological concept of meaning, marketing, and audience capturing to a large degree. This work looks into the reasons

*Duck University
†Chicken Institute of Technology
‡Correspondence should be addressed to Louie Duck, Duck University, Institute of Quackerism and Empirical Meta Studies. E-mail: louie.duck@du.edu

[1]The interested reader is referred to Osiris' biography for a full coverage of the events that followed the paper.

1

Figure 2: A possible result

Why & What LaTeX

LaTeX [$\sim$ Latech or Latec] is:

- Developed by academics and professionals for academics and professionals
- Scripting-language (in plain-text) not WYSIWYG[1]
- Open-source (No cost, huge community: tex.stackexchange.com)
- Offers a wide variety of packages for mostly everything
- Stylesheets for most academic journals
- Beautiful tables directly exported from R/stata/Matlab/SPSS(?)
- Vector graphics (pdf-import, tikz)

[1] "What-You-See-Is-What-You-Get" i.e., MS Word.

Installation

LaTeX included in *MikTeX* (Windows only), *MacTeX* (Mac), or *TexLive* (Ubuntu).

Similar to R & RStudio, LaTeX works best with an IDE. Many alternatives available, I use *TexMaker* (alternatives include *TexStudio*, *TexitEasy*).

Download and Installation:

- *MikTeX* (Windows): `http://miktex.org/download`
- *MacTeX* (Mac): `http://www.tug.org/mactex/mactex-download.html`
- *TeXLive* (Ubuntu): `sudo apt-get install texlive-full`
- *TexMaker*: `http://www.xm1math.net/texmaker/download.html` or: `sudo apt-get install texmaker`

Article: MWE

A Minimal-Working Example (MWE)

```latex
\documentclass[12pt,a4paper]{article}
\usepackage[utf8]{inputenc}
\usepackage{amsmath}
\usepackage{amsfonts}
\usepackage{amssymb}

\begin{document}
This is a text
\end{document}
```

Listing 1: MWE LaTeX

An MWE Annotated

```
1  % DOCUMENTCLASS: article, beamer. font-size and
       paper-size
2  \documentclass[12pt,a4paper]{article}
3  % HEADER: Commands for the compiler
4  \usepackage[utf8]{inputenc} % proper use of
       special characters
5  \usepackage{amsmath} % beautiful math
6  \usepackage{amsfonts} % beautiful fonts
7  \usepackage{amssymb} % beautiful symbols
8
9  % DOCUMENT: Where the actual text goes
10 \begin{document}
11 This is a text
12 \end{document}
```

Listing 2: MWE Annotated

An MWE Result

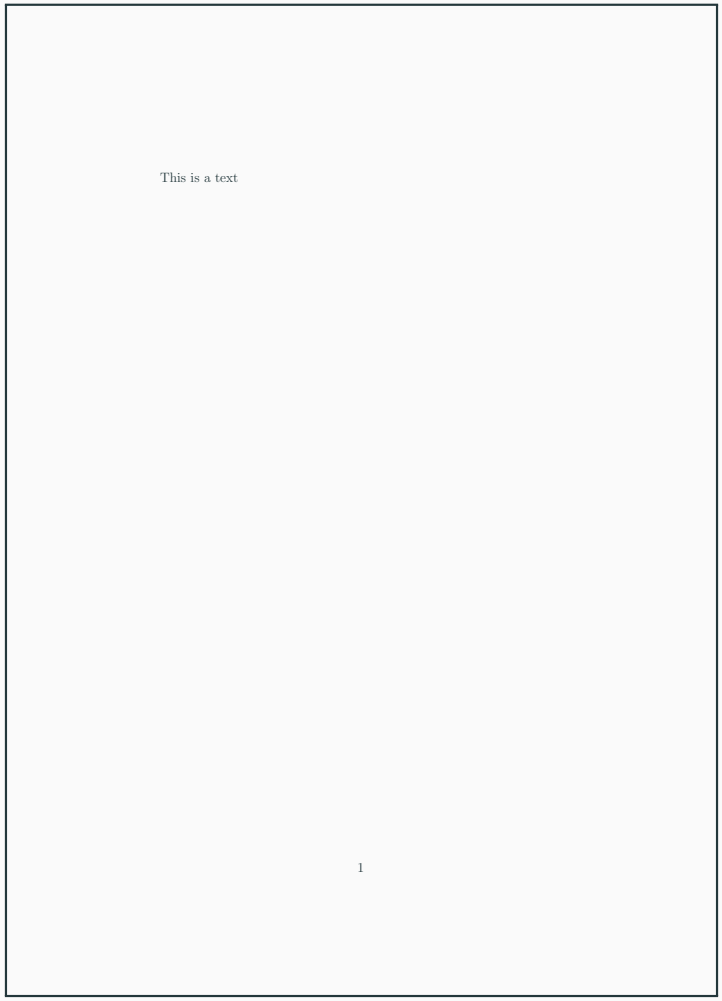


Figure 3: An MWE (Annotated) Result

Text

Text

Everything between `\begin{document}` and `\end{document}`

LaTeX -script (in `main.tex`):

```
1 This is a text
2
3 This text goes in the
    next paragraph,
4 whereas this text
    continues
```

Output (in `main.pdf`):

This is a text

This text goes in the next paragraph,
whereas this text continues

Text cont'd

Special characters ($ % & { } _ etc) need escaping using the "\"-operator[2]

LaTeX -script (in `main.tex`):

```
1 The stock ABC \& Inc.
    rose by 10\%
2
3 The new value is
    104.23\$ (curly
    braces are included
    like this \{tada\})
```

Output (in `main.pdf`)

The stock ABC & Inc. rose by 10%
The new value is 104.23$ (curly braces are included like this {tada})

---

[2] More information: `https://en.wikibooks.org/wiki/LaTeX/Special_Characters`

Title and maketitle

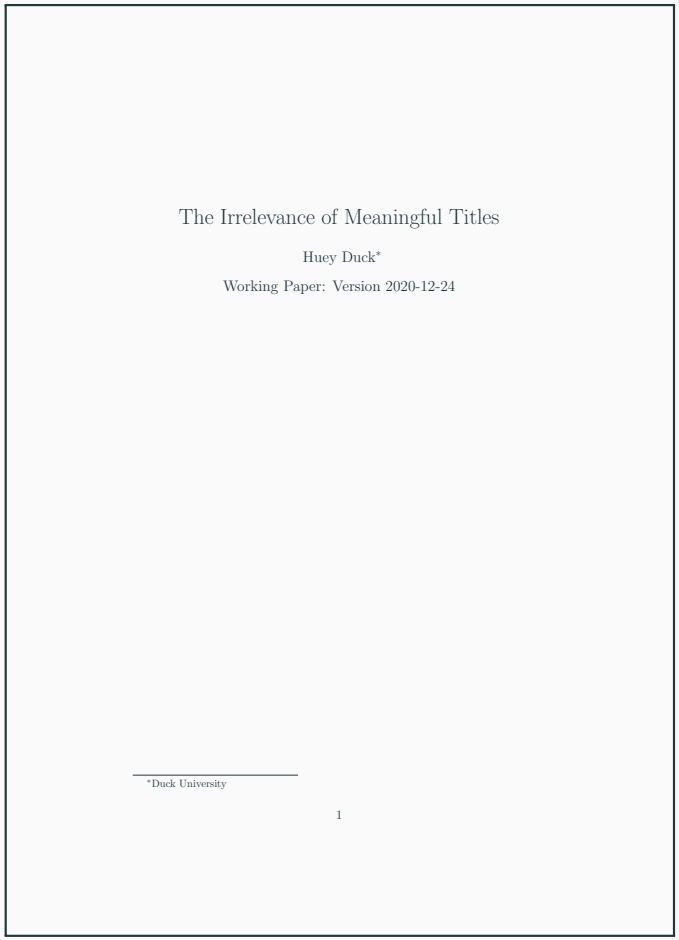
Title, author and affiliation are declared in the header (before `\begin{document}`), `\maketitle` (inside document) creates the title

---

LaTeX -script (parts of `main.tex`):

```
6  % ...
7  \title{The Irrelevance
       of Meaningful Titles}
8  \author{Huey Duck\thanks
       {Duck University}}
9  \date{Working Paper:
       Version 2020-12-24}
10
11 \begin{document}
12 \maketitle
13 % ...
```

Output (`main.pdf`):



Environments

Environments

```
\begin{ENVNAME}
ENVCONTENT (i.e., Text)
\end{ENVNAME}
```

`\begin(ENVNAME)` creates a new `ENVNAME`, `\end{ENVNAME}` ends the environment.

Environments are used for lists, figures, tables, equations, etc.

Lists (Itemize)

Listings are created using the `itemize`-environment. Items are created using the `\item`-command.

LaTeX -script (in `main.tex`):

```
1 \begin{itemize}
2 \item This is item 1
3 \item This is another
    item
4 \item[--] now with a
    dash
5 \item[a] and an ``a''
6 \end{itemize}
```

Output (in `main.pdf`)

- This is item 1
- This is another item

– now with a dash

a and an “a”

Numbered Lists (Enumerate)

Numbered listings are created using the `enumerate`-environment. Items are created using the `\item`-command.

LaTeX -script (in `main.tex`):

```
\begin{enumerate}
\item This is item 1
\item This is another item
\item[10] now with a number 10
\item and lastly this
\end{enumerate}
```

Output (in `main.pdf`)

1. This is item 1
2. This is another item

10 now with a number 10

3. and lastly this

Sections, Subsections & Table of Contents

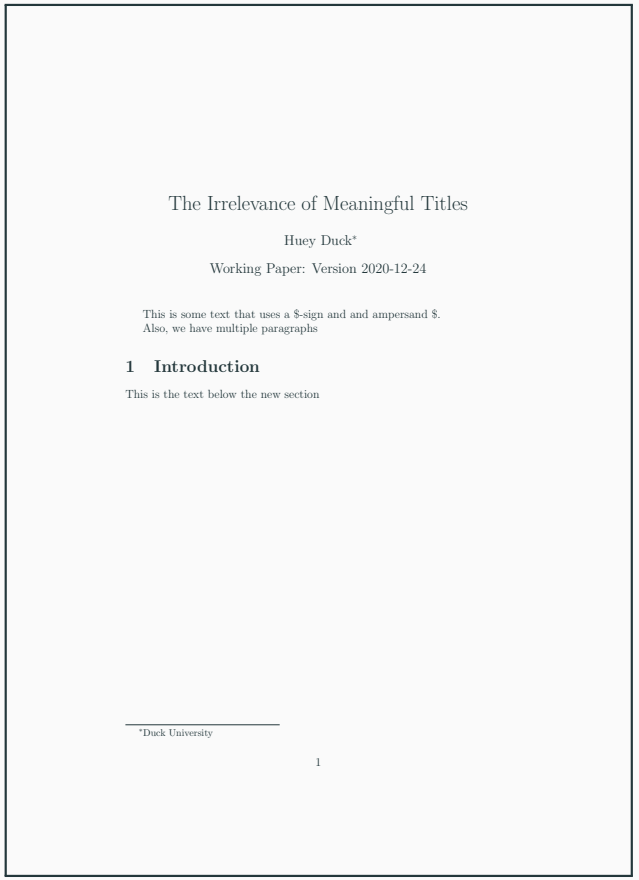
Sections

Sections are created using the `\section{}`-command and are numbered automatically. To surpress numbers (and appearance in the table-of-contents) use `\section*{}`.

LaTeX -script (parts of `main.tex`):

```
17 % ...
18 \section{Introduction}
19 This is the text below
       the new section
20 % ...
```

Output (`main.pdf`):



The Irrelevance of Meaningful Titles

Huey Duck*

Working Paper: Version 2020-12-24

This is some text that uses a $-sign and and ampersand $.
Also, we have multiple paragraphs

1 Introduction

This is the text below the new section

*Duck University

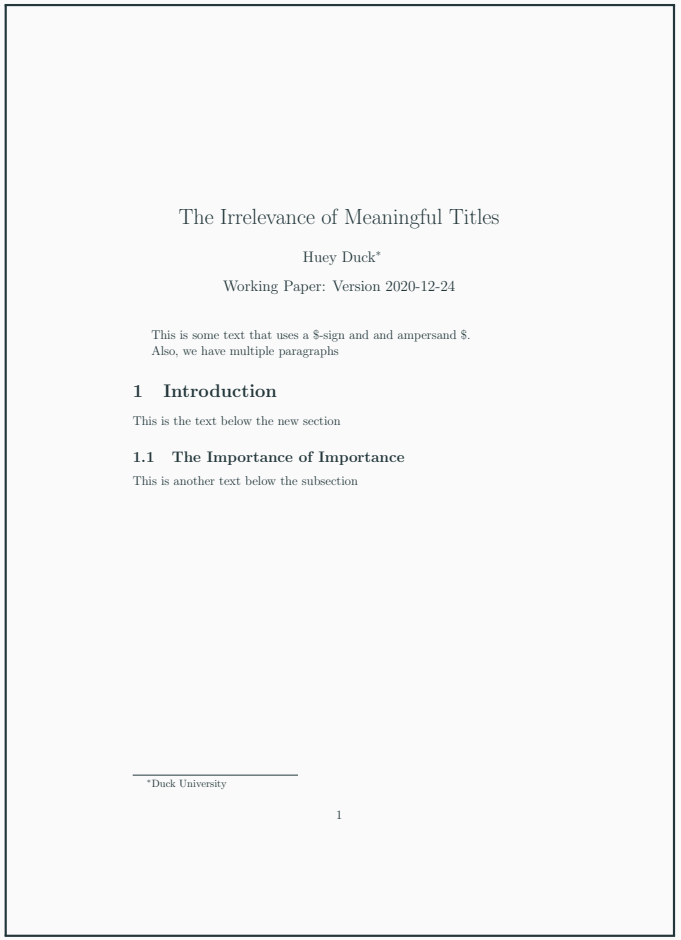
1

Subsections

LaTeX -script (parts of `main.tex`):

```
20 % ...
21 \subsection{The
       Importance of
       Importance}
22 This is another text
       below the subsection
23 % ...
```

Output (`main.pdf`):



The Irrelevance of Meaningful Titles

Huey Duck*

Working Paper: Version 2020-12-24

This is some text that uses a $-sign and and ampersand $.
Also, we have multiple paragraphs

1 Introduction

This is the text below the new section

1.1 The Importance of Importance

This is another text below the subsection

*Duck University

1

Table of Contents

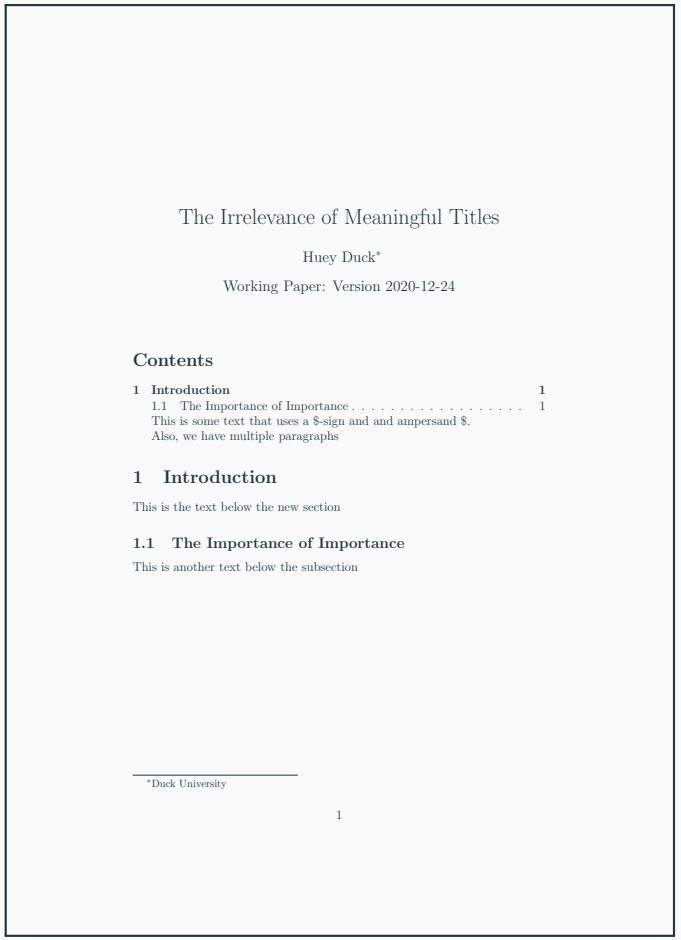
After using `\tableofcontents`, you need to run the compiler twice.

Also useful: `\newpage` after `\tableofcontents`, to start a new page.

LaTeX -script (parts of `main.tex`):

```
13 %... after \maketitle
14 \tableofcontents
15 %... before the first
       text
```

Output (`main.pdf`):



The Irrelevance of Meaningful Titles

Huey Duck*

Working Paper: Version 2020-12-24

Contents

1 Introduction

This is the text below the new section

1.1 The Importance of Importance

This is another text below the subsection

*Duck University

1

Figures

Figures

Figures are environments as well. Options include the placing `[ht]`[3].

LaTeX -script (parts of `main.tex`):

```
28 %... \usepackage{graphicx}
29 \begin{figure}[ht]
30 \centering
31 \includegraphics[width=0.5\
       textwidth]{pictures/
       homer} % picture in
       folder "pictures"; no .
       png
32 \caption{Homer Simpson}
33 \end{figure}
34 %...
```

Output (`main.pdf`):

Figure 1: Homer Simpson

2

[3] *h* for here first, then *t* top. Also *b* for bottom or *p* for appendix. More Information: `https://en.wikibooks.org/wiki/LaTeX/Floats,_Figures_and_Captions`

Tables

Tables

Tables are environments as well. But the actual table is included in a *tabular*-environment[4]. Next col: `&`; next row: `\\`; line `\hline`

LaTeX -script (in `main.tex`):

```
\begin{table}[ht] % same
\caption{A nice table}
\begin{tabular}{lcr}
\hline
Name & Dir & Sales \\
\hline
Alice & C & \$12,000\\
Bob & R & \$17,000\\
\hline
\end{tabular}
\end{table}
```

Output (in `main.pdf`)

Table 1: A nice table

Name	Dir	Sales
Alice	C	$12,000
Bob	R	$17,000

[4]Options include **l**eft, **c**enter, and **r**ight orientation for columns. 5 cols would be **rrrrr**.
More Info: `https://en.wikibooks.org/wiki/LaTeX/Tables`

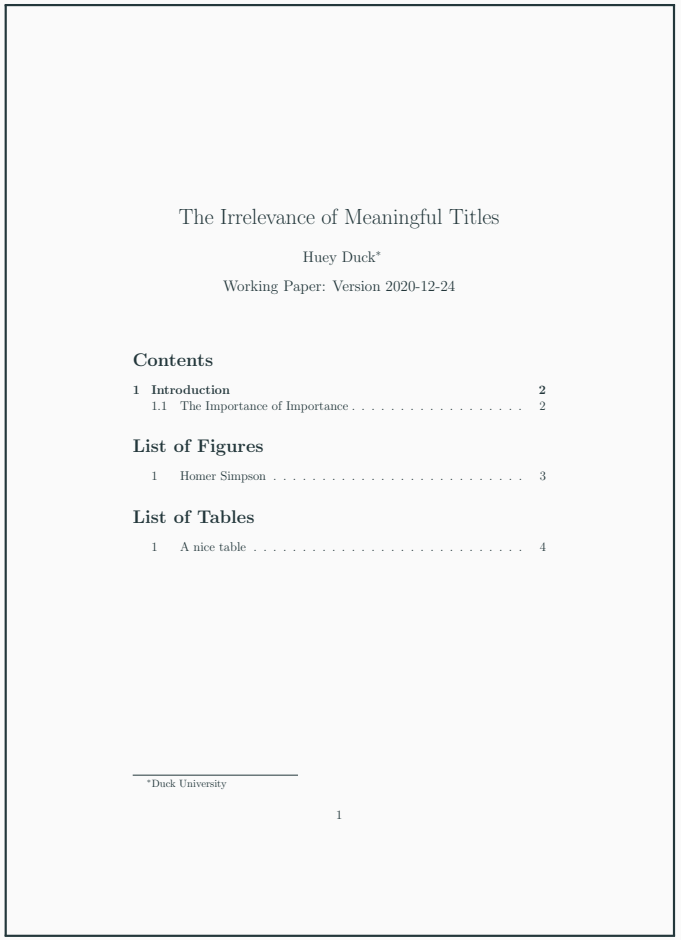
List of Figures and List of Tables

List of figures/list of tables can be included with `\listoffigures` and `\listoftables`.

LaTeX -script (parts of `main.tex`):

```
14 %... after \maketitle
15 \tableofcontents
16 \listoffigures
17 \listoftables
18 \newpage
19 %... before the first
       text
```

Output (`main.pdf`):



The Irrelevance of Meaningful Titles

Huey Duck*

Working Paper: Version 2020-12-24

Contents

*Duck University

1

Mathematics

Mathematics

Maths: the equation-environment. Everything you need to know:
https://en.wikibooks.org/wiki/LaTeX/Mathematics

LaTeX -script (in main.tex):

```
\begin{equation}
x = \sqrt[2]{y^2 + z^2}
\end{equation}

\begin{equation}
\mu = \frac{1}{n}\
    displaystyle\sum_{i
    =1}^n x
\end{equation}
% Inline Math
Given $ x = 3 $ we ...
```

Output (in main.pdf)

$$x = \sqrt[2]{y^2 + z^2} \tag{1}$$

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x \tag{2}$$

Given $x = 3$ we ...

References

References

Want to reference a formula from the last section? Or a section itself? Or some other environment (table, picture, ...) → references (`\label{}` & `\ref{}`).[5] Also: links!

LaTeX -script (in `main.tex`):

```
\begin{equation}
\lim_{x \to \infty} \exp
    (-x) = 0
\label{eq:limits} %
    after!
\end{equation}

Equation \ref{eq:limits}
    refers to a ...
```

Output (in `main.pdf`)

$$\lim_{x \to \infty} \exp(-x) = 0 \qquad (3)$$

Equation 3 refers to a ...

[5] After adding references, the work needs to be compiled twice (doesn't add with TOCs, i.e., twice not four-times). If you get **??** recompile and/or check names (keys).

Citation and Bibliography

Academic Citations – The Bibfile

Idea: Use **BibLaTeX** to organize citations and formatting. Setting-up BibLaTeX (aka. Biber) [6]

Create another file with the ending *.bib* (your bib-file), which serves as a database, which we will populate with possible entries for our work.

Pro Tip: Google Scholar & citation export to BibTeX, or Export from Mendeley/Citavi/....

Include the packages and point to the bib-file with

```
1 % in the header:
2 \usepackage[style=authoryear, backend=biber,
    maxnames=3]{biblatex} % for the citation,
3 \addbibresource{my_bib_file.bib}
```

[6] **Win:** http://tex.stackexchange.com/questions/63039/install-biber-in-miktex-on-a-64-bit-version-of-windows
Mac: http://tex.stackexchange.com/questions/153359/setting-up-texmaker-on-mac-to-work-with-biber

Filling the Bibfile

A Minimum Working Example Bibfile in *my_bib_file.bib*

```
@article{Coase1937, %<- This is the key
  title={The nature of the firm},
  author={Coase, Ronald H}, % <- multiple
      authors with "and", i.e,
  %{Coase, R H and Williamson, O E}
  journal={economica},
  volume={4},
  number={16},
  pages={386--405}, % long dash --, not minus -
  year={1937},
  publisher={Wiley Online Library}
}
```

More formats: https://en.wikibooks.org/wiki/LaTeX/Bibliography_Management#BibTeX

Citing a Work

`\textcite{KEY}`, `\parencite{KEY}`, or `\citeauthor{KEY}` [7]

Options can specify page i.e., `\textcite[p. 390]{Coase1937}`

LaTeX -script (in `main.tex`):

```
1 \textcite{Coase1937}
    said that
2
3 yada \parencite{Coase
    1937}.
4
5 \citeauthor{Coase1937}'s
    argument
6
7 \textcite{Coase1936}
    argued that
```

Output (in `main.pdf`)

Coase (1937) said that
yada (Coase 1937).
Coase's argument ...
Coase1936 argued that

[7] Needs compilations: normal + bibtex (biber) + normal + normal

Bibliography

Inserting bibliography at the end of the work with `\printbibliography`

LaTeX -script (in `main.tex`):

```
1 \printbibliography
```

Output (in `main.pdf`)

Omitted here, see next slide

References

 Coase, Ronald H (1937). "The nature of the firm". In: *Economica* 4.16, pp. 386–405.

Reproducible Research

Workflow of a Project

Idea: Create an automated workflow, that allows for reproduction (applies mostly for quantitative work)

- Document everything that you do[8]
- Fragmentize your (quantitative) work into scripts. One script per task
- Create a clear workflow of the project. What script depends on which input and produces which output?
- Well-documented functions/workflow including the dependencies and requirements (README.txt/.md)[9]
- If possible, use git or other version control software, and makefiles to automate the build-process

More information: `http://kbroman.org/steps2rr/`

[8] Don't touch the raw data with anything other than a script, and don't overwrite it.

[9] A README.md (markdown) example can be found here:
`https://github.com/DavZim/Independent-Event-Identifier-IEI-`

Typical Workflow

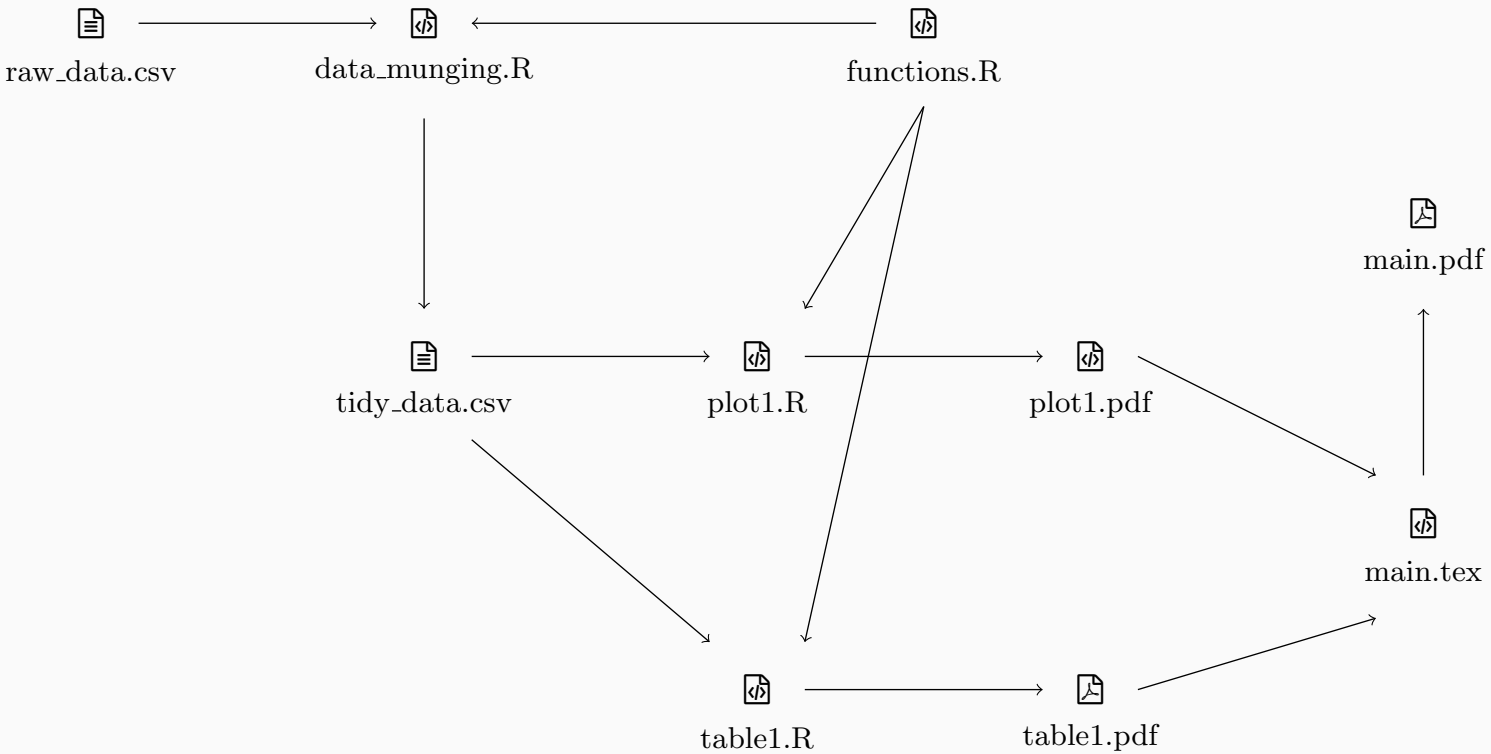


Figure 4: R/LaTeX -Workflow

R substitutional with any other scripting-language (Python, Stata, Matlab, . . .)

File Structure

```
.
+-- data
|   +-- raw_data.csv
|   +-- tidy_data.csv (created by tidy_data.R)
+-- tex-files
|   +-- plots
|       +-- plot1.pdf (created by plot1.R)
|   +-- tables
|       +-- table1.tex (created by table1.R)
|   +-- main.tex
|   +-- ... (other tex-files)
|   +-- my_bib_file.bib
|   +-- main.pdf
+-- R
|   +-- functions.R (containing all functions used)
|   +-- tidy_data.R (clean and tidy the data)
|   +-- plot1.R (creates plot1.pdf)
|   +-- table1.R (creates table.tex)
+-- R_projects_data.RProj
+-- README.MD (this file)
+-- README.html (this file in HTML)
(+-- Makefile)
```

Figure 5: A Typical Filestructure

Plots with R/LaTeX

Use R to create a plot, save it as a `.pdf` and include it in LaTeX.

R-script (in `plot1.R`):

```r
library(ggplot2) # plots
library(readr) # data io
dt <- read_csv("data/tidy_data.csv")

plot1 <- ggplot(dt, aes(x = price, y = carat, color = cut)) +
geom_point()

ggsave("plots/plot1.pdf", plot1)
```

LaTeX -script (in `main.tex`):

```latex
\begin{figure}
\centering
\includegraphics[width = \textwidth]{plots/plot1}
\caption{A nice, R-exported plot}
\end{figure}
```

Output on the next slide

Plots with R/LaTeX Output

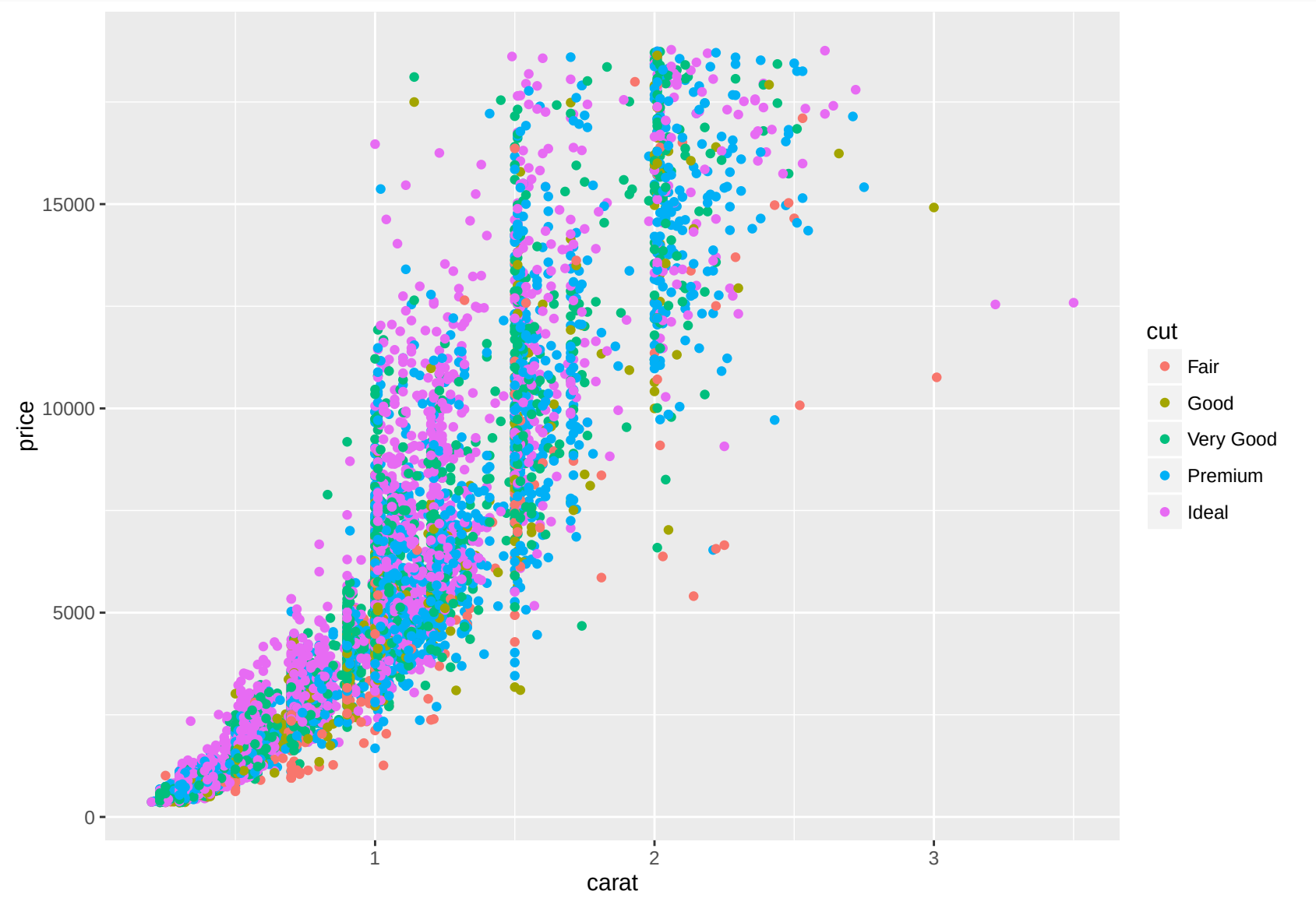


Figure 6: A nice (yet basic), R-exported plot

Tables with R/LaTeX

Use R to create a regression-table, save it as a `.tex`-file and include it in LaTeX. R-packages: `texreg`, `stargazer`, or others

---

R-script (in `table1.R`):

```r
library(texreg)
library(readr)
dt <- read_csv("data/tidy_data.csv")

reg1 <- lm(data = dt, price ~ carat)
reg2 <- lm(data = dt, price ~ carat + x)
texreg(list(reg1, reg2), table = F, file = "tables/table1.tex")
```

LaTeX -script (in `main.tex`):

```latex
\begin{table}
\input{tables/table1.tex}
\caption{A nice R-created table}
\end{table}
```

Output on the next slide

Tables with R/LaTeX Output

	Model 1	Model 2
(Intercept)	-2292.86^{***}	2952.55^{***}
	(29.87)	(240.93)
carat	7814.95^{***}	10935.96^{***}
	(32.00)	(145.69)
x		-1350.22^{***}
		(61.56)
R^2	0.86	0.86
Adj. R^2	0.86	0.86
Num. obs.	10000	10000
RMSE	1527.45	1492.05

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: A nice R-created table

Typical Workflow

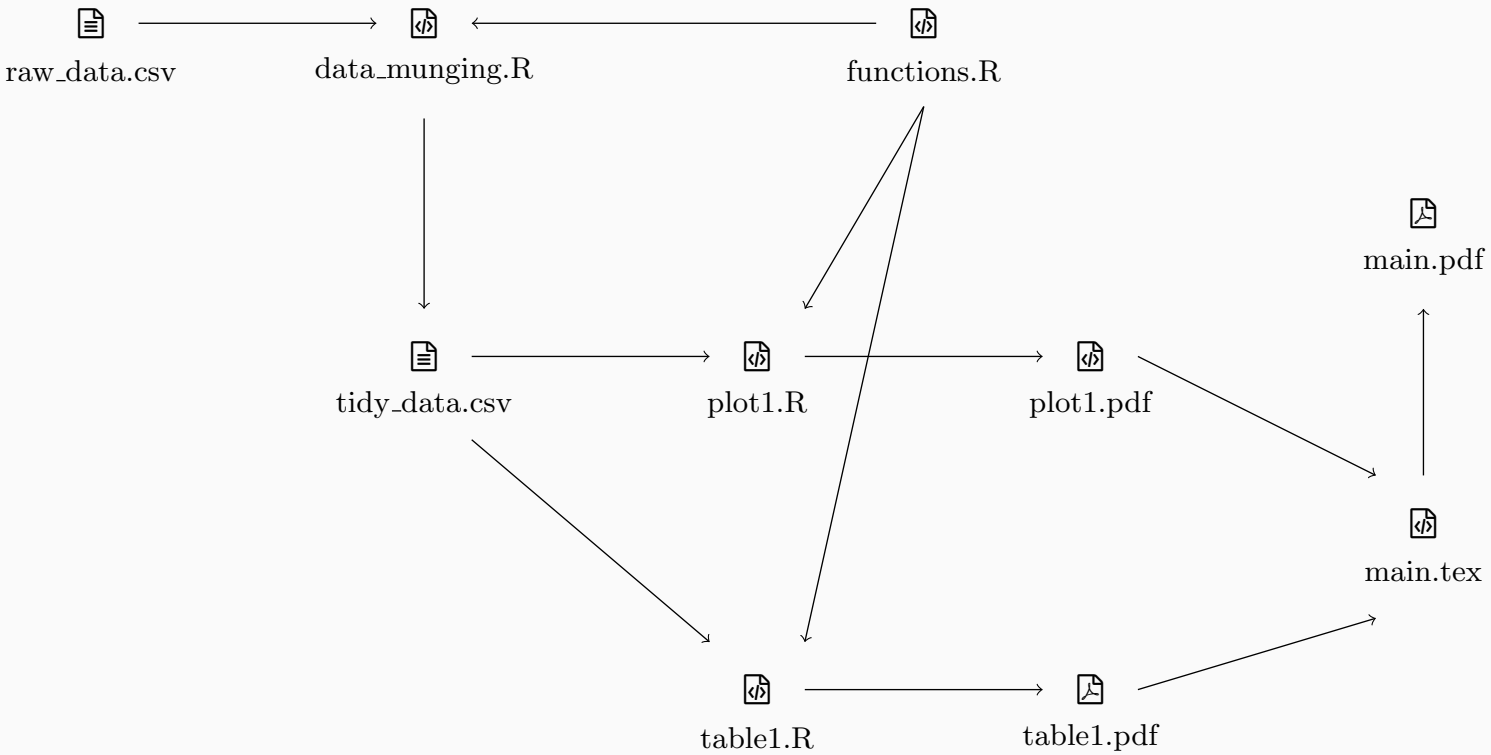


Figure 7: R/LaTeX -Workflow

Additional Infos

Further Reading

Useful links:

- Wikibooks ftw: `https://en.wikibooks.org/wiki/LaTeX`
- Questions: Google and `tex.stackexchange.com`
- Online TeX-editor: `www.overleaf.com`

Templates:

- ZU: Assignments, BA/MA-Thesis, Make-example: `https://github.com/DavZim/Templates`
- Assignments: `https://www.overleaf.com/read/twpcrktdhwqy`
- BA-Thesis: `https://www.overleaf.com/read/fdmrpcbmxfqx`
- MA-Thesis: `https://www.overleaf.com/read/jxsyvpcbtbss`

Further Reading Advanced

Reproducible Research:

- Reproducible reseach intro: `https://ropensci.org/blog/2014/06/09/reproducibility/`
- Makefile (R): `http://kbroman.org/minimal_make/`
- Makefiles: `https://bost.ocks.org/mike/make/`

Questions?