

Introduction to R

SpelkeLab R workshop

Irene Canudas Grabolosa, 07/18/24



Agenda for today

1. Getting ready (15 mins)
2. Basic notions about data analysis (30 mins)
 1. Types of data & why do we care
 2. What type of data are you collecting this summer?
 3. Statistical analysis in R
 4. Presentation of our dataset
3. Hands-on! (45 mins)
 1. Importing data
 2. Tidying data
 3. Transforming data
 4. Visualizing data
 5. Modeling data

Getting ready

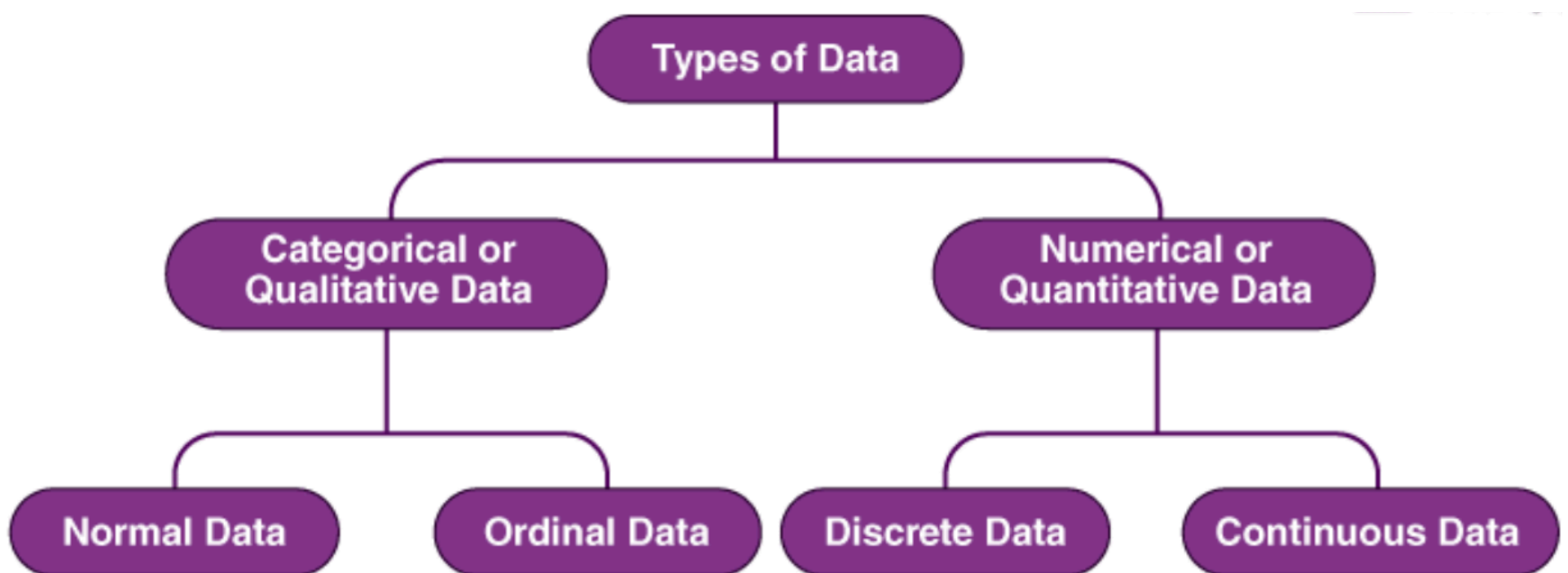
- Create an account at Posit Cloud (if you haven't done it)
- Download the data and the script from: <https://github.com/irenecanudas/RWorkshop>



Recapping: Types of data

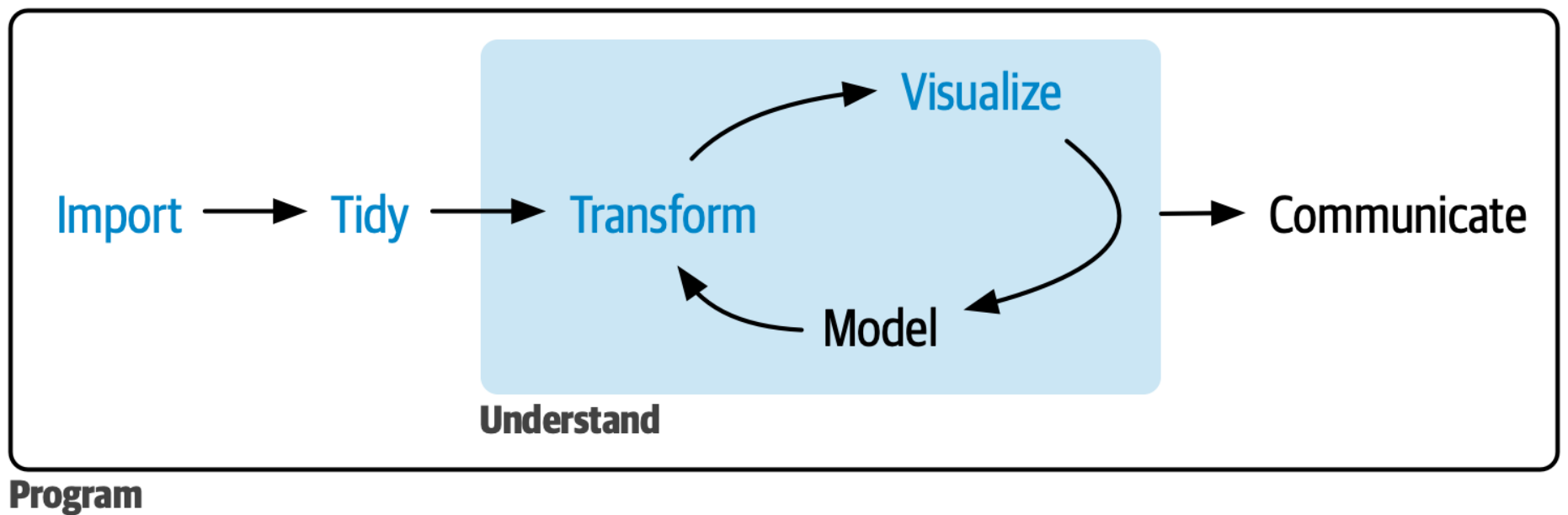
Fast reminder

- **DV (dependent variable):** the response/outcome we're measuring
- **IV (independent variable):** Variables not affected by any other variables measured by the study



- The type of DV and IV will influence the statistical test we use
- Which kind of data are you collecting?

Statistical analysis with R



- R has:
 - Variables: *placeholders for information*
 - Different types:
 - characters [~text]
 - integers & numbers
 - factors [labels]
 - ...
 - Functions: *commands to transform the data*

One important thing to keep in mind!

• R is stupid

- If you don't close a parenthesis/quotation marks/brackets...
error
- Pay attention to the variables' type! A variable containing a 2 BUT labeled as a "character" cannot be summed.
- name ≠
 - Name
 - NAME
 - nme
- "name" = string of letters "n" "a" "m" "e"
Vs
name = a variable you want to access its content
- name-of
 - That means, take the value of the variable "name" and subtract the value of the variable "of" from it
- If you ask R to open a file that is not directly in the folder (even if it's in the immediate subfolder) it will give you an **error**.

• R is powerful

- If you ask it to perform an inadequate statistical test for your data, it will do it. It won't warn you that that's the wrong thing to do!

Our dataset: dragons

- We're dragon trainers
- **Aim:** We want to understand the effect of training in dragons.
- **The dataset:**
 - Two dragon populations:
Southern & Maritime dragons
 - Both male and female dragons
 - Also collecting their age
 - 2 data points: pre & post training
 - IQ score (TestScore)
 - Body Weight
- **Hypotheses:**
 - *H1:* The training should work equally across populations:
 - *H1.1.* Dragons will have higher Test scores after training than before
 - *H1.2.* Where a dragon is from should have no effect on their improvement.
 - *H2:* Being smarter will impact on their ability to hunt: smarter dragons will hunt more, eat more, and therefore will weight more:
 - *H2.1:* Dragons will weight more after training than before
 - *H2.2:* The smarter a dragon is, the heavier it will be

} Independent variables

} Dependent variables



1. Importing the data

- Importing the data into R
- Visually inspect the dataset to make sure it was correctly imported.

2. Tidy the data

- Format and tidy the data so R can understand it
- All variables are assigned to the correct type
- Columns contain information about ONE variable

ID	Words produced
865934	3,car, bus and dog
583945	0
328492	4, chair, table, piano, palnt
58374	2, computer, glasses
201834	5; basketball, football, phone, party, bus
838739	6; car, dog, cat, jellyfish, key, clock



2. Tidy the data

- Format and tidy the data so R can understand it
- All variables are assigned to the correct type
- Columns contain information about ONE variable

ID	# words produced	Words produced
865934	3	car, bus, dog
583945	0	
328492	4	chair, table, piano, palnt
58374	2	computer, glasses
201834	5	basketball, football, phone, party, bus
838739	6	car, dog, cat, jellyfish, key, clock



3. Transform the data

- Format and the data so R can use it for the analysis you want
- General rule of thumb: one row per observation

ID	Trial 1	Trial 2	Trial 3	Trial 4
865934	1	3	8	3
583945	5	4	6	5
328492	7	2	4	3
58374	3	5	7	6
201834	6	6	2	7
838739	3	7	1	4



3. Transform the data

- Format and the data so R can use it for the analysis you want
- General rule of thumb: one row per observation

ID	Trial	Value
58374	1	3
58374	2	5
58374	3	7
58374	4	6
201834	1	6
201834	2	6
201834	3	2
201834	4	7
328492	1	7
328492	2	2
328492	3	4
328492	4	3
583945	1	5
583945	2	4
583945	3	6
583945	4	5
838739	1	3
838739	2	7
838739	3	1
838739	4	4
865934	1	1
865934	2	3
865934	3	8
865934	4	3



4. Visualize the data

- Think about what you want to see from the data
- Usually the best way to intuitively understand whether you had the effect you were looking for

5. Model the data

- Stick to the hypotheses: it's very easy to get lost.
- *H1*: The training should work equally across populations:
 - *H1.1*. Dragons will have higher Test scores after training than before
 - *H1.2*. Where a dragon is from should have no effect on their improvement.
- *H2*: Being smarter will impact on their ability to hunt: smarter dragons will hunt more, eat more, and therefore will weight more:
 - *H2.1*: Dragons will weight more after training than before
 - *H2.2*: The smarter a dragon is, the heavier it will be
- What variables do we need to consider to analyze the hypotheses?
- Exploratory analyses will come later

Useful links

- R for data science: <https://r4ds.hadley.nz/>
- Introduction to linear models: <https://gkhajduk.github.io/2017-03-09-mixed-models/>
- Another nice intro to linear models: <https://pagepiccinini.com/2016/01/08/introduction-and-linear-models-part-1/>