

What r doesn't tell you

What r doesn't tell you

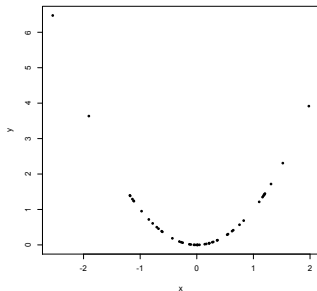
Association is not causation.

What r doesn't tell you

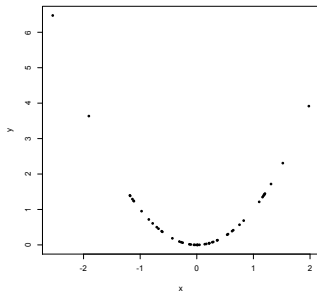
Association is not causation.

If two variables have a non-zero correlation, then they are related to each other in some way, but that doesn't mean that one causes the other.

r measures *linear* association

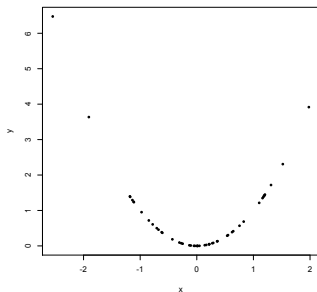


r measures *linear* association



These two variables are clearly related, but r is close to 0.

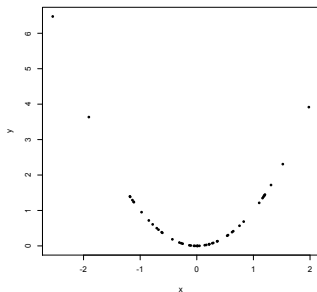
r measures *linear* association



These two variables are clearly related, but r is close to 0.

r measures **linear association**. Don't use it if the scatter diagram is non-linear.

r measures *linear* association



These two variables are clearly related, but r is close to 0.

r measures **linear association**. Don't use it if the scatter diagram is non-linear.

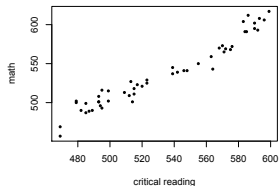
- **correlated:** **linearly** related

Watch out for outliers

Even one outlier can have a noticeable effect on r .

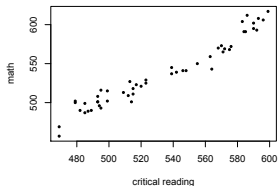
Ecological correlation

Average SAT reading and math scores for each state and D.C., 2011



Ecological correlation

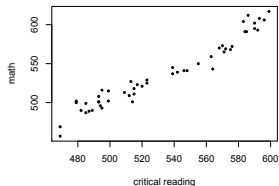
Average SAT reading and math scores for each state and D.C., 2011



$$r = 0.97$$

Ecological correlation

Average SAT reading and math scores for each state and D.C., 2011

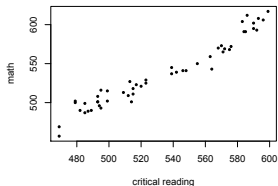


$$r = 0.97$$

Is the correlation between reading and math SAT scores really that high?

Ecological correlation

Average SAT reading and math scores for each state and D.C., 2011



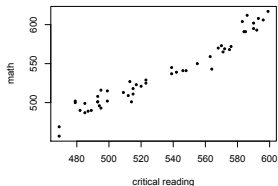
$$r = 0.97$$

Is the correlation between reading and math SAT scores really that high?

- States don't take exams.

Ecological correlation

Average SAT reading and math scores for each state and D.C., 2011



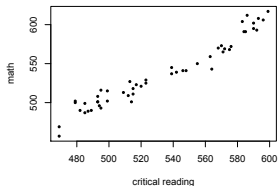
$$r = 0.97$$

Is the correlation between reading and math SAT scores really that high?

- States don't take exams.
- Students do.

Ecological correlation

Average SAT reading and math scores for each state and D.C., 2011



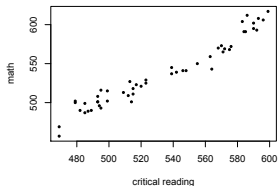
$$r = 0.97$$

Is the correlation between reading and math SAT scores really that high?

- **States don't take exams.**
- **Students do.**
- **Students in each state are more variable than the state average.**

Ecological correlation

Average SAT reading and math scores for each state and D.C., 2011



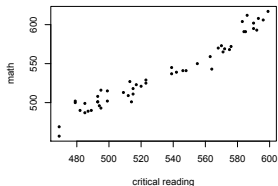
$$r = 0.97$$

Is the correlation between reading and math SAT scores really that high?

- States don't take exams.
- Students do.
- Students in each state are more variable than the state average.
- The correlation between the reading and math scores of all the students is quite a bit lower than 0.97.

Ecological correlation

Average SAT reading and math scores for each state and D.C., 2011



$$r = 0.97$$

Is the correlation between reading and math SAT scores really that high?

- States don't take exams.
- Students do.
- Students in each state are more variable than the state average.
- The correlation between the reading and math scores of all the students is quite a bit lower than 0.97.
- Replacing students by averages can artificially increase clustering.