

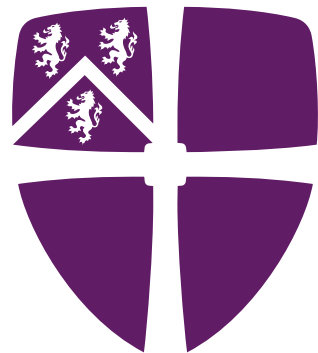
Jonathan Cumming, Department of Mathematical Sciences

Statistics, Data, and Science

Introduction

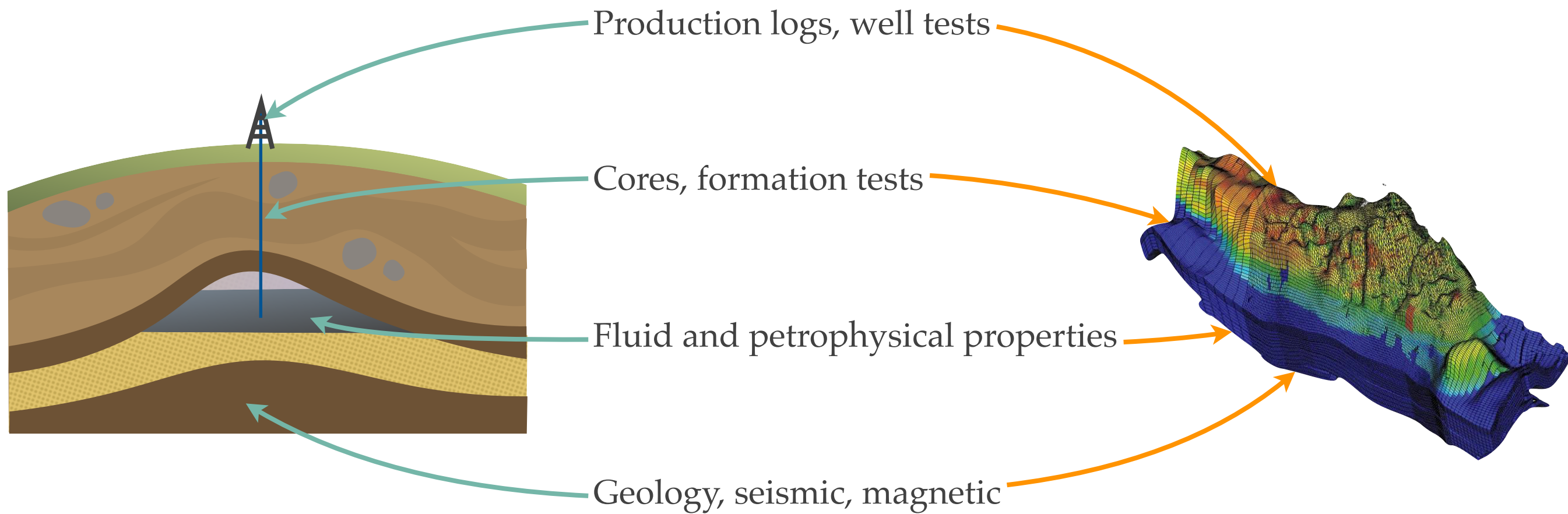
- ❖ **Statistics** is the **science** of learning from **data**, and dealing with the **uncertainties** in that process.
- ❖ Two case studies
 - ❖ Learning about petroleum reservoirs
 - ❖ Online browsing behaviour in digital commerce

Wells, Oil, and Some Maths

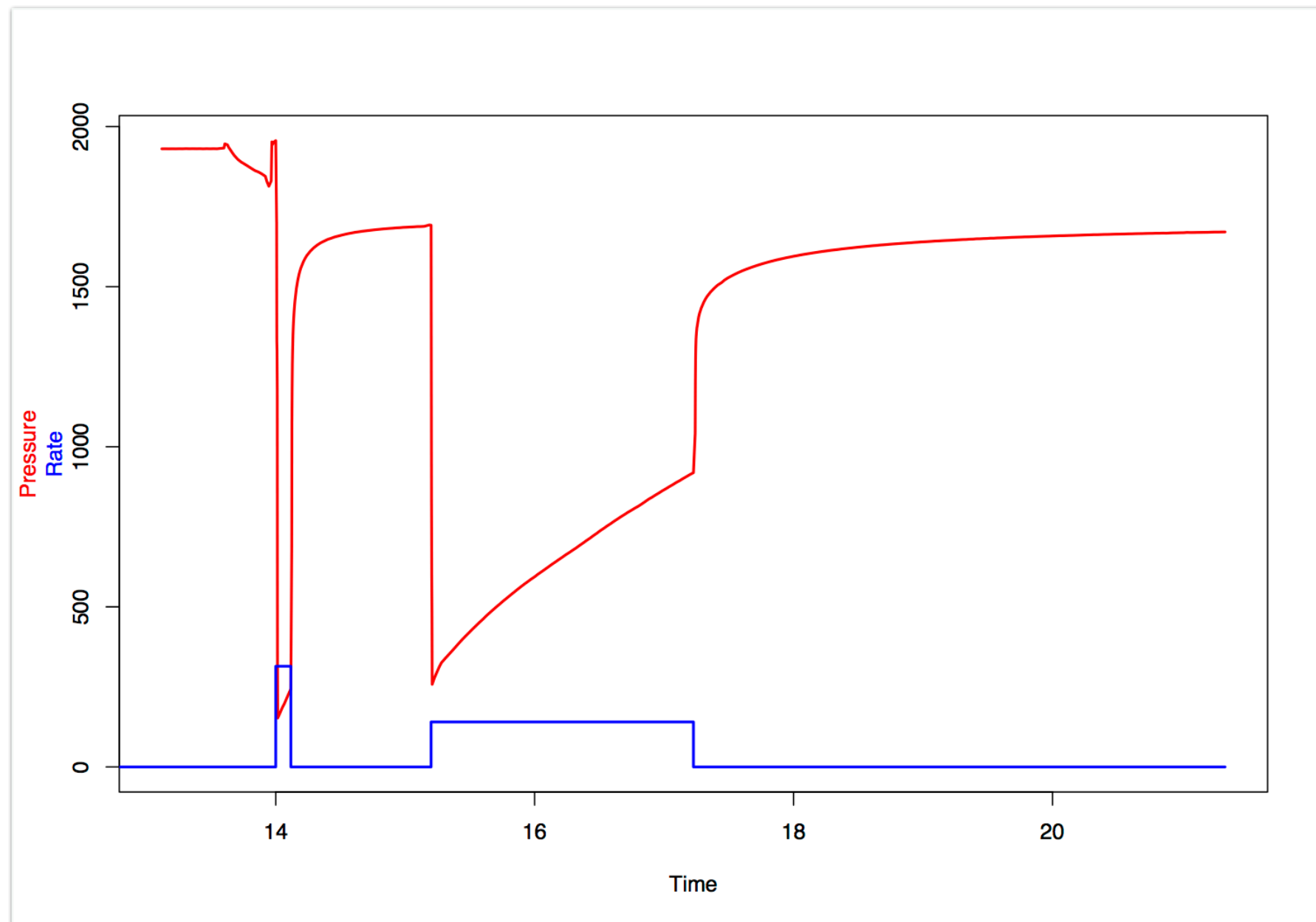


Imperial College
London

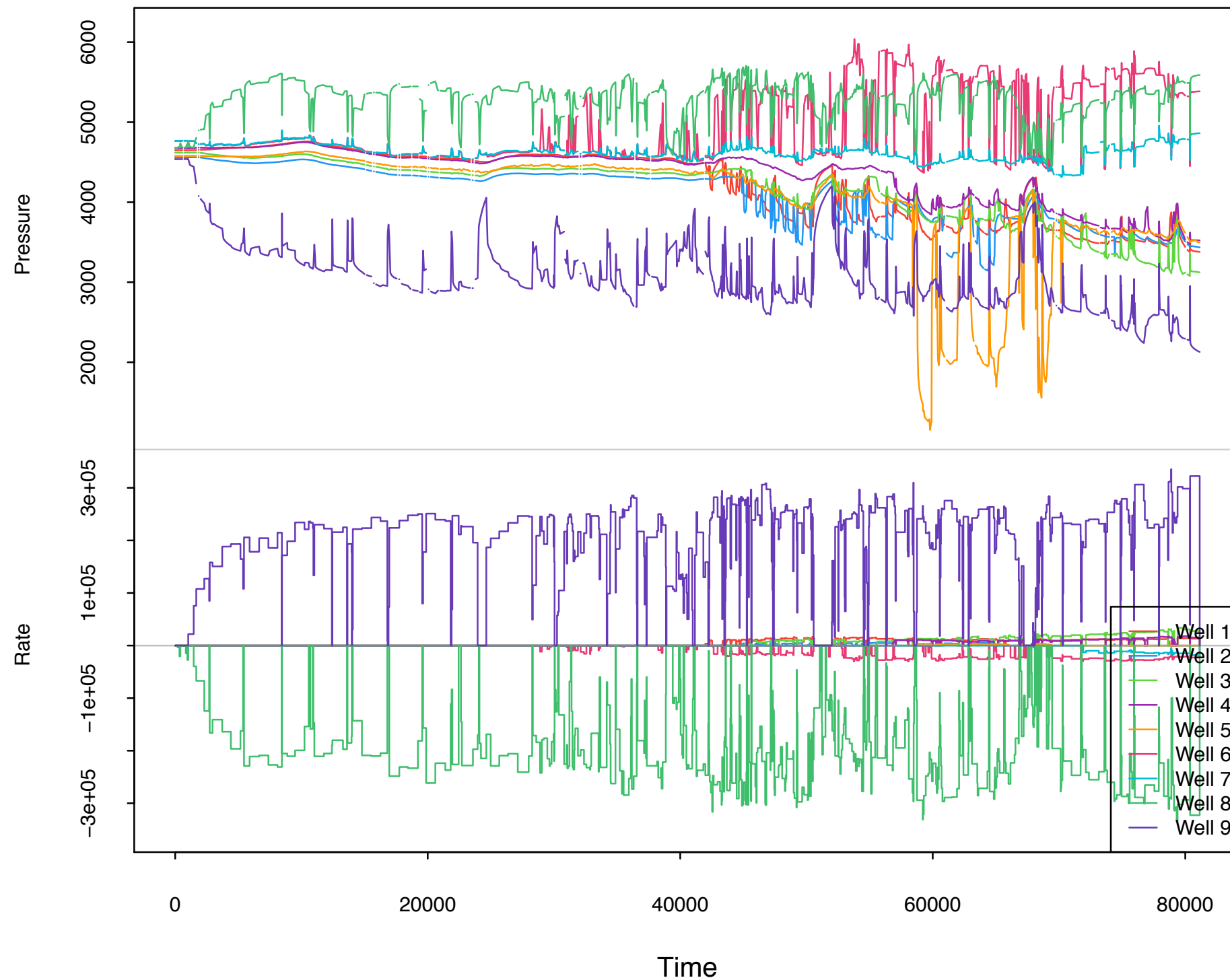
Rich in data



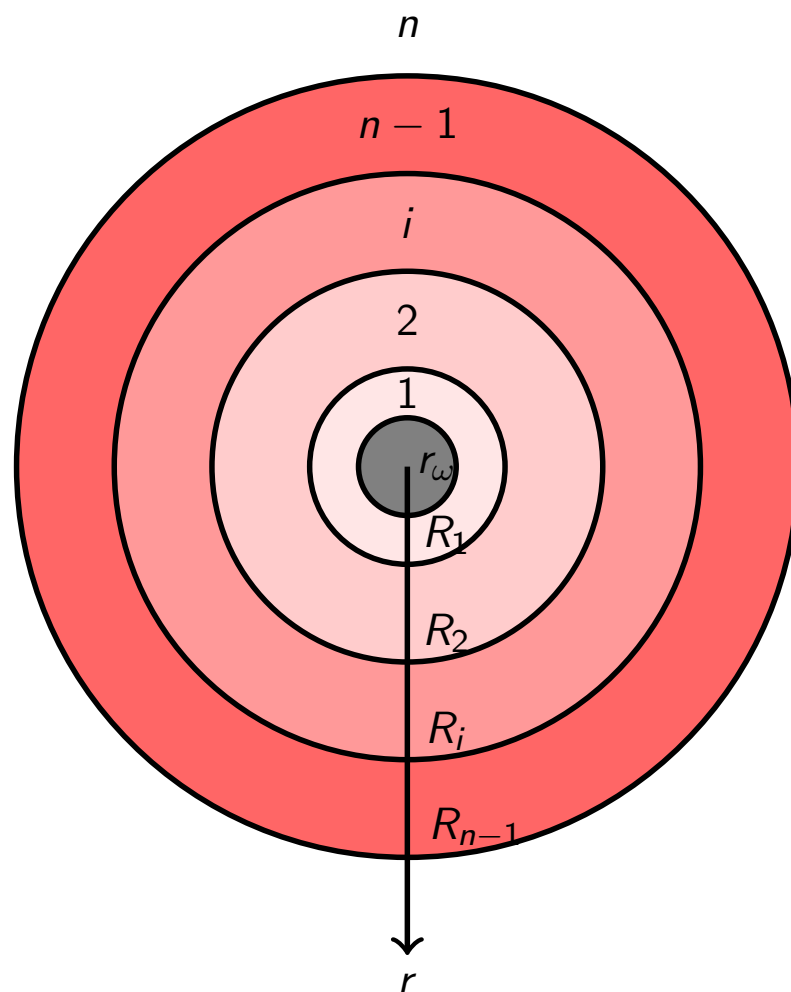
Small data problems



Not-so-small data problems



Mathematical basis



A fairly scary model for the reservoir:

$$\frac{\partial^2 P_{D1}}{\partial r_D^2} + \frac{1}{r_D} \frac{\partial P_{D1}}{\partial r_D} - \frac{1}{C_D e^{2S}} \frac{\partial P_{D1}}{\partial t_D} = 0$$

$$\frac{\partial^2 P_{D2}}{\partial r_D^2} + \frac{1}{r_D} \frac{\partial P_{D2}}{\partial r_D} - \frac{\eta}{C_D e^{2S}} \frac{\partial P_{D2}}{\partial t_D} = 0$$

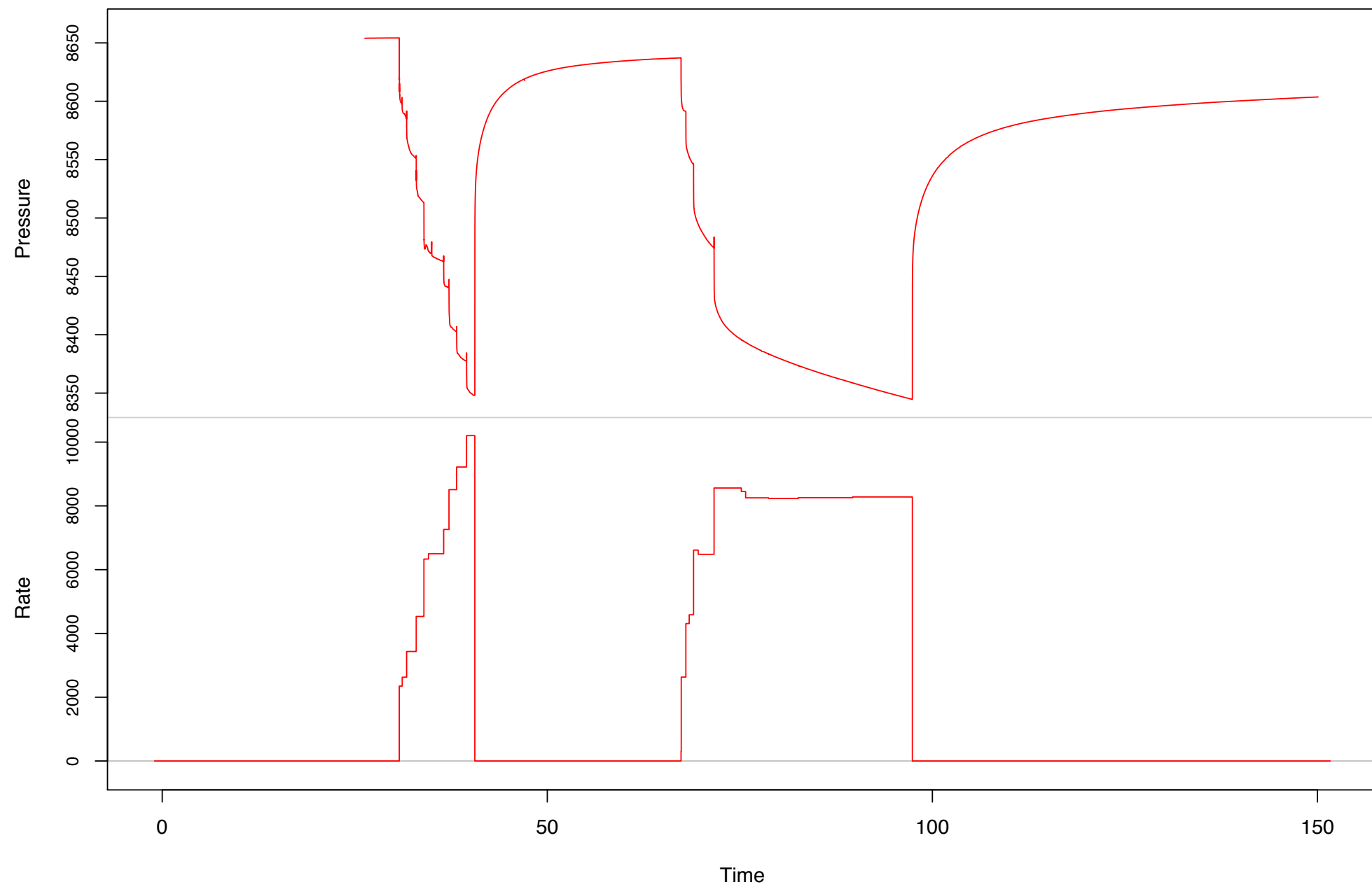
A slightly less scary relation between pressure and rate:

$$p(t) = p_0 - \int_0^t g(t - t'; \theta) q(t') dt'$$

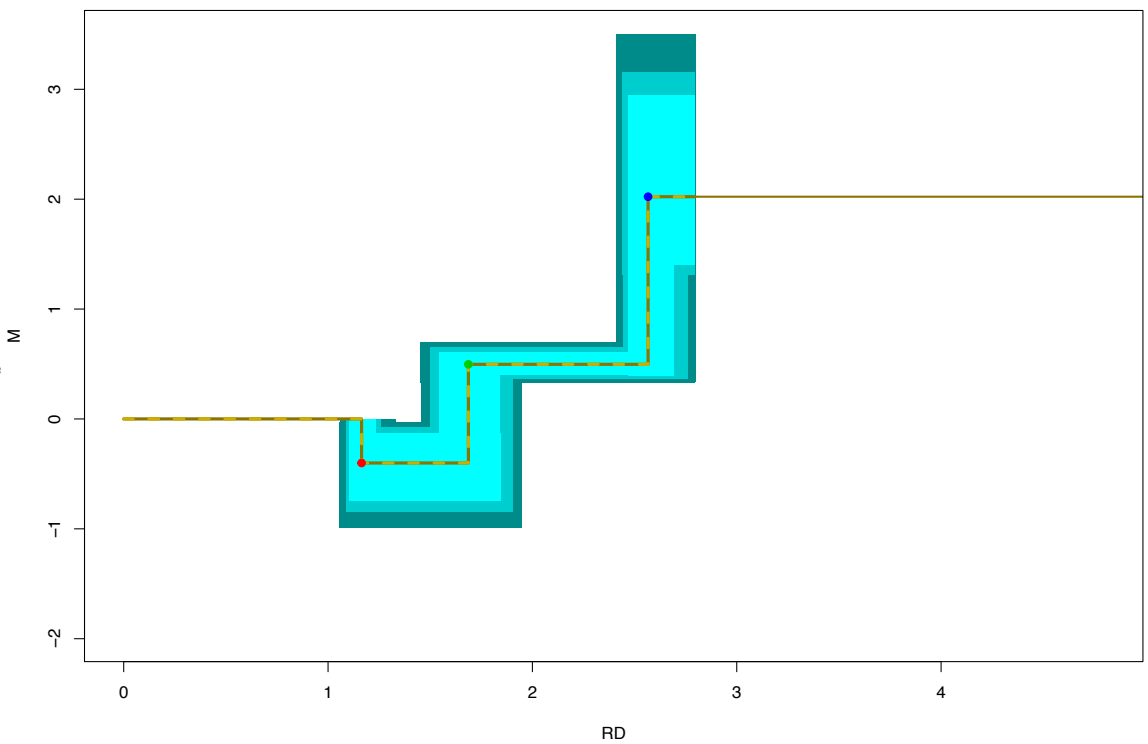
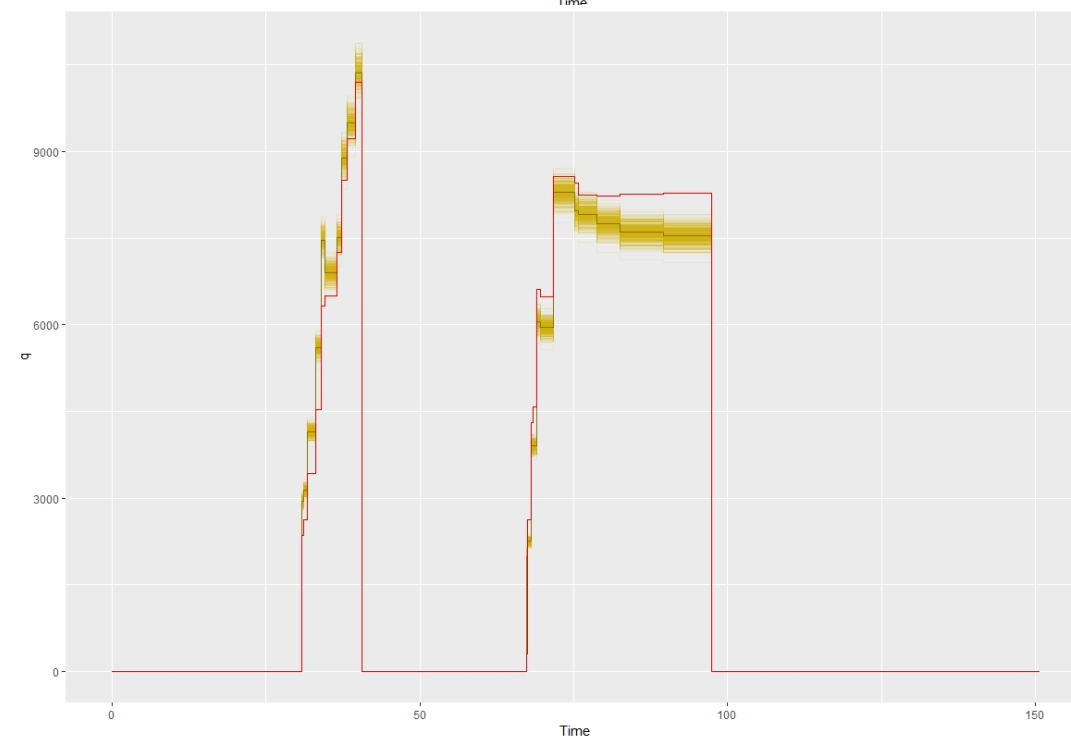
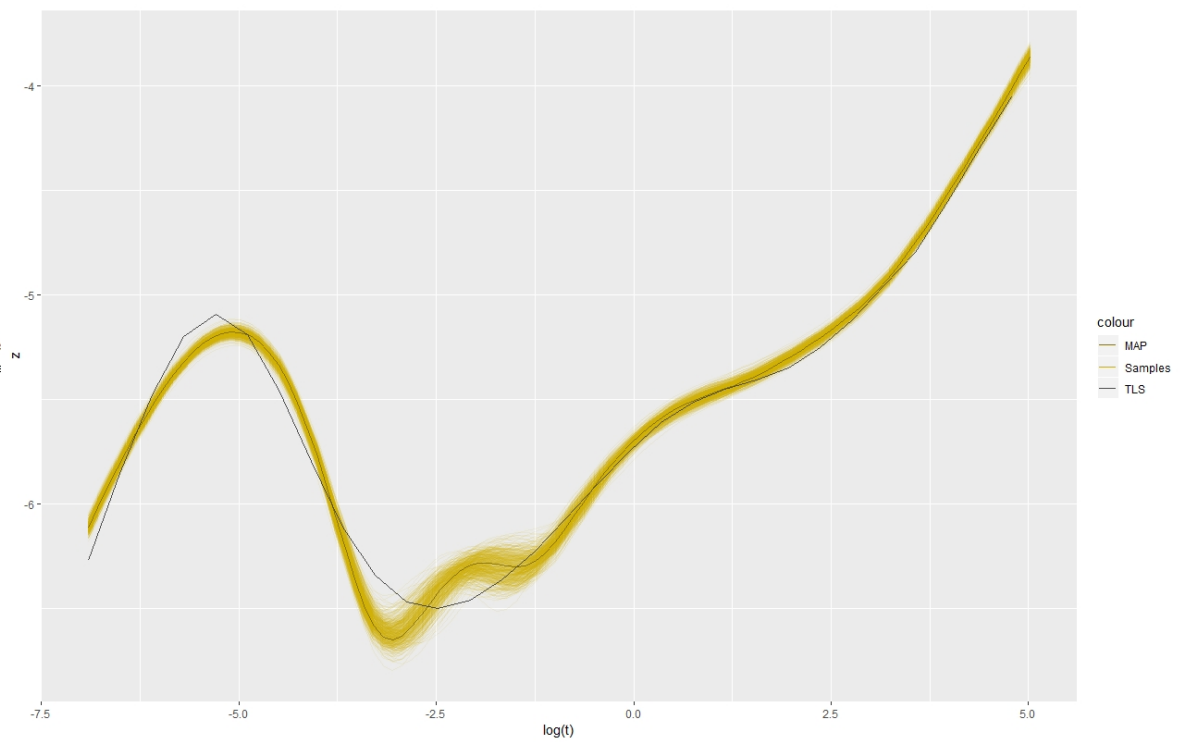
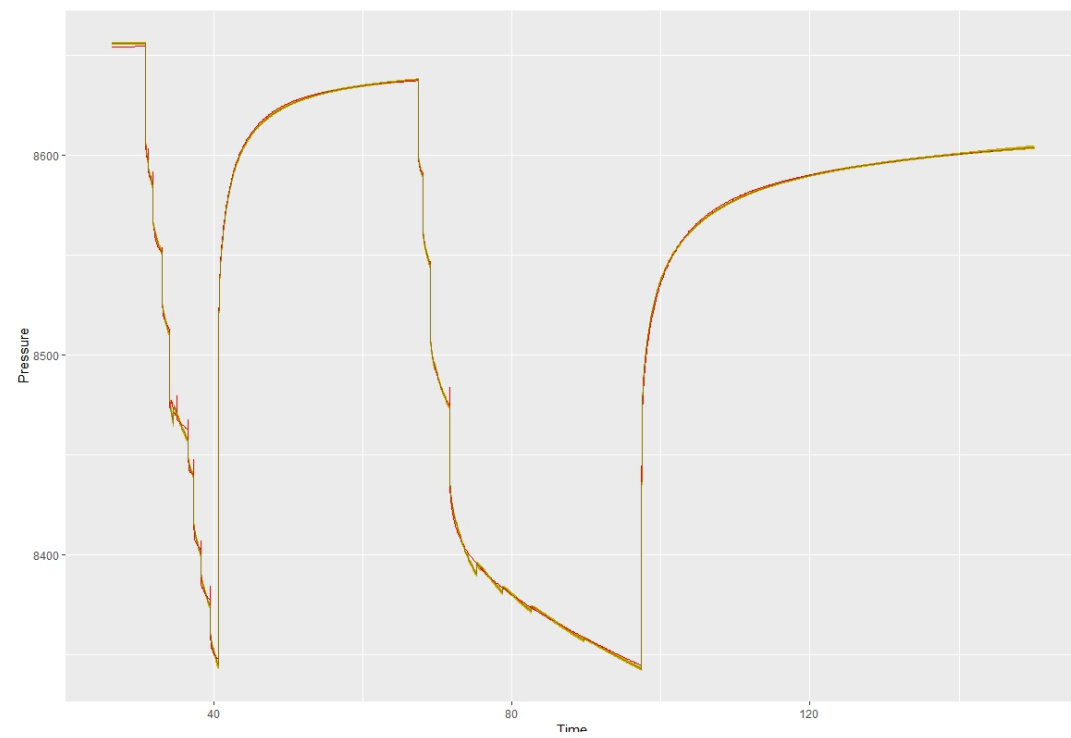
A Bayesian approach

- ❖ **Uncertain** (often very wrong) data
- ❖ **Uncertain** mathematical model for the reservoir
- ❖ **Bayesian statistics** represents uncertainty as probability (or expectation), and learns from data by conditioning

Example data



Results



Clicks & Hoovers

Knowledge Transfer
Partnerships

Innovate UK
Technology Strategy Board



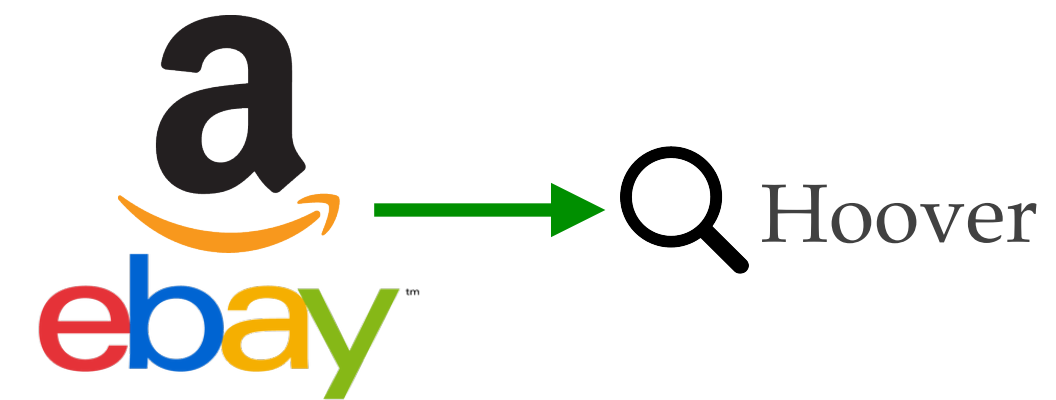
carbon
by  clicksco

Online browsing behaviour

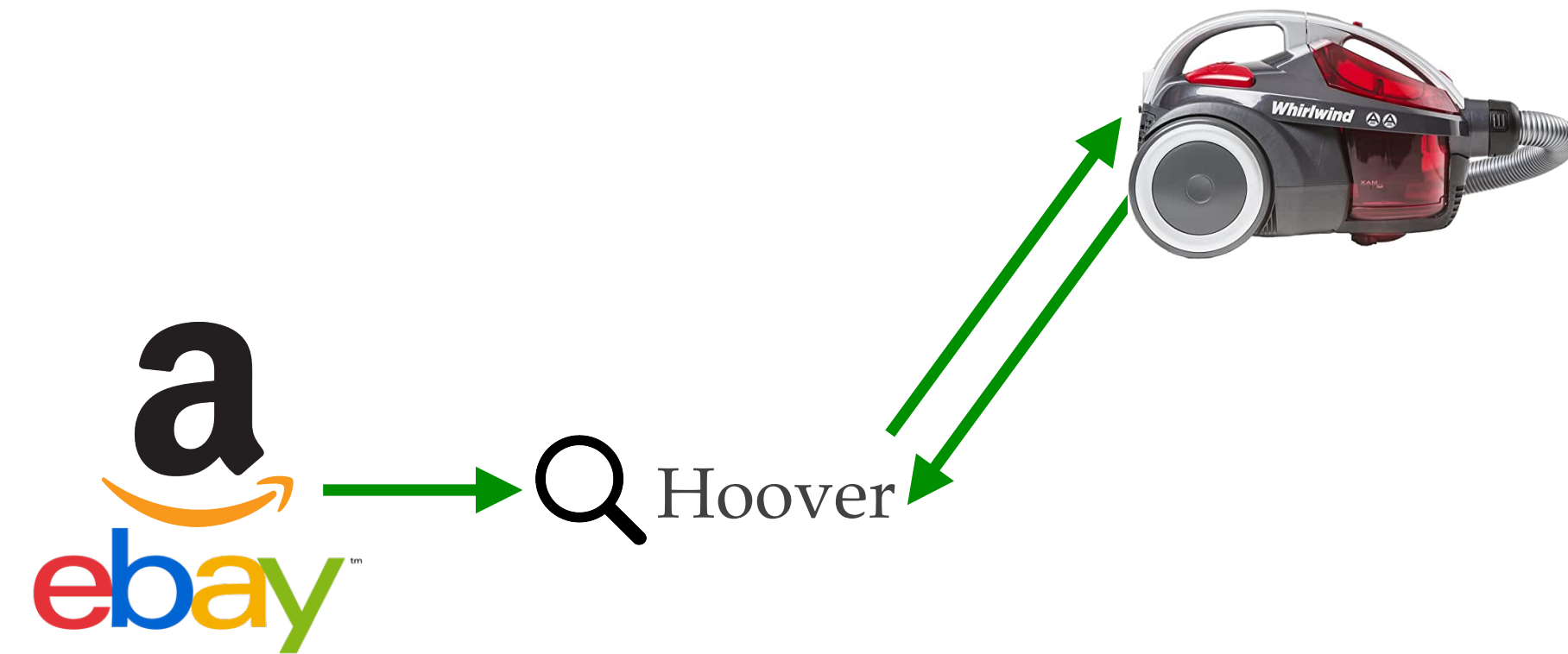
Online browsing behaviour



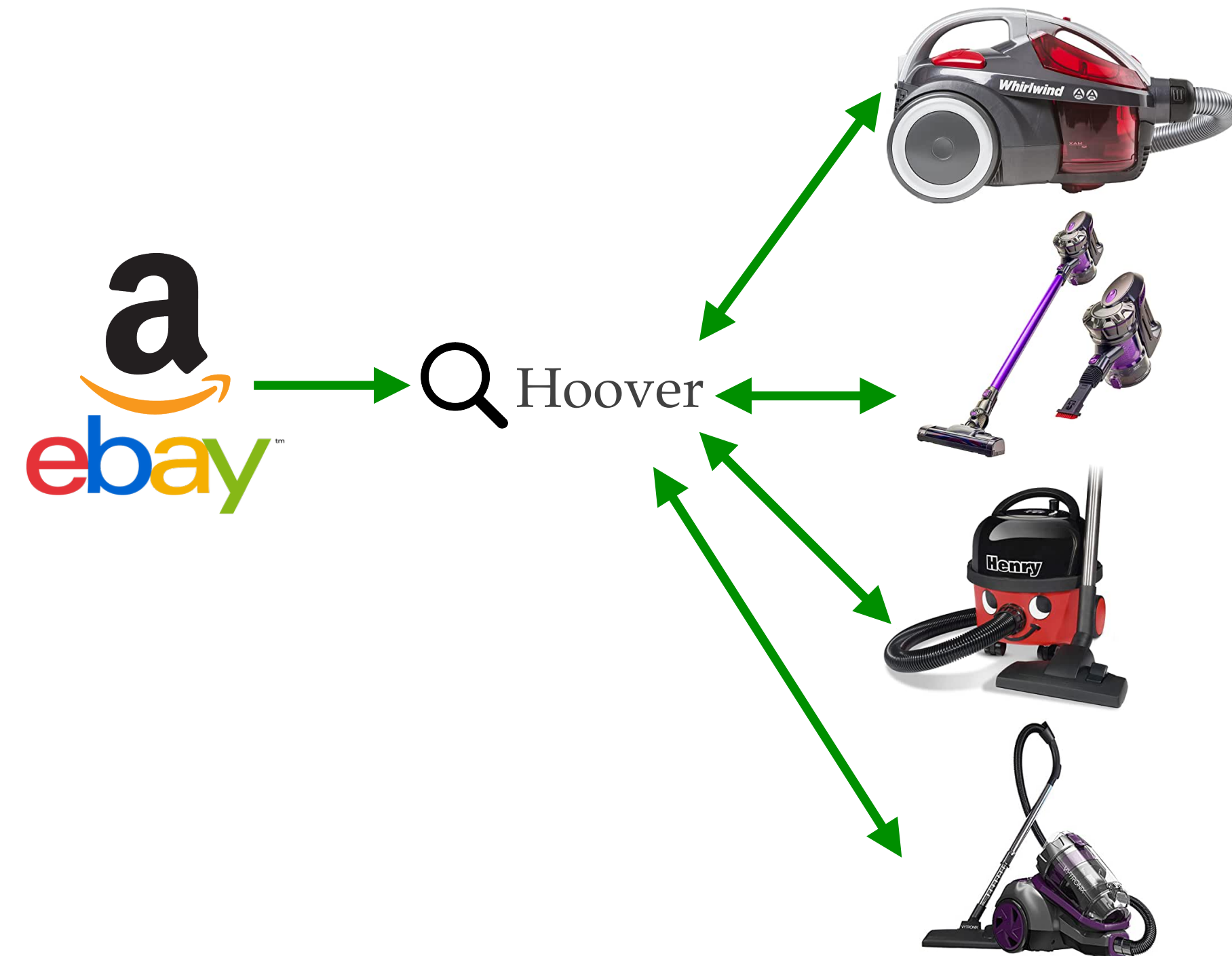
Online browsing behaviour



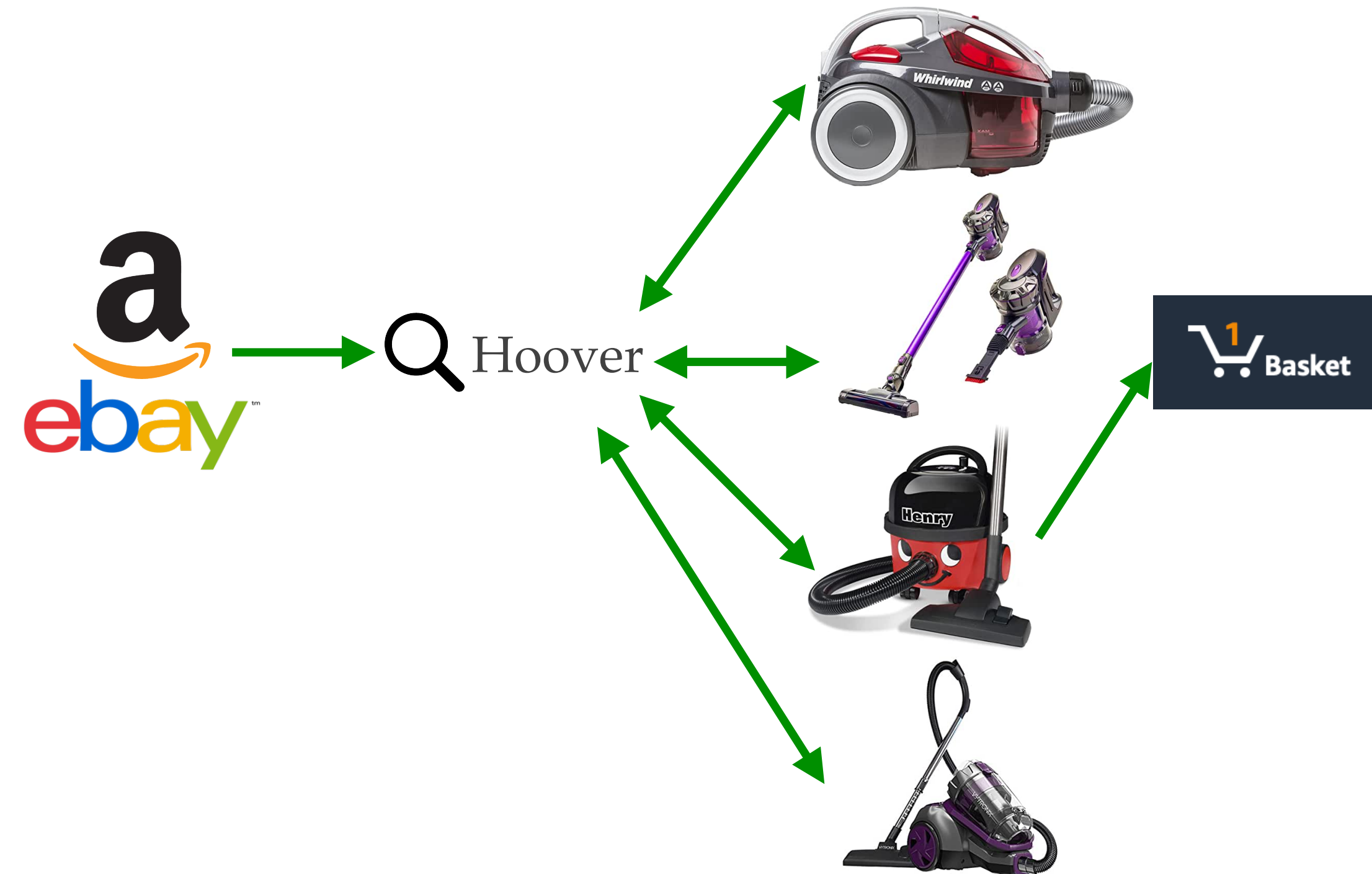
Online browsing behaviour



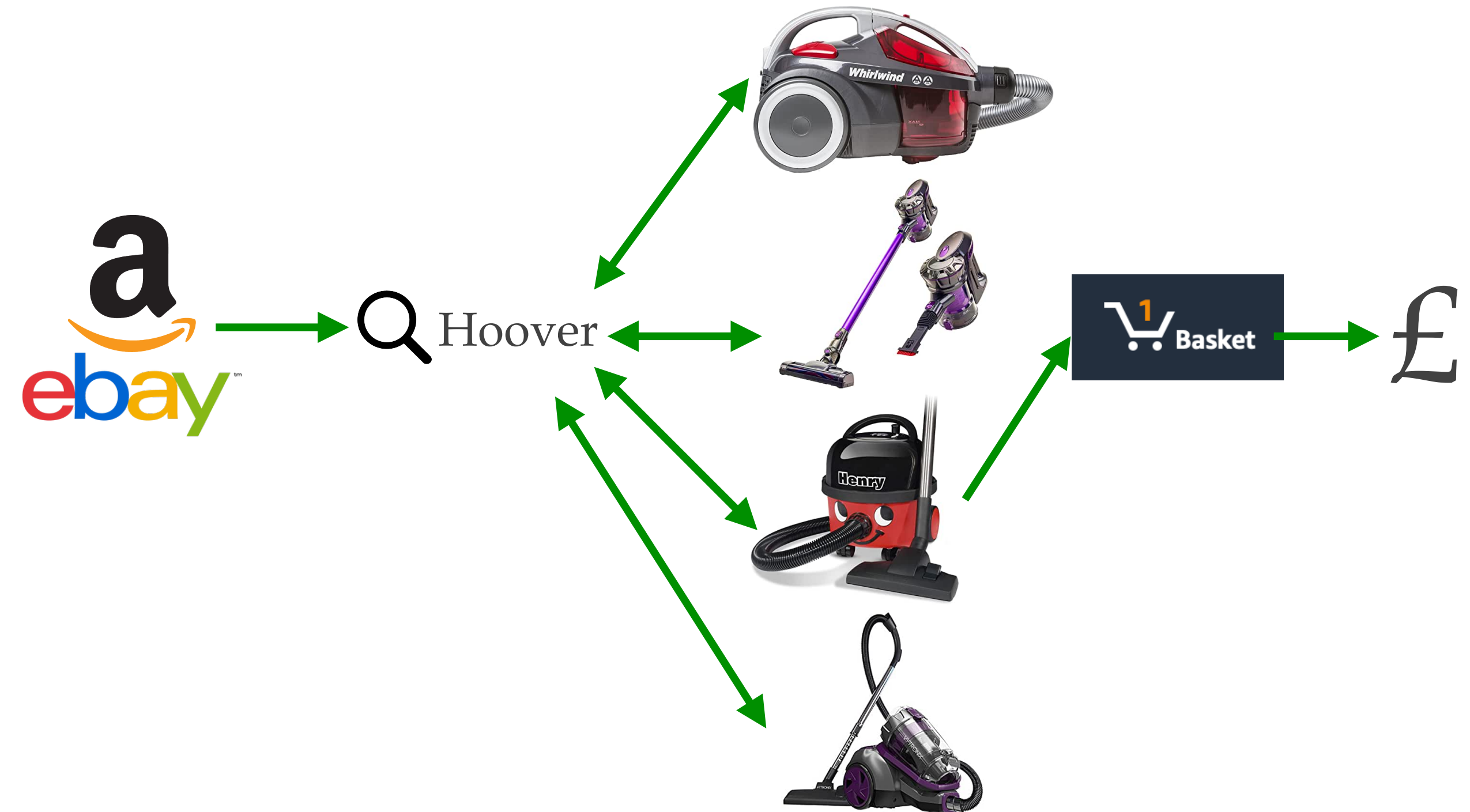
Online browsing behaviour



Online browsing behaviour



Online browsing behaviour

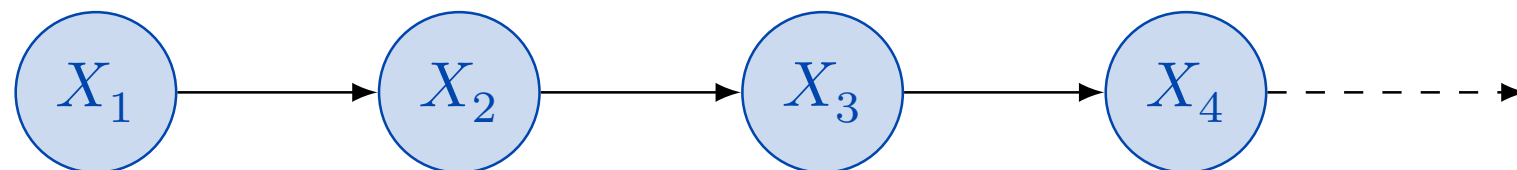


Clickstream Markov Chain

- ❖ The sequence of pages gives a **clickstream**:
H S L P L P L P L P B X
- ❖ **Huge** dataset of millions of clickstreams a day

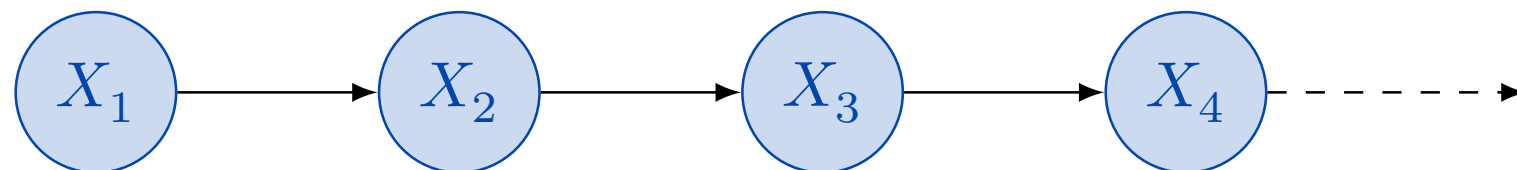
Clickstream Markov Chain

- ❖ The sequence of pages gives a **clickstream**:
H S L P L P L P L P B X
- ❖ **Huge** dataset of millions of clickstreams a day
- ❖ Can think of this as a random jump between pages

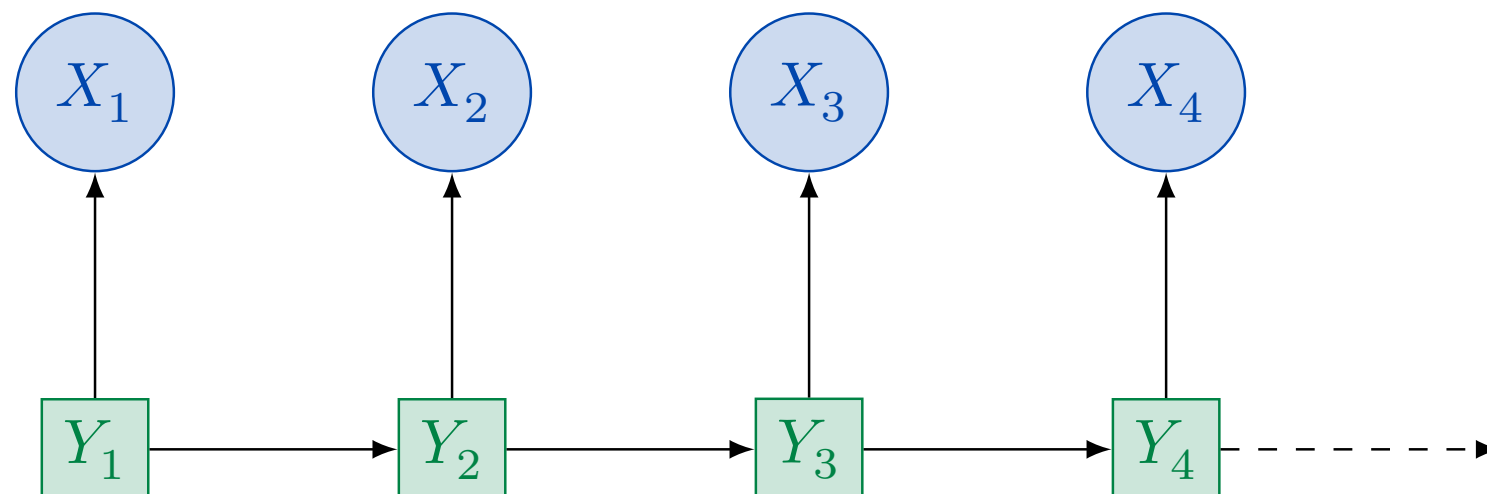


Clickstream Markov Chain

- ❖ The sequence of pages gives a **clickstream**:
H S L P L P L P L P B X
- ❖ **Huge** dataset of millions of clickstreams a day
- ❖ Can think of this as a random jump between pages

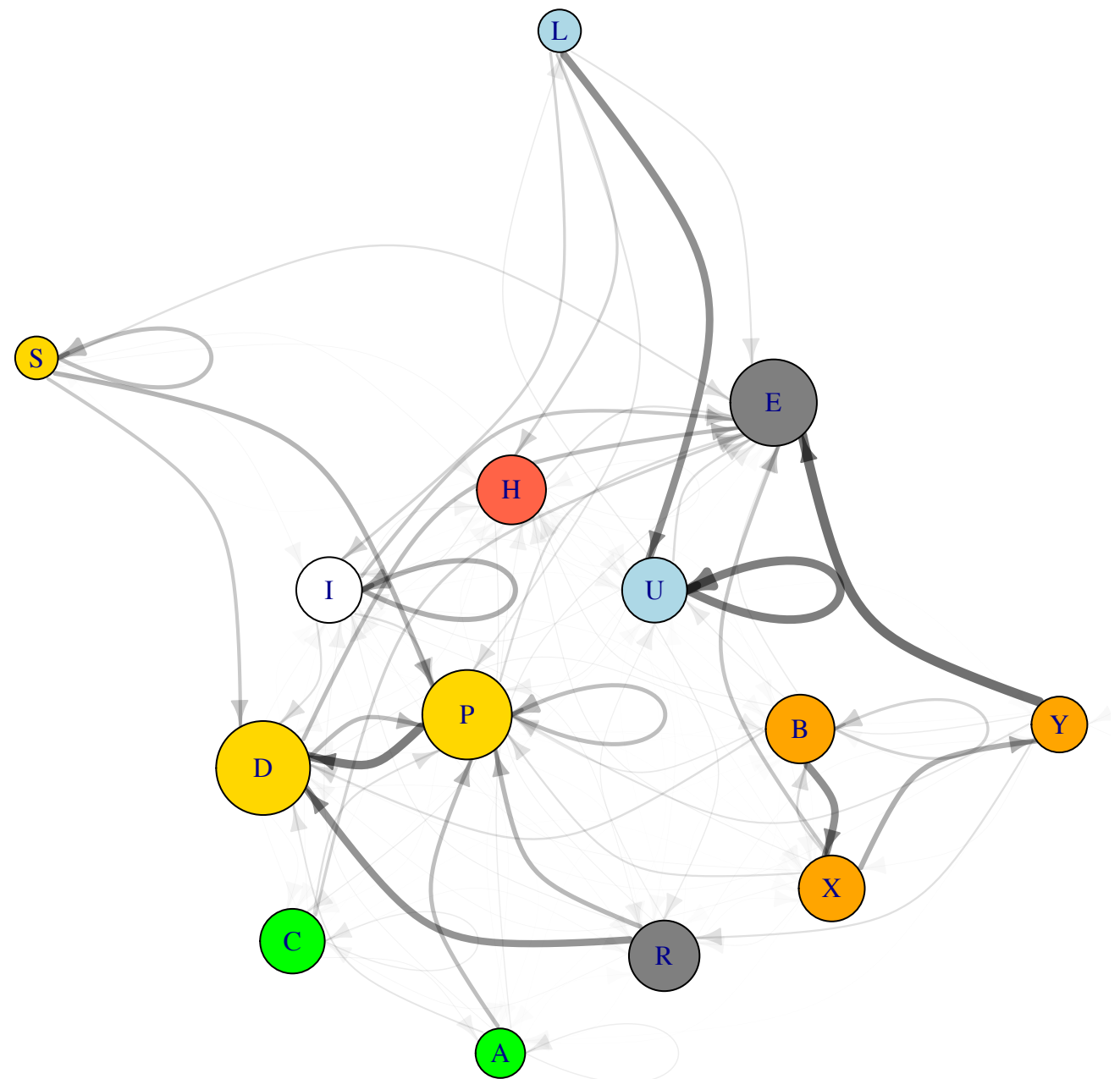
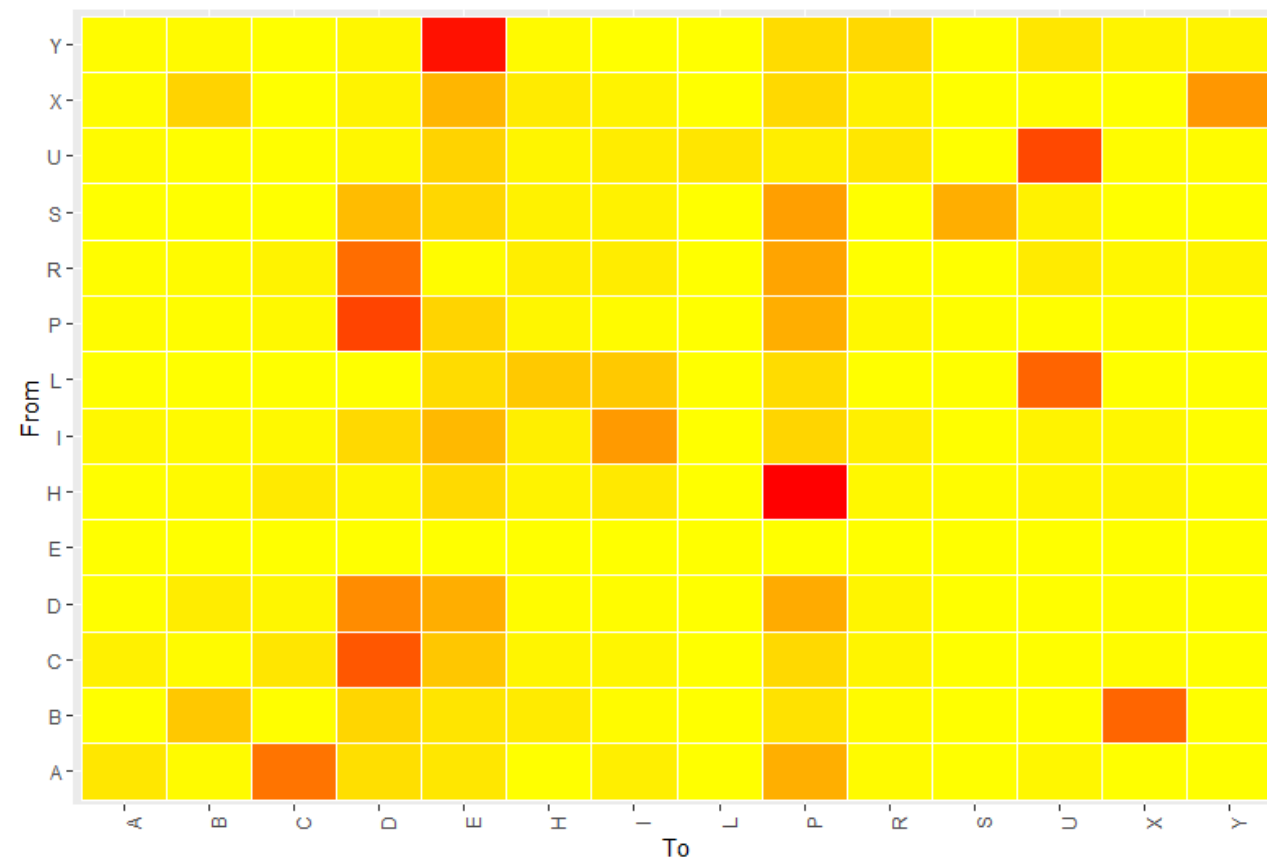


- ❖ Or something more complex

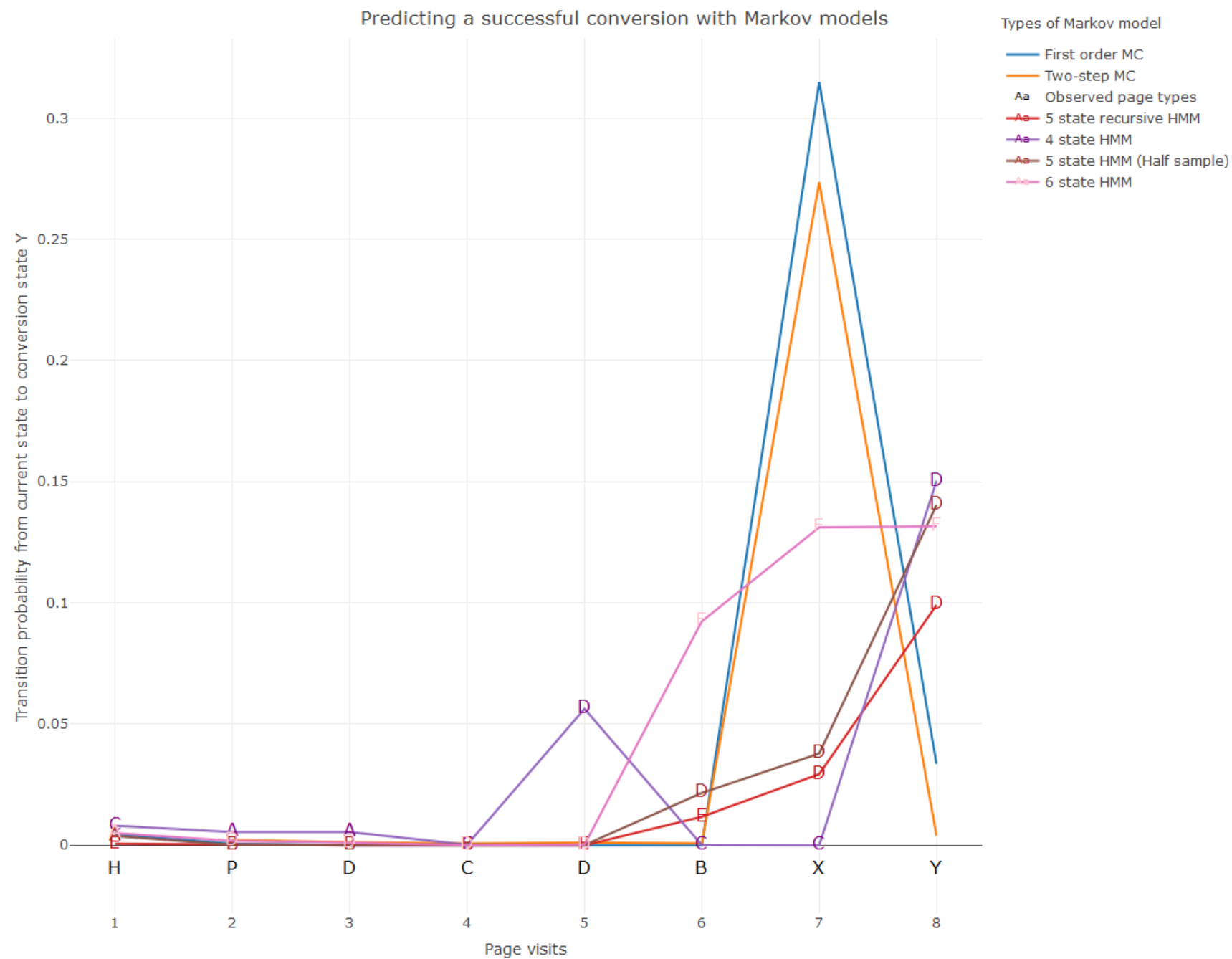


Results

First order Markov Chain transition matrix



Predicting a jump to “Checkout”



Final comment

Data + Science + Statistics
=
Many interesting problems