

A Visual Recommender Framework for Exploratory Data Analytics

Alan Tan, Yue Lin, Ralf Gommers
20th Feb 2019



Permanent Sample Plot (PSP) Database

■ What is it?

- Field measurements of tree plots geographically distributed across New Zealand.

■ Scale of the Data

- More than 100 years of field measurements.
- Over 100 measured and derived features on trees, silvicultural practices and other aspects of a tree stand.
- Approximately over 29,000 tree sample plots geographically distributed across NZ.

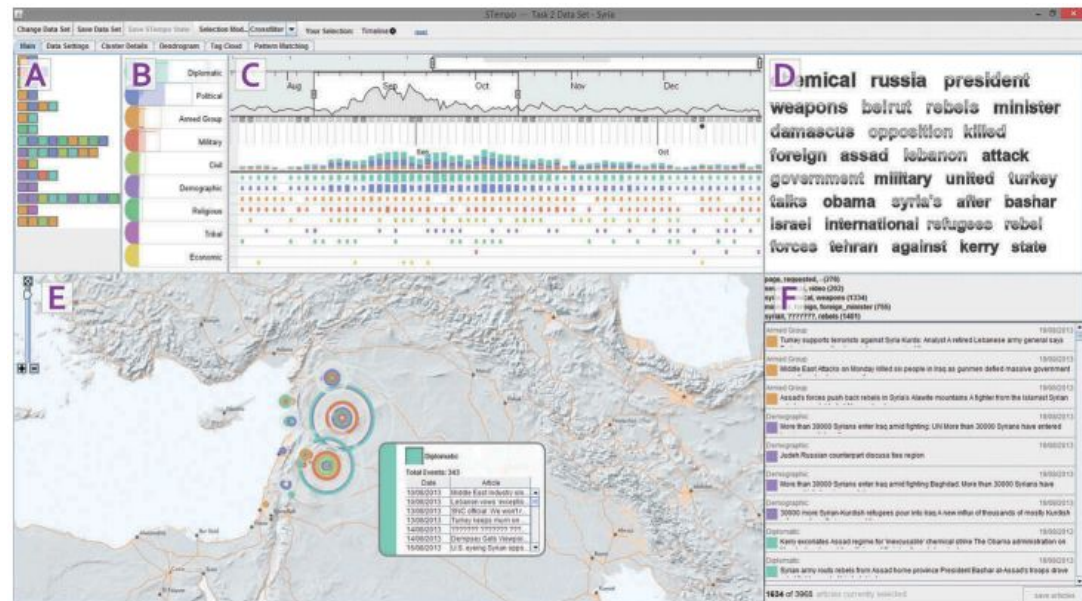


■ The issue

- Difficult to conduct holistic analysis on the data

Challenges

- Challenging to explore complex spatio-temporal datasets (i.e. PSP dataset)
 - Time consuming
 - Lack of computational tools
- Existing tools emphasised on manual efforts for identifying and testing for relationships¹
 - E.g. STempo²



[1] N. Andrienko, G. Andrienko, and P. Gatalsky, "Exploratory spatio-temporal visualization: an analytical review," *Journal of Visual Languages & Computing*, vol. 14, no. 6, pp. 503-541, 2003/12/01/ 2003

[2] 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

Our Approach

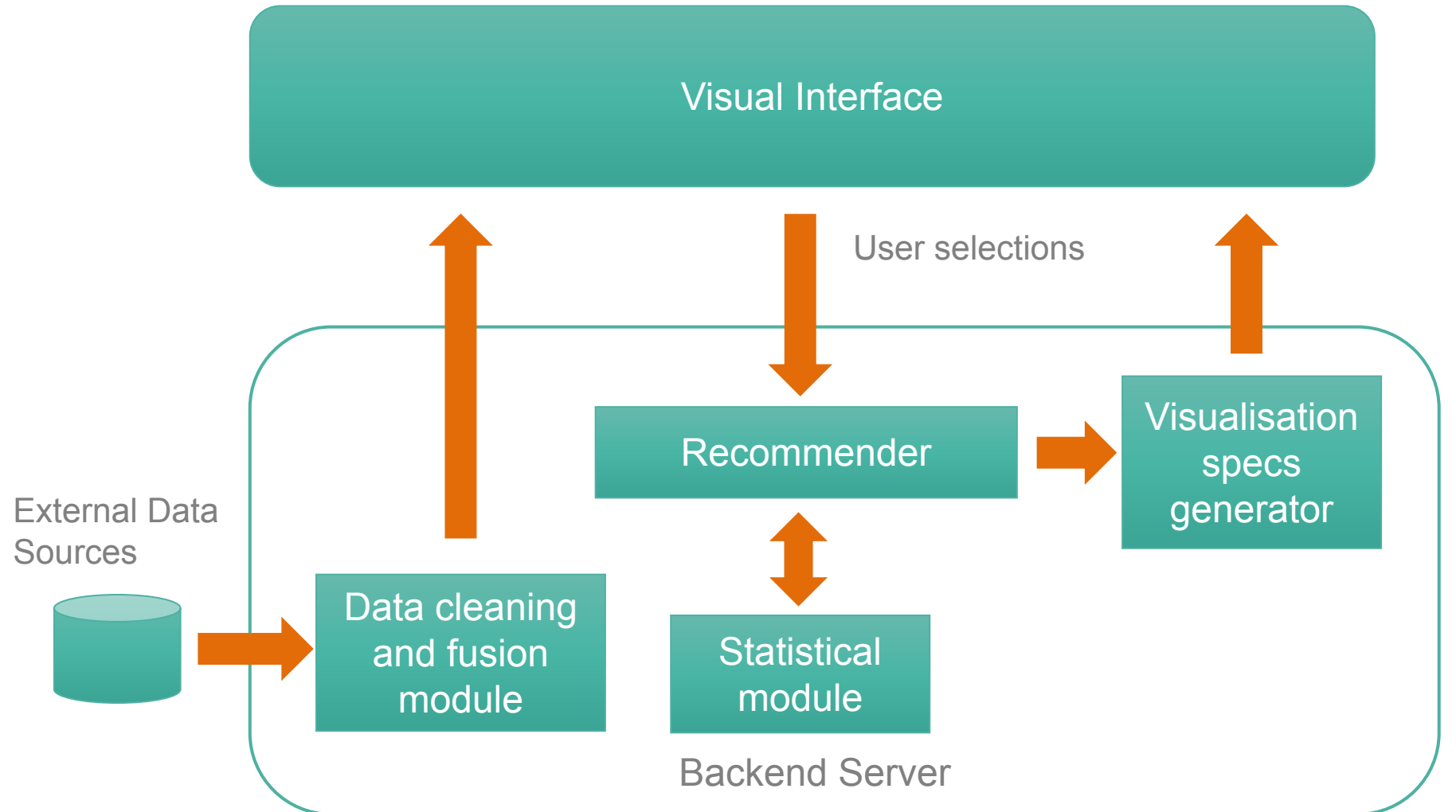
Objectives

- Explore new methods for allowing exploratory analytics of complex spatio-temporal datasets.
- Improve significantly the breadth of data explored by users per time spent.

Approach

- Interactive visualisation + Recommender.

Visual Recommender Architecture



Visual Recommender User Interface

Variable panel

Spatial Map

Dataset: [SFTI_PSP_v2]

Data Fusion: [Select dataset]

Variables --- [Data Type] : [DQI]

Select variables (independent) to analyse:

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]

Stem_locations --- [categorical] : [1]

Meas_count --- [numerical] : [1]

Thin_count --- [numerical] : [1]

Prune_count --- [numerical] : [1]

Fert_count --- [numerical] : [1]

Select dependent variable (optional):

Fert_count --- [numerical] : [1]

Region --- [categorical] : [1]

Meas_key --- [numerical] : [1]

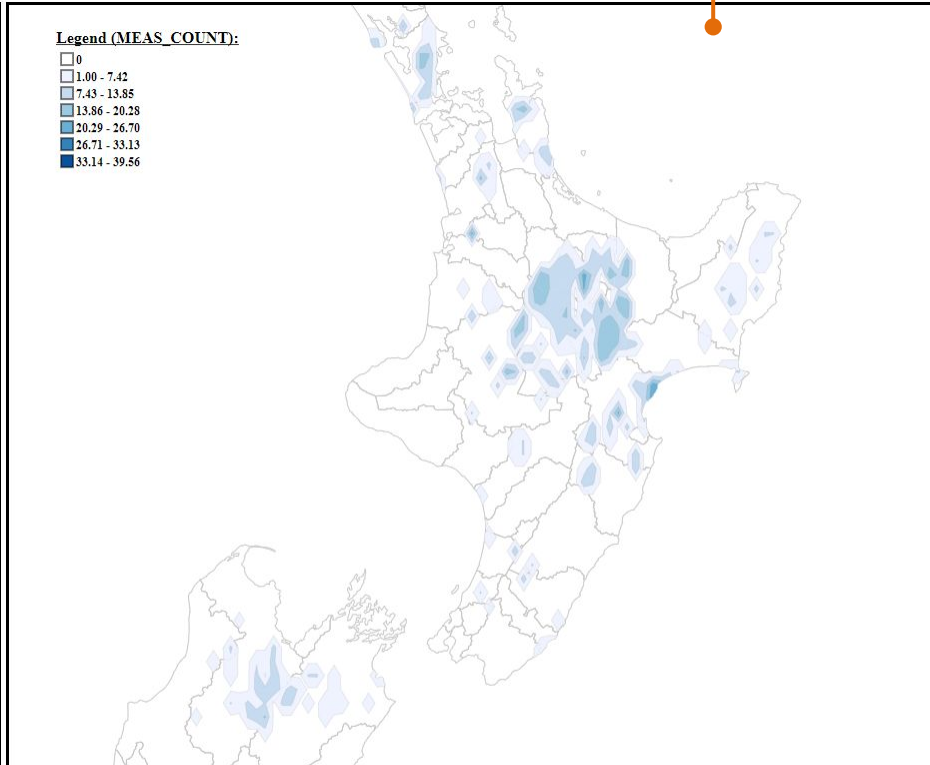
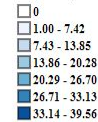
Sph_live --- [numerical] : [0.99]

Select analysis type: [Univariate]

Select metric type: [Statistical]

Generate Facet View

Legend (MEAS_COUNT):



Meas_year:1985

Play

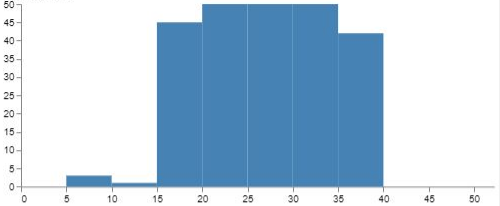
1922

1989

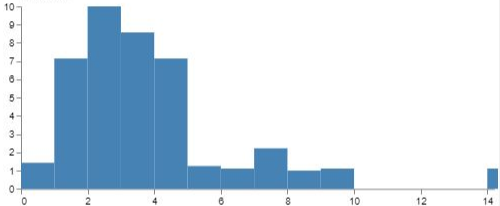
Time Panel

Facet view:

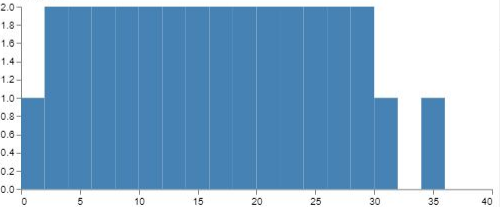
x-axis (SITE_INDEX): 25 -- 30
count: 50



x-axis (ROW_SPACE): 1 -- 2
count: 5



x-axis (MEAS_COUNT): 6 -- 8
count: 2



x-axis (THIN_COUNT):

Facet View

Variable Panel

Dataset: SFTI_PSP_v2

Data Fusion: Select dataset

Variables --- [Data Type] : [DQI]

Select variables (independent) to analyse:

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]
Stem_locations --- [categorical] : [1]
Meas_count --- [numerical] : [1]
Thin_count --- [numerical] : [1]
Prune_count --- [numerical] : [1]
Fert_count --- [numerical] : [1]

Select dependent variable (optional):

Plot_size --- [numerical] : [1]
Plot_age --- [numerical] : [1]
Age_year --- [numerical] : [1]
Mai_dbh --- [numerical] : [0.98]

Select analysis type: Univariate

Select metric type: Statistical

Generate Facet View

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

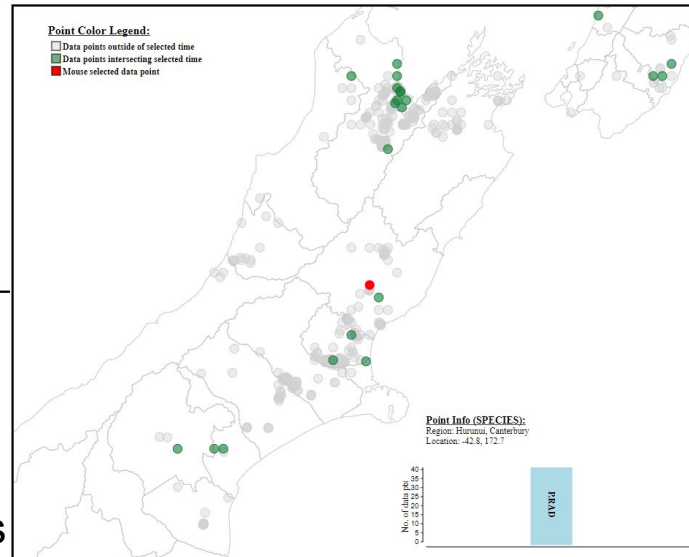
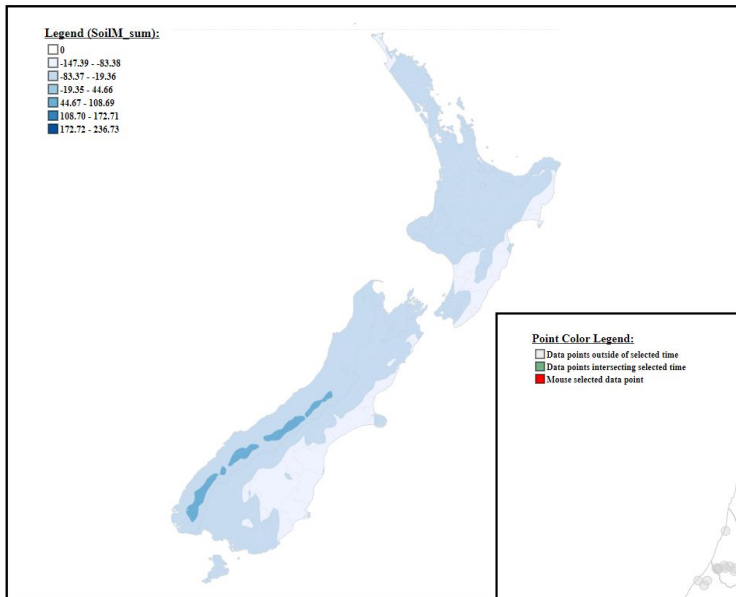
- Control types and mode of analysis

Spatial Map

- Different modes of spatial visualisation

Heatmap

- Numerical analysis

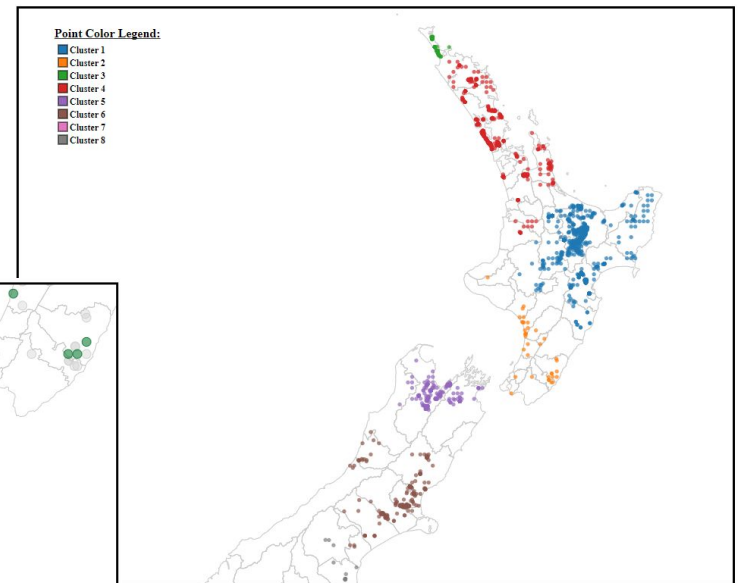


Scatter map

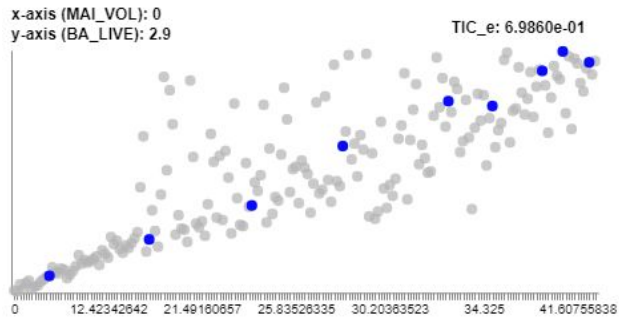
- Geo-location analysis

Spatial cluster map

- Spatio-temporal analysis



Facet View

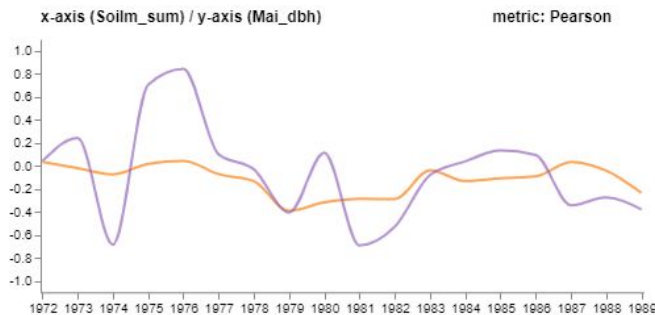
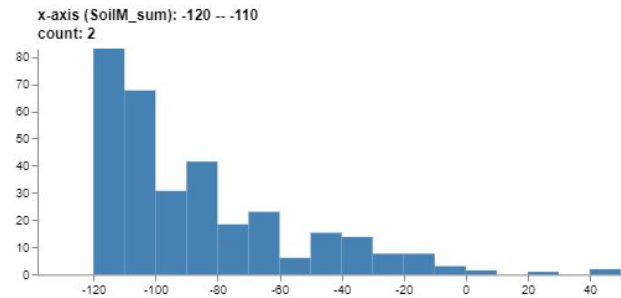


Scatter plots

- Categorical data analysis
- Exploring data relationships

Histograms

- Visualising data distribution



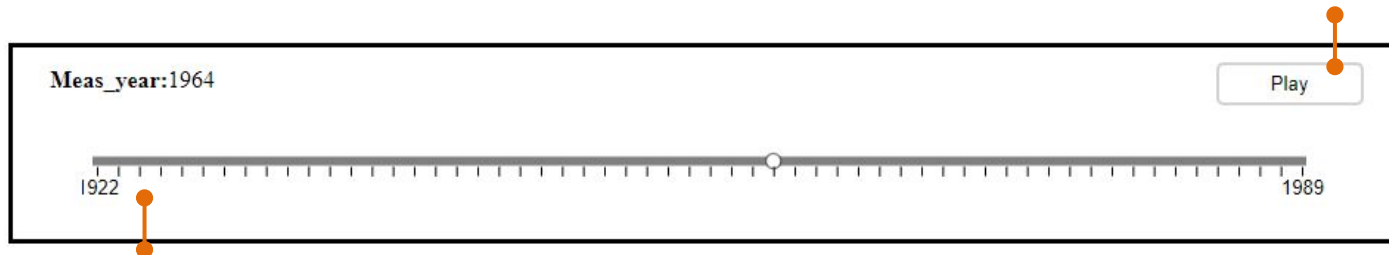
Time-series plot

- Temporal pattern analysis

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

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

[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

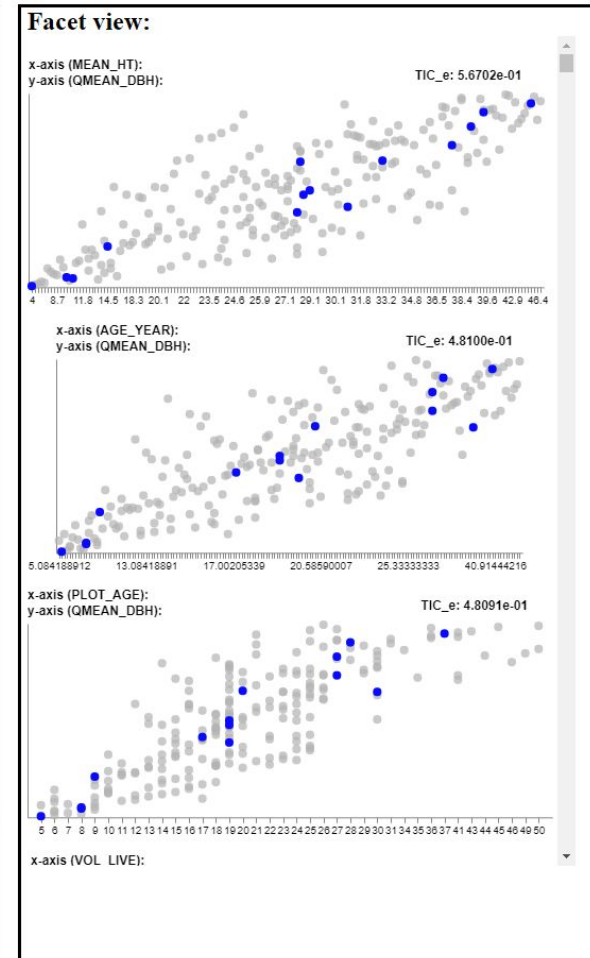
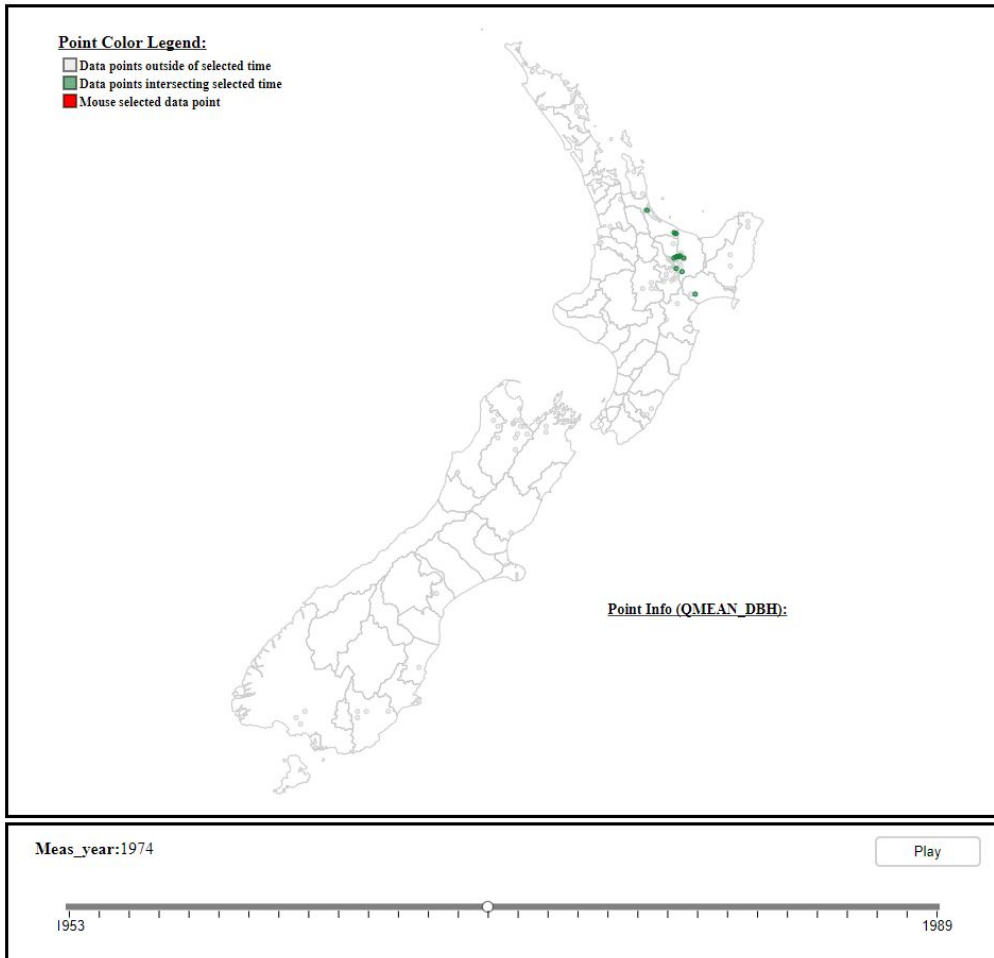
Software Stack

- 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

Example – Statistical Analysis

MIC

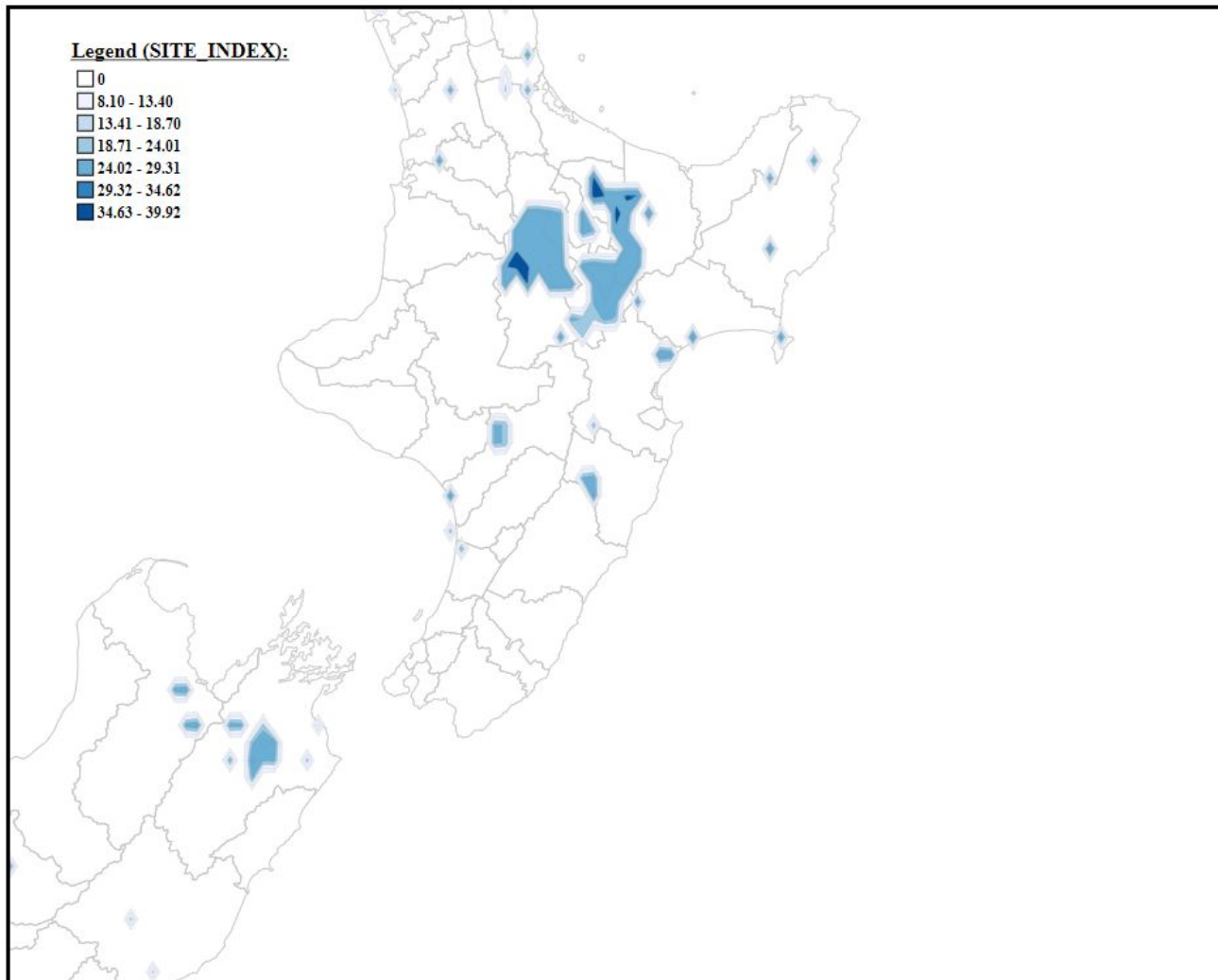
- Detecting bivariate relationships



Example – Spatial Analysis

Moran's I

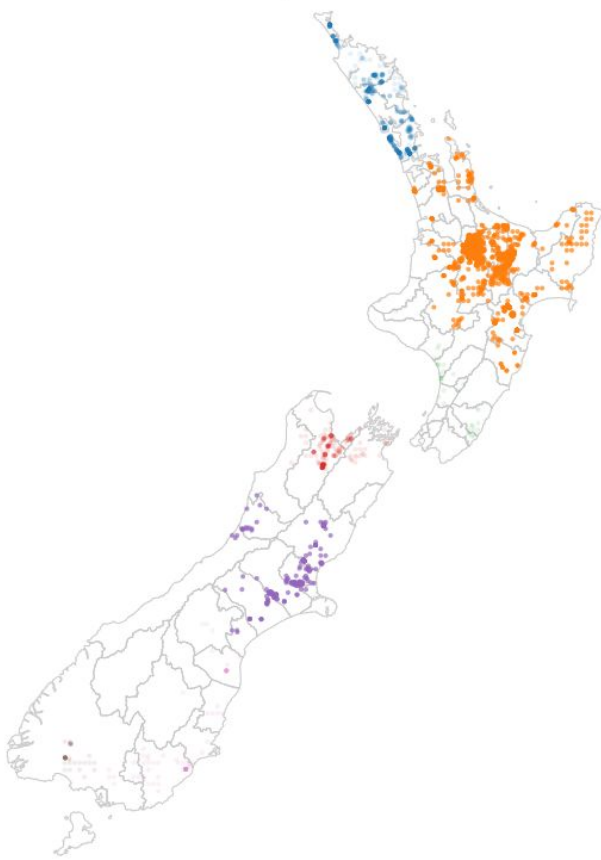
- Cluster / Dispersion of values across different spatial locations



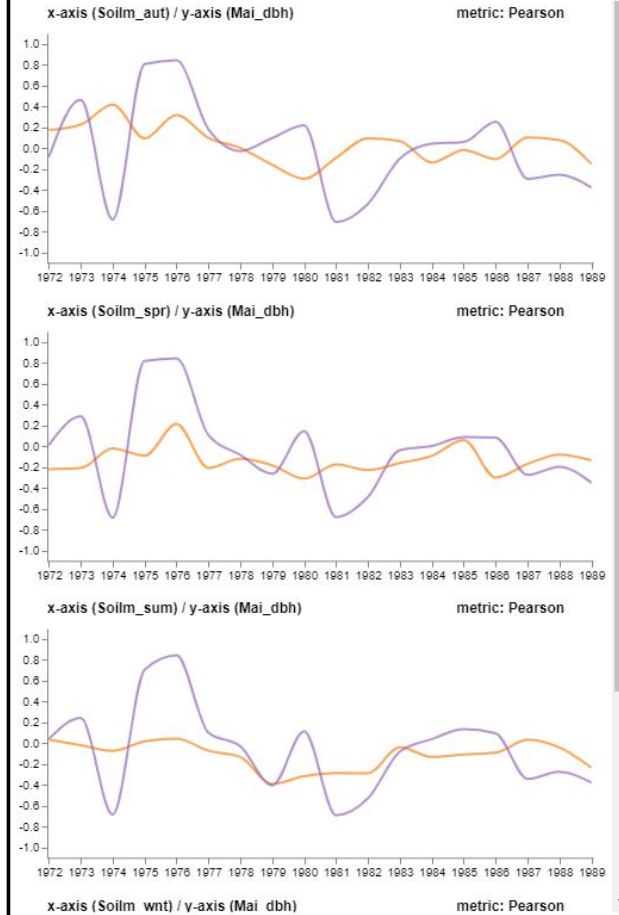
Example – Spatio-temporal Analysis

Point Color Legend:

- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5
- Cluster 6
- Cluster 7

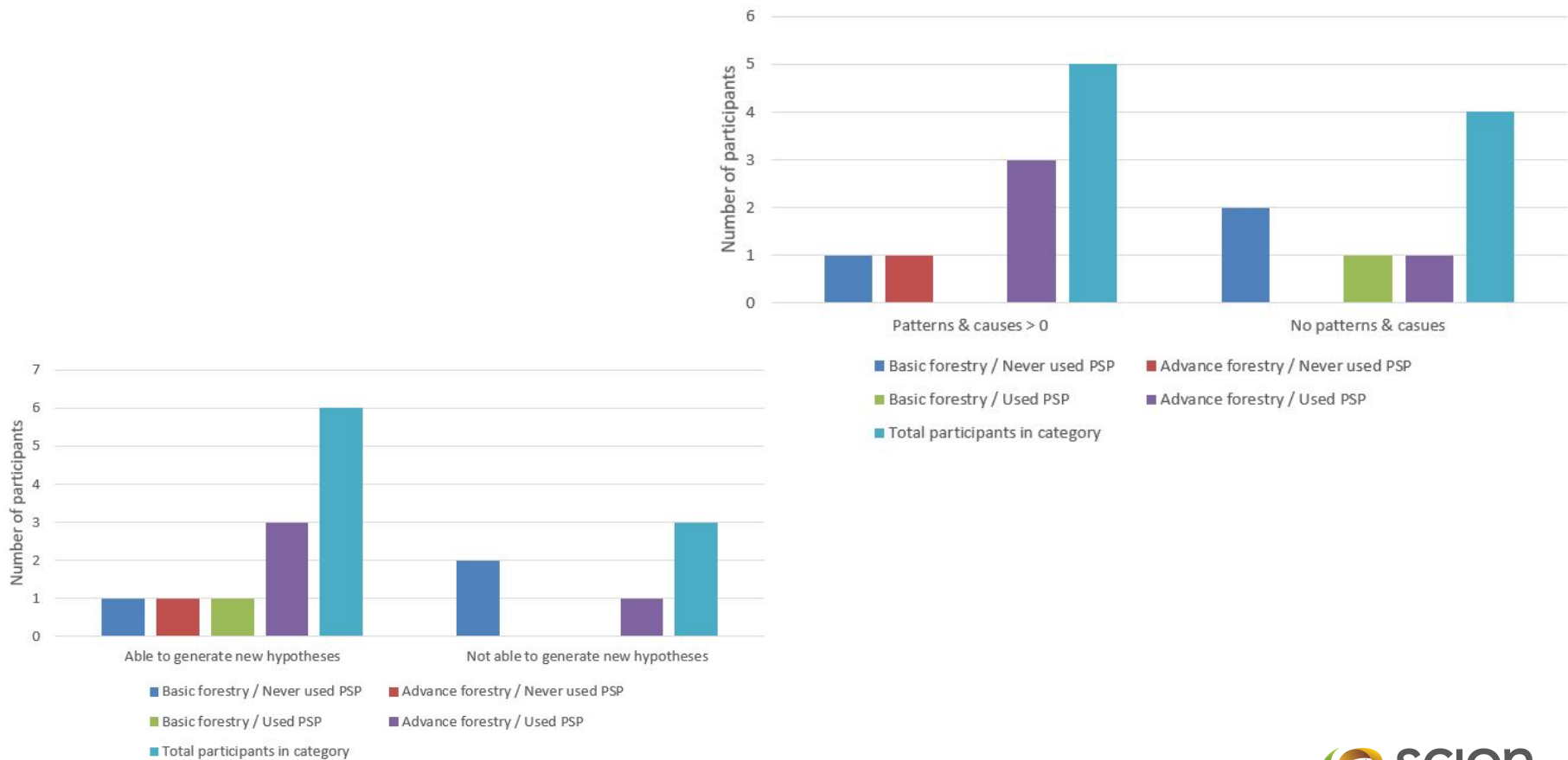


Facet view:



Results of User Study

- Pilot user study
- Participants with different expertise and skill sets



Future Work

- Methods for time-series analysis
- Advanced recommendation algorithms
- Incorporate user sessions
 - Data access control
 - Customising recommendations to different type of users

Acknowledgements

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

www.scionresearch.com

Alan Tan

alan.tan@scionresearch.com



Prosperity from trees *Mai i te ngahere oranga*

Scion is the trading name of the New Zealand Forest Research Institute Limited