types of variables

prof. dr. nicolas meseth

1. observations and variables
2. four scales
3. scales in R

a ruler, a thermometer, a scale and a tally sheet all measure, but what you may do with their numbers is not the same.



image: ai-generated (gpt-image-2.5-flare)

part 1

observations and variables

what a row and a column mean

one row per game, one column per fact

game	prize_door	contestant_choice	monty_opens	decision
1	2	2	1	stay
2	1	3	2	switch
3	3	1	2	stay
4	3	1	2	switch

an observation is one game, and a variable is one fact that can differ from game to game.

after Döring (2023), p. 577, and R for Data Science, chapter "data tidying"

one row per game, one column per fact

game	prize_door	contestant_choice	monty_opens	decision
1	2	2	1	stay
2	1	3	2	switch
3	3	1	2	stay
4	3	1	2	switch

observation (row)

an observation is one game, and a variable is one fact that can differ from game to game.

after Döring (2023), p. 577, and R for Data Science, chapter "data tidying"

one row per game, one column per fact

game	prize_door	contestant_choice	monty_opens	decision
1	2	2	1	stay
2	1	3	2	switch
3	3	1	2	stay
4	3	1	2	switch

observation (row)

variable (column)

an observation is one game, and a variable is one fact that can differ from game to game.

after Döring (2023), p. 577, and R for Data Science, chapter "data tidying"

one row per game, one column per fact

game	prize_door	contestant_choice	monty_opens	decision
1	2	2	1	stay
2	1	3	2	switch
3	3	1	2	stay
4	3	1	2	switch

observation (row)

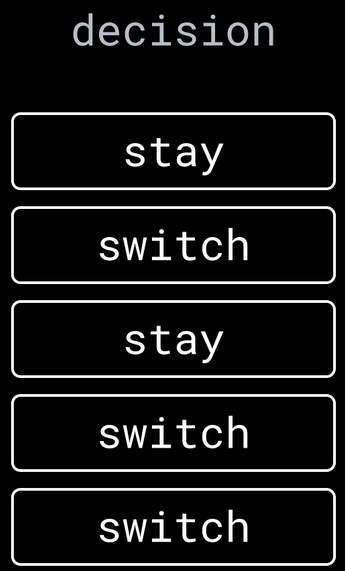
value (cell)

variable (column)

an observation is one game, and a variable is one fact that can differ from game to game.

after Döring (2023), p. 577, and R for Data Science, chapter "data tidying"

a variable varies



variable

differs from game to game



constant

the same in every game

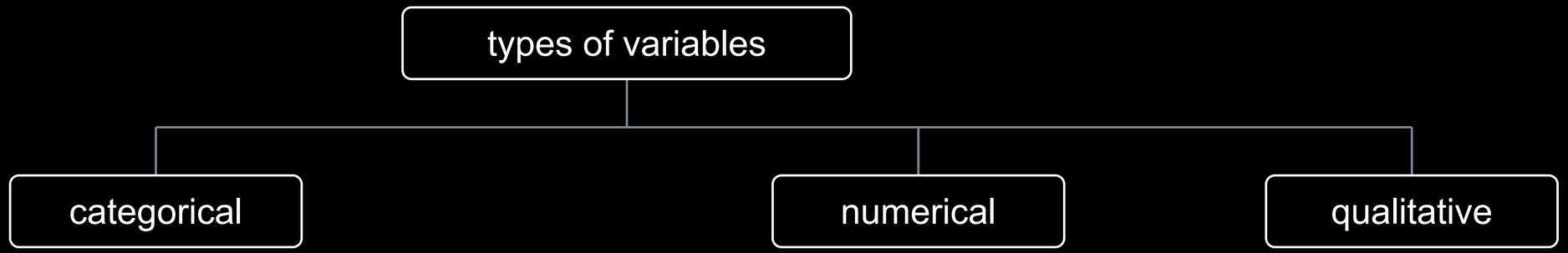
a fact that is the same in every game is a constant, and a constant needs no column.

part 2

four scales

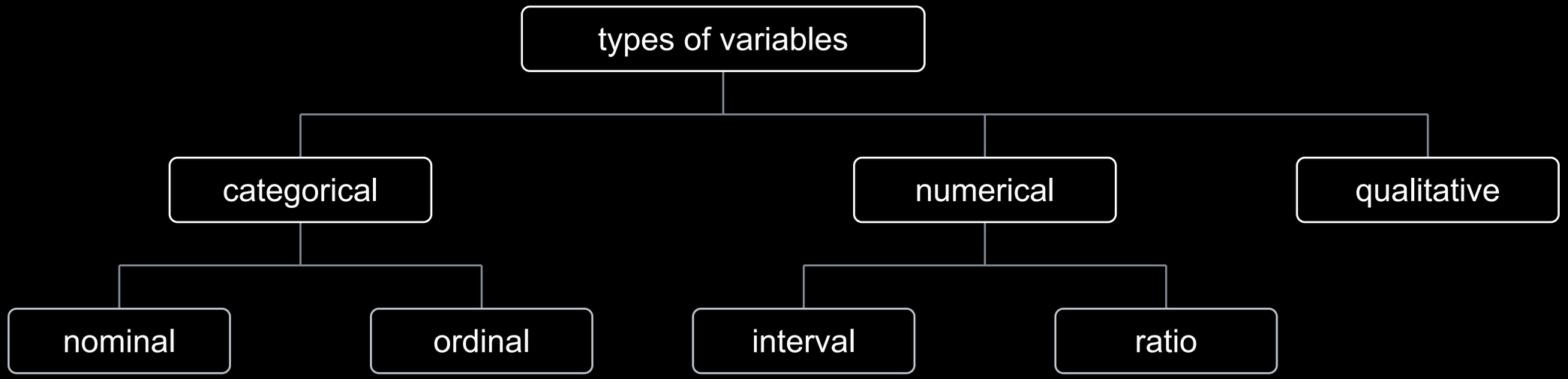
what the numbers may mean

types of variables

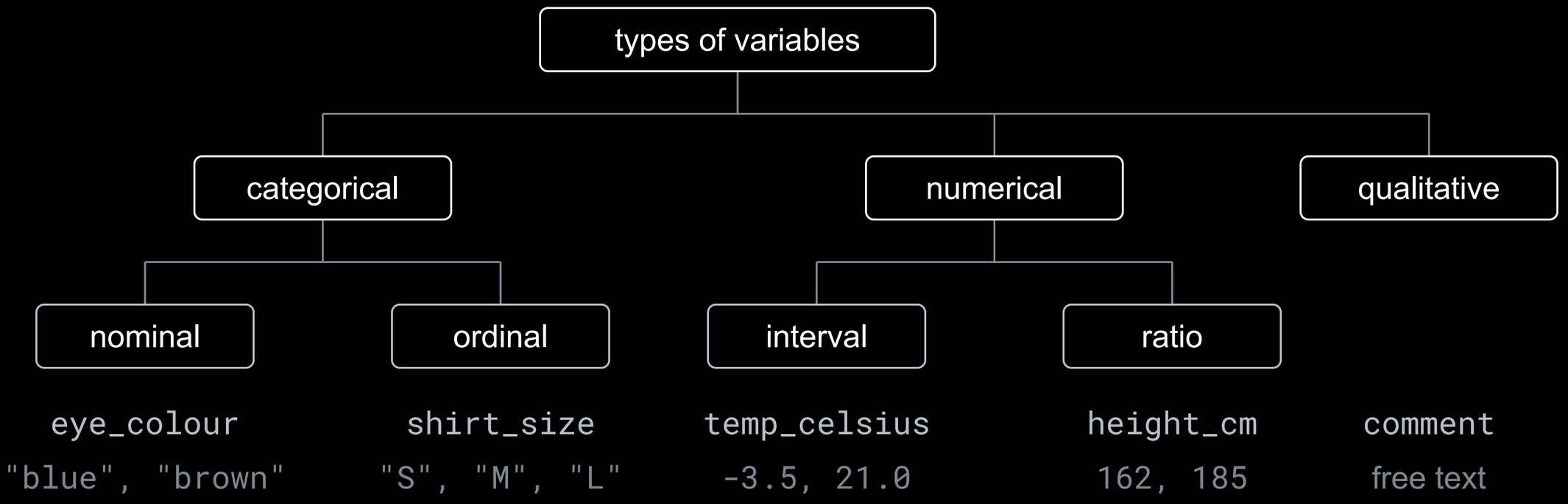


after Döring (2023), chapter 8.4

types of variables



after Döring (2023), chapter 8.4



after Döring (2023), chapter 8.4

what each scale allows

	equal or not	order	differences	ratios	example
nominal	✓	—	—	—	`eye_colour`

each scale keeps everything the one above it allows and adds one more question you may ask.

what each scale allows

	equal or not	order	differences	ratios	example
nominal	✓	—	—	—	`eye_colour`
ordinal	✓	✓	—	—	`shirt_size`

each scale keeps everything the one above it allows and adds one more question you may ask.

what each scale allows

	equal or not	order	differences	ratios	example
nominal	✓	—	—	—	`eye_colour`
ordinal	✓	✓	—	—	`shirt_size`
interval	✓	✓	✓	—	`temp_celsius`

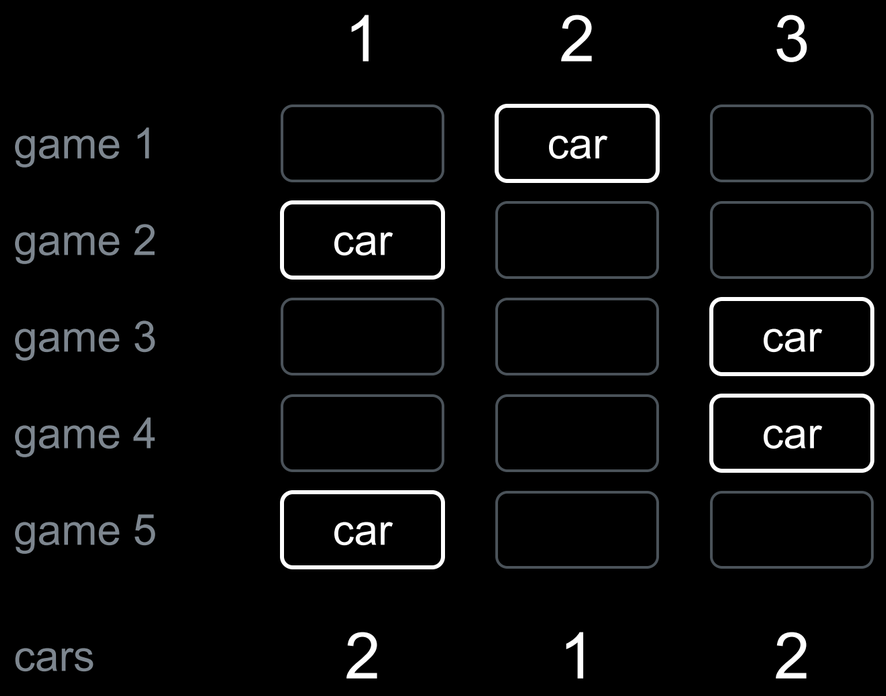
each scale keeps everything the one above it allows and adds one more question you may ask.

what each scale allows

	equal or not	order	differences	ratios	example
nominal	✓	—	—	—	`eye_colour`
ordinal	✓	✓	—	—	`shirt_size`
interval	✓	✓	✓	—	`temp_celsius`
ratio	✓	✓	✓	✓	`height_cm`

each scale keeps everything the one above it allows and adds one more question you may ask.

door 3 is not three times door 1




```
mean(prize_door) ⟶ 2
```

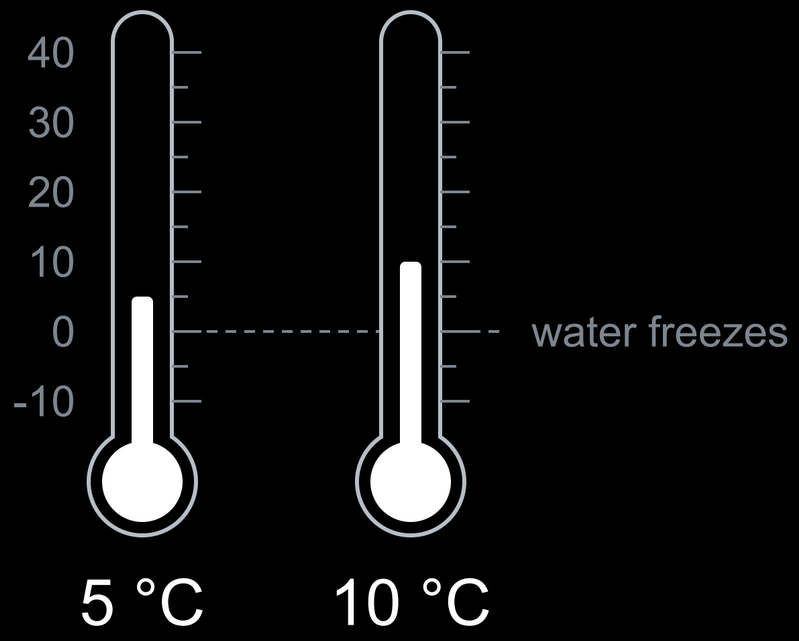
```
table(prize_door)
```

```
1 2 3
2 1 2
```

a number used as a name is still a name, so count it instead of averaging it.

ten degrees are not twice as warm as five

temperature



10 is twice 5

price

1 €

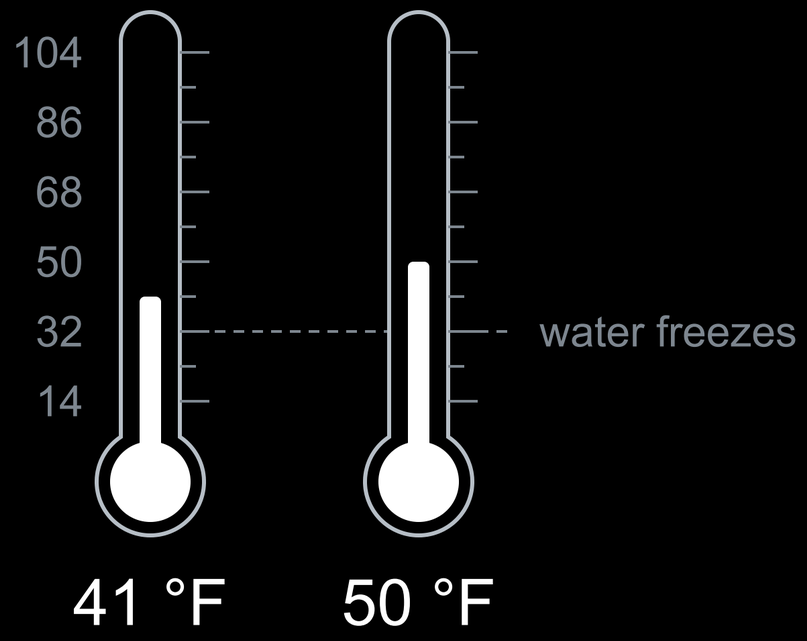
2 €

2 is twice 1

ratios need a true zero, which a price has and a temperature in celsius does not.

ten degrees are not twice as warm as five

temperature



50 is not twice 41

price

100 ct

200 ct

200 is still twice 100

ratios need a true zero, which a price has and a temperature in celsius does not.

is a rating scale ordinal or interval?



as an ordinal scale

```
median(organic_share)  # 2
```

as an interval scale

```
mean(organic_share)  # 2.366667
```

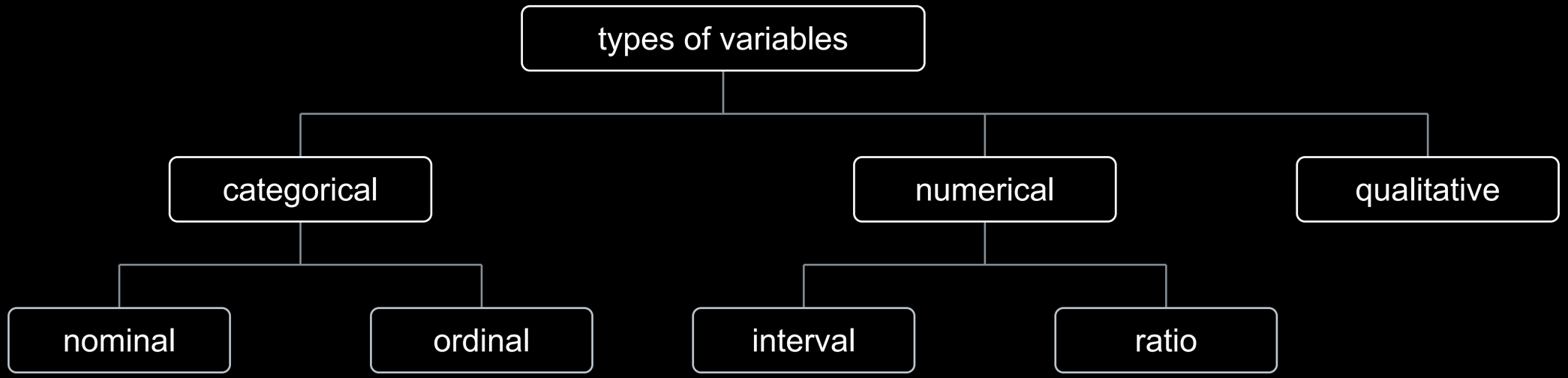
if in doubt, compute both, and say which one you report.

part 3

scales in R

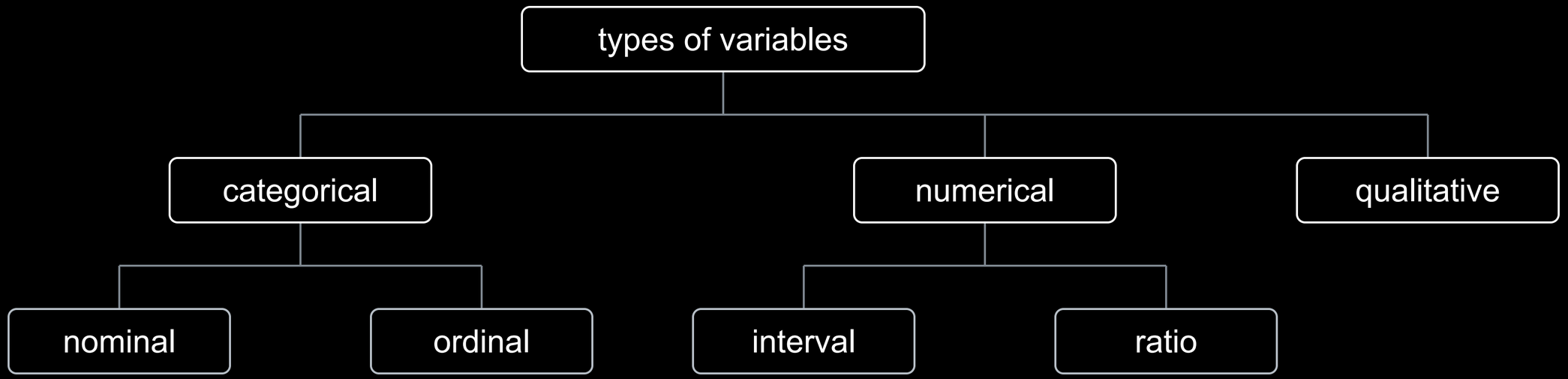
from a scale to a data type

every scale has its type in R



a factor is how R remembers that a column is categorical.

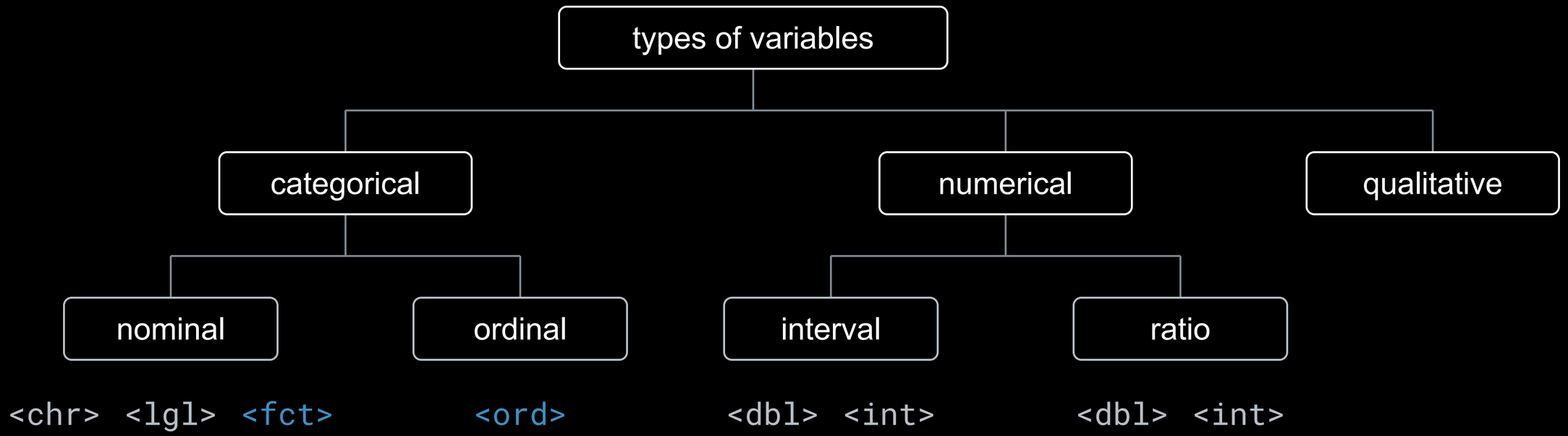
every scale has its type in R



`<chr>` `<lgl>` `<fct>` `<ord>`

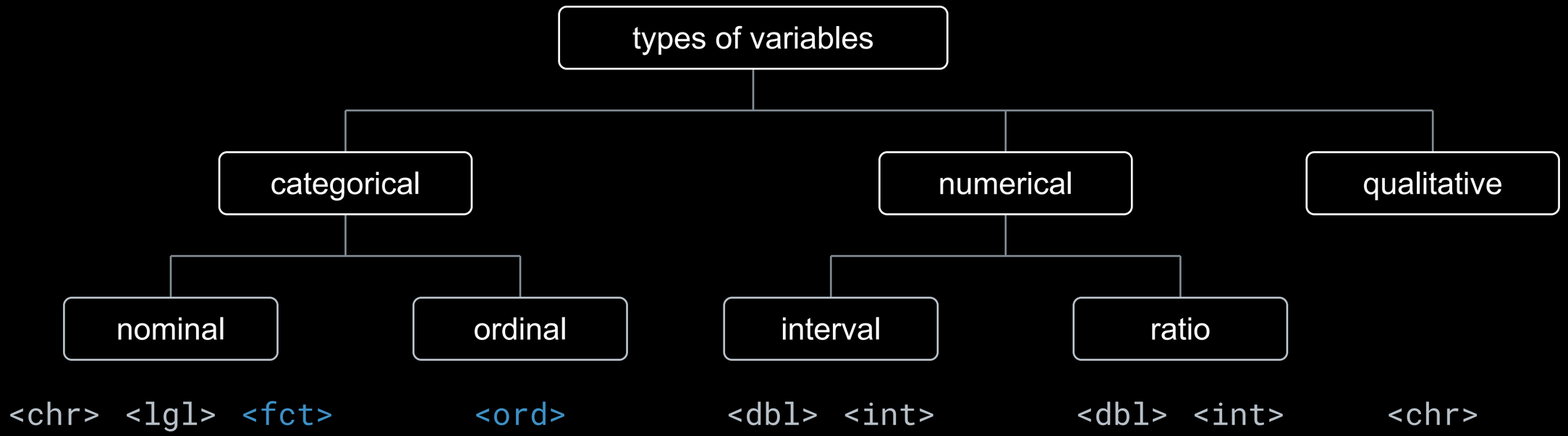
a factor is how R remembers that a column is categorical.

every scale has its type in R



a factor is how R remembers that a column is categorical.

every scale has its type in R



a factor is how R remembers that a column is categorical.

telling R what a column is

```
library(dplyr)

monty_hall <- monty_hall |>
  mutate(decision = factor(decision, levels = c("stay", "switch")))

survey <- survey |>
  mutate(organic_share = factor(organic_share, levels = 1:5, ordered = TRUE))
```

`factor()` marks a column as nominal, `ordered = TRUE` as ordinal, and numbers stay double or integer.

what you may compute

	count	location	spread	ratios	picture
nominal	`count()`				bar chart

the scale tells you which function makes sense, while R will run all of them anyway.

what you may compute

	count	location	spread	ratios	picture
nominal	`count()`				bar chart
ordinal	`count()`	`median()`	`quantile()`		ordered bars

the scale tells you which function makes sense, while R will run all of them anyway.

what you may compute

	count	location		spread		ratios	picture
nominal	`count()`						bar chart
ordinal	`count()`	`median()`		`quantile()`			ordered bars
interval	`count()`	`median()`	`mean()`	`quantile()`	`sd()`		histogram

the scale tells you which function makes sense, while R will run all of them anyway.

what you may compute

	count	location		spread		ratios	picture
nominal	`count()`						bar chart
ordinal	`count()`	`median()`		`quantile()`			ordered bars
interval	`count()`	`median()`	`mean()`	`quantile()`	`sd()`		histogram
ratio	`count()`	`median()`	`mean()`	`quantile()`	`sd()`	`new / old`	histogram

the scale tells you which function makes sense, while R will run all of them anyway.

the scale of a variable decides what you may compute with it, and R's type should say so.

the scale of a variable decides what you may compute with it, and R's type should say so.

to read on: Döring (2023), Forschungsmethoden und Evaluation, chapter 8.4, and R for Data Science, chapter "factors"