

Digital Research Toolkit for Linguists

Week 2: Looking at data

Anna Pryslopska

April 15, 2024

Psycholinguistics and Cognitive Modeling Lab

Questions?

Table of contents

1. Where are we this week?

2. Experiments

3. R

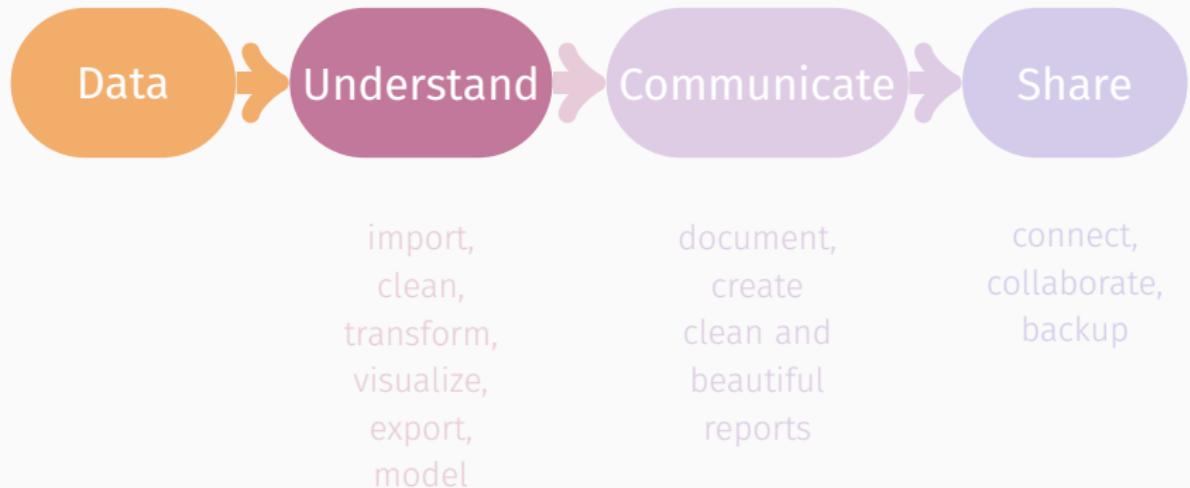
4. Working with RStudio

5. R packages

6. Wrap-up

Where are we this week?

Workflow



Experiments

Experiment 1: Continuation

Q: Can a man marry his widow's sister?

No 🙅‍♂️💀💡

Q: According to the Bible, how many animals of each kind did Moses take on the ark?

can't answer

Q: How many numbers are in the English alphabet?

can't answer



Moses Illusions or semantic illusions occur when readers fail to recognize inconsistency in a text even if they were warned and know the correct word (Erickson and Mattson 1981).

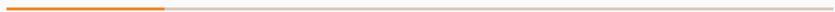
Experiment 2: Self-paced reading with acceptability judgments

- The cook baked Lucy a cake. plausible
 - The cook baked Lucy for a cake. implausible
 - The cook baked a cake for Lucy. plausible
 - **The cook baked a cake Lucy.** implausible

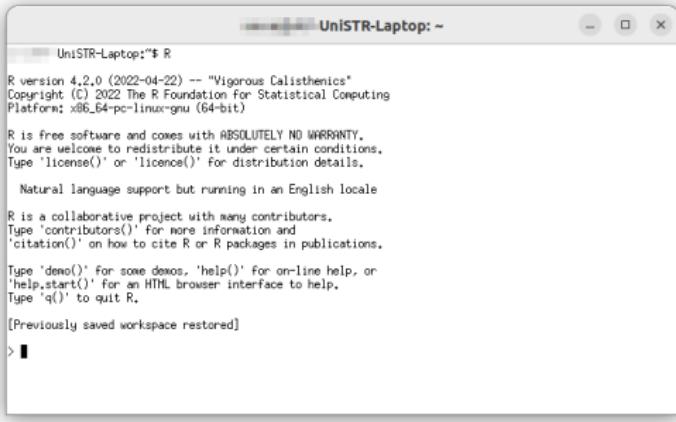
Noisy channel: Humans understand language even in noisy environments and can recover meaning from imperfect utterances. Semantic cues can pull a comprehender towards plausible meanings, but too much noise makes comprehenders switch to the literal interpretation.

We only did two conditions for bigger effects. Materials were kindly shared by Ted Gibson (Gibson et al. 2013).

R



What R we doing



```
UniSTR-Laptop:$ R
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.

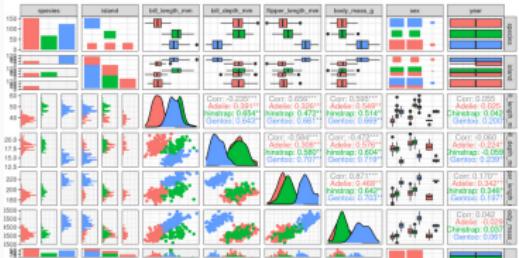
[Previously saved workspace restored]
> █
```

- language & environment for statistical computing & graphics
- used by scientists, data miners & statisticians for data analysis & developing statistical software
- developed by R Core Team & the R Foundation for Statistical Computing <https://www.r-project.org/>
- base + packages

R: Why bother?



most popular in data science



powerful visualizations



used in almost every industry



Free as in Freedom

free, open source

R: Why bother?



(relatively) easy to learn



gateway to programming and lucrative careers

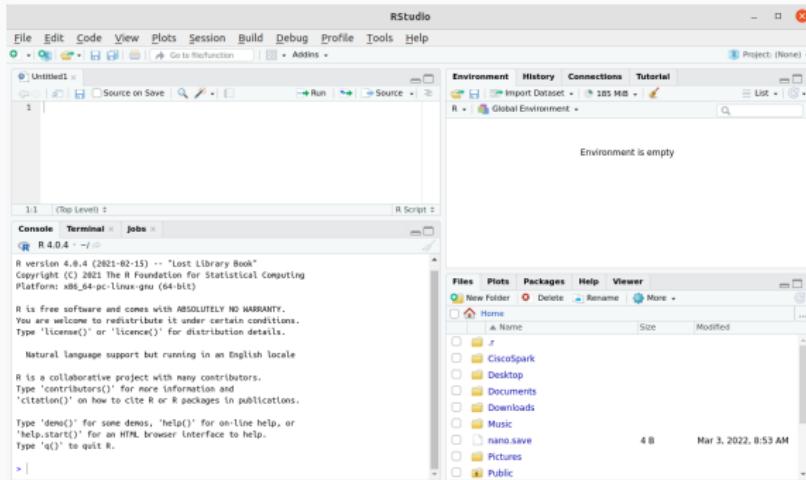
Bonus: you can develop cool web apps

<https://ceefluz.shinyapps.io/radar/>

<https://gallery.shinyapps.io/087-crandash/>

Working with RStudio

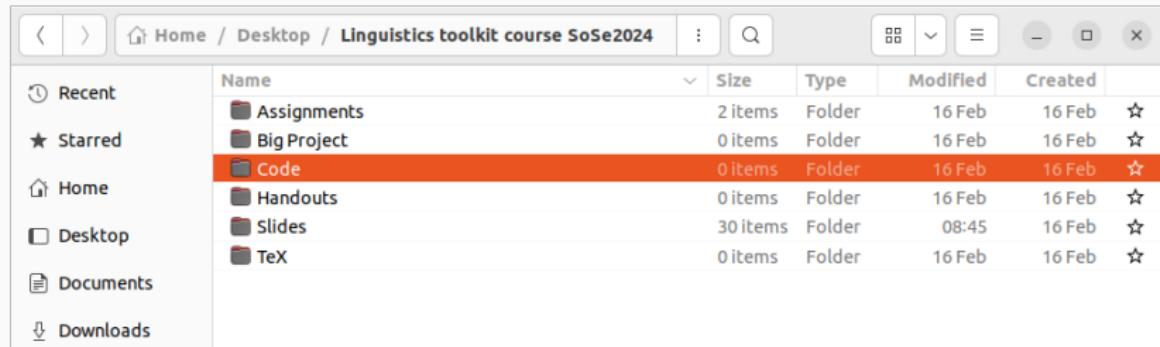
RStudio



- Integrated Development Environment (=IDE) for R (& Python)
- free, open source, “friendlier” than pure R
- developed by **Posit**, who develop, maintain, and promote many awesome packages (e.g. tidyverse, RMarkdown, knitr) and Quarto
- source, console, environment, plots, packages, help

Make a home for your project

Create a new folder and set it as the working directory

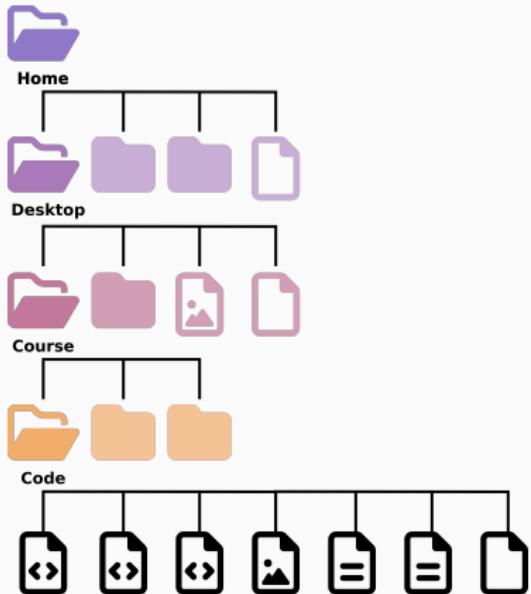


- where R will look for files
- where R will save visible and hidden files
- where R will automatically load files from

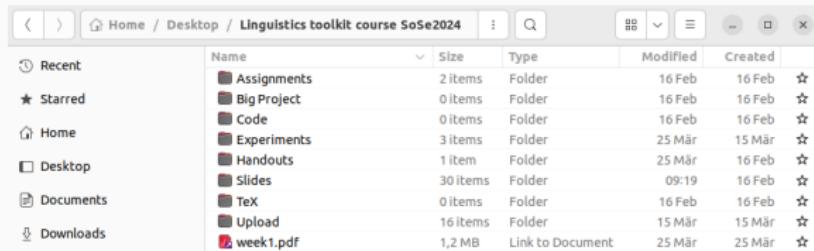
File structures and directories

A **directory** or **folder** is a container for storing files or other folders.

File structure or file hierarchy or **folder organization** is the way these containers are organized.



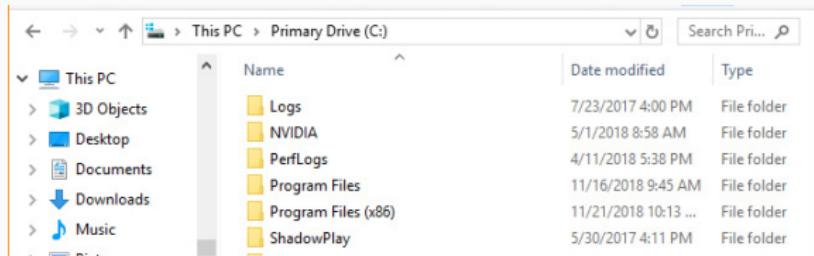
We're Different, We're the Same



A screenshot of the Unity desktop environment's file manager. On the left is a sidebar with links to 'Recent', 'Starred', 'Home', 'Desktop', 'Documents', and 'Downloads'. The main area shows a list of files and folders under 'Linguistics toolkit course SoSe2024'. The columns are 'Name', 'Size', 'Type', 'Modified', and 'Created'. A star icon is present in the 'Created' column.

	Name	Size	Type	Modified	Created
① Recent	Assignments	2 items	Folder	16 Feb	16 Feb
★ Starred	Big Project	0 items	Folder	16 Feb	16 Feb
⌚ Home	Code	0 items	Folder	16 Feb	16 Feb
📁 Desktop	Experiments	3 items	Folder	25 Mär	15 Mär
📄 Documents	Handouts	1 item	Folder	25 Mär	16 Feb
⬇️ Downloads	Slides	30 items	Folder	09:19	16 Feb
	TeX	0 items	Folder	16 Feb	16 Feb
	Upload	16 items	Folder	15 Mär	15 Mär
	week1.pdf	1,2 MB	Link to Document	25 Mär	25 Mär

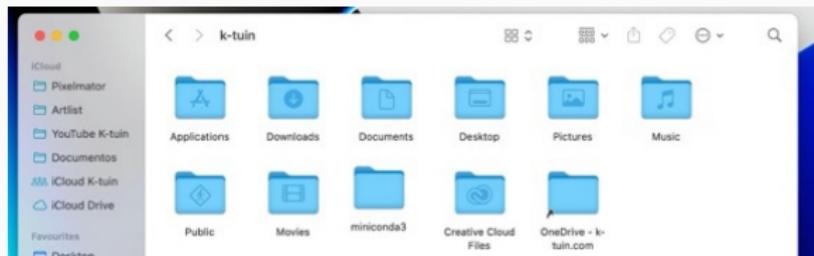
Ubuntu



A screenshot of the Windows File Explorer. The left sidebar shows 'This PC' with options like '3D Objects', 'Desktop', 'Documents', 'Downloads', and 'Music'. The main area lists files and folders on 'Primary Drive (C:)'. The columns are 'Name', 'Date modified', and 'Type'.

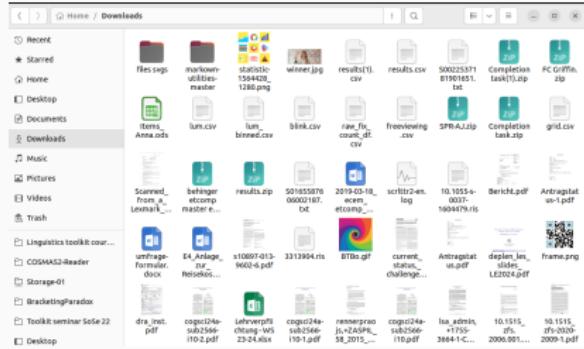
Name	Date modified	Type
Logs	7/23/2017 4:00 PM	File folder
NVIDIA	5/1/2018 8:58 AM	File folder
PerfLogs	4/11/2018 5:38 PM	File folder
Program Files	11/16/2018 9:45 AM	File folder
Program Files (x86)	11/21/2018 10:13 ...	File folder
ShadowPlay	5/30/2017 4:11 PM	File folder

Windows



Mac

Why it matters and best practices



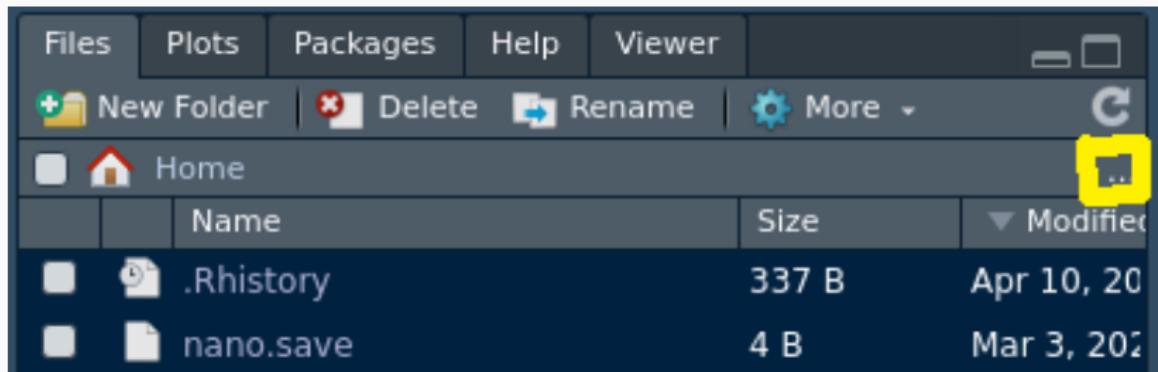
Save space and time, establish expectations and be consistent, avoid confusion and repetition.

Do this for yourself, your future self, and your collaborators.

→ We will return to this later.

→ This is important and will be on the test.

Make a home for your project



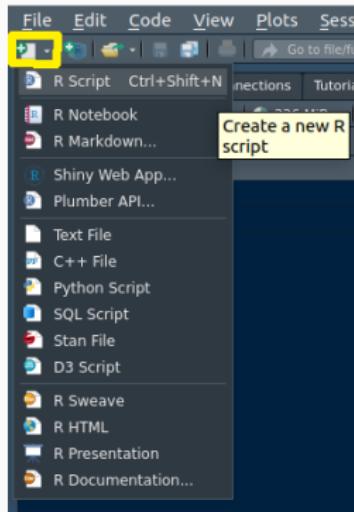
```
setwd("PATH")
```

set working directory

```
getwd()
```

check working directory

Don't You Forget About Me



	Windows	Mac
use auto-completion	TAB	TAB
run selected code	CTRL+ENTER	CMD+RETURN
leave comments #	CTRL+SHIFT+C	CMD+SHIFT+C
make sections	CTRL+SHIFT+R	CMD+SHIFT+R
work with scripts	CTRL+SHIFT+N	CMD+SHIFT+N

Hello world

```
print("Hello World")
print("Twas brillig,
and the slithy toves
did gyre and gimble
in the wabe: all mimsy
were the borogoves,
and the mome raths
outgrabe.")
```



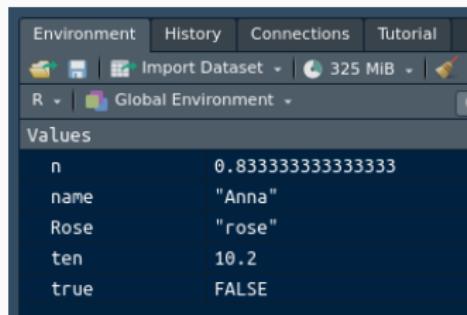
Assignment

Saves information for later, keeps it in the environment, but overwrites existing values.

```
x <- value          ten <- 10.2
value -> x         "rose" -> Rose

x = value           name = "Anna"

x <<- value        true <<- FALSE
value ->> x       13/12 ->> n
```



The screenshot shows the RStudio interface with the Global Environment tab selected. The Values section displays the following assignments:

	Values
n	0.8333333333333333
name	"Anna"
Rose	"rose"
ten	10.2
true	FALSE

```
jabberwocky <- print("Twas brillig ...")
```

❓ What are some good names?

R packages

What's in the package

Packages are collections of functions and/or datasets with a common theme (e.g. statistics, spatial analysis, plotting).

Anyone can develop packages (you too!) and share them with others.

Most packages are available through the **Comprehensive R Archive Network (CRAN)** and on **GitHub**.

The screenshot shows the homepage of the R Package Documentation. At the top, there is a search bar with the placeholder "Search for anything R related". Below the search bar, there is a search input field containing the word "gaussian" and a green "Search" button. To the right of the search bar, there are four large numbers representing the count of packages in different sources: 25160 CRAN PACKAGES, 2130 BIOCONDUCTOR PACKAGES, 2225 R-FORGE PACKAGES, and 85708 GITHUB PACKAGES. At the bottom, there are three blue buttons with white text: "Find an R package", "Run R code online", and "Browse R language docs".

R Package Documentation
A comprehensive index of R packages and documentation from CRAN, Bioconductor, GitHub and R-Forge.

Search for anything R related
Find an R package by name, find package documentation, find R documentation, find R functions, search R source code...

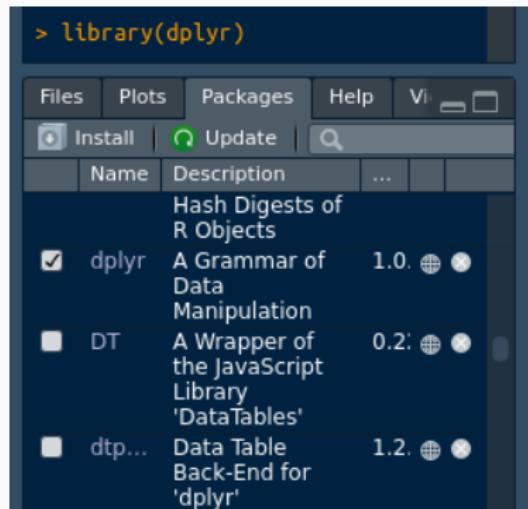
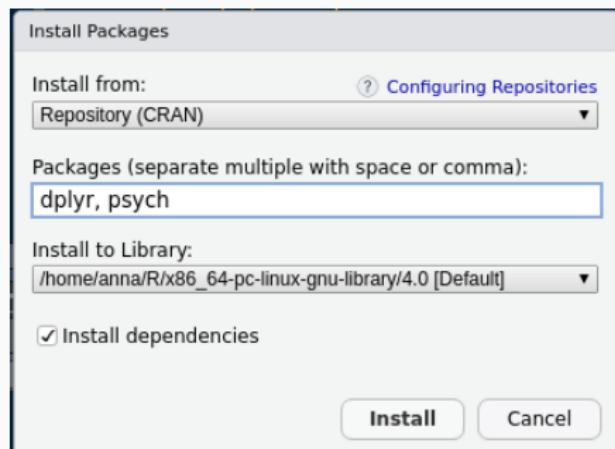
gaussian Search

25160 CRAN PACKAGES 2130 BIOCONDUCTOR PACKAGES 2225 R-FORGE PACKAGES 85708 GITHUB PACKAGES

Find an R package Run R code online
Over 19,000 packages are preinstalled! Browse R language docs

Installing, loading, checking

`install.packages("NAME", "NAME2")` install from repository
`library(NAME)` load into workspace (=activate)
`sessionInfo()` collect information about current R session



Questions?

Wrap-up

Summary

- ✓ R and RStudio IDE
 - ✓ packages (installing and loading)
 - ✓ working directory
 - ✓ directories and file hierarchy
 - ✓ scripts
 - ✓ assignment
- Next week: Looking at (your) data and R programming basics
- 💡 Want to read more? QCBS R Workshop Series

Homework assignment due April 19th 15:30

- ➊ Read chapters 2, 6 and 7 of *R for Data Science* (Wickham et al. 2023)
- ➋ Complete assignment 1 (\rightarrow ILIAS)

References

-  Erickson, Thomas D and Mark E Mattson (1981). “From words to meaning: A semantic illusion”. In: *Journal of Verbal Learning and Verbal Behavior* 20.5, pp. 540–551. DOI:
[10.1016/s0022-5371\(81\)90165-1](https://doi.org/10.1016/s0022-5371(81)90165-1).
-  Gibson, Edward, Leon Bergen, and Steven T Piantadosi (2013). “Rational integration of noisy evidence and prior semantic expectations in sentence interpretation”. In: *Proceedings of the National Academy of Sciences* 110.20, pp. 8051–8056. DOI:
[10.1073/pnas.1216438110](https://doi.org/10.1073/pnas.1216438110).
-  Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Grolemund (2023). *R for data science: import, tidy, transform, visualize, and model data*. 2nd ed. O'Reilly Media, Inc. URL:
<https://r4ds.hadley.nz/>.