

## Two-variables statistics:

- **Explanatory** variable (sometimes called “independent”)
- **Response** variable (“dependent”)

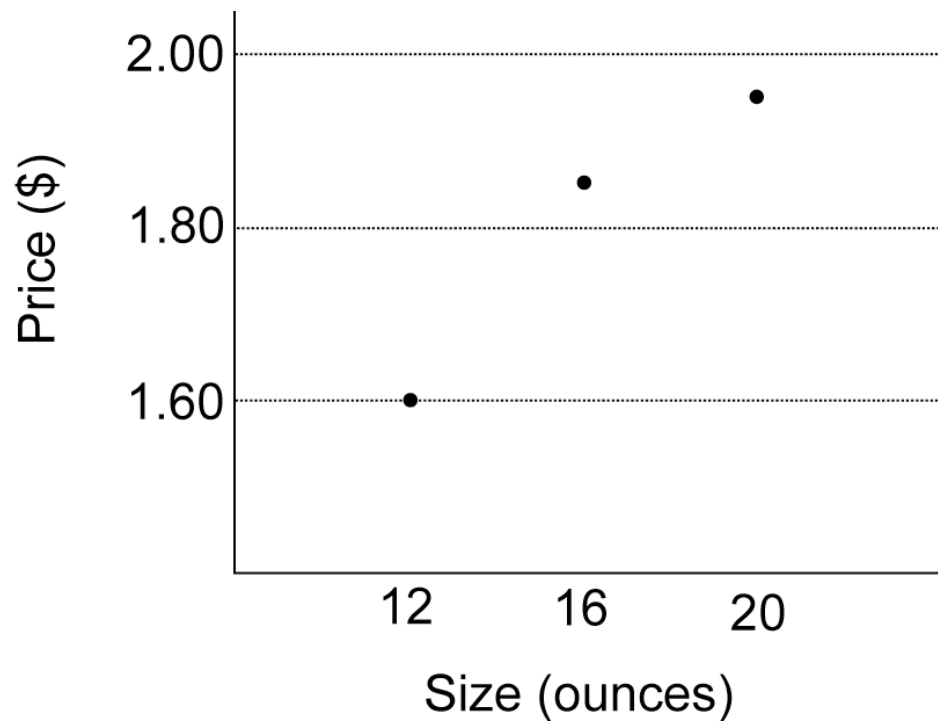
### Examples:

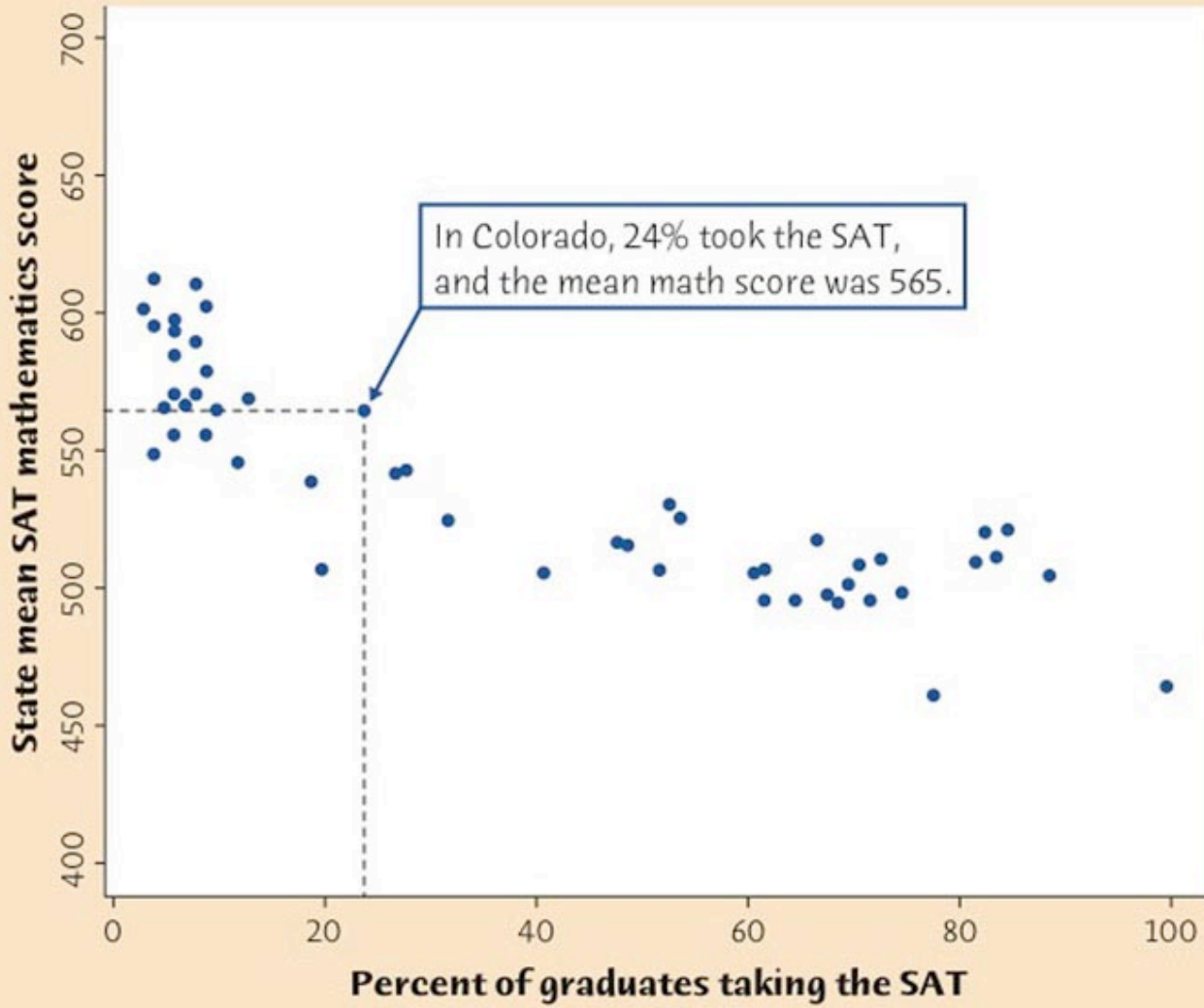
- Rainfall levels (explanatory)/agricultural output (response)
- Rainfall levels (explanatory)/food prices (response)
- Education level (explanatory)/income (response)
- Mean driving speed (explanatory)/gas mileage (response)
- Math SAT score (?)/Verbal SAT score (?)

We display the relationship between two quantitative variables with a **scatterplot**.

## Example: Starbucks coffee:

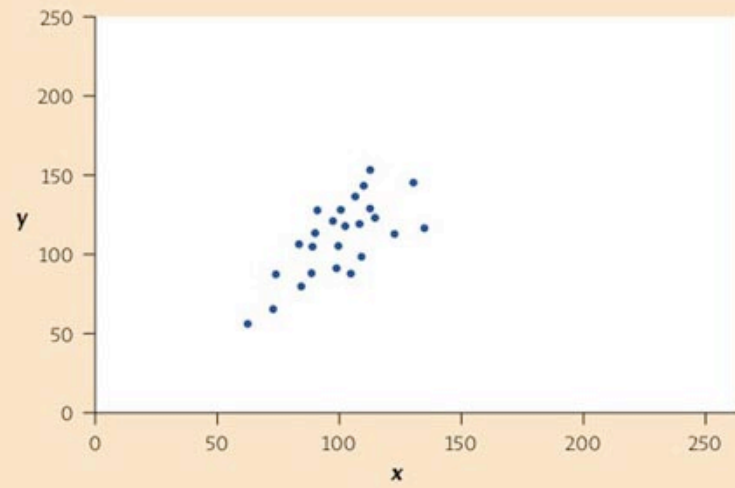
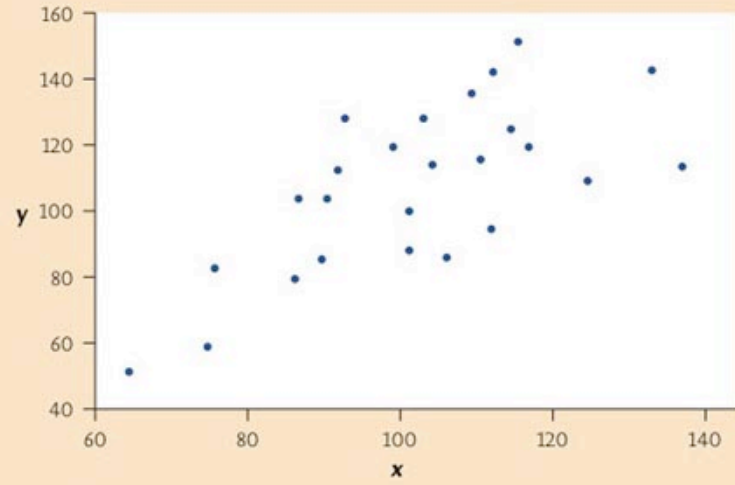
	Size (ounces):	Price (\$):
Tall	12	1.60
Grande	16	1.85
Venti	20	1.95



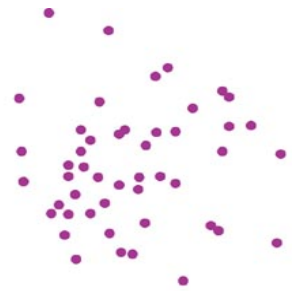


## About scatterplots:

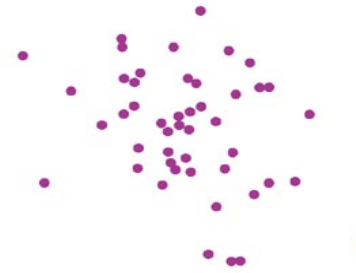
- Always put explanatory on the x-axis, response on the y-axis
- **Direction:** **positive** or **negative** association?
- **Form:** **linear** or **curved**?
- **Strength:** how closely do the points follow a clear form?



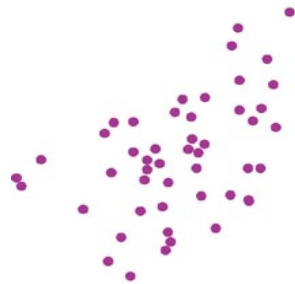
To measure the strength of a linear relationship, look at the **correlation coefficient  $r$** .



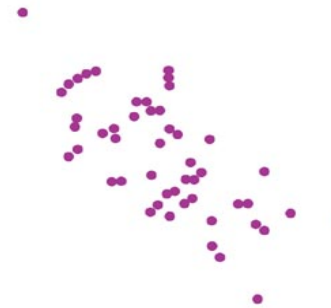
Correlation  $r = 0$



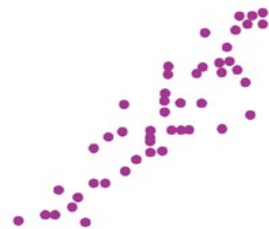
Correlation  $r = -0.3$



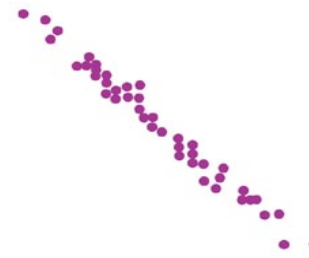
Correlation  $r = 0.5$



Correlation  $r = -0.7$



Correlation  $r = 0.9$



Correlation  $r = -0.99$

## Facts about $r$ :

1.  $r$  is always between  $-1$  and  $1$ .
  - Close to  $-1$  or  $1$   $\rightarrow$  strong linear relationship
  - Close to  $0$   $\rightarrow$  weak linear relationship
2.  $r$  positive  $\rightarrow$  positive association  
 $r$  negative  $\rightarrow$  negative association
3.  $r$  does not change if you change the units of  $x$  or  $y$ .  
 $r$  itself does not have units.

The formula for  $r$  ... (don't panic!)

Explanatory:	$x_1$	$x_2$	...	$x_n$	$\longrightarrow \bar{x}, s_x$
Response:	$y_1$	$y_2$	...	$y_n$	$\longrightarrow \bar{y}, s_y$

$$r = \frac{1}{n-1} \left[ \left( \frac{x_1 - \bar{x}}{s_x} \right) \left( \frac{y_1 - \bar{y}}{s_y} \right) + \left( \frac{x_2 - \bar{x}}{s_x} \right) \left( \frac{y_2 - \bar{y}}{s_y} \right) + \dots + \left( \frac{x_n - \bar{x}}{s_x} \right) \left( \frac{y_n - \bar{y}}{s_y} \right) \right]$$

or

$$r = \frac{1}{n-1} \left[ \sum \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right) \right]$$

## TI-83 or 84:

0. One time preliminary step:  
**CATALOG** (2nd+'0') → **DiagnosticOn** (Enter twice)
1. **STAT** → **EDIT** → **1:Edit**,  
enter explanatory in  $L_1$ , response in  $L_2$
2. **STAT** → **CALC** → **8:LinReg(a+bx)** (Enter)

### Quick practice Example: Starbucks data

$L_1$ :	12	16	20
$L_2$ :	1.60	1.85	1.95

**Answer:**  $r = .970725$

## More facts about $r$ :

1.  $r$  requires both variables to be quantitative
2.  $r$  only measures the strength of linear relationships, not curved ones.
3.  $r$  is strongly affected by outliers.