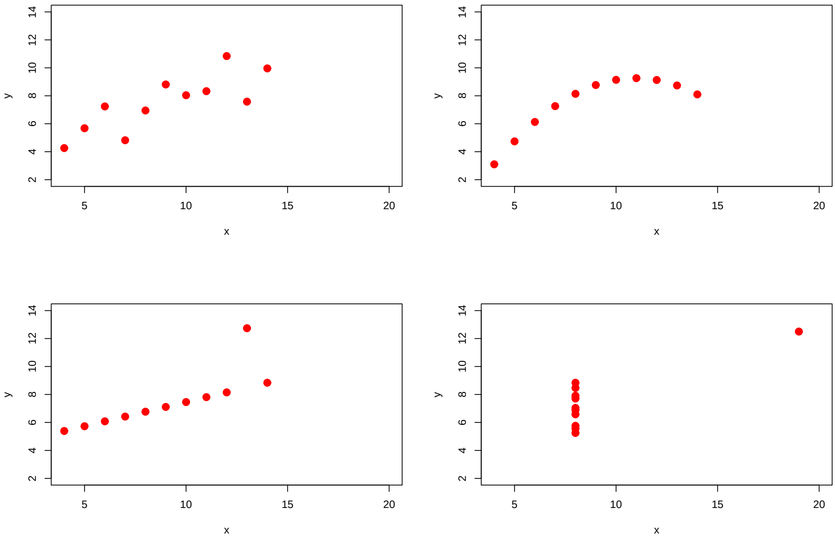
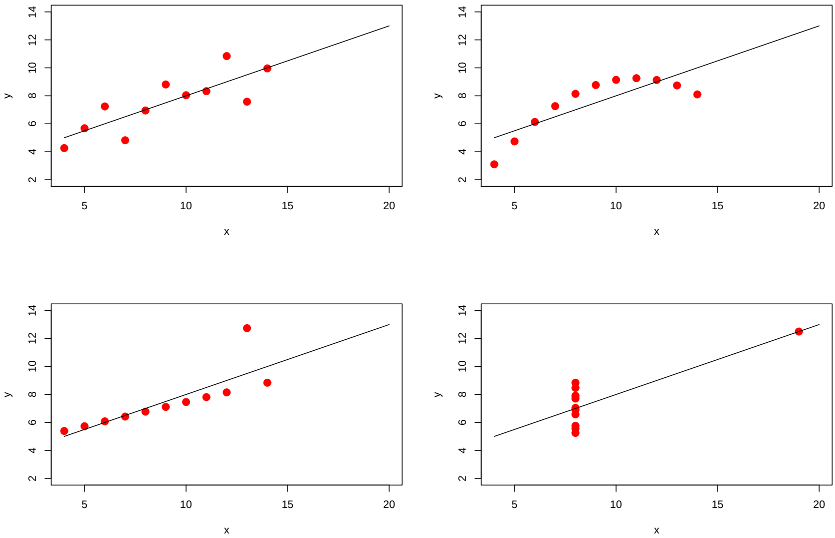
# Residuals: Anscombe's quartet



# Residuals: Anscombe's quartet



## Defining residuals

DATA = FITTED VALUE + RESIDUAL

## Anscombe's quartet

- all four data-sets have:
  - the same best-fitting straight line
  - the same residual sum of squares
- but the residuals tell four different stories.

| | res1 | res2 | res3 | res4 |
|---|---|---|---|---|
| 1 | -0.740 | -1.901 | 0.389 | 0.000 |
| 2 | 0.179 | -0.761 | 0.229 | -0.111 |
| 3 | 1.239 | 0.129 | 0.079 | -1.751 |
| 4 | -1.681 | 0.759 | -0.081 | 0.909 |
| 5 | -0.051 | 1.139 | -0.230 | -1.241 |
| 6 | 1.309 | 1.269 | -0.390 | 1.839 |
| 7 | 0.039 | 1.139 | -0.540 | -0.421 |
| 8 | -0.171 | 0.759 | -0.689 | 1.469 |
| 9 | 1.839 | 0.129 | -0.849 | -1.441 |
| 10 | -1.921 | -0.761 | 3.241 | 0.709 |
| 11 | -0.041 | -1.901 | -1.159 | 0.039 |

**Why are the four sets of residuals different?**

- obvious if you have only one explanatory variable
- less obvious when you have many

## Analysing residuals

- check that their average value is (close to) zero
- plot them against fitted values
- plot them against explanatory variables in the model
- plot them against explanatory variables **not** in the model (for example, residuals against time-order)