

Vorlesung

Daten

Vom Zettel nach Excel nach R

Version: 2.0

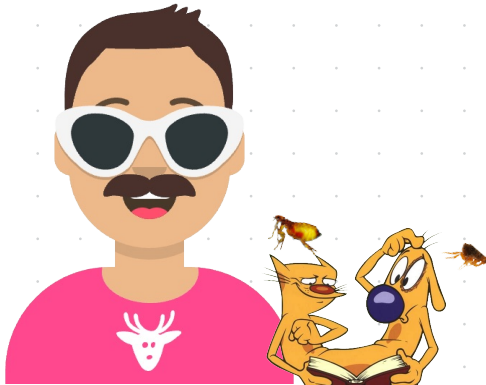


Unterscheiden sich die Sprungweiten von

Hunde- und Katzenflöhen?

↓ animal
x "exchange"
ω - Worte

→ jump-length
Y "yield"
Messwert



Messwerte y
"yield"

X "exchange"

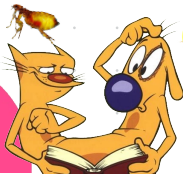
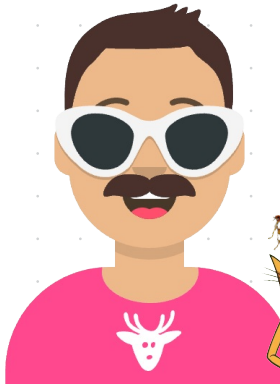
W- Worte

Was Womit

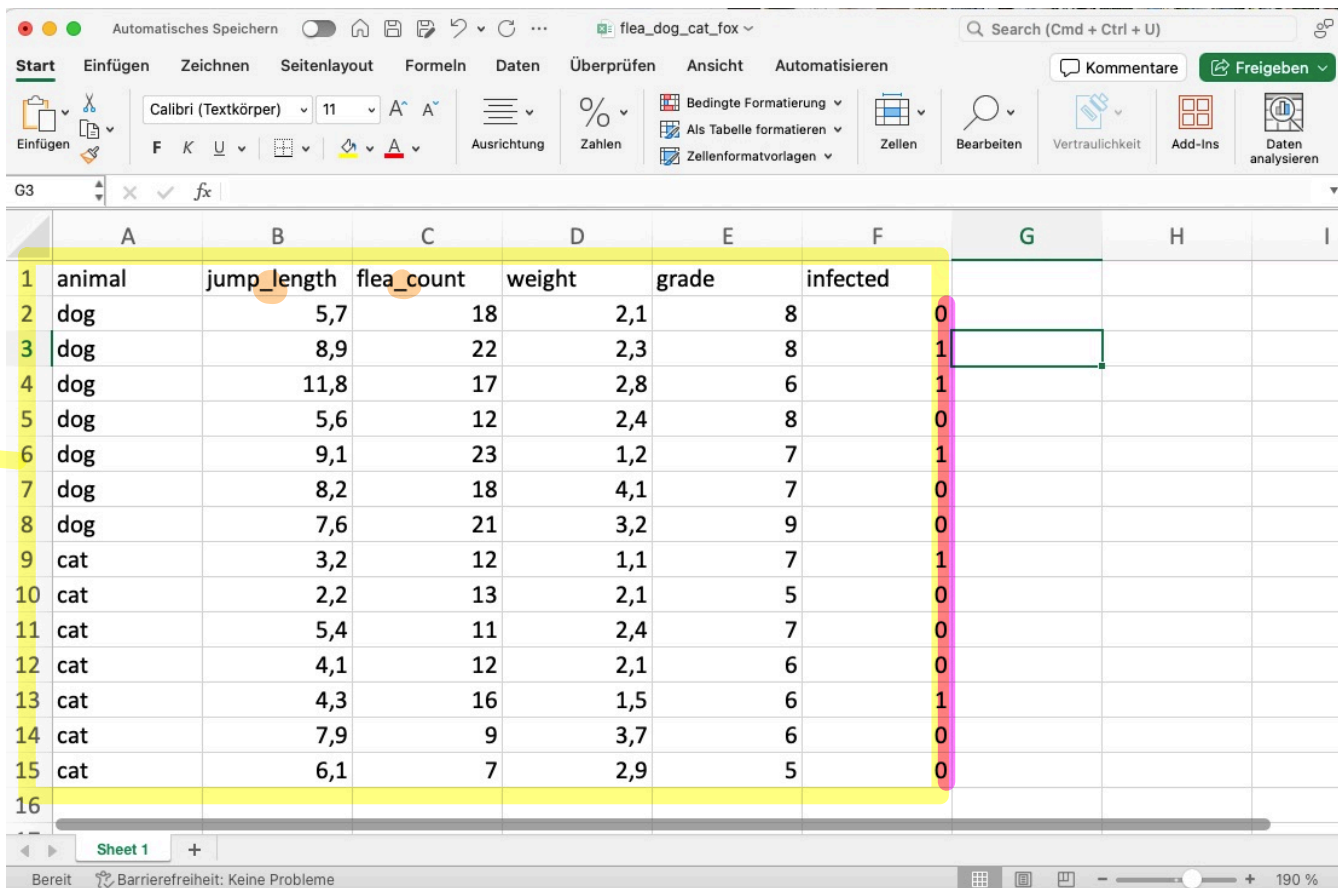
Wo
Wie
Wann

Tabelle 9.1— Tabelle von sieben Hunde- und Katzenflöhen mit der Sprunglänge [cm], Anzahl an Haaren am rechten Flohbein, Gewicht der Flöhe, Boniturnote sowie der Infektionsstatus für Flohschnupfen. Die erste Spalte `animal` gibt an, dass wir es hier mit Hunde- und Katzenflöhe zu tun haben. Die Tabelle ist im Long-Format dargestellt.

animal	jump_length	flea_count	weight	grade	infected
dog	5.7	18	2.1	8	FALSE
dog	8.9	22	2.3	8	TRUE
dog	11.8	17	2.8	6	TRUE
dog	5.6	12	2.4	8	FALSE
dog	9.1	23	1.2	7	TRUE
dog	8.2	18	4.1	7	FALSE
dog	7.6	21	3.2	9	FALSE
cat	3.2	12	1.1	7	TRUE
cat	2.2	13	2.1	5	FALSE
cat	5.4	11	2.4	7	FALSE
cat	4.1	12	2.1	6	FALSE
cat	4.3	16	1.5	6	TRUE
cat	7.9	9	3.7	6	FALSE
cat	6.1	7	2.9	5	FALSE



Daten in Excel



Automatisches Speichern

Search (Cmd + Ctrl + U)

Start Einfügen Zeichnen Seitenlayout Formeln Daten Überprüfen Ansicht Automatisieren

Einfügen Einfügen

Calibri (Textkörper) 11 A⁺ A⁻

F K U Ausrichtung Zahlen Bedingte Formatierung Als Tabelle formatieren Zellenformatvorlagen Zellen Bearbeiten Vertraulichkeit Add-Ins Daten analysieren

G3

	A	B	C	D	E	F	G	H	I
1	animal	jump_length	flea_count	weight	grade	infected			
2	dog	5,7	18	2,1	8	0			
3	dog	8,9	22	2,3	8	1			
4	dog	11,8	17	2,8	6	1			
5	dog	5,6	12	2,4	8	0			
6	dog	9,1	23	1,2	7	1			
7	dog	8,2	18	4,1	7	0			
8	dog	7,6	21	3,2	9	0			
9	cat	3,2	12	1,1	7	1			
10	cat	2,2	13	2,1	5	0			
11	cat	5,4	11	2,4	7	0			
12	cat	4,1	12	2,1	6	0			
13	cat	4,3	16	1,5	6	1			
14	cat	7,9	9	3,7	6	0			
15	cat	6,1	7	2,9	5	0			
16									

Sheet 1

Bereit Barrierefreiheit: Keine Probleme

190 %

Tidy

Daten als tibble in R

Programmieren ist wie eine
Sprache lernen!

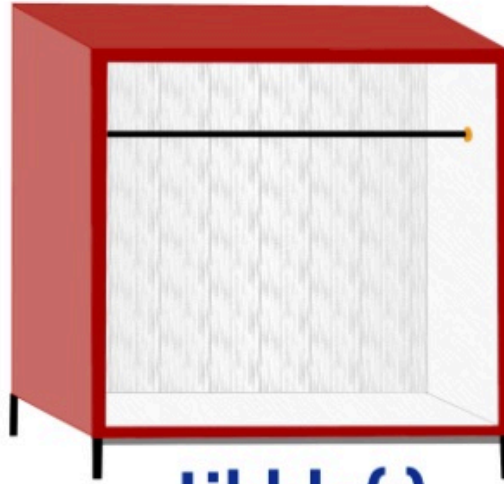


① Ein **tibble** ist **tidy** (deu. *sauber*)

Das Datenformat, was wir hier erarbeiten wollen, nennen wir auch *tidy* (deu. *sauber*) nach dem R Paket tidyr was dann auch später mit die Basis für unsere Analysen in R sein wird. Wenn ein Datensatz *tidy* ist, dann erfüllt er folgende Bedingungen.

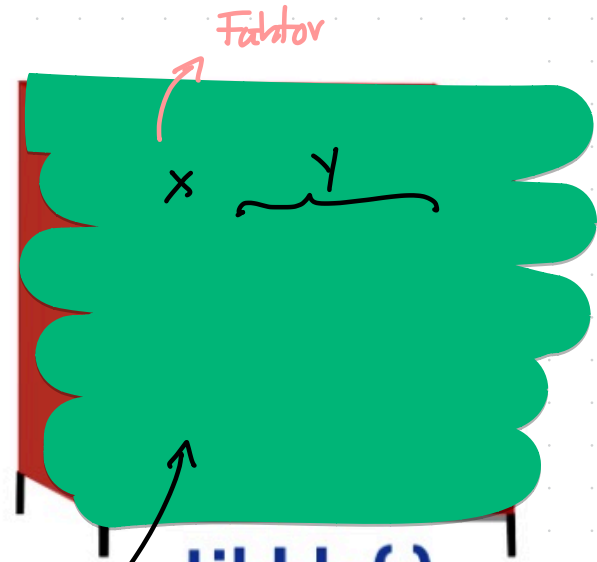
1. **Jede Variable ist eine Spalte**; jede Spalte ist eine Variable.
2. **Jede Beobachtung ist eine Zeile**; jede Zeile ist eine Beobachtung.
3. **Jeder Wert ist eine Zelle**; jede Zelle ist ein einziger Wert.

Nach diesen Regeln bauen wir dann jeden Datensatz auf, den wir in einem Experiment gemessen haben.



tibble()

(a) Ein leerer `tibble` ist wie ein Schrank der befüllt werden kann.



tibble()

(b) Ein `tibble` mit drei Variablen, die wie Kleiderbügel in den Schrank hängen.

Variable

Zahlen

0/1

A tibble: 14 × 6

Wörter

Gruppe 1

Gruppe 2

animal jump_length flea_count weight grade infected

<chr>

<dbl>

<int>

<dbl>

<dbl>

<lgl>

1 dog

5.7

18

2.1

8

FALSE

2 dog

8.9

22

2.3

8

TRUE

3 dog

11.8

17

2.8

6

TRUE

4 dog

5.6

12

2.4

8

FALSE

5 dog

9.1

23

1.2

7

TRUE

6 dog

8.2

18

4.1

7

FALSE

7 dog

7.6

21

3.2

9

FALSE

8 cat

3.2

12

1.1

7

TRUE

9 cat

2.2

13

2.1

5

FALSE

10 cat

5.4

11

2.4

7

FALSE

11 cat

4.1

12

2.1

6

FALSE

12 cat

4.3

16

1.5

6

TRUE

13 cat

7.9

9

3.7

6

FALSE

14 cat

6.1

7

2.9

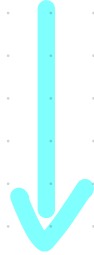
5

FALSE



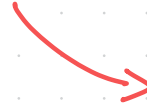
DATEN

Y
Messwerte
"yield"



R

"tibbles"



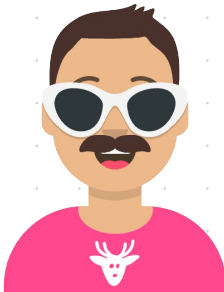
X

Faktoren → Snuppe

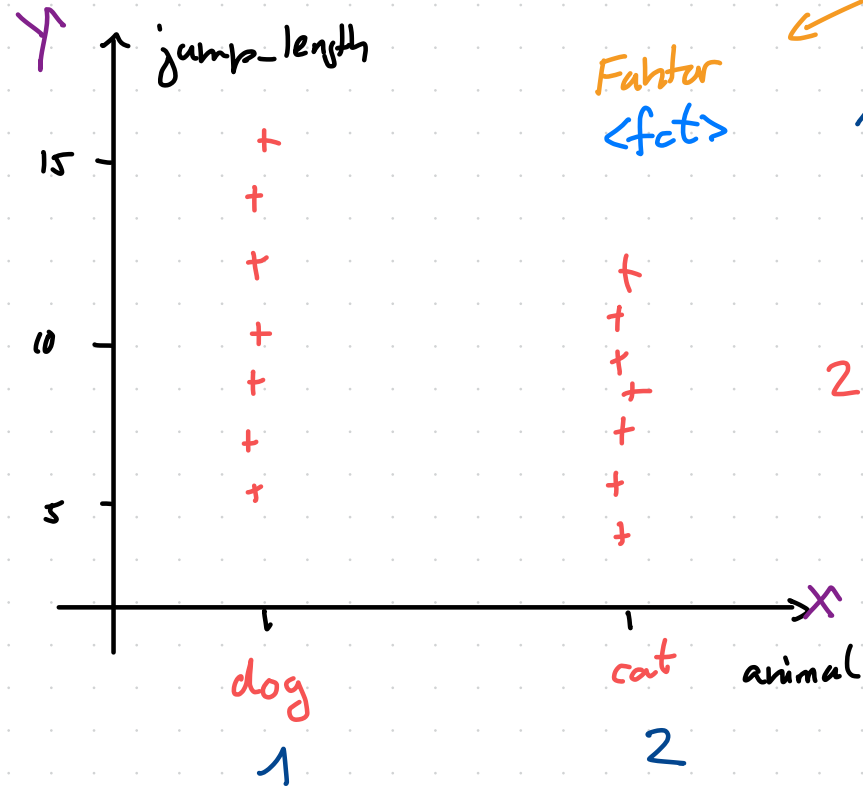
"exchange" animal

ω - Worte

- Wo
 - Was
 - Womit
 - Wann
- } Aussagen über die Beobachtung



Faktoren als Wörter zu Zahlen



A tibble: 14 x 6

animal	jump_length	flea_count	weight	grade	infected
1 dog	5.7	18	2.1	8	FALSE
2 dog	8.9	22	2.3	8	TRUE
3 dog	11.8	17	2.8	6	TRUE
4 dog	5.6	12	2.4	8	FALSE
5 dog	9.1	23	1.2	7	TRUE
6 dog	8.2	18	4.1	7	FALSE
7 dog	7.6	21	3.2	9	FALSE
8 cat	3.2	12	1.1	7	TRUE
9 cat	2.2	13	2.1	5	FALSE
10 cat	5.4	11	2.4	7	FALSE
11 cat	4.1	12	2.1	6	FALSE
12 cat	4.3	16	1.5	6	TRUE
13 cat	7.9	9	3.7	6	FALSE
14 cat	6.1	7	2.9	5	FALSE

character
Wörter

double
Komma
zahlen

integer
Ganz
zahlen

logical
ja/nein
1 0

Faktor

dog cat → Level 1, 2

Von Wörtern und Objekten und Funktionen

`as_tibble()`



Function



Erstellt
Objekte

`animal`



fct

Variable

Objekt

`"dog"`



chr

`"cat"`



chr