Parallel Coordinate Plots
for Fun and Profit

Kevin Pulo
kevin.pulo@anu.edu.au

Australian National University Supercomputing Facility (ANUSF),
National Computational Infrastructure National Facility (NCI NF),
Canberra, Australia

2010-12-13
Introduction
Mapping data to space

- Graphs map variables *directly* to 2D/3D space
Mapping data to space

- Graphs map variables *directly* to 2D/3D space

2 variables:

\[ x, y \]
Mapping data to space

- Graphs map variables *directly* to 2D/3D space

2 variables:

\[ x, y \]
Mapping data to space

- Graphs map variables \textit{directly} to 2D/3D space
Mapping data to space

- Graphs map variables *directly* to 2D/3D space
Mapping data to space

- Graphs map variables *directly* to 2D/3D space
Mapping data to space

- Graphs map variables *directly* to 2D/3D space
But what if we have hundreds of independent variables?
Visualising high dimensional data

- But what if we have hundreds of independent variables?
- Cannot possibly map this directly to hundred-dimensional space
Visualising high dimensional data

- But what if we have hundreds of independent variables?
- Cannot possibly map this directly to hundred-dimensional space
- Need some way of mapping this high-dimensional data to low-dimensional space
• But what if we have hundreds of independent variables?
• Cannot possibly map this directly to hundred-dimensional space
• Need some way of mapping this high-dimensional data to low-dimensional space

Parallel Coordinate Plots
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Parallel Coordinate Plots

- Dimensions (axes) are laid out in parallel
- Use lines to join variable values
Advantages

- Allows many dimensions
- Straightforward mapping
- Human perceptual system is good at seeing connectedness and patterns
Challenge: Ambiguity

Common values cause ambiguity
Challenge: Ambiguity

Common values cause ambiguity
Solution: (a) Curves
Challenge: Ambiguity

Common values cause ambiguity
Solution: (a) Curves
Challenge: Ambiguity

Common values cause ambiguity
Solution: (a) Curves      (b) Spreading
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation

Solution: Interactive manipulation
Social Science
Social Science and Visualisation

- Social science tends to be “undervisualised”
  - Analytic/statistical techniques
  - Simple graphs
- Despite large and rich datasets
- Excellent scope for using sophisticated and advanced visualisation techniques to better exploit the data
<table>
<thead>
<tr>
<th>$\bar{x}_1$</th>
<th>9.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Var}(x_1)$</td>
<td>10.0</td>
</tr>
<tr>
<td>$\bar{y}_1$</td>
<td>7.5</td>
</tr>
<tr>
<td>$\text{Var}(y_1)$</td>
<td>3.75</td>
</tr>
<tr>
<td>$\text{Corr}(x_1, y_1)$</td>
<td>0.816</td>
</tr>
<tr>
<td>Regression $y_1$</td>
<td>$3 + 0.5x_1$</td>
</tr>
</tbody>
</table>
### Why Visualise? Ask Anscombe

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$y_1$</th>
<th>$x_2$</th>
<th>$y_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}_1 = 9.0$</td>
<td>$\bar{y}_1 = 7.5$</td>
<td>$\bar{x}_2 = 9.0$</td>
<td>$\bar{y}_2 = 7.5$</td>
</tr>
<tr>
<td>$\text{Var}(x_1) = 10.0$</td>
<td>$\text{Var}(y_1) = 3.75$</td>
<td>$\text{Var}(x_2) = 10.0$</td>
<td>$\text{Var}(y_2) = 3.75$</td>
</tr>
<tr>
<td>$\text{Corr}(x_1, y_1) = 0.816$</td>
<td>$\text{Regression } y_1 = 3 + 0.5x_1$</td>
<td>$\text{Corr}(x_2, y_2) = 0.816$</td>
<td>$\text{Regression } y_2 = 3 + 0.5x_2$</td>
</tr>
</tbody>
</table>
### Why Visualise? Ask Anscombe

<table>
<thead>
<tr>
<th></th>
<th>( x_1 )</th>
<th>( y_1 )</th>
<th>( x_2 )</th>
<th>( y_2 )</th>
<th>( x_3 )</th>
<th>( y_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>9.0</td>
<td>7.5</td>
<td>9.0</td>
<td>7.5</td>
<td>9.0</td>
<td>7.5</td>
</tr>
<tr>
<td>Var</td>
<td>10.0</td>
<td>3.75</td>
<td>10.0</td>
<td>3.75</td>
<td>10.0</td>
<td>3.75</td>
</tr>
<tr>
<td>Corr</td>
<td>0.816</td>
<td></td>
<td>0.816</td>
<td></td>
<td>0.816</td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td>( y_1 = 3 + 0.5x_1 )</td>
<td>( y_2 = 3 + 0.5x_2 )</td>
<td>( y_3 = 3 + 0.5x_3 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Why Visualise? Ask Anscombe

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}_1 = 9.0$</td>
<td>$\bar{x}_2 = 9.0$</td>
<td>$\bar{x}_3 = 9.0$</td>
<td>$\bar{x}_4 = 9.0$</td>
</tr>
<tr>
<td>$\text{Var}(x_1) = 10.0$</td>
<td>$\text{Var}(x_2) = 10.0$</td>
<td>$\text{Var}(x_3) = 10.0$</td>
<td>$\text{Var}(x_4) = 10.0$</td>
</tr>
<tr>
<td>$\bar{y}_1 = 7.5$</td>
<td>$\bar{y}_2 = 7.5$</td>
<td>$\bar{y}_3 = 7.5$</td>
<td>$\bar{y}_4 = 7.5$</td>
</tr>
<tr>
<td>$\text{Var}(y_1) = 3.75$</td>
<td>$\text{Var}(y_2) = 3.75$</td>
<td>$\text{Var}(y_3) = 3.75$</td>
<td>$\text{Var}(y_4) = 3.75$</td>
</tr>
<tr>
<td>$\text{Corr}(x_1, y_1) = 0.816$</td>
<td>$\text{Corr}(x_2, y_2) = 0.816$</td>
<td>$\text{Corr}(x_3, y_3) = 0.816$</td>
<td>$\text{Corr}(x_4, y_4) = 0.816$</td>
</tr>
<tr>
<td>Regression $y_1 = 3 + 0.5x_1$</td>
<td>Regression $y_2 = 3 + 0.5x_2$</td>
<td>Regression $y_3 = 3 + 0.5x_3$</td>
<td>Regression $y_4 = 3 + 0.5x_4$</td>
</tr>
</tbody>
</table>
Why Visualise? Ask Anscombe

\begin{align*}
\bar{x}_2 &= 9.0 \\
\text{Var}(x_2) &= 10.0 \\
\bar{y}_2 &= 7.5 \\
\text{Var}(y_2) &= 3.75 \\
\text{Corr}(x_2, y_2) &= 0.816 \\
\text{Regression: } y_2 &= 3 + 0.5x_2 \\
\bar{x}_3 &= 9.0 \\
\text{Var}(x_3) &= 10.0 \\
\bar{y}_3 &= 7.5 \\
\text{Var}(y_3) &= 3.75 \\
\text{Corr}(x_3, y_3) &= 0.816 \\
\text{Regression: } y_3 &= 3 + 0.5x_3 \\
\bar{x}_4 &= 9.0 \\
\text{Var}(x_4) &= 10.0 \\
\bar{y}_4 &= 7.5 \\
\text{Var}(y_4) &= 3.75 \\
\text{Corr}(x_4, y_4) &= 0.816 \\
\text{Regression: } y_4 &= 3 + 0.5x_4
\end{align*}
Why Visualise? Ask Anscombe

\[ x_3 = 9.0 \]
\[ \text{Var}(x_3) = 10.0 \]
\[ y_3 = 7.5 \]
\[ \text{Var}(y_3) = 3.75 \]
\[ \text{Corr}(x_3, y_3) = 0.816 \]
\[ \text{Regression } y_3 = 3 + 0.5x_3 \]

\[ x_4 = 9.0 \]
\[ \text{Var}(x_4) = 10.0 \]
\[ y_4 = 7.5 \]
\[ \text{Var}(y_4) = 3.75 \]
\[ \text{Corr}(x_4, y_4) = 0.816 \]
\[ \text{Regression } y_4 = 3 + 0.5x_4 \]
Why Visualise? Ask Anscombe

\[
\begin{align*}
\bar{x}_4 &= 9.0 \\
\text{Var}(x_4) &= 10.0 \\
\bar{y}_4 &= 7.5 \\
\text{Var}(y_4) &= 3.75 \\
\text{Corr}(x_4, y_4) &= 0.816 \\
\text{Regression } y_4 &= 3 + 0.5x_4
\end{align*}
\]
**Longitudinal/panel survey data**

- Follows the same set of individuals over time
- Eg. employment, study and relationship status over the past 10 years for a group of people (“respondents”)

Longitudinal/panel survey data

- Follows the same set of individuals over time
- Eg. employment, study and relationship status over the past 10 years for a group of people (“respondents”)

<table>
<thead>
<tr>
<th>person</th>
<th>year</th>
<th>employment</th>
<th>study</th>
<th>relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2008</td>
<td>none</td>
<td>full-time</td>
<td>single</td>
</tr>
<tr>
<td>1</td>
<td>2009</td>
<td>part-time</td>
<td>part-time</td>
<td>cohabitating</td>
</tr>
<tr>
<td>1</td>
<td>2010</td>
<td>full-time</td>
<td>none</td>
<td>married</td>
</tr>
<tr>
<td>2</td>
<td>2008</td>
<td>full-time</td>
<td>part-time</td>
<td>cohabitating</td>
</tr>
<tr>
<td>2</td>
<td>2009</td>
<td>full-time</td>
<td>none</td>
<td>cohabitating</td>
</tr>
<tr>
<td>2</td>
<td>2010</td>
<td>full-time</td>
<td>none</td>
<td>single</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Longitudinal/panel survey data

- Follows the same set of individuals over time
- Eg. employment, study and relationship status over the past 10 years for a group of people (“respondents”)

<table>
<thead>
<tr>
<th>year</th>
<th>person</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>n,f,s</td>
<td>f,p,c</td>
<td>...</td>
</tr>
<tr>
<td>2009</td>
<td>p,p,c</td>
<td>f,n,c</td>
<td>...</td>
</tr>
<tr>
<td>2010</td>
<td>f,n,m</td>
<td>f,n,s</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>
Goals

• Data contains:
  - Thousands of respondents, answering
  - Hundreds of questions, over
  - Multiple waves

• Direct visualisation

• Support interactive exploration
Main use cases

1. Initial familiarisation
2. Researchers looking for interesting features
   • integration with Australian Data Archive (ADA) website
3. Archivists performing data cleaning
   • via desktop application
Negotiating the Life Course (NLC)

- Interested in:
  - “... the changing life courses ... as the family and society move from male breadwinner orientation in the direction of higher levels of gender equity.”

- 4 waves, unbalanced
  - Wave 1 (1997): 2231 respondents
  - Wave 2 (2000): 1768 respondents
  - Wave 3 (2003): 1192 respondents

- Noise added to address confidentiality

Pilot software tool

Panimalia

- Based on “parvis” InfoVis research software
- Written in Java
  - Web (applet) usage
  - Desktop (application) usage
- Still under development
  - Interactivity (responsiveness, usability)
  - Web integration
  - Data input/output (over web, native files)
- Work progressing on web-enabled version
Labour force status
Labour force status

![Diagram showing labour force status analysis and visualisation of large and complex data.](image-url)
Work/study: interleaved by year
Work/study: interleaved by year
Work/study: interleaved by year
Work/study: separately by year
Work/study: separately by year
Work/study: combined by year
Work/study: combined by age
Work/study: combined by age
Work/study: combined by age
Work/study: combined by age
Work/study: combined by age
Work/study: combined transitions
Work/study: combined transitions
Relationship path to first birth
Relationship path to first birth
Stockmarket Data
2009 eResearch vis challenge

- Sponsored by Sirca
- NYSE/NASDAQ and Reuters data
- 30 Dow Jones stocks
- From 29 Sept – 3 Oct 2008
  - When GFC became fully apparent
- Promote meaningful visualisations
- $5000 first prize, $500 second/third
Data description

- 1Gb CSV file, sorted by stock/time
- 19,050,304 records (3 types):
  - Trades: 3,634,444 records
  - Quotes: 15,413,586 records
  - News: 2,274 records
Data description

- 1Gb CSV file, sorted by stock/time
- 19,050,304 records (3 types):
  - Trades: 3,634,444 records
  - Quotes: 15,413,586 records
  - News: 2,274 records
- Each record has:
  - Stock name, date/time (milliseconds)
  - Trades: price/volume
  - Quotes: bid/ask, price/volume
  - News: headline (free text)
- Quotes visually indistinguishable from trades
Sample data

AA.N, 29-SEP-2008, 18:33:10.103, -4, Quote, , 21.11, 21.14,
AA.N, 29-SEP-2008, 18:33:10.533, -4, Quote, , 21.09, 21.14,
AA.N, 29-SEP-2008, 18:33:10.556, -4, Trade, 21.11, 100, ,
AA.N, 29-SEP-2008, 18:33:11.081, -4, Trade, 21.11, 200, ,
AA.N, 29-SEP-2008, 18:33:11.125, -4, Quote, , 21.09, 21.14,
US-Wall St routed as House rejects bailout bill"
AA.N, 29-SEP-2008, 18:33:11.995, -4, Quote, , 21.1, 21.12,
AA.N, 29-SEP-2008, 18:33:12.037, -4, Quote, , 21.11, 21.12,
AA.N, 29-SEP-2008, 18:33:12.094, -4, Quote, , 21.1, 21.12,
AA.N, 29-SEP-2008, 18:33:12.155, -4, Quote, , 21.09, 21.12,
AA.N, 29-SEP-2008, 18:33:12.199, -4, Trade, 21.11, 100, ,
AA.N, 29-SEP-2008, 18:33:13.003, -4, Quote, , 21.09, 21.11,
Dow Jones Animated Multiverse

Analysis and Visualisation of Large and Complex Data — Kevin Pulo
• **[The] Judges felt this entry was terrific ...**

• **A really simple rendering of complex datasets, it invites you in to try to change the list of stocks being displayed, and to try to pause and start the time-slide to take a better look at the way things are panning out ...**

• **It provides a great way of replaying a market event ... [and] also has great “real time” promise in the way a trader or analyst would monitor a market segment ...**

• **This is a platform that has potential to move forward commercially and academically.**
Conclusion
Conclusion

1. Use Parallel Coordinate Plots
Conclusion

1. Use Parallel Coordinate Plots

2. ...
Conclusion

1. Use Parallel Coordinate Plots
2. ...
3. Profit!