

Developing Good Research Habits

Rob J Hyndman

21 March 2023

bit.ly/research_habits



Outline

1 Citing

2 Searching

3 Scripting

4 Writing

Alternative titles

There's more to being a researcher than writing a thesis

Alternative titles

There's more to being a researcher than writing a thesis

What you should know but probably don't

Alternative titles

There's more to being a researcher than writing a thesis

What you should know but probably don't

Listen up, young padawans



Outline

1 Citing

2 Searching

3 Scripting

4 Writing

Managing references

Zotero

- ▶ Free and on all operating systems
- ▶ Web-version and local version synced
- ▶ Browser extension for adding papers/books
- ▶ Attach notes and annotations to papers.
- ▶ Works with Word, LibreOffice or LaTeX.
- ▶ Generate bibliography automatically
- ▶ Handles all formatting for you.

The logo for Zotero, featuring the word "zotero" in a lowercase, sans-serif font. The letter "z" is colored red, while the rest of the letters are black.

To install:

- Set up account at www.zotero.org
- Download from www.zotero.org

Managing references

Mendeley

- ▶ Free and on all operating systems
- ▶ Web-version and local version synced
- ▶ Browser extension for adding papers/books
- ▶ Attach notes and annotations to papers.
- ▶ Works with Word, LibreOffice or LaTeX.
- ▶ Generate bibliography automatically
- ▶ Handles all formatting for you.



To install:

- Set up account at www.mendeley.com
- Download from www.mendeley.com

Managing references

Paperpile

- ➔ \$3 per month and runs on Google Chrome
- ➔ Papers stored on Google Drive
- ➔ Browser extension for adding papers/books
- ➔ Works with Google Docs or LaTeX.
- ➔ Generate bibliography automatically
- ➔ Handles all formatting for you.
- ➔ Amazingly fast



Paperpile

To install:

- Set up account at paperpile.com
- Download Google chrome browser extension

What to cite?

- Cite what is important.
- Cite (only) what is relevant.
- Avoid lists of gratuitous references.
- Include proper citations for all packages and software you use.



What to cite?

- Cite what is important.
- Cite (only) what is relevant.
- Avoid lists of gratuitous references.
- Include proper citations for all packages and software you use.



When using R

```
citation("packagename")
```

Sight what you cite

- Every article cited should be sighted, & preferably read.
- At the very least, check that the article cited really does say what you think it says.
- Type the reference information yourself.
- Don't just cite what other people say about citations.
- Store accurate reference info from the start.
- Give credit where it is due.



Sight what you cite

- Every article cited should be sighted, & preferably read.
- At the very least, check that the article cited really does say what you think it says.
- Type the reference information yourself.
- Don't just cite what other people say about citations.
- Store accurate reference info from the start.
- Give credit where it is due.
 - ▶ Diebold did not invent PITs.
 - ▶ Hyndman did not invent exponential smoothing or ARIMA models.



Outline

1 Citing

2 Searching

3 Scripting

4 Writing

Google Scholar

- Searching for papers
- Use advanced search
- Link GS to your reference manager
- Track citations of key papers
- Star papers for your own library
- Check recommended articles
- Check author profiles, especially highly cited authors
- Create your own GS profile once you have (at least) one paper
- Follow key authors in your area

Use the magic button!

The screenshot shows a web browser window with the URL [sciedirect.com/science/article/pii/S0169207019301633](https://www.sciencedirect.com/science/article/pii/S0169207019301633). The page is from the International Journal of Forecasting, Volume 36, Issue 2, April–June 2020, Pages 324-333. The title of the article is "Temperature anomaly detection for electric load forecasting". The authors listed are Masoud Sobhani^a, Tao Hong^a, Claude Martin^b. Below the title, there is a "Get Access" button, a "Share" button, and an "Export" button. To the right of the article summary, there is a "Search ScienceDirect" search bar and a "Advanced" search button. On the left side, there is a sidebar with links to "Outline", "Abstract", "1. Introduction", "2. Data", "3. A load-based temperature prediction model", "4. Model-based anomaly detection", "5. Case study", "6. Conclusion", "Acknowledgment", "Appendix A. The trend variable", and "Appendix B. Supplementary data". At the top of the page, there is a navigation bar with various icons and a "Pandemic..." link. The top right corner features "Register" and "Sign in" buttons.

Use the magic button!

The screenshot shows a ScienceDirect article page for the International Journal of Forecasting. A red circle highlights the 'Get Access' button, and a large red arrow points upwards from the bottom right towards the 'Search ScienceDirect' bar.

Outline

- Abstract
- 1. Introduction
- 2. Data
- 3. A load-based temperature prediction model
- 4. Model-based anomaly detection
- 5. Case study
- 6. Conclusion
- Acknowledgment
- Appendix A. The trend variable
- Appendix B. Supplementary data

ScienceDirect

Get Access Share Export

International Journal of Forecasting
Volume 36, Issue 2, April–June 2020, Pages 324-333

Temperature anomaly detection for electric load forecasting

Masoud Sobhani ^a, Tao Hong ^a✉, Claude Martin ^b

Show more

<https://doi.org/10.1016/j.jforecast.2019.04.022>

Get rights and content

Recommended articles

- Comparing density forecasts in a risk m...
International Journal of Forecasting, Volume 3...
- Purchase PDF View details
- Five dimensions of the uncertainty-disa...
International Journal of Forecasting, Volume 3...
- Purchase PDF View details
- Semi-parametric dynamic asymmetric L...
International Journal of Forecasting, Volume 3...
- Purchase PDF View details

1 2 Next >

Semantic Scholar

 Rob J Hyndman ▾



SEMANTIC SCHOLAR

A free, AI-powered research tool for scientific literature

Search 211,093,141 papers from all fields of science

Search Q.

Try: [Jean Louise Cohen](#) • [Market Structure](#) • [Cultural Universals](#)

Outline

1 Citing

2 Searching

3 Scripting

4 Writing

Reproducibility

Not reproducible:

- Data edited in a spreadsheet
- Click and point analysis
- Copy and paste graphs and tables
- Tables typed by hand

Reproducible

- All data edits scripted
- All analysis scripted
- Graphs and tables automatically pulled in to the thesis
- Tables generated with scripts



**STOP CLICKING
AND START SCRIPTING**

Reproducibility

Someone should be able to reproduce your thesis without having to guess what software you had installed, what versions, which files do what, etc.

Reproducibility

Someone should be able to reproduce your thesis without having to guess what software you had installed, what versions, which files do what, etc.

- Stay organized.
- One system for doing this using R is to write your thesis in an Rmarkdown or Quarto file.
- Track software versions

Version control

- `thesis_v1`, `thesis_v2`, etc., is not adequate version control.
- You need to track changes over time, have a *remote* repository, and be able to roll back as required.
- Your repository should contain *everything* required to produce your thesis including computer code, references, writing.
- Your repository should have an obvious structure and be fully documented.
- **Github** solves these problems
- Read “Happy git with R”: happygitwithr.com



Version control with git

- RStudio integrates with github, so version control is built in.
- But github can be used with *any* text-based language including Stata, Python, LaTeX, R, Rmarkdown, Quarto, markdown, etc.
- Git allows you to:
 - ▶ track changes
 - ▶ experiment in branches
 - ▶ undo
- Github provides:
 - ▶ backup and restore
 - ▶ synchronisation



Outline

1 Citing

2 Searching

3 Scripting

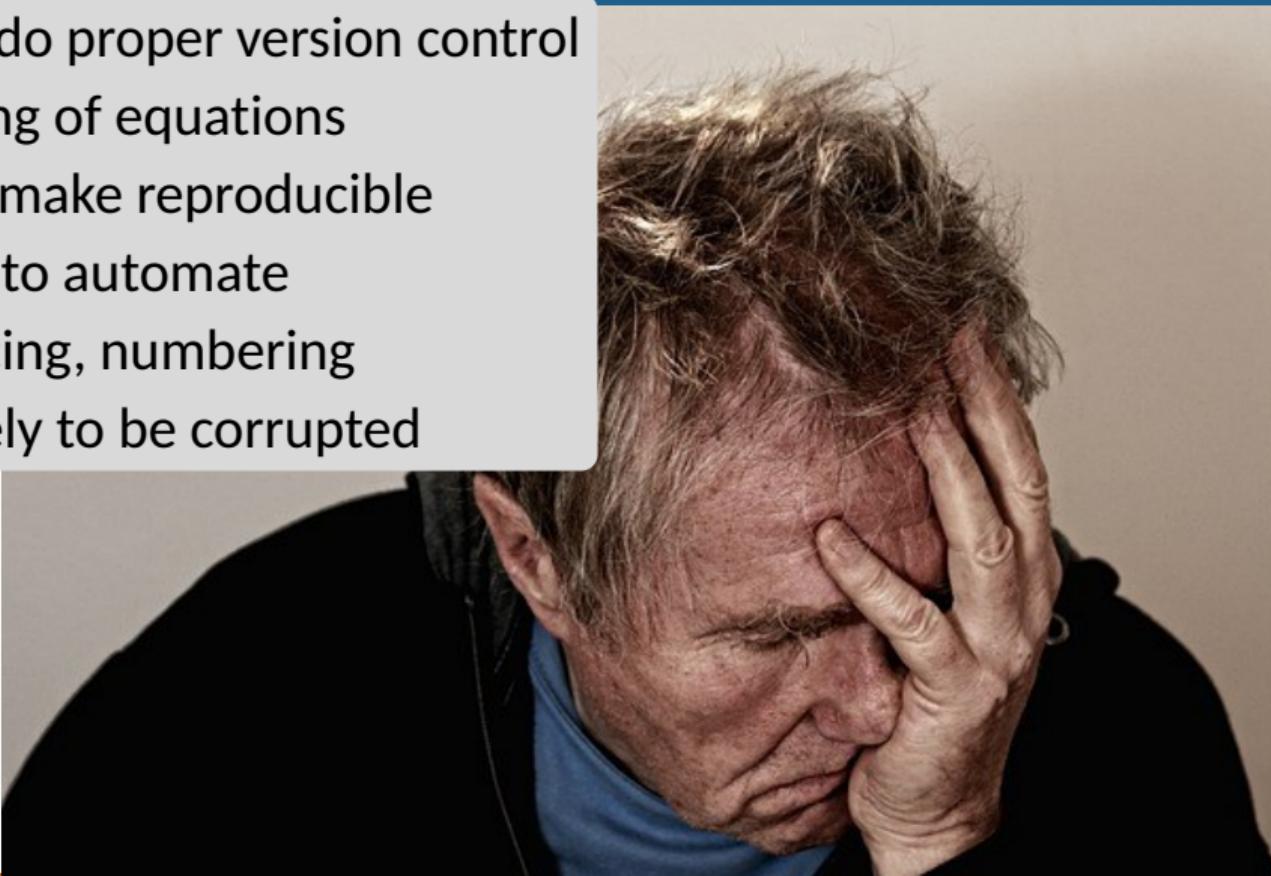
4 Writing

Microsoft Word



Microsoft Word

- Impossible to do proper version control
- Poor formatting of equations
- Impossible to make reproducible
- More difficult to automate cross-referencing, numbering
- Files more likely to be corrupted



LATEX

To install:

- Download MikTeX,
MacTeX or TeXlive.
- Download TeXStudio
from texstudio.sourceforge.net/

Document processing

- ➔ Free and open-source
- ➔ Available on all operating systems
- ➔ Used by every mathematical publisher
- ➔ Separate content from style
- ➔ Format complex equations
- ➔ Automatic numbering
- ➔ Automatic bibliography
- ➔ Almost every language

RMarkdown

My Thesis

A thesis submitted for the degree of
Bachelor of Commerce (Honours)

by
Sarah Baggart



Department of Econometrics and Business Statistics
Monash University
Australia

April 2017

RMarkdown

- ➔ Combines R and LaTeX into one system
- ➔ Reproducible research
- ➔ Monash Honours Template via monash package
github.com/numbats/monash
- ➔ Useful for assignments too
- ➔ See rmarkdown.rstudio.com for help

Quarto



This is my thesis

Susan Su

B.Sc. (Hons), University of Tangambalanga

A thesis submitted for the degree of Doctor of Philosophy at Monash University

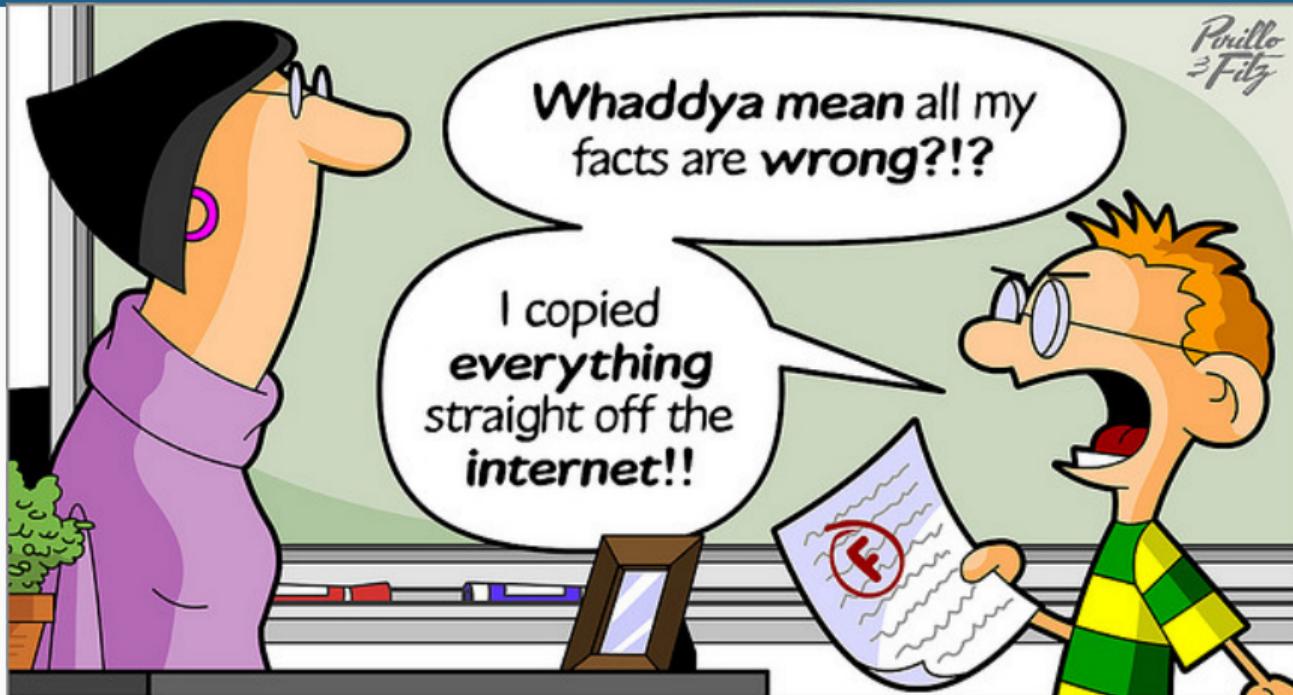
In 2022

Department of Econometrics & Business Statistics

Quarto

- ➔ Combines R, Python and LaTeX into one system
- ➔ Reproducible research
- ➔ Monash Thesis Template via github.com/numbats/monash-quarto-thesis
- ➔ Useful for assignments too
- ➔ See quarto.org for help

Plagiarism



- Never plagiarise from other papers — not even sentence fragments. Use your own words.

Resources

Slides and links:

bit.ly/research_habits