

Data visualization via pairwise displays

Catherine Hurley NUI Maynooth

joint work with Wayne Oldford (Waterloo)

catherine.hurley@nuim.ie

April 15 2008



NUI MAYNOOTH

Ollscoil na hÉireann M^á Nuad

NUI Maynooth



- ▶ 1795: Founded. St Patrick's College Maynooth: became largest Catholic seminary in the world!
- ▶ 1910: Recognised college of NUI
- ▶ 1965: Accepted lay students
- ▶ 1997: NUI Maynooth became separate University
- ▶ Smallest of 7 Univs in Ireland. 5500 students

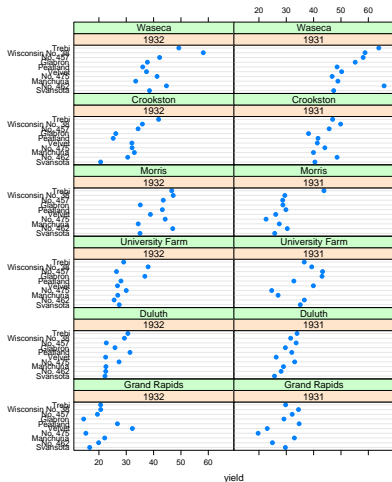


NUI MAYNOOTH

Ollscoil na hÉireann M^a Nuad

Component ordering problem

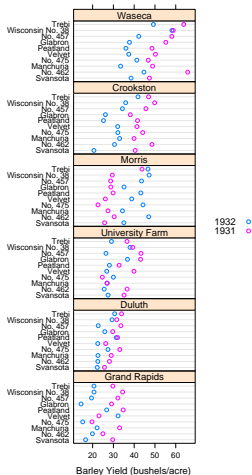
Barley data



- ▶ Trellis display (Cleveland)
- ▶ Order variety, site and year by median
- ▶ Anomaly in Morris data immediately evident

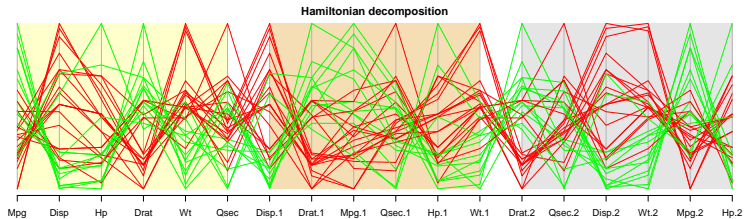
Component ordering problem

Barley data



- ▶ Trellis display (Cleveland)
- ▶ Order variety, site and year by median
- ▶ Anomaly in Morris data immediately evident

Parallel coordinates



Shows all pairs of variables adjacently.

Tufte's principles of graphical excellence

Displays should

- ▶ reveal the data
- ▶ encourage data comparisons
- ▶ and make large datasets coherent

Barley display: **encourages data comparisons** by placing similar sites and varieties adjacently

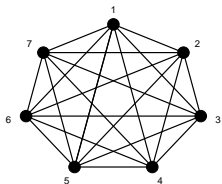
Cars data: **reveal** by showing all pairwise relationships

Outline

- ▶ Ordering: a graph theory approach
- ▶ Examples
- ▶ Algorithms

Graphs: vertices, edges and weights

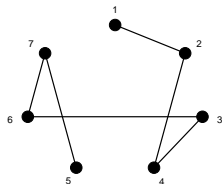
- ▶ n variables, cases, factor levels, boxplots: identify with nodes of graph
- ▶ All possible pairings are of interest: place an undirected edge between each pair of nodes
- ▶ Graph is complete, K_n



- ▶ Dissimilarity measure: edge weight

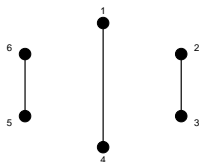
Hamiltonian and Eulerian paths

- ▶ **Hamiltonian path** gives a permutation of vertices
- ▶ 1243675 is a hamiltonian - visits only $n - 1 = 6$ of $\binom{n}{2}$ available edges
- ▶ **Eulerian path** visits all edges
- ▶ eg $T_0 = 1234567461427157352631$ is a closed eulerian **tour** on K_7
- ▶ Which path?



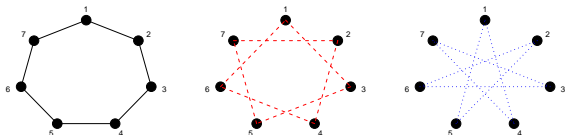
Eulerian tours and paths

- ▶ Eulerian tour exists when every vertex is even. ie for K_{2m+1}
- ▶ About 130 million choices for K_7 , discounting cyclic permutations
- ▶ Do not exist for K_{2m}
- ▶ Eulerian paths (open) exists when two vertices are odd.
- ▶ For K_{2m} form a new augmented graph by adding edges $n/2 - 1$ edges so eulerian path exists (or $n/2$ for a tour).
- ▶ Prefer eulerians where low-weight edges occur early on.



Eulerian tours via Hamiltonian decomposition

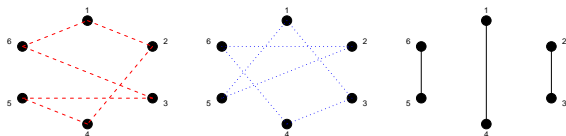
- ▶ **Hamiltonian decomposition**: an Eulerian tour composed of edge-distinct Hamiltonian cycles



- ▶ eg K_7
 $T_1 = 1\ 234567\ 1\ 357246\ 1\ 473625\ 1$
- ▶ Vary node labels, vary cycle order, direction and contact point to produce different tours.
- ▶ Colburn (1982) lists canonical forms for non-isomorphic Hamiltonian decompositions: eg 2 for K_7 , 45,000+ for K_{11} .
- ▶ Not practical to find “best” decomposition.
- ▶ Finding “best” Hamiltonian (TSP) is NP hard.

Eulerian tours via Hamiltonian decomposition

- ▶ No eulerian tour and hence no hamiltonian decomposition exists for K_{2m}
- ▶ eg K_6



Form $T_{open} = 124536 \ 134625 \ 143265$, which is an eulerian path on K_6^e .

Decomposition of complete graphs

K_n can be decomposed as follows:

- ▶ For $n = 2m + 1$, into either
 - ▶ m hamiltonian cycles, or
 - ▶ m hamiltonian paths and an almost-one factor.
- ▶ For $n = 2m$ into either
 - ▶ m hamiltonian paths, or
 - ▶ $m - 1$ hamiltonian cycles and a 1-factor (or perfect matching).

Lucas-Walecki (1892) Alspach(1990)

Applications

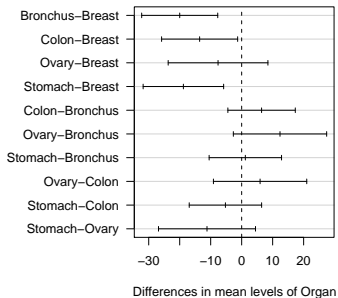
- ▶ Pairwise comparison of treatments
- ▶ Interaction plots
- ▶ Star glyphs of multivariate data
- ▶ Parallel coordinates

Pairwise comparison of treatments

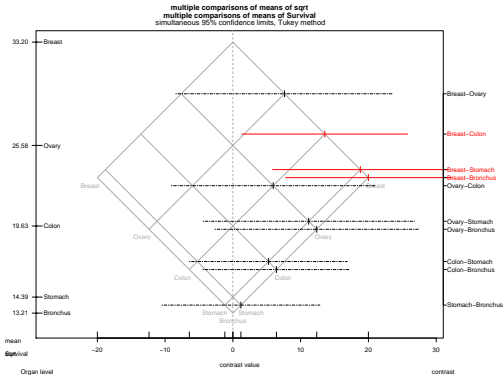
Study on sqrt survival time of vit-C treated cancer patients
(Cameron and Pauling 1978)

Which pairs are significantly different?

95% family-wise confidence level



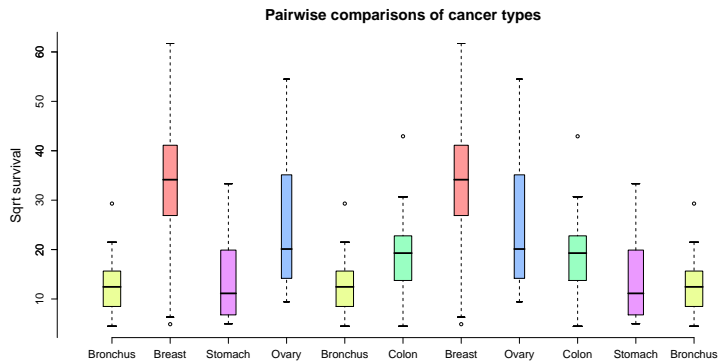
Improved version



- ▶ Shows treatment means
- ▶ Significant differences coloured red.
- ▶ CIs located vertically at treatment average- danger of overlapping.
- ▶ Background square-chartchunk?
- ▶ Visually complicated

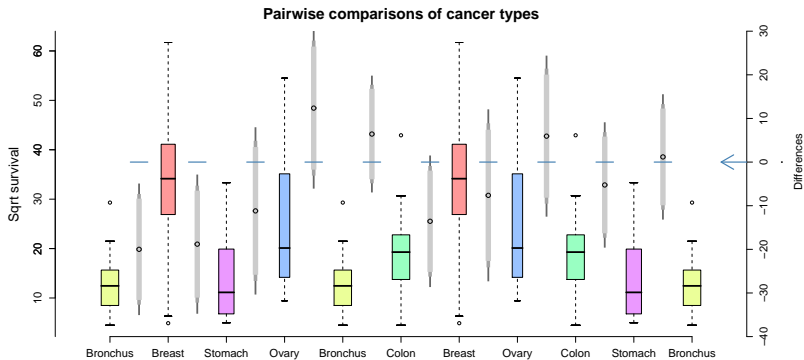
Hsu, Perigga (1994), Heiberger and Holland (2006)

New pairwise comparison display



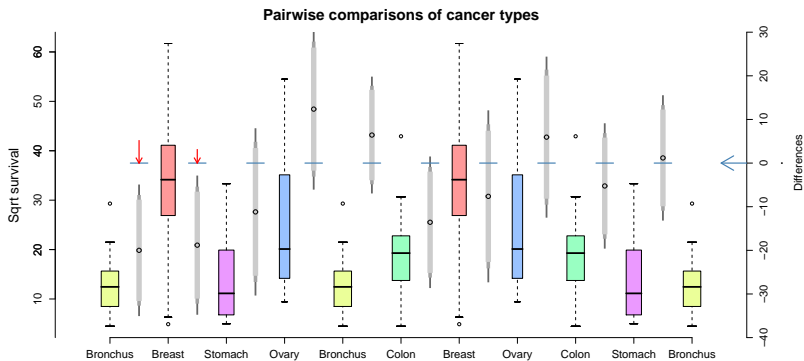
- Show all boxplot pairs adjacently via an eulerian on K_5

New pairwise comparison display



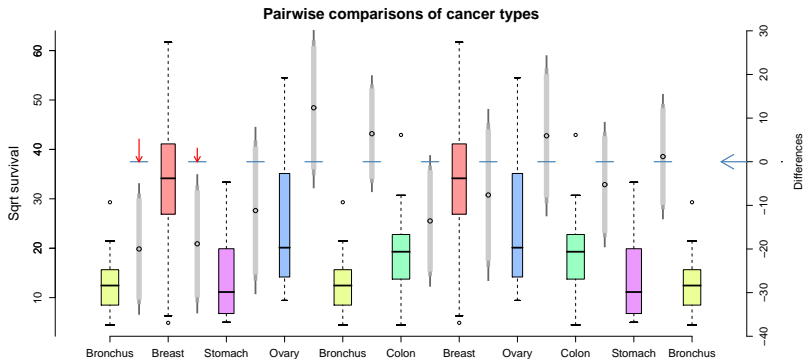
- ▶ Show all boxplot pairs adjacently via an eulerian on K_5
- ▶ Overlay CIs

New pairwise comparison display



- ▶ Show all boxplot pairs adjacently via an eulerian on K_5
- ▶ Overlay CIs
- ▶ Red arrow: significantly different comparisons

New pairwise comparison display



- ▶ Eulerian designed so significant pairs appear on lhs
- ▶ Simple yet informative

Star glyph displays

Construction

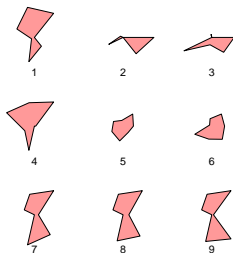
- ▶ One glyph per case
- ▶ Assign variables to equi-space radii
- ▶ Plot scaled observation on each radius
- ▶ Drawing options (`stars` in R)
 - ▶ draw rays
 - ▶ draw polygon
 - ▶ draw filled polygon
- ▶ Here we use filled polygons, no rays, for better overall comparison and clustering of cases.

Cars data

- ▶ Task: visually cluster cases

Default ordering of variables.

Dataset order H0



789 look similar, and to 1?

Other groups: 23, 56

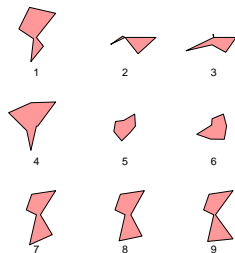
4 on its own

Cars data

- ▶ Task: visually cluster cases

Default ordering of variables.

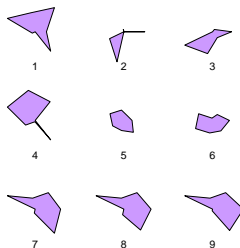
Dataset order H0



789 look similar, and to 1?
Other groups: 23, 56
4 on its own

Another hamiltonian

Order H3



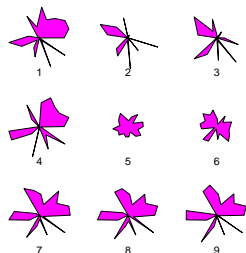
14 look similar
23 look different

Conclusions are order dependent!

Cars data

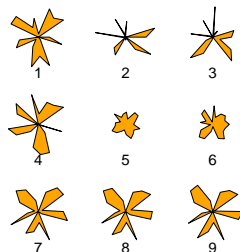
Eulerian order

Eulerian order



Another hamiltonian

Hamiltonian decomp, H1:H2:H3



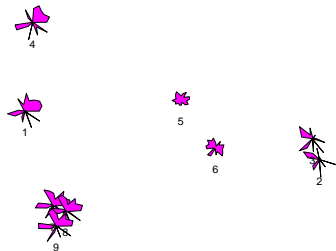
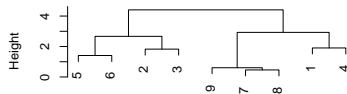
Groups: 789,23,56,14

Less shape variation between orderings.
Conclusions are less order dependent!

Cars data

Verify visual clusterings

Groups: 789,23,56,14



Glyph PC scatterplot

Algorithms: Euler path on K_n^e

- ▶ Complete Graph whose nodes are variables/factor levels etc
- ▶ Path traversing all edges exactly once is an Euler trail
- ▶ Exists when n is odd
- ▶ When n is even $n/2 - 1$ edges must be visited twice
- ▶ In our examples, we used
 1. General euler tour algorithm-modified for weights
 2. Hamiltonian decomposition algorithm-modified for weights

Euler tour algorithm

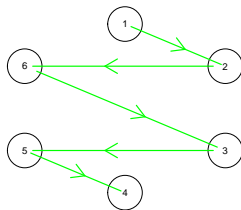
- ▶ Hierholzer (1873) (similar to Fleury 1883)
- ▶ Greedy modification prefers low-weight edges

Require: A connected graph G that is even or that has exactly two odd vertices.

- 1: Choose a starting vertex v from one of the odd vertices connected by the *lowest weight edge*, using the next *lowest weight edge* in their vertex sets to decide between them.
- 2: Starting at v construct a path T in G , always moving to the *lowest weight unused edge*, stopping when a vertex is reached without an unused edge.
- 3: **while** there are edges of G not already in path T **do**
- 4: Choose the *last* vertex w in T that is incident on an unused edge.
- 5: Starting at w , construct a path D of unused edges, *always moving to the lowest weight unused edge* and stopping when a node is reached without any unused edges.
- 6: Enlarge T by splicing path D into T at vertex w .
- 7: **end while**
- 8: **return** T

Hamiltonian decomposition algorithm

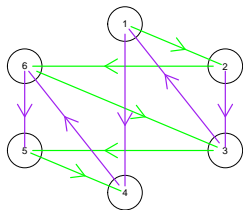
- ▶ When n is even $n/2 - 1$ edges must be visited twice
- ▶ Lucas-Walecki construction (1892)
- ▶ Construction: n even



1 2 6 3 5 4

Hamiltonian decomposition algorithm

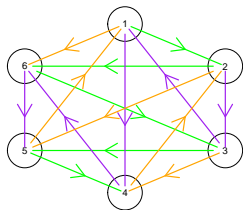
- ▶ When n is even $n/2 - 1$ edges must be visited twice
- ▶ Lucas-Walecki construction (1892)
- ▶ Construction: n even



1 2 6 3 5 4
2 3 1 4 6 5

Hamiltonian decomposition algorithm

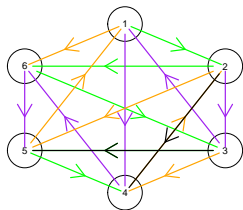
- ▶ When n is even $n/2 - 1$ edges must be visited twice
- ▶ Lucas-Walecki construction (1892)
- ▶ Construction: n even



1 2 6 3 5 4
2 3 1 4 6 5
3 4 2 5 1 6

Hamiltonian decomposition algorithm

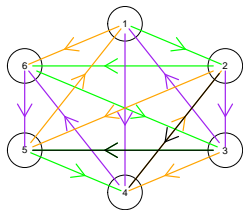
- ▶ When n is even $n/2 - 1$ edges must be visited twice
- ▶ Lucas-Walecki construction (1892)
- ▶ Construction: n even



1 2 6 3 5 4
2 3 1 4 6 5
3 4 2 5 1 6

Hamiltonian decomposition algorithm

- ▶ When n is even $n/2 - 1$ edges must be visited twice
- ▶ Lucas-Walecki construction (1892)
- ▶ Construction: n even



1 2 6 3 5 4
2 3 1 4 6 5
3 4 2 5 1 6

- ▶ Construction: n odd

7 1 2 6 3 5 4
7 2 3 1 4 6 5
7 3 4 2 5 1 6 7

Hamiltonian decomposition- weighted graphs

- ▶ Canonical forms for non-isomorphic hamiltonian decompositions: eg 2 for K_7 , 45,000+ for K_{11} (Colburn)
- ▶ Algorithm produces only one.
- ▶ For T composed of hamiltonian cycles:
vary node labels, vary cycle order, direction and contact point.
- ▶ For T composed of hamiltonian paths:
vary node labels, vary cycle order, direction and contact point.
- ▶ Use TSP algorithm to find the lowest-weight hamiltonian, use this as the first cycle or path.

Hamiltonian decomposition- weighted graphs

- ▶ $n = 7$. TSP gives best ham cycle as 4361257
- ▶ Best start and direction(measured by how much weights increase) is 3475162

Original

7 1 2 6 3 5 4
7 2 3 1 4 6 5
7 3 4 2 5 1 6 7

Relabel

3 4 7 5 1 6 2
3 7 1 4 2 5 6
3 1 2 7 6 4 5

Switch rows, allow
reversals

3 4 7 5 1 6 2
3 5 4 6 7 2 1
3 7 1 4 2 5 6

Concluding remarks

- ▶ Other applications: parallel coordinate displays, profile displays, Andrew's curves
- ▶ Problems: large p - use scagnostics to select “interesting” paths
- ▶ Wegman(1990) - LW hamiltonian path algorithm in parallel coordinate displays
- ▶ Bailey et al (2003)- Hamiltonian cycles, in DOE
- ▶ Software EulerViz R-package
- ▶ Uses TSP(Hahsler et al), scagnostics (Hofman et al)
- ▶ **Another hamiltonian: the Hamilton walk—>**

Annual NUIM event- all welcome!

HAMILTON WALK



Hamilton Walk

An annual walk on 16th October, formerly known as the Quaternion Walk. Please book a place with Dr. Fiacre Ó Cairbre (+353 1 708 3763).

2007: The bus from Maynooth to Dunsink will leave from under the flyover (connecting the old campus to the new campus) at 1.20p.m. sharp on October 16. The bus is only one-way and is free. There is a train station at Broombridge for returning to Maynooth.

Recently it has been graced by famous guests: the Nobel Prize winner Murray Gell-Mann (2002) and Andrew Wiles of Fermat's Last Theorem fame (2003).

Sir William Rowan Hamilton
(1805 - 1865)

In 1843, on a bright October day, Sir William Rowan Hamilton, the greatest Irish Mathematician of all time, discovered the numbers called Quaternions.

For many years he had been trying to find a satisfactory way to multiply points in three dimensions, in such a way as to allow division. The idea of using four dimensions instead, and the way to do it, came to him in a flash, as he walked with his wife by the Royal Canal. Since it is one of the few major mathematical discoveries which is precisely located in time and circumstances, the event is very well-known in the international mathematical community, and people from all over the world know about "Hamilton's Bridge". When he made the discovery, Hamilton paused under Brougham (or Broom) Bridge, took out his penknife, and scratched the fundamental formula:

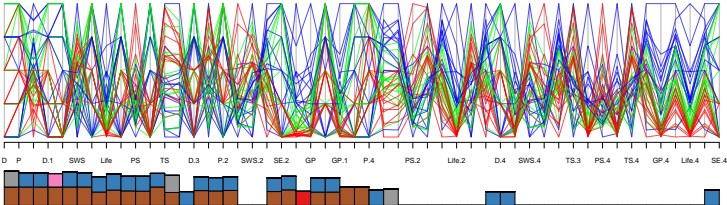
$$i^2 = j^2 = k^2 = ijk = -1$$

into the stone. The day was the 16th of October, 1843. No trace of this can be found today, but in 1958 a plaque was erected on the site, commemorating the discovery and displaying the formula.

Hamilton lived and worked at Dunsink Observatory, and was on his way into Dublin to attend a meeting at the Royal Irish Academy.

More on PCA- sleep data

Eulerian on all scagnostics.



First 18 panels of Eulerian on all scagnostics.

