CAUSALITY
& CORRELATION

How are continuous variables related
and are these relations causal?

IHenrik von Wehrden

CAUSALITY OR CORRELATION

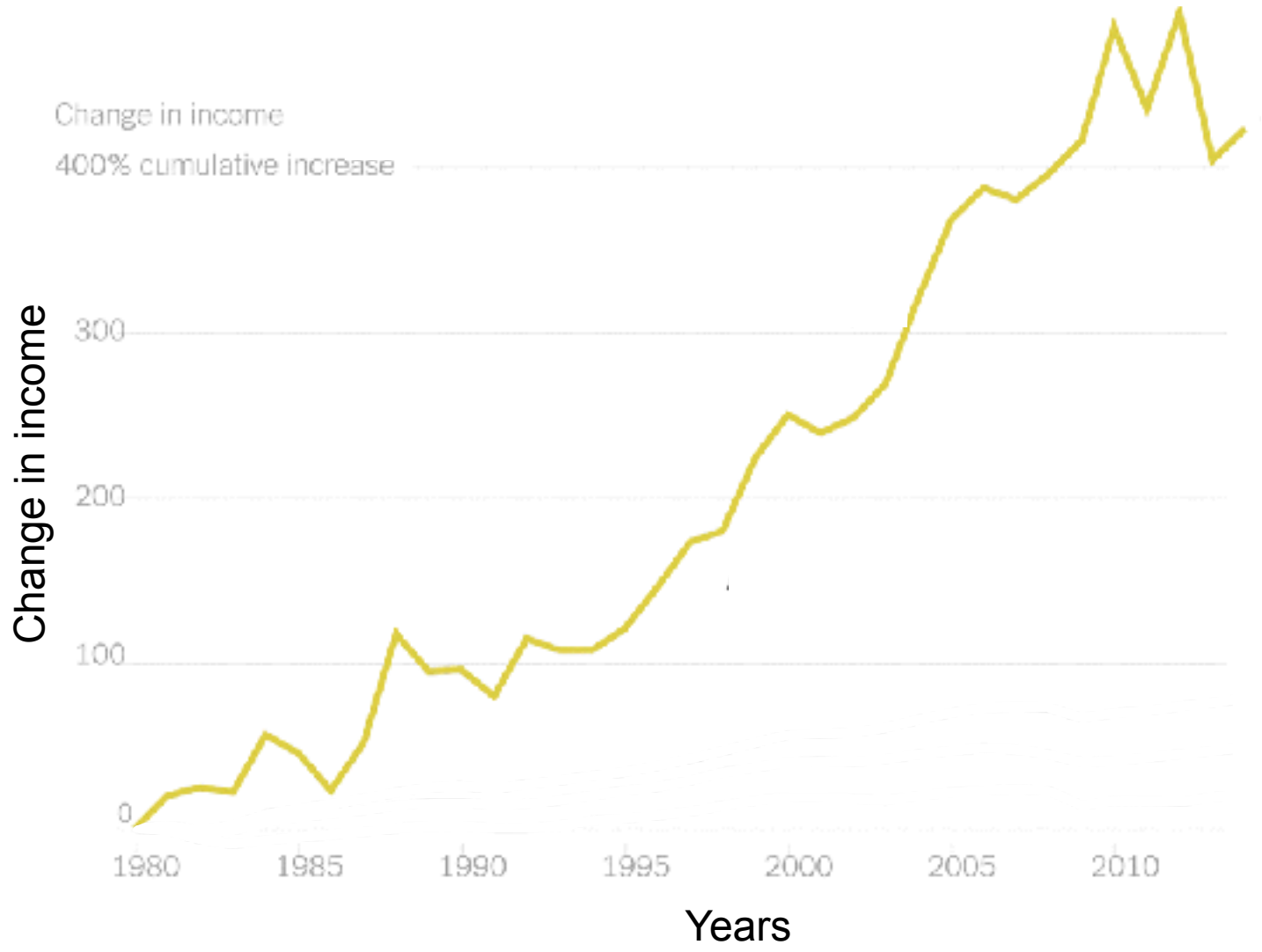
Are relations really causal?
And does this matter?

→ Henrik von Wehrden

LEUPHANA
UNIVERSITÄT LÜNEBURG

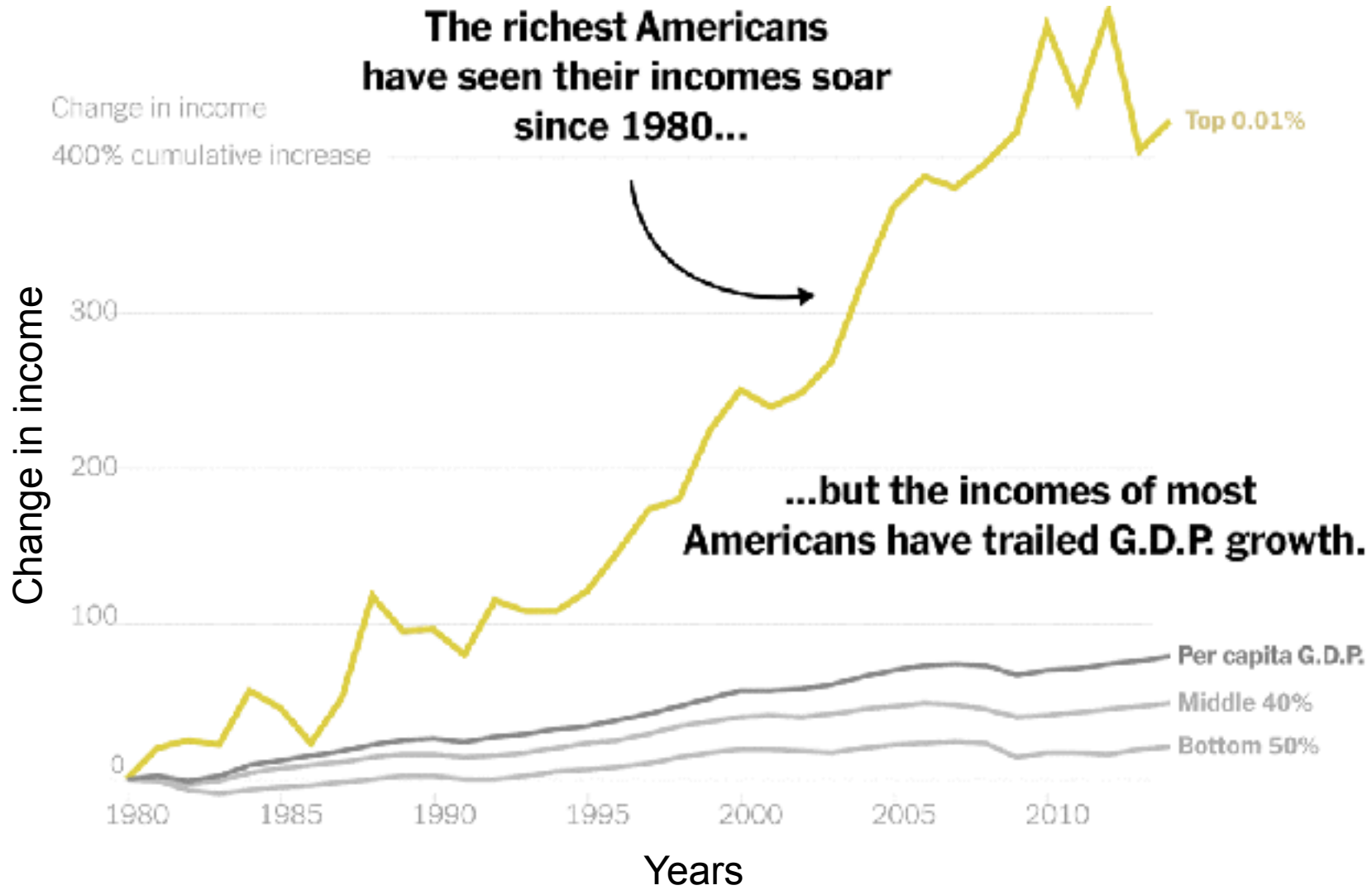
LET US START SIMPLE: STATISTICAL CORRELATIONS

How are two continuous variables in relation with each other?



LET US START SIMPLE: STATISTICAL CORRELATIONS

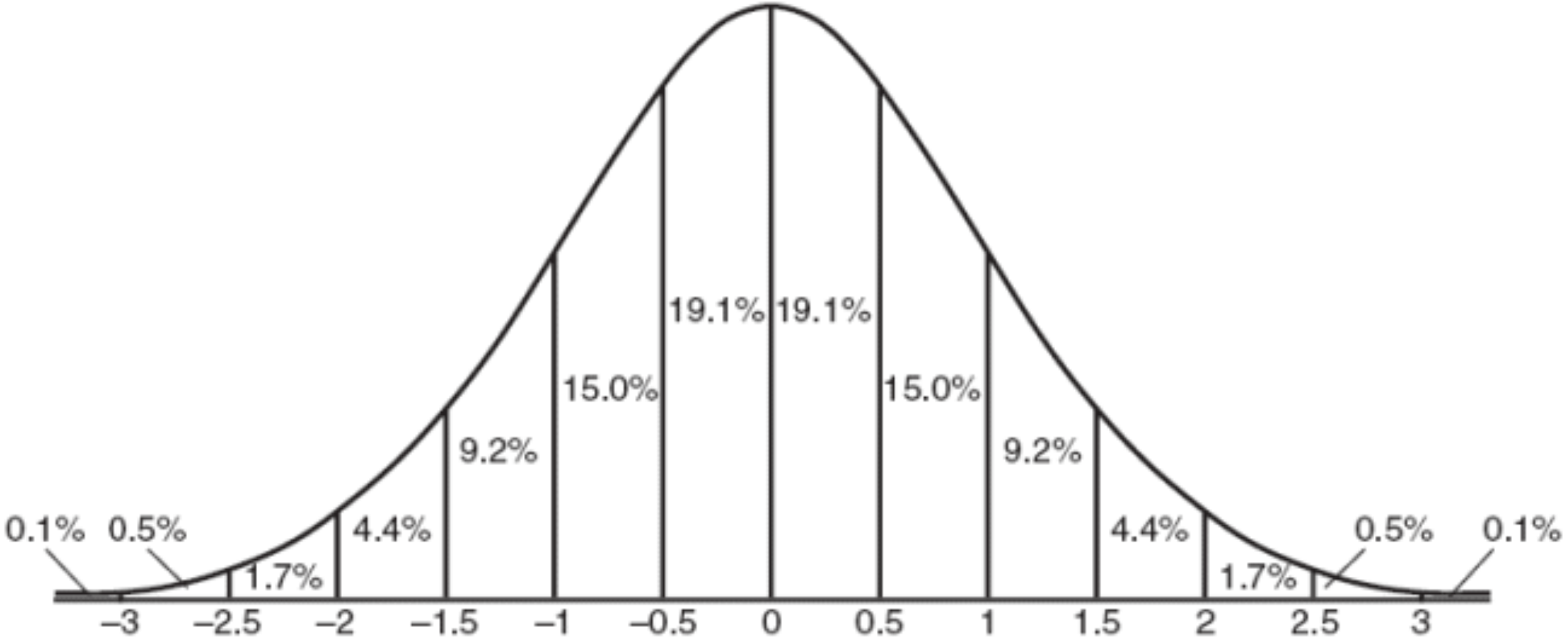
How are two continuous variables in relation with each other?



HISTORY OF CORRELATIONS

- Early focus by Pearson widely theoretical
- Fisher unraveled linear relations
- Explanatory measures have no real threshold
- Statistical power is highly relevant
- Shift from hard to soft science

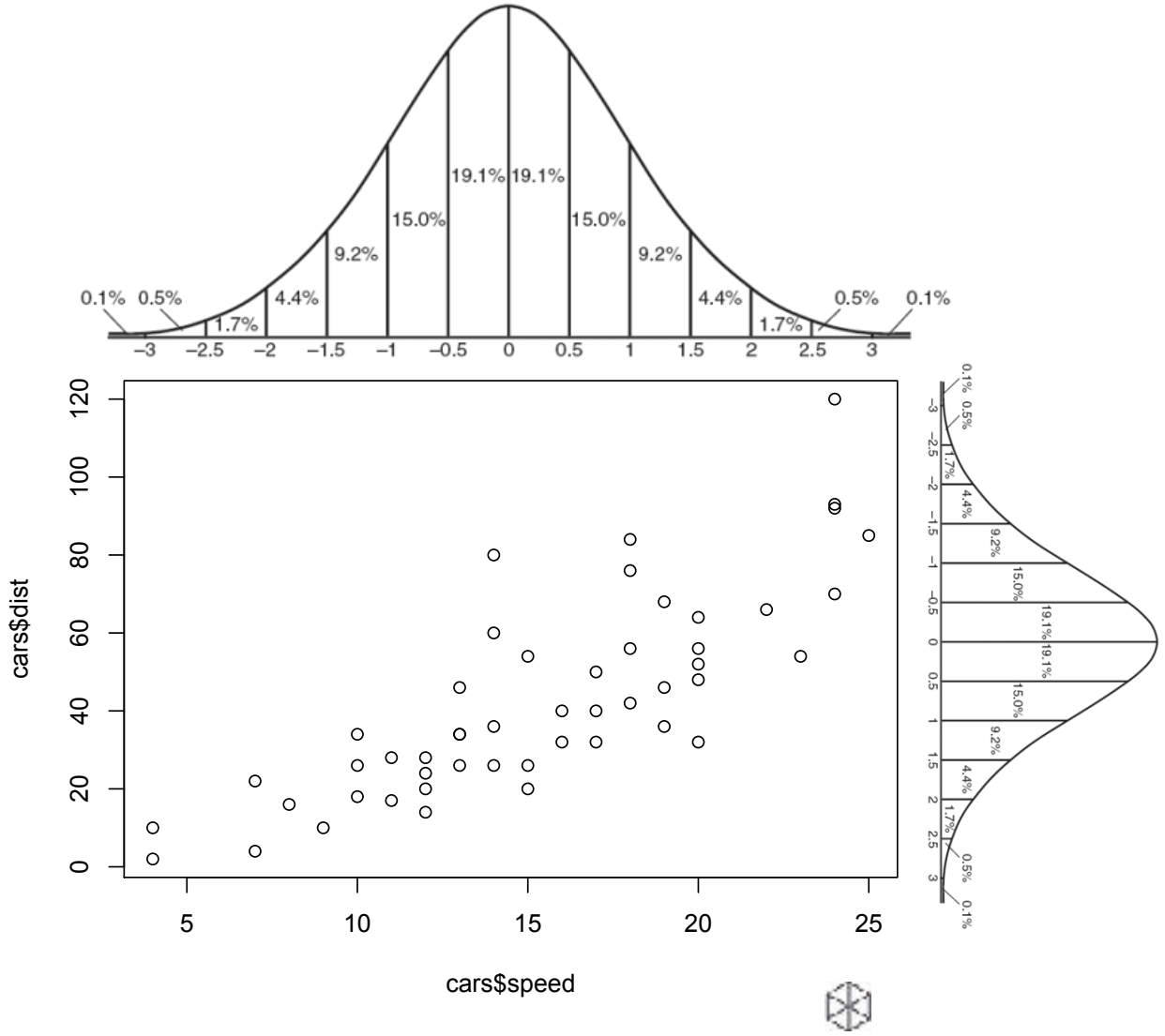
THE NORMAL DISTRIBUTION



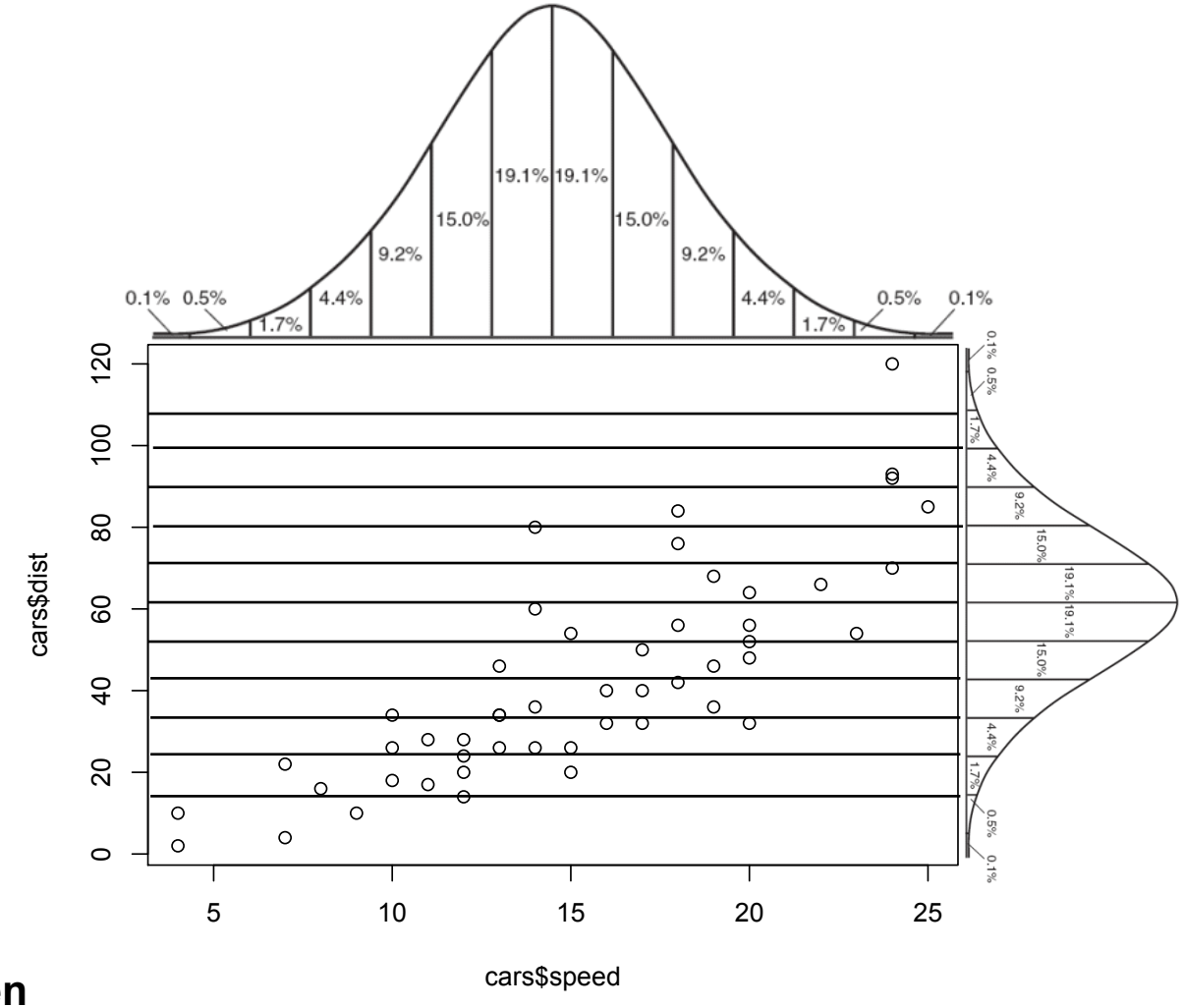
LET US START SIMPLE: STATISTICAL CORRELATIONS

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

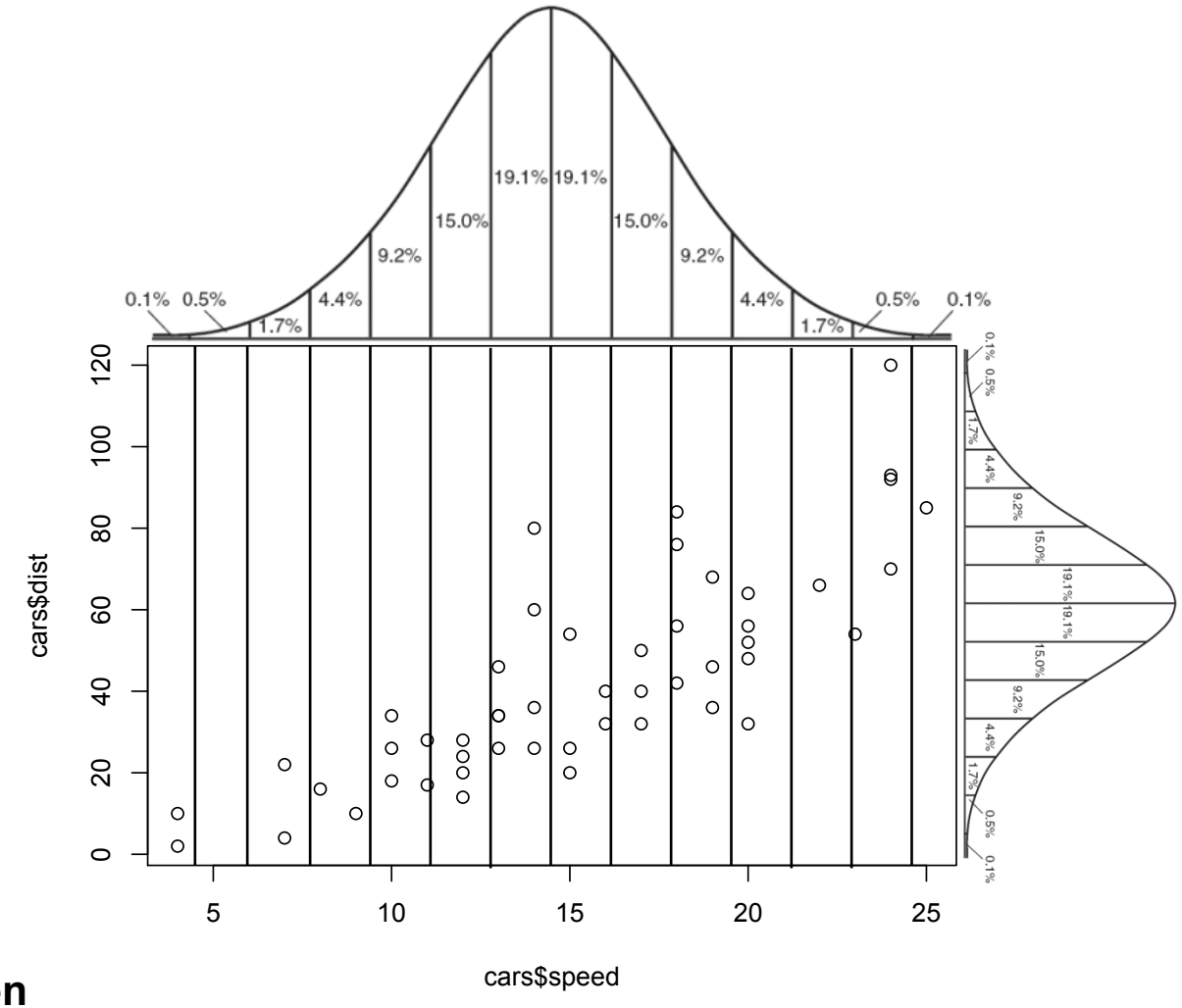
- r_{xy} – the correlation coefficient of the linear relationship between the variables x and y
- x_i – the values of the x-variable in a sample
- $\bar{x}$ – the mean of the values of the x-variable
- y_i – the values of the y-variable in a sample
- $\bar{y}$ – the mean of the values of the y-variable



LET US START SIMPLE: STATISTICAL CORRELATIONS



LET US START SIMPLE: STATISTICAL CORRELATIONS



PROBABILITY - STATISTICS

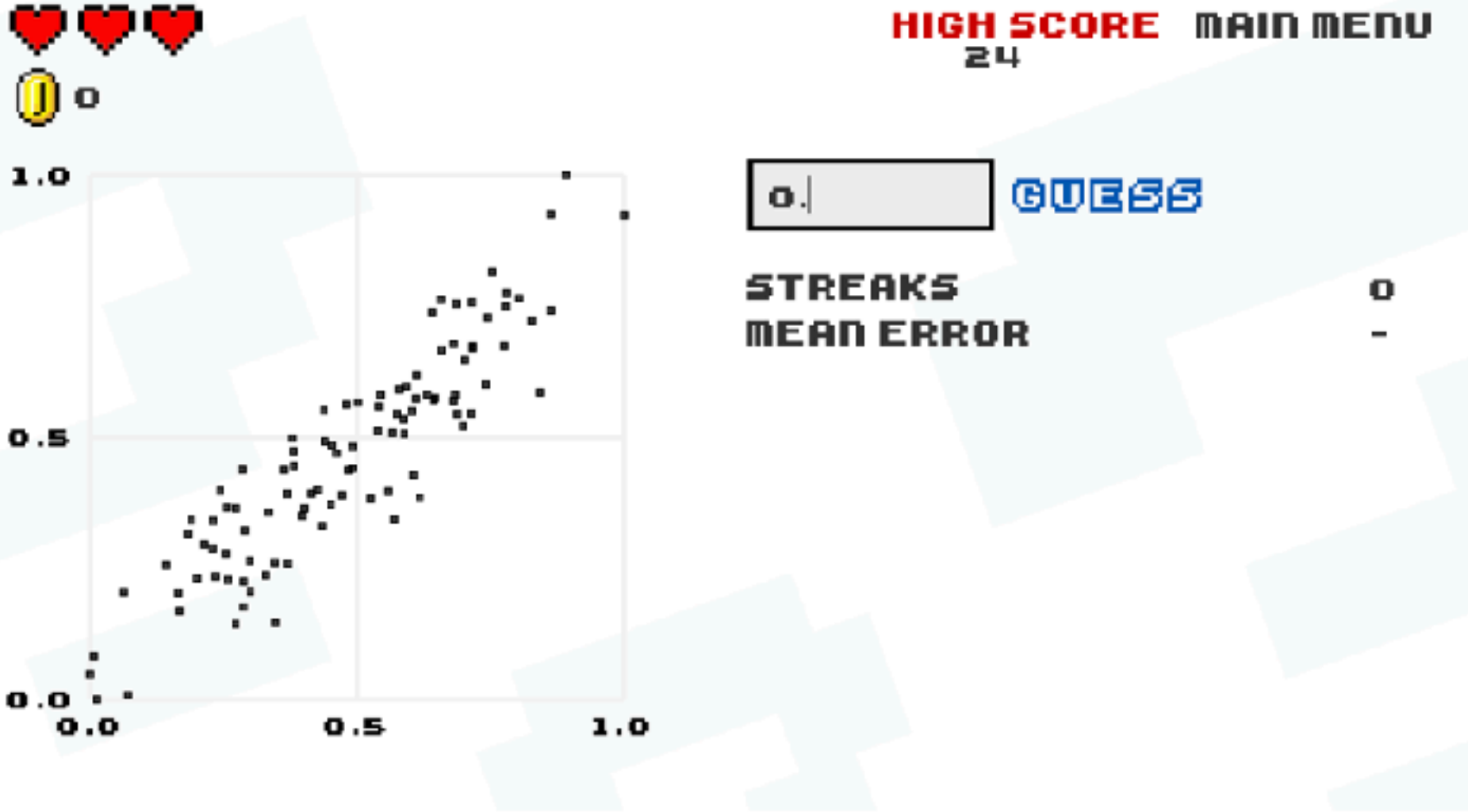
The lady tasting tea

What is the probability?

HOW MUCH DOES A MODEL EXPLAIN?

http://guessthecorrelation.com

HOW MUCH DOES A MODEL EXPLAIN?



http://guessthecorrelation.com

STATISTICAL CORRELATIONS: ONE VARIABLE RELATES ON ONE VARIABLE.

```
> cor.test(cars$dist,cars$speed)

        Pearson's product-moment correlation

data:  cars$dist and cars$speed
t = 9.464, df = 48, p-value = 1.49e-12
alternative hypothesis: true correlation is not equal
0
95 percent confidence interval:
 0.6816422 0.8862036
sample estimates:
      cor
0.8068949
```

P-VALUE:
0.000000000000149

CORRELATION COEFFICIENT
RANGES FROM -1 TO 1

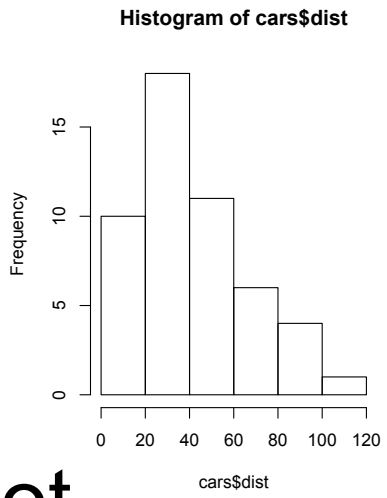
STATISTICAL CORRELATIONS: ONE VARIABLE RELATES ON ONE VARIABLE.

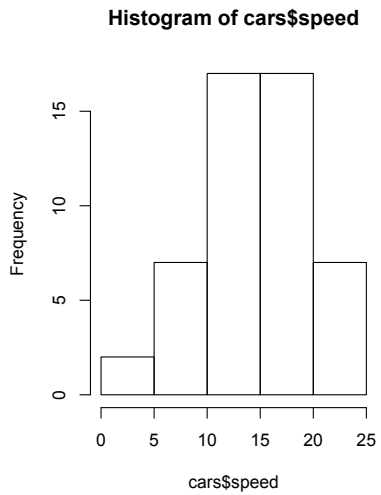
Correlation

Correlation coefficient ranges from -1 to 1
Inspect the relation for flaws
Correlations can be inductive and deductive

Check for data density
Data gaps
Changes in the relation

Diagnostics:
x-y-plot, Normal distribution?,Scatterplot





STATISTICAL CORRELATIONS: ONE VARIABLE RELATES ON ONE VARIABLE.

Data	
Normal distribution	Pearson‘s R
Non normal distribution	Spearman‘s R

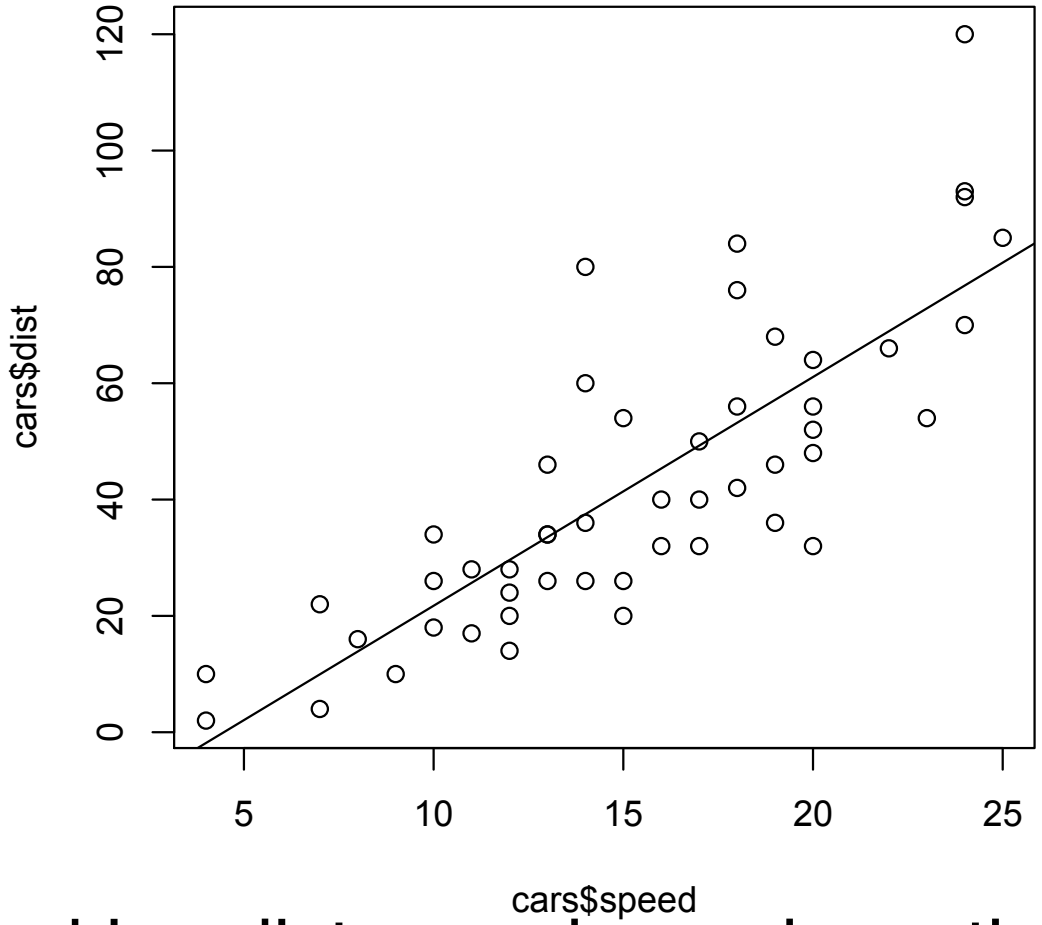
STATISTICAL CORRELATIONS: ONE VARIABLE RELATES ON ONE VARIABLE.

Pearson correlations

linear relation between two variables, normal distribution?; $-1 \leq R \leq 1$.

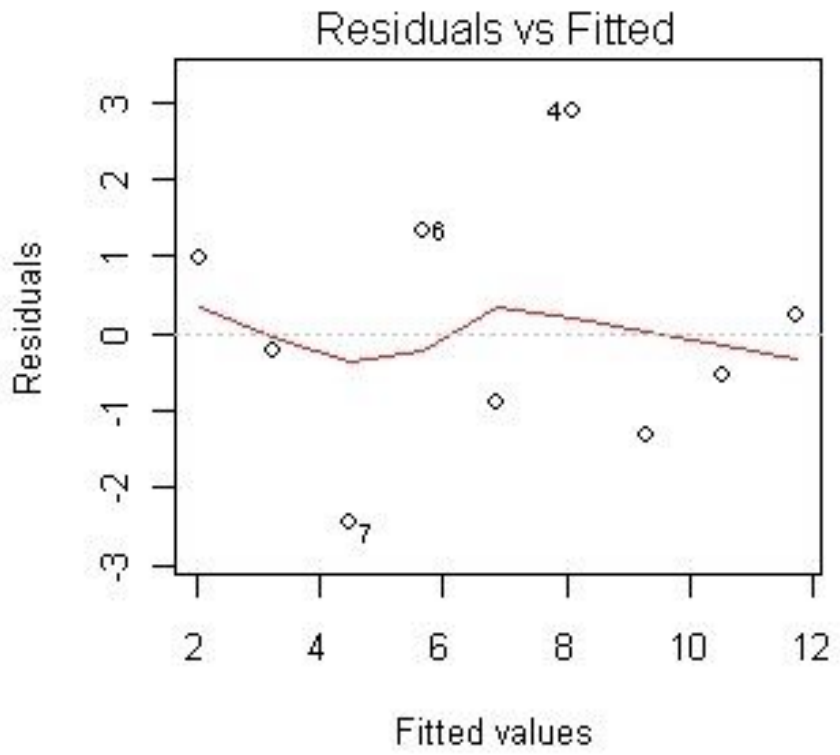
- cor(x,y) → simple values
- cor(data) → simple values, correlation
- matrix
- cor.test(x,y) → extensive result with many details

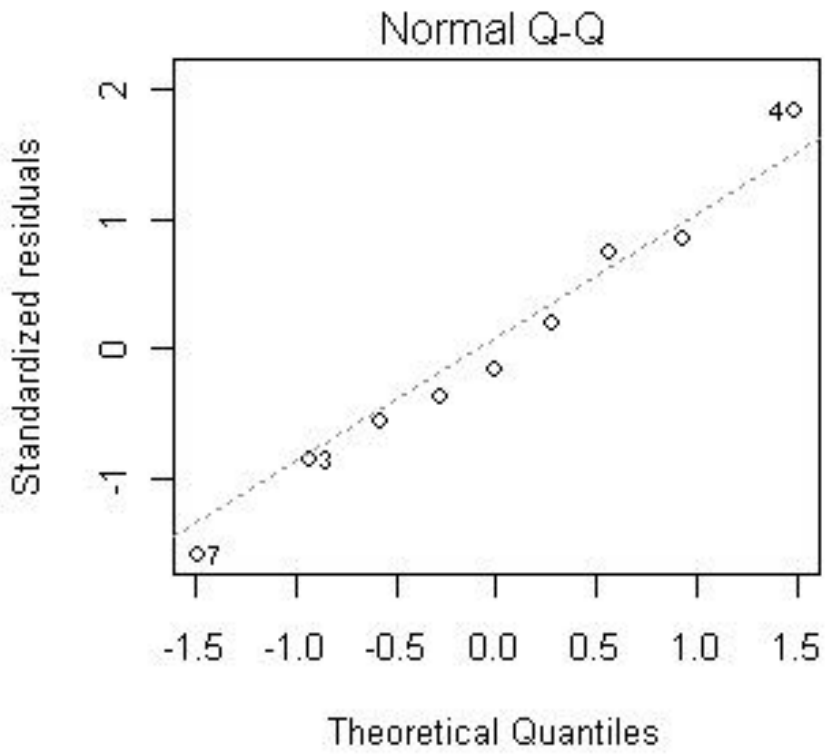
AND NOW THE REGRESSION: ONE VARIABLE IS EXPLAINED BY ANOTHER VARIABLE, AND HENCE DEPENDS ON IT



Breaking distance depends on the speed!

Model inspection





Assumptions: of a linear regression

- independend and dependend Variables are continuous
- dependend variables and residuals are normally distributed (Transformations, GLM)
- Homogeneity variances (Transformations, GLM)

REGRESSION ON A SHOESTRING

Linear regression

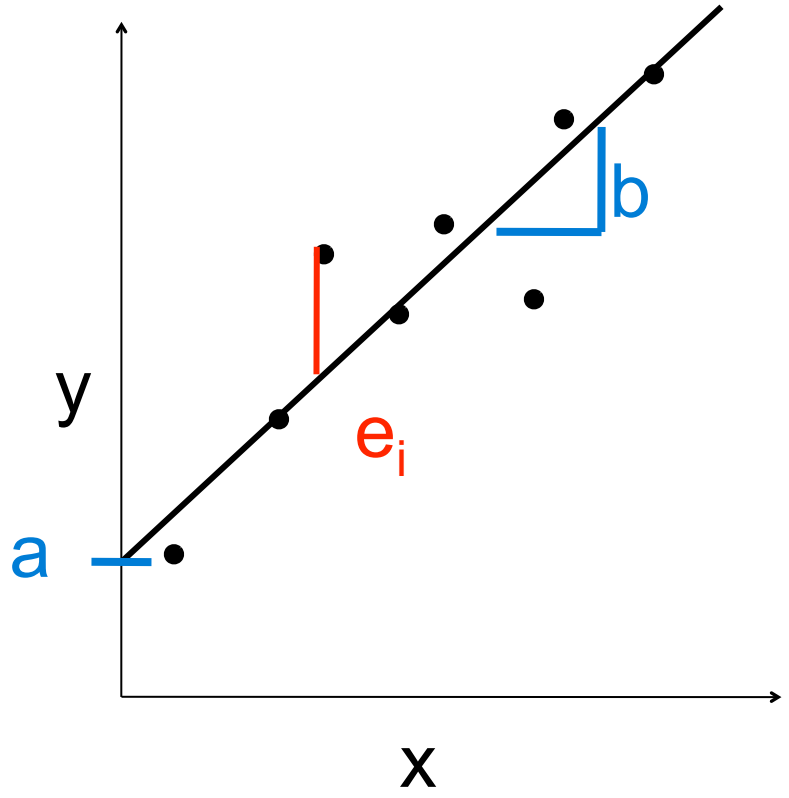
- Relation between an independent variable **x** and a dependend variable **y**

$$y = a + bx$$

a Intercept with the y-axis

b Estimate of the regression

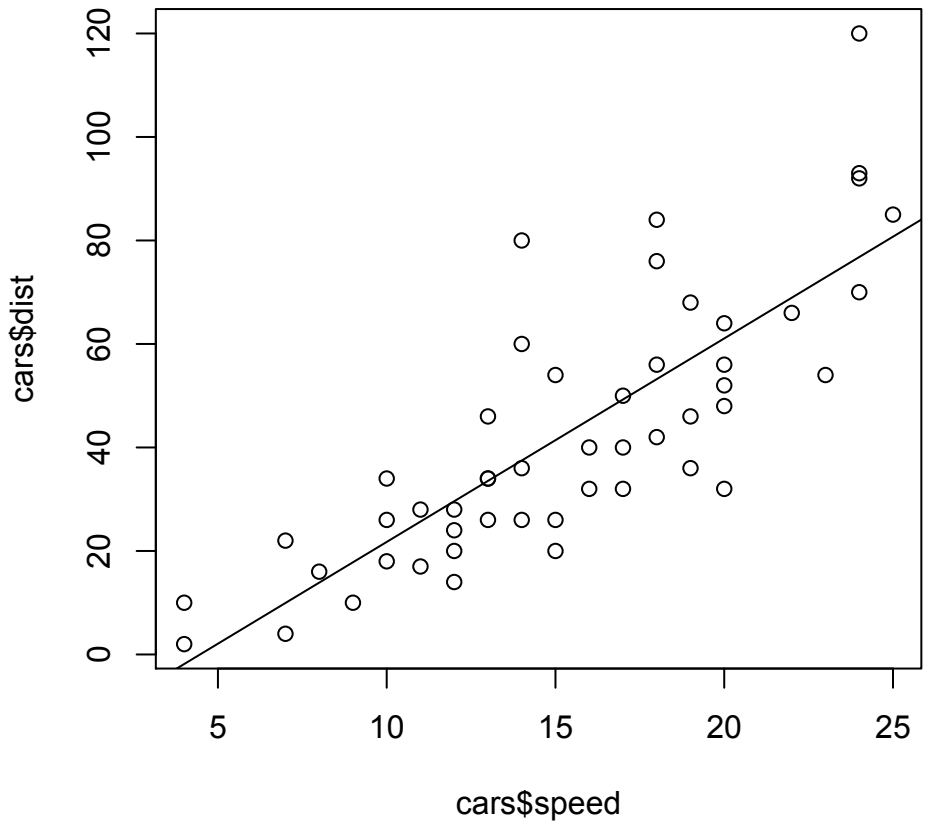
REGRESSION AS A FIGURE AND A FORMULA



$$y = a + bx$$

Residual e_i

PREDICTIONS: SOMETIMES OUTSIDE OF THE DATA



?

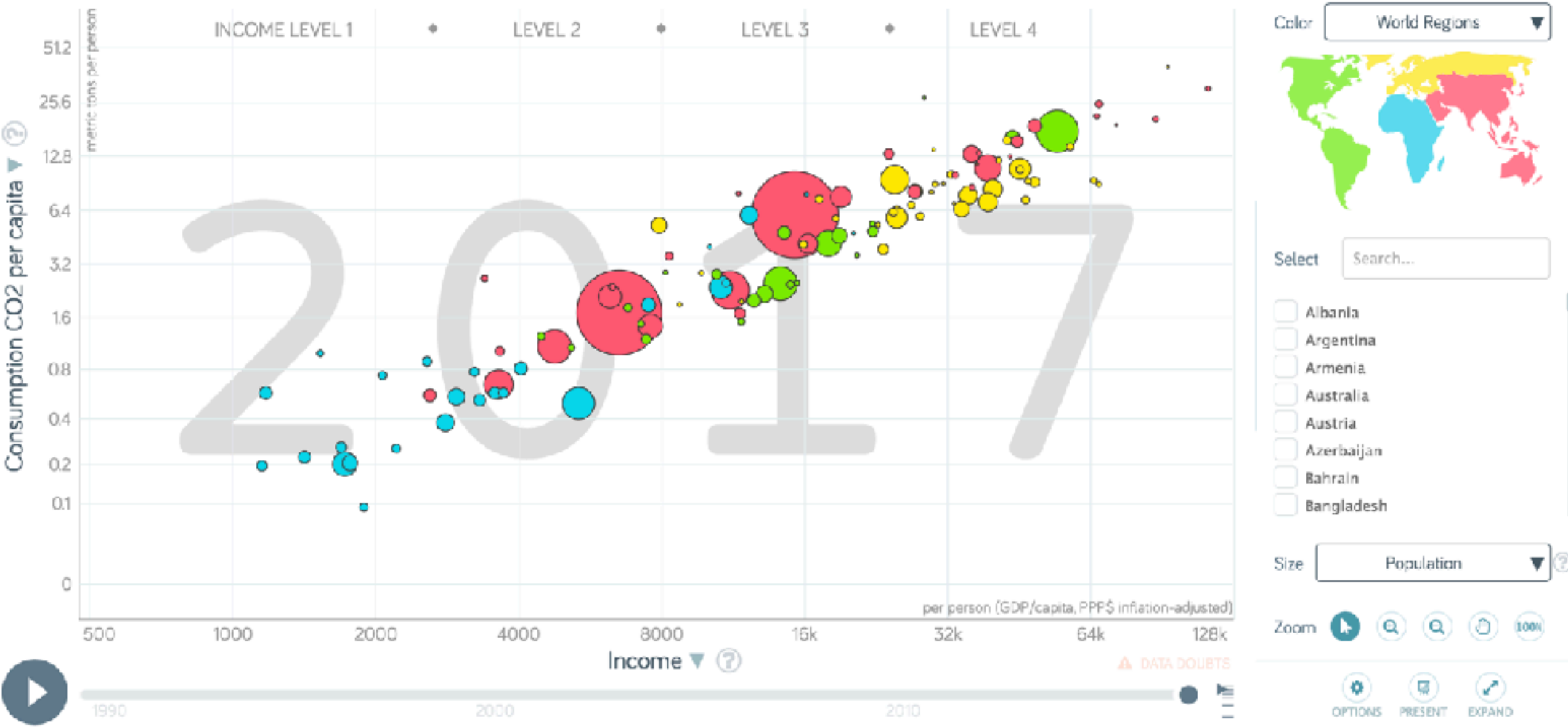
UNCERTAINTY - HOW CERTAIN ARE WE?

There are measures of uncertainty!
Context matters!
What could be unknown?
No model is perfect!

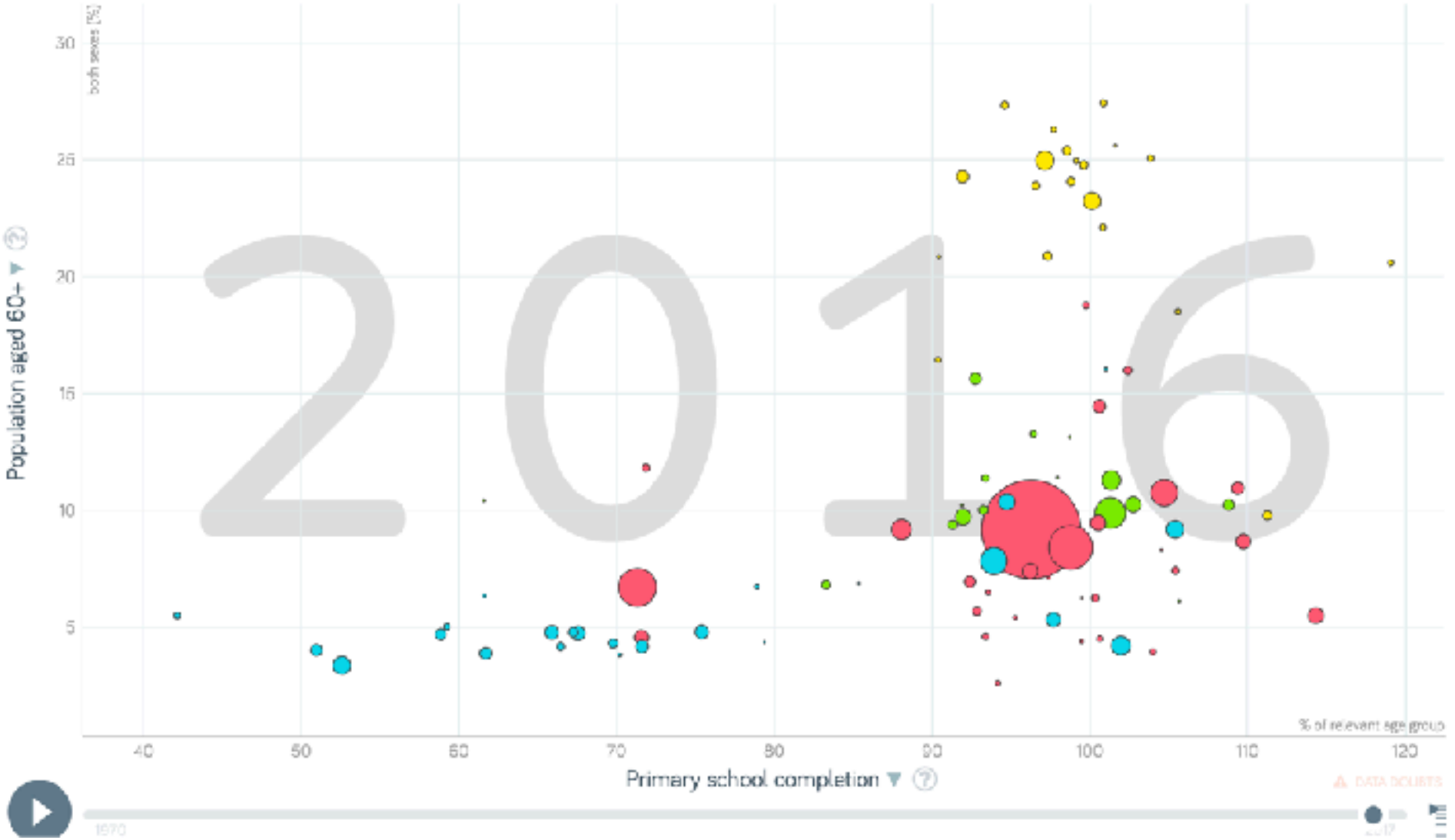
R EXAMPLE OF A LINEAR REGRESSION

```
> summary(model)

Call:
lm(formula = growth ~ tannin)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4556 -0.8889 -0.2389  0.9778  2.8944

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.7556     1.0408  11.295 9.54e-06 ***
tannin       -1.2167     0.2186  -5.565 0.000846 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.693 on 7 degrees of freedom
Multiple R-Squared: 0.8157,     Adjusted R-squared: 0.7893
F-statistic: 30.97 on 1 and 7 DF,  p-value: 0.000846
```

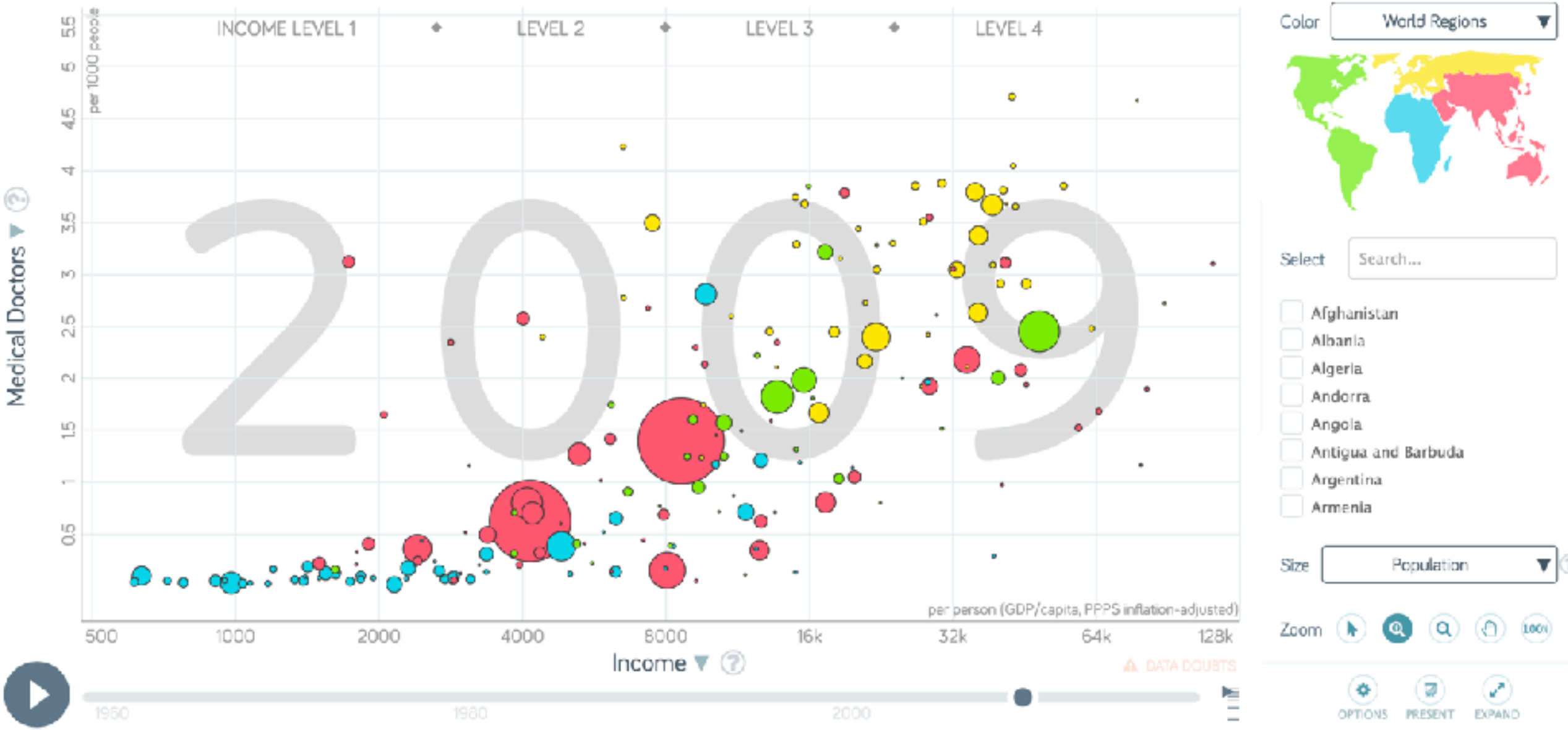

LEARN TO READ CORRELATION PLOTS



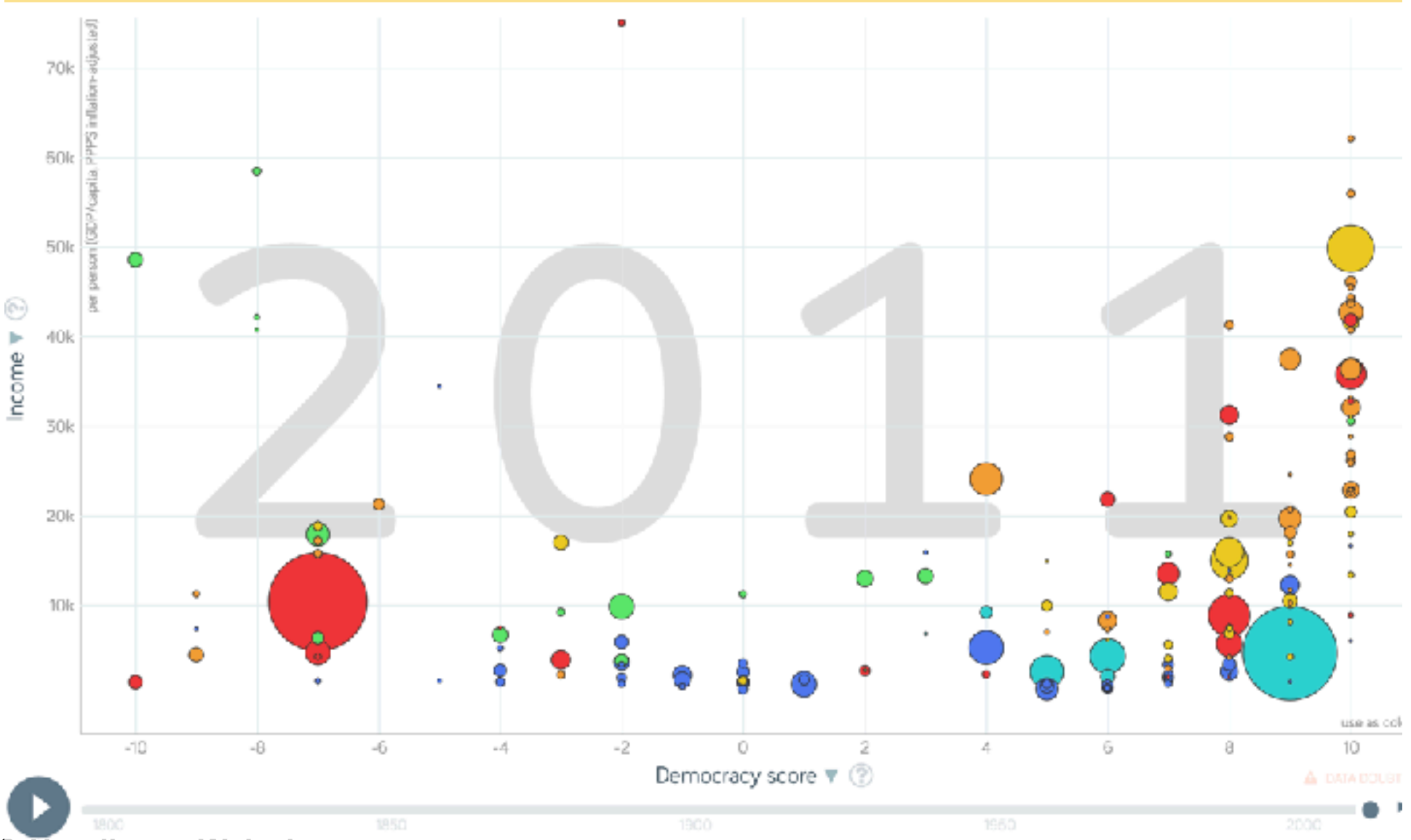
LEARN TO READ CORRELATION PLOTS



LEARN TO READ CORRELATION PLOTS



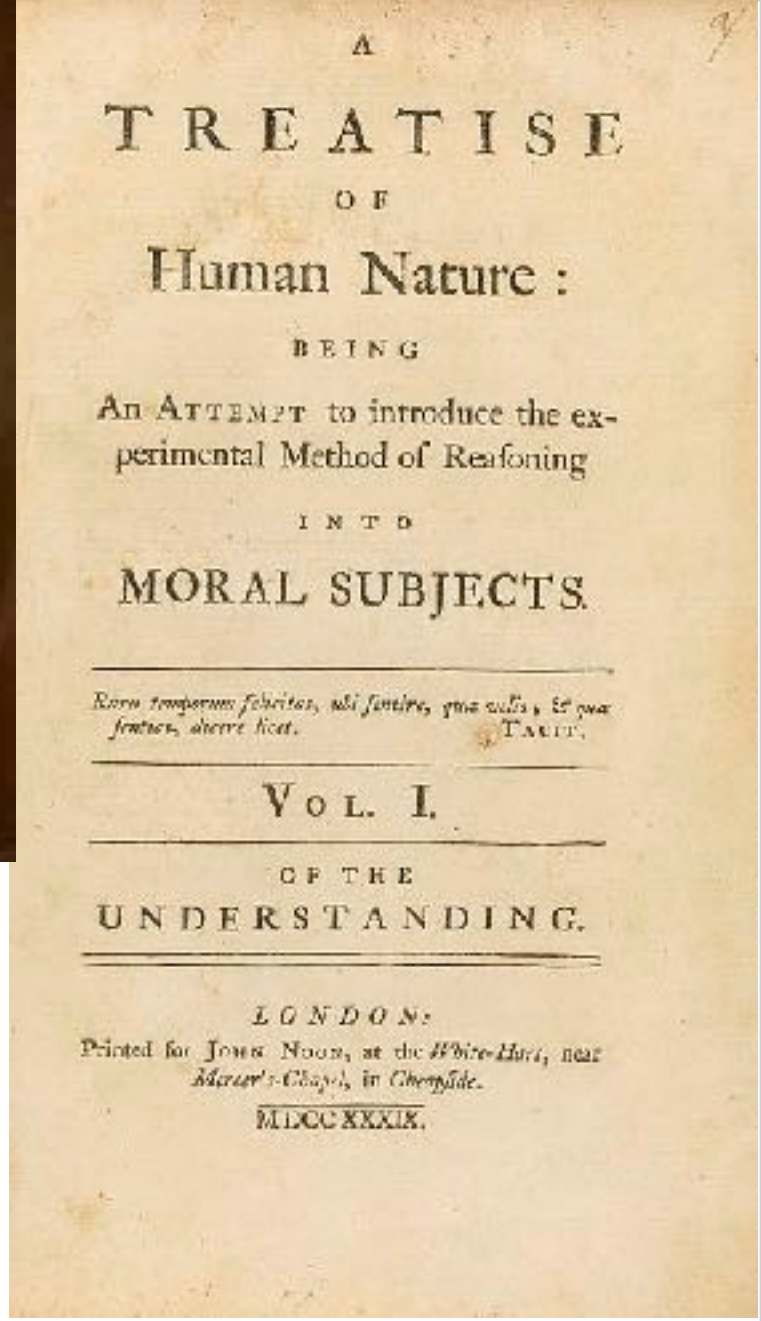
LET US START SIMPLE: STATISTICAL CORRELATIONS



CAUSALITY

?

Hume


A

TREATISE

OF

Human Nature:

BEING

An ATTEMPT to introduce the experimental Method of Reasoning

INTO

MORAL SUBJECTS.

Rara temporum felicitas, ubi sentire, quæ velis, & quæ sentias, dicere licet. TACIT.

VOL. I.

OF THE

UNDERSTANDING.

LONDON:

Printed for JOHN NOON, at the *White-Hart*, near *Mercer's-Chapel*, in *Cheapside*.

MDCCXXXIX.

Knowledge and probability

	Immediate	Inferential
Relations of ideas	intuition	demonstrative reasoning
Matters of fact	perception	probable reasoning

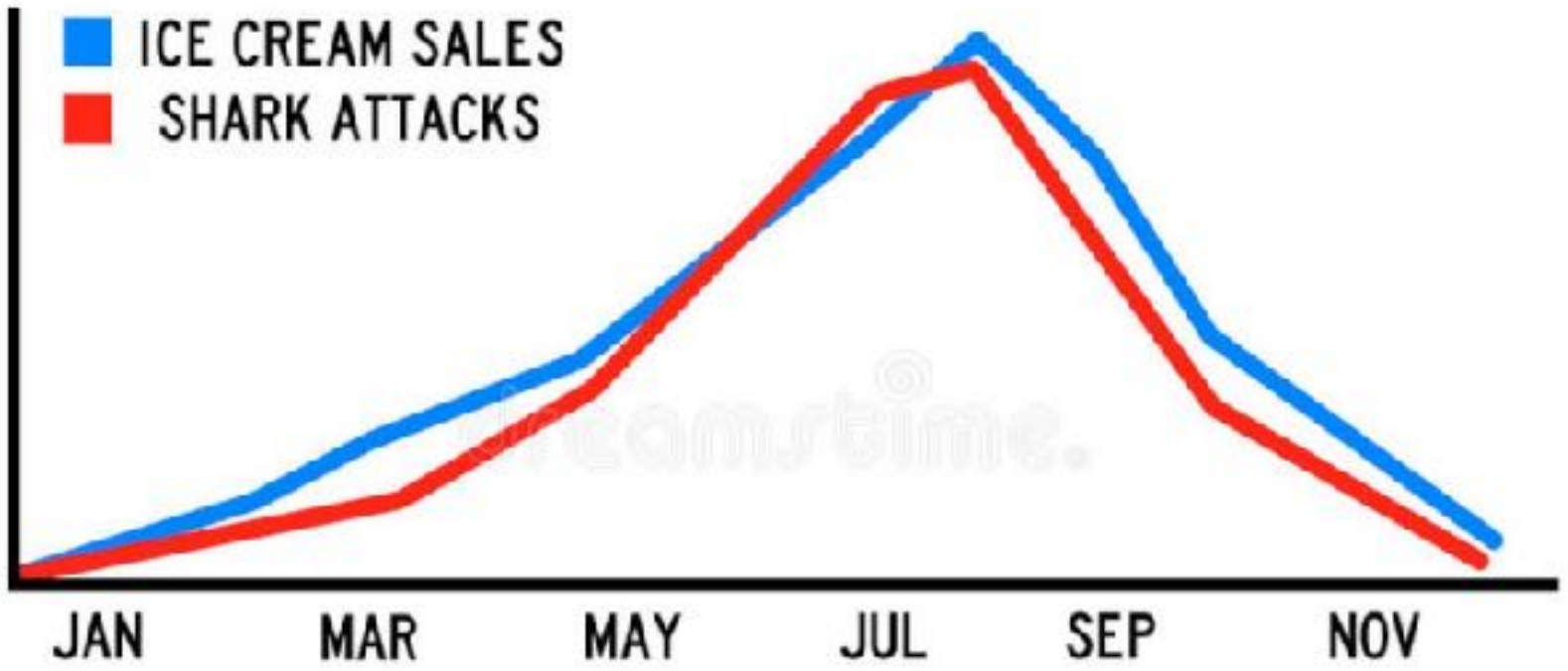
A GLIMPSE AT HUME'S CAUSALITY CRITERIA

1. The same cause produces the same effect.

2. If several objects create the same effect, then there must be a uniting criterion among them causing the effect.

3. If two objects have a different effect, there must be a reason that explains the difference.

DO YOU WANT TO PREDICT OR TO EXPLAIN?

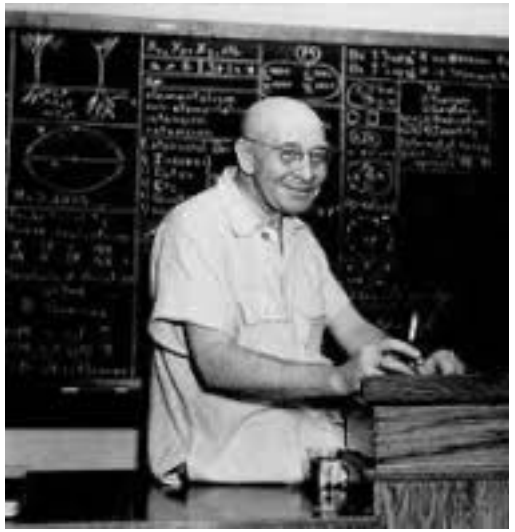


Both ice cream sales and shark attacks increase when the weather is hot and sunny, but they are not caused by each other (they are caused by good weather, with lots of people at the beach, both eating ice cream and having a swim in the sea)

SOME GENERAL LAWS THAT MAY BE VALUABLE

All models are wrong
Some models are useful.

Everything needs to be as simple as possible, and as complex as necessary.

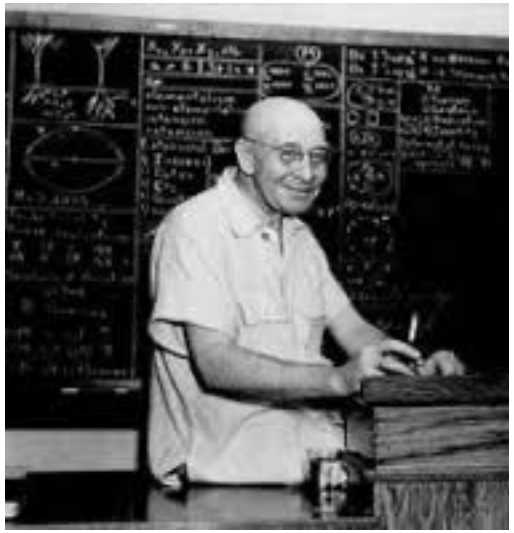


A map is not the territory it represents, but, if correct, it has a similar structure to the territory, which accounts for its usefulness.

WHAT SHOULD MODELS BE?

Georg Box

William of Ockham



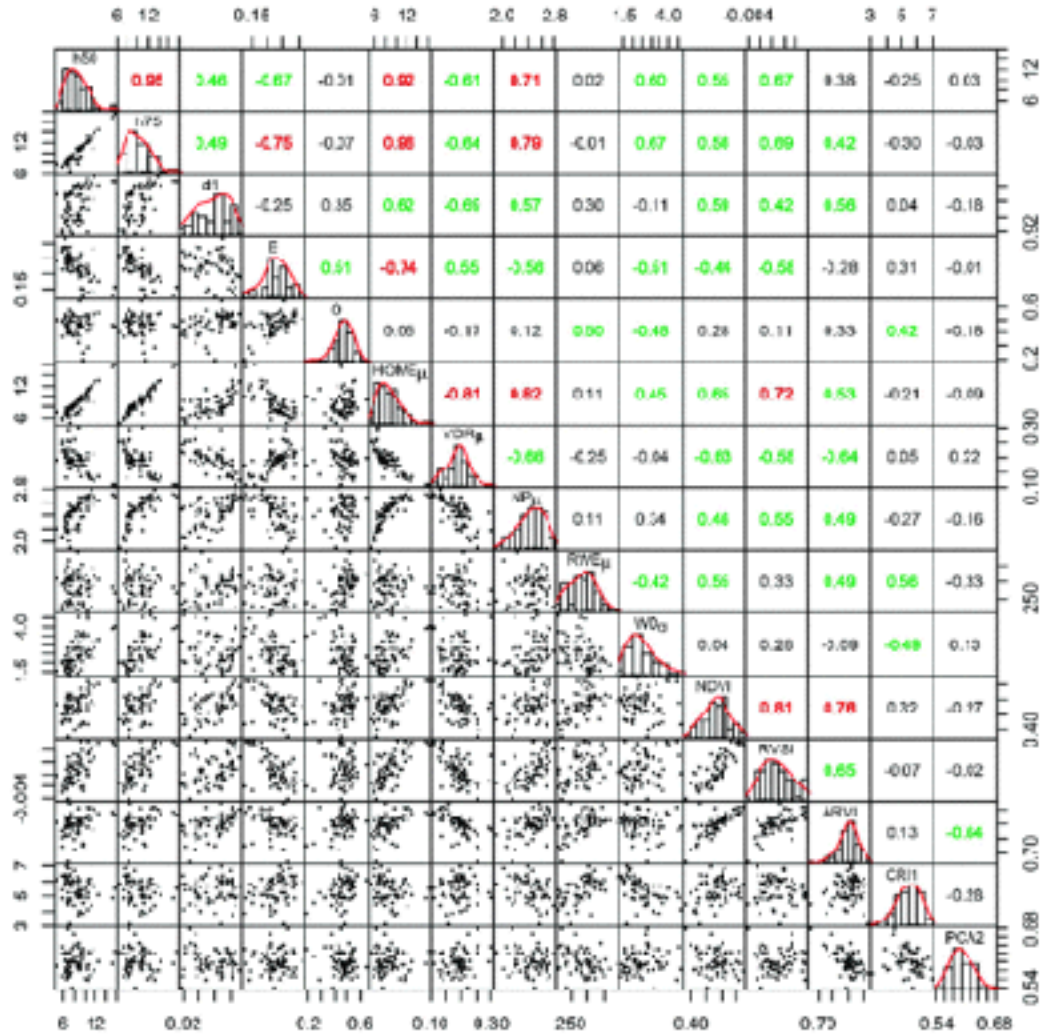
Alfred Korzybski

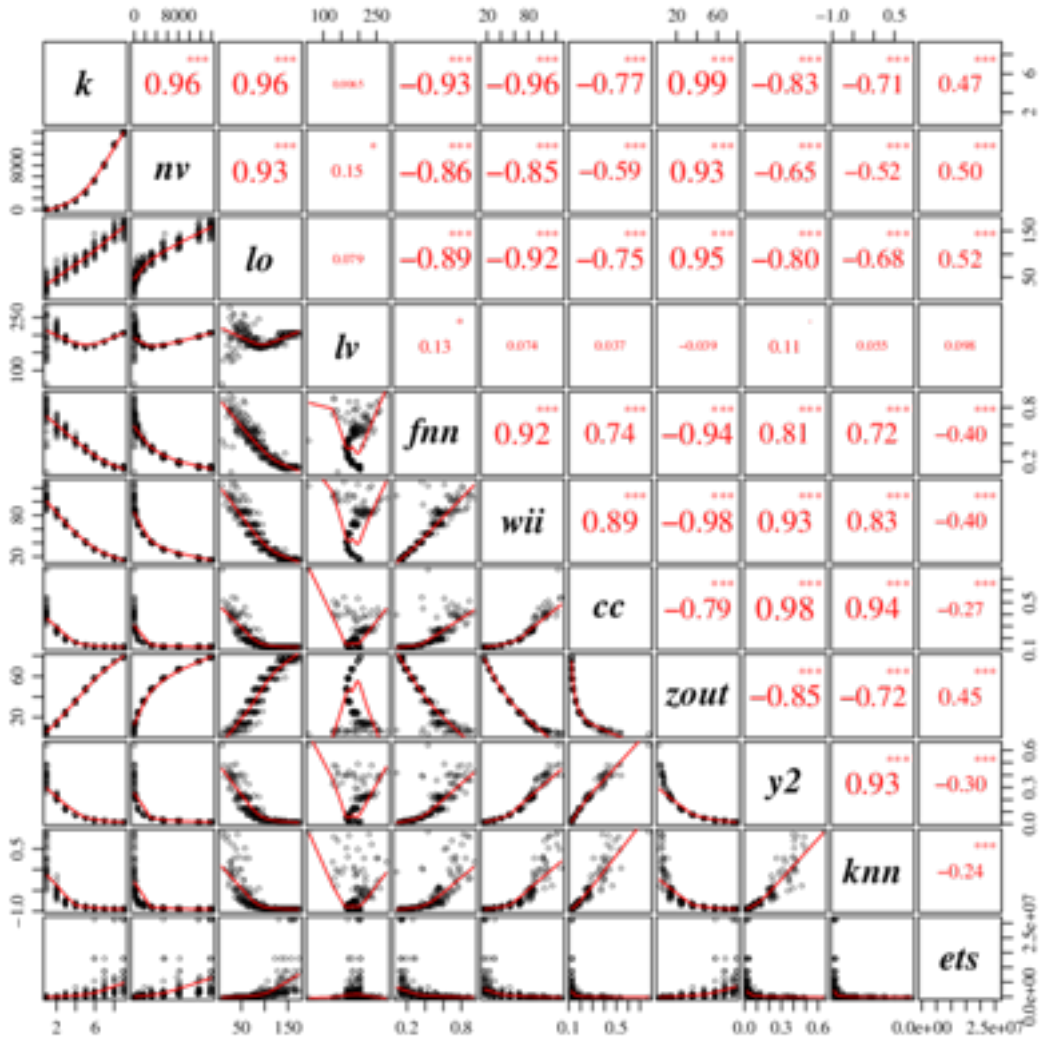
Approximations?

Parsimonious!

Generalisation?

AFTER TEN THOUSANDS OF PLOTS AND MODELS, INTUITION





CAUSALITY OR CORRELATION? ONE SUMMARY.

- In a century of numbers, correlations mattered
- Correlation can predict, and may even help to explain
- Correlation models are generalisation, approximations, and (ideally) parsimonious
- Causality can be altogether different thing
- Suggestions of causality are rooted in logic
- Whether you may ultimately understand relations between to continuous variables is a matter of theory

CONTACT

Henrik von Wehrden | Faculty of Sustainability
Universitätsallee 1 | 21335 Lüneburg
Fon 04131.677-1671 | henrik.von_wehrden@leuphana.de
https://henrikvonwehrden.web.leuphana.de