Explorable data visualisations

Roly Perera
rperera@turing.ac.uk
The Alan Turing Institute

Tomas Petricek
t.petricek@kent.ac.uk
University of Kent & The Alan Turing Institute

James Cheney
j.cheney@inf.ed.ac.uk
University of Edinburgh

Project goals

Help data scientists using interactive notebooks:

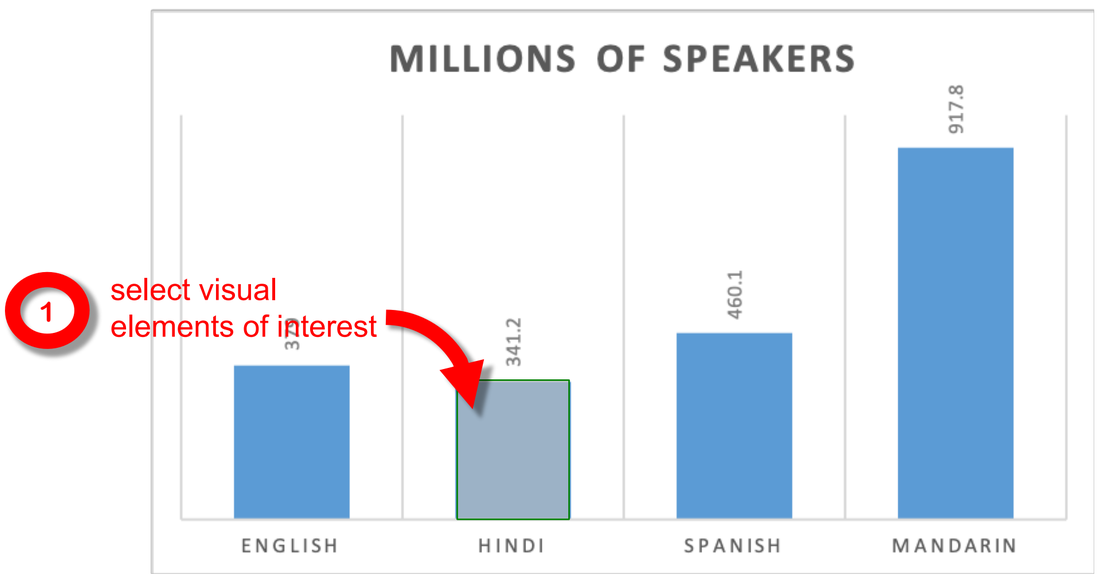
- **understand** data analyses by connecting parts of visualisations to data they depend on
- **explore** consequences of changes to models or data in order to test hypotheses
- **debug** causes of unexpected outcomes

Current: exploring dependencies

Adapt program slicing techiques for dependency tracking to the **data visualisation** domain:

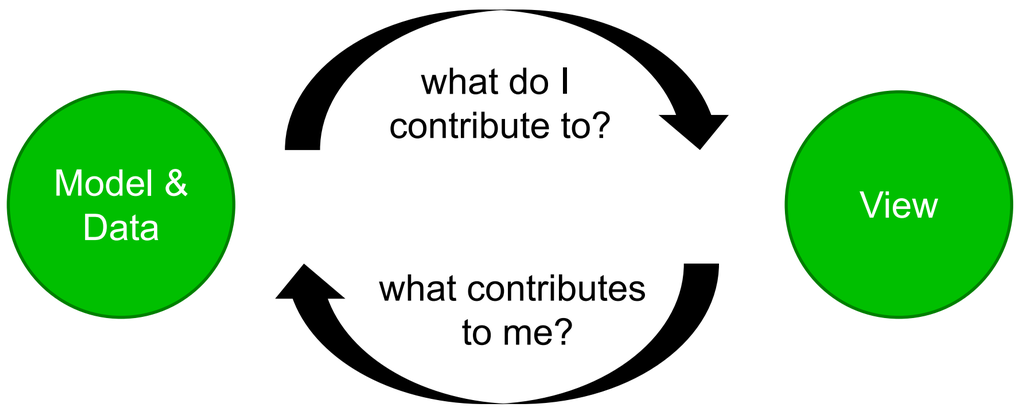
```
let data = [[English, 379.0], [Hindi, 341.2], ...]
let invalid = -1
data ▷ filter (λ(lang, n) => n <> invalid)
     ▷ barChart
```

2 backward dependencies are highlighted



Lattice-theoretic foundation

Our approach to data provenance is based on **Galois connections** and analyses dependencies in both directions:



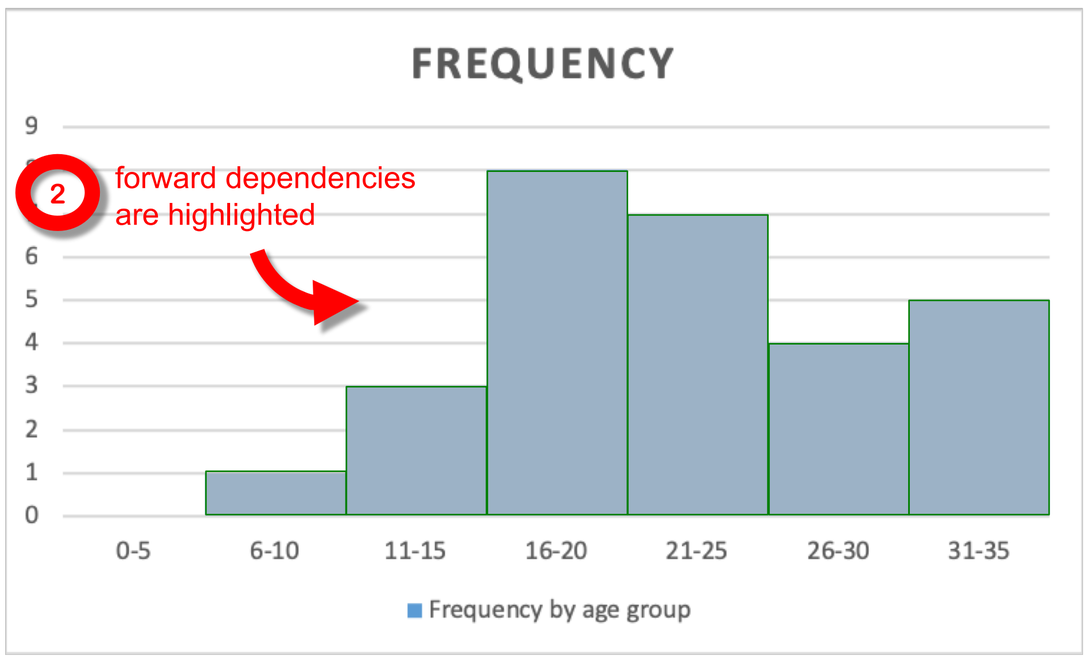
$$source(v_\alpha) = \bigwedge \{ d_{\alpha'} \mid visualise(d_{\alpha'}) \geq v_\alpha \}$$

Future: exploring changes

Analyses involving aggregate properties such as **binning** or **averaging** often have many dependencies:

```
let data = [[yes, 18], [no, 12], [yes, 34], ...]
let bins = [5, 10, 15, 20, 25, 30, 35]
data ▷ groupBy bins
     ▷ histogram
```

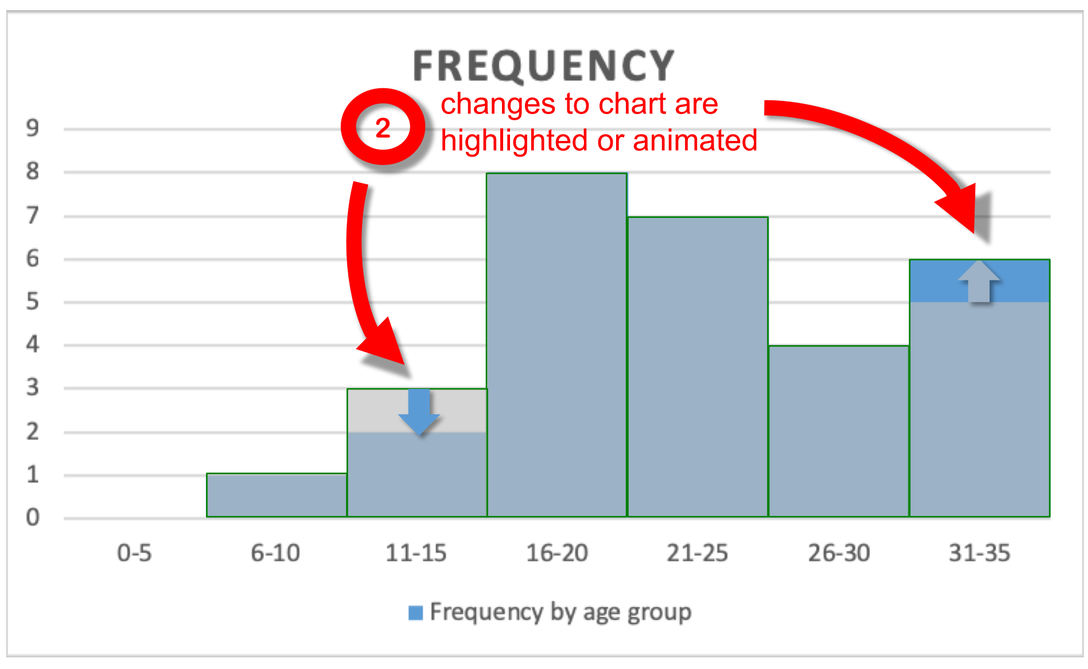
1 select data or code of interest



Generalise prior techniques to associate **changes in data or analyses** to changes in visualisations:

```
let data = [[yes, 18], [no, 12 32 es, 34], ...]
let bins = [5, 10, 15, 20, 25, 30, 35]
data ▷ groupBy bins
     ▷ histogram
```

1 modify data or code



Connections to explore

- probabilistic programming
- incremental computation