

# Essential Research Toolkit for the Humanities

## Week 6: Data visualization

---

Anna Pryslopska

May 13, 2024

Psycholinguistics and Cognitive Modeling Lab

Register for the exam starting tomorrow!  
Things are complicated.

# Homework

## General comments

`dplyr` and `readr` are loaded together with `tidyverse`. You don't need to load `psych` if you're not using it.

You can save/assign to variable and then call the variable to shorten code. What you don't assign, disappears.

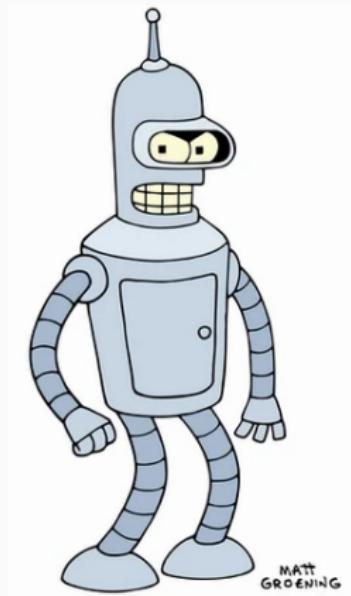
`arrange()` by default arranges by first column (or more).

Test your code on a clean slate/empty environment (sometimes stuff isn't actually working):

```
df_perc <- summarise(moses.preproc, Count=n()) |>  
  mutate(Perc=Count/sum(df_perc$Count)*100)
```

If nothing in your code works, read the error messages and look for help (Correct working directory? Loaded required packages? Documentation? Slides? Forum? Email?).

# AI use



I can tell when you use AI for help. If you're smart about it, I'll turn a blind eye to it. If you're dumb about it, I will fail the assignment.

If you use the `magrittr` pipe I expect you to explain why.

## Moses illusion

The conditions were not “correct, incorrect, don’t know”.

I prefer %, but counts are ✓

I prefer % adding to 100% per condition, but 100% overall is ✓

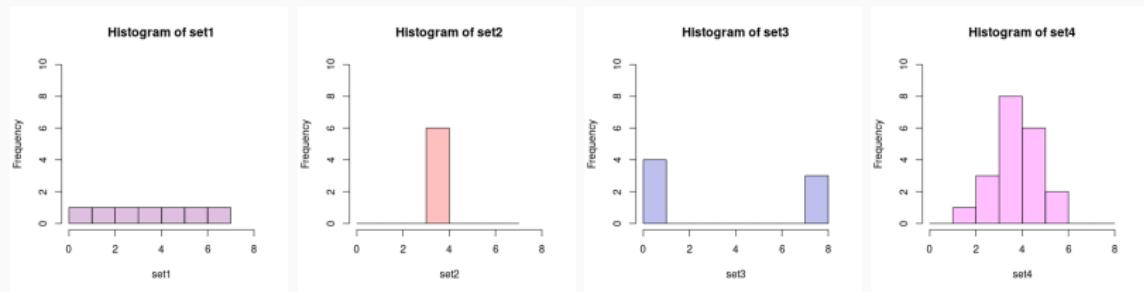
Asked for %, got proportions (accepted this time).

Count “don’t know” and “incorrect” together.

When counting incorrect answers, higher number = worse accuracy.

# Noisy channel

I wanted to see the standard deviation or similar.



Questions?

# Table of contents

1. Where are we this week?

2. Data visualization

3. Accessibility

4. Choice of visualization

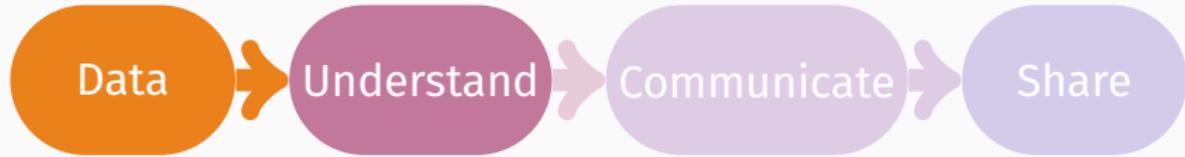
5. Plotting in R

6. Wrap-up

Where are we this week?

---

# Recap



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

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

document,  
create clean  
and beautiful  
reports

connect,  
collaborate,  
backup

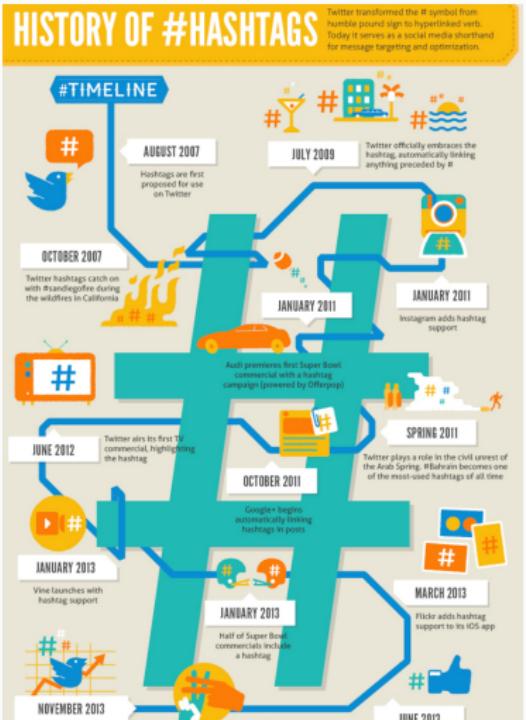
# Data visualization

---

# What is data visualization?

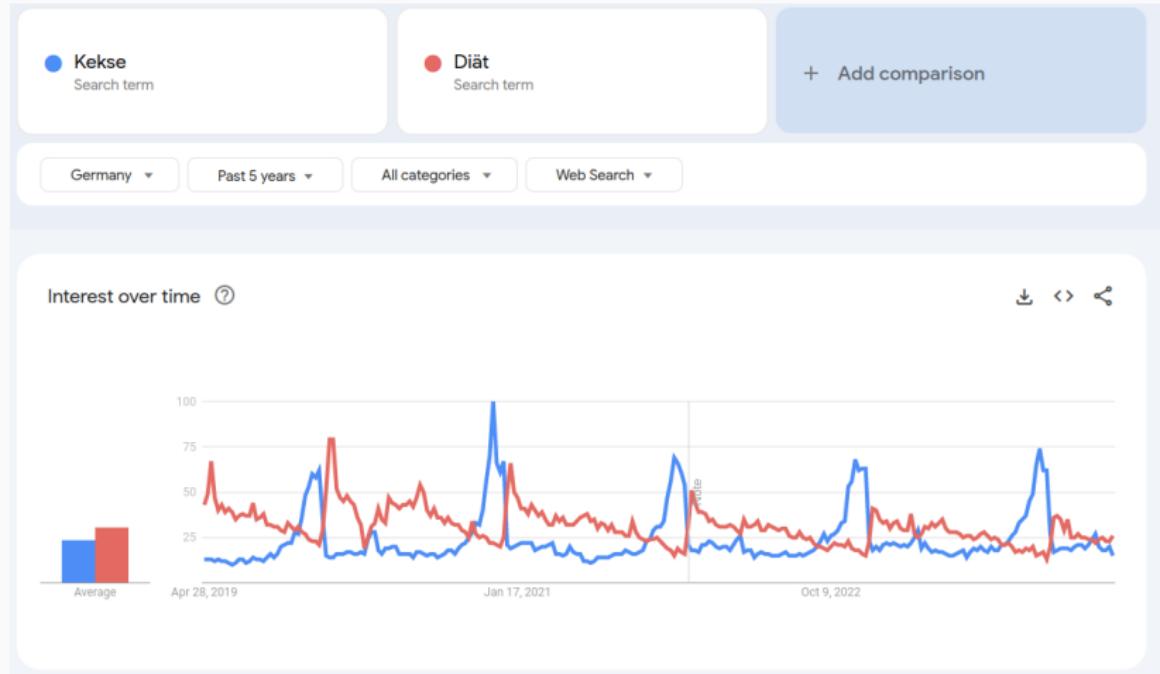


Interdisciplinary field



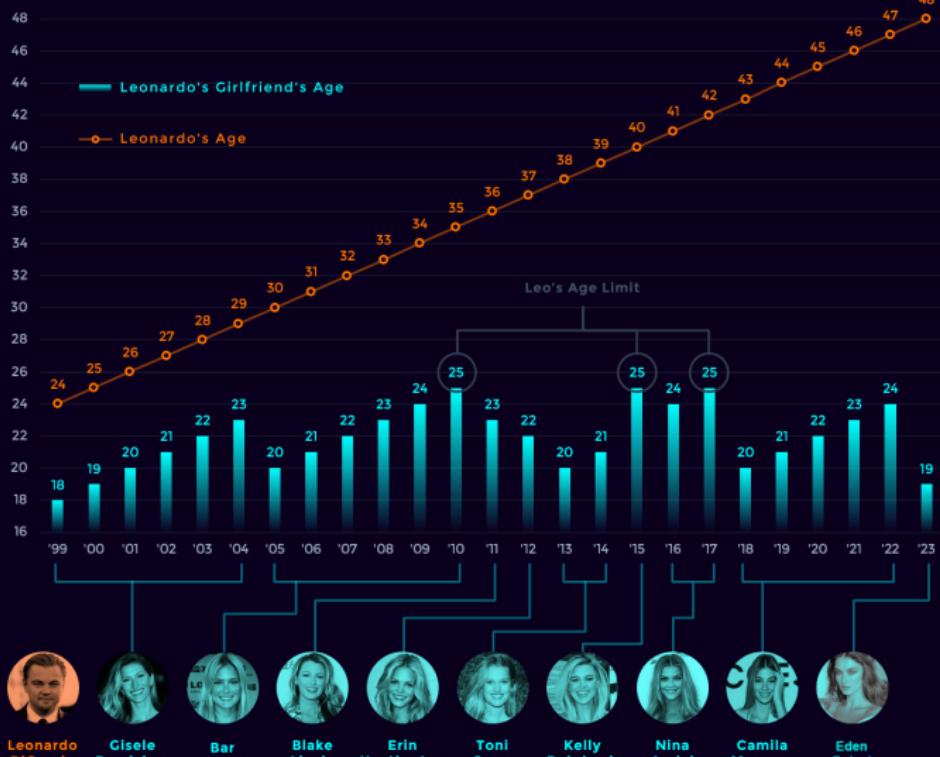
Graphical representation of data & information

# What is data visualization?



Efficient way of communicating (Google Trends cookies vs. diet)

### LEONARDO DICAPRIO REFUSES TO DATE A WOMAN OVER 25



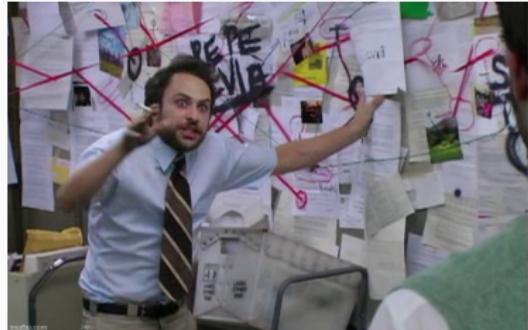
Original Design : u/TrustLittleBrother

Communicate complex relationships & insights in a comprehensible way

# Data visualization goals



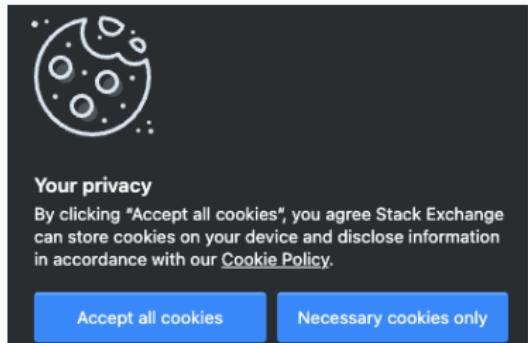
Communication & understanding



Analysis & exploration



Decision making

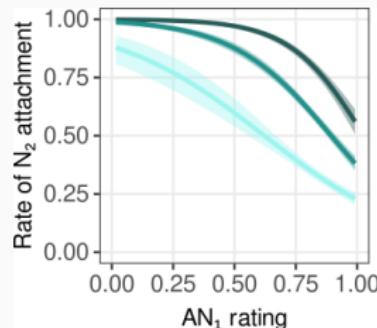


It's likely required in your career

# Visualization types

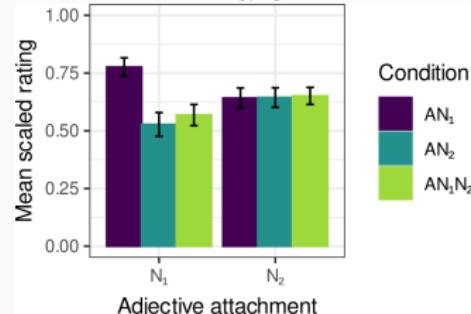
## Tables

item	sentence	N1	N2	unclear
3	sommerliche Arbeitskleidung	p 20	p	
11	silberne Bahnhofsuhren	p 20	p	
12	amtliche Baugenehmigung	p 19	1	
16	hohler Bienenstock	p 20	p	
19	gefüttertes Brillenetui	p 20	p	
22	lackiertes Bücherregal	p 20	p	
23	evangelischer Büroangestellter	p 20	p	

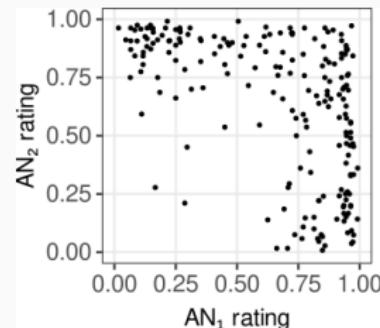


Lines

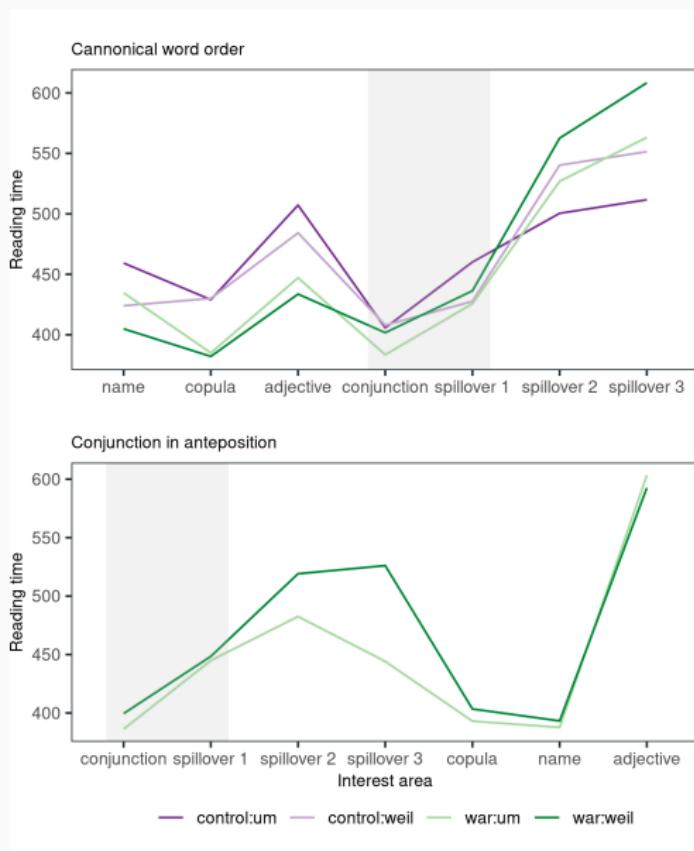
## Bars



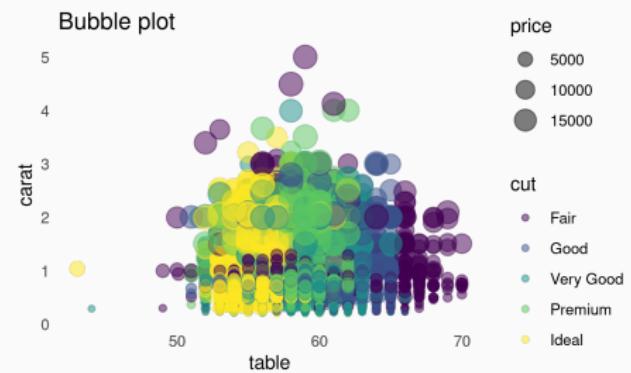
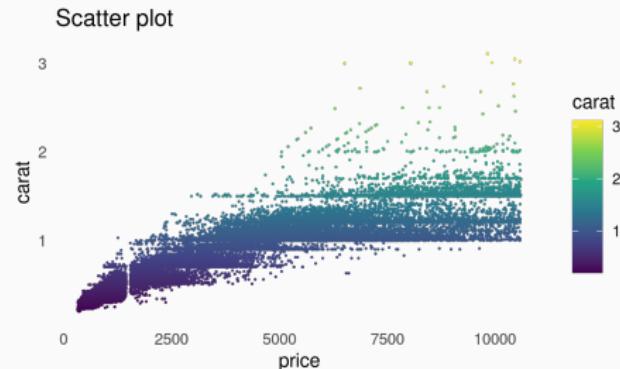
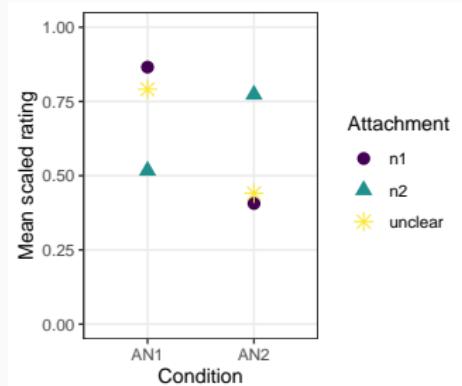
Points



# Lines



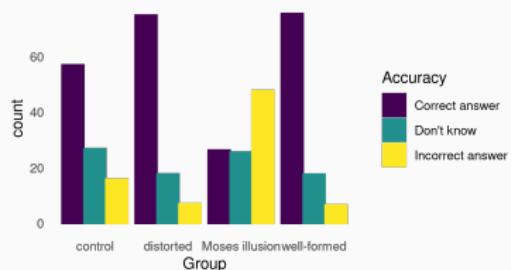
# Points



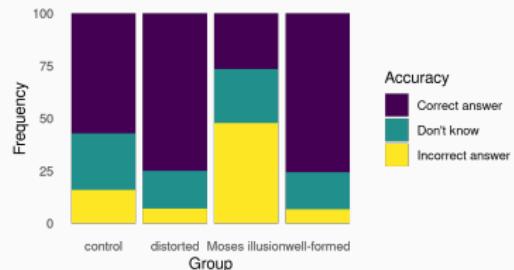
→<https://r-graph-gallery.com/>

# Bars

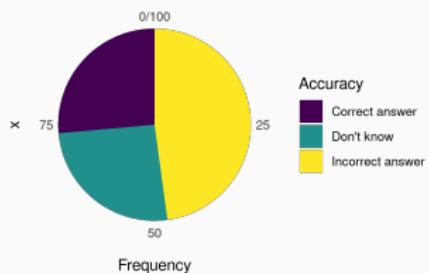
Bar plot



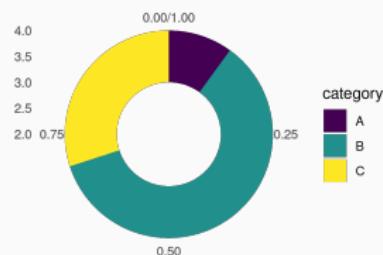
Stacked bar plot



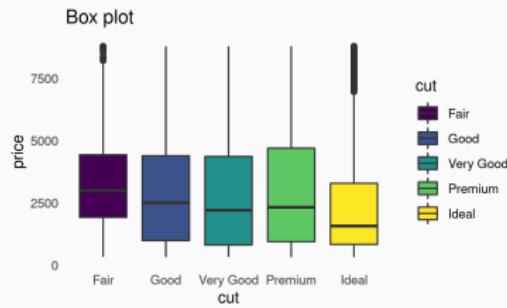
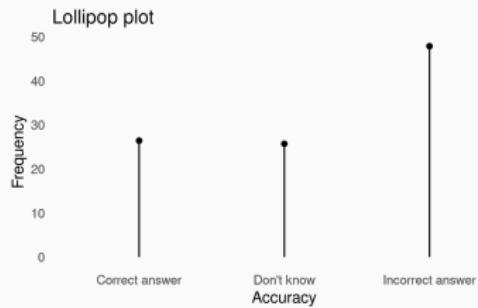
Pie plot



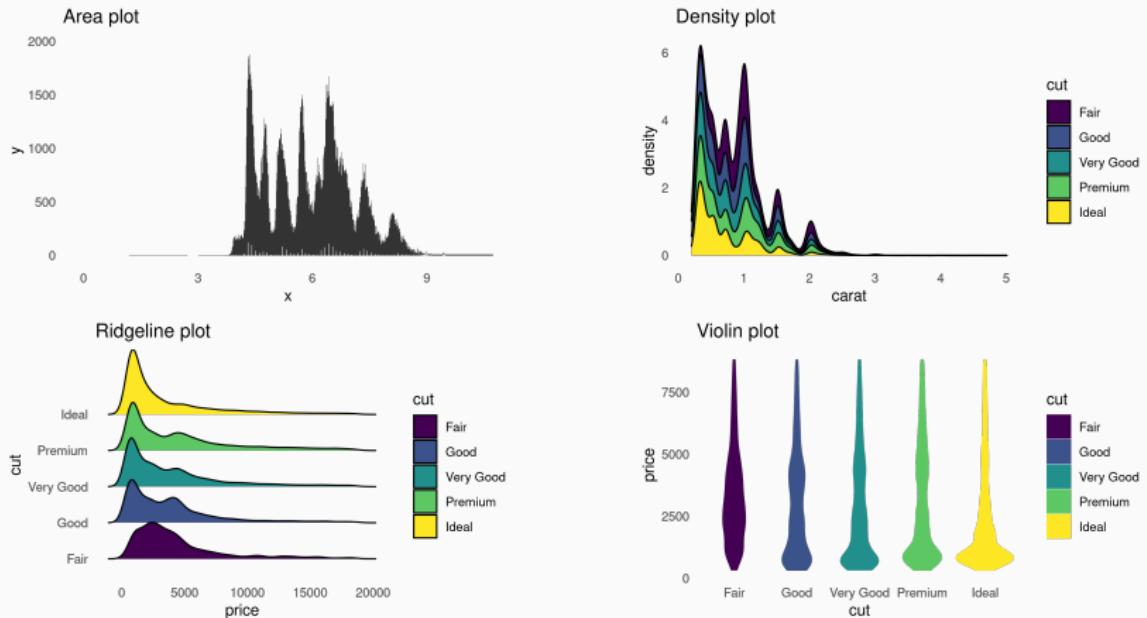
Doughnut plot



# Hybrids



# Hybrids

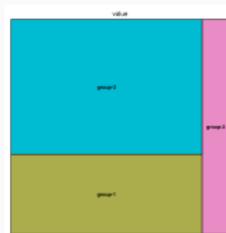


# Others

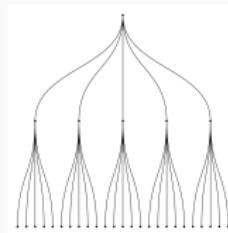
Radar



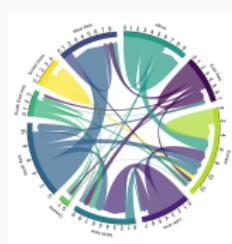
Treemap



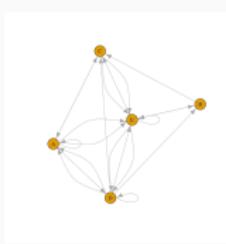
Dendrogram



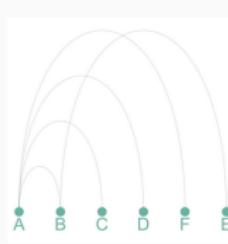
Wordcloud



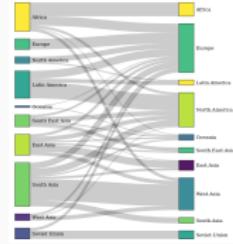
Chord



Network



Arc



Sankey

Source: [www.data-to-viz.com](http://www.data-to-viz.com)

# Accessibility

---

# The 4 Principles of Accessibility

## Web Content Accessibility Guidelines (WCAG)



**Perceivable:** Information and interface components are perceivable by users (understandable and not hidden from any senses).

**Operable:** User interface components and navigation should be operable (usable and cannot require tasks beyond users' abilities).

**Understandable:** Information and the operation of user interfaces must be understandable (understand the information and how to interact with the interface).

**Robust:** Content must be so robust that most users can interpret it (as technology progresses, content should remain accessible, including via assistive technologies).

Best honest effort

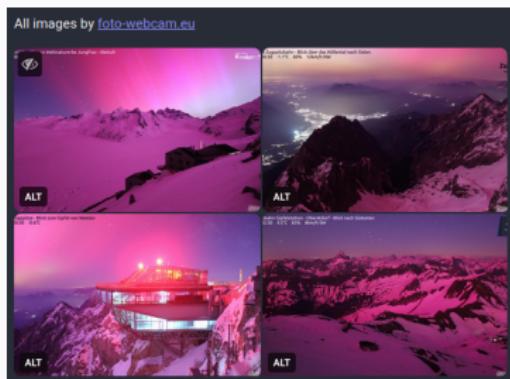


# Perceivable Application

Provide **text alternatives** for non-text content (e.g. ALT, subtitles) that can be changed into LARGE PRINT, Braille, speech, translations, simple language.

Create content that can be modified **without losing information or structure** (e.g. simple layout, reader view).

Choose a form that makes **seeing and hearing content easy** (e.g. foreground vs. colors).



The screenshot shows a news article from the German newspaper 'taz'. The header reads 'Pilotprojekt Grundeinkommen: Geld bedeutet Selbstbestimmung'. Below the header, there is a summary: 'Drei Jahre lang erhielten 122 Personen Geld, einfach so. Zwei von ihnen ziehen jetzt ein erstes Fazit. Doch das Konzept wird zunehmend kritisiert.' The author's name 'Hannes Koch' and a duration of '5-10 minutes' are also visible.

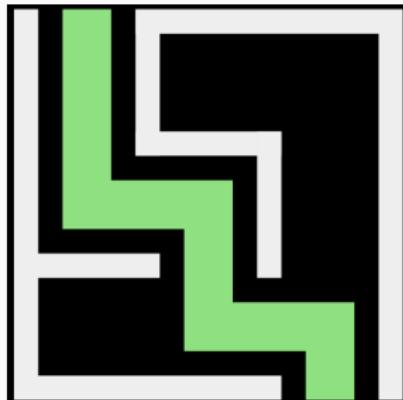
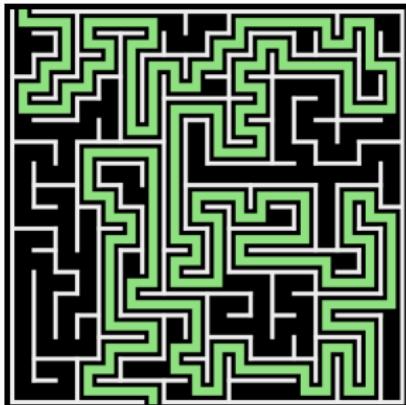
# Operable Application

Make all functionality **keyboard accessible** and allow different input modalities.

Give **enough time** to read and use content.

Don't design content in a way that is known to cause **seizures or physical reactions**.

Make content easy to find and navigate.



# Understandable Application

Provide **labels and instructions**.  
Make content readable and understandable.

Make your content appear and operate in **predictable ways**.

Provide **input assistance** (to avoid/correct mistakes).



# Robust Application

Maximize compatibility with current and future software and hardware.

Make compatible with assistive technologies.

Simple UI > complex UI.

Use simple markup languages.

Test content for crashes.

Include meta information (session info).



## Choice of visualization

---

# Plot choice examples

These are loose recommendations. Every case is different.

**Change over time**

line plot

**Part-to-whole**

stacked bar or area plot

**Data distribution**

bar plot, histogram, density curve, box plot,  
violin plot

**Comparing groups**

bar plot, dot plot, line plot, box plot, violin  
plot

**Relationships**

**between variables**

scatter plot, bubble plot

# Plotting in R

---

# ggplot2: A Layered Grammar of Graphics

Themes  
Coordinates  
Statistics  
Facets  
Geometries  
Aesthetics  
Data



Created by Hadley Wickham.

Design and construct graphics in a structured manner from data upwards.

“A **layer** is a collection of geometric elements and statistical transformations”  
(Wickham et al. n.d.).

# Assembly



1. Create a plot object
2. Add aesthetic mapping
3. Add geometric objects
4. Add appearance and statistics  
... repeat 3 and 4 until you're happy
5. Print and profit!

```
my.plot <- ggplot() +  
  aes() +  
  geom_*( ) +  
  coord_*( ) + theme()
```

## Data

	CONDITION	IA	MeanRT	SD
	<chr>	<chr>	<dbl>	<dbl>
1	implausible	1	823.	119.
2	implausible	2	759.	107.
3	implausible	3	818.	179.
4	implausible	4	1023.	133.
5	plausible	1	781.	106.
6	plausible	2	700.	64.8
7	plausible	3	783.	115.
8	plausible	4	1092.	137.

# Aesthetics

Aesthetics `aes()` translates data into visual elements:

<code>x, y</code>	variables
<code>colour</code>	colours the lines of geometries
<code>fill</code>	fill geometries
<code>group</code>	groups based on the data
<code>shape</code>	defines the shape (point, triangles)
<code>linetype</code>	defines the type of line (solid, dashed)
<code>size</code>	define sizes of elements
<code>alpha</code>	changes the transparency

# Geometries

Geometries (`geom_*`( )) provide the shapes and patterns to the data:

`geom_point()`

dot-plot (e.g. scatter-plot)

`geom_line()`

lines connecting (invisible) points

`geom_bar()`

bar charts for categorical x axis

`geom_histogram()`

histogram for continuous x axis

`geom_boxplot()`

box plot for categorical variables

# Facets

Facets (`facet_*`) divide your plot into smaller sections across panels.

`facet_wrap()` based on one variable, it arranges plots in a single row or column, adding more rows/columns as needed

`facet_grid()` does the same, but based on two variables

# Statistics

Statistics (`stat_*`( )) compute and add statistical transformations to the data (e.g. means, counts, linear models)

# Coordinates

Coordinates (`coord_*( )`) change the coordinate system and axes.

<code>coord_cartesian()</code>	sets the limits of the coordinate system
<code>coord_polar()</code>	circular plots
<code>coord_map()</code>	map projections
<code>coord_flip()</code>	flips the coordinates

# Themes

Themes (`theme_*`( )) fine-tune the overall appearance and style:

- fonts,
- background color,
- line width,
- legend,
- axis and tick label sizes and colors,
- ... and so much more!

## Honorable mention: Labels

`labs()` allows you to set titles for the x-axis, y-axis, plot title, and other labels.

<code>x</code>	x-axis
<code>y</code>	y-axis
<code>title</code>	plot title
<code>subtitle</code>	plot subtitle
<code>caption</code>	an additional caption
<code>colour</code>	color legend title
<code>size</code>	size legend title
<code>fill</code>	fill legend title



memegenerator.net

# esquisse: a rough or preliminary sketch



**esquisse** lets you explore data interactively. Uses **ggplot2** for visualization.

You can export the generated graph and save the code to generate it.

Available in many languages (incl. French, Turkish, Korean, German).

It has its limitations but is useful to get a first impression/overview.

## Wrap-up

---

# Summary

- ✓ data visualization goals
- ✓ accessibility and WCOG
- ✓ plot types and choice of visualization
- ✓ plotting in R with `ggplot2` and `esquisse`
- no class next week, then more data viz

# Homework assignment due May 17th 15:30

Submit 1 R script.

- ➊ Complete assignment 5 ( $\rightarrow$  ILIAS)
- ➋ Cookbook for R: <http://www.cookbook-r.com/>
- ⌃ R Graph Gallery: <https://r-graph-gallery.com/>
- ⌄ Tidyverse documentation:  
<https://ggplot2.tidyverse.org/>
- ⌅ Wickham et al. (n.d.)

## References

---

-  Wickham, Hadley, Danielle Navarro, and Thomas Lin Pedersen (n.d.). *ggplot2: Elegant Graphics for Data Analysis*. URL: <https://ggplot2-book.org/>.