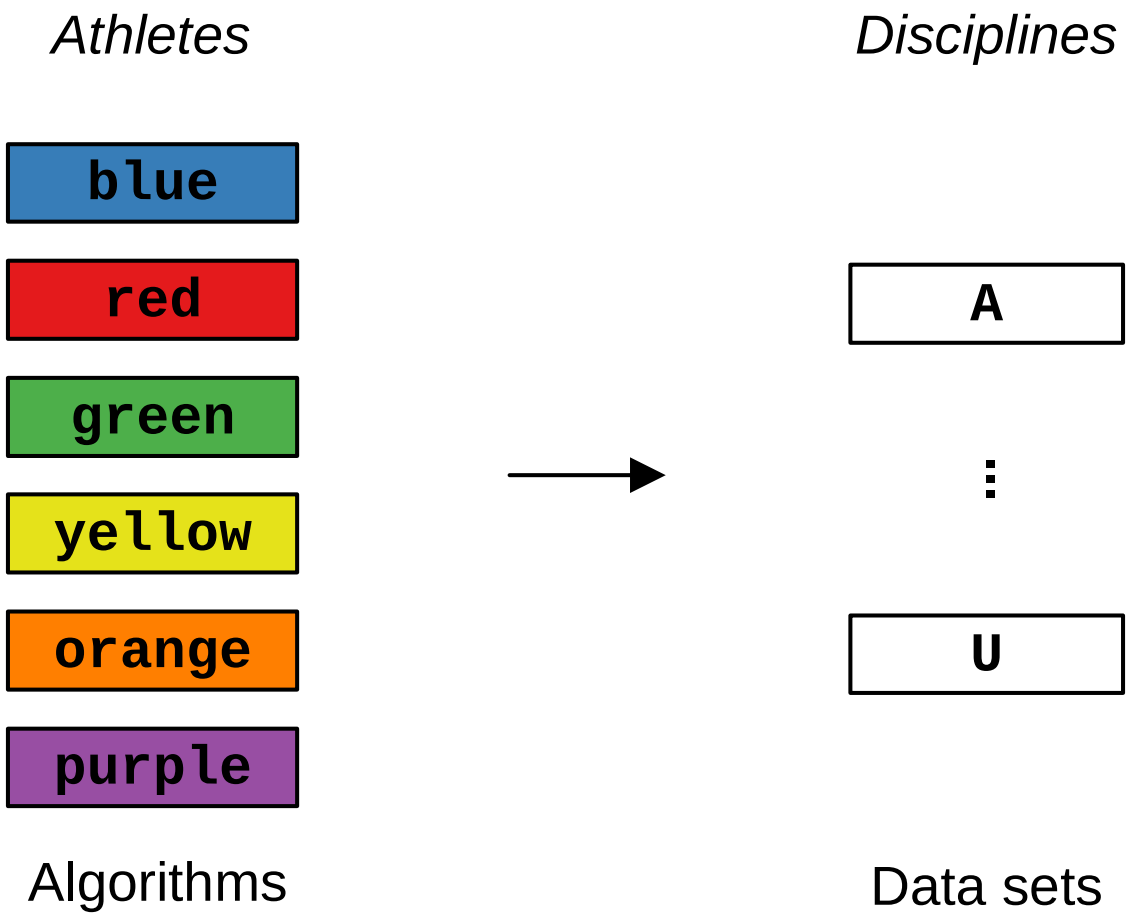
Visualisation of Benchmark Experiments

Manuel J. A. Eugster and Friedrich Leisch

Departement for Statistics
Ludwig-Maximiliams-Universität München

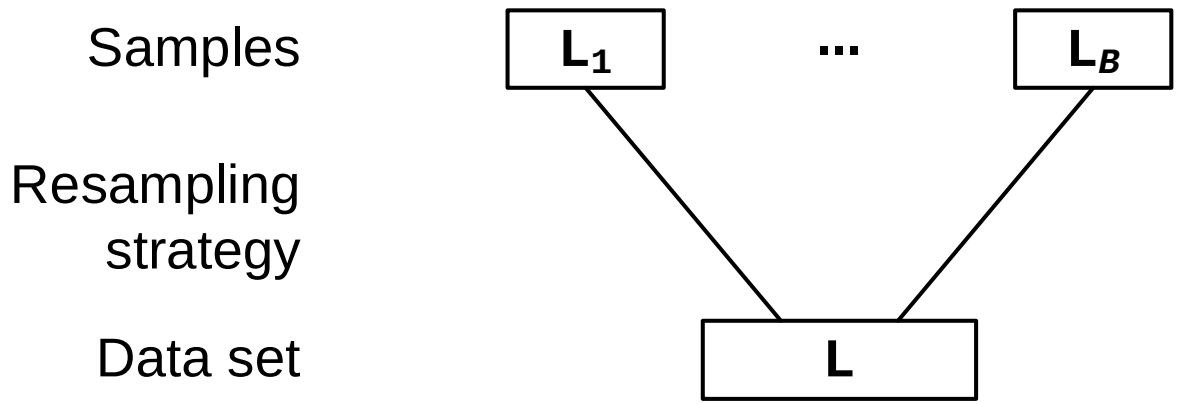
Statistical Graphics: Data and Information Visualization in Today's Multimedia Society, 2008

“Statistical decathlon”

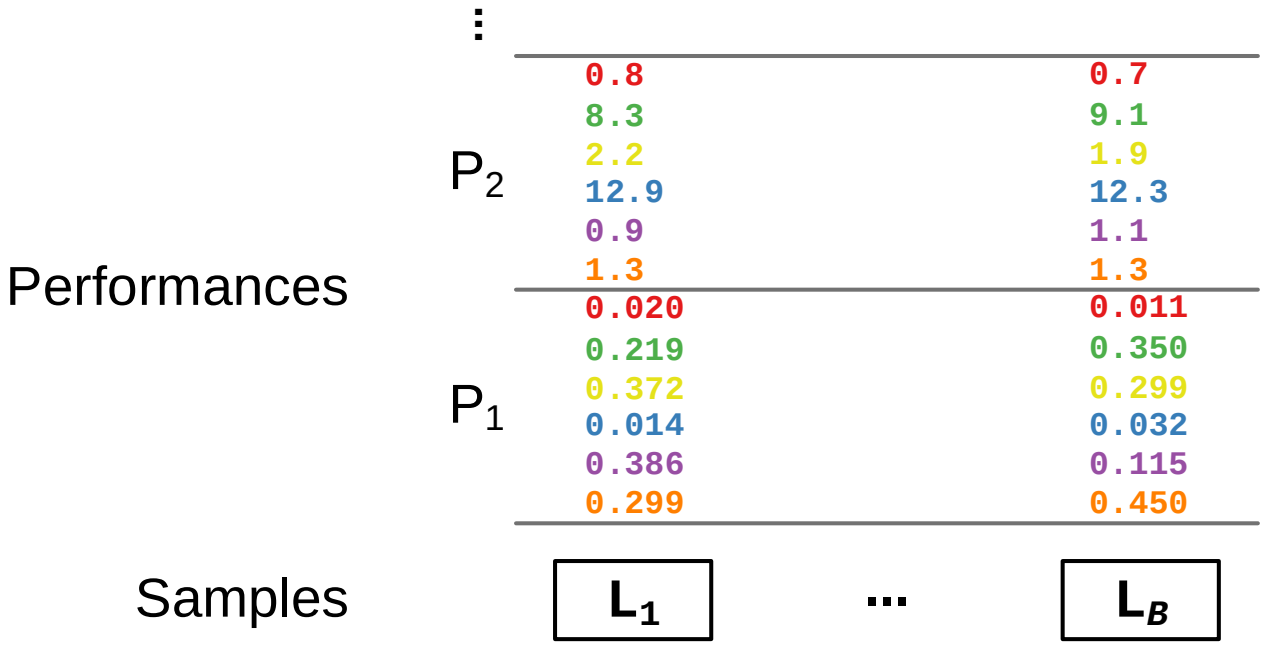


Layers of abstraction

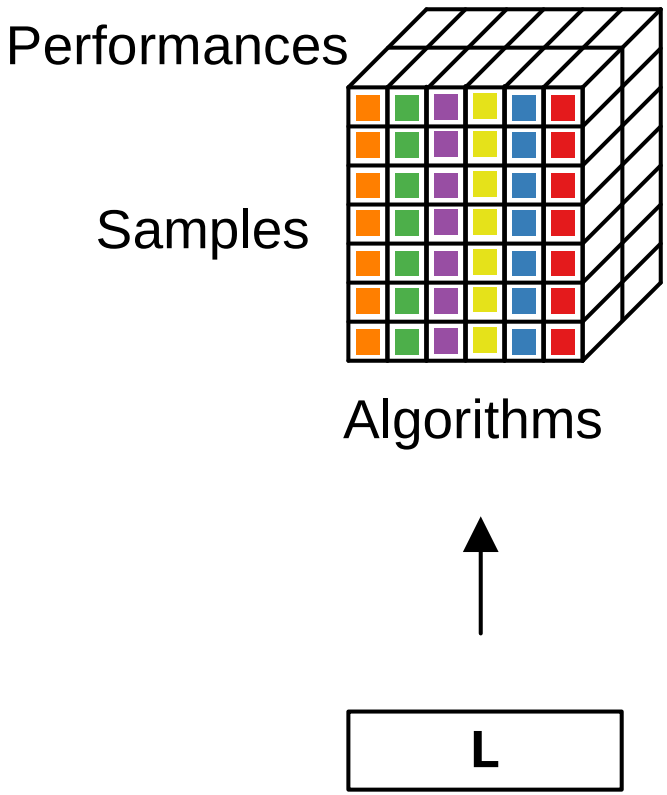
Layer One: Setup



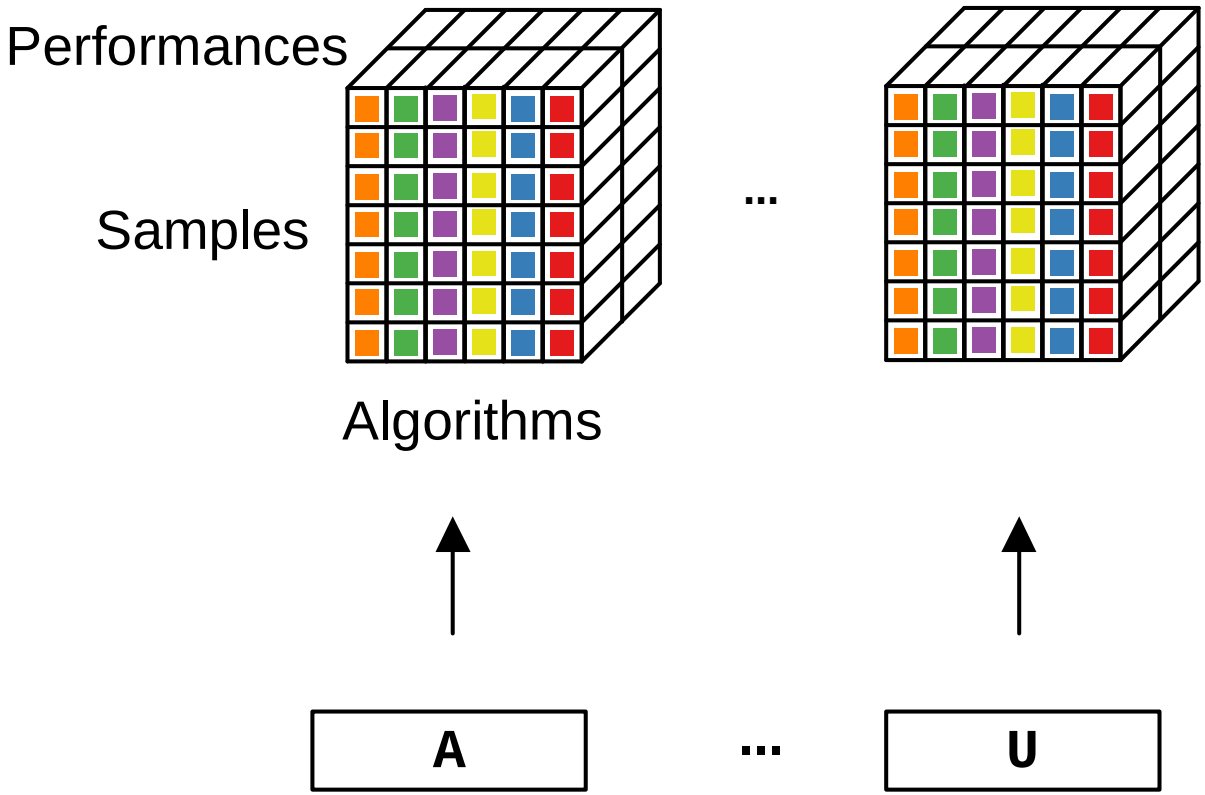
Layer Two: Execution (1)



Layer Two: Execution (2)



Layer Two: Execution (3)



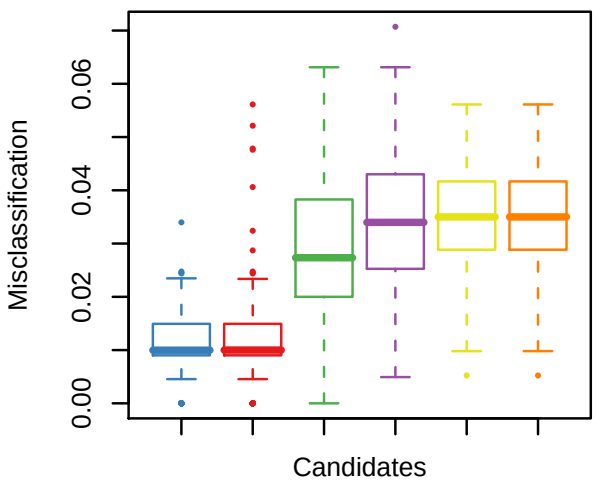
Layer Three: Analysis

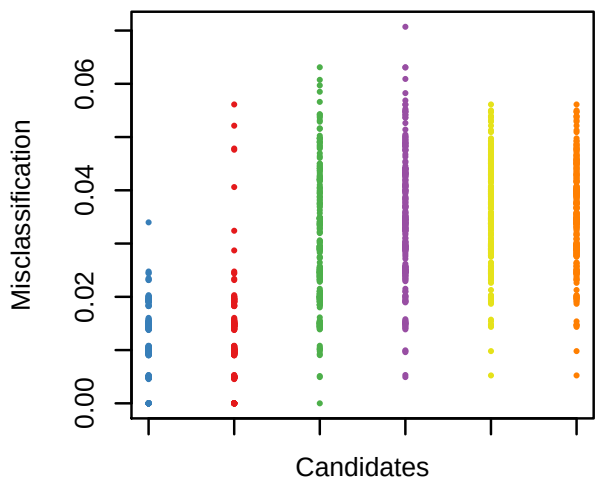
Exploratory: get a better understanding of the benchmark experiment, "dig" for interesting information.

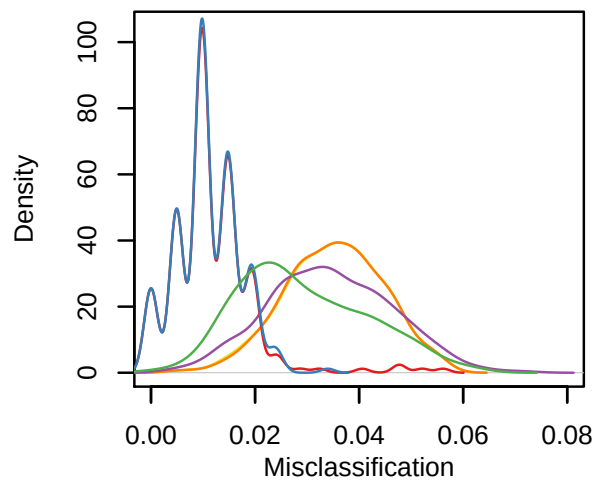
Inferential: test hypotheses of interest, infer a statistically correct order.

Exploratory Analysis: Visualisation

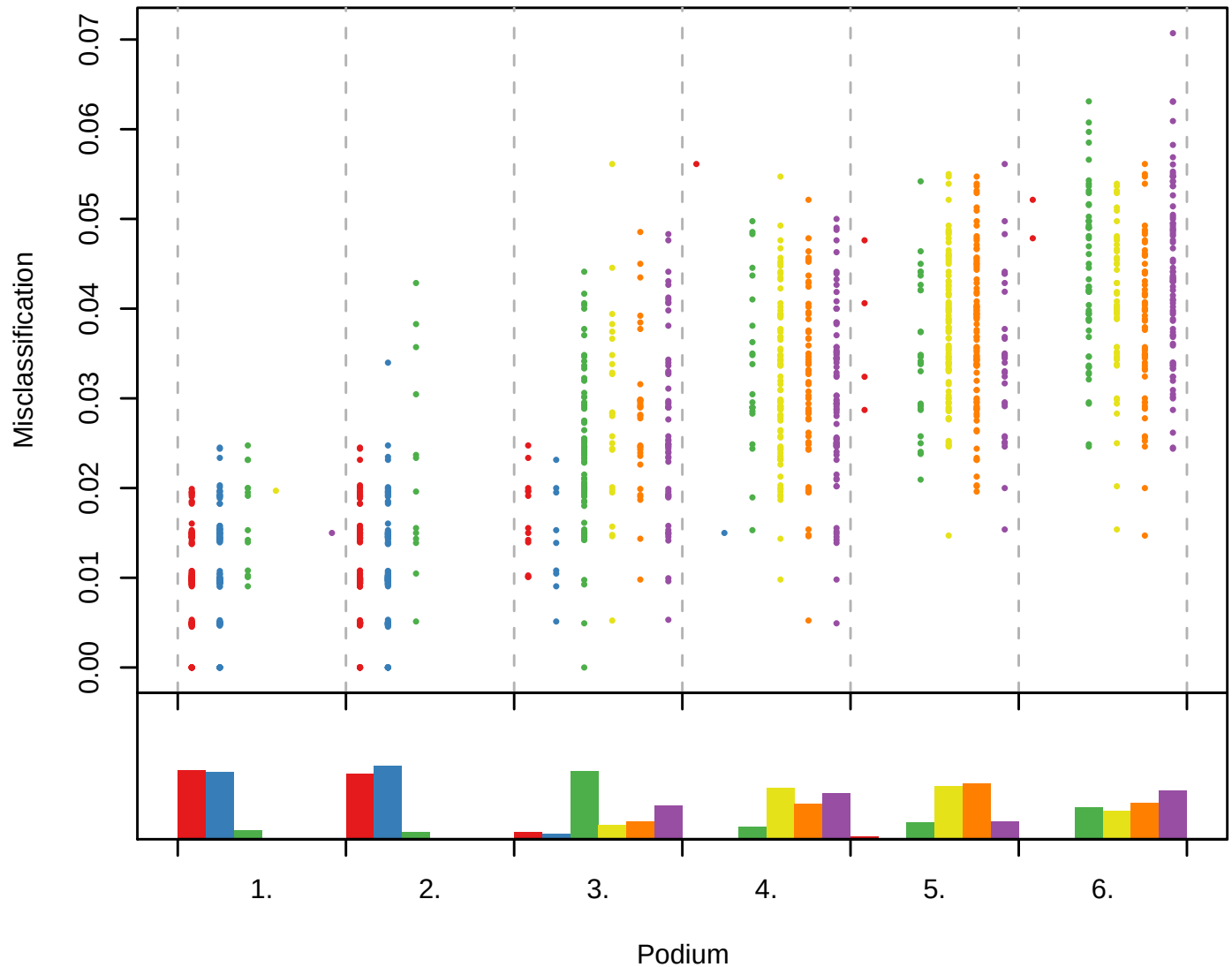
Common visualisation tools



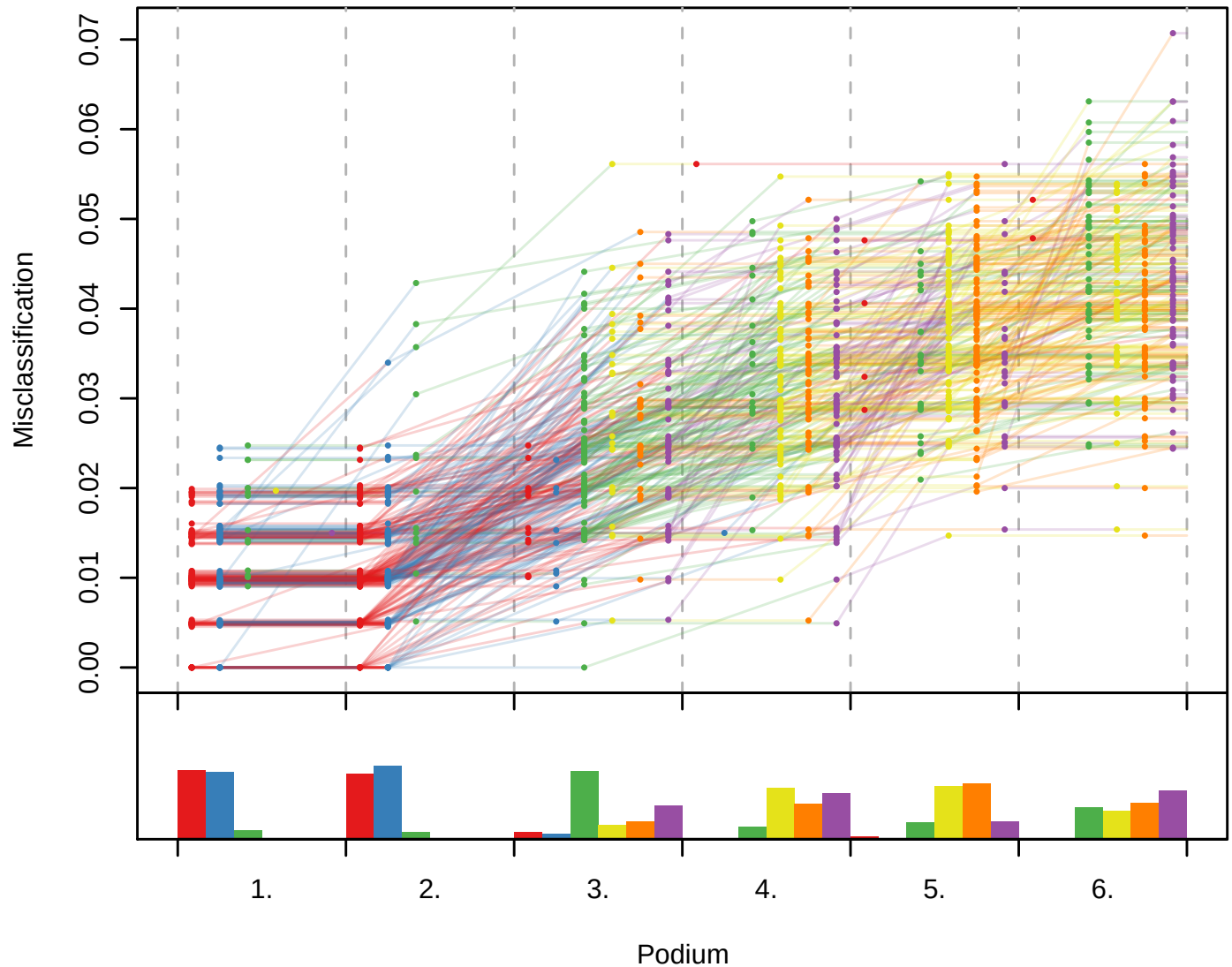




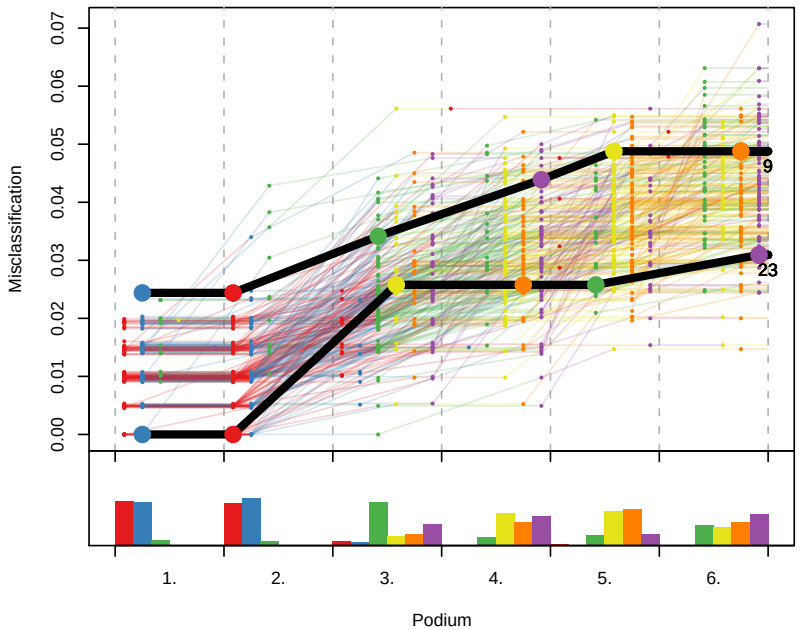
Benchmark experiment plot



“Full” benchmark experiment plot



Matter for deeper analyses ...



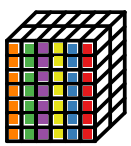
... are the *bootstrap samples* different?

... are the *models* different?

Formal analysis

Statistically correct order*:

Modelling of the *random block design*.

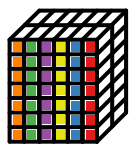


$\rightarrow$ $\texttt{blue} \approx \texttt{red} < \texttt{green} < \texttt{yellow} \approx \texttt{orange} \approx \texttt{purple}$

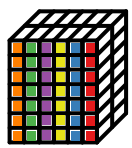
Formal analysis

Statistically correct order*:

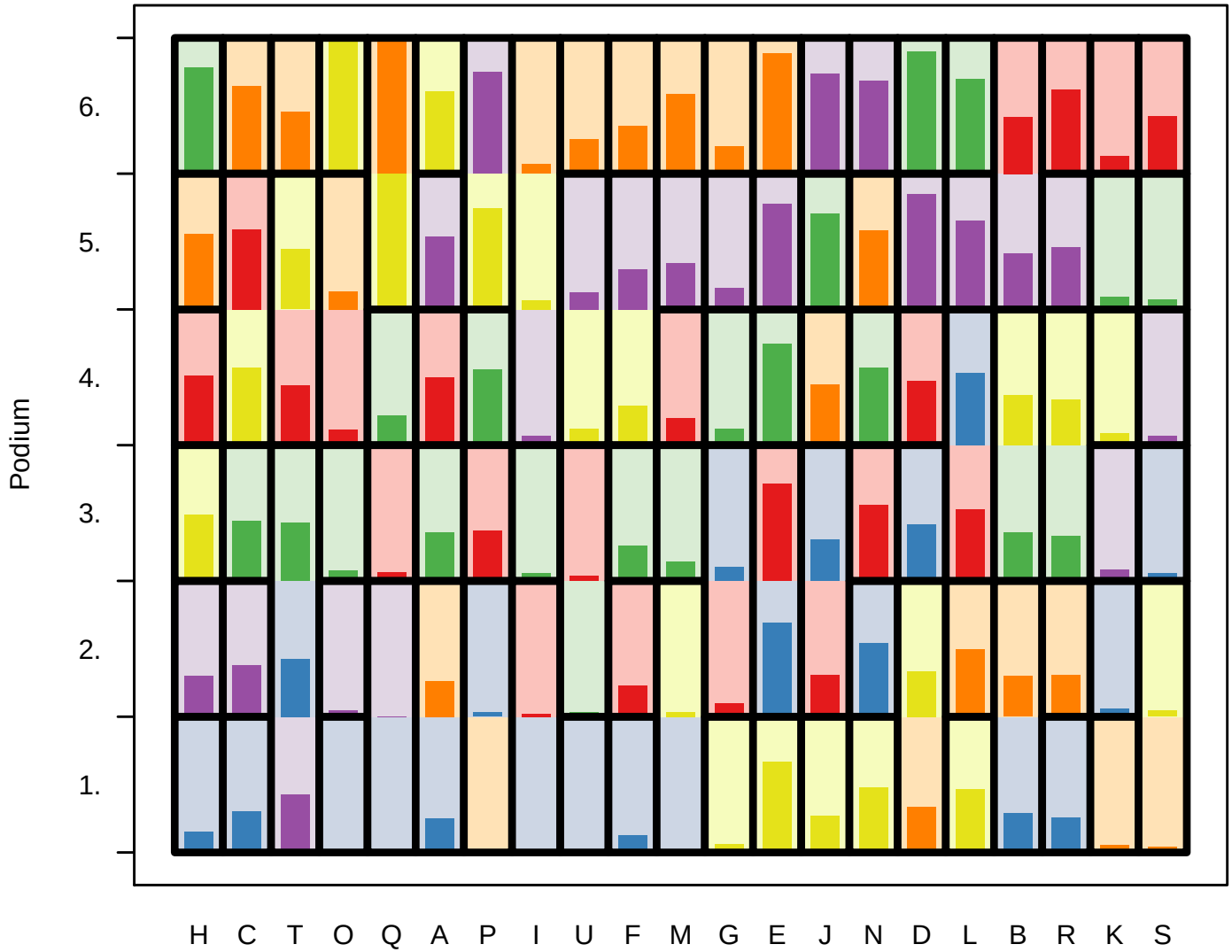
Modelling of the *random block design*.

A  → $\texttt{blue} < \texttt{purple} < \texttt{yellow} < \texttt{red} < \texttt{orange} \approx \texttt{green}$

⋮

U  → $\texttt{orange} \approx \texttt{yellow} < \texttt{blue} < \texttt{purple} < \texttt{green} \approx \texttt{red}$

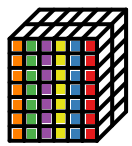
Benchmark summary plot



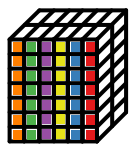
Overall order

Consensus ranking***:**

Aggregation of "preferences of voters" with all order relations equally relevant.

A  → $\texttt{blue} < \texttt{purple} < \texttt{yellow} < \texttt{red} < \texttt{orange} \approx \texttt{green}$

⋮

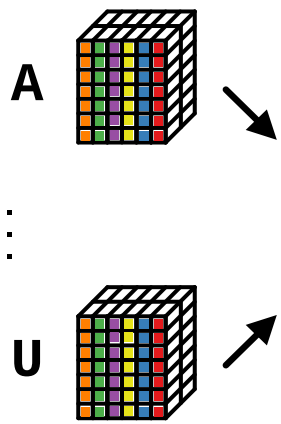
U  → $\texttt{orange} \approx \texttt{yellow} < \texttt{blue} < \texttt{purple} < \texttt{green} \approx \texttt{red}$

$$\texttt{blue} \approx \texttt{yellow} < \texttt{red} < \texttt{green} < \texttt{purple} < \texttt{orange}$$

Overall order

Statistical modelling***:**

Modelling of the *design with two experimental factors, their interactions and blocking factors at two levels.*



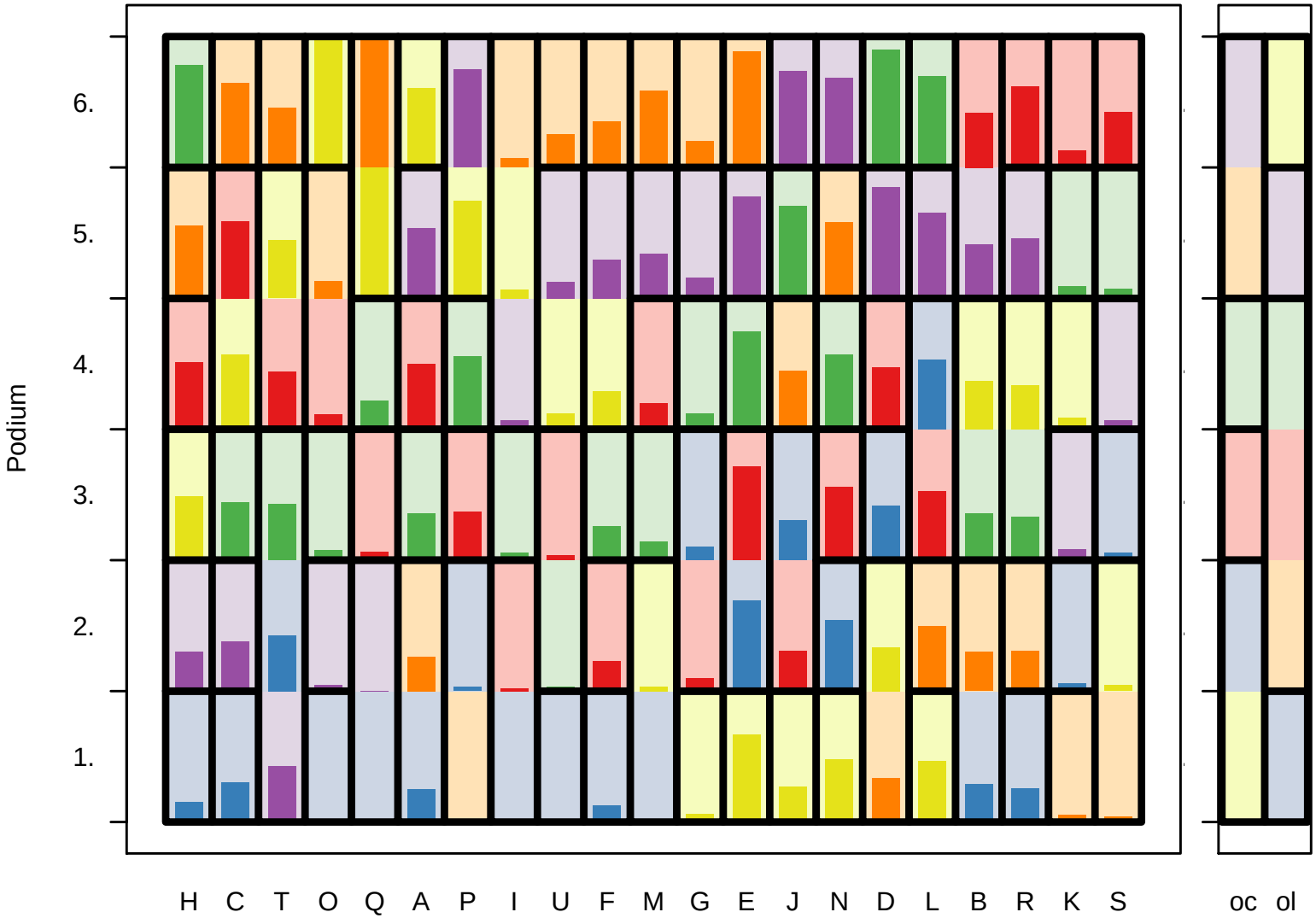
$\texttt{blue} < \texttt{red} \approx \texttt{orange} \approx \texttt{green} < \texttt{purple} < \texttt{yellow}$

Overall order

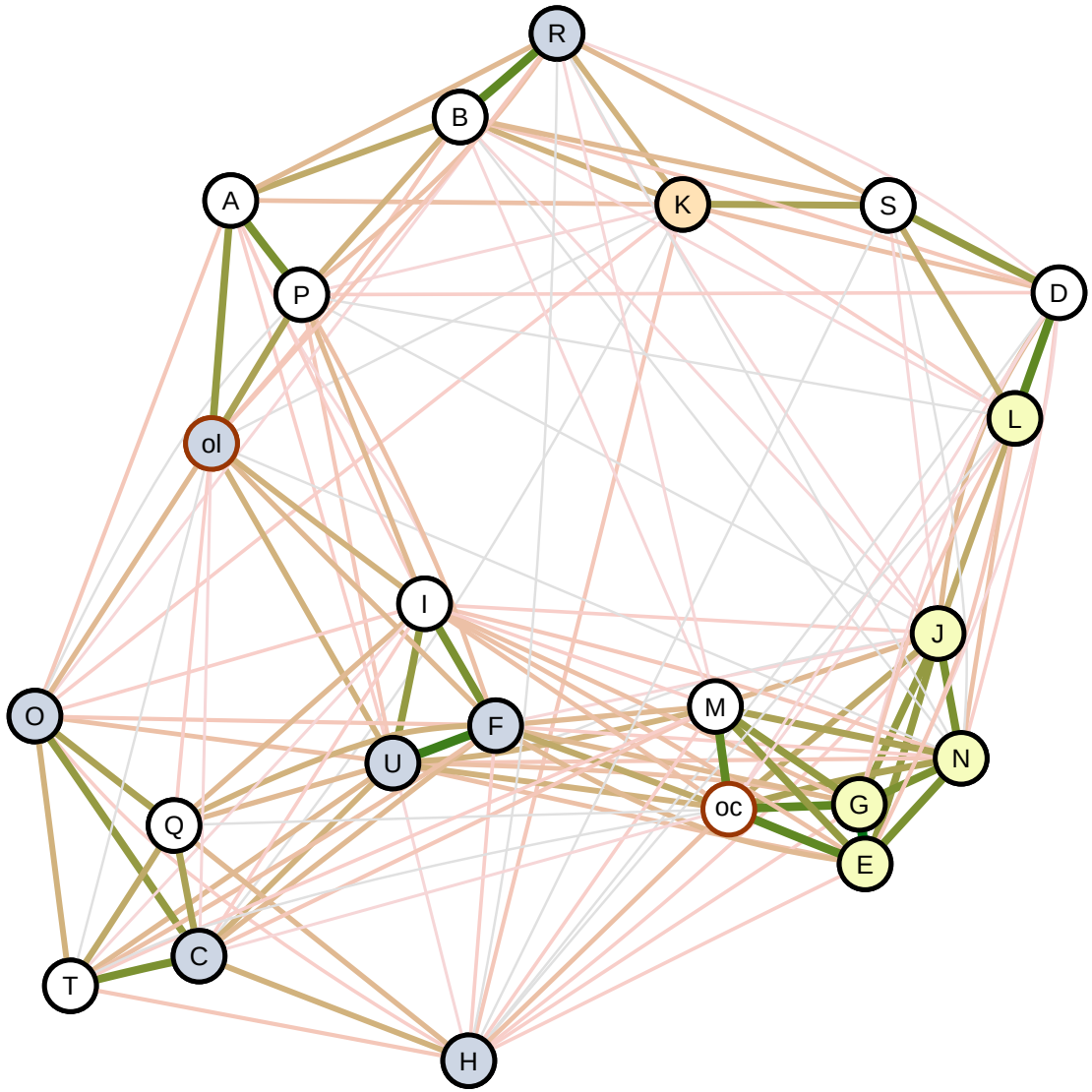
$$\text{Consensus}: \texttt{blue} \approx \texttt{yellow} < \texttt{red} < \texttt{green} < \texttt{purple} < \texttt{orange}$$

$$\text{LME}: \texttt{blue} < \texttt{red} \approx \texttt{orange} \approx \texttt{green} < \texttt{purple} < \texttt{yellow}$$

Overall order interpretation



Overall order interpretation



End titles

Goals and outlook

Ultimate goal: develop a comprehensive toolbox for exploratory and inferential analysis of benchmark experiments; set a quasi-standard for the comparison of learning methods.

Visualisation: interactivity, interactivity, interactivity.

References

Bench Plot and Mixed Effects Models: First steps toward a comprehensiv benchmark analysis toolbox.
Manuel J. A. Eugster and Friedrich Leisch. Technical Report 26, LMU Munich.
Accepted for the Compstat 2008-Proceedings in Computational Statistics.

(*) *Exploratory and Inferential Analysis of Benchmark Experiments*.
Manuel J. A. Eugster, Torsten Hothorn and Friedrich Leisch. Technical Report 30, LMU Munich.

`http://www.statistik.lmu.de/~eugster`