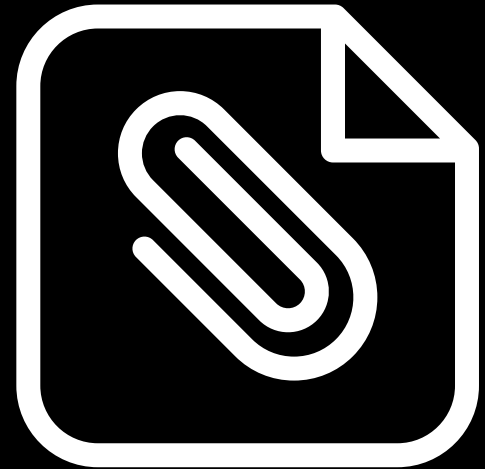


Reproduceerbaar Analyseren voor Praktijkgerichte Onderzoekers:

Dag 2

Programma

- Introductie Rstudio
- Introductie R
- Beginnen met data
- Morgen: Data Wrangling
- Overmorgen: Data visualisatie

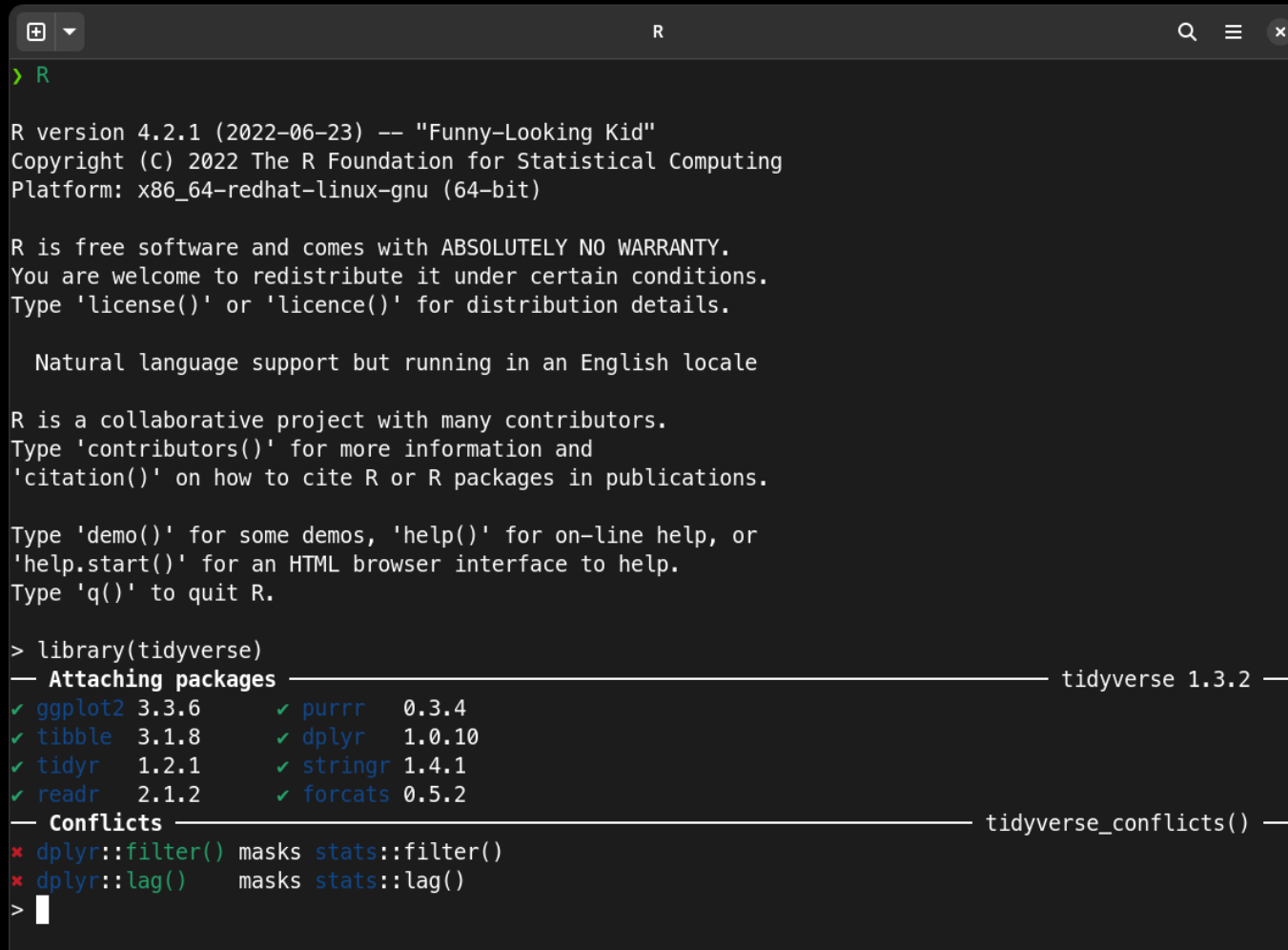


Waarom R?

- Geen klikken
- Reproduceerbaar
- Interdisciplinair
- Werkt op alle data
- Hoge kwaliteit figuren
- Vrije en open source software



Waarom RStudio?



```
R

R version 4.2.1 (2022-06-23) -- "Funny-Looking Kid"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-redhat-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.

> library(tidyverse)
— Attaching packages — tidyverse 1.3.2 —
✔ ggplot2 3.3.6    ✔ purrr  0.3.4
✔ tibble  3.1.8    ✔ dplyr  1.0.10
✔ tidyr   1.2.1    ✔ stringr 1.4.1
✔ readr   2.1.2    ✔ forcats 0.5.2
— Conflicts — tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()     masks stats::lag()
> 
```

Waarom RStudio?

The screenshot displays the RStudio IDE interface. The main editor window on the left contains R code for creating a boxplot and a bar chart. The Environment pane on the right shows the current data objects. The bottom pane is split into a Console window and an R Documentation window.

```
124 ggplot(aes(x = respondent_wall_type, y = rooms)) +  
125   geom_boxplot() +  
126   geom_jitter(height = 0,  
127             alpha = 0.5,  
128             color = "pink")  
129  
130 interviews_plotting %>%  
131   ggplot(aes(x = respondent_wall_type)) +  
132   geom_bar(aes(fill = village), position = "dodge") +  
133   labs(title = "Title",  
134        x = "X",  
135        y = "Y")  
136  
137 # percent_wall_type <- interviews_plotting %>%  
138 percent_wall_type <- interviews_plotting %>%  
139   filter(respondent_wall_type != "cement") %>%  
140   count(village, respondent_wall_type) %>%  
141   group_by(village) %>%  
142   mutate(percent = (n / sum(n)) * 100) %>%  
143   ungroup()  
144  
145 percent_wall_type %>%  
146   ggplot(aes(x = ))  
147  
148  
149  
150  
151  
152
```

Environment

Object	Class	Attributes
interviews	data.frame	131 obs. of 14 variables
interviews_filter...	data.frame	39 obs. of 5 variables
interviews_plotti...	data.frame	131 obs. of 45 variables
interviews_wide	data.frame	131 obs. of 17 variables
members	data.frame	33 obs. of 5 variables
people_per_room	data.frame	131 obs. of 15 variables
village_info	data.frame	131 obs. of 3 variables

Values

Variable	Value
area_hectares	1
hh_members	num [1:4] 2 3 10 6

Console

```
R 4.2.1 - ~/Projects/r-socialsci-workshop/  
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 ~/Projects/r-socialsci-workshop/.RData]  
> |
```

R Documentation: Bar charts

Description

There are two types of bar charts: `geom_bar()` and `geom_col()`. `geom_bar()` makes the height of the bar proportional to the number of cases in each group (or if the weight aesthetic is supplied, the sum of the weights). If you want the heights of the bars to represent values in the data, use `geom_col()` instead. `geom_bar()` uses `stat_count()` by default: it counts the number of cases at each x position. `geom_col()` uses `stat_identity()`: it leaves the data as is.

Usage

```
geom_bar(  
  mapping = NULL,
```

Rondleiding

Opdracht

Gebruik het Console en de Packages tab om te bevestigen dat de “tidyverse” is geïnstalleerd.

Introductie R

- Rekenen
- Comments
- Functies
- Vectors
- Missing data

Opdracht: variabelen

Wat staat er nu in `area_acres`?
123.5 of 6.175?

Opdracht: functies

Typ `?round` in de Console en kijk naar de output in de help pagina. Welke functies bestaan er nog meer die lijken op `round`? Wat doet de `floor` functie?

Opdracht: vectors

Wat gebeurt er als je datatypes
door elkaar gebruikt?

Opdracht: missing values

- Maak de volgende vector aan en verwijder de NAs:
`rooms <- c(1, 2, 1, 1, NA, 3, 1, 3, 2, 1, 1, 8, 3, 1, NA, 1)`
- Gebruik de functie `median()` om de mediaan te berekenen.
- Kijk hoe veel huishoudens meer dan 2 kamers hebben.

Starting with Data

data frame

columns

rows	key_id	village	has_bicycle
	1	"God"	TRUE
	7	"God"	FALSE
	3	"Chirodzo"	TRUE
	numeric	character	logical

De data

column_name	description
key_id	Added to provide a unique Id for each observation. (The InstanceID field does this as well but it is not as convenient to use)
village	Village name
interview_date	Date of interview
no_membrs	How many members in the household?
years_liv	How many years have you been living in this village or neighboring village?
respondent_wall_type	What type of walls does their house have (from list)
rooms	How many rooms in the main house are used for sleeping?
memb_assoc	Are you a member of an irrigation association?
affect_conflicts	Have you been affected by conflicts with other irrigators in the area?
liv_count	Number of livestock owned.
items_owned	Which of the following items are owned by the household? (list)
no_meals	How many meals do people in your household normally eat in a day?
months_lack_food	Indicate which months, In the last 12 months have you faced a situation when you did not have enough food to feed the household?
instanceID	Unique identifier for the form data submission