ANOVA (analysis of variance)

Table 10.12 ANOVA, one factor, more than two treatments

Item	Description
Input	a samples: $x_{11}, x_{12}, \ldots x_{1n_1}; x_{21}, x_{22}, \ldots x_{2n_2}; \ldots; x_{a1}, x_{a2}, \ldots x_{an_a}$
H_0	$\mu_{x_1} = \mu_{x_2} = \ldots = \mu_{x_a}$, i.e. all expected means are equal
Calculations	$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n_i} x_{ij}^2 - \frac{x_{..}^2}{N}$ $SS_{Treatment} = \sum_{i=1}^{a} \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{N}$ $SS_{Error} = SS_T - SS_{Treatment}$ $MS_{Treatment} = SS_{Treatment}/(a-1)$ $MS_{Error} = SS_{Error}/(N-a)$ $F_0 = MS_{Treatment}/MS_{Error}$ where N is the total number of measurements and a dot index denotes a summation over the dotted index, e.g. $x_{i.} = \sum_j x_{ij}$
Criterion	Reject H_0 if $F_0 > F_{\alpha,a-1,N-a}$. Here, F_{α,f_1,f_2} is the upper α percentage point of the F distribution with f_1 and f_2 degrees of freedom, which is tabulated, for example, in Table B.5 and by Montgomery [125] and Marascuilo and Serlin [119]

Source: Experimentation in Software Engineering, Springer, 2012

ANOVA (analysis of variance)

Table 10.13 ANOVA table for the ANOVA test described above

Source of variation	Sum of squares	Degrees of freedom	Mean square	F_0
Between treatments	$SS_{Treatment}$	$a-1$	$MS_{Treatment}$	$F_0 = \frac{MS_{Treatment}}{MS_{Error}}$
Error[a]	SS_{Error}	$N-a$	MS_{Error}	
Total	SS_T	$N-1$		

[a] This is sometimes denoted within treatments

ANOVA in MATLAB

- The modules sizes in three different programs have been measured.
  - P = {221, 159, 191, 194, 156, 238, 220, 197, 197, 194, 200, 198}
  - Q = {173, 171, 168, 286, 206, 140, 226, 248, 189, 208, 213, 205}
  - R = {234, 188, 181, 207, 266, 153, 190, 195, 181, 238, 191, 260}

```
> P = [221 159 191 194 156 238 220 197 197 194 200 198].';
> Q = [173 171 168 286 206 140 226 248 189 208 213 205].';
> R = [234 188 181 207 266 153 190 195 181 238 191 260].';
> [p, TABLE] = anova2([P Q R])
    p = 0.7730 0.5692
    TABLE = 'Source'     'SS'            'df'    'MS'            'F'          'Prob>F'
            'Columns'    [   594.0556]   [ 2]    [   297.0278]   [0.2604]     [0.7730]
            'Rows'       [1.1077e+04]    [11]    [1.0070e+03]    [0.8830]     [0.5692]
            'Error'      [2.5090e+04]    [22]    [1.1405e+03]          []           []
            'Total'      [3.6761e+04]    [35]              []          []           []
```

Kruskal-Wallis test

Table 10.15 Kruskal-Wallis

Item	Description
Input	a samples: $x_{11}, x_{12}, \ldots x_{1n_1}; x_{21}, x_{22}, \ldots x_{2n_2}; \ldots; x_{a1}, x_{a2}, \ldots x_{an_a}$
H_0	The population medians of the a samples are equal.
Calculations	All measures are ranked in one series $(1, 2, \ldots n_1 + n_2 + \ldots + n_a)$, and the calculations are based on these ranks. See for example [119, 157].
Criterion	See, for example, Siegel and Castellan [157] and Marascuilo and Serlin [119]

Source: Experimentation in Software Engineering, Springer, 2012

Kruskal-Wallis in MATLAB

- The modules sizes in three different programs have been measured.
  - P = {221, 159, 191, 194, 156, 238, 220, 197, 197, 194, 200, 198}
  - Q = {173, 171, 168, 286, 206, 140, 226, 248, 189, 208, 213, 205}
  - R = {234, 188, 181, 207, 266, 153, 190, 195, 181, 238, 191, 260}

```
> P = [221 159 191 194 156 238 220 197 197 194 200 198].';
> Q = [173 171 168 286 206 140 226 248 189 208 213 205].';
> R = [234 188 181 207 266 153 190 195 181 238 191 260].';
> [p, TABLE] = kruskalwallis([P Q R])
    p = 0.9792
    TABLE = 'Source'     'SS'            'df'    'MS'              'F'         'Prob>F'
            'Columns'    [     4.6667]   [ 2]    [     2.3333]     [0.0421]    [0.9792]
            'Error'      [3.8778e+03]    [33]    [   117.5101]           []          []
            'Total'      [3.8825e+03]    [35]               []           []          []
```

pairwise analysis

- comparing treatments in pairs to judge which pair is statistically significantly different

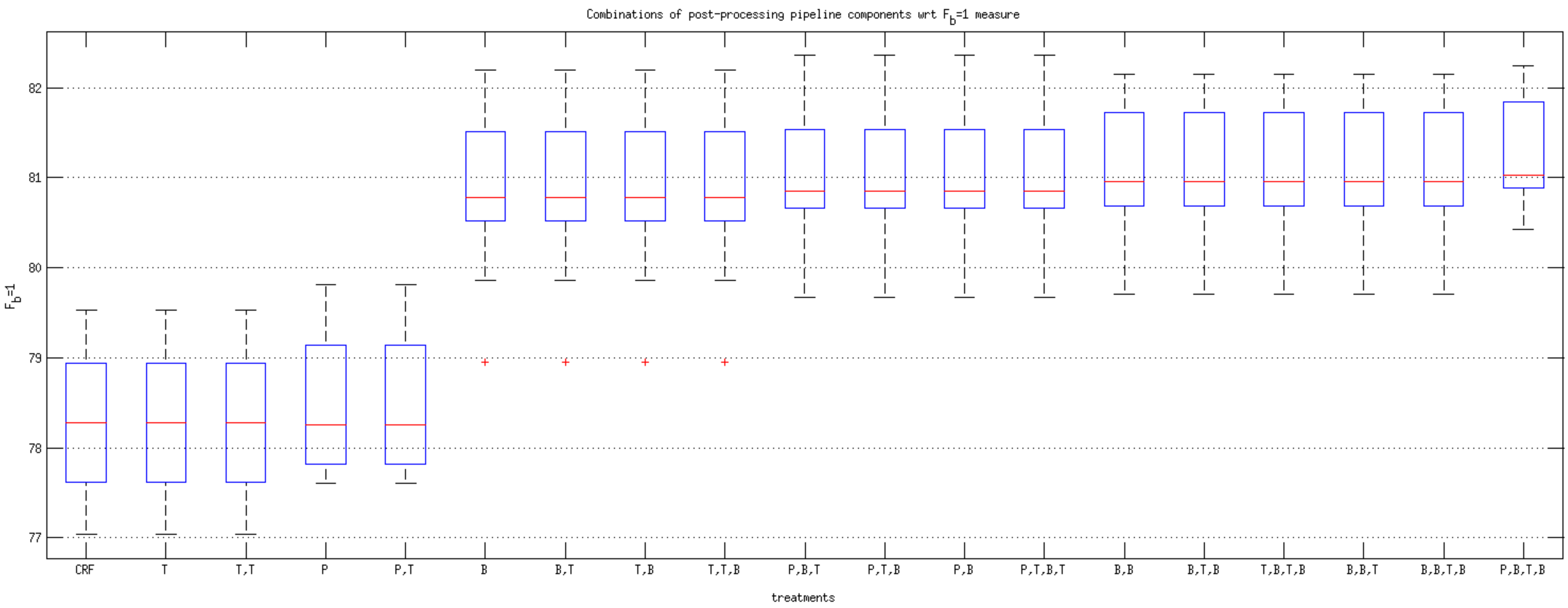
Michele Filannino

pairwise analysis

Shuffling	Fold	CRF	T	T,T	P	P,T	B	B,T	T,B	T,T,B	P,B,T	P,T,B	P,B	P,T,B,T	B,B	B,T,B	T,B,T,B	B,B,T	B,B,T,B	P,B,T,B
0	0	78.47	78.47	78.47	78.51	78.51	80.91	80.91	80.91	80.91	80.94	80.94	80.94	80.94	80.93	80.93	80.93	80.93	80.93	80.91
	1	78.96	78.96	78.96	79.14	79.14	81.51	81.51	81.51	81.51	81.67	81.67	81.67	81.67	81.72	81.72	81.72	81.72	81.72	81.84
	2	79.53	79.53	79.53	79.81	79.81	82.19	82.19	82.19	82.19	82.36	82.36	82.36	82.36	82.11	82.11	82.11	82.11	82.11	82.24
	3	78.10	78.10	78.10	78.27	78.27	80.52	80.52	80.52	80.52	80.66	80.66	80.66	80.66	80.69	80.69	80.69	80.69	80.69	80.83
	4	78.45	78.45	78.45	78.24	78.24	81.66	81.66	81.66	81.66	81.53	81.53	81.53	81.53	82.15	82.15	82.15	82.15	82.15	82.01
	5	77.04	77.04	77.04	77.7	77.7	78.95	78.95	78.95	78.95	79.67	79.67	79.67	79.67	79.71	79.71	79.71	79.71	79.71	80.43
	6	78.94	78.94	78.94	79.22	79.22	81.07	81.07	81.07	81.07	81.34	81.34	81.34	81.34	81.49	81.49	81.49	81.49	81.49	81.71
	7	77.42	77.42	77.42	77.82	77.82	79.86	79.86	79.86	79.86	80.37	80.37	80.37	80.37	80.39	80.39	80.39	80.39	80.39	80.89
	8	77.62	77.62	77.62	77.61	77.61	80.66	80.66	80.66	80.66	80.74	80.74	80.74	80.74	80.98	80.98	80.98	80.98	80.98	81.00
	9	77.74	77.74	77.74	77.95	77.95	80.62	80.62	80.62	80.62	80.76	80.76	80.76	80.76	80.92	80.92	80.92	80.92	80.92	81.06
1	0	78.45	78.45	78.45	78.24	78.24	81.66	81.66	81.66	81.66	81.53	81.53	81.53	81.53	82.15	82.15	82.15	82.15	82.15	82.01
	1	77.42	77.42	77.42	77.82	77.82	79.86	79.86	79.86	79.86	80.37	80.37	80.37	80.37	80.39	80.39	80.39	80.39	80.39	80.89
	2	77.74	77.74	77.74	77.95	77.95	80.62	80.62	80.62	80.62	80.76	80.76	80.76	80.76	80.92	80.92	80.92	80.92	80.92	81.06
	3	78.10	78.10	78.10	78.27	78.27	80.52	80.52	80.52	80.52	80.66	80.66	80.66	80.66	80.69	80.69	80.69	80.69	80.69	80.83
	4	78.96	78.96	78.96	79.14	79.14	81.51	81.51	81.51	81.51	81.67	81.67	81.67	81.67	81.72	81.72	81.72	81.72	81.72	81.84
	5	79.53	79.53	79.53	79.81	79.81	82.19	82.19	82.19	82.19	82.36	82.36	82.36	82.36	82.11	82.11	82.11	82.11	82.11	82.24
	6	77.04	77.04	77.04	77.7	77.7	78.95	78.95	78.95	78.95	79.67	79.67	79.67	79.67	79.71	79.71	79.71	79.71	79.71	80.43
	7	78.94	78.94	78.94	79.22	79.22	81.07	81.07	81.07	81.07	81.34	81.34	81.34	81.34	81.49	81.49	81.49	81.49	81.49	81.71
	8	77.62	77.62	77.62	77.61	77.61	80.66	80.66	80.66	80.66	80.74	80.74	80.74	80.74	80.98	80.98	80.98	80.98	80.98	81.00
	9	78.47	78.47	78.47	78.51	78.51	80.91	80.91	80.91	80.91	80.94	80.94	80.94	80.94	80.93	80.93	80.93	80.93	80.93	80.91
2	0	77.42	77.42	77.42	77.82	77.82	79.86	79.86	79.86	79.86	80.37	80.37	80.37	80.37	80.39	80.39	80.39	80.39	80.39	80.89
	1	78.94	78.94	78.94	79.22	79.22	81.07	81.07	81.07	81.07	81.34	81.34	81.34	81.34	81.49	81.49	81.49	81.49	81.49	81.71
	2	78.47	78.47	78.47	78.51	78.51	80.91	80.91	80.91	80.91	80.94	80.94	80.94	80.94	80.93	80.93	80.93	80.93	80.93	80.91
	3	78.45	78.45	78.45	78.24	78.24	81.66	81.66	81.66	81.66	81.53	81.53	81.53	81.53	82.15	82.15	82.15	82.15	82.15	82.01
	4	77.62	77.62	77.62	77.61	77.61	80.66	80.66	80.66	80.66	80.74	80.74	80.74	80.74	80.98	80.98	80.98	80.98	80.98	81.00
	5	77.04	77.04	77.04	77.7	77.7	78.95	78.95	78.95	78.95	79.67	79.67	79.67	79.67	79.71	79.71	79.71	79.71	79.71	80.43
	6	78.96	78.96	78.96	79.14	79.14	81.51	81.51	81.51	81.51	81.67	81.67	81.67	81.67	81.72	81.72	81.72	81.72	81.72	81.84
	7	79.53	79.53	79.53	79.81	79.81	82.19	82.19	82.19	82.19	82.36	82.36	82.36	82.36	82.11	82.11	82.11	82.11	82.11	82.24
	8	78.10	78.10	78.10	78.27	78.27	80.52	80.52	80.52	80.52	80.66	80.66	80.66	80.66	80.69	80.69	80.69	80.69	80.69	80.83
	9	77.74	77.74	77.74	77.95	77.95	80.62	80.62	80.62	80.62	80.76	80.76	80.76	80.76	80.92	80.92	80.92	80.92	80.92	81.06
3	0	77.74	77.74	77.74	77.95	77.95	80.62	80.62	80.62	80.62	80.76	80.76	80.76	80.76	80.92	80.92	80.92	80.92	80.92	81.06
	1	77.04	77.04	77.04	77.7	77.7	78.95	78.95	78.95	78.95	79.67	79.67	79.67	79.67	79.71	79.71	79.71	79.71	79.71	80.43
	2	79.53	79.53	79.53	79.81	79.81	82.19	82.19	82.19	82.19	82.36	82.36	82.36	82.36	82.11	82.11	82.11	82.11	82.11	82.24
	3	78.45	78.45	78.45	78.24	78.24	81.66	81.66	81.66	81.66	81.53	81.53	81.53	81.53	82.15	82.15	82.15	82.15	82.15	82.01
	4	77.62	77.62	77.62	77.61	77.61	80.66	80.66	80.66	80.66	80.74	80.74	80.74	80.74	80.98	80.98	80.98	80.98	80.98	81.00
	5	78.94	78.94	78.94	79.22	79.22	81.07	81.07	81.07	81.07	81.34	81.34	81.34	81.34	81.49	81.49	81.49	81.49	81.49	81.71
	6	77.42	77.42	77.42	77.82	77.82	79.86	79.86	79.86	79.86	80.37	80.37	80.37	80.37	80.39	80.39	80.39	80.39	80.39	80.89
	7	78.10	78.10	78.10	78.27	78.27	80.52	80.52	80.52	80.52	80.66	80.66	80.66	80.66	80.69	80.69	80.69	80.69	80.69	80.83
	8	78.47	78.47	78.47	78.51	78.51	80.91	80.91	80.91	80.91	80.94	80.94	80.94	80.94	80.93	80.93	80.93	80.93	80.93	80.91
	9	78.96	78.96	78.96	79.14	79.14	81.51	81.51	81.51	81.51	81.67	81.67	81.67	81.67	81.72	81.72	81.72	81.72	81.72	81.84
4	0	77.62	77.62	77.62	77.61	77.61	80.66	80.66	80.66	80.66	80.74	80.74	80.74	80.74	80.98	80.98	80.98	80.98	80.98	81.00
	1	79.53	79.53	79.53	79.81	79.81	82.19	82.19	82.19	82.19	82.36	82.36	82.36	82.36	82.11	82.11	82.11	82.11	82.11	82.24
	2	77.74	77.74	77.74	77.95	77.95	80.62	80.62	80.62	80.62	80.76	80.76	80.76	80.76	80.92	80.92	80.92	80.92	80.92	81.06
	3	78.47	78.47	78.47	78.51	78.51	80.91	80.91	80.91	80.91	80.94	80.94	80.94	80.94	80.93	80.93	80.93	80.93	80.93	80.91
	4	77.42	77.42	77.42	77.82	77.82	79.86	79.86	79.86	79.86	80.37	80.37	80.37	80.37	80.39	80.39	80.39	80.39	80.39	80.89
	5	77.04	77.04	77.04	77.7	77.7	78.95	78.95	78.95	78.95	79.67	79.67	79.67	79.67	79.71	79.71	79.71	79.71	79.71	80.43
	6	78.10	78.10	78.10	78.27	78.27	80.52	80.52	80.52	80.52	80.66	80.66	80.66	80.66	80.69	80.69	80.69	80.69	80.69	80.83
	7	78.45	78.45	78.45	78.24	78.24	81.66	81.66	81.66	81.66	81.53	81.53	81.53	81.53	82.15	82.15	82.15	82.15	82.15	82.01
	8	78.94	78.94	78.94	79.22	79.22	81.07	81.07	81.07	81.07	81.34	81.34	81.34	81.34	81.49	81.49	81.49	81.49	81.49	81.71
	9	78.96	78.96	78.96	79.14	79.14	81.51	81.51	81.51	81.51	81.67	81.67	81.67	81.67	81.72	81.72	81.72	81.72	81.72	81.84
Average		**78.22700**	**78.22700**	**78.22700**	**78.42700**	**78.42700**	**80.79500**	**80.79500**	**80.79500**	**80.79500**	**81.00400**	**81.00400**	**81.00400**	**81.00400**	**81.10900**	**81.10900**	**81.109000**	**81.10900**	**81.10900**	**81.29200**
St. Deviation		**0.751475**	**0.751475**	**0.751475**	**0.707713**	**0.707713**	**0.886235**	**0.886235**	**0.886235**	**0.886235**	**0.721750**	**0.721750**	**0.721750**	**0.721750**	**0.738054**	**0.738054**	**0.738054**	**0.738054**	**0.738054**	**0.57962**

MANCHESTER 1824
The University of Manchester

pairwise analysis



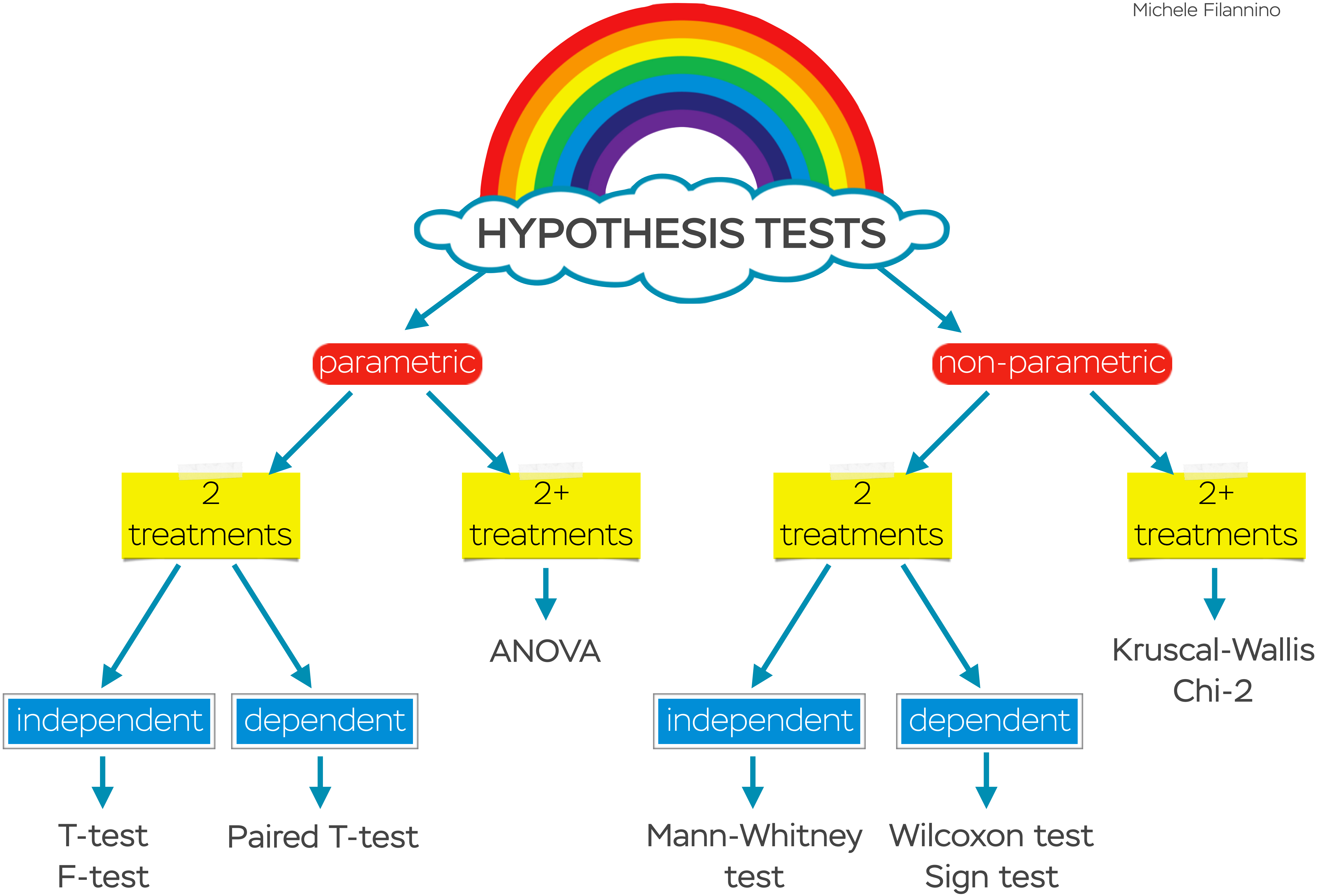
MANCHESTER 1824
The University of Manchester

pairwise analysis

treatments	1	2	3	4	5	6	7	8	...
1	-	p-value	p-value	p-value	p-value	p-value	p-value	p-value	p-value
2	-	-	p-value	p-value	p-value*	p-value	p-value	p-value*	p-value
3	-	-	-	p-value	p-value	p-value	p-value	p-value	p-value
4	-	-	-	-	p-value*	p-value	p-value*	p-value	p-value
5	-	-	-	-	-	p-value	p-value	p-value	p-value
6	-	-	-	-	-	-	p-value	p-value	p-value
7	-	-	-	-	-	-	-	p-value	p-value
8	-	-	-	-	-	-	-	-	p-value
...	-	-	-	-	-	-	-	-	-

Note: Indicate what test has been used for the pairwise assessment.

Michele Filannino



MANCHESTER 1824
The University of Manchester

drawing conclusions

- if we can reject H_0
  - given that the experiment is valid
  - influence of the independent vars. on the dependent ones
  - can we generalise it?
- else
  - we have shown that there is no statistical difference between treatments

Note: Sometimes (the else branch) is a very nice thing to show!

joke: p-value

Note: This is just a joke.

low p-value <> happiness

- We show that method X is statistically significantly 2% more cost effective than method Y.
- Is changing from method X to method Y still cost effective?

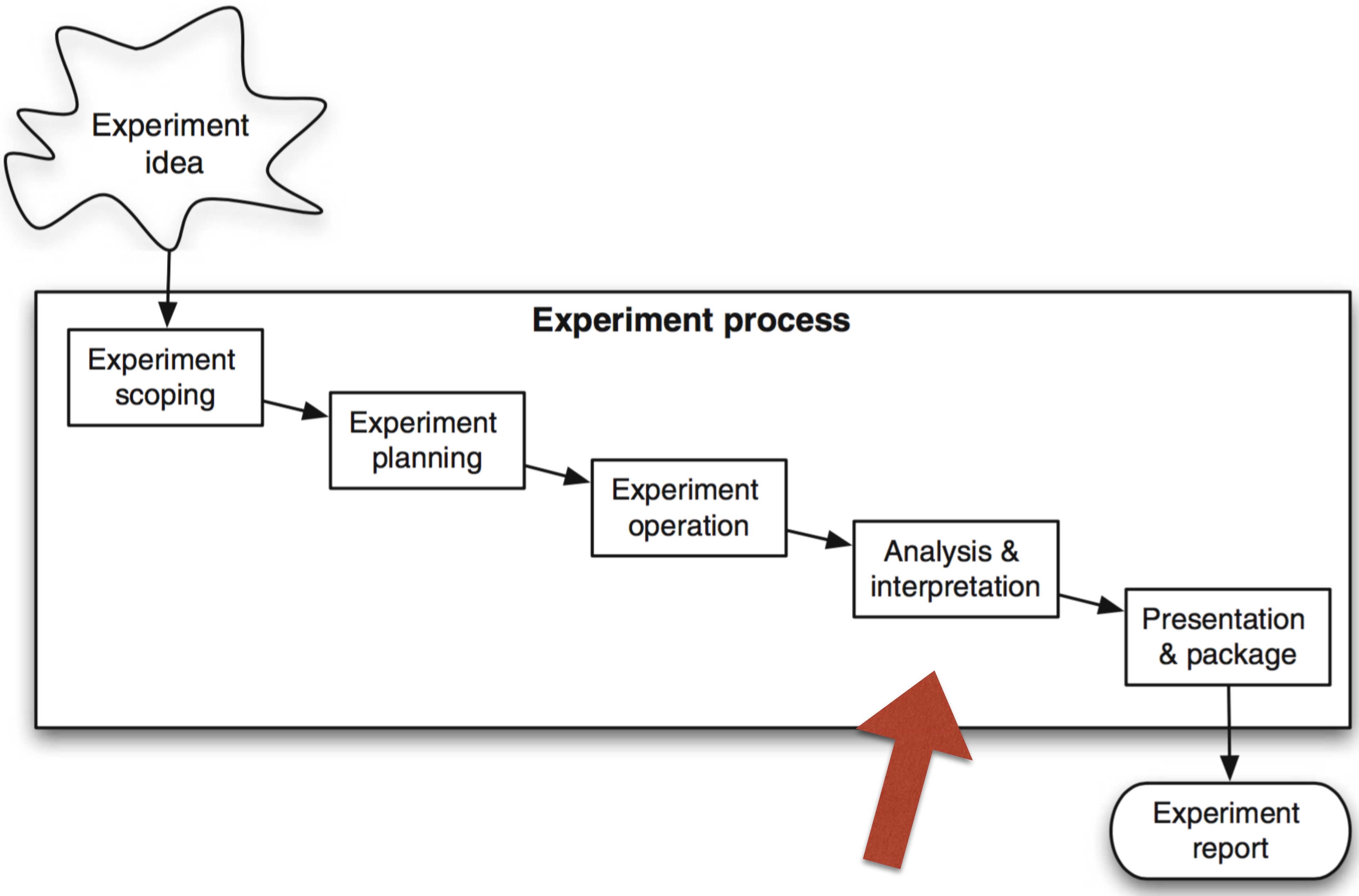
Michele Filannino

joke: p-value

Note: This is just a joke.

high p-value <> sadness

- non-statistically-significant results may still be of practical importance
  - the experiment may not be well designed
  - the samples may be too few



Source: Experimentation in Software Engineering, Springer, 2012

different audiences

- conference paper
- journal paper
- report for decision-makers
- package for replication of the experiment
- educational material

structured abstract

- background or context
- objectives or aims
- method
- results
- conclusions

motivation

- set the scope
- define the objectives
- primarily reports the outcome of the scoping phase
  - statement (why?)

related work

- provides the picture
- how this experiment is related to the previous work?
- not necessarily need a systematic literature review
  - although extremely beneficial!

experimental design

- outcome of the planning phase
  - detailed description of the hypotheses (null and alternative)
  - experimental design (type, variables, tools)
  - data gathering technique[s]
  - characterisation of the subjects
  - threats to validity (all the 4 types)

execution

- how the operation is prepared
- description that will ease the replication
- validation procedures used on the data
    - outliers? how many? why? what? when?

analysis

- graphs
- sample sizes
- independent vs. dependent
- parametric vs. non-parametric
- hypothesis testing choice
    - assumptions
    - confidence levels

interpretation

- hypothesis testing outcome
  - reject of accept H_0?
- interpretation
  - practical importance vs. statistical significance
  - so what?

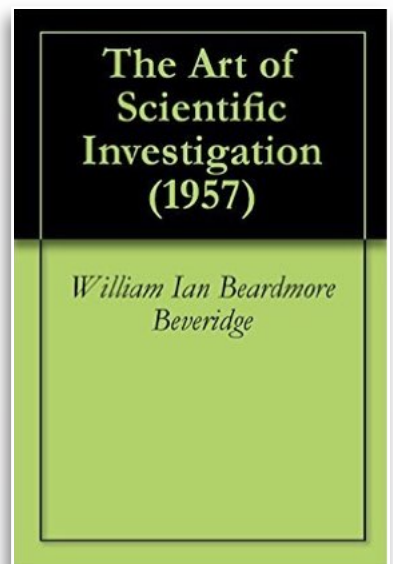
conclusions and further work

- summary of findings
- possible limitations
  - limits of the generalisability (write them!)
- relate to the work reported in the “related works”
  - similarities, differences
- references to the resources for replication
  - “To aid replicability of this work, the source code, the B and the C are available at [URL]”

appendices

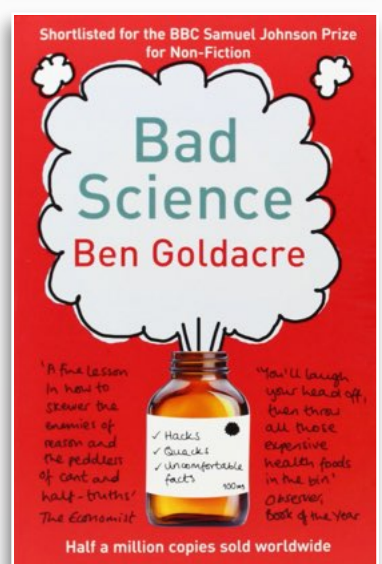
- collected data
- technical details about the execution
- more info about subjects and objects
- links to resources
  - GitHub, BitBucket …
- screenshots of tools
  - online DEMO

sources: science



- **The art of scientific investigation,**

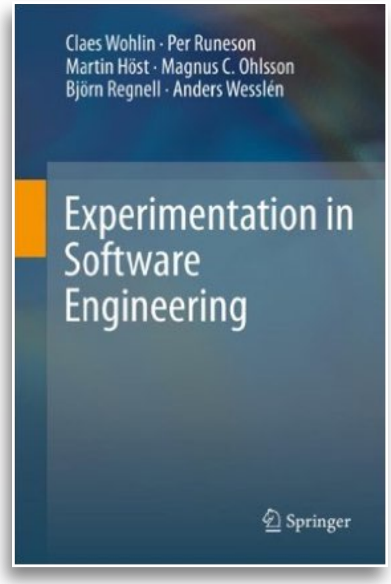
  Beveridge, Blackburn press, 1957

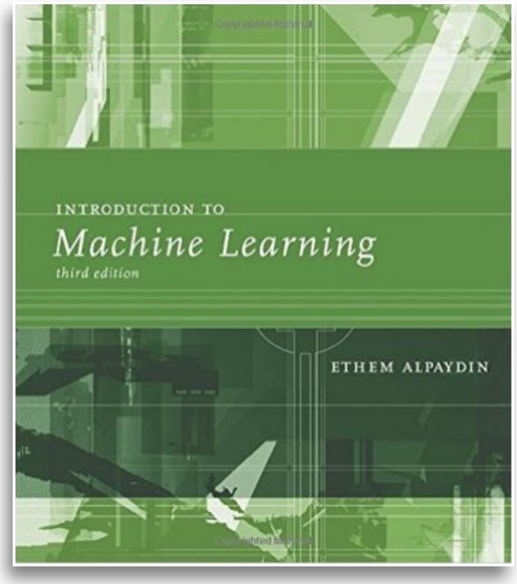


- **Bad Science,**

  Goldrace, Fourth Estate, 2009

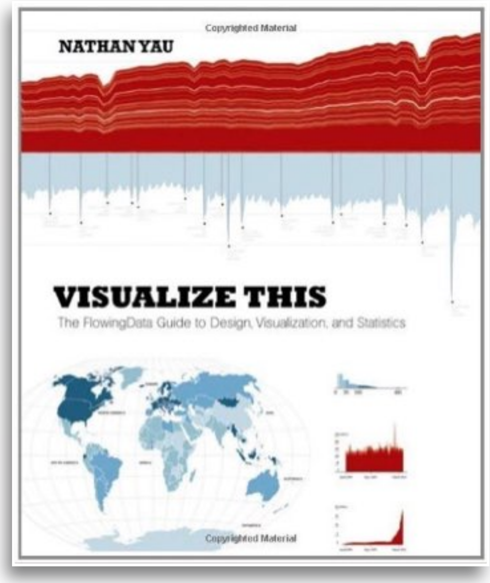
sources: experimentation





- **Experimentation in software engineering**, Wohlin et al., Springer, 2012 (Chap. 7-11)
- **Introduction to Machine Learning**, Alpaydin, MIT press, 2004 (Chap. 19)

sources: visualisation



- **Visualize this,**

  Yau, John Wiley & Sons, 2011

sources: people

- Giuseppe Visaggio
  SERlab, University of Bari, Italy

- Danilo Caivano
  SERlab, University of Bari, Italy

- Teresa Baldassarre
  SERlab, University of Bari, Italy

Note: They brilliantly and passionately taught me 75% of what I've covered in this course. ❤️

thanks ;)