

Digital Research Toolkit for Linguists

Week 13: Literature management, the command line

Anna Pryslopska

July 1, 2024

Psycholinguistics and Cognitive Modeling Lab

Exam: 3 sessions left

MY EXAM DAY	
morning	 A portrait of Albus Dumbledore with a long white beard and hair, looking very tired or ill.
exam starts	 A close-up of Gollum's face, looking shocked or scared with wide eyes and a pale expression.
exam ends	 A close-up of Gandalf's face, shouting with a wide-open mouth and a determined expression.
I passed	 Students at a table cheering and raising their hands in excitement.
I failed	 Students at a table looking dejected and disappointed.

Exam

Exam date

Monday, July 29th 14:00–15:00, Kepler 17 (K2), M 17.52 (KP1705M 17.52)
Same time, different room! Should be in Campus.

Earlier exam date in justifiable cases

Wednesday, July 24th 11:00–12:00, Kepler 17 (K2), M 17.81 (KP1780GAM 17.81)

Different time, same room. Should be in Campus.

Term paper

You will receive new data on a different subject. Task: clean, analyse, summarize, do literature research, write a 15 page report in \LaTeX , upload data to Git (Quarto report, tex files, plots etc.)



Homework

Some of you knocked it out of the park!



Homework

Reference all the citations in the text, so that there is at least one of each of these:

- as a parenthetical reference
- as a textual reference
- reference only the author
- reference only the publication year
- reference only the title
- reference a book without a citation but include in bibliography

Homework

Continue writing the **Moses illusion** report: make a style file and add a reference section. Upload your TEX, PDF, BIB, image, and style files.

“After collecting demographic information like first language and student number, the experiment looped through an empty page with an asterisk, then cycled through parts of sentences at the press of the space bar. At the end of every sentence, the participants were asked to rate how natural the sentence sounds.”

“As demonstrated in figure 3.1, as the amount of noise introduced increases, so does the literal interpretation of the sentence. These findings highlight human adaptability n language processing.”

Glosses

Which internal organs can be seen through the toes?

Welche einheimischen Orgeln können sein gesehen durch die Zehen?

Welche inneren Organe kann man durch die Zehen erkennen?

- (1) Which internal organs can be seen through the
Welche einheimischen Orgeln können sein gesehen durch die
toes?
Zehen?

Welche inneren Organe kann man durch die Zehen erkennen?

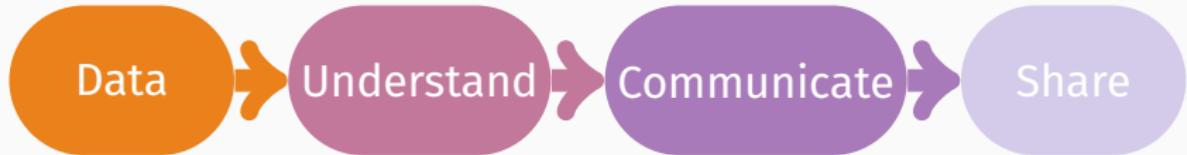
Glosses

```
\begin{exe}
\ex
Which continent's name starts with the letter X?
\glt ¿El nombre de qué continente comienza con la letra X?
\end{exe}
```

- (2) chew the fat
kauen das Fett
ein lockeres Gespräch führen

```
\begin{exe}
\ex
\gll chew the fat\\
      kauen das Fett\\
\glt ein lockeres Gespräch führen
\end{exe}
```

Questions?



R & RStudio,
packages, data
types, formats,
encoding

import from
workspace,
assign values,
operations,
clean, filter,
arrange,
select,
merge, group,
summarize,
export,
visualize

document,
research,
create clean
and beautiful
reports

connect,
collaborate,
backup

Table of contents

1. Reference managers
2. Literature research
3. Mastering the command line
4. Homework assignment

Reference managers

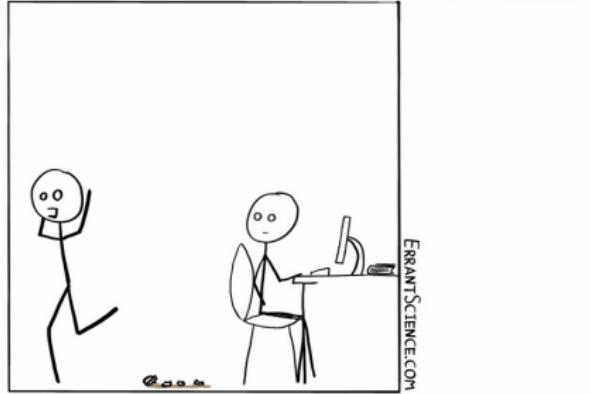
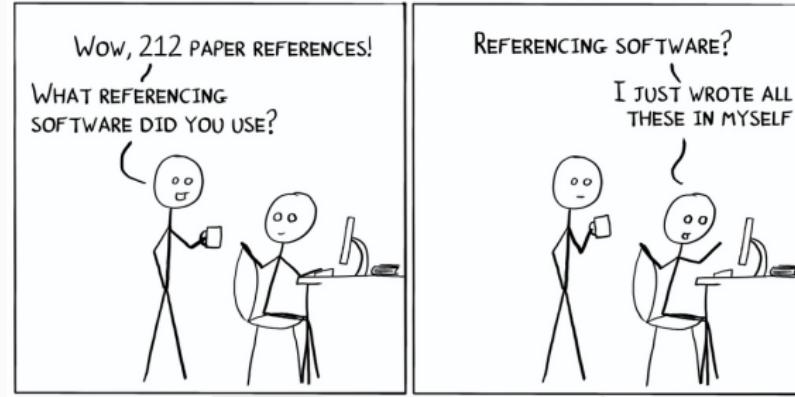
Managing large .bib files

You can keep your entire bibliography in a huge .bib file...

```
1 % ===== A
2 @article{almor1999,
3   Title      = {Noun-Phrase Anaphora and Focus: The Informational Load
4   Hypothesis},
5   Author     = {Almor, Amit},
6   Journal    = {Psychological Review},
7   Volume     = {106},
8   Number     = {4},
9   Pages      = {748},
10  Year       = {1999},
11  Publisher  = {American Psychological Association},
12  doi        = {10.1037/0033-295X.106.4.748}
13 }
14
15 @inproceedings{arche2017,
16   Title      = {Towards a Unified Treatment of Spanish Copulas},
17   Author     = {Arche, Mar\'ia J and F\'abregas, Antonio and Mar\'ia
18   n, Rafael},
19   Booktitle  = {Romance Languages and Linguistic Theory 11: Selected Papers
20   from the 44th Linguistics Symposium on Romance Languages},
21   Pages      = {33--52},
22   Year       = {2017},
23   Publisher  = {John Benjamins Publishing Company},
24   doi        = {10.1075/rllt.11.02arc}
25 }
26
27 @incollection{asher2000,
28   Title      = {Events, Facts, Propositions, and Evolutive Anaphora},
29   Author     = {Asher, Nicholas},
30   Booktitle  = {Speaking of Events},
31   Pages      = {123--158},
32   Year       = {2000},
33   Editor     = {Higginbotham, James and Pianesi, Fabio and Varzi, Achille},
34   Publisher  = {Oxford University Press},
35   Address    = {New York, Oxford}
36 }
37
38 @book{asher2011,
39   Title      = {Lexical Meaning in Context: A Web of Words},
40   Author     = {Asher, Nicholas},
41   Year       = {2011},
42   Publisher  = {Cambridge University Press},
43   Address    = {Cambridge}
44 }
45
46 @incollection{auer2006,
47   Title      = {Construction [G]rammar Meets [C]onversation: Einige
48   \"Uberlegungen am [B]ispiel von ''so''-[K]onstruktionen},
49   Author     = {Auer, Peter},
50   Pages      = {291--314},
51   Booktitle  = {Konstruktion in der Interaktion},
```

Reference manager

...or use a reference management software.



Reference managers: Overview

Software	Windows	Mac OS	Linux
Citavi	✓	✗	✗
EndNote	✓	✓	✗
Mendeley	✓	✓	✓
Zotero	✓	✓	✓
JabRef	✓	✓	✓

For normal people: **Zotero**

For academics with gargantuan libraries: **EndNote**

There are others out there (e.g. BibDesk, KBibTeX
→ Wikipedia overview)

	Citavi	EndNote	JabRef	Mendeley	Zotero
Notes	✓	✓	✓	✓	✓
Sharing	✓	✓	✓	✓	✓
Duplicate detection	✓	✓	✓	✓	✓
Browser extension	✓	✓	✓	✓	✓
Export (Word & T _E X)	✓	✓	T _E X	✓	✓
Annotation	✓	✓	✗	✓	✗
Spellcheck	✓	✓	✗	✗	✗
Cost ¹	€73 1 year	€131 1 time	free	free-ish	free-ish
Cloud storage	unlimited	unlimited	✗	2GB, pay yearly for more	300MB, pay yearly for more
Is not evil 😺	✗-ish	✗-ish	✓	✗	✓

¹Some licenses may be provided by your university/institute. Ask, make them pay.

When to use it?



For literature research and review, during reading and writing

Zotero

thesis - Zotero

File Edit View Tools Help

My Library thesis

Duplicate Item... Unfiled Items Trash

No tags to display

ggsignif: Significance Brackets for 'ggplot2' Ahlmann-Elte and Patil

rmarkdown: Dynamic Documents for R Allaire et al.

qqplotr: Quantile-Quantile Plot Extensions for 'g... Almeida et al.

Noun-Phrase Anaphora and Focus: The Informati... Almor

Towards a Unified Treatment of Spanish Copulas Arche et al.

Events, Facts, Propositions, and Evolutive Anaph... Asher

Lexical Meaning in Context: A Web of Words Asher

Construction Grammar Meets Conversational Lingu... Auer

gridExtra: Miscellaneous Functions for "Grid" Gra... Auguie

The R Package `\texttt{textrmnorm}`: The Multivariate N... Azzalini and Genz

languages: Analyzing Linguistic Data: A Practical I... Baayen and Shafaei-Bajestan

magrittr: A Forward-Pipe Operator for R Bache and Wickham

Components of Fluent Reading Baddeley et al.

Coercion and Compositionality Baggio et al.

Computing and Recomputing Discourse Models: ... Baggio et al.

Fitting Linear Mixed-Effects Models Using `lme4` Bates et al.

Matrix: Sparse and Dense Matrix Classes and Met... Bates and Maechler

minqa: Derivative-Free Optimization Algorithms ... Bates et al.

The Development of the Copula in Child English: T... Becker

Copula Omission Is a Grammatical Reflex Becker

Isn't Becker

effects: Estimation of Effect Size Indices and S... Ben-Shachar et al.

tmvnsim: Truncated Multivariate Normal Simulati... Bhattacharjee

Duden Online-Wörterbuch Bibliographisches Institut Gm...

Formal and Lexical Semantics Bierwisch

Semantische und konzeptuelle Repräsentation le... Bierwisch

Lexical Information from a Minimalist Point of Vi... Bierwisch

Effects of Syntactic Prominence on Eye Movemen... Birch and Rayner

Kausale Satzverknüpfungen im Deutschen Blöhorn

A Semantic Typology of Sentence Connectives Blöhorn

Semantische Unterbestimmtheit bei Konnektoren Blöhorn

Some Aspects of Optimality in Natural Language ... Blutner

All Fields & Tags

Info Notes Tags Related

Item Type Conference Paper

Title Towards a Unified Treatment of Spanish Copulas

Author Arche, Maria J

Author Fábregas, Antonio

Author Marin, Rafael

Abstract Date 2017

Proceedings Title Romance Languages and Linguistic Theory 11: Selected Papers from the 44th Linguistics Symposium on Romance Languages

Conference Name Place

Publisher John Benjamins Publishing Company

Volume

Pages 33–52

Series

Language DOI 10.1075/rllt.11.02arc

ISBN

Short Title URL

Accessed Archive

Loc. in Archive

Library Catalog

Call Number Rights

Extra

Date Added 6/27/2022, 2:51:31 PM

Modified 6/27/2022, 2:51:31 PM

Benefits

-  active community of users and developers
-  open source, cross-platform
-  (mostly) no typing necessary
-  automatic formatting
-  works with \LaTeX , Word, LibreOffice
-  Notifies you about retracted papers

Variants

-  Desktop-version
-  Browser plugin
-  Text processor plugin
-  Zotfile plugin to manage your attachments: rename, move, and attach PDFs

AI for BibTeX

You can ask AI to make a BibTeX reference for you, but it's on par with website export → Check for mistakes (e.g. authors, book title)

If given a citation, it may do better (but still check)

Prompt

Create a BibLaTeX entry for these:

Dima C, Ma J, Bücking S, Buscher F, Herdtfelder J, Lukassek J, Pryslopska A, Hinrichs EW, De Kok D, Maienborn C (2017). "A Corpus-Based Model of Semantic Plausibility for German Bracketing Paradoxes". In: Proceedings of Corpora in the Digital Humanities (CDH). Bloomington, Indiana, pp. 64–70.

<https://link.springer.com/article/10.1007/s10936-016-9466-7>

<https://en.wikipedia.org/wiki/Bridgerton>

AI for BibTeX

Result

1. Conference Paper:

bibtex

 Copy code

```
@inproceedings{Dima2017,
    author      = {Dima, C. and Ma, J. and Bücking, S. and Buscher, F. and Herdtfelder,
    title       = {A Corpus-Based Model of Semantic Plausibility for German Bracketing},
    booktitle   = {Proceedings of Corpora in the Digital Humanities (CDH)},
    year        = {2017},
    address     = {Bloomington, Indiana},
    pages       = {64--70}
}
```

Result

2. Journal Article:

```
bibtex Copy code  
  
@article{10.1007/s10936-016-9466-7,  
  author    = {},  
  title     = {},  
  journal   = {},  
  year      = {},  
  volume    = {},  
  number    = {},  
  pages     = {},  
  doi       = {10.1007/s10936-016-9466-7},  
  url       = {https://link.springer.com/article/10.1007/s10936-016-9466-7}  
}
```

Result

3. Wikipedia Article:

bibtex

 Copy code

```
@misc{WikipediaBridgerton,
    author      = {Wikipedia contributors},
    title       = {Bridgerton --- Wikipedia{}, The Free Encyclopedia},
    year        = {2024},
    howpublished = {\url{https://en.wikipedia.org/wiki/Bridgerton}},
    note         = {[Online; accessed 1-July-2024]}
}
```

Summary



Look at them, choose one, and stick with it.

Citavi tutorial:

<https://www.youtube.com/watch?v=9GDYEGWooXE>

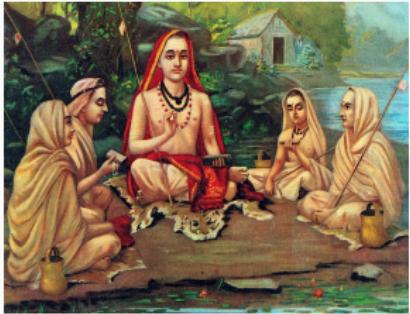
EndNote tutorial:

<https://www.youtube.com/watch?v=IKPZfGPxXtg>

Questions?

Literature research

Starting research



Ask your mentor/advisor

- look at overview articles & their references
- Library website & ask librarians:
<https://www.ub.uni-stuttgart.de/>
- Google scholar: scholar.google.com
- Connected papers: <https://www.connectedpapers.com/>
- Elicit: <https://elicit.org/>

References [edit]

1. ^ a b Legally, German from 3 July 1938: Verordnung über die deutsche Staatsangehörigkeit im Lande Österreich, 3 July 1938@.
2. ^ GLOCK, HANS-JOHANN. "WAS WITTGENSTEIN AN ANALYTIC PHILOSOPHER?" Metaphysics, vol. 35, no. 4, 2004, pp. 419–44, <http://www.jstor.org/stable/24439710?g=.> Accessed 17 Apr. 2022.
3. ^ Rodnyi, Victor. "Wittgenstein's Philosophy of Mathematics"§. In Zalta, Edward N. (ed.), Stanford Encyclopedia of Philosophy (Spring 2018 ed.).
4. ^ Graham, George. "Behaviorism"§. In Zalta, Edward N. (ed.), Stanford Encyclopedia of Philosophy (Spring 2019 ed.).
5. ^ Tang, Han (February 2014). "It is not a something, but not a nothing either! —McDowell on Wittgenstein". Synthese, 191 (3): 557–567. doi:10.1007/s11229-013-0291-3@. JSTOR 24021447@. S2CID 29141239@.
6. ^ Magee, Bryan (14 August 1997). *The Philosophy of Schopenhauer*§. Oxford University Press. doi:10.1093/0198237227.003.0014@. ISBN 978-0-19-823722-8.
7. ^ "Wittgenstein Reads Weininger"§. ndpr.nd.edu. Retrieved 4 February 2021.
138. ^ a b McGuinness 1988, p. 118.
139. ^ Monk 1990, p. 76.
140. ^ McGuinness 2008, p. 39.
141. ^ a b McGuinness 2008, p. 59.
142. ^ Monk 1990, p. 75.
143. ^ McGuinness 2008, p. 63.
144. ^ Flowers & Ground 2015, p. 151.
145. ^ a b Monk 1990, p. 85.
146. ^ Monk 1990, pp. 583–586.
147. ^ Monk 1990, p. 369.
148. ^ Monk 1990, pp. 238–240, 258, 280–281, 294, 318–319.
149. ^ Goldstein 1999, p. 179.
150. ^ von Wright 1990, p. 221.
151. ^ von Wright 1990, p. 230.
152. ^ Monk 1990, p. 58ff.
153. ^ a b von Wright 1990, p. 88.
154. ^ a b Kanterian 2007, p. 40.
155. ^ Monk 1990, p. 71.
156. ^

Check Wiki references

AI for literature research and documentation

You can ask AI for a list of references, but there are better methods,
AI tends to hallucinate articles/books/etc.

AI is poor at summarizing articles → may be enough to get a feel for relevance of an article, but AI often fails to get the point.

Don't use it for summaries/overview → AI fails to get the point, easy to unknowingly plagiarize, your lecturer can tell.

For technical documentation, try PerplexityAI

AI is decent at documenting code → good draft, but needs editing

Get that paper: Dos and donts

- Angel 🕊 University website
- Angel 🕊 Library
- Angel 🕊 University access
- Angel 🕊 Write to authors or check their websites
- Skull 🚨 Library Genesis: <https://libgen.is/>
- Skull 🚨 Sci-hub: <https://www.sci-hub.wf/>
- Skull 🚨 Anna's Archive: <https://annas-archive.org/>
- Skull 🚨 Z-library: <https://z-lib.io/>
- Skull 🚨 Reddit: <https://www.reddit.com/r/scholar>

~~LATEX~~

LATEX-FREE ONLY

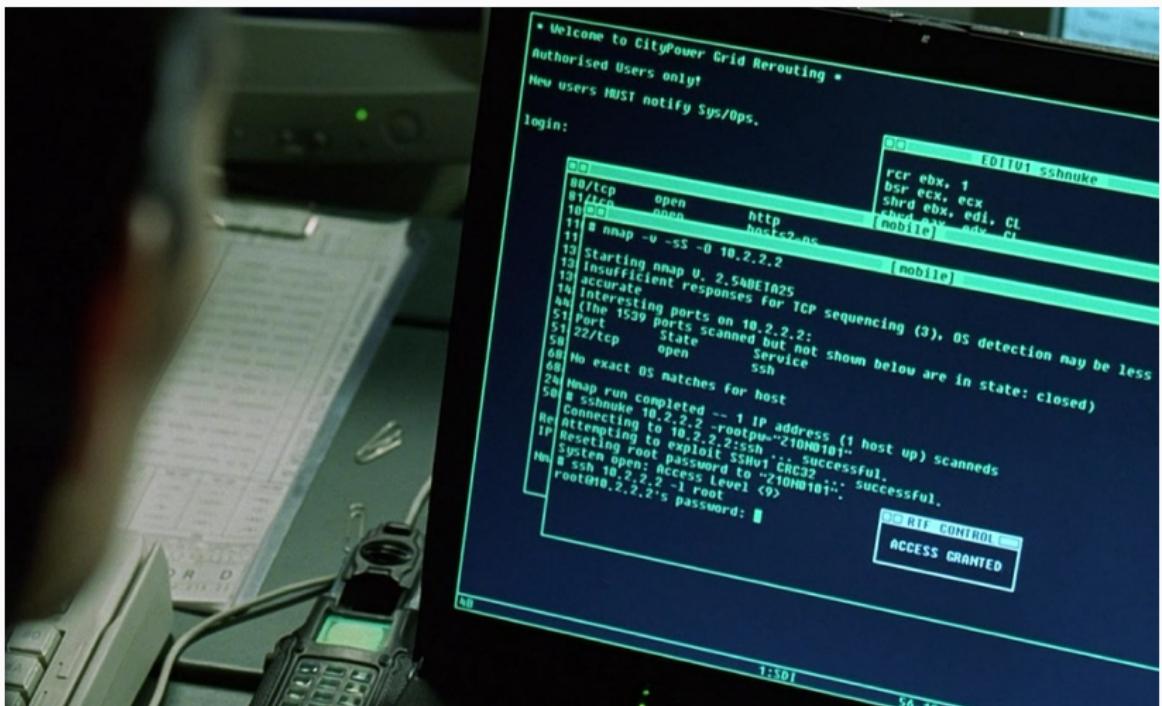
(DO NOT STORE ITEMS CONTAINING LATEX HERE)



Questions?

Mastering the command line

Terminal, console, shell, or command line?



The **command line interface (CLI)** = GUI but in text form (how everything worked way back when).

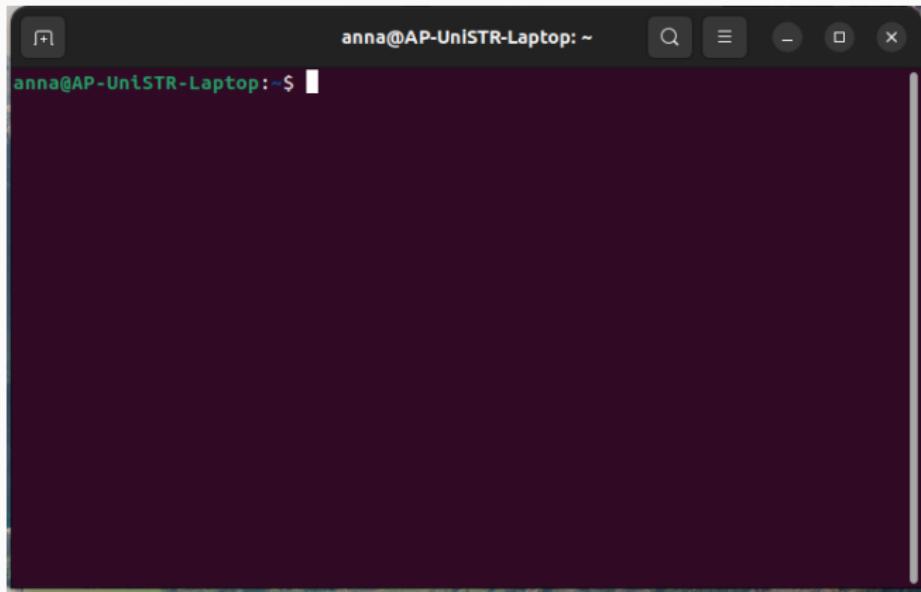
Terminal, console, shell, or command line?



Lat. *terminus* = end, boundary (bus, airport, battery, patient, ...)

A **terminal** is a text input and output environment (e.g. command prompt, gnome-terminal). It used to be physical device, located at the *end of the cable* to a computer.

Terminal, console, shell, or command line?



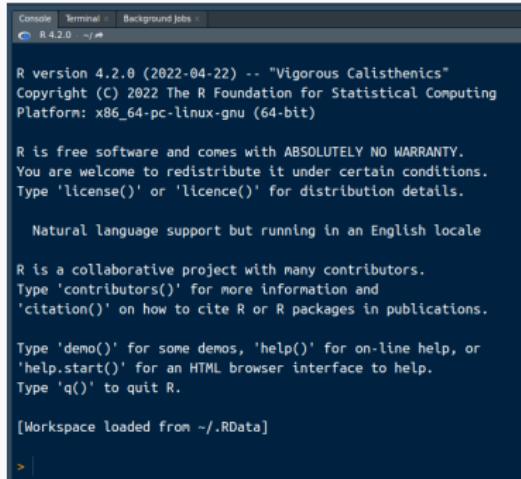
A **terminal window** (or terminal emulator) is a text-only window that emulates a terminal/console in a GUI. It runs a shell.

Terminal, console, shell, or command line?



A **console** is (generally) a type of physical terminal. It is sometimes synonymous with a terminal (e.g. for direct communication at a low level with the OS). Terminal vs. console = rap vs. hip hop

Console vs Terminal in RStudio



R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

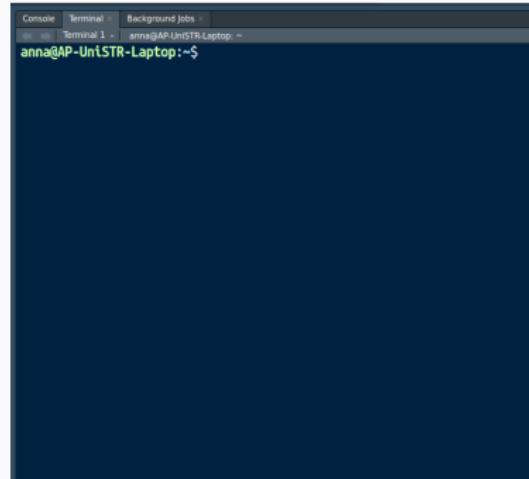
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> |

Console → R code

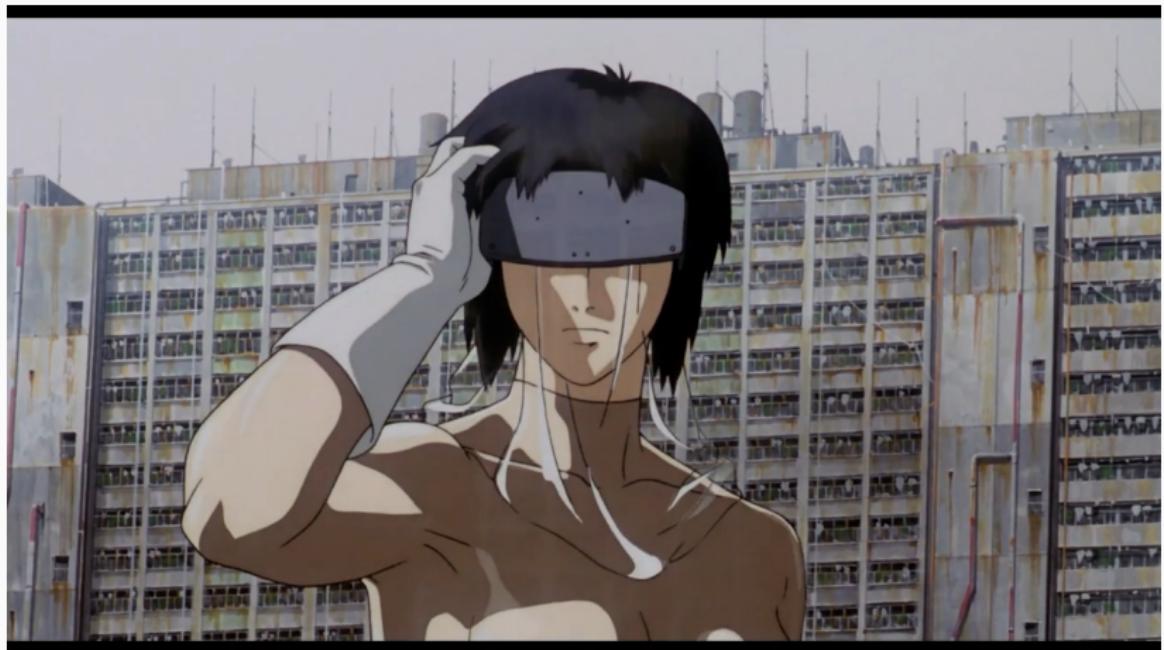


Console Terminal Background Jobs

Terminal 1 - anna@AP-UniSTR-Laptop:~\$

Terminal → access to system shell

Terminal, console, shell, or command line?



A **shell** is a command-line interpreter (e.g. Bash, Power Shell, CMD). It's a program that runs in the terminal. It reads and processes commands, and outputs the results.

Terminal, console, shell, or command line?

```
anna@AP-UniSTR-Laptop:~$
```

A **command line** is an area to the right of the command prompt where you type commands.

Terminal



Console

The **command line** is what you input into your shell. The **shell** is what processes and executes the command. The **terminal** is the window in which the shell runs, which is often the same as the **console**.  

Why is my evil lecturer making me learn it?

Advantages of a CLI

- greater control of an OS or application
- more informative
- helps with troubleshooting (e.g. network connection issues)
- faster than clicking
- required for many devices (e.g. eye-trackers, servers)
- handle big data
- easy to automate processes
- looks cool

Disadvantages of a CLI

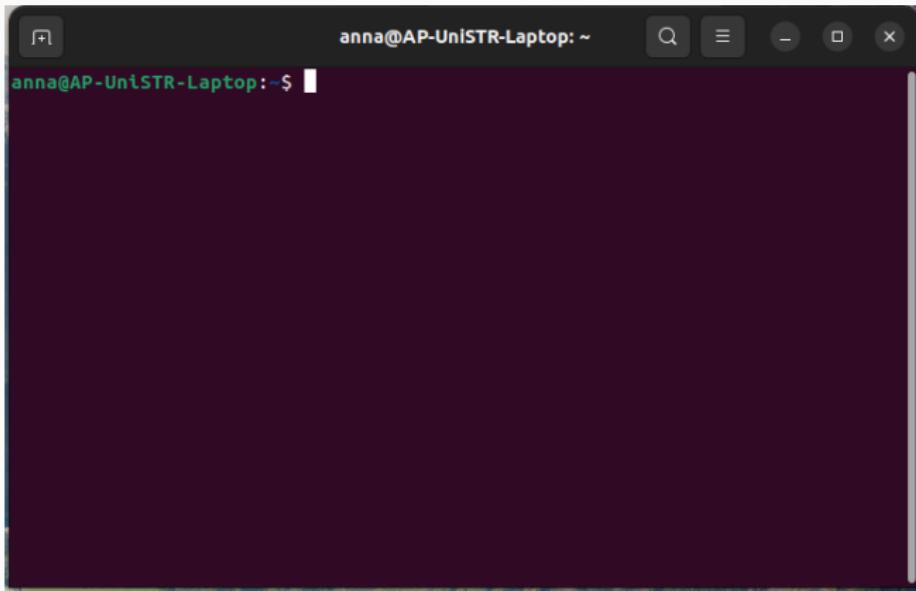
- GUI is more user-friendly
- steep(er) learning curve
- need to memorize different commands which can differ between shells
- no undo

How to access command line

Mac OS and Linux are both Unix, Windows is *special*.

Windows users: Install  Git for Windows & use Git BASH

Mac OS & Linux users: open Terminal (install Git later)



Getting around

<code>pwd</code>	print working directory
<code>ls</code>	list all files or directories in the current directory
<code>cd</code>	change directory to somewhere else: <code>cd /home/anna/</code>
<code>cd ..</code>	change directory to one up: <code>cd ..</code> → <code>/home/</code>
<code>mkdir</code>	make directory : <code>mkdir Hazard</code>
<code>cp</code>	copy file(s): <code>cp virus.txt Hazard/</code>
<code>mv</code>	move file(s): <code>mv virus.txt Hazard/</code>
<code>diff</code> and <code>sdiff</code>	list differences (side-by-side) between files: <code>diff virus.txt virus_evil.txt</code>
<code>rm</code>	remove file(s) forever: <code>rm virus.txt</code>



Type a simple command into the console



Press up key dozens of times until you find it.

Command line L^AT_EX

1. Compile the document with X_LT_EX
`xelatex report.tex`
2. Run biber for bibliography processing
`biber report`
3. Compile again to integrate bibliography
`xelatex report.tex`
4. Compile again to resolve references, make TOC etc.
`xelatex report.tex`

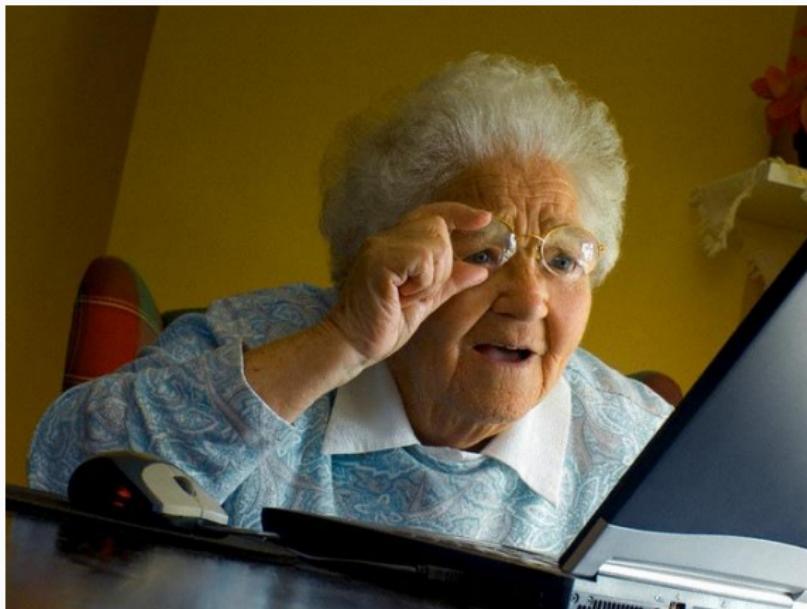
Save only the code as a shell script `compile.sh` and execute `bash compile.sh`

Command line R and Quarto

```
Rscript analysis.R
```

```
quarto render presentation.qmd
```

Looking for stuff: grep (global regular expression print)



Search for a sequence of characters in a specified file or files

Looking for stuff: grep

```
grep OPTIONS PATTERN WHERE
grep Winter virus_benign.txt
grep "ha ha" virus.txt virus_evil.txt
grep virus *
grep -w Winter virus_benign.txt
grep -v "ha ha" virus.txt virus_evil.txt
grep -i platz virus_benign.txt
grep -c am *
grep -c -i -w Winter *
```

option	alternative	description
-c	--count	Count matching lines
-i	--ignore-case	Case-insensitive search
-v	--invert-match	Invert match
-w	--word-regexp	Match the whole word
-e	--regexp=PATTERN	Match regular expression
-h	--help	Get help
-l	--files-with-matches	List the file names
-o	--only-matching	Show only matching part
-r	--recursive	Search directories recursively

```
grep -c -i -w Winter *
```

```
grep -c -i -w Winter * | wc
```

count words

Other tools

wc	word count (nr of lines, words, and bytes)
ping	ring the door bell and see how long until it opens <i>Am I online? Is the website up? Is the response slow/fast?</i>
	ping google.com
date	date and time now
clear	clear the terminal
ps aux	show all processes for all users
kill	stop a process with a PID
exit	close the terminal
shutdown	shutdown the computer
whois	who is the owner of this website? whois pryslopska.com

Questions?

Summary

- ✓ reference managers
- ✓ literature research
- ✓ command line, terminal, console, and shell
- ✓ Render/run \LaTeX , R, and Quarto via CLI
- ✓ GREP
- ✓ Most common command line commands
- Text editors, git

Homework assignment

Homework assignment due July 5th at 15:30

- ➊ Complete assignment 11 (\rightarrow ILIAS)