

```
> sessionInfo()  
[1] "June 30 - July 3, 2015"  
[2] "Aalborg, Denmark"
```

New interactive visualization tools for exploring high dimensional data in R

Wayne Oldford & Adrian Waddell

UNIVERSITY OF
WATERLOO



Challenge

- p values on each of n individuals
- modern data: n , or p , or both, can be very large



- can have non-obvious variables, complex, unanticipated structure, ...

Challenge

Large p

- visually, we're constrained to small p
 - ✦ locations: $p < 4$
- comprehension depends on only a few dimensions
 - ... at a time
- need to connect many smaller dimensional views

Approach

exploratory

- interactively slice and dice
 - ✦ find interesting low dimensional structure
 - ✦ cut away uninteresting subsets
 - ✦ dynamic linking across data
- integrate with statistical methods and models
 - ✦ scagnostics, dimension reduction, model building, ...
- mix direct manipulation and scripting in analysis
 - ✦ whichever is more powerful/natural
 - ✦ extendability is essential



Loon R
demo uses



... a whirlwind tour

These slides contain videos.

If your PDF reader cannot display the videos on the following pages use the Adobe Acrobat reader:

get.adobe.com/reader

loon

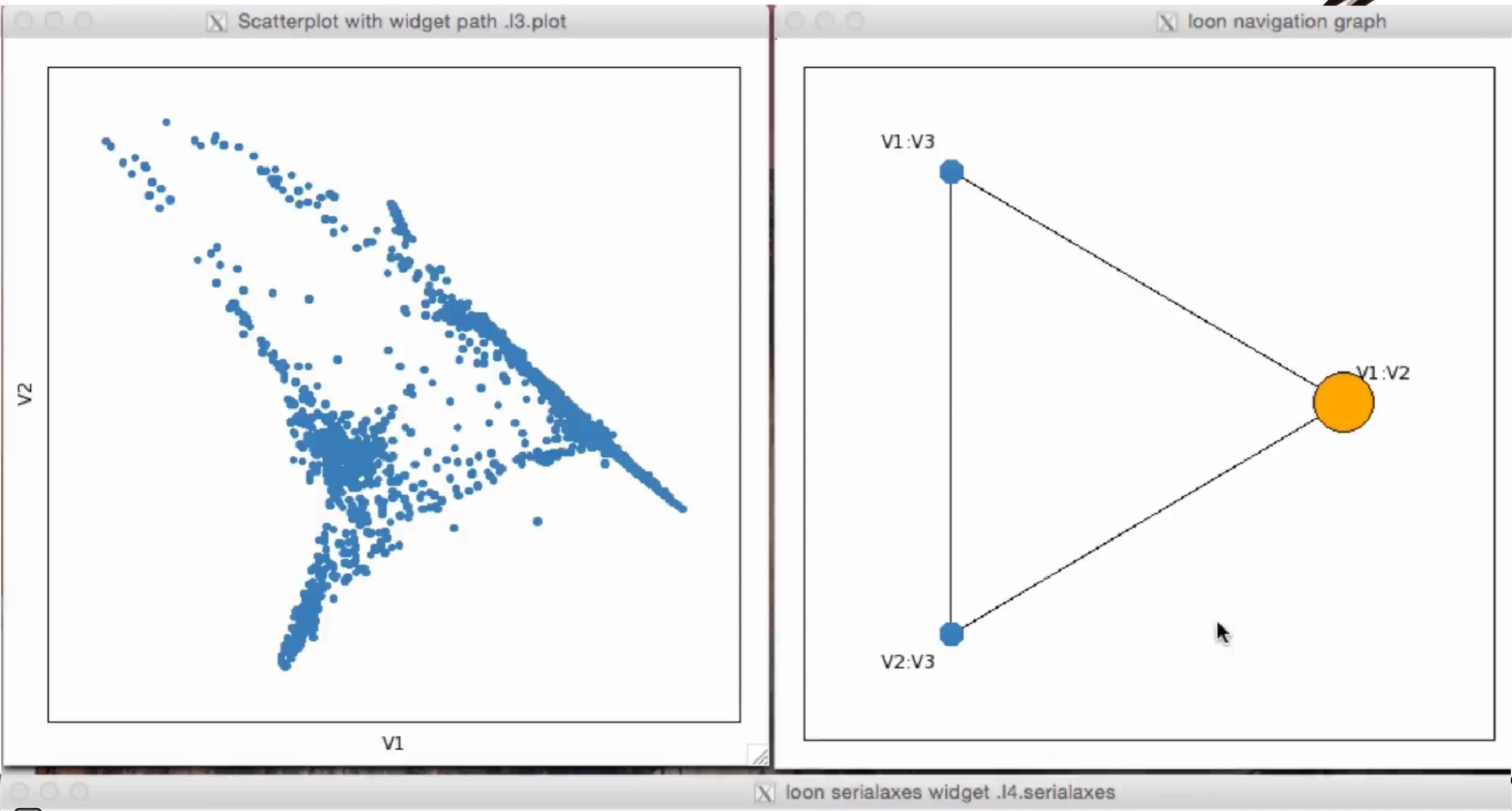
scagnostics, RDRTtoolbox

1,965 images



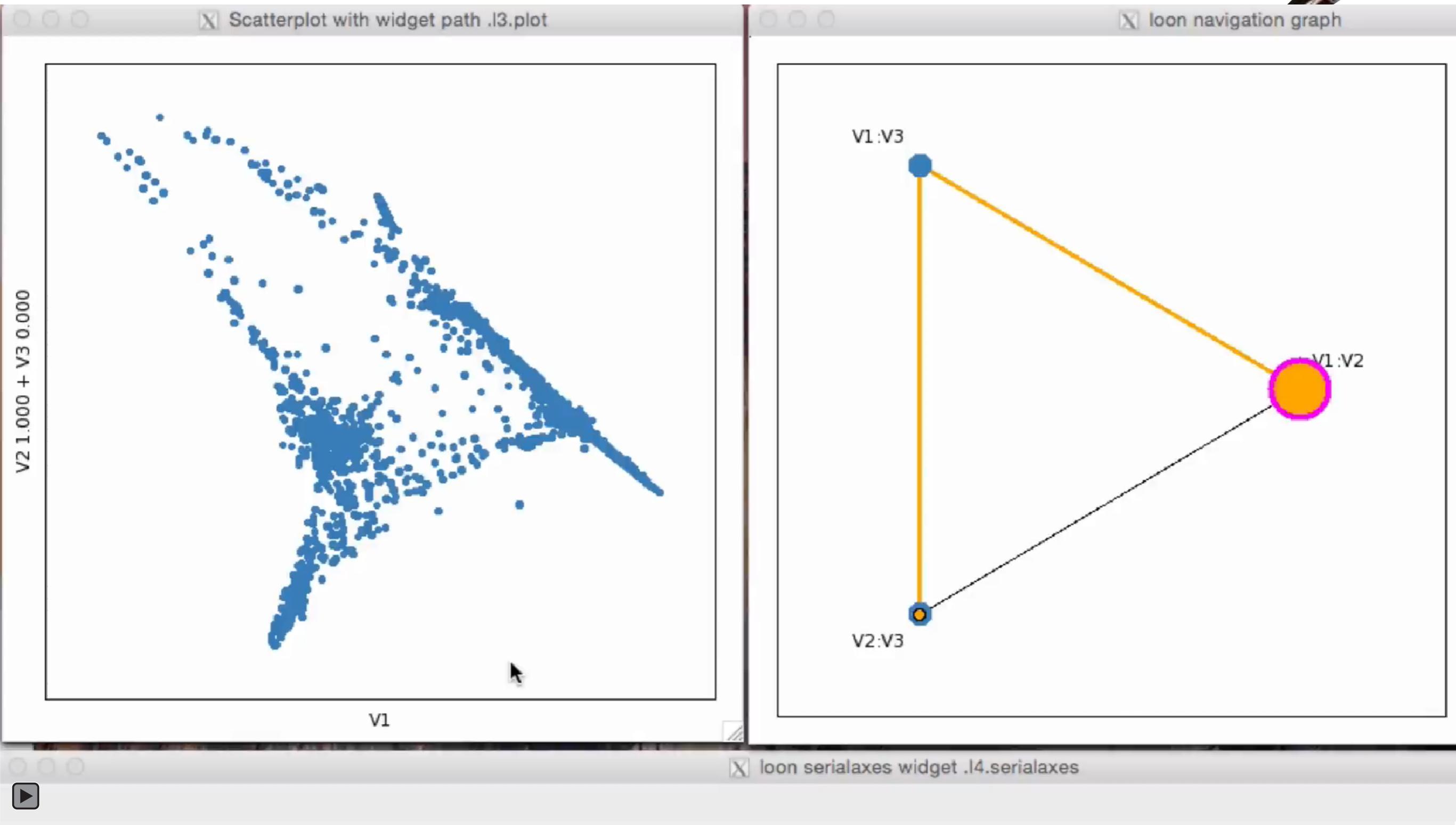
Example: images

loon



travel from one 2d space to another

loon



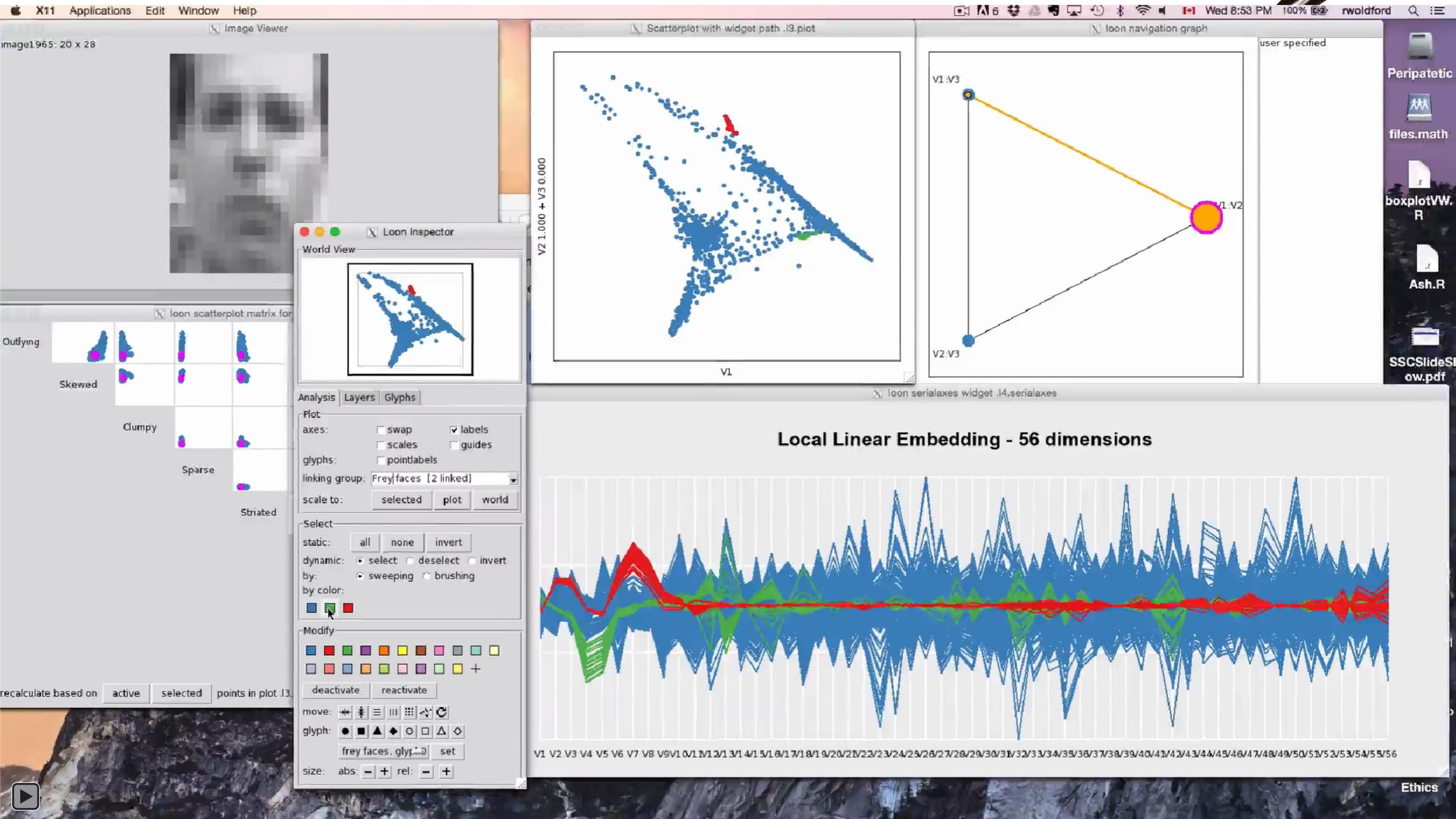
link the scatterplot to a parallel coordinate plot

loon

The screenshot displays the loon software interface on a Mac OS X desktop. The main window is titled "Scatterplot with widget path .I3.plot" and shows a scatterplot of data points in a 2D space defined by axes V1 and V2:V3 + V3:0.000. A red group of points is highlighted. To the right, a "loon navigation graph" shows a network of nodes and edges, with a yellow node labeled V1:V2 and a blue node labeled V1:V3. Below the scatterplot, a "loon serialaxes widget .I4.serialaxes" displays a time-series plot titled "Local Linear Embedding - 56 dimensions". The plot shows multiple overlapping time series, with a red series highlighted. The x-axis is labeled with indices from V1 to V56. A "Loon Inspector" window is open in the foreground, showing various settings for the plot, including axes, glyphs, linking group, and select options. The desktop background features a penguin illustration in the top right corner and a dock with several application icons.

Focus on the red group

loon



Focus on the green group

Challenge

Look at all scatterplots

- p ... overall dimensionality (images, $p=56$ and 560)

p	10	20	56	560
$\binom{p}{2}$	45	190	1,540	156,520

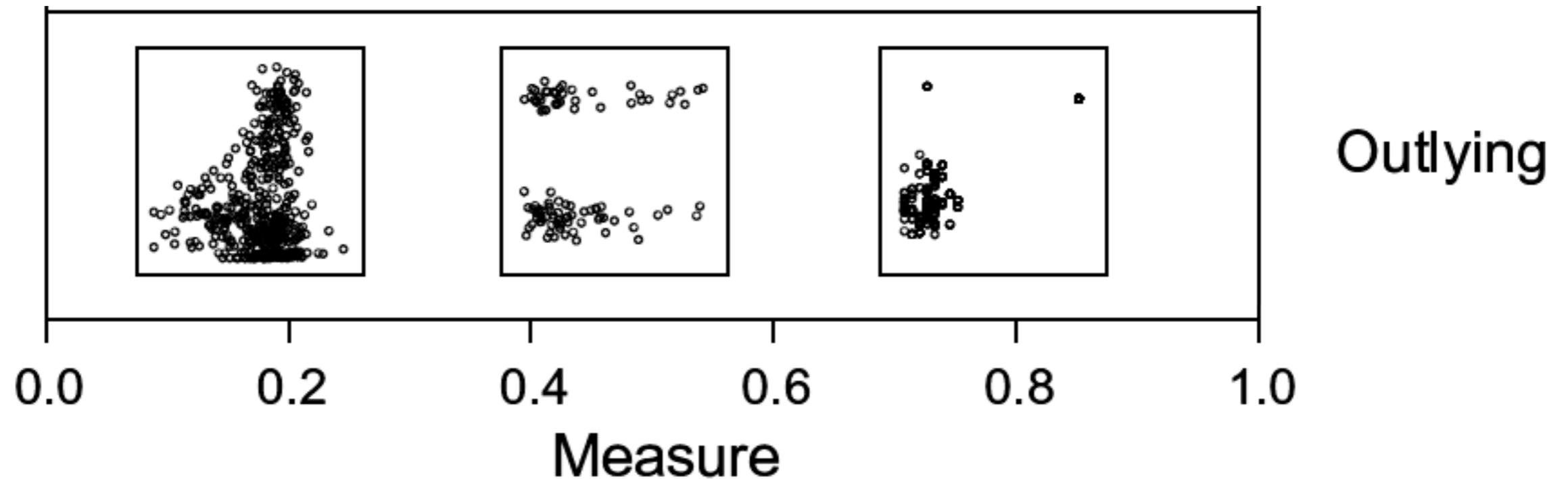
Need to look at the most interesting ones

Interesting variable pairs

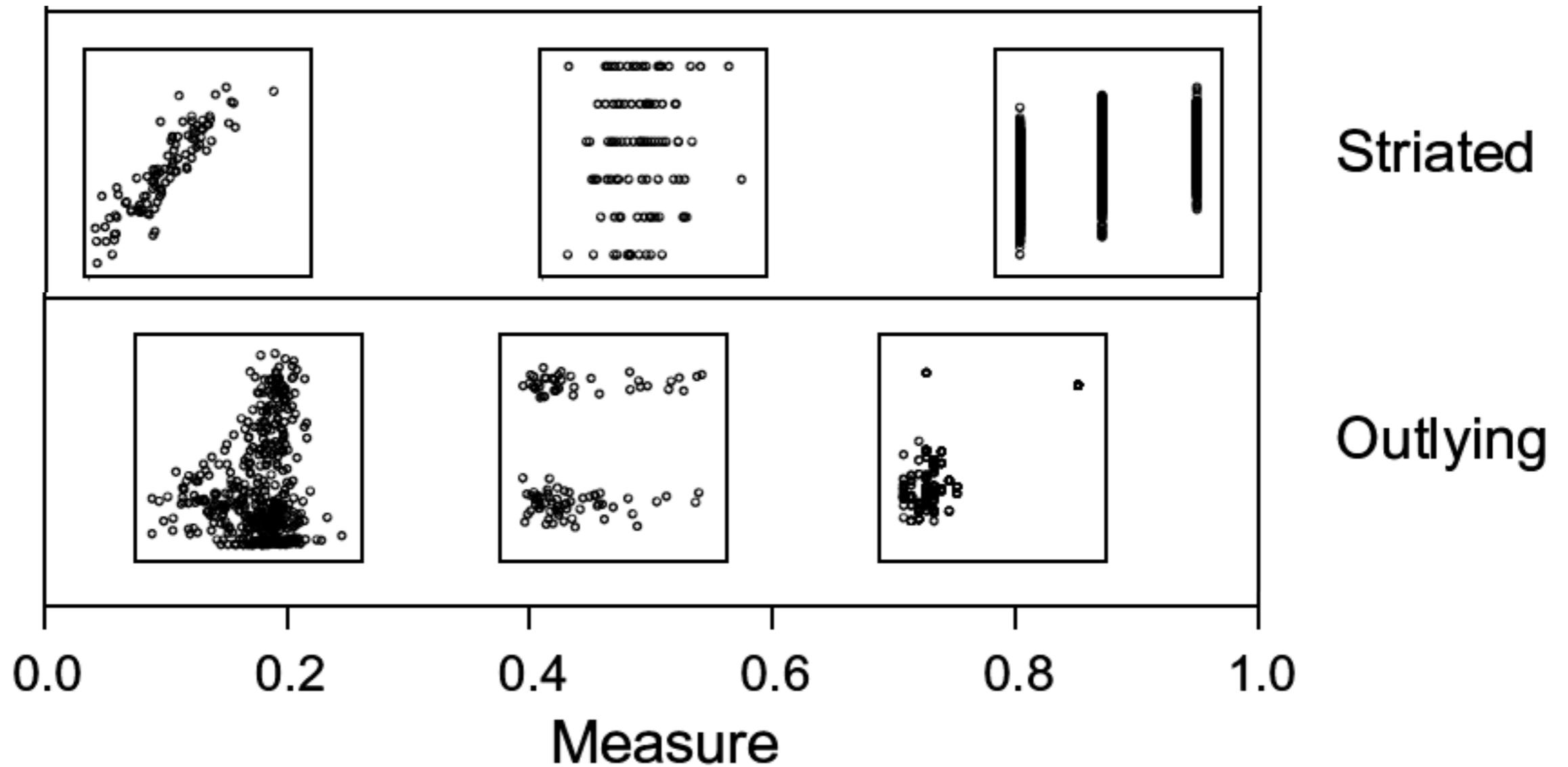
Cognostics (**Computer aided diagnostics**)

Scagnostics ... **Scatterplot cognostics** (Tukey & Tukey, 1985)

Wilkinson et al (2006+) provide such measures

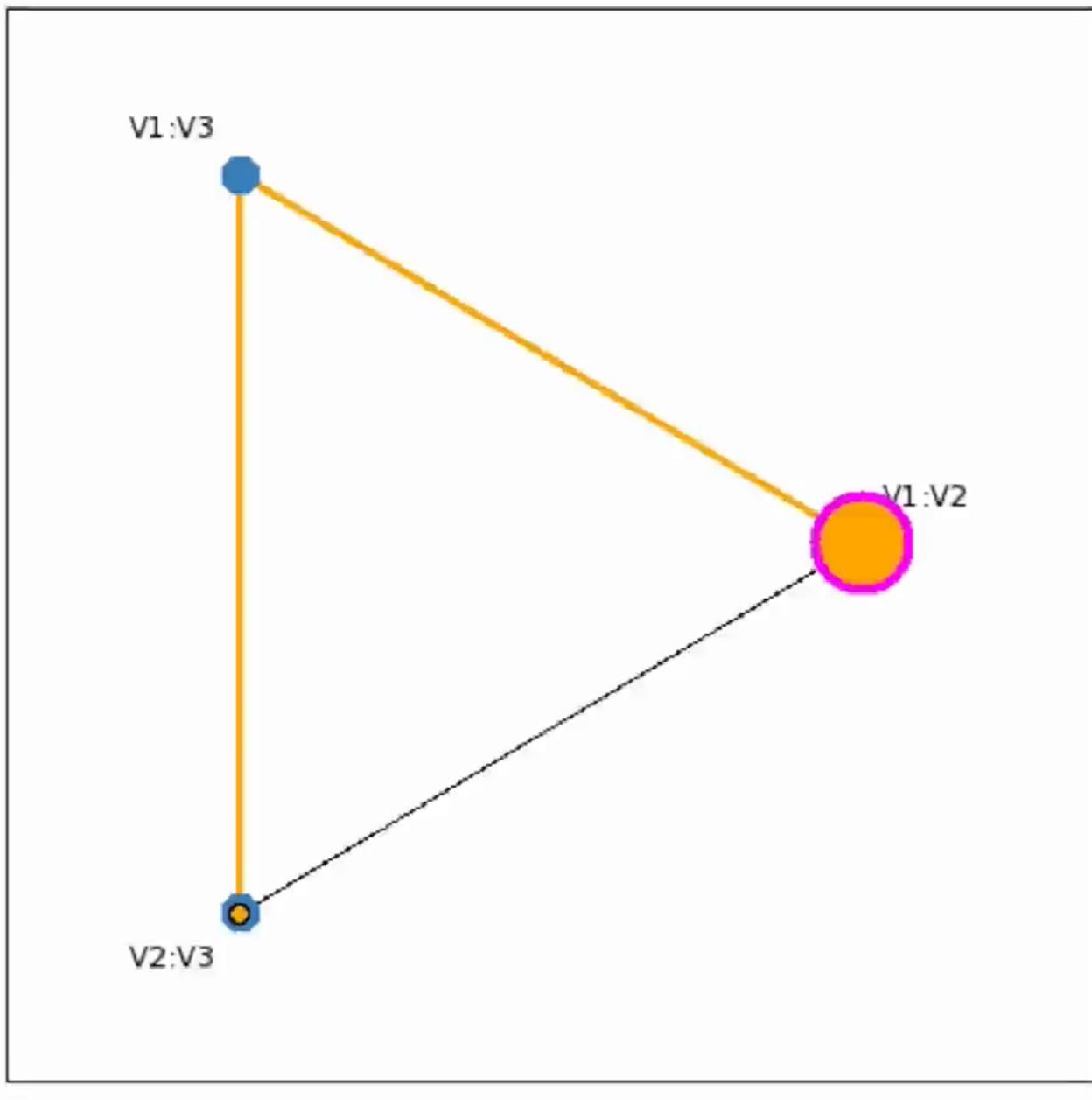
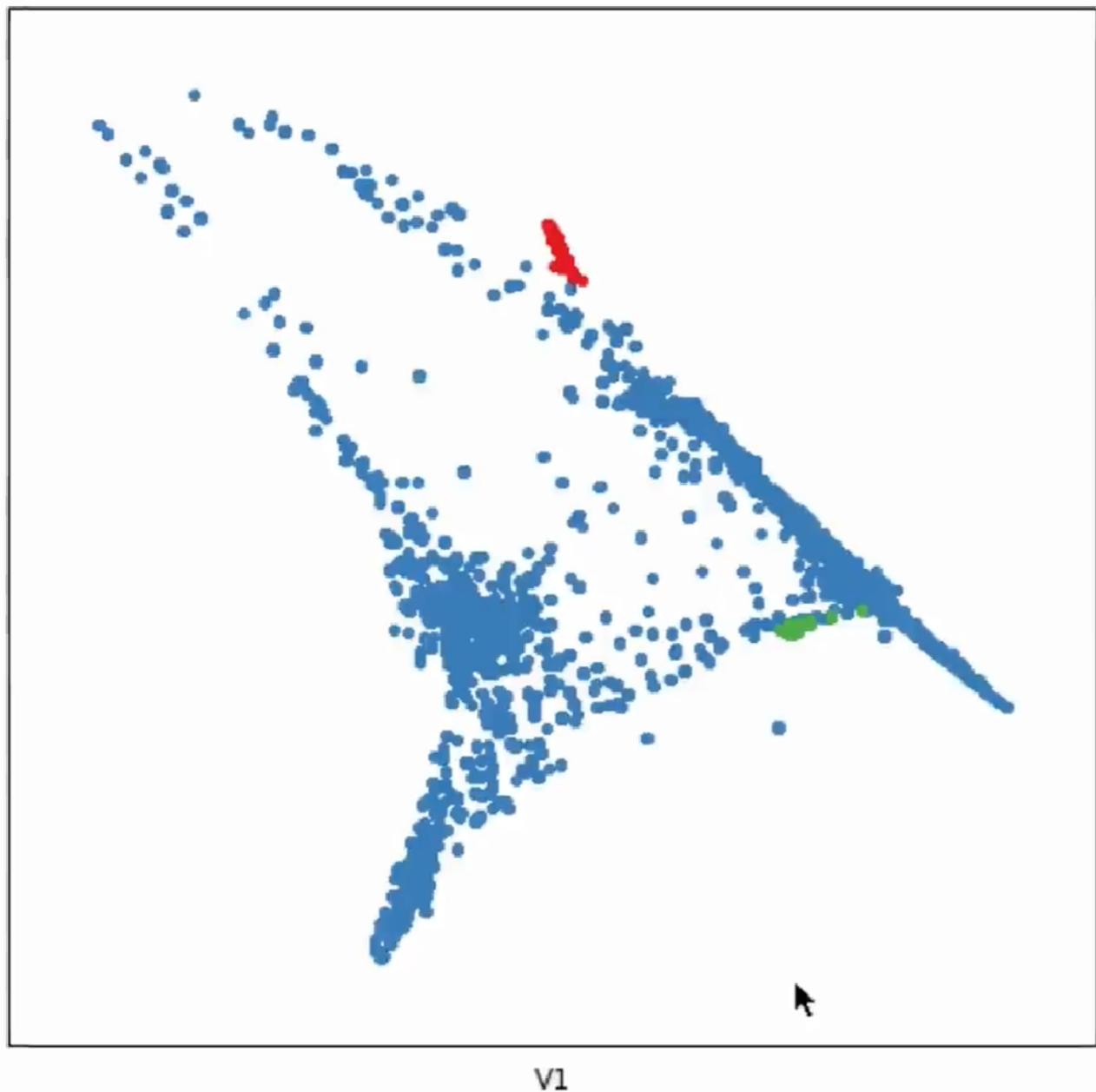


Interesting node pairs



There are nine such scagnostics measures

loon

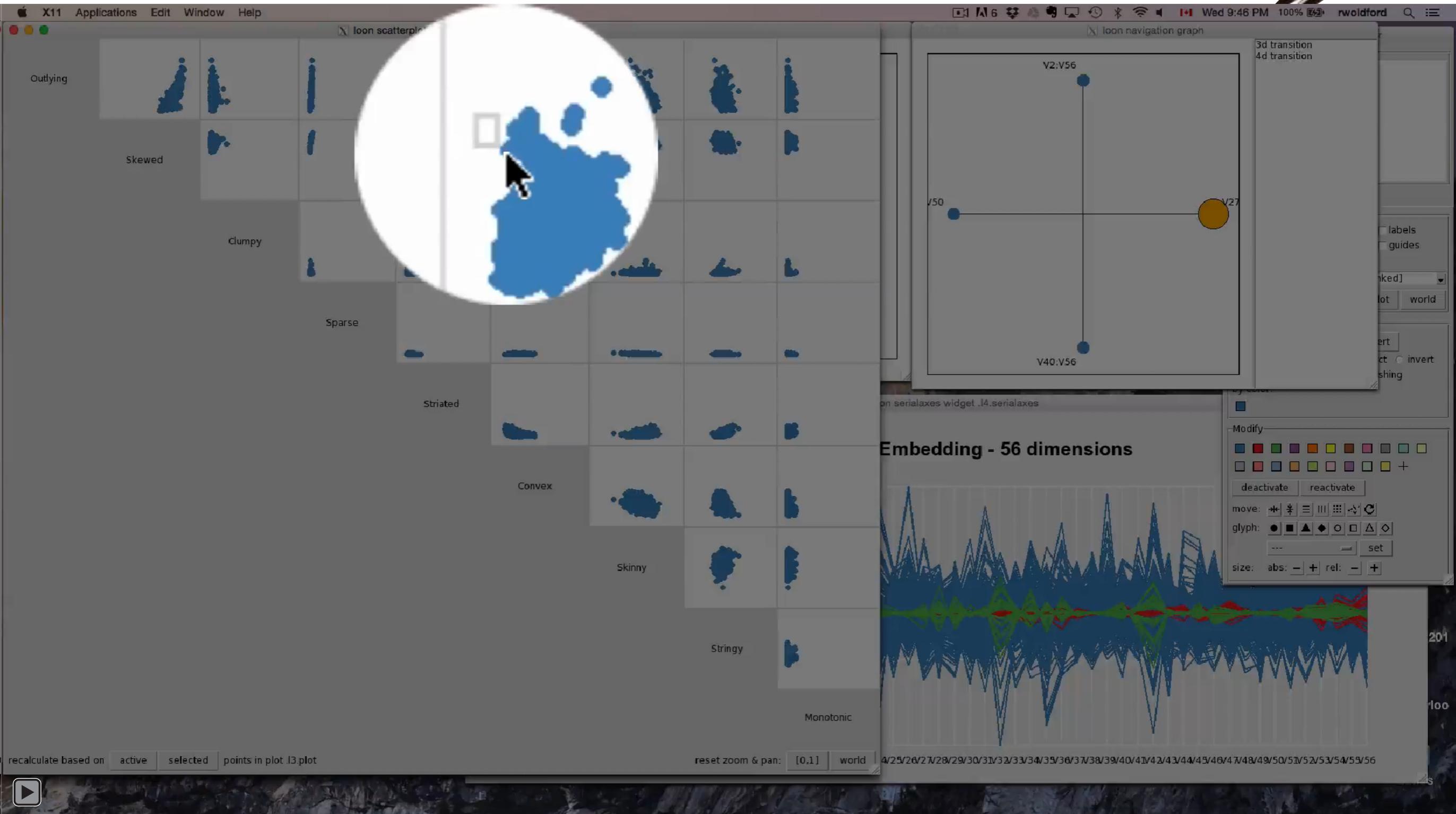


loon serialaxes widget .l4.serialaxes

Local Linear Embedding 56 dimensions

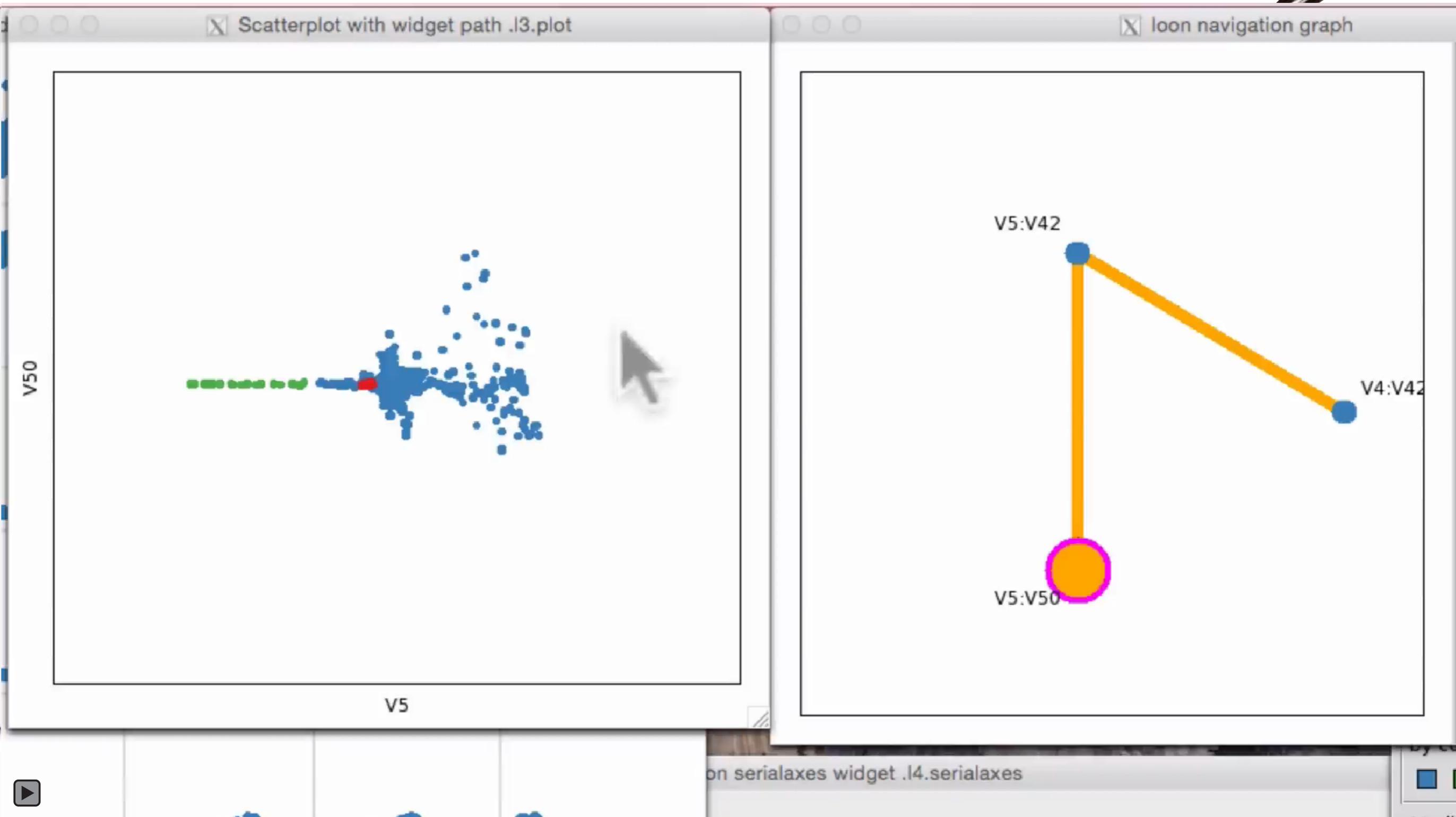
every point in the scatterplot matrix is a scatterplot!

loon



high outlying and low convexity

loon



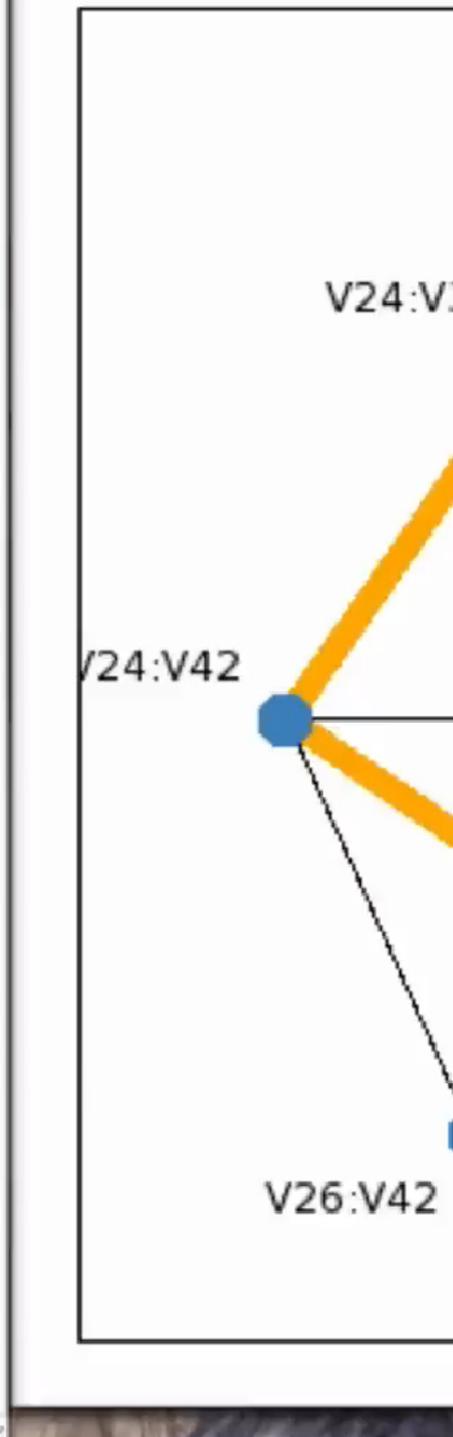
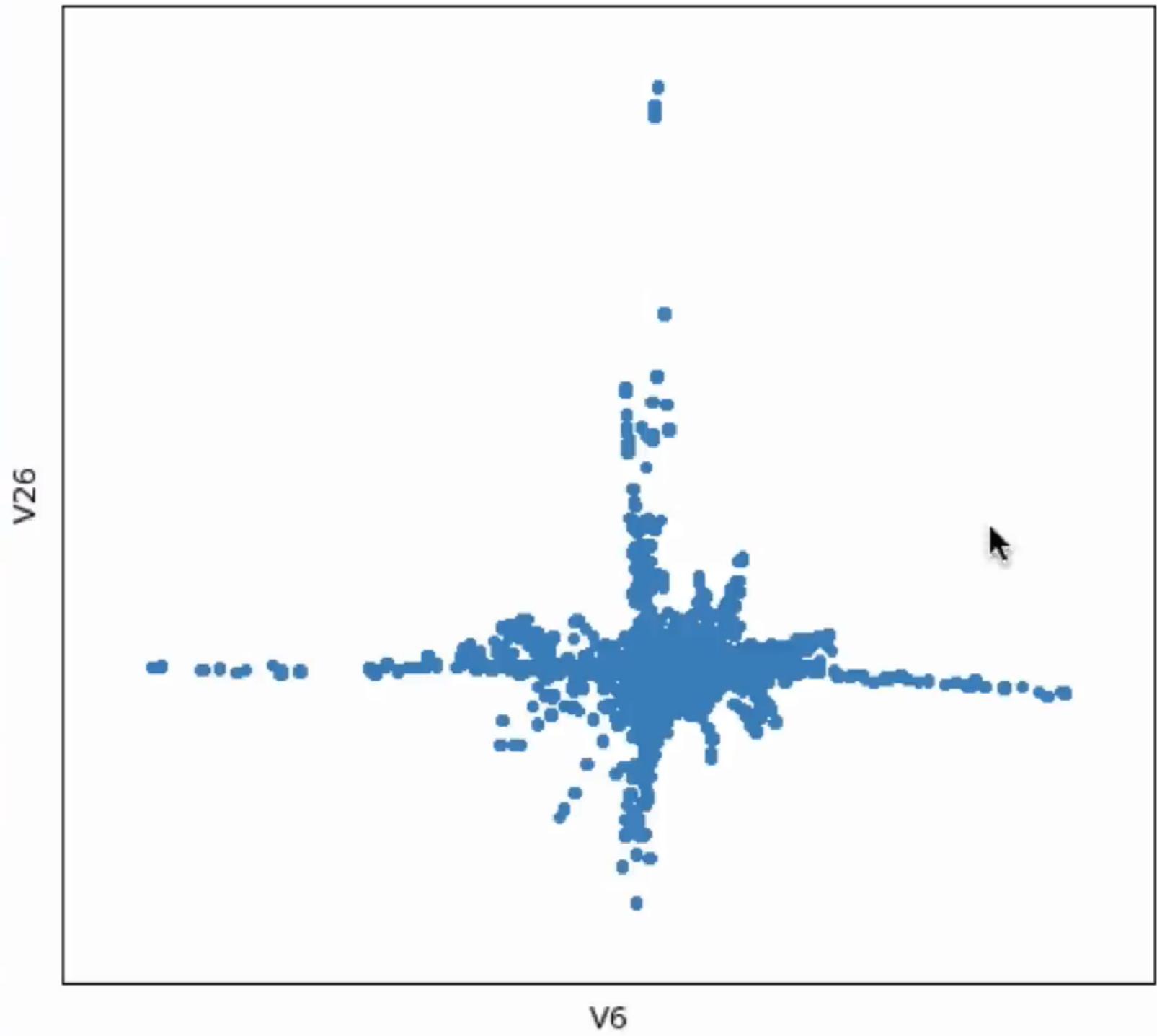
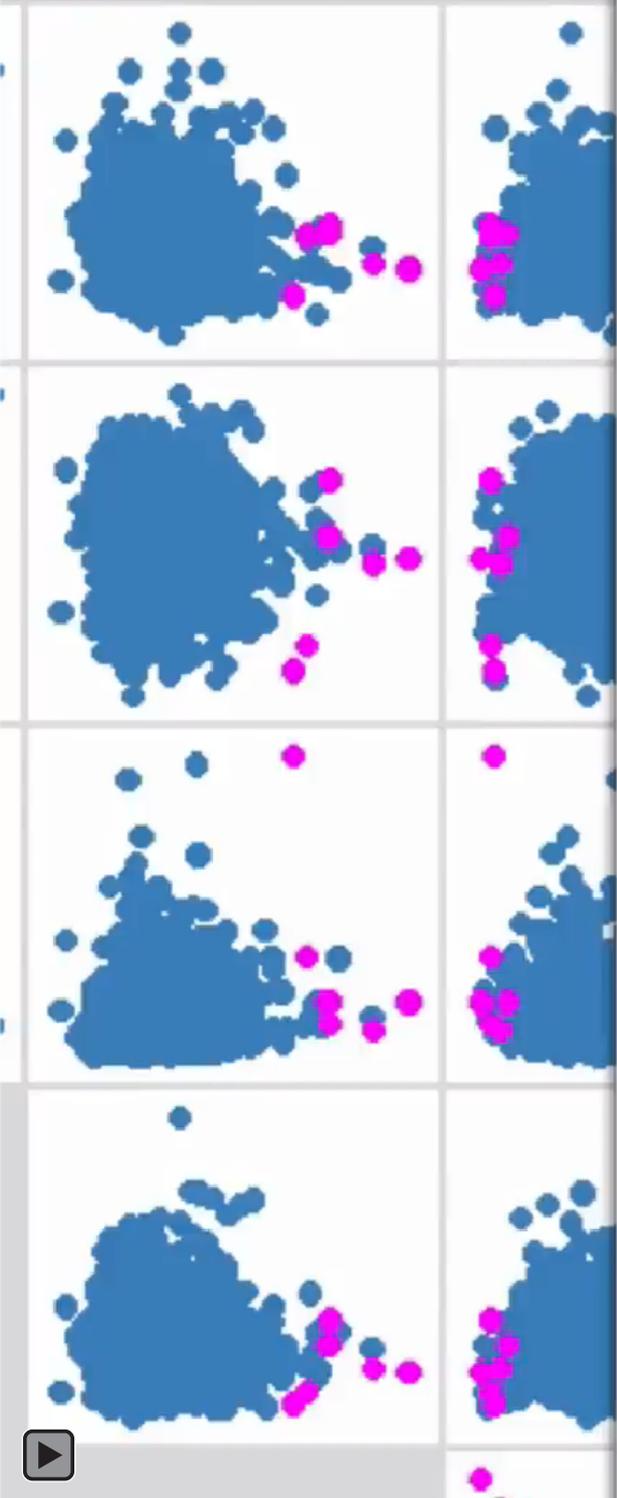
focus on structure of a subset

loon



Scatterplot matrix for measures of

Scatterplot with widget path .13.plot



focus on structure of a subset

Geographical ethnomusicology

1,059 tracks covering 33 countries/areas

traditional, ethnic, or 'world' music only

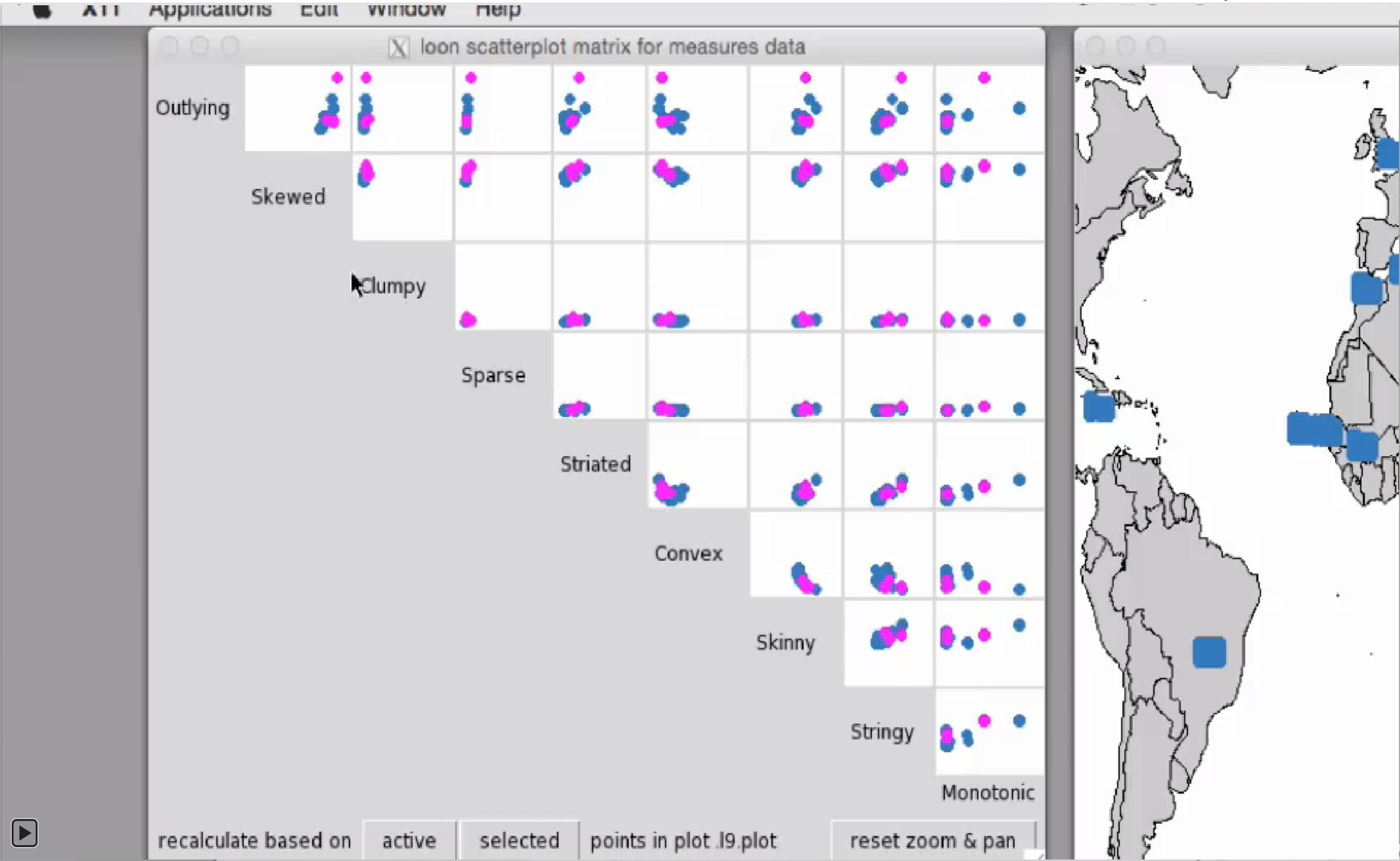
no 'western' music ... too global

geographic location (lat and long)

68 audio descriptors from each track
(using MARSYAS default settings)

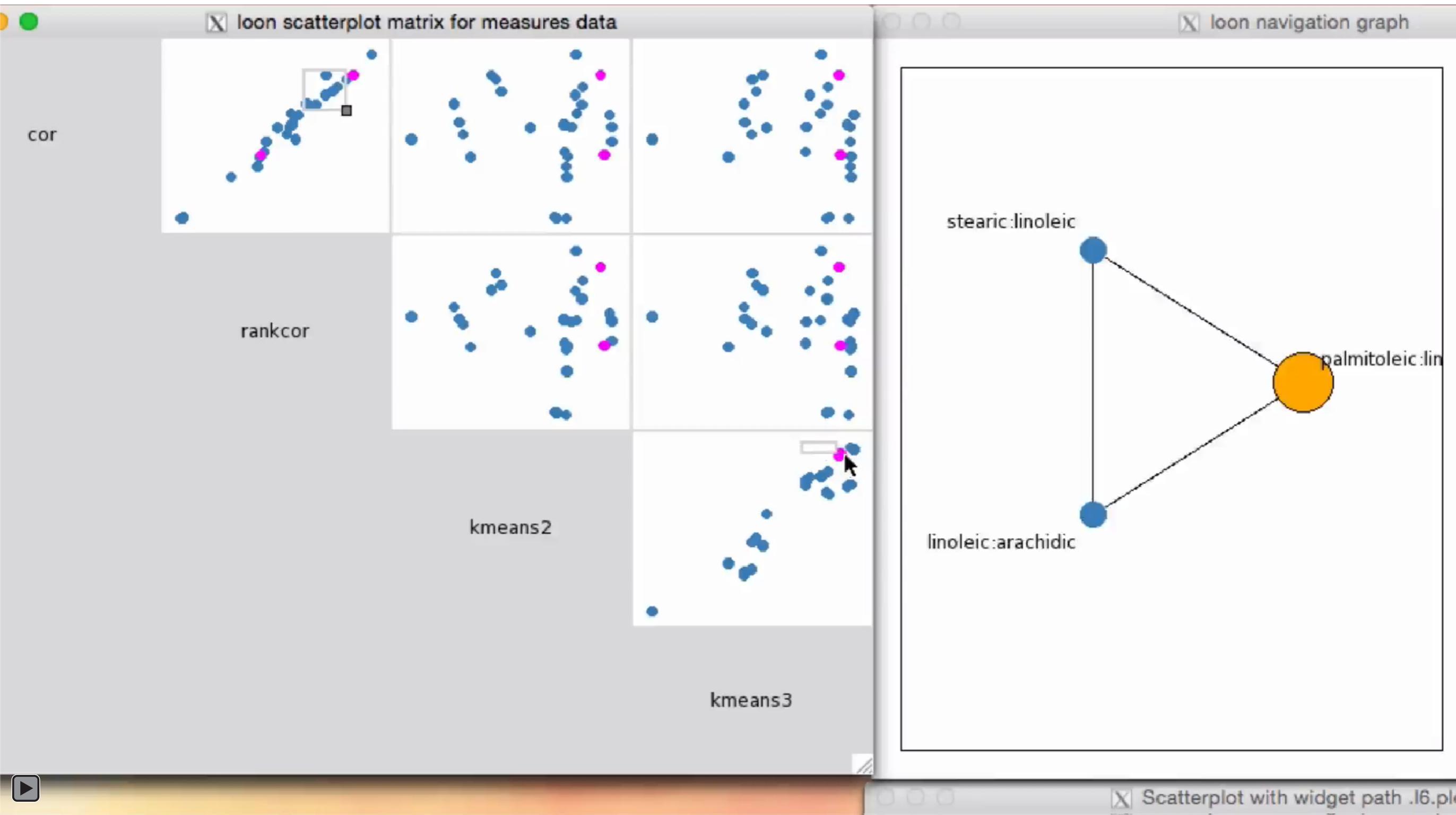
loon

scagnostics, RDRTtoolbox



Build your own

User defined 2d-measures



Build your own

User defined 1d-measures

The screenshot displays a software interface with several windows and panels:

- Top Panel:** A row of five histograms labeled "range", "iqr", "skew", "tailweight", and "gdp". Each histogram has "Frequency" on the y-axis. A mouse cursor is hovering over the "gdp" histogram.
- Loon Inspector:** A window titled "Loon Inspector" with a "World View" section showing a single histogram.
- loon navigation graph:** A window showing a network graph with three nodes: "palmitic:linoleic" (top left, blue), "oleic:linoleic" (bottom left, blue), and "palm" (right, orange). Lines connect "palmitic:linoleic" to "oleic:linoleic", "palmitic:linoleic" to "palm", and "oleic:linoleic" to "palm".
- Left Panel:** A scatter plot titled "with widget path .I7.plot" showing a cluster of blue dots on a grid.
- Right Panel:** A control panel with various options:
 - Buttons: "3d transition", "4d transition", "invert", "deselect", "brushing", "private".
 - Checkboxes: "scales", "labels", "bin handle".
 - Radio buttons: "density", "spinogram".
 - Dropdown menu: "linked]".
 - Color palette: A row of colored squares (brown, pink, grey, teal, yellow, light blue) and a plus sign.

Summary

exploratory ... data diving

- interactively slice and dice
 - ✦ linkable interactive displays and inspector
 - ✦ deactivate/reactivate cases — focus on subsets
 - ✦ zoom, pan, ...
- integrate with statistical methods
 - ✦ apply to selected subsets and update displays
- mix direct manipulation and scripting in analysis
 - ✦ use R functionality and methods
 - ✦ Loon provides event hooks to call back to any R function

request a beta version:

navgraph.com/loon

exploratory ... data diving

