Interpretation of Correlation

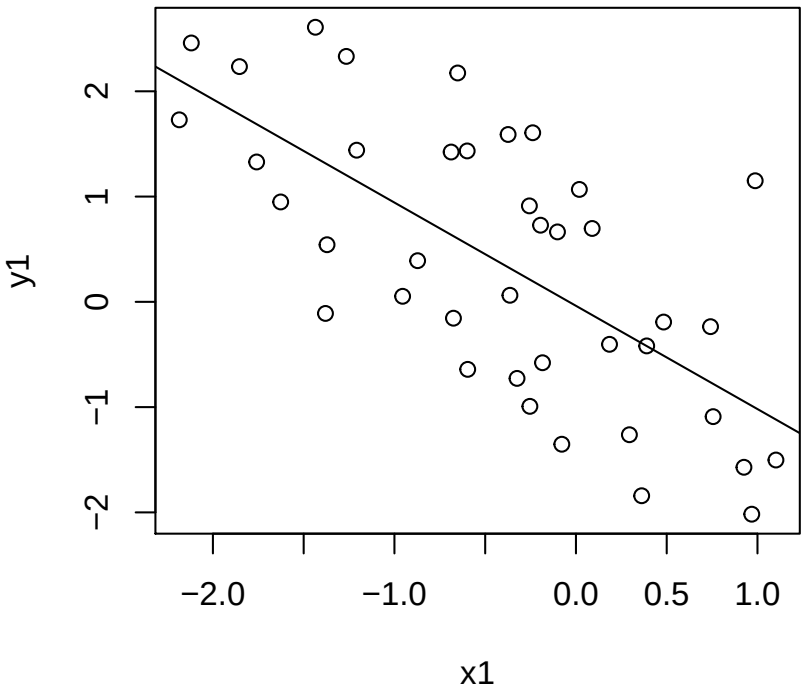
Te Rutherford

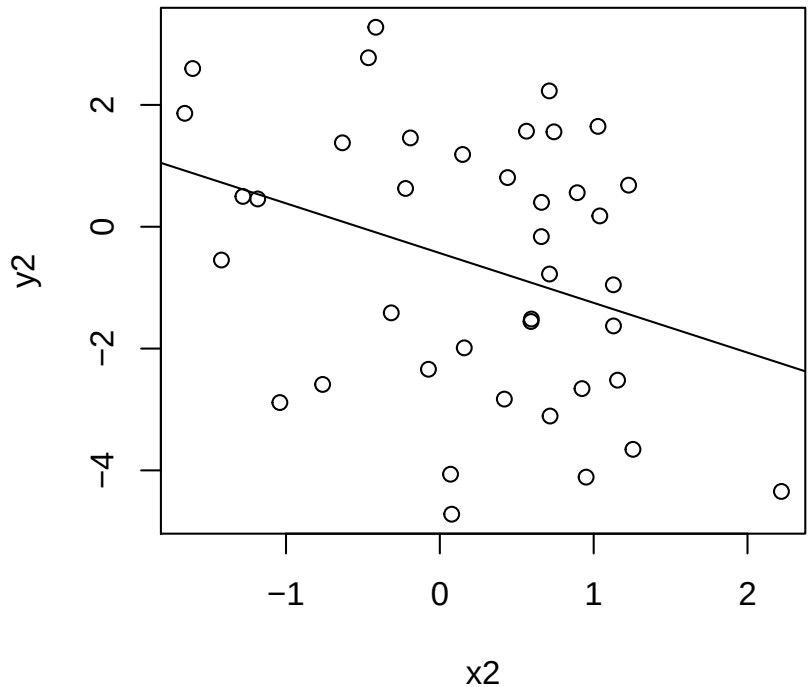
February 26, 2015

Agenda

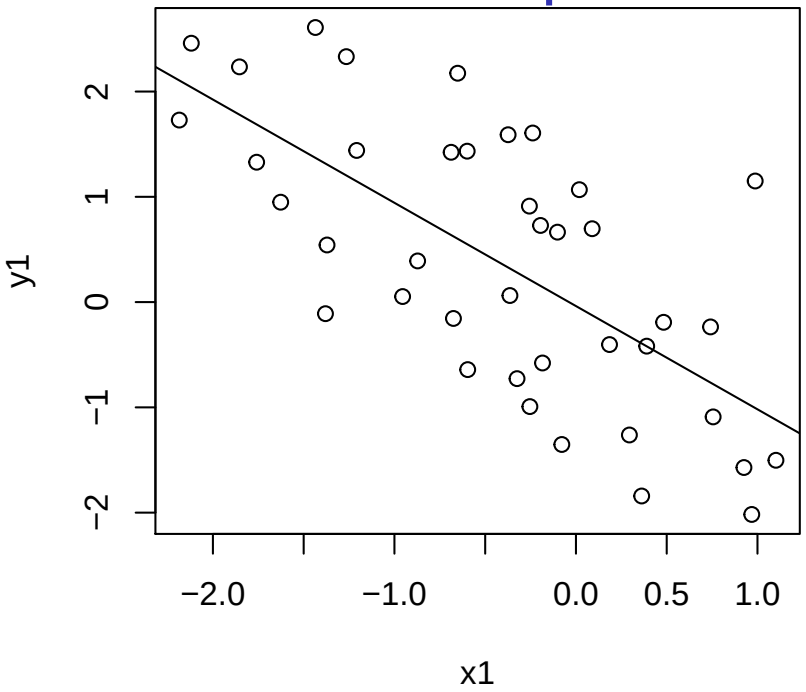
- Last time
  - Positive correlation
  - Negative correlation
  - Scatterplot and linear trend line
- Today
  - Interpretation of correlation
  - Causation

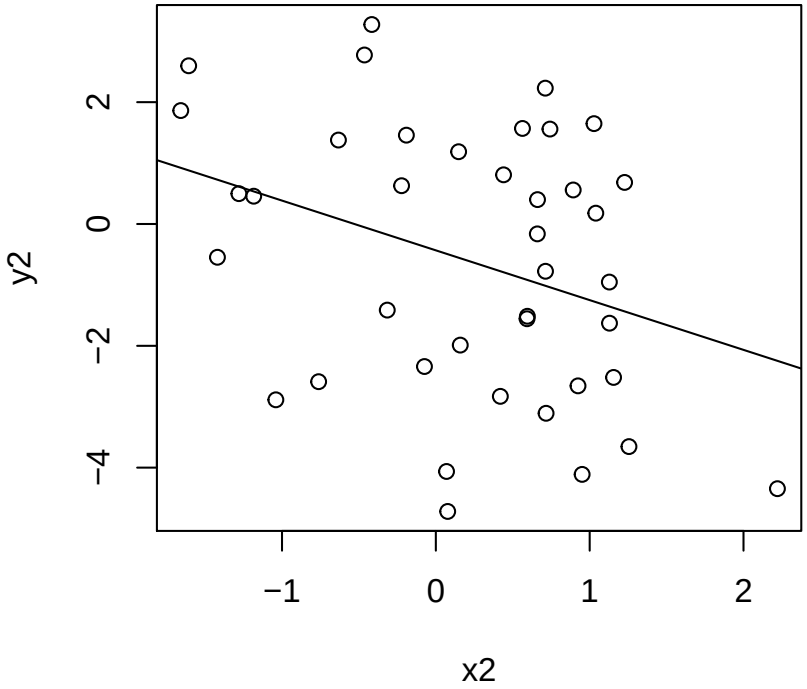
50 shades of scatterplot





50 shades of scatterplot






```
cor(x1, y1)
```

```
## [1] -0.6787
```

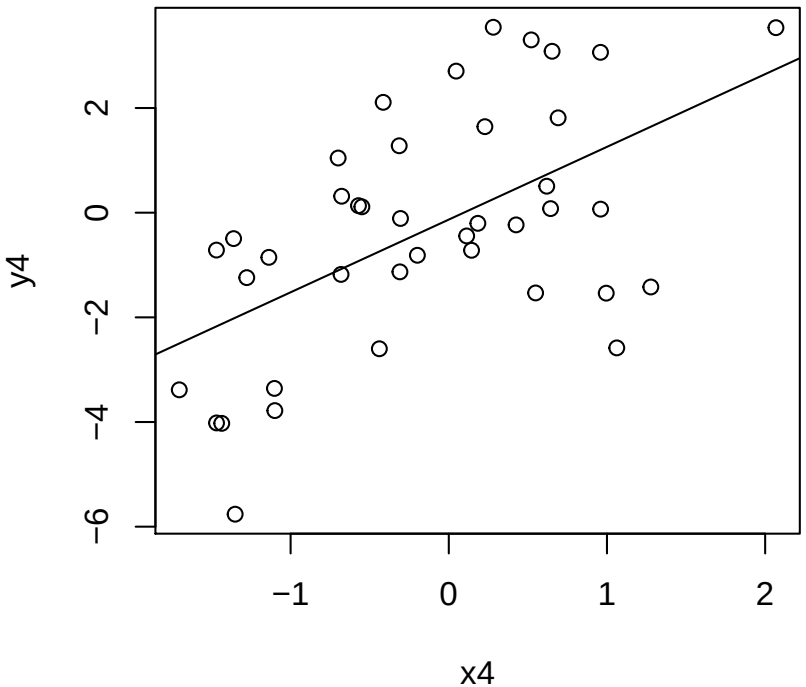
```
cor(x2, y2)
```

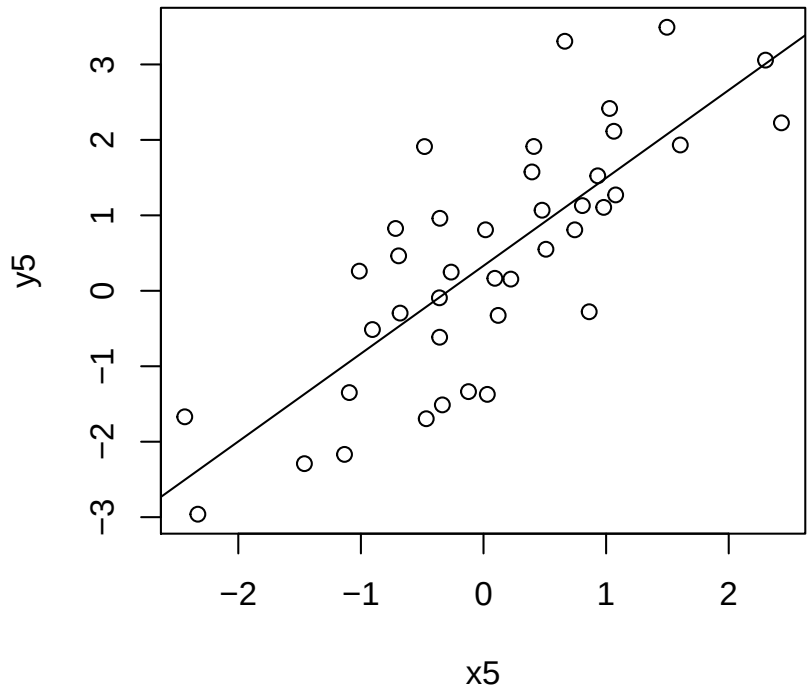
```
## [1] -0.3366
```

Negative Correlation : Interpretation

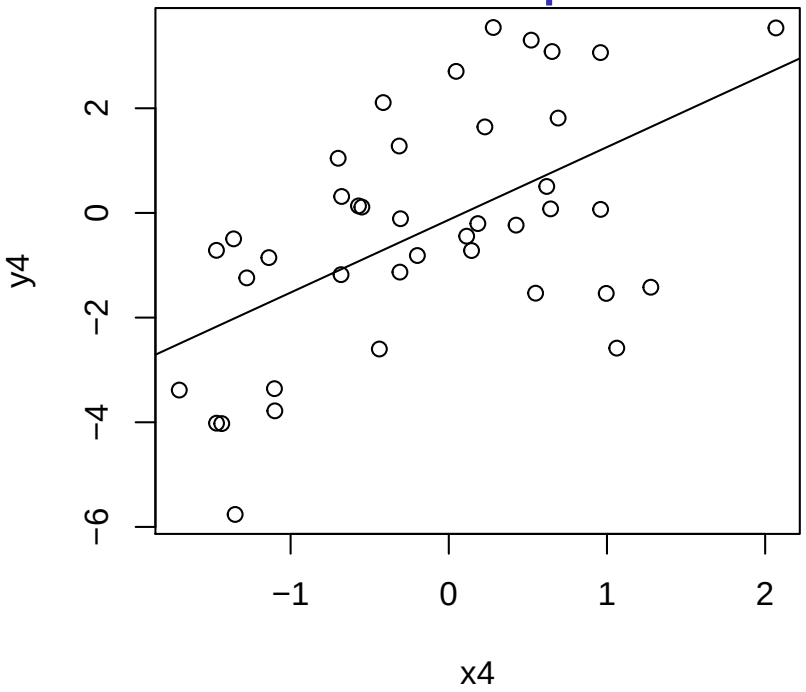
- $-1 < r < -0.7$ x and y are strongly negatively correlated
  - If x is high, then y is low. But there are not many exceptions.
- $-0.7 < r < -0.3$ x and y are weakly negatively correlated
  - If x is high, then y is low. But there are a lot of exceptions.

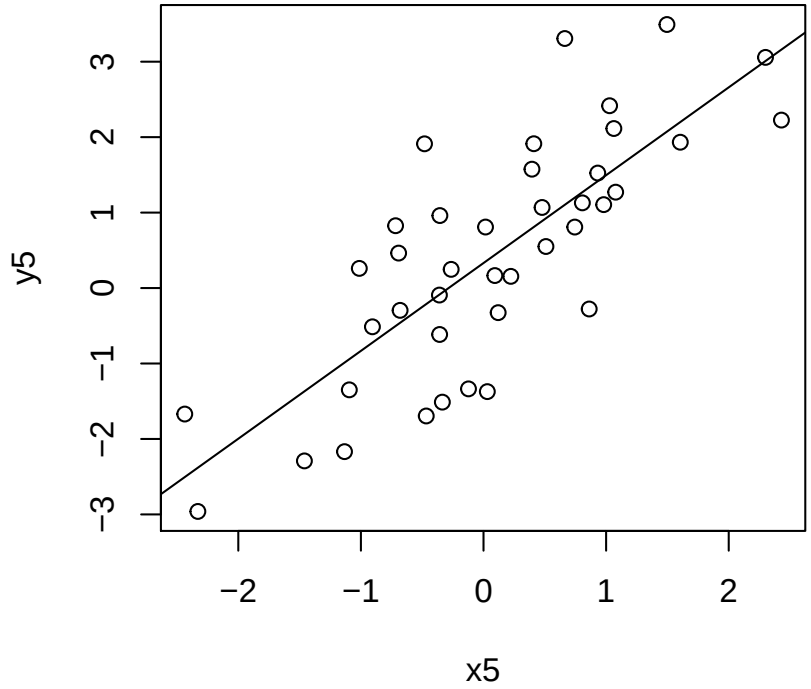
50 shades of scatterplot





50 shades of scatterplot






```
cor(x4, y4)
```

```
## [1] 0.5599
```

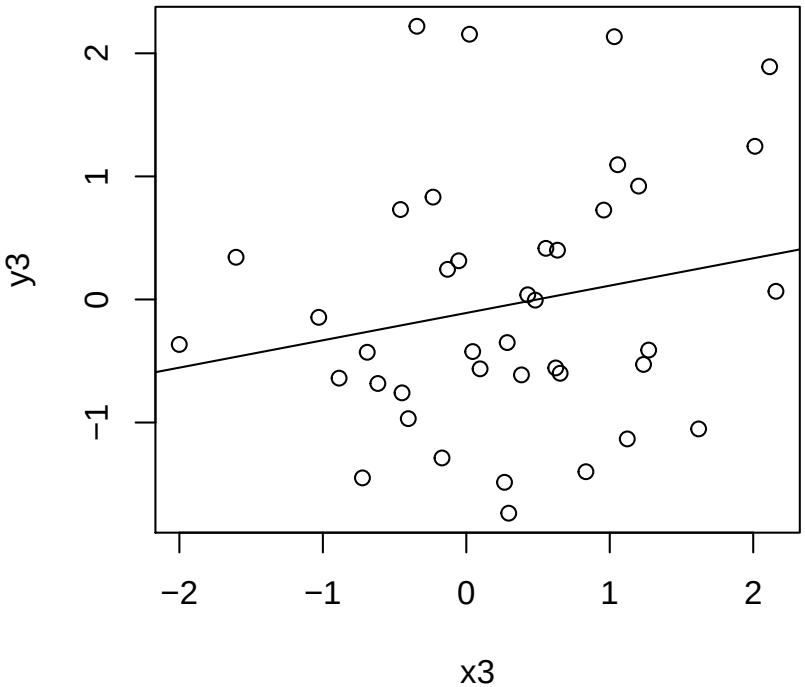
```
cor(x5, y5)
```

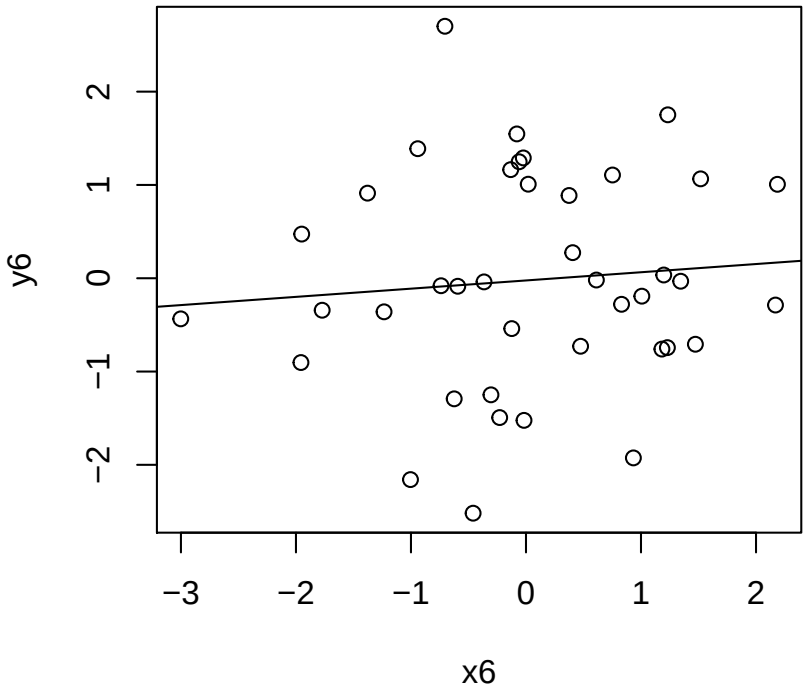
```
## [1] 0.7831
```

Positive Correlation : Interpretation

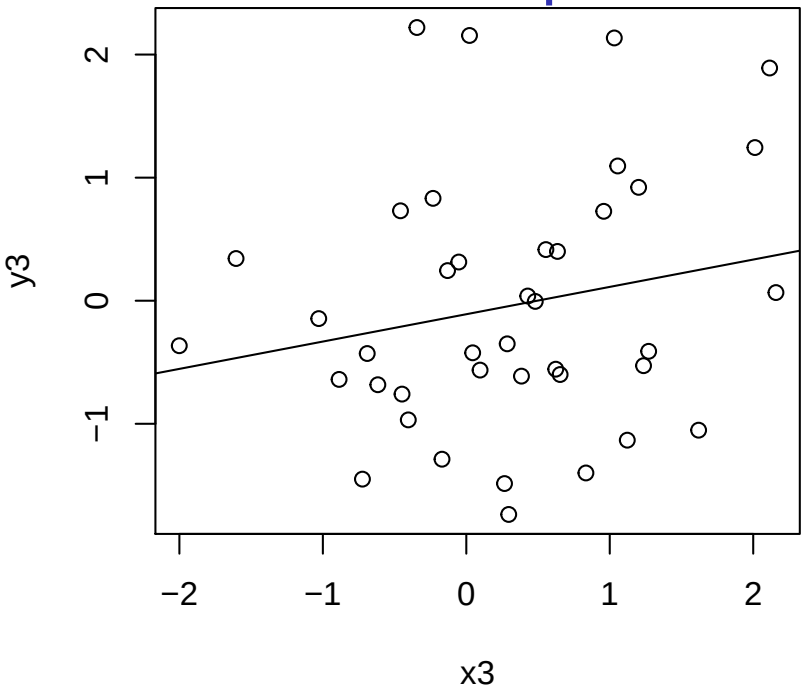
- $0.7 < r < 1$ x and y are strongly positively correlated
  - If x is high, then y is high. But there are not many exceptions.
- $0.3 < r < 0.7$ x and y are weakly positively correlated
  - If x is high, then y is high. But there are a lot of exceptions.

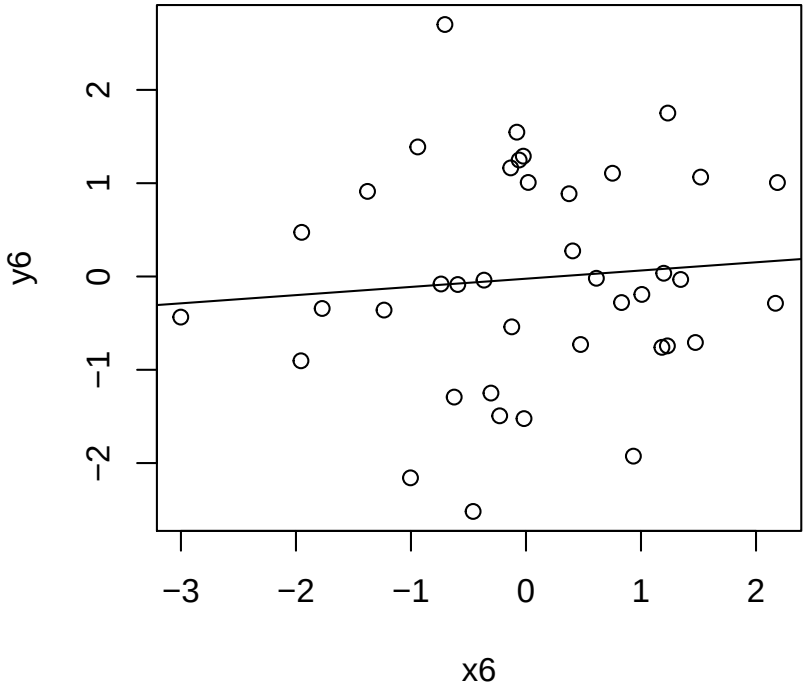
50 shades of scatterplot





50 shades of scatterplot






```
cor(x3, y3)
```

```
## [1] 0.2034
```

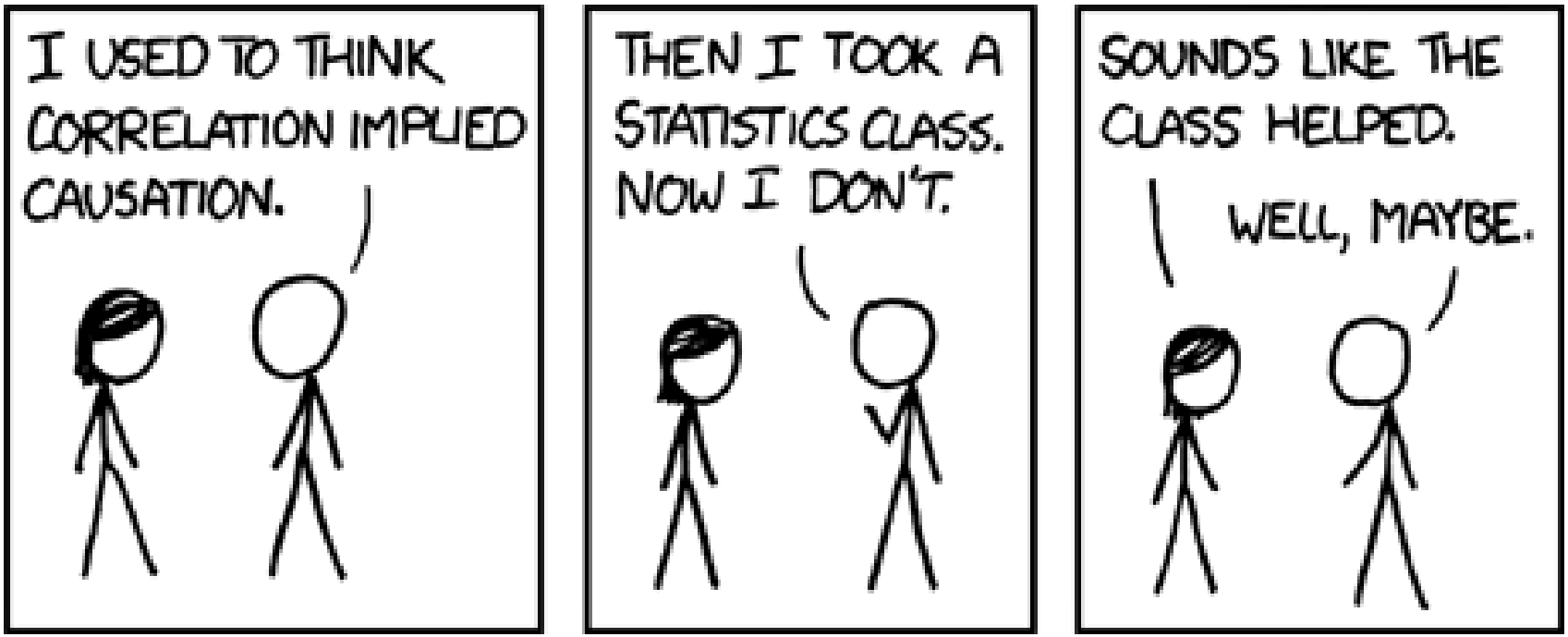
```
cor(x6, y6)
```

```
## [1] 0.08888
```

Lack of Correlation : Interpretation

- $-0.3 < r < 0.3$ x and y are not correlated.
  - There is no relationship between x and y.

Correlation does not always imply causation



- It is a crime not to know this mantra. So let me say it again: correlation does not imply causation.

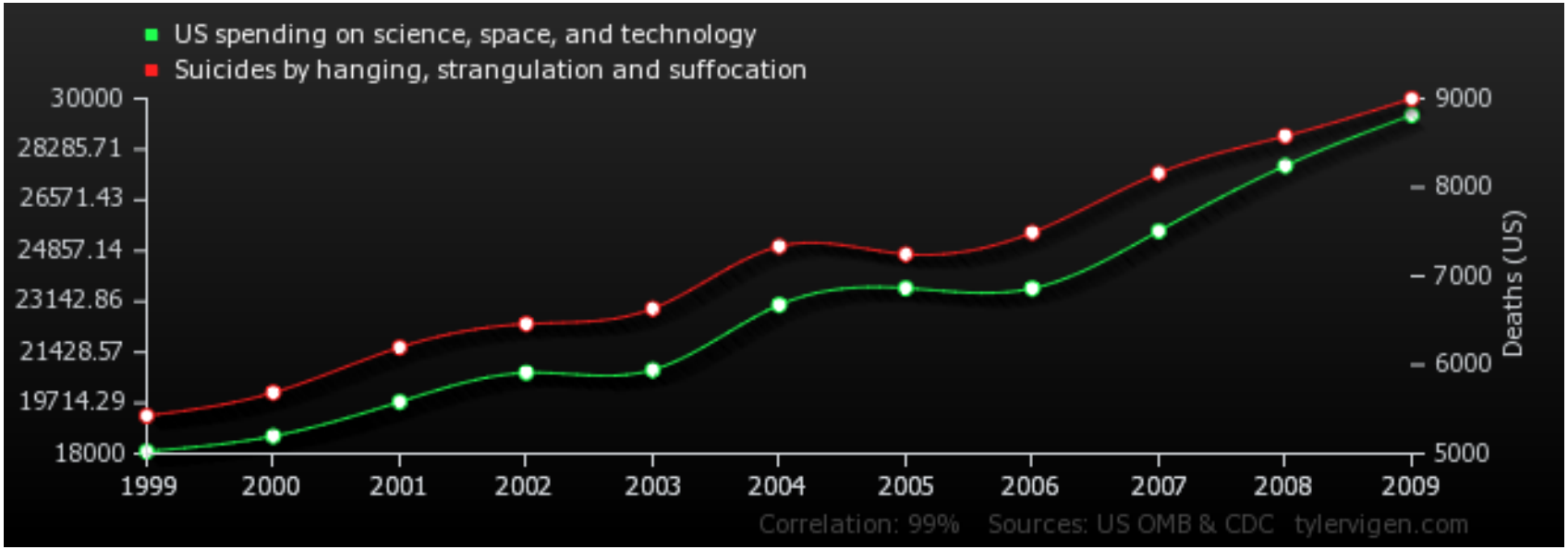
Example

- Waiting duration and the number of people at the stop are negatively correlated.
- We cannot infer that the number of people CAUSES the waiting time to go down.
- Correlation does not imply causation.

Example

- High school GPA and college GPA are positively correlated.
- If we hack into the high school database and change our GPA to be very high, the college GPA won't increase.
- We cannot infer that the high school GPA causes the college GPA to go up or down
- Correlation does not imply causation.

To be even more absurd, look at this correlation.



- We have the correlation, but we cannot infer that the US government has increased the spending on science to cause suicides.
- Or we cannot infer that the US spending in science makes people suicidal.

Other correlation

- The number of donkeys in a state is negatively correlated with the number of PhD graduates.
- Foot sizes are positively correlated with brain sizes.
- High Fructose Corn Syrup consumption is correlated with body fat.
- Household income is correlated with SAT score.