

---

# The fundamentals of data visualization in R

## Our topics and goals

Guilherme D. Garcia

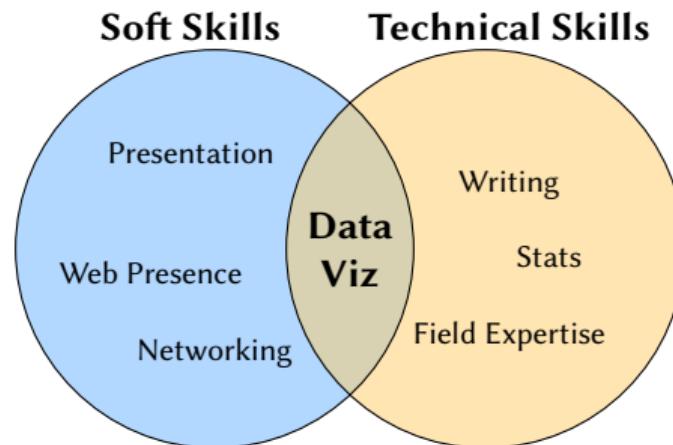
[gdgarcia.ca](http://gdgarcia.ca) ↗

University of Alberta, Edmonton, May 2025

# Welcome!

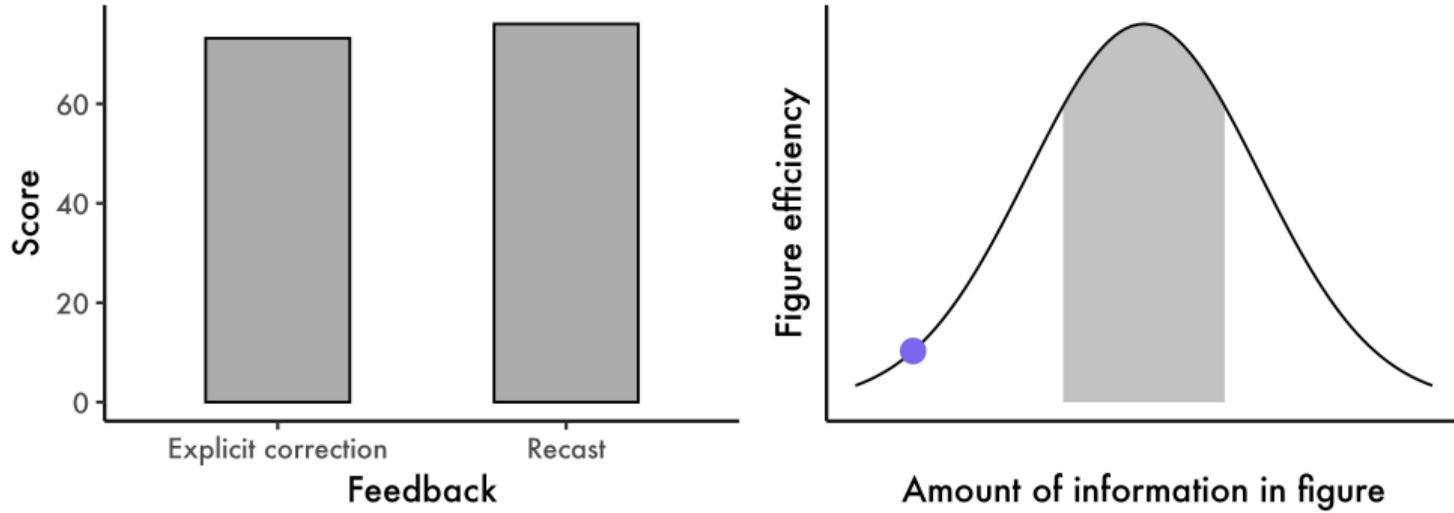
## Why visualize data?

- **Communicating** results is a key element in research (and in the job market)
- Good visuals are a crucial but often underrated skill in academia
- ☞ In some sense, where soft and technical skills meet:



# What's a “good figure”?

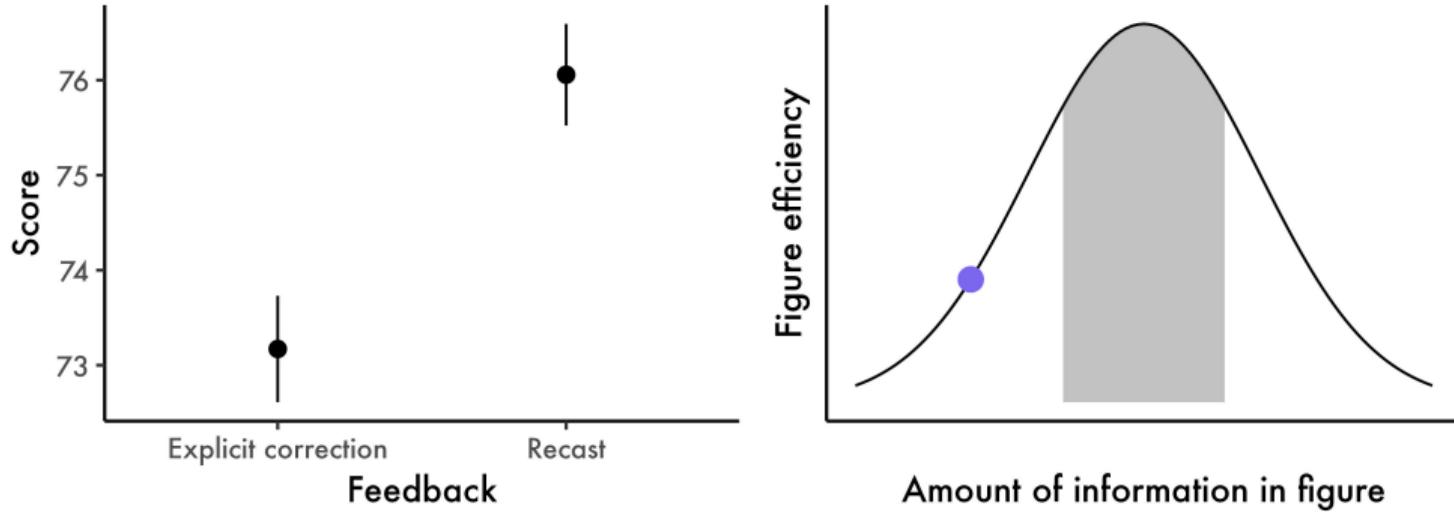
What geomets are being used? What issues do you notice?



**Figure 1:** Figure vs. amount of information

# What's a “good figure”?

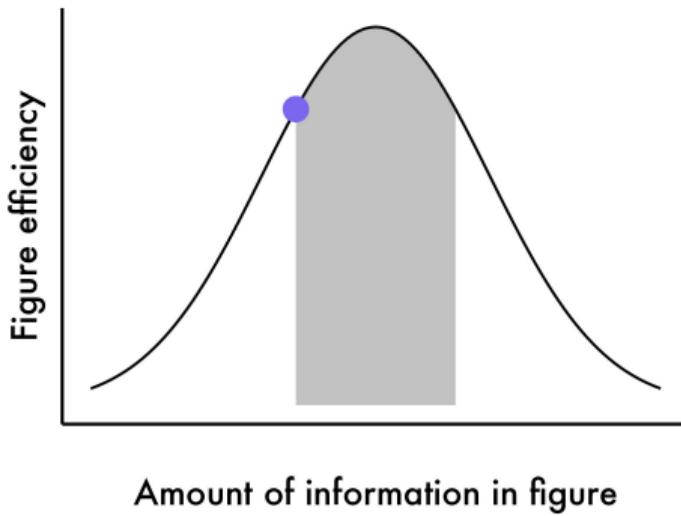
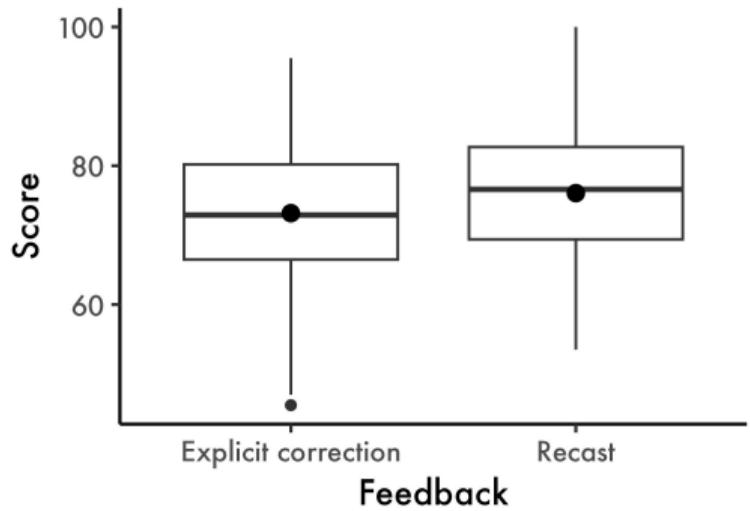
What geomets are being used? What issues do you notice?



**Figure 2:** Figure vs. amount of information

# What's a “good figure”?

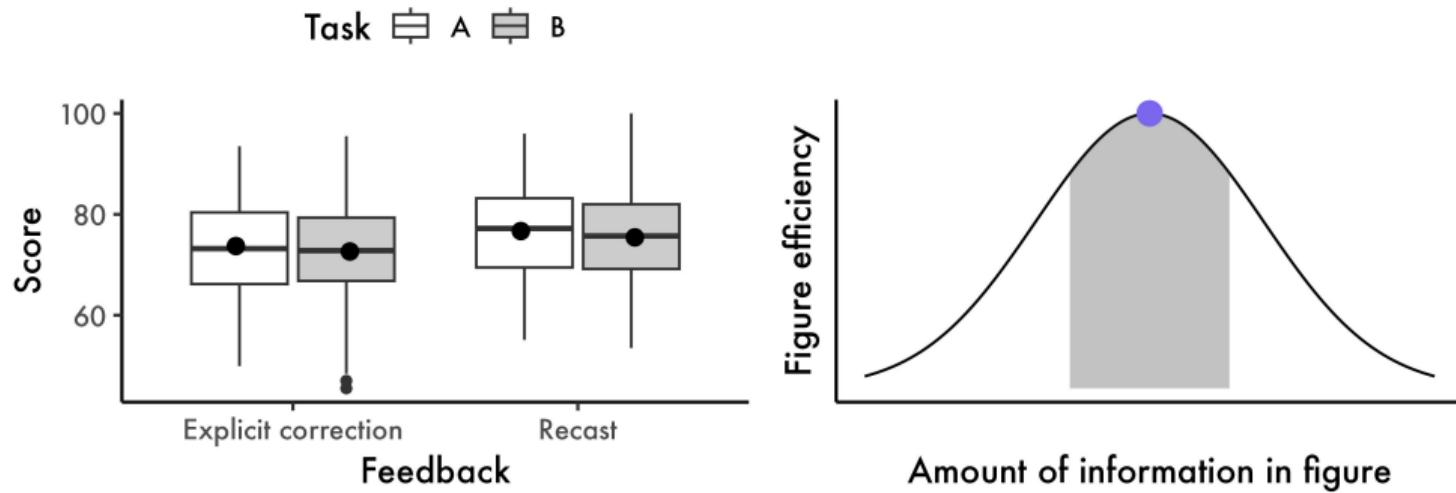
What geom are being used? What issues do you notice?



**Figure 3:** Figure vs. amount of information

# What's a “good figure”?

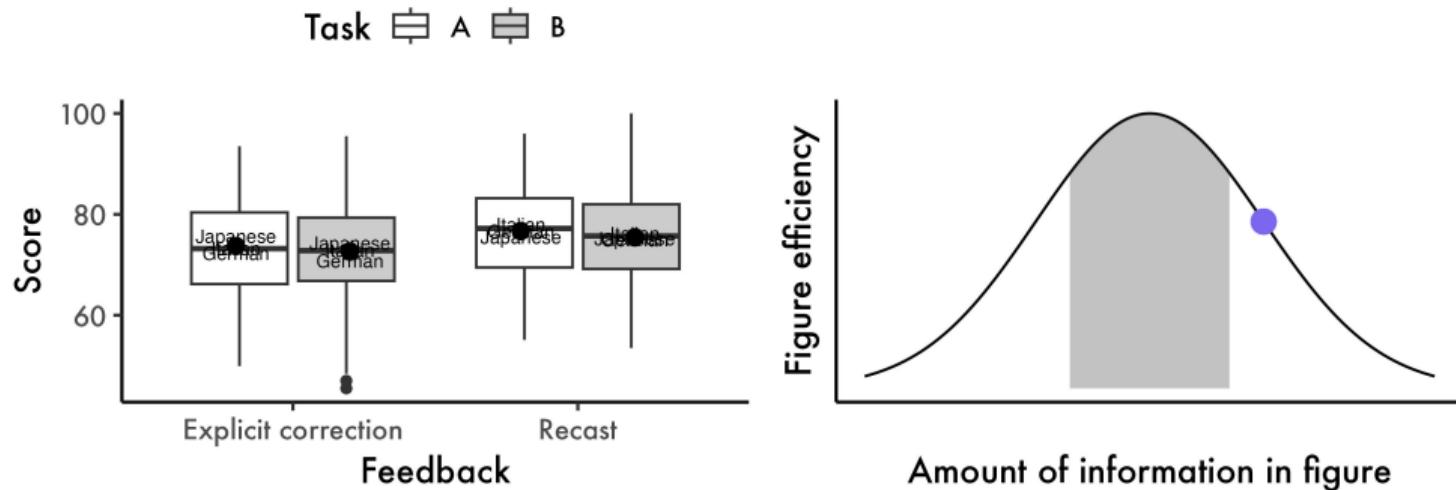
What geomets are being used? What issues do you notice?



**Figure 4:** Figure vs. amount of information

# What's a “good figure”?

What geoms are being used? What issues do you notice?

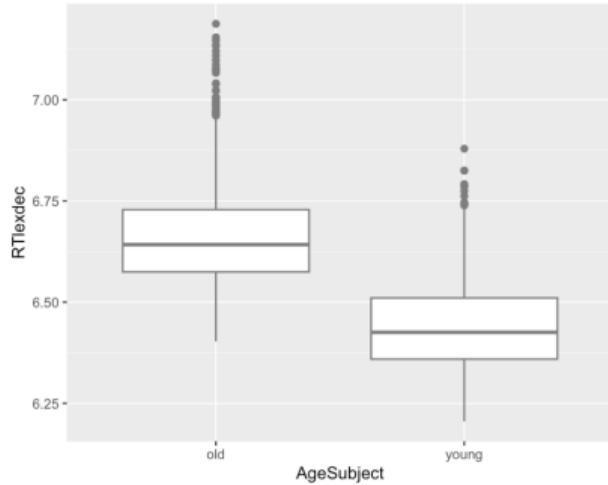


**Figure 5:** Figure vs. amount of information

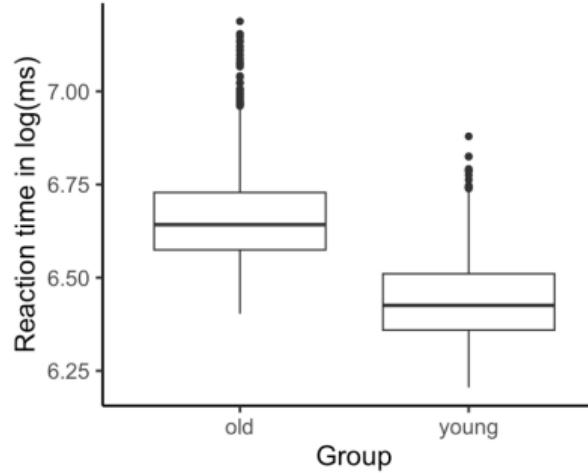
# What's a “good figure”?

Form also matters

## Examples



(a) Font is **too small**; contrast is **too low**



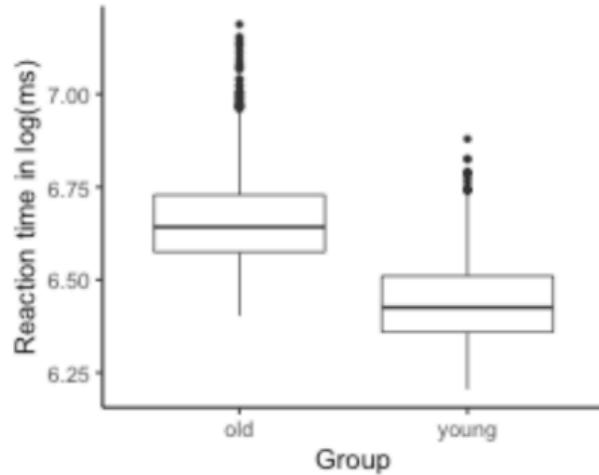
(b) Good font size; good contrast; better labels

**Figure 6:** From the guidelines at SLR (1)

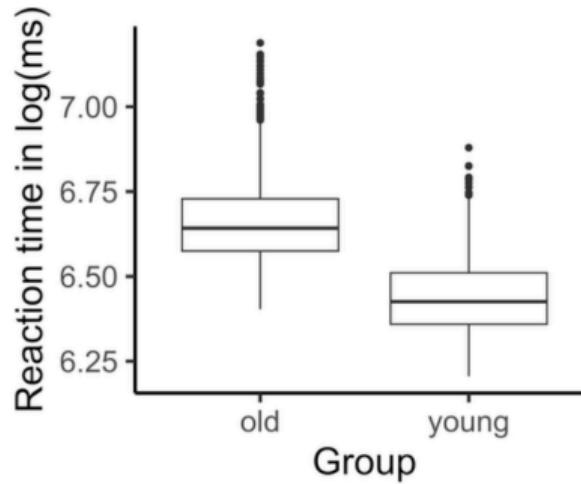
# What's a “good figure”?

Form also matters

Examples



(c) Font size and contrast are appropriate, but resolution is **too low** (25 dpi)



(d) Resolution is good (400 dpi), but font is too large

**Figure 7:** From the guidelines at SLR (2)

# Our itinerary

Topics we will cover in three days

## 1. Basic principles behind data visualization

Data cleaning/preparation; `ggplot2` and its core structure; essential `geoms`

**Practice**

## 2. Data transformation for visualization

Scales, facets, and grids; aesthetics; `dplyr` and conditionals

**Practice**

## 3. Plotting individual variation

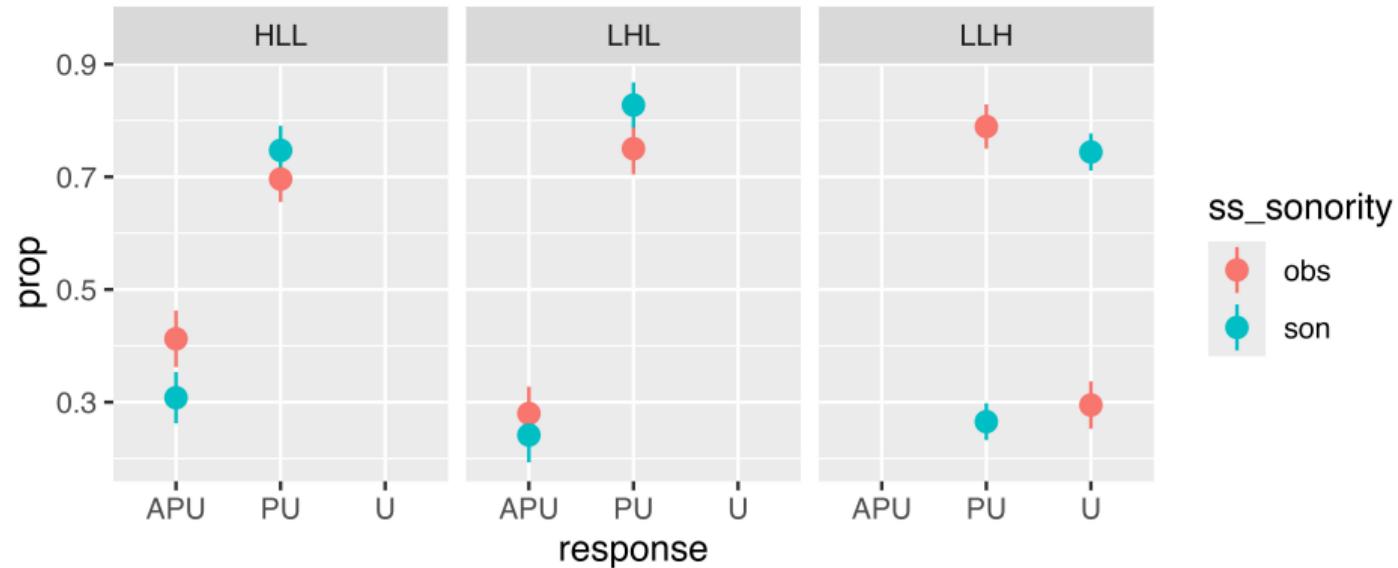
Visualizing model estimates; a brief intro to Quarto

Interactive plots and extra packages

**Practice**

# Example 1

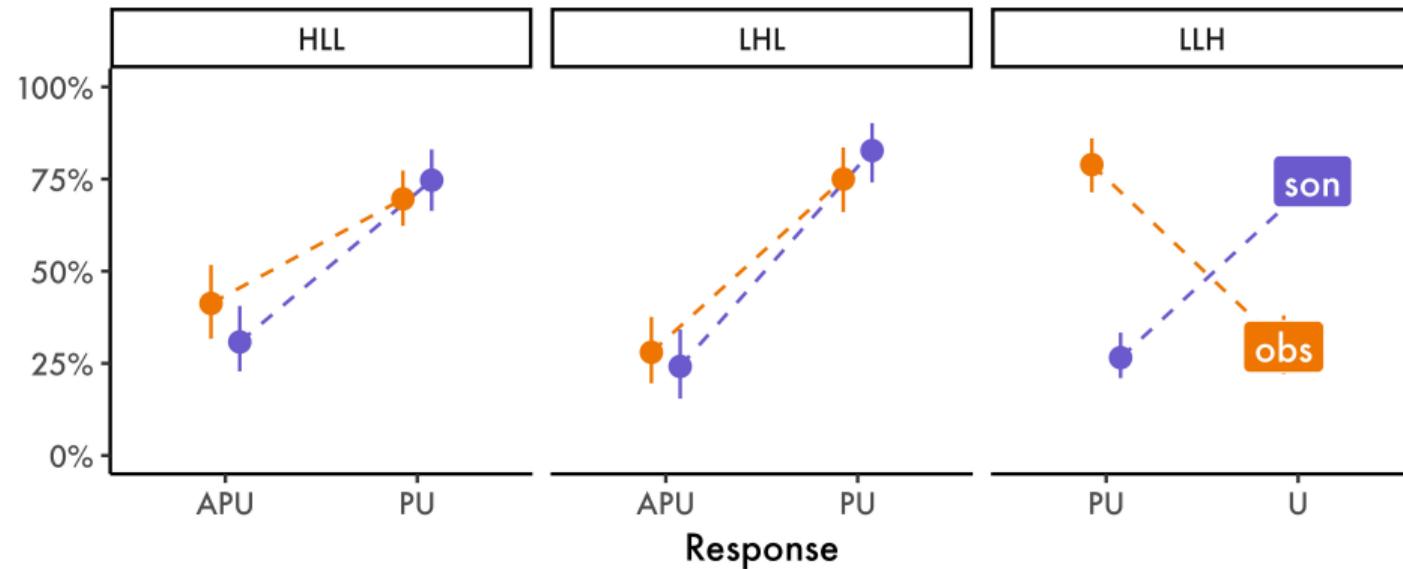
## A basic figure



**Figure 8:** Response preferences as a function of coda sonority in heavy (H) syllables

# Example 1

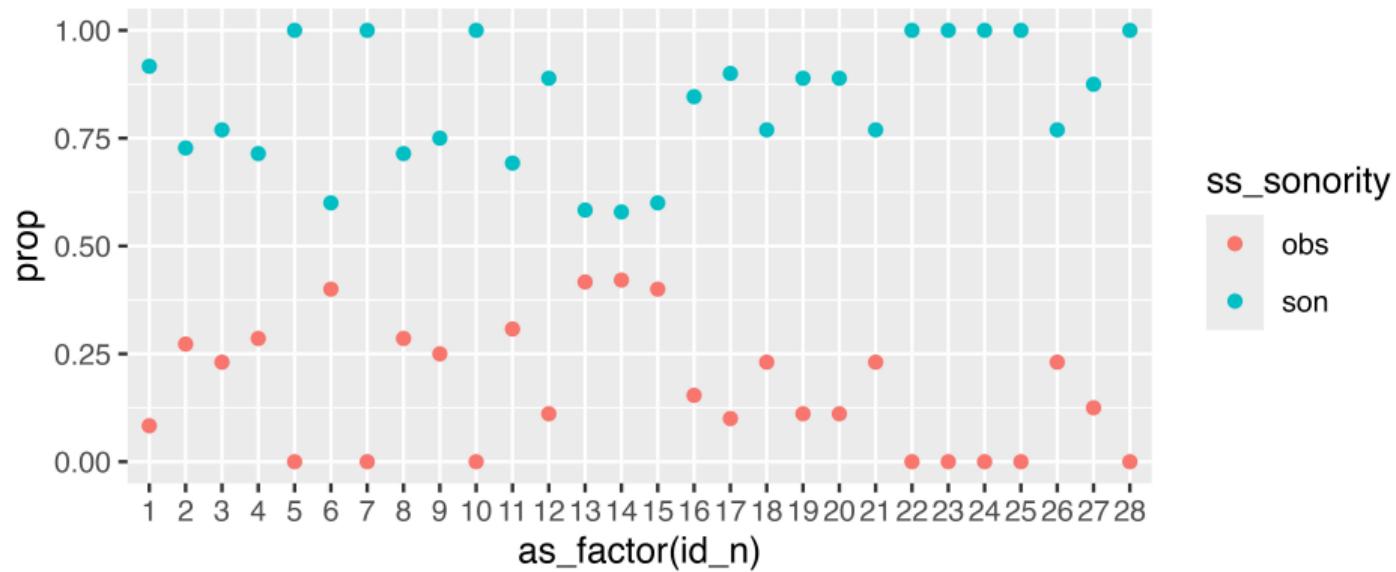
A better version



**Figure 9:** Legend is not inside the plot and labels appear only once. Better?

## Example 2

A basic figure



**Figure 10:** Individual variation. This works, but isn't too clean

## Example 2

A better version

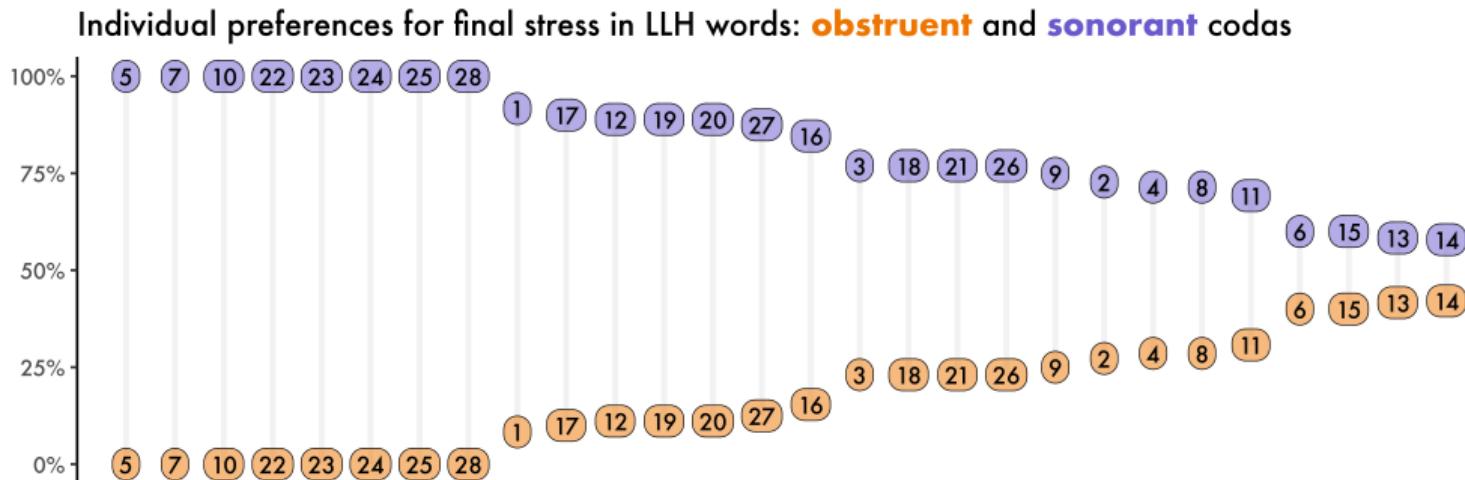


Figure 11: This is cleaner and easier to understand

# Materials and logistics

☞ **IDE:** many options (**Positron**, RStudio, Sublime, VSCode, nvim, etc.)

- If you prefer, you can also use [posit.cloud](#) ↗
- We will focus on **scripts**, *not* slides
- Our main package will be **tidyverse**, but we will certainly use others

(Wickham et al. 2019)

**What you need:** an IDE + **tidyverse** installed

## Positron vs. RStudio

- Positron is the new IDE from Posit, and will likely replace RStudio
  - So you might as well switch to Positron these days
- ☞ It's a personal choice: for our course, it won't matter which IDE you use

# Survey

[forms.cloud.microsoft/r/m6mCwa4d5m](https://forms.cloud.microsoft/r/m6mCwa4d5m)



## References I

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.