

Data Basics

Slides developed by Mine Çetinkaya-Rundel of OpenIntro
The slides may be copied, edited, and/or shared via the [CC BY-SA license](#)

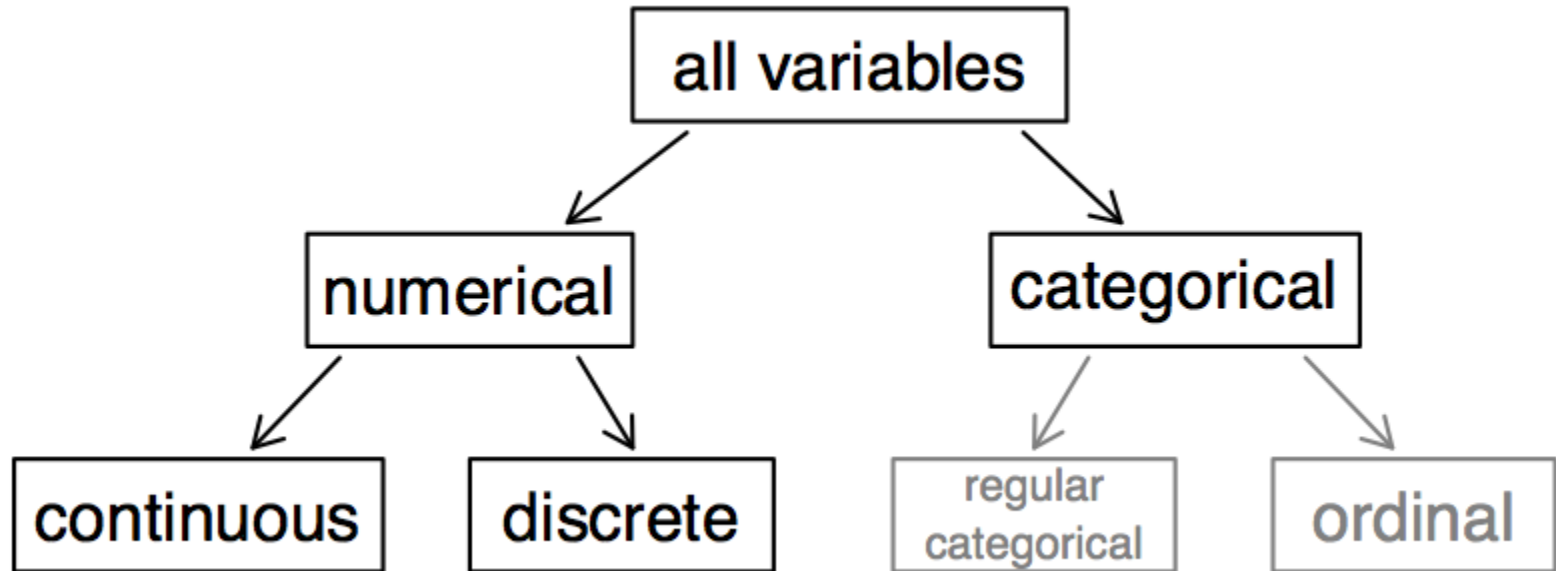
Data matrix

Data collected on students in a statistics class on a variety of variables:

variable
↓

Stu.	gender	intro_extra	...	dread	
1	male	extravert	...	3	
2	female	extravert	...	2	
3	female	introvert	...	4	←
4	female	extravert	...	2	<i>observation</i>
⋮	⋮	⋮	⋮	⋮	
86	male	extravert	...	3	

Types of variables



Types of variables (cont.)

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

gender - categorical

sleep - numerical, continuous

bedtime - categorical, ordinal

countries - numerical, discrete

dread - categorical, ordinal (could also be used as numerical)

Practice

What type of variable is a telephone area code?

- (a) numerical, continuous
- (b) numerical, discrete
- (c) categorical
- (d) categorical, ordinal

Practice

What type of variable is a telephone area code?

(a) numerical, continuous

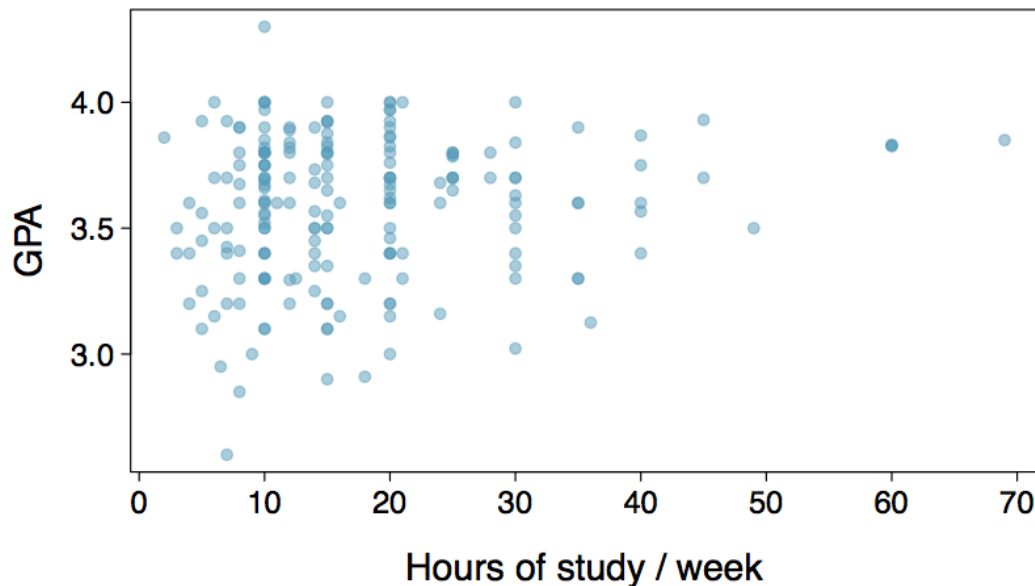
(b) numerical, discrete

(c) *categorical*

(d) categorical, ordinal

Relationships among variables

Does there appear to be a relationship between number of alcoholic drinks consumed per week and age at first alcohol consumption?

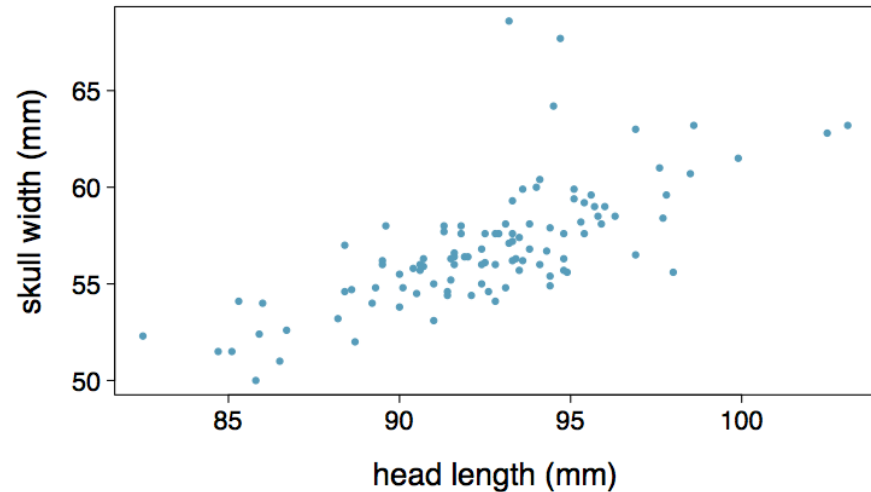


Can you spot anything unusual about any of the data points?

There is one student with $\text{GPA} > 4.0$, this is likely a data error.

Practice

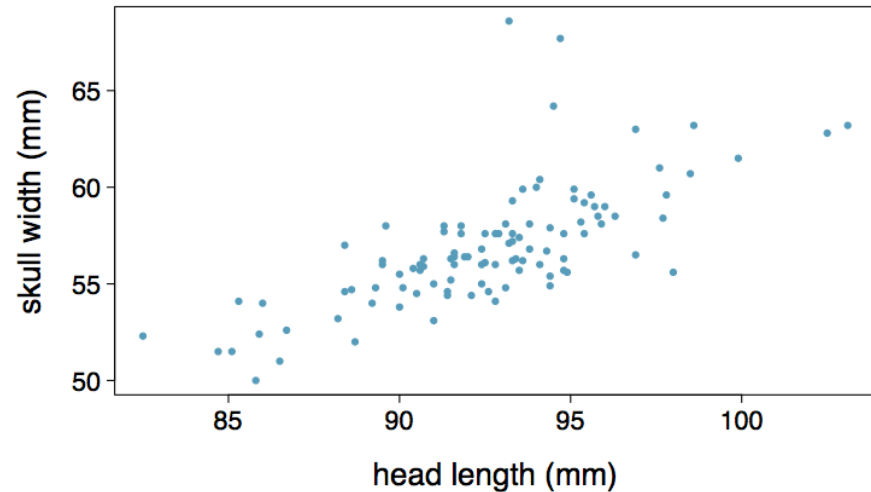
Based on the scatterplot on the right, which of the following statements is correct about the head and skull lengths of possums?



- (a) There is no relationship between head length and skull width, i.e. the variables are independent.
- (b) Head length and skull width are positively associated.
- (c) Skull width and head length are negatively associated.
- (d) A longer head causes the skull to be wider.
- (e) A wider skull causes the head to be longer.

Practice

Based on the scatterplot on the right, which of the following statements is correct about the head and skull lengths of possums?



- (a) There is no relationship between head length and skull width, i.e. the variables are independent.
- (b) *Head length and skull width are positively associated.***
- (c) Skull width and head length are negatively associated.
- (d) A longer head causes the skull to be wider.
- (e) A wider skull causes the head to be longer.

Associated vs. independent

- When two variables show some connection with one another, they are called **associated** variables.
 - Associated variables can also be called **dependent** variables and vice-versa.
- If two variables are not associated, i.e. there is no evident connection between the two, then they are said to be **independent**.