SCION
FORESTS ▪ PRODUCTS ▪ INNOVATION

Building a Visual Analytics System for Spatio-temporal Analysis

Alan Tan, Yue Lin, Ralf Gommers
5th Sep 2019

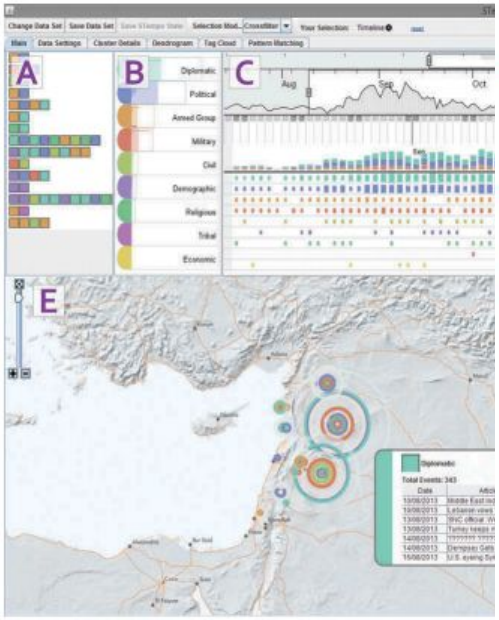
Problem

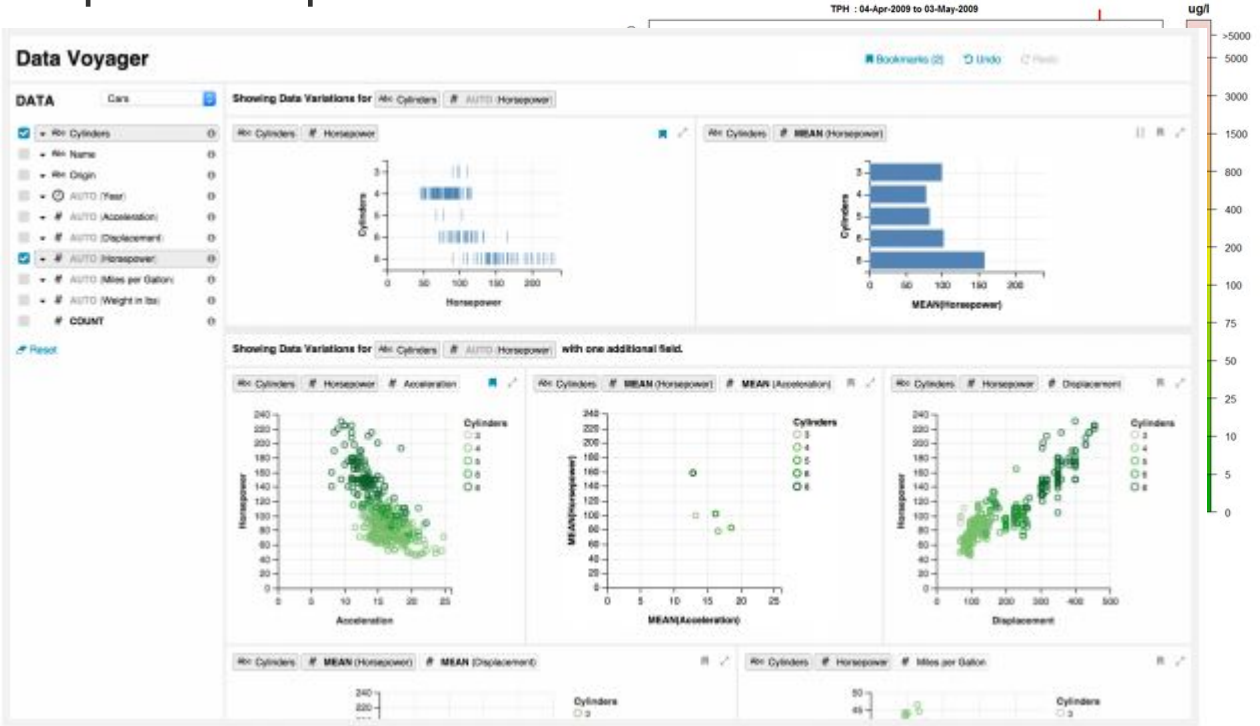
- Many real-world data is of spatio-temporal natured
- Fundamentally challenging to explore and discover data relationships in complex spatio-temporal datasets
- Permanent Sample Plot (PSP) Database
  - Database capturing field measurements from tree plots geographically distributed across New Zealand
  - More than 100 years of field measurements with over 100 measured and derived variab

scion

Existing tools

- Fit for purpose or data tools
  - STempo[1]
  - Groundwater Spatio-temporal Data Analysis Tool[2]
  - Voyager[3]





[1] A. C. Robinson, D. J. Peuquet, S. Pezanowski, F. A. Hardisty, and B. Swedberg, "Design and evaluation of a geovisual analytics system for uncovering patterns in spatio-temporal event data," *Cartography and Geographic Information Science,* vol. 44, no. 3, pp. 216-228, 2017/05/04 2017

[2] W.R. Jones, M. Bonte, K. Cady, "The Groundwater Spatiotemporal Data Analysis Tool for Groundwater Quality Analyses", CL:AIRE technical bulletin, July 2019

[3] Wongsuphasawat, K., Moritz, D., Anand, A., Mackinlay, J., Howe, B., Heer J., "Voyager: Exploratory Analysis via Faceted Browsing of Visualisation Recommendations. IEEE Transactions on Visualisation and Computing Graphics 22,1, doi: 10.1109/TVCG.2015.2467191

scion

Goals

- Robust tool that allows user to explore different facets of a complex spatio-temporal dataset
  - Different facets (i.e. statistical, spatial, temporal, spatio-temporal)
  - Large dimensionality (e.g. PSP > 100 dimensions/variables)
  - Historically rich datasets (i.e. dynamic temporal patterns)
- Ease-of-use and Interactive

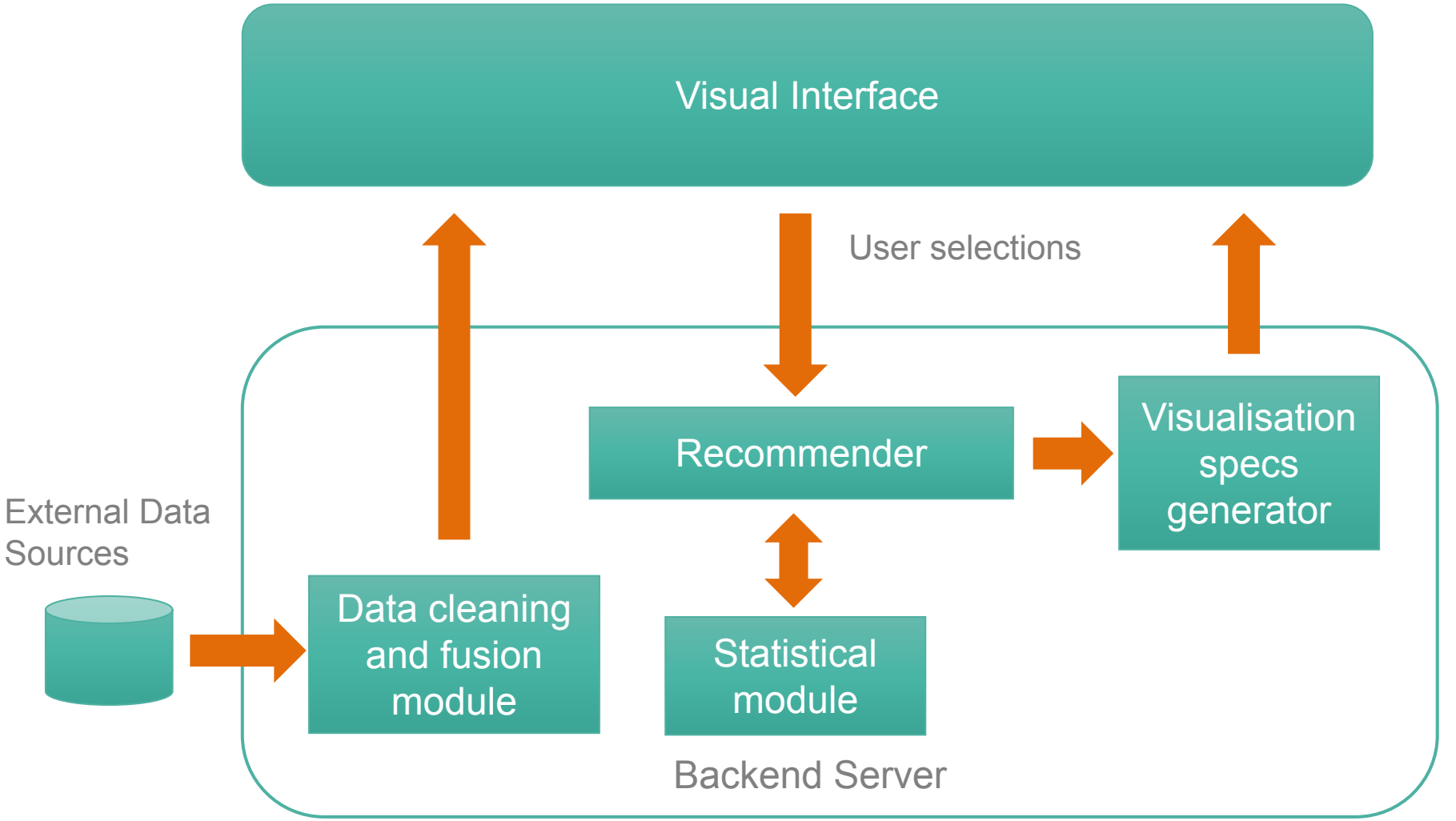
scion

Challenges

- Presentation of information
  - Different data types
  - Different information – spatial, temporal, spatio-temporal patterns
- Allowing users to dynamically focus on different aspects of the dataset
  - Variables
  - Types of analysis
- Interactive capabilities and data linkage
- Data computation
- Allowing users to quickly identify or discover patterns or data relationships that are of interest

SCION

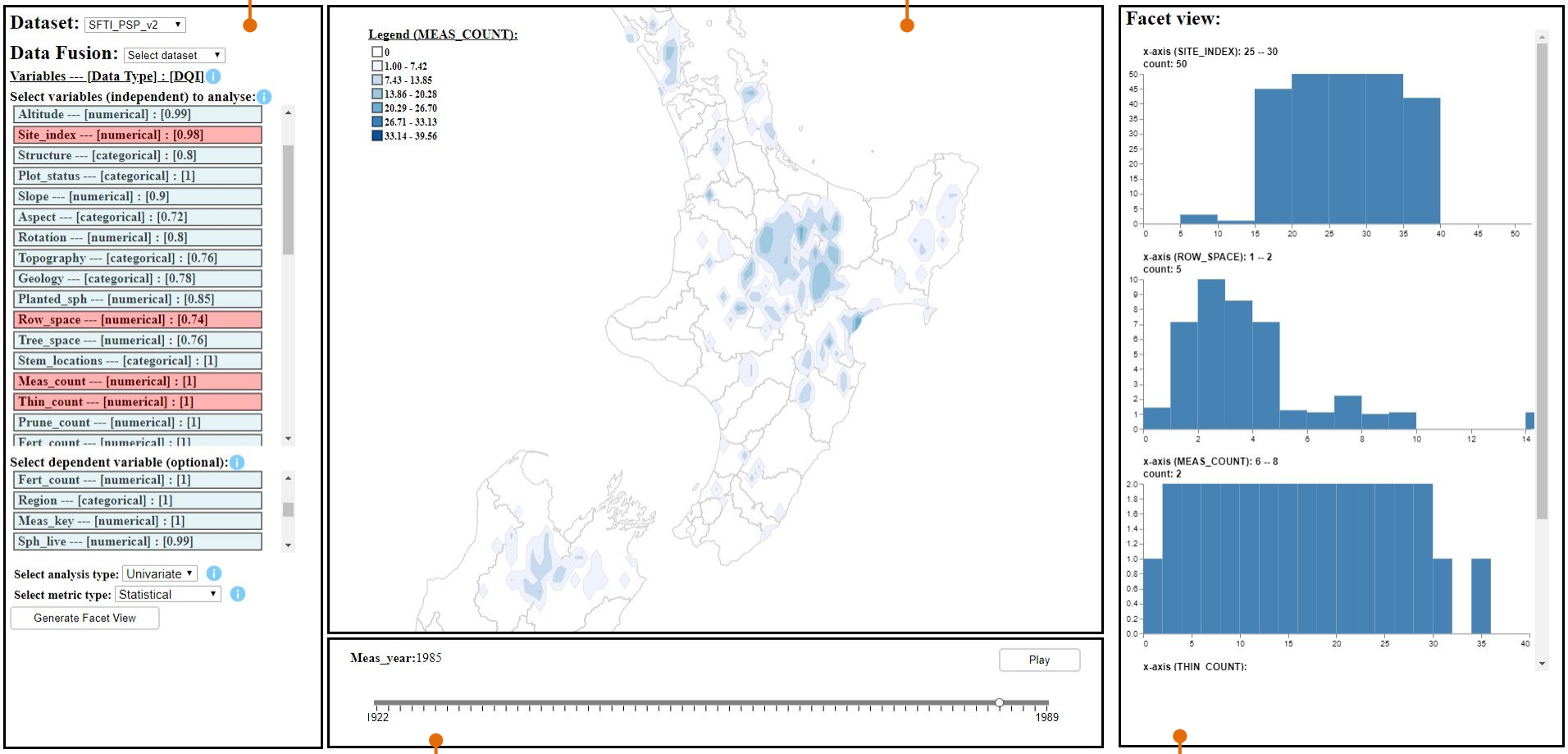
Visual Recommender Architecture



Visual Recommender User Interface

Variable panel

Spatial Map

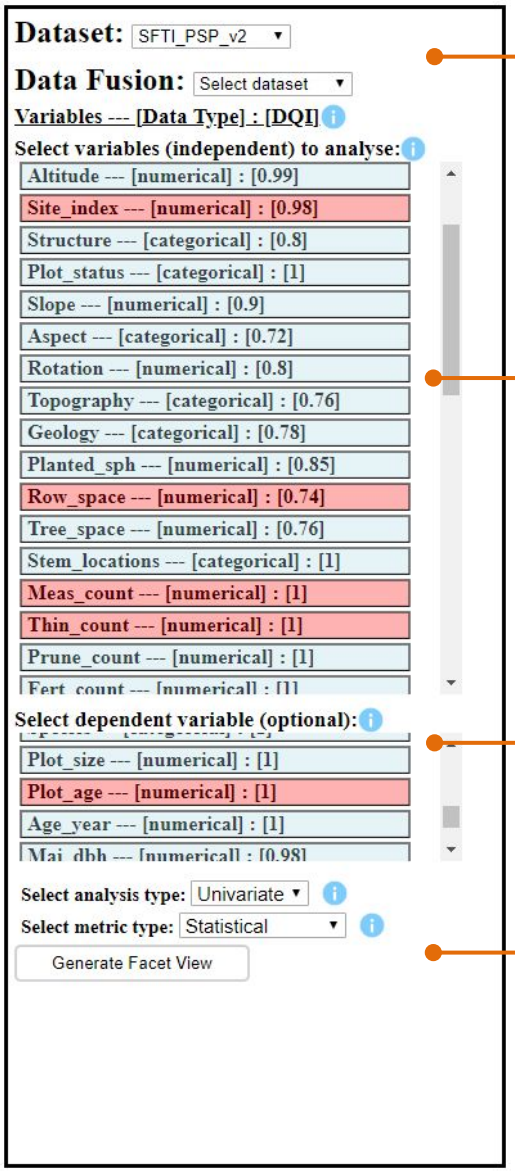


Time Panel

Facet View

scion

Variable Panel



Dataset selection

- Select datasets for analysis and for data fusion

Independent variable selection

- Choosing of variables for exhaustive pair-wise analysis

Dependent variable selection

- Select datasets for pair-wise analysis against all selected independent variables

Mode controls

- Control types and mode of analysis

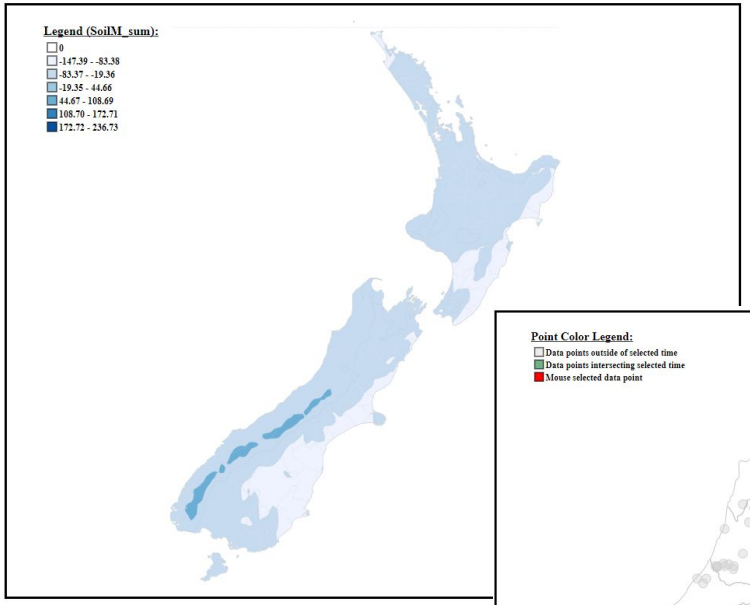
scion

Spatial Map

- Different modes of spatial visualisation

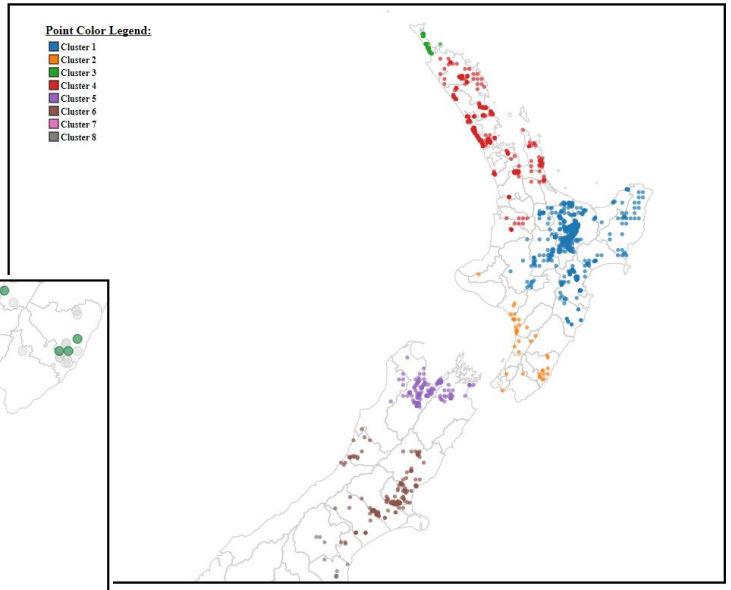
Heatmap

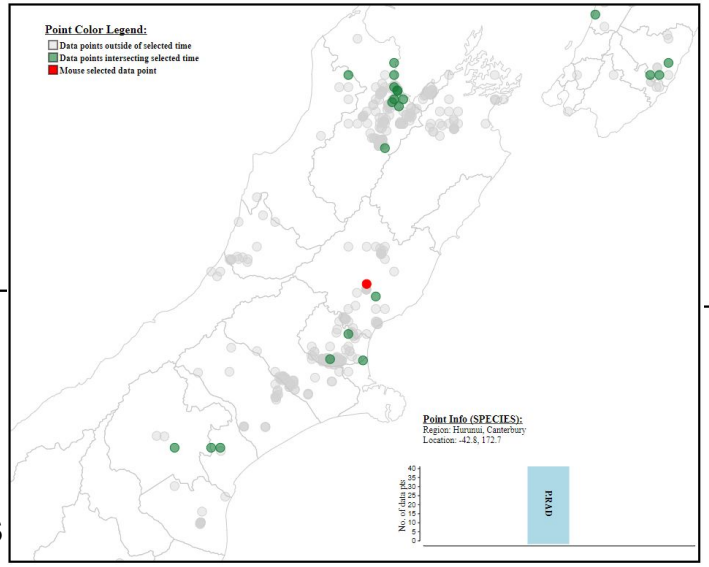
- Numerical analysis



Spatial cluster map

- Spatio-temporal analysis



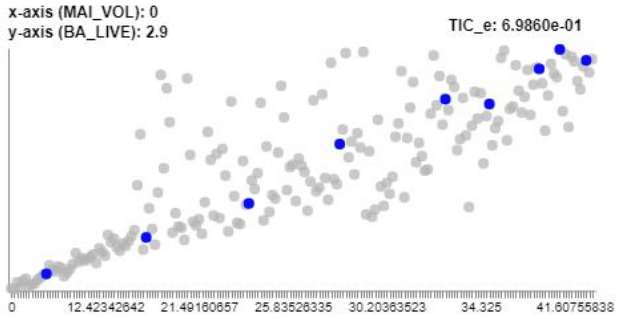


Scatter map

- Geo-location analysis

scion

Facet View

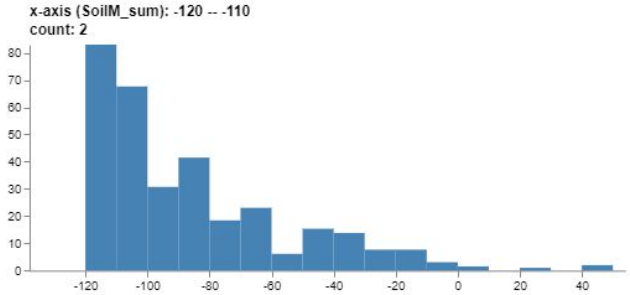


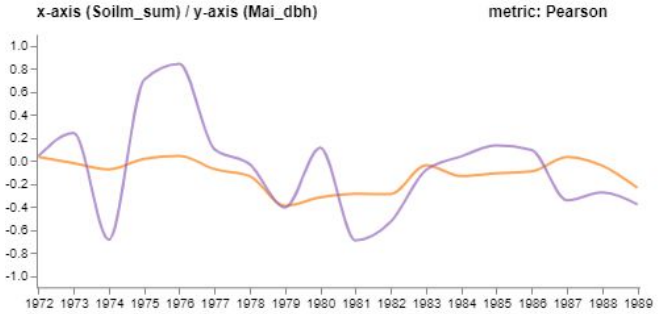
Scatter plots

- Categorical data analysis
- Exploring data relationships

Histograms

- Visualising data distribution





Time-series plot

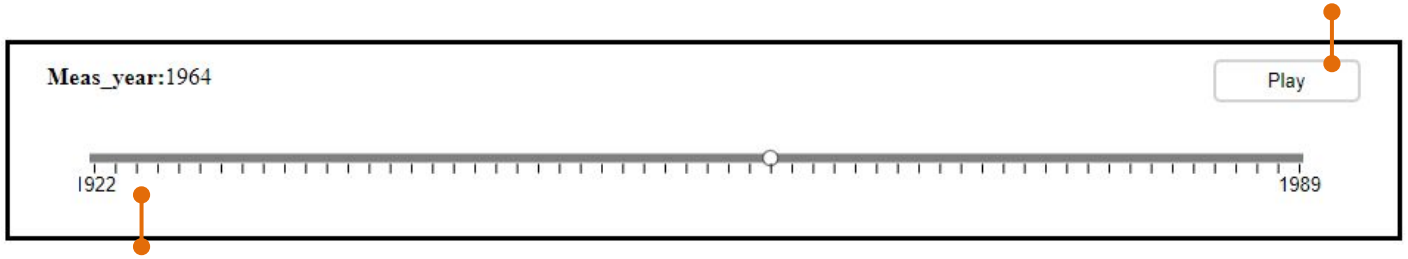
- Temporal pattern analysis

scion

Time Panel

‘Play’ button

- automatic traversal across temporal dimension



Time slider

- Select time points along the temporal dimension
- Interactive analysis with the spatial map and facet view

Allow users to interact and change data represented in both the Facet view and Spatial map along the temporal dimension

scion

Statistical Frameworks

- Statistical analysis
  - Maximal Information Coefficient (MIC)[1] – Linear, non-linear, complex relationship testing
- Spatial analysis
  - Moran's I – Spatial autocorrelation analysis
- Spatio-temporal analysis
  - Hierarchical clustering – Spatial points clustering (allow adaptive clustering of spatial points)
  - Pearson – Quick intra-cluster linear relationship testing between variables

[1] D. N. Reshef *et al.*, "Detecting Novel Associations in Large Data Sets," *Science,* vol. 334, no. 6062, pp. 1518-24, Dec 16 2011

[2] Moran, P. A. P. (1950). "Notes on Continuous Stochastic Phenomena." Biometrika, 37(1): 17—23 doi:10.2307/2332142 JSTOR 2332142

SCION

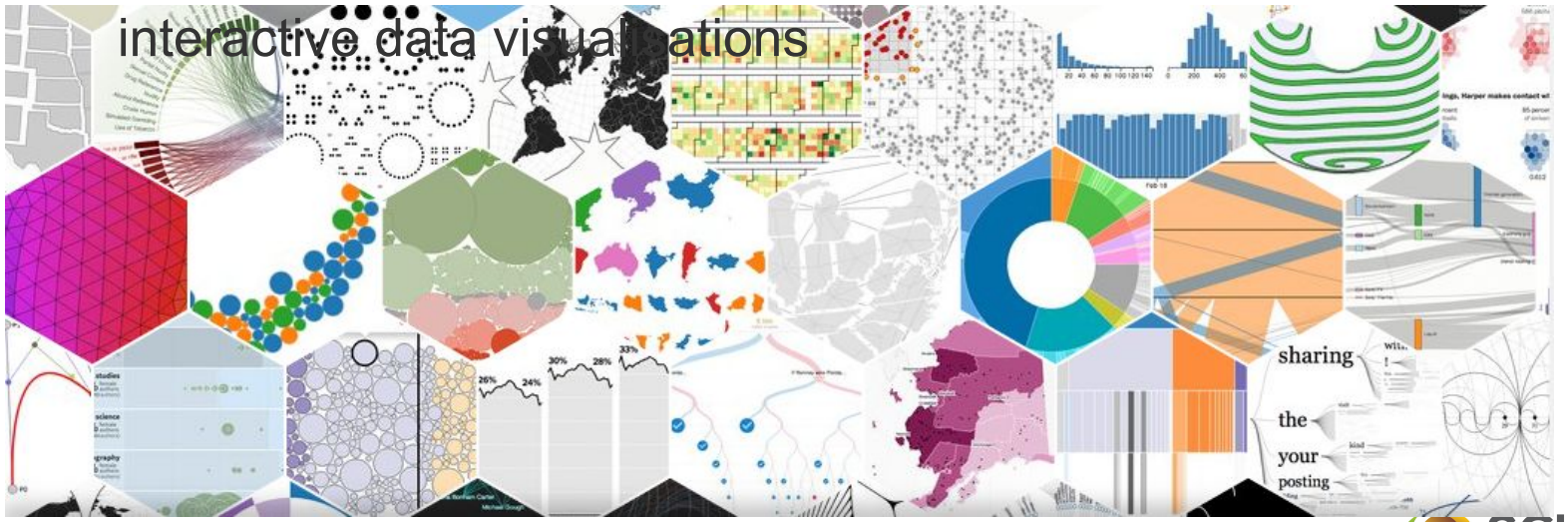
Software Stack

- Python – Backend server and data wrangling
- Scipy + other APIs – Statistical module
- Scikit-learn – Recommender engine
- Vega – Visualisation specification generation
- Javascript + D3 – Visual interface and data visualisation

scion

Data Visualisation – Vega + D3

- Toolkits for building an interactive and dynamic front-end data visualisation interface
- Both APIs are data-driven:
  - APIs responsible for figuring out what elements to add or remove to the visualisation based on changes in the data
  - Simplifies rendering on front-end, allowing responsive and interactive data visualisations

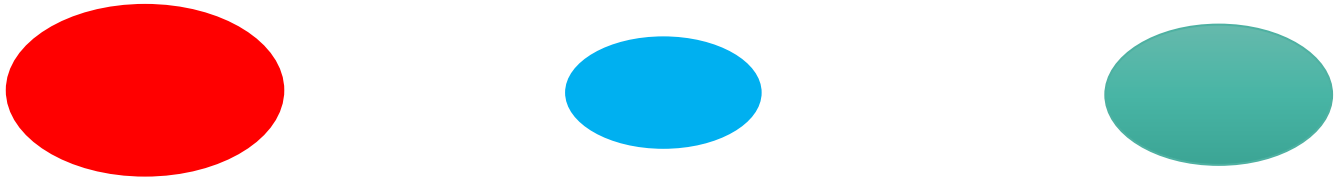


scion

D3 – Data Objects

- Parses arrays of data into data objects

```
var dataset = [{name: Richard, speakerID: 5} ,{name: Wolfgang, speakerID: 2 }]
```



- Manipulates HTML Document Object Model (DOM) instances based on changes in data
  - Enter() – Add new DOM elements when it detects new data objects
  - Update() – Update properties of existing elements based on changes in values for each object
  - Exit() - Remove elements with no corresponding data objects in the dataset

SCION

D3 - Selection

- Robust control over created elements

```
element.append("rect")
    .attr("x", -5)
    .attr("y", -bb.height+ 8)
    .attr("width", panel_width + 10)
    .attr("height", bb.height * 1.2)
    .attr("fill", "lightblue")
    .attr("fill-opacity", 0.3)
    .attr("stroke", "black")
    .attr("stroke-width", 1)
    .on("click", function(){
        var this_box = d3.select(this);
```

```
var element = d3.select("#attributes_selector")
```

```
▼<div id="attributes_selector">
  ▶<g transform="translate(10,30)">…</g>
  ▶<g transform="translate(10,55)">…</g>
  ▶<g transform="translate(10,80)">…</g>
  ▶<g transform="translate(10,105)">…</g>
  ▶<g transform="translate(10,130)">…</g>
  ▶<g transform="translate(10,155)">…</g>
  ▶<g transform="translate(10,180)">…</g>
  ▶<g transform="translate(10,205)">…</g>
  ▶<g transform="translate(10,230)">…</g>
  ▶<g transform="translate(10,255)">…</g>
  ▶<g transform="translate(10,280)">…</g>
  ▶<g transform="translate(10,305)">…</g>
  ▶<g transform="translate(10,330)">…</g>
  ▶<g transform="translate(10,355)">…</g>
  ▶<g transform="translate(10,380)">…</g>
  ▶<g transform="translate(10,405)">…</g>
  ▶<g transform="translate(10,430)">…</g>
  ▶<g transform="translate(10,455)">…</g>
  ▶<g transform="translate(10,480)">…</g>
  ▶<g transform="translate(10,505)">…</g>
  ▶<g transform="translate(10,530)">…</g>
  ▶<g transform="translate(10,555)">…</g>
  ▶<g transform="translate(10,580)">…</g>
  ▶<g transform="translate(10,605)">…</g>
  ▶<g transform="translate(10,630)">…</g>
```

Plot_id --- [categorical] : [1]
Controller --- [categorical] : [1]
Forest --- [categorical] : [1]
Altitude --- [numerical] : [0.99]
Site_index --- [numerical] : [0.98]
Structure --- [categorical] : [0.8]
Plot_status --- [categorical] : [1]
Slope --- [numerical] : [0.9]
Aspect --- [categorical] : [0.72]
Rotation --- [numerical] : [0.8]
Topography --- [categorical] : [0.76]
Geology --- [categorical] : [0.78]
Planted_sph --- [numerical] : [0.85]
Row_space --- [numerical] : [0.74]
Tree_space --- [numerical] : [0.76]

scion

D3 – Other functions

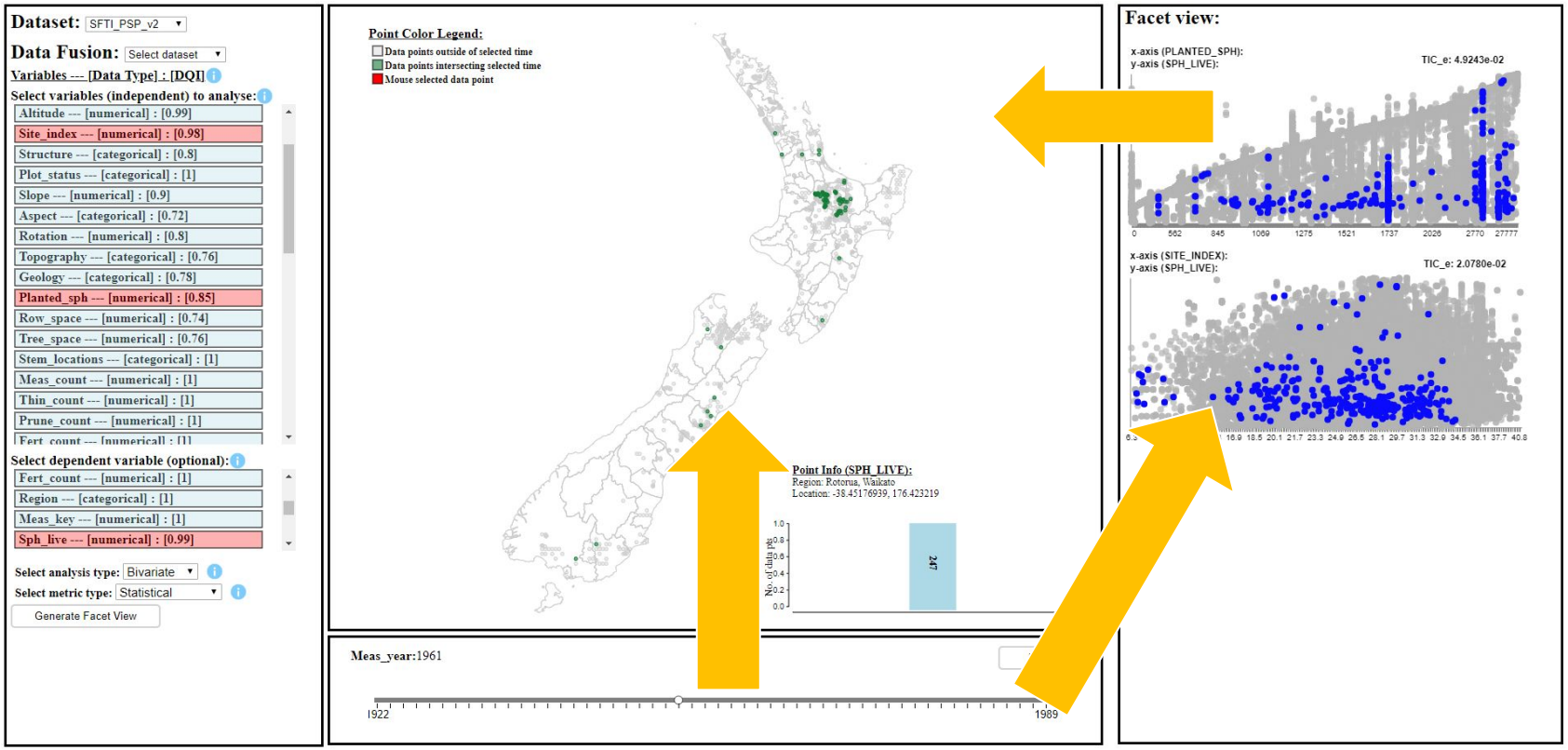
- Smooth visual transitions and an
  - Transition() - timers and dela transitions
  - On() – event handlers to rea ‘click’, ‘mouseover’, ‘mouseo
- Whole list of functions to assist data manipulation and construct intuitive visualisations

- Arrays (Statistics, Search, Transformations, Histograms)
- Axes
- Brushes
- Chords
- Collections (Objects, Maps, Sets, Nests)
- Colors
- Color Schemes
- Contours
- Dispatches
- Dragging
- Delimiter-Separated Values
- Easings
- Fetches
- Forces
- Number Formats
- Geographies (Paths, Projections, Spherical Math, Spherical Shapes, Streams, Transforms)
- Hierarchies
- Interpolators
- Paths
- Polygons
- Quadtrees
- Random Numbers
- Scales (Continuous, Sequential, Diverging, Quantize, Ordinal)
- Selections (Selecting, Modifying, Data, Events, Control, Local Variables, Namespaces)
- Shapes (Arcs, Pies, Lines, Areas, Curves, Links, Symbols, Stacks)
- Time Formats
- Time Intervals
- Timers
- Transitions
- Voronoi Diagrams
- Zooming

D3

- Useful for working with visualising and interacting with large amount of data points

Visualise spatial points for different variables



Manipulate visualisation as data to visualise changes across time

scion

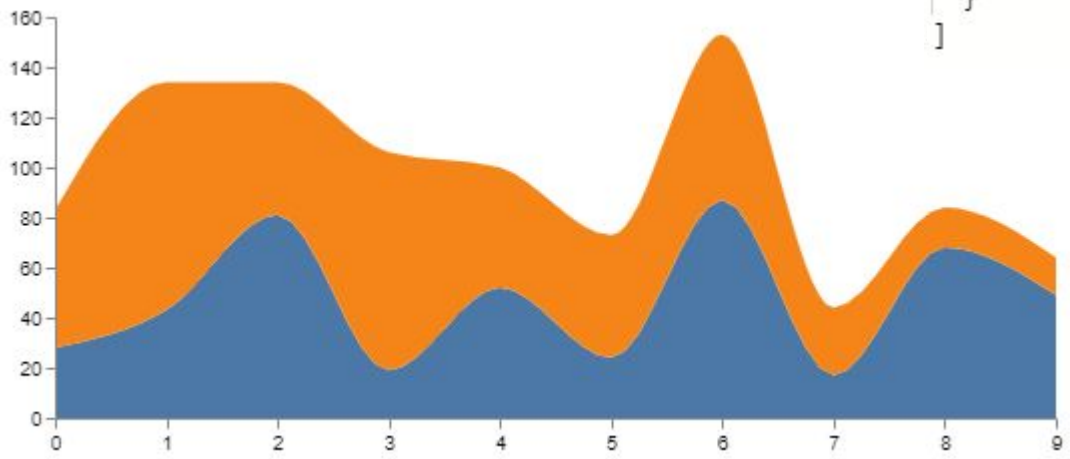
Vega

- Built on D3 – runtime interpreter for a JSON-based visualisation grammar
- Declarative language to ‘describe’ visualisations – abstracting the implementation
- Promotes reusable visualisation design and interoperability
- Great for generating different facet views of the data
  - By dimension
  - By “category” within a variable (i.e. how does student perform across each class)

Vega – describing visualisations

```
"$schema": "https://vega.github.io/schema/vega/v5.json",
"width": 500,
"height": 200,
"padding": 5,

"data": [ ···
],

"scales": [ ··
],

"axes": [ ···
],

"marks": [ ···
]
```

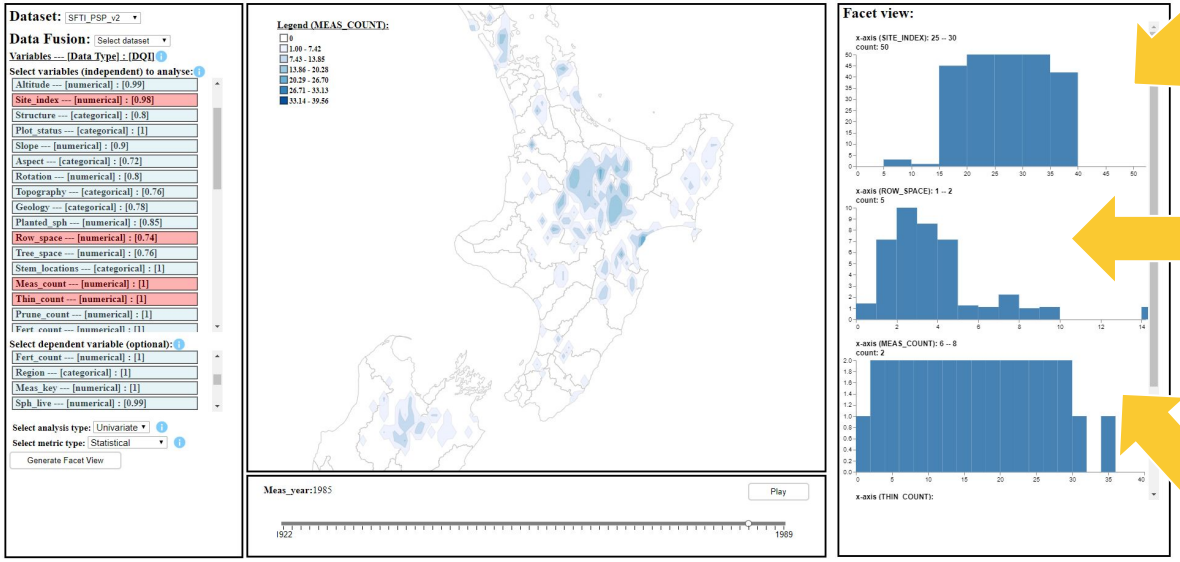
```
"marks": [
  {
    "type": "area",
    "from": {"data": "series"},
    "encode": {
      "enter": {
        "interpolate": {"value": "monotone"},
        "x": {"scale": "x", "field": "x"},
        "y": {"scale": "y", "field": "y0"},
        "y2": {"scale": "y", "field": "y1"},
        "fill": {"scale": "color", "field": "c"}
      },
      "update": {
        "fillOpacity": {"value": 1}
      },
      "hover": {
        "fillOpacity": {"value": 0.5}
      }
    }
  }
]
```

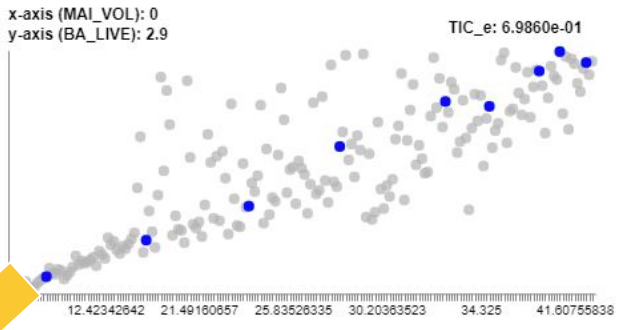


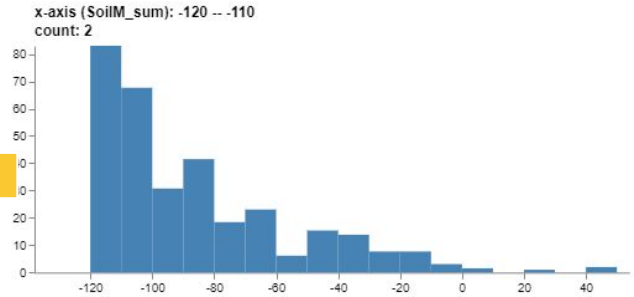
scion

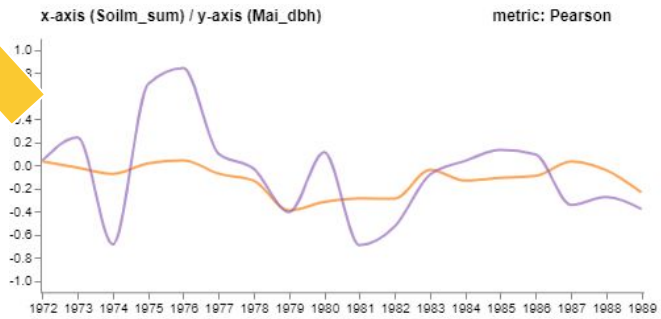
Vega

- Handling visualisation of different data types





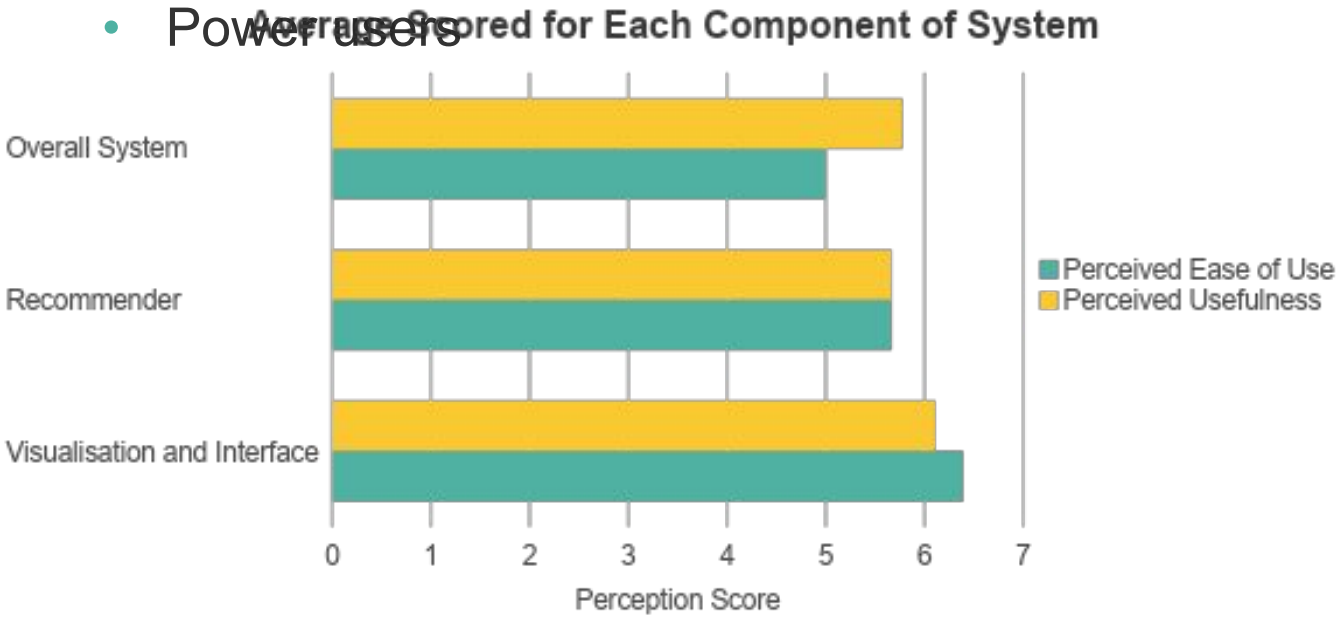




scion

User study

- 2 user studies conducted across the project duration
  - Perceived usefulness of system
  - Facilitating data exploratory efforts
- Different groups of users
  - Non-data analysts
  - Power users



scion

D3 / Vega – Cons

- Steep learning curve
  - Require an awareness of how the data is structured when implementing the visualisation
  - Different kind of thinking – how can I generalise my implementation to work with different data
- Vega – still lack robust support for spatial data visualisation
  - custom maps
- Toolkits still restricted by resources of browsers
  - Memory, bandwidth
- Data needs to be sent to client-side
  - Challenges with sensitive data

Acknowledgements

- Science for Technological Innovation National Science Challenge program (SfTI).
- Dr Stephen MacDonell, AUT
- Christine Dodunski, Scion PSP administrator

scion

www.scionresearch.com

Prosperity from trees *Mai i te ngahere oranga*