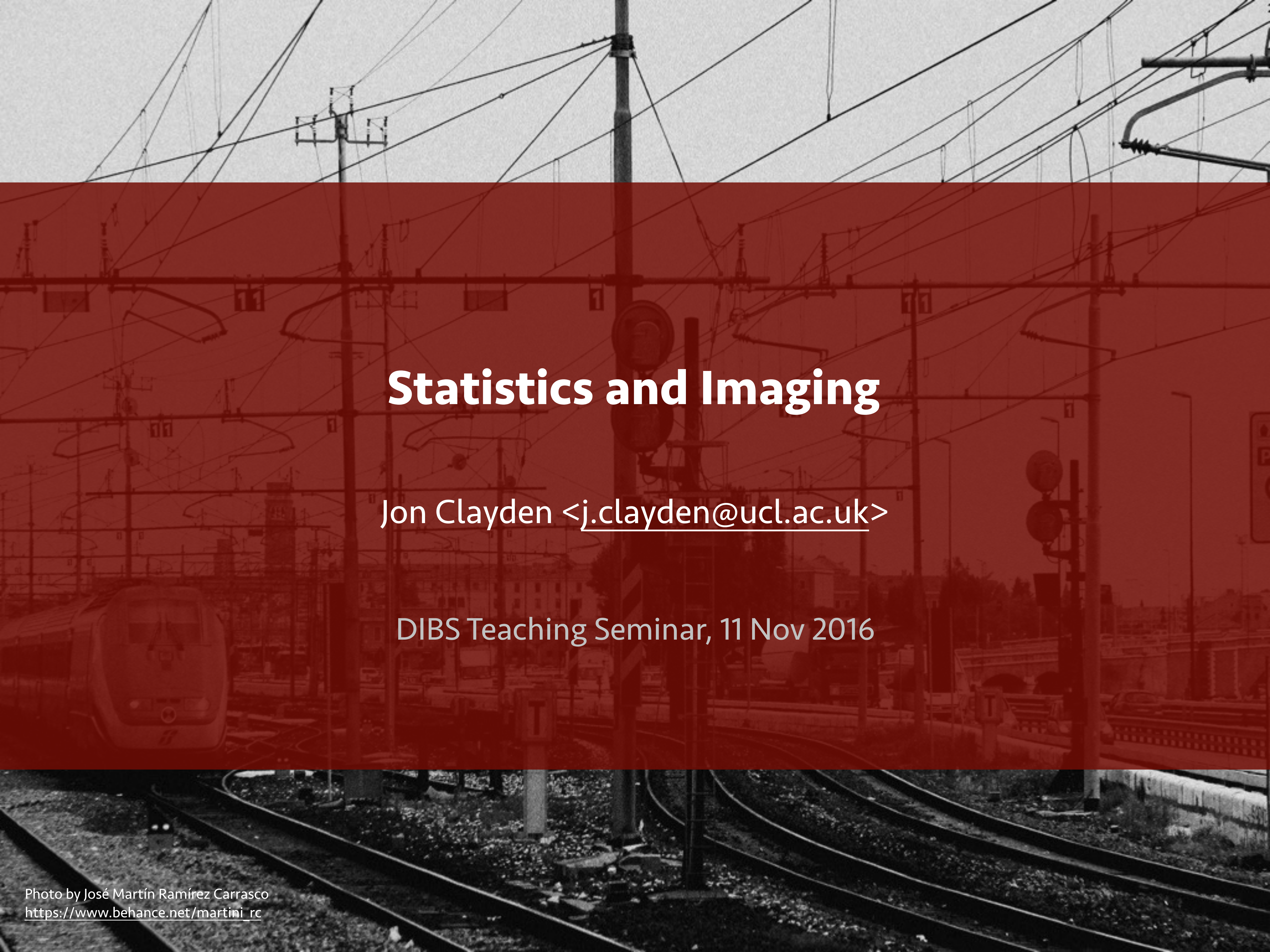


Statistics and Imaging

Jon Clayden <j.clayden@ucl.ac.uk>

DIBS Teaching Seminar, 11 Nov 2016

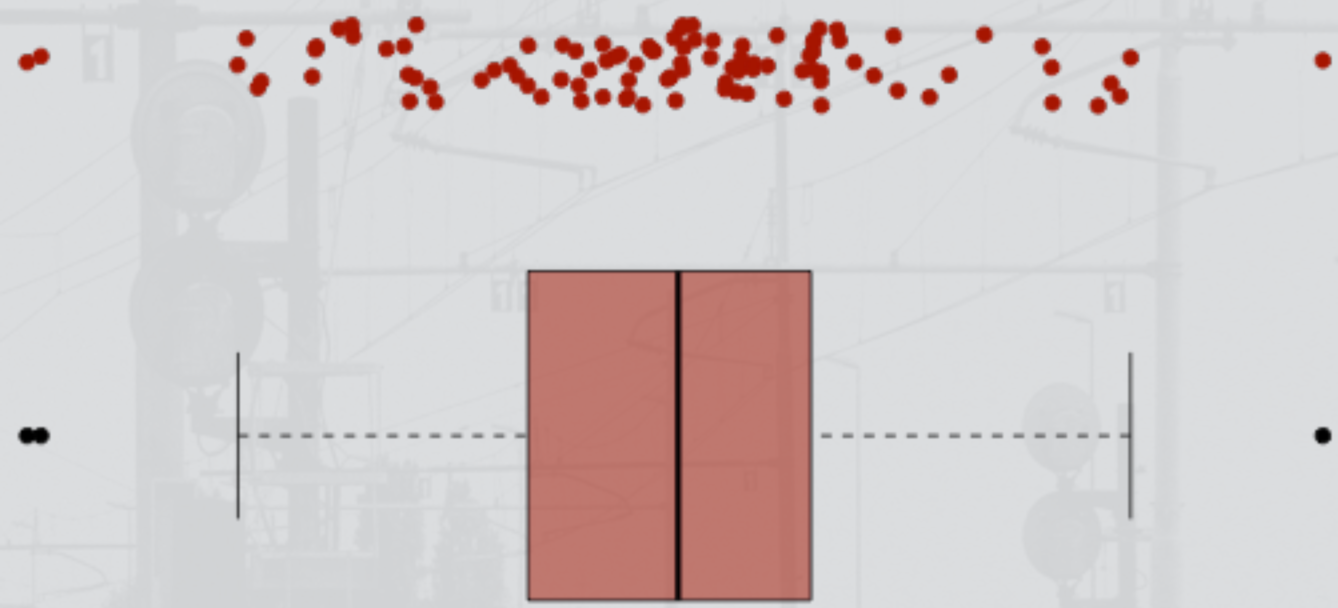
Photo by José Martín Ramírez Carrasco
https://www.behance.net/martini_rc

“Statistics is a subject that many medics find easy, but most statisticians find difficult”

— Stephen Senn (attrib.)

Purposes

- Summarising data, describing features such as central tendency and dispersion



- Making inferences about the population that a given sample was drawn from

Hypothesis testing

- A null hypothesis is a default position (no effect, no difference, no relationship, etc.)
- This is set against an alternative hypothesis, generally the opposite of the null
- A hypothesis test estimates the probability, p, of observing data at least as extreme as the sample, under the assumption that the null is true
- If this p-value is less than a threshold, α, usually 0.05, then the null is rejected and treated as false
- 5% of rejections are therefore expected to be false positives
- The rate at which the null hypothesis is correctly rejected is the power
- NB: Failing to reject the null hypothesis does not constitute strong evidence in support of it

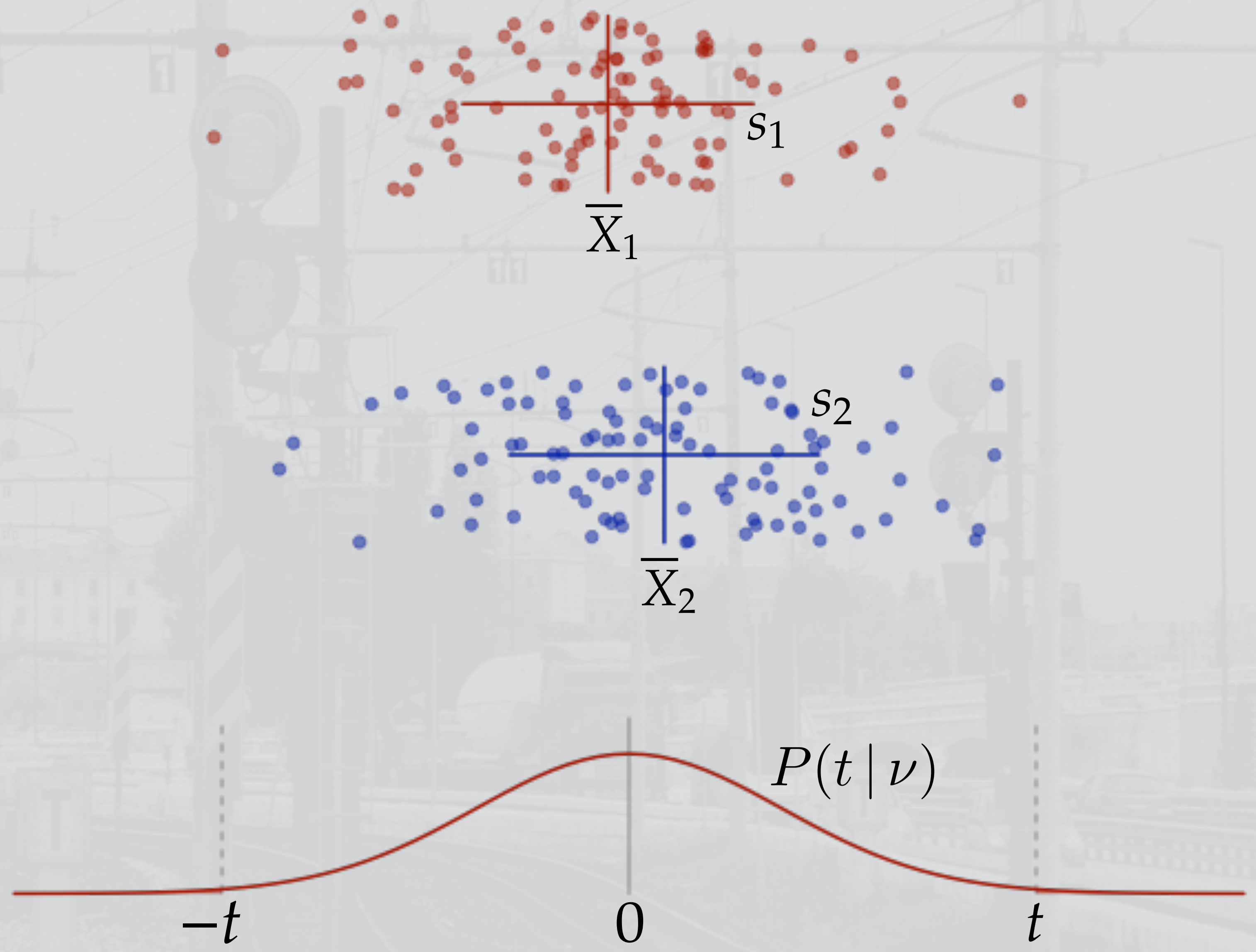
The *t*-test

- A test for a difference in means …
- … which may be of a particular sign (one-tailed) or either sign (two-tailed) …
- … either between two groups of observations (two sample), or one group and a fixed value, often zero (one sample) …
- … which is valid under the assumptions that the groups are approximately normally distributed, independently sampled and (for some implementations) have equal population variance

Anatomy of a test

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2 \left(\frac{1}{n_1 - 1}\right) + \left(\frac{s_2^2}{n_2}\right)^2 \left(\frac{1}{n_2 - 1}\right)}$$



In R

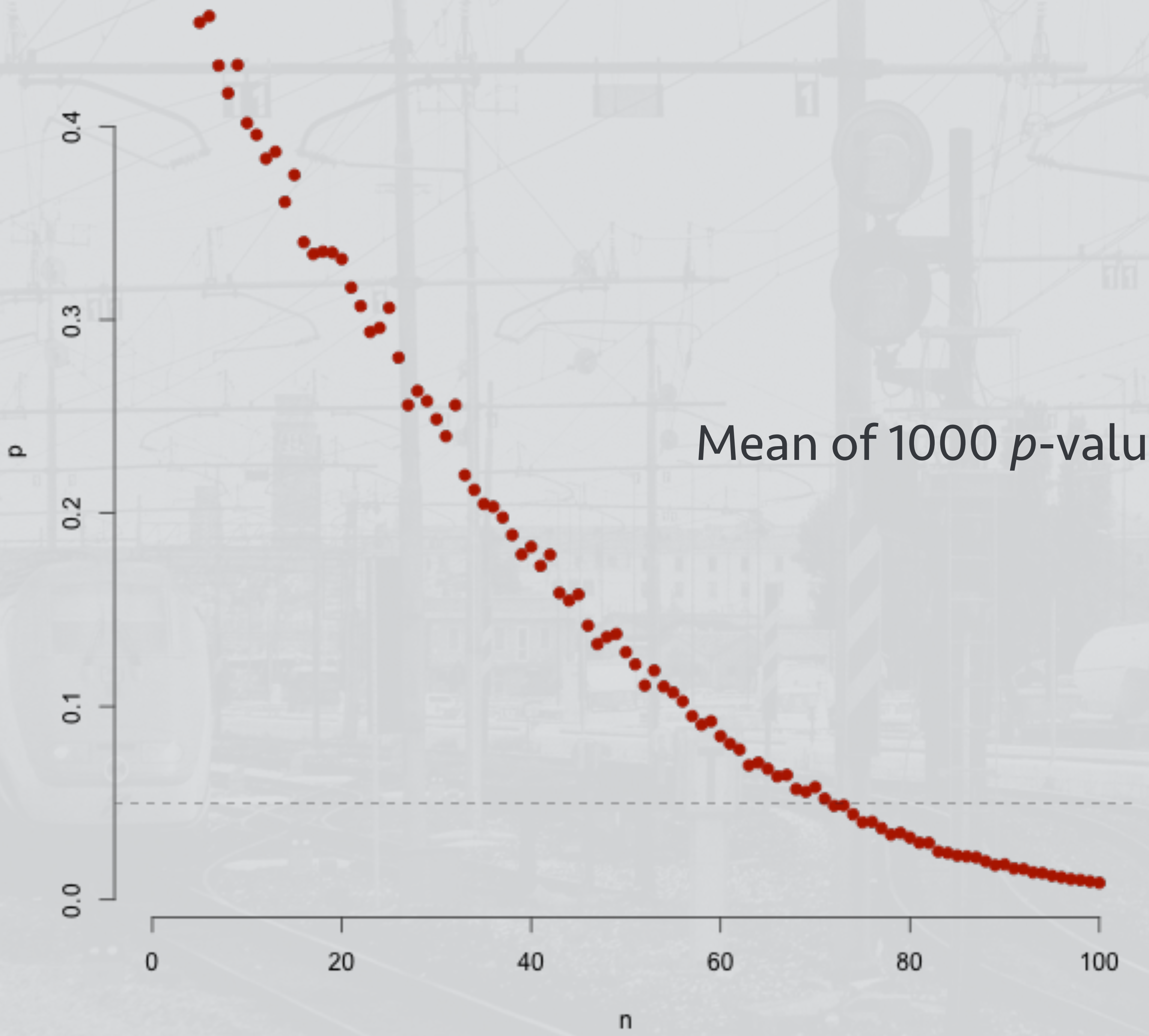
```
> t.test(a, b)

    Welch Two Sample t-test

data:  a and b
t = -2.6492, df = 197.232, p-value = 0.008722
alternative hypothesis: true difference in
means is not equal to 0
95 percent confidence interval:
 -0.63820792 -0.09351402
sample estimates:
 mean of x  mean of y
-0.1366332  0.2292278
```

```
> se2.a <- var(a) / length(a)
> se2.b <- var(b) / length(b)
> t <- (mean(a) - mean(b)) / sqrt(se2.a + se2.b)
> t
[1] -2.6492
> df <- (se2.a + se2.b)^2 / ((se2.a^2)/
(length(a)-1) + (se2.b^2)/(length(b)-1))
> df
[1] 197.2316
> pt(t, df) * 2
[1] 0.00872208
```

Effect of sample size



Mean of 1000 *p*-values at each *n*

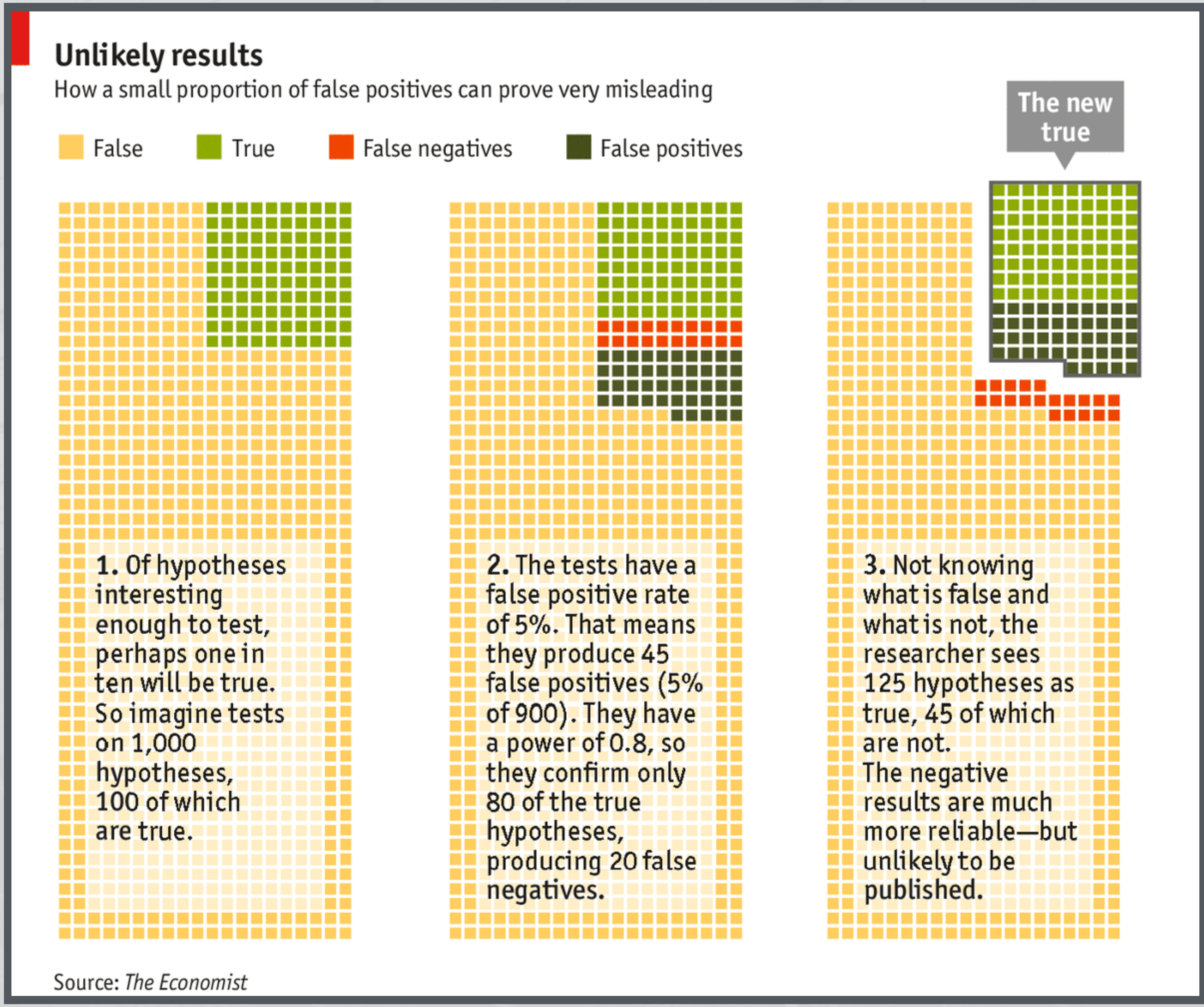
Other common hypothesis tests

- *t*-test for significant correlation coefficient
- *t*-test for significant regression coefficient
- *F*-test for difference between multiple means
- *F*-test for model comparison
- Nonparametric equivalents, e.g. signed-rank test
- Robustness to violations of assumptions varies

Issues with significance tests

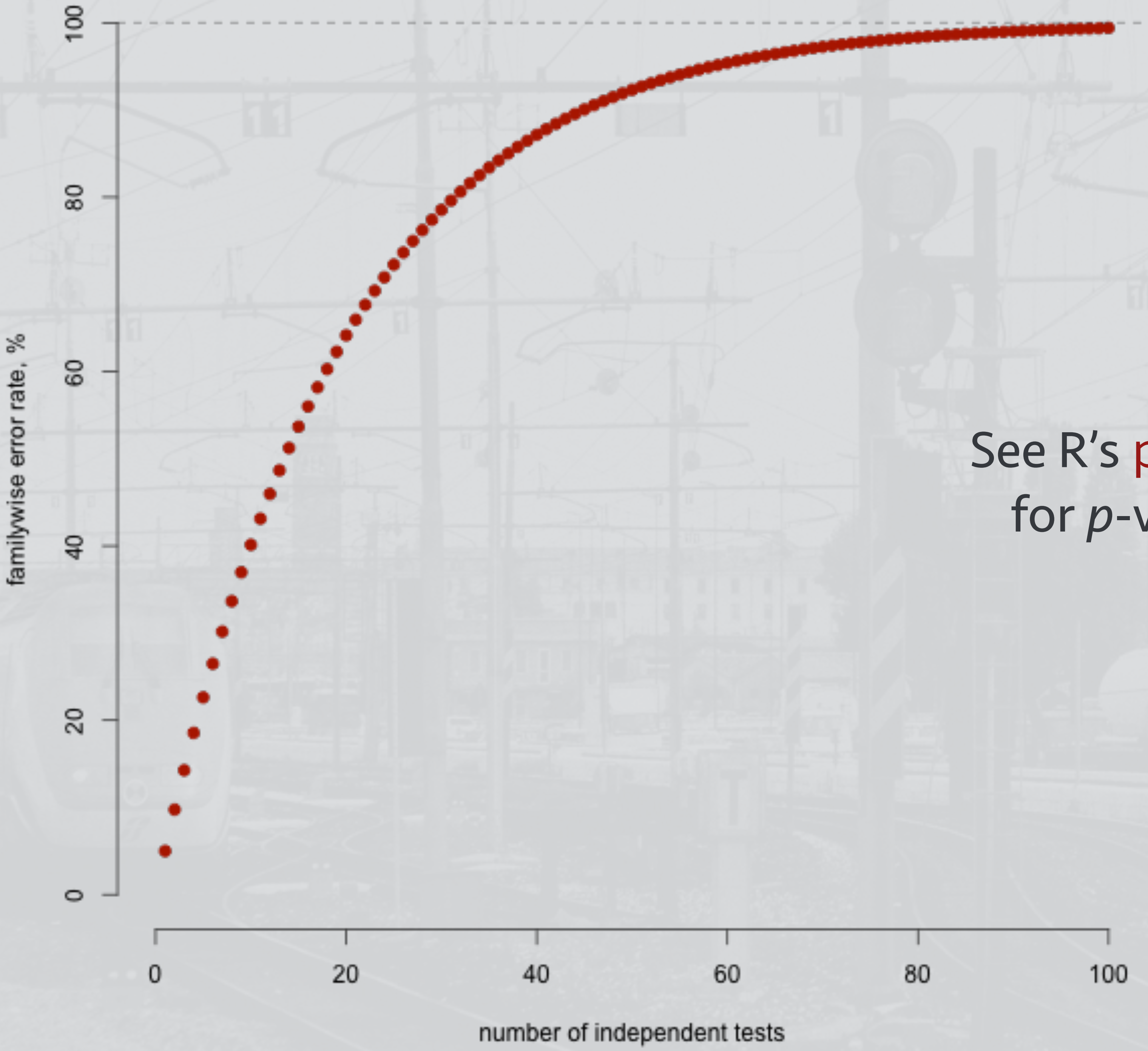
- Arbitrary *p*-value threshold
- Significance vs effect size, especially with many observations
- Publication bias: non-significant results are rarely published
- Incentives for *p*-hacking
- Choice of null hypothesis can be controversial
- Ignores any prior information
- Probability of observing data under the null hypothesis (obtained) vs probability that hypothesis is correct (often desired)

The big-picture problem



The Economist, 19th October 2013

Multiple comparisons



See R's `p.adjust` function for *p*-value adjustments

The picture in imaging

- Hypothesis tests may be performed on a variety of scales
- Worth carefully considering the appropriate scale for the research question
- Dimensionality reduction can be helpful
- Mass univariate testing (e.g. voxelwise) produces a major multiple comparisons issue

Linear (regression) models

- We have some measurement, y, for each subject
- We have some predictor variables, x_1, x_2, x_3, etc., for which we have measurements for each subject
- We want to know β_1, β_2, β_3, etc., the influences of each x on y
- We use the model

  $$y^i = \beta_0 + \beta_1 x_1^i + \ldots + \beta_p x_p^i + \varepsilon^i$$

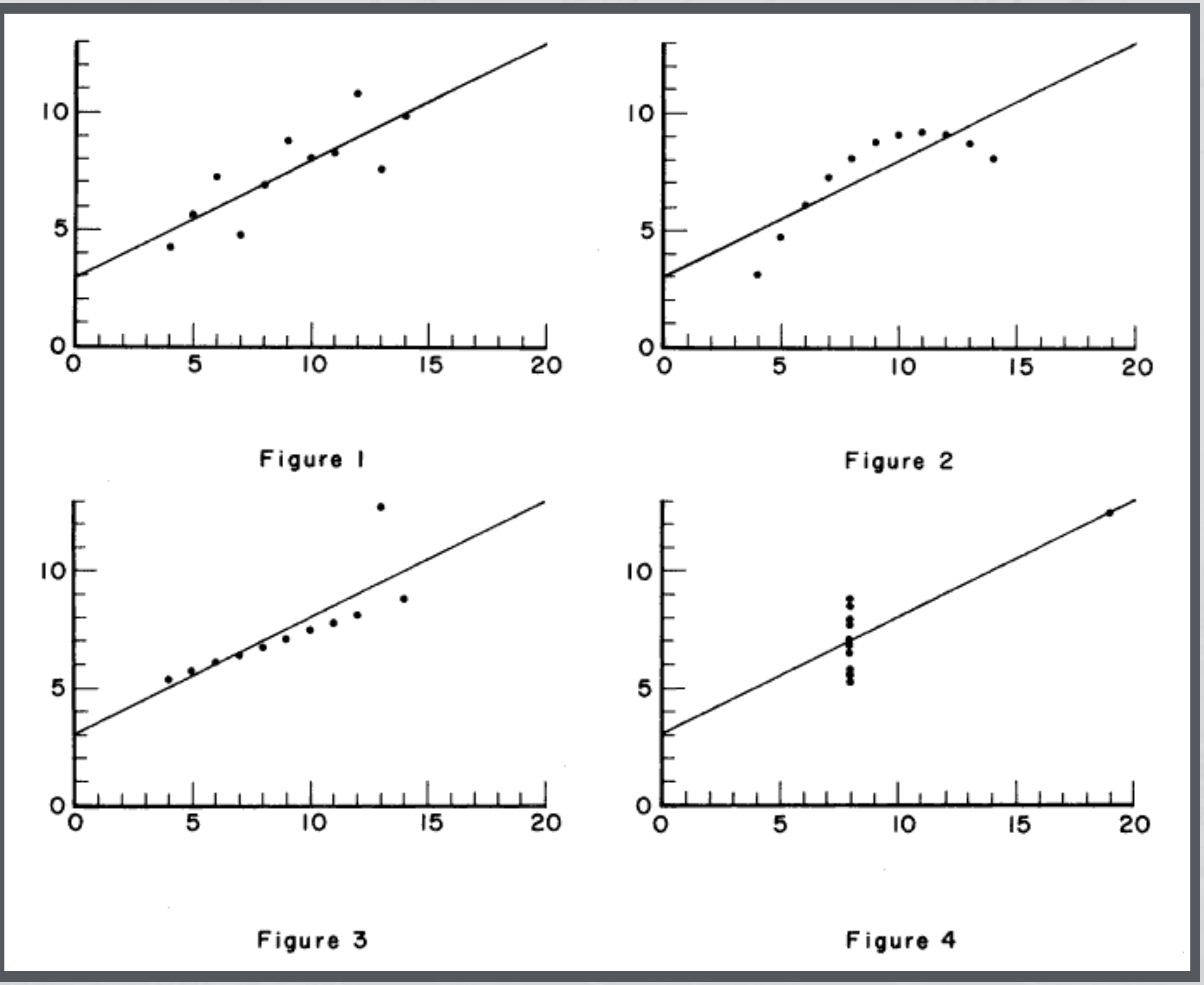
  where the errors (or residuals), ε^i, are assumed to be normally distributed with zero mean
- Typically fitted with ordinary least squares, a simple matrix operation
- Assumes constant variance, independent errors, noncollinearity in predictors

A versatile tool

- With one predictor, a regression model is closely related to (Pearson) correlation or *t*-test
- With more predictors, also covers analysis of (co)variance
- Extension to multivariate outcomes (general linear model) covers MANOVA, MANCOVA

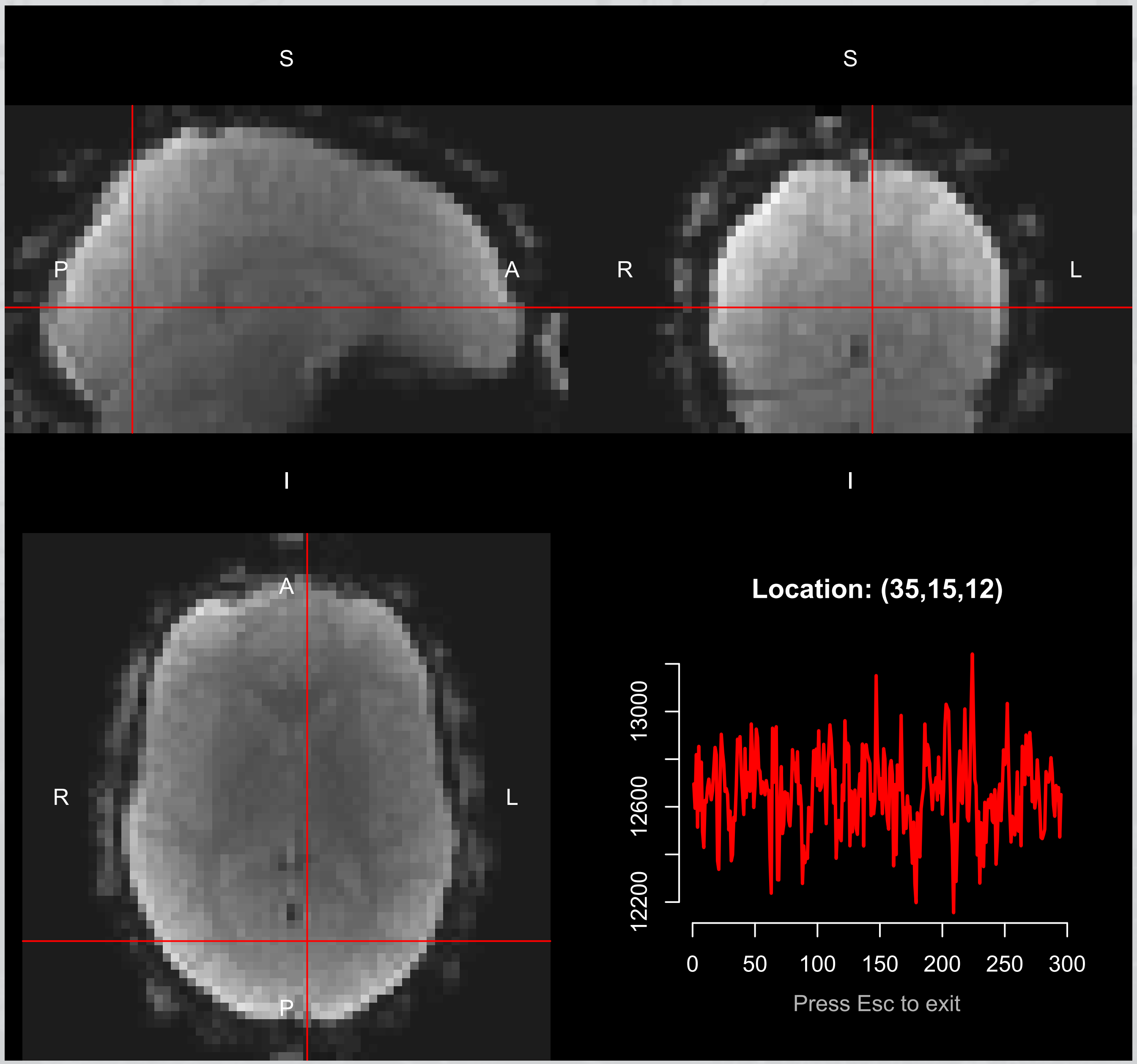
Anscombe's quartet, or, why you should look at your data

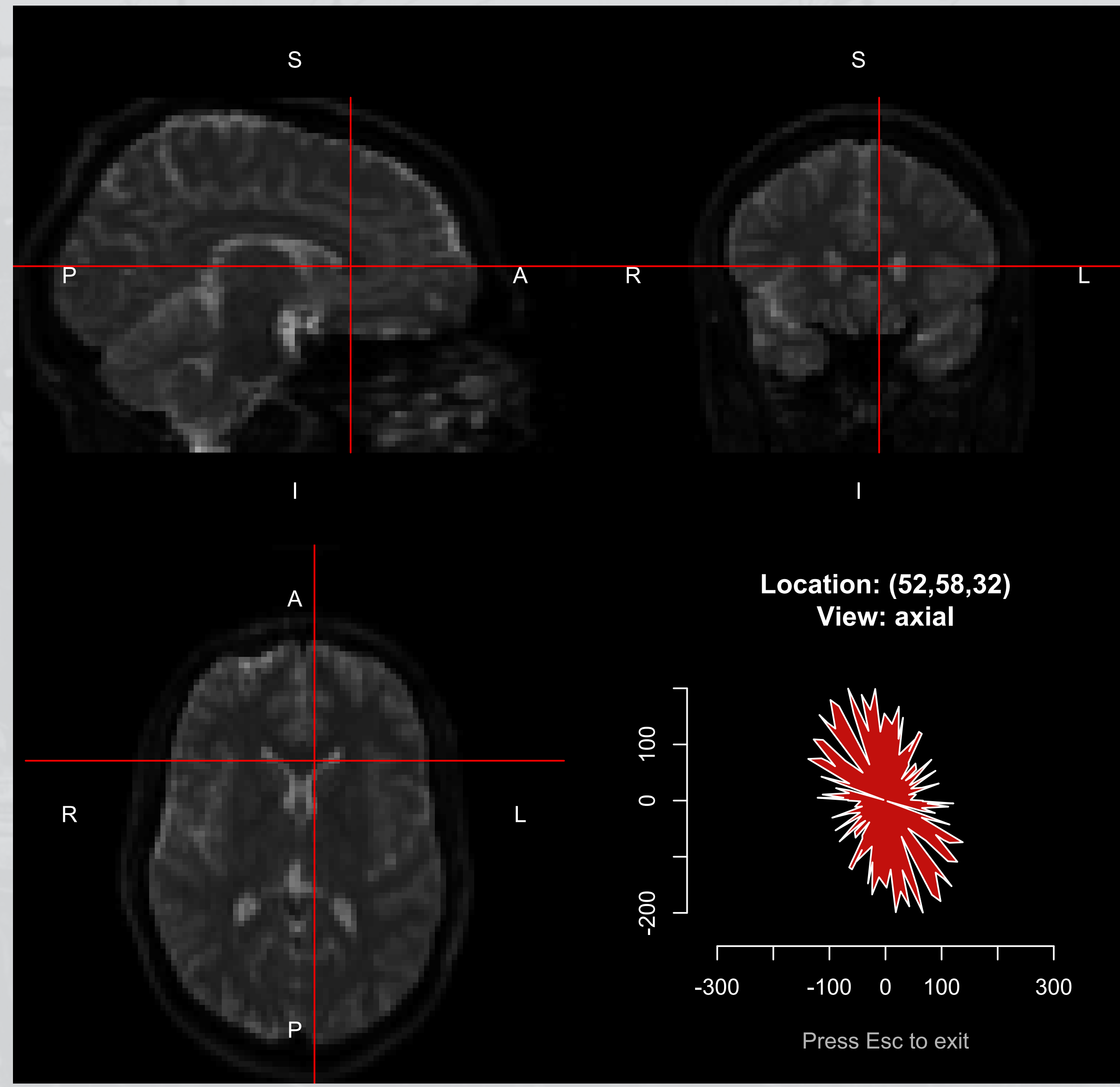
- Same mean
- Same variance
- Same correlation coefficient
- Same regression line



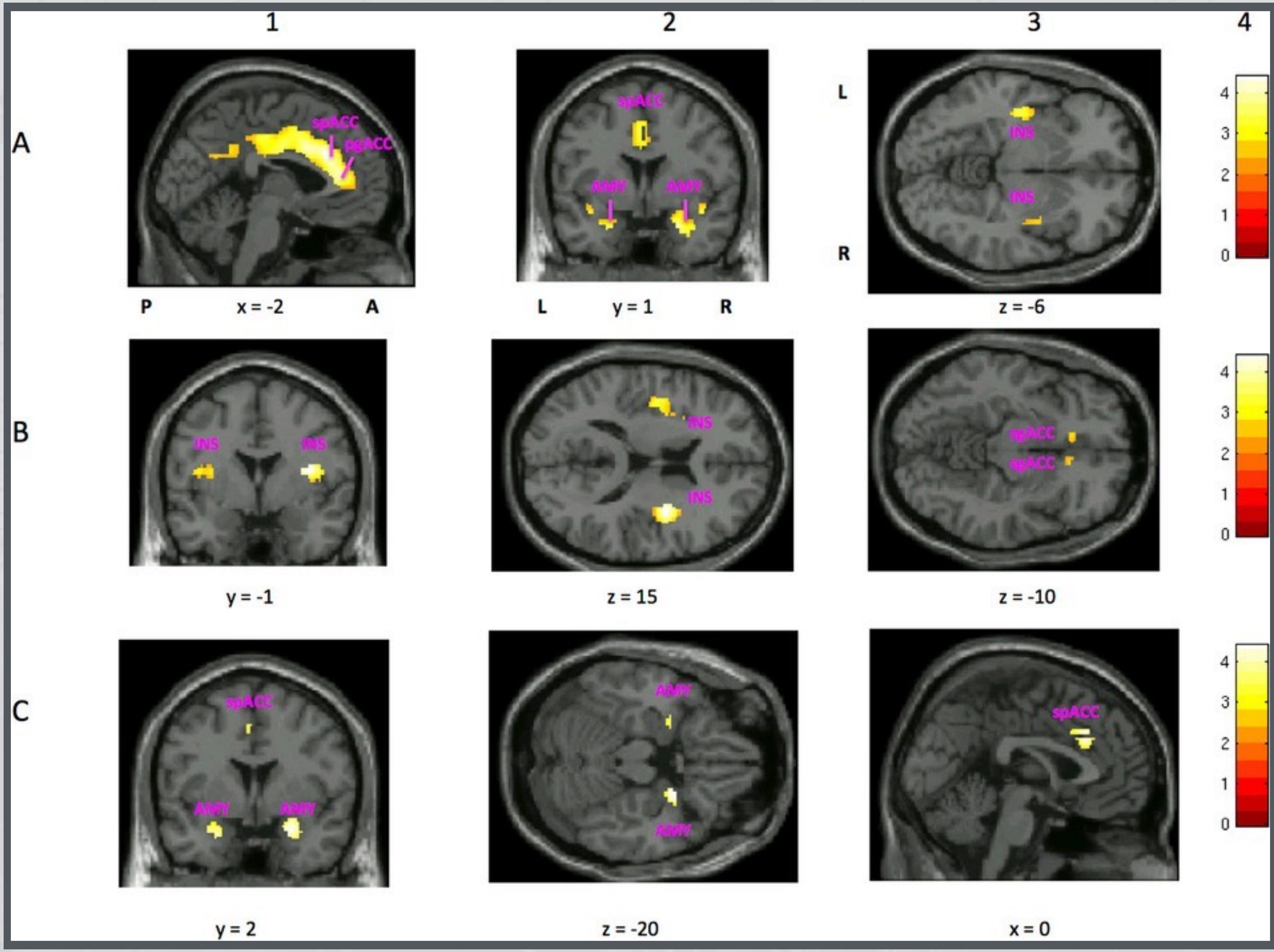
Anscombe, *Amer Stat*, 1973

Visualising complex image data





SPM

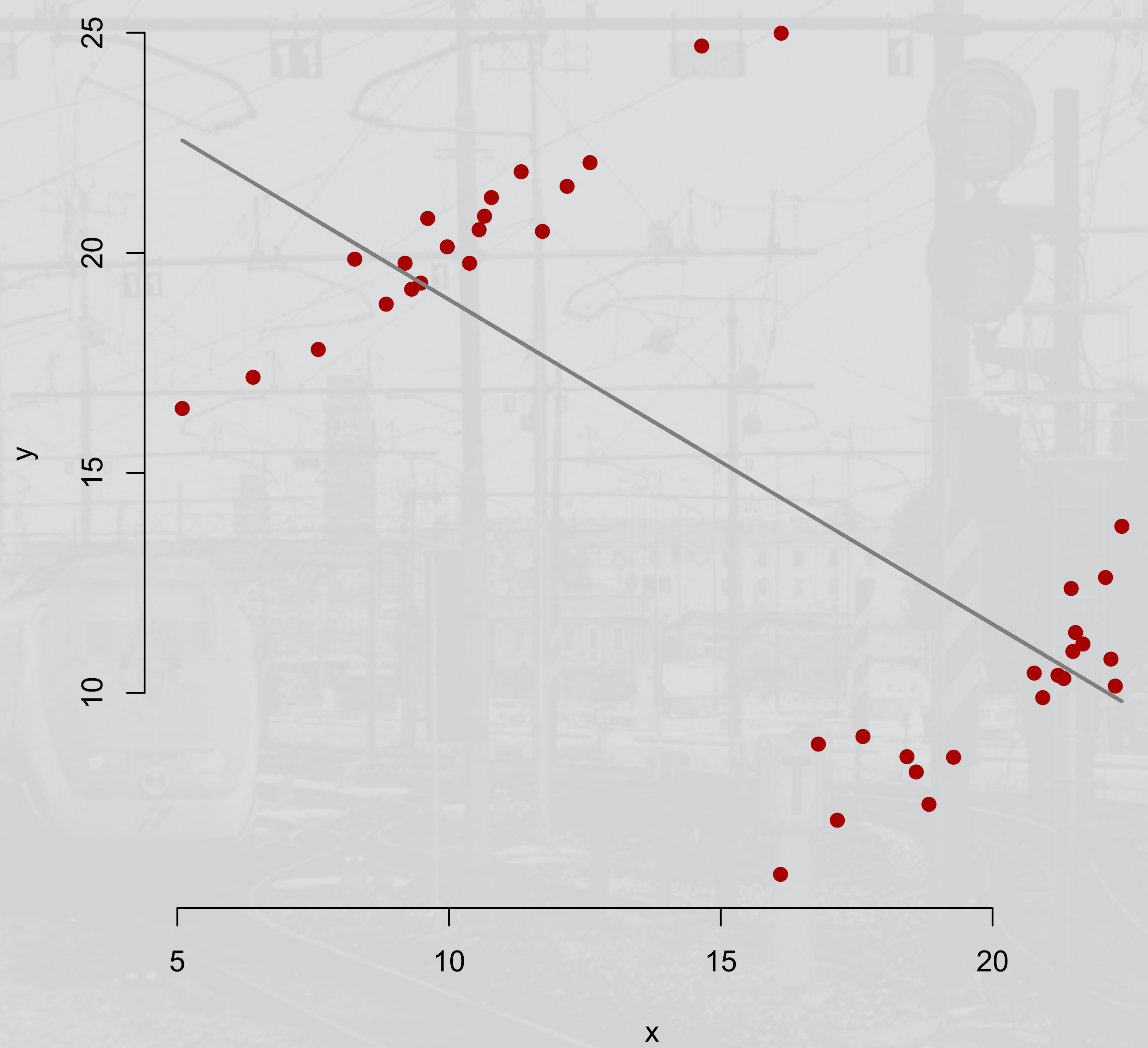


Savitz et al., *Sci Reports*, 2012

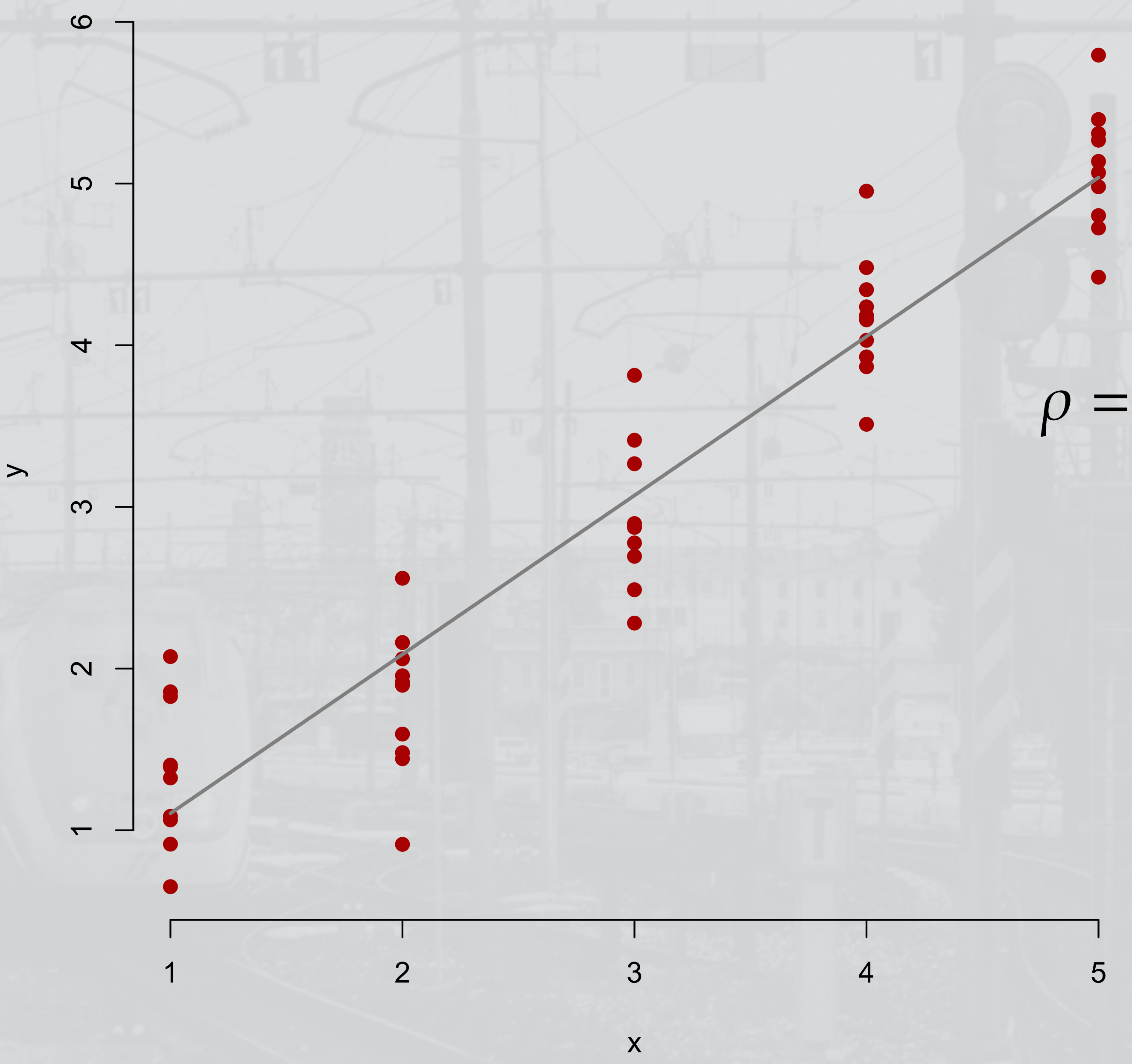
Beyond hypothesis tests

- Models of data as outcomes, plus derivatives such as reference ranges
- Parameter estimates, confidences intervals, etc.
- Model comparison via likelihood, information theory approaches
- Clustering
- Predictive power, e.g. ROC analysis
- Measures of uncertainty via resampling methods
- Bayesian inference: prior and posterior distributions

Simpson's paradox

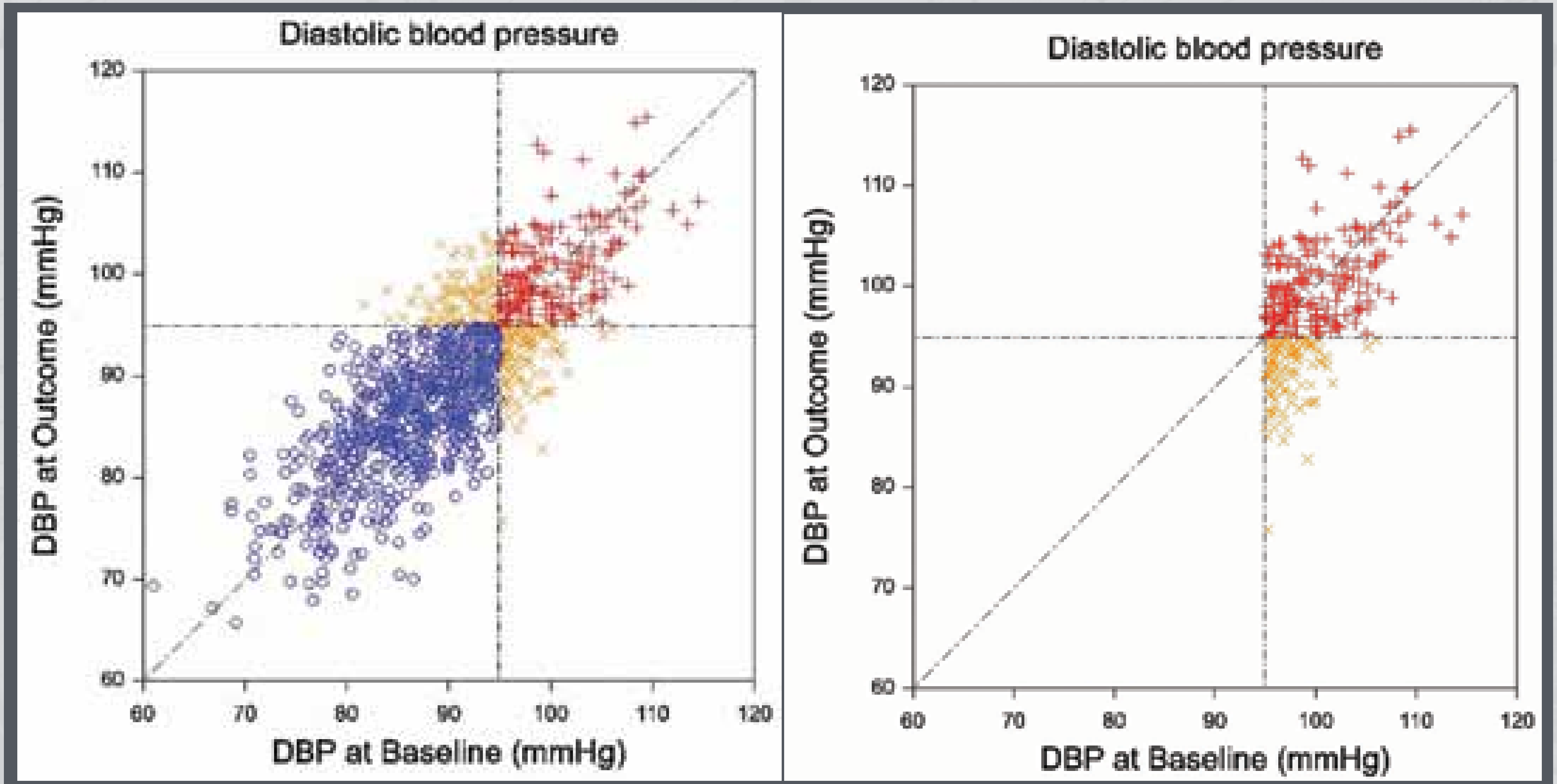


Categorical variables, ties and correlation



$\rho = 0.95$

Regression to the mean



Senn, *Write Stuff*, 2009

Some advice

- Plan ahead
- Be clear what you really want to know
- Use R
- Visualise and understand your data
- Save scripts
- Keep statistical tests to a minimum
- Be aware of sources of bias
- Use available resources at ICH and beyond