

The tour

Hadley Wickham

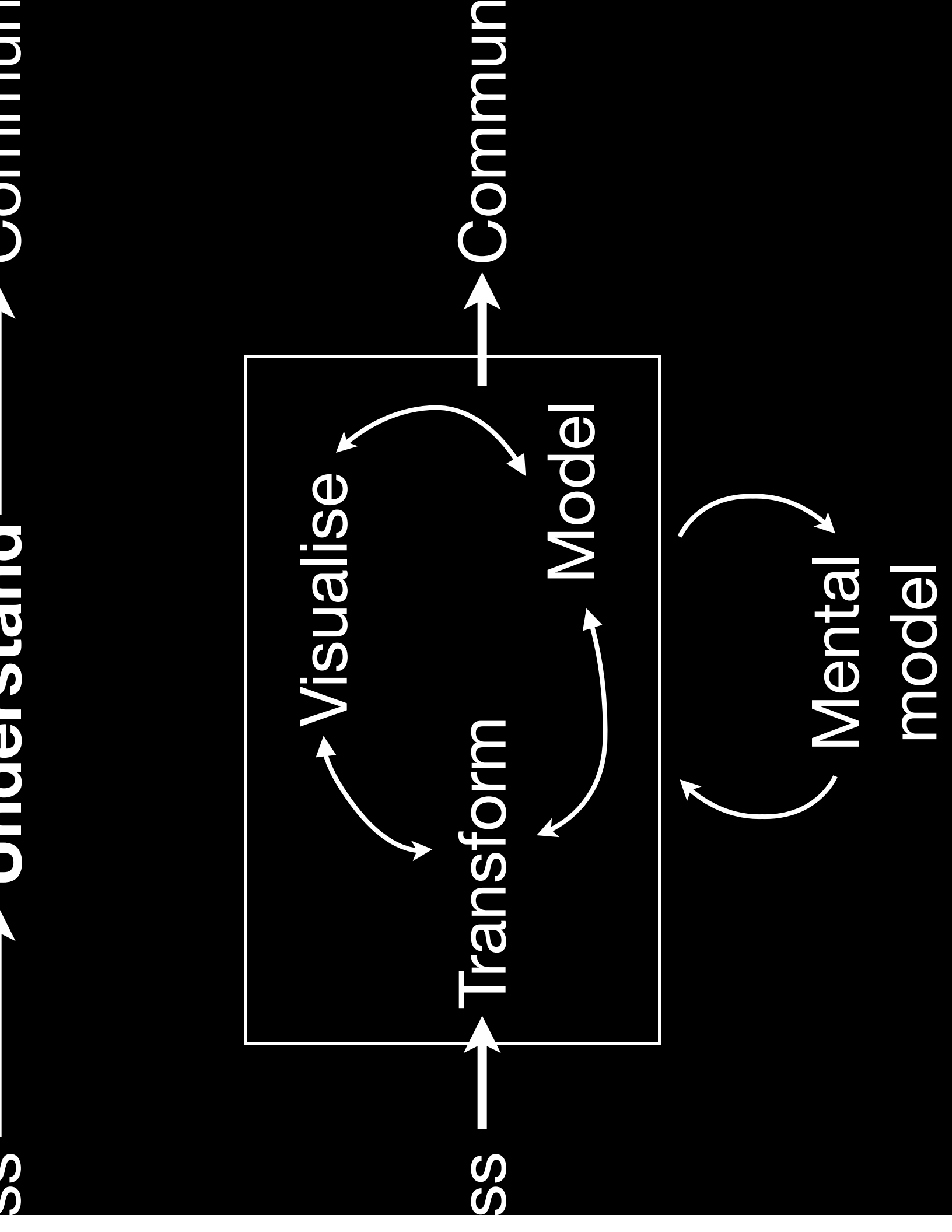
Assistant Professor / Dobelman Family Junior Chair
Department of Statistics / Rice University

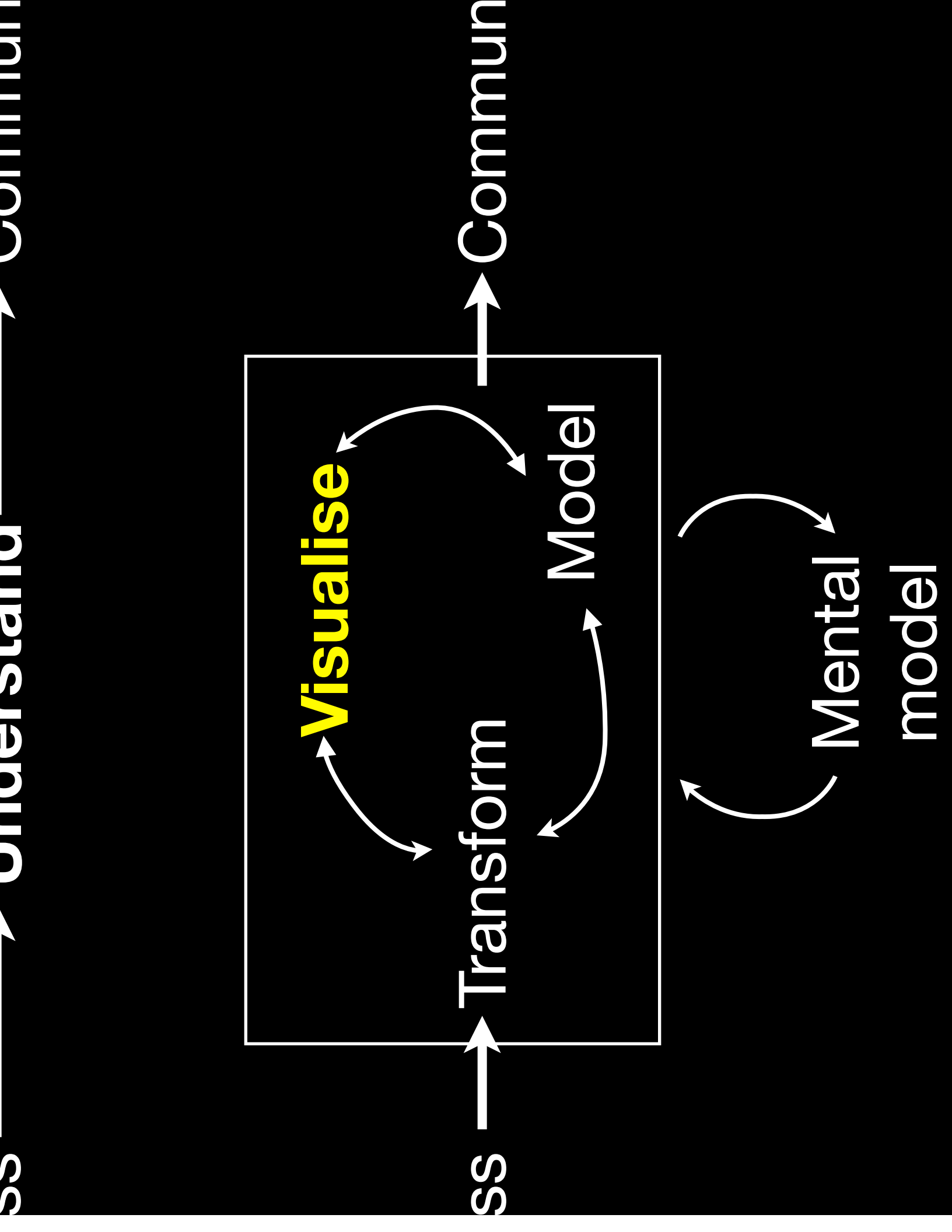
[tp://had.co.nz/pasi-to](http://had.co.nz/pasi-to)

Slides, code and data,
recommended reading

1. Overview of data analysis and where the tour fits in
2. Introduction to the tour (and to linked brushing)
3. Motivation: how the tour can help us better understand data
4. The tour + models
5. What's next

million
→
understand
→
ess





Your turn

Take two minutes to brainstorm (in pairs)
on the pros and cons of numerical models
vs. visual methods.

The tour

A. Buja, D. Cook, and D. F. Swayne. *Interactive high-dimensional data visualization*. Journal of Computational and Graphical Statistics, 5 (1):78–99, 1996.

Tools

GGobi (<http://ggobi.org/>): GGobi is an open source visualization program for exploring high-dimensional data.

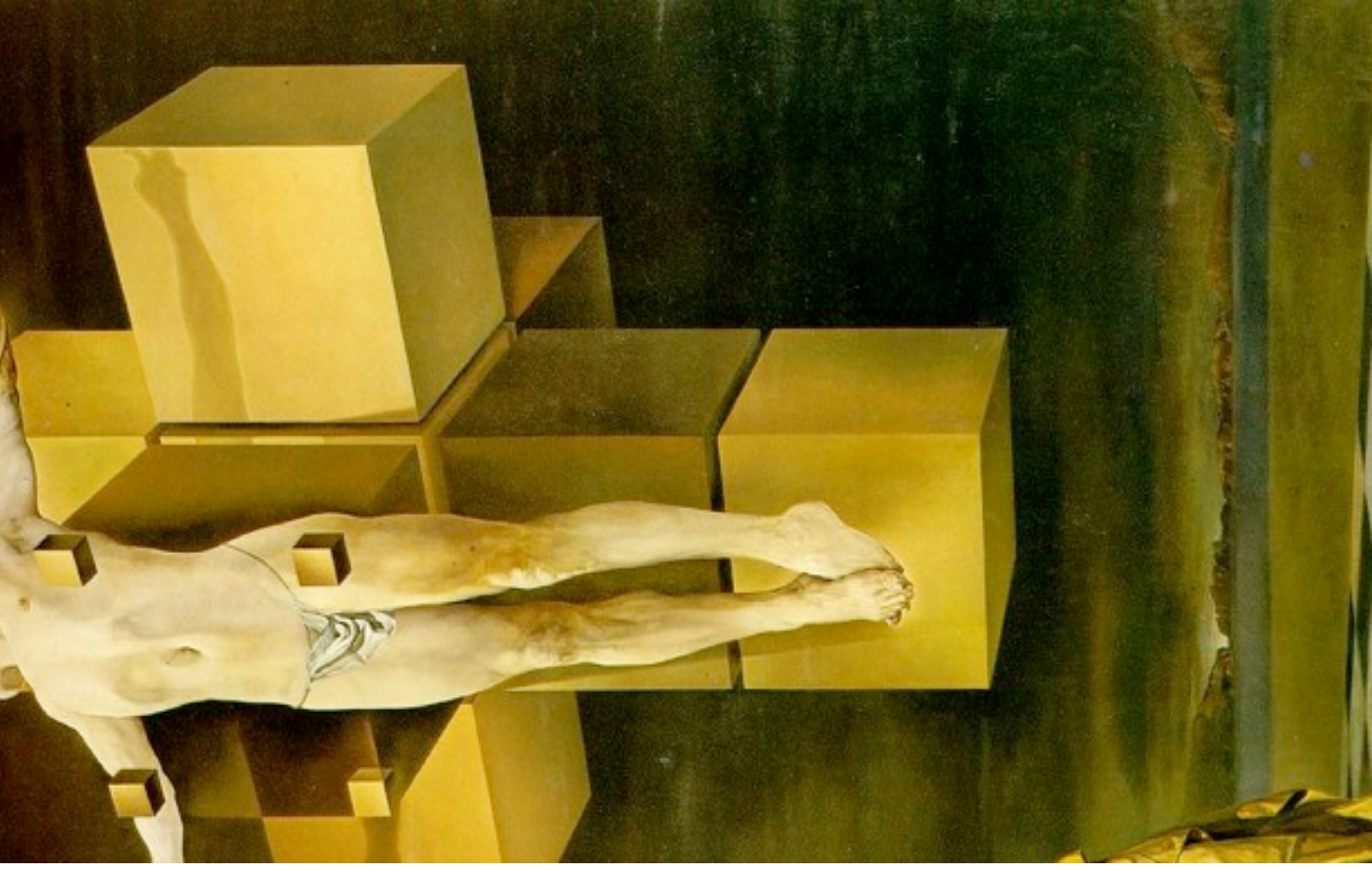
tourr: A new package for R that focusses exclusively on the tour. Aims to be a convenient framework for research on the tour. Hands on tomorrow.

The idea

We understand 3d objects by looking at them from all sides.

Why not do the same for 4d objects?

Three examples of the **grand tour**.



Corpus Hypercubus
Crucifixion) Salvador

Experimental design

Visualising a hypercube not just an academic exercise.

Can also use it to visualise factorial experimental design and detect interactions.

Example of manual controls.

How the tour works

Pick new angle randomly.

Rotate smoothly to that new angle, minimising inter-frame rotation.

Intuitive in 3d dimensions. Mathematics trivial in any number of dimensions (just matrix multiplication)

Interpretation

How can we interpret individual projections within the tour?
(Like a biplot!)

We can take unit vectors corresponding to each original axes and project. These give us the **tour axes**.

AT&T laser data

Wanted to establish performance characteristics of new laser.

Inputs: front current, back current

Outputs: energy, wavelength

4 variables x 64 observations

Plots

Experimental design: ifront vs iback

Marginal relationships to power:

ifront vs power, iback vs power

Joint relationship with power:

ifront + iback + power.

Mix in a little of the wavelength. Outliers

Explorations with brushing.

Your turn

Given what you've seen so far, what might you use the tour for? Can you see any limitations?

Some uses

- Clustering and classification
- See multivariate outliers
- Creating better mental models of the data
- Model diagnostics

Tour + Models

Numerical models

Never use the tour by itself!

Always a process of iteration between explicit numerical models and visualisation.

Want to both understand our data and our models.

Examples

Adding in clustering information.

Displaying classification boundaries.

Show the hierarchical clustering in the high-dimensional data space.

Outline

Aims

Get you interested in the tour, and give you the tools to contribute to the field.

Focus on seeing the tour in practice, and on building an intuition about how it works.

For more details, see the papers on the website.

What's next

The mathematics of the tour

Other types of tour

Tomorrow

Programming the tour.

(How much R do you know?)

s work is licensed under the Creative Commons Attribution-Noncommercial 3.0 United States License. To view a copy of this license visit <http://creativecommons.org/licenses/by-nc/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.