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

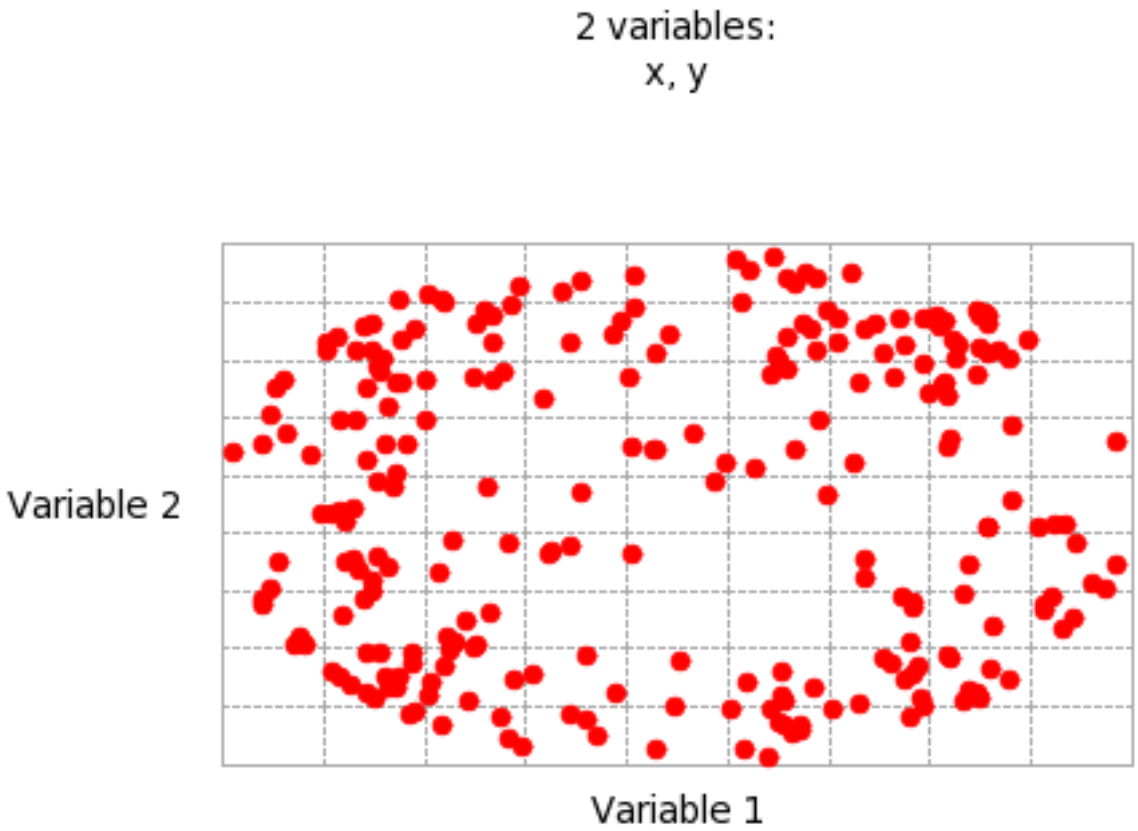
Introduction

Mapping data to space

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

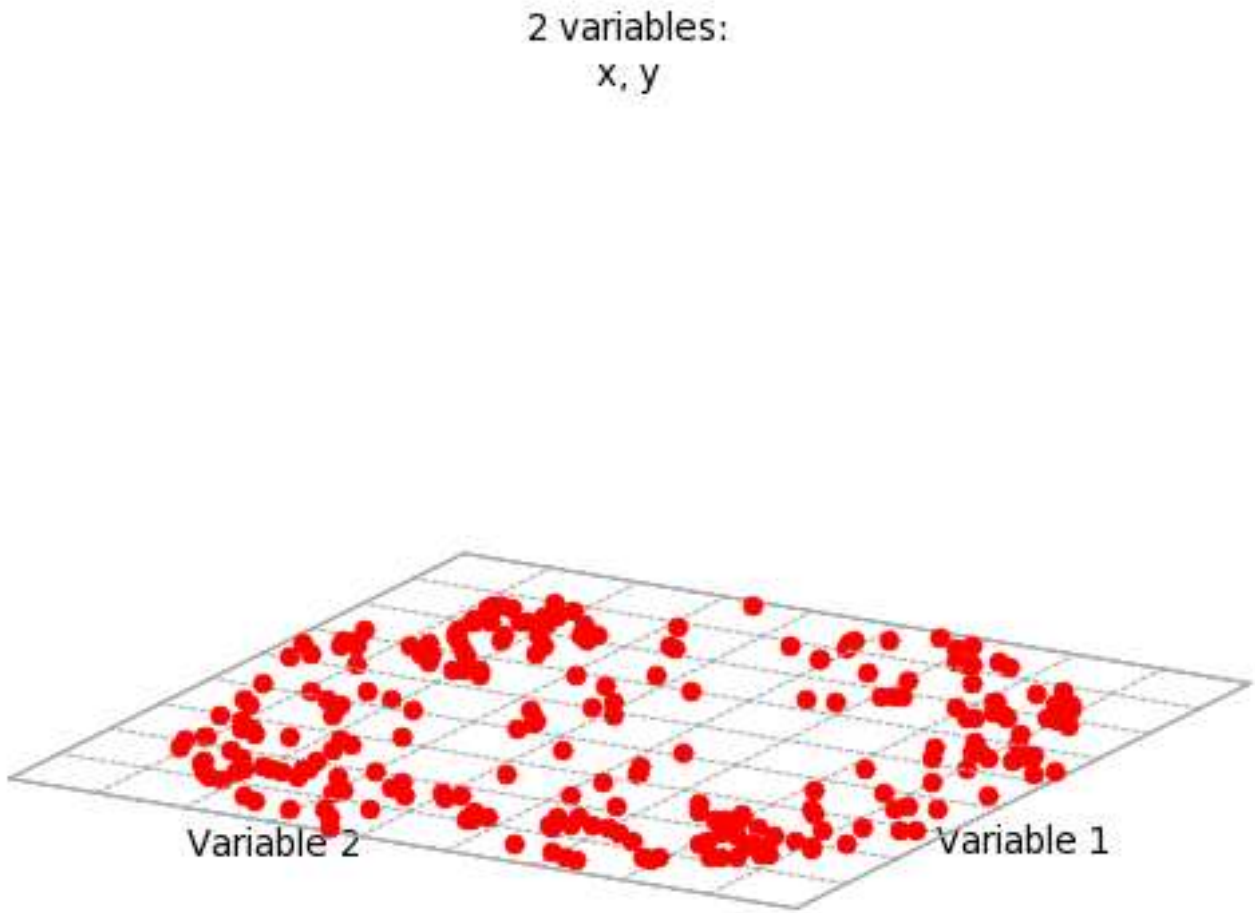
Mapping data to space

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



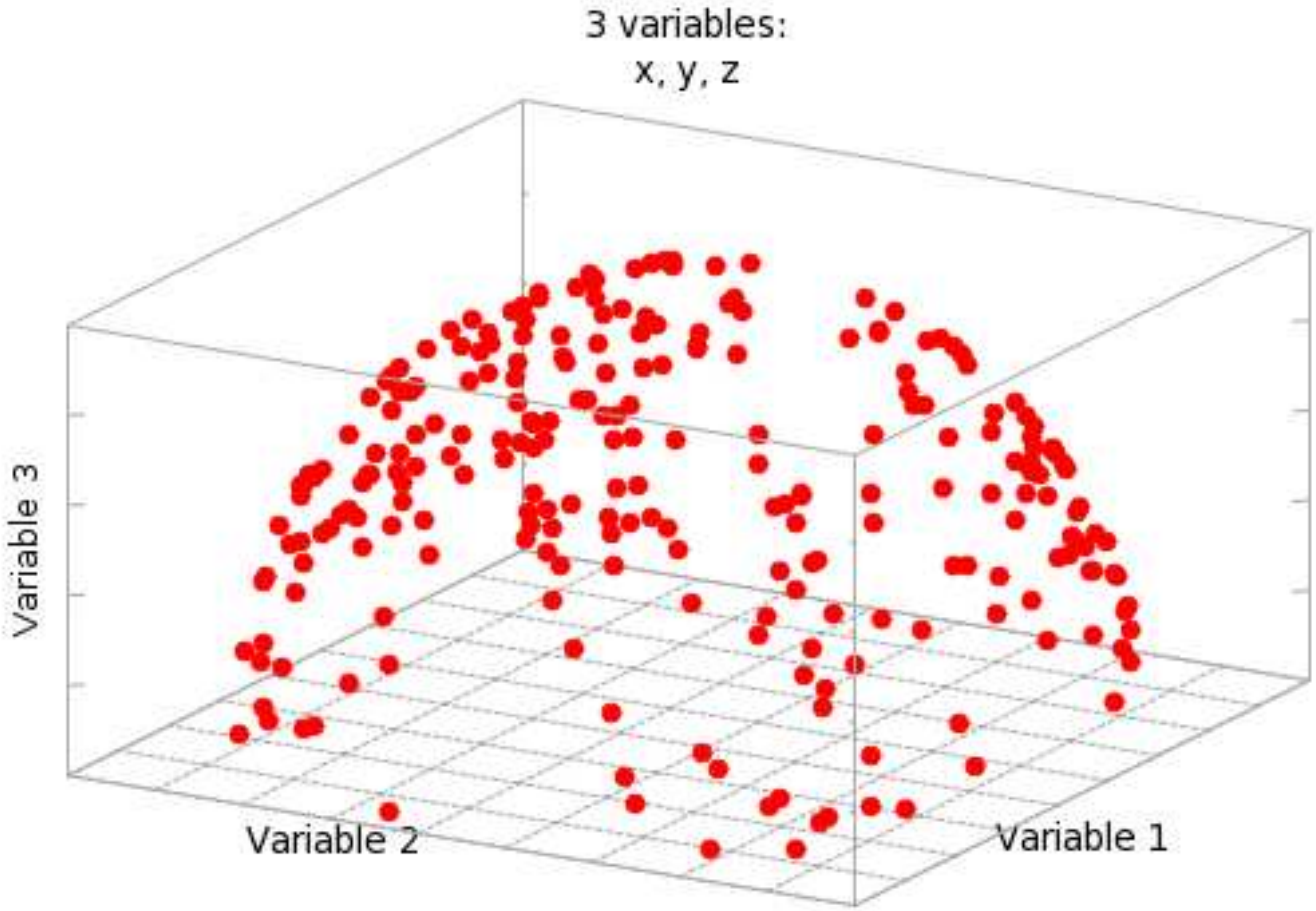
Mapping data to space

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



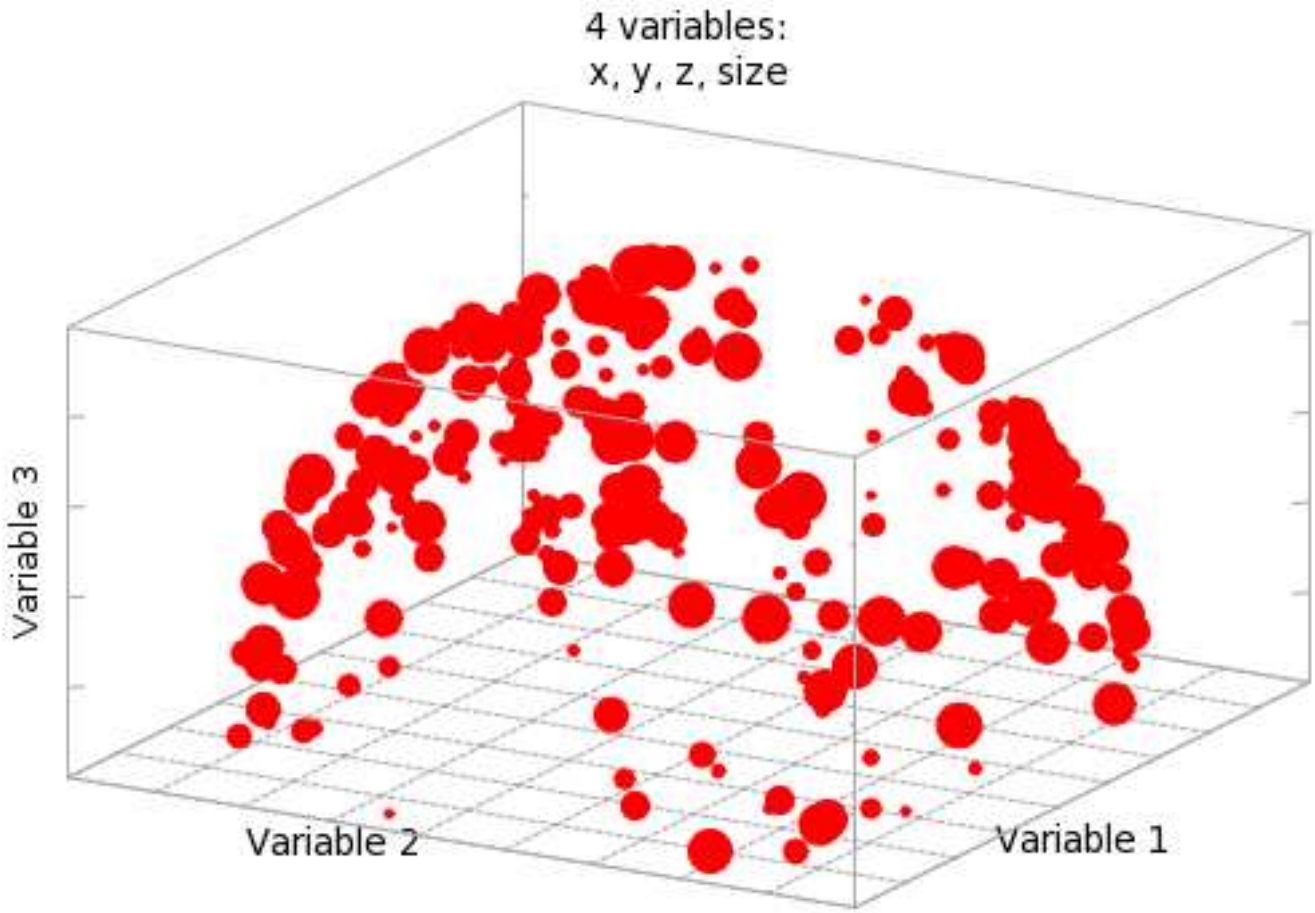
Mapping data to space

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



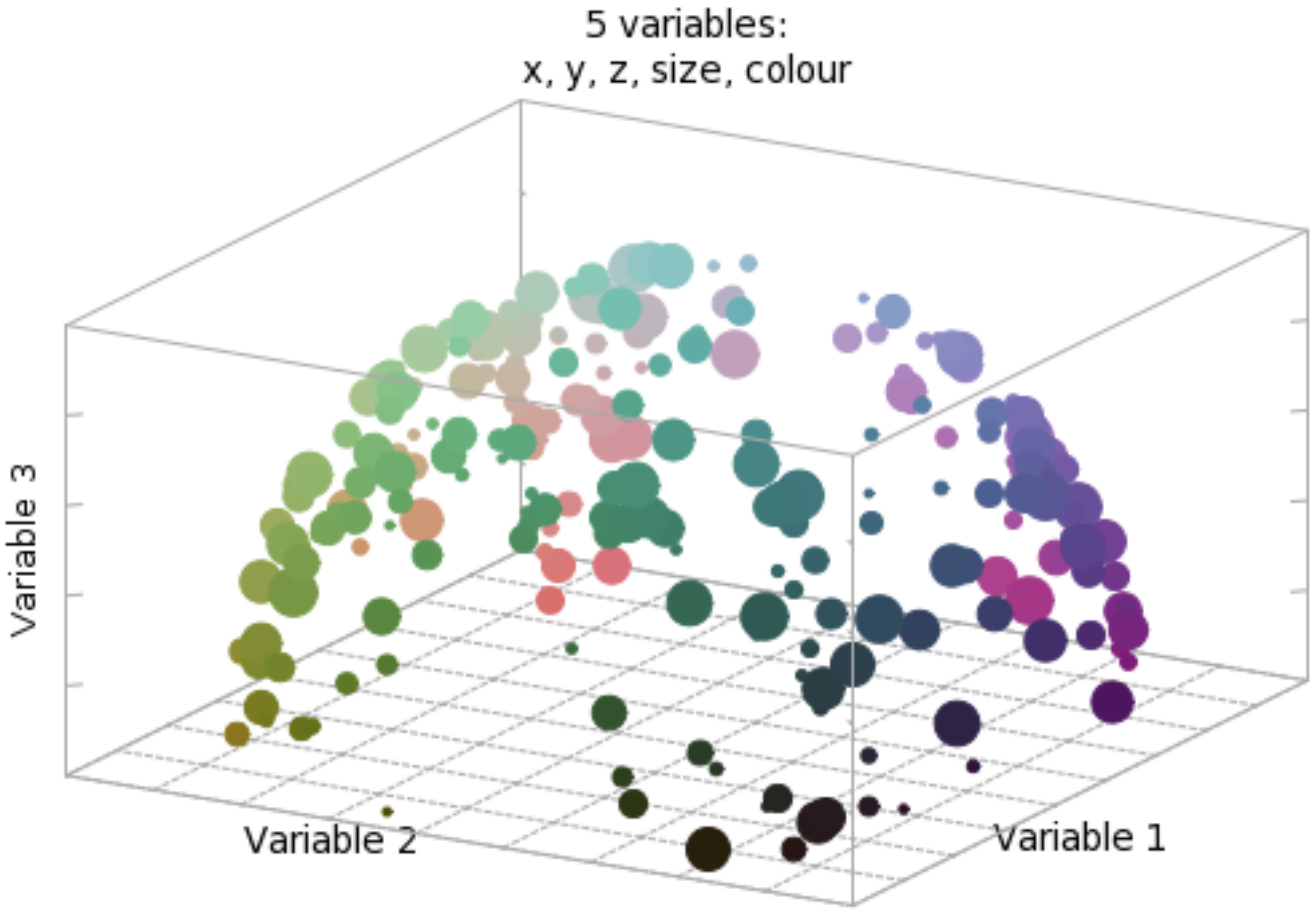
Mapping data to space

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



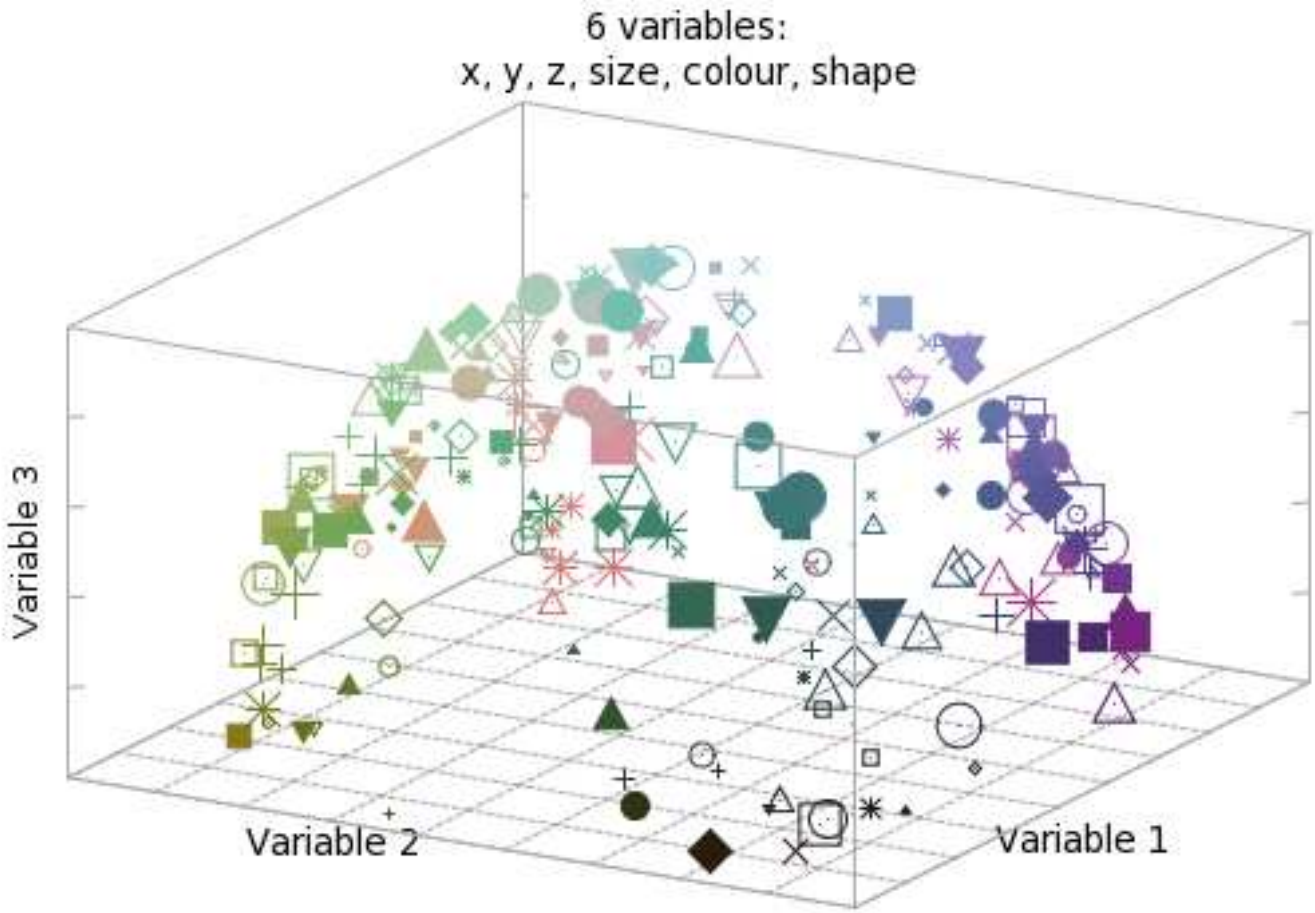
Mapping data to space

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



Mapping data to space

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



Visualising high dimensional data

- 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

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

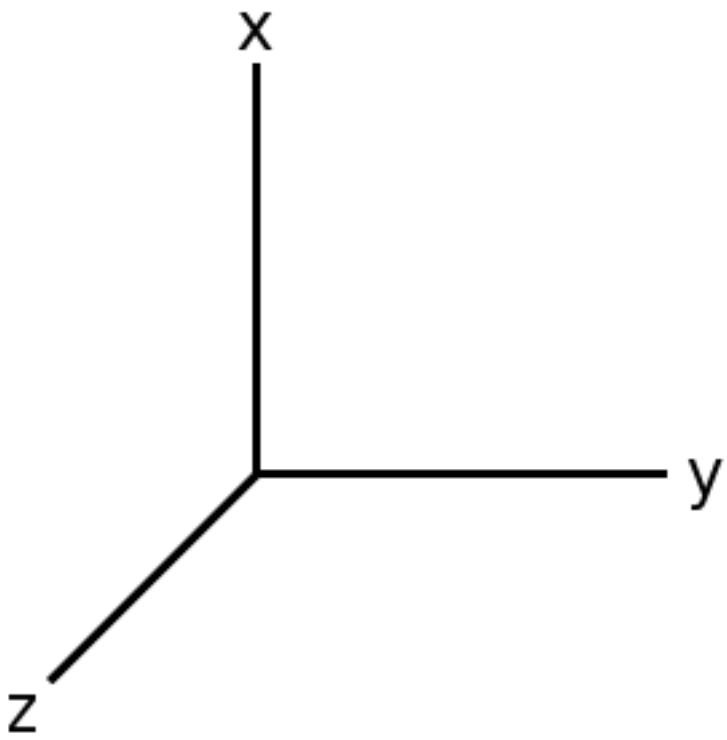
Parallel Coordinate Plots

Parallel Coordinate Plots

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

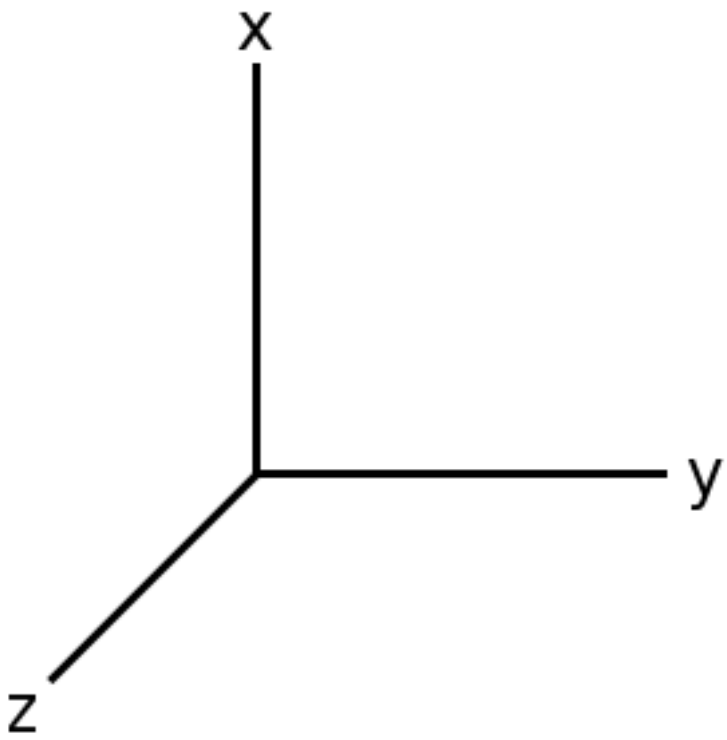
Parallel Coordinate Plots

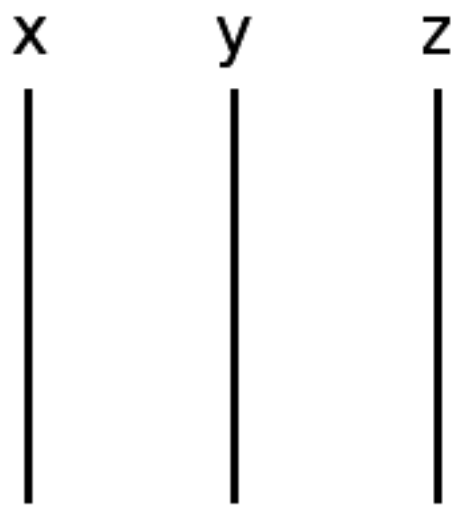
- Dimensions (axes) are laid out in parallel
- Use lines to join variable values



Parallel Coordinate Plots

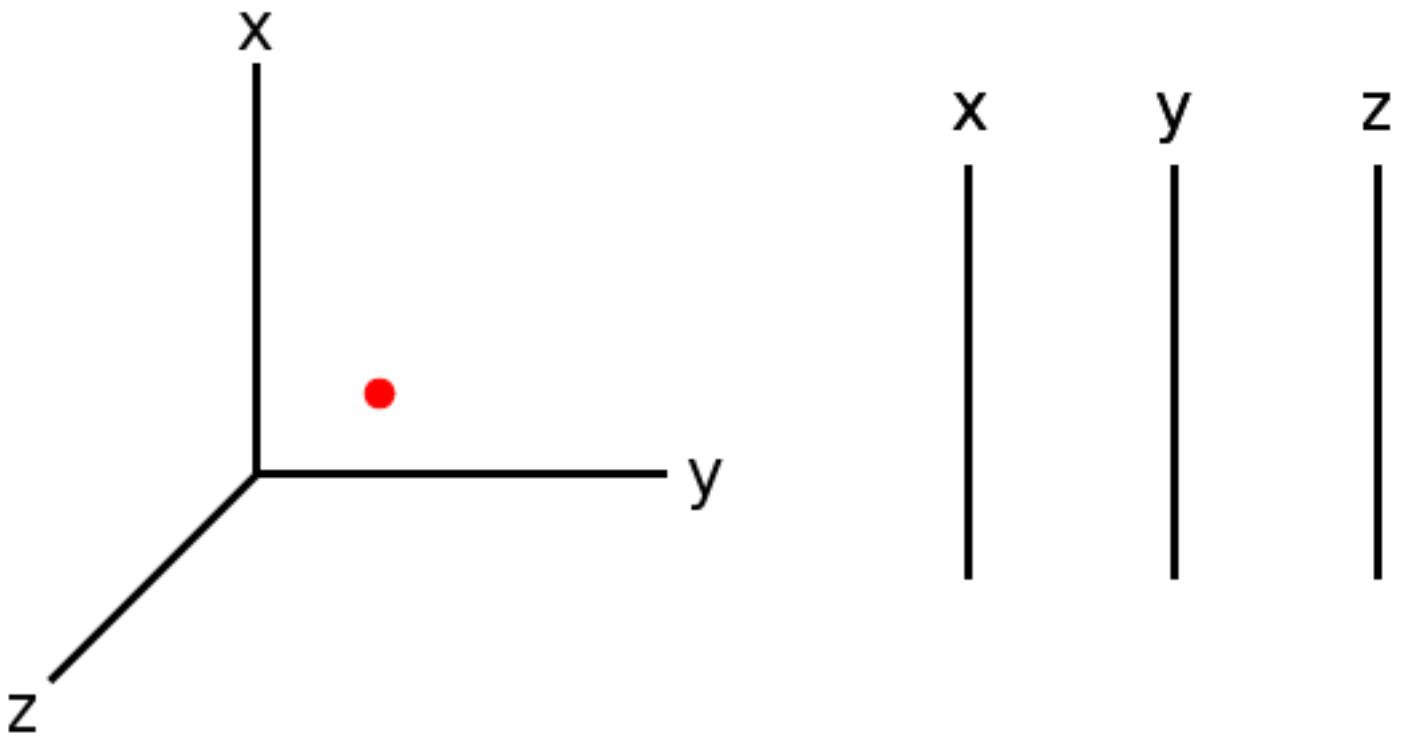
- Dimensions (axes) are laid out in parallel
- Use lines to join variable values





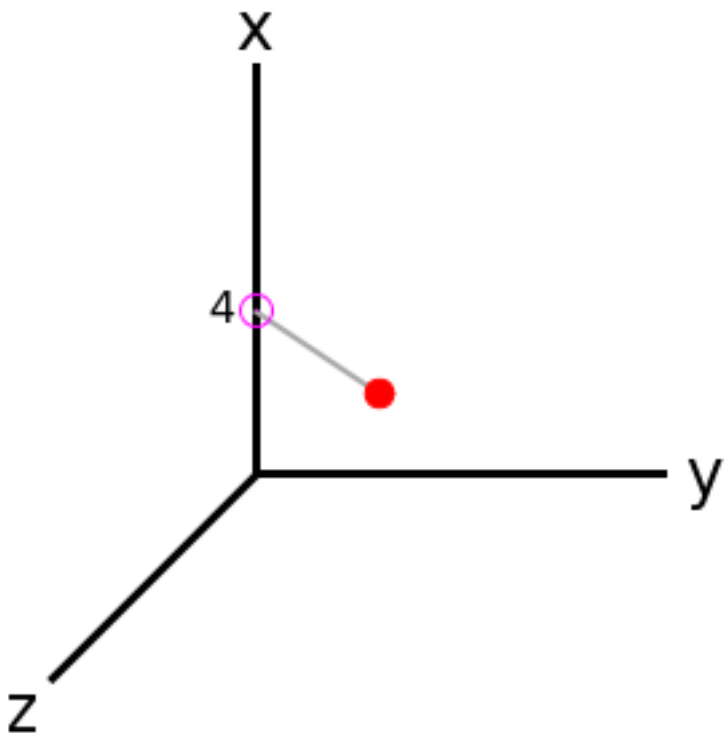
Parallel Coordinate Plots

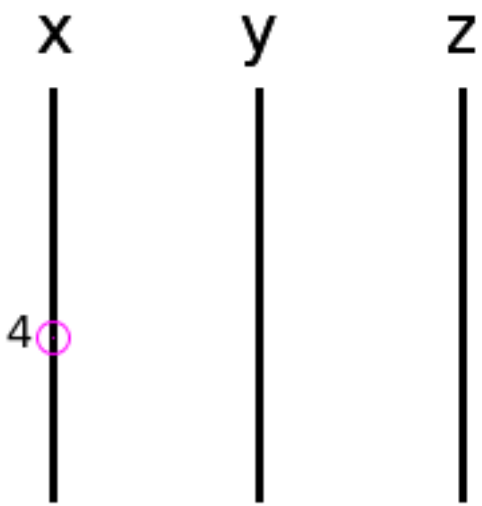
- Dimensions (axes) are laid out in parallel
- Use lines to join variable values



Parallel Coordinate Plots

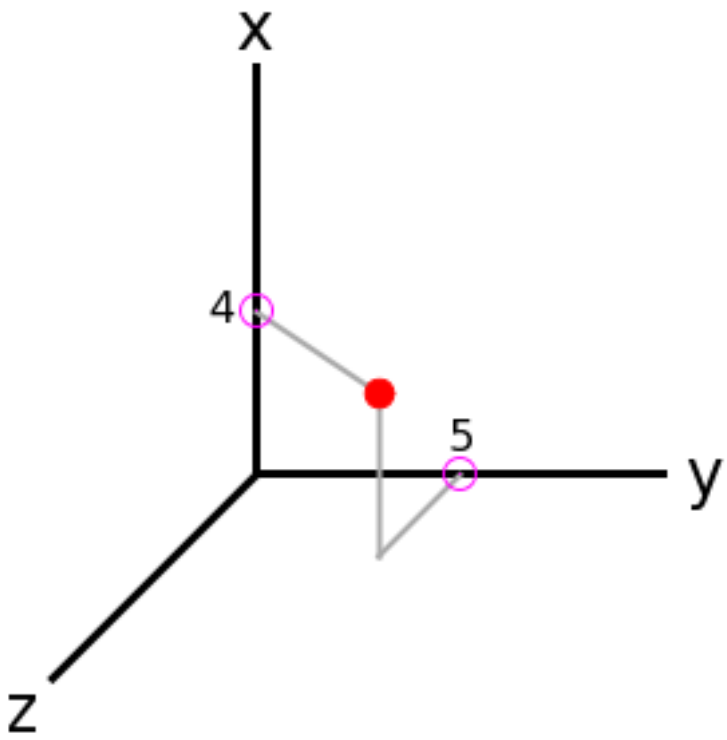
- Dimensions (axes) are laid out in parallel
- Use lines to join variable values

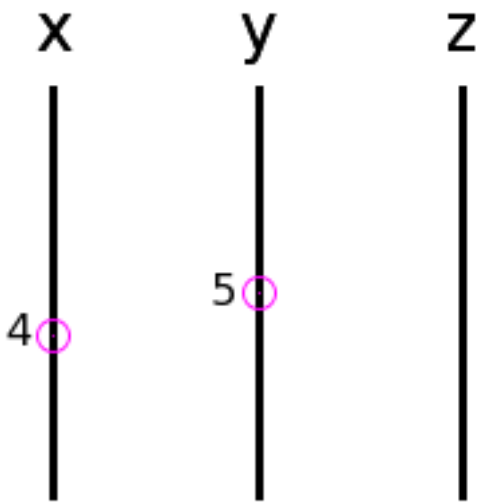




Parallel Coordinate Plots

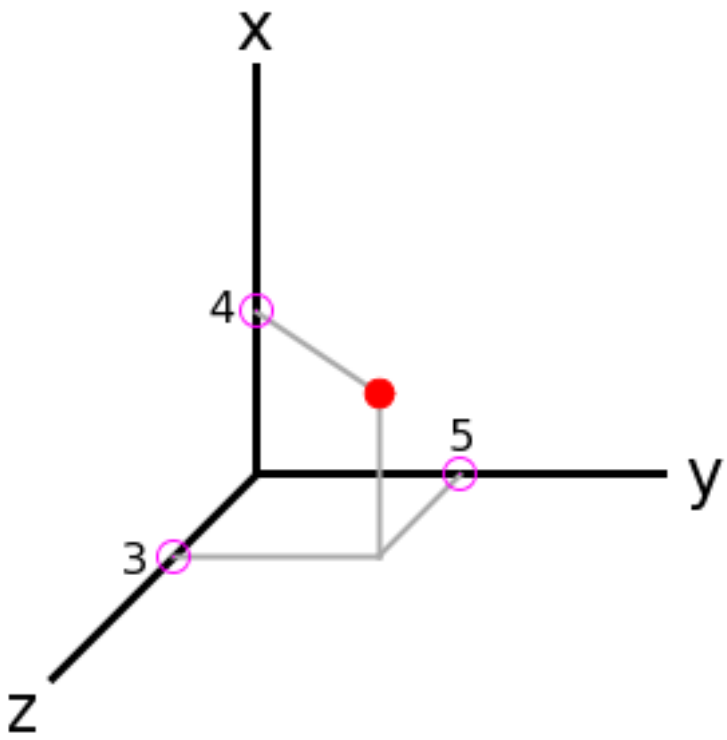
- Dimensions (axes) are laid out in parallel
- Use lines to join variable values

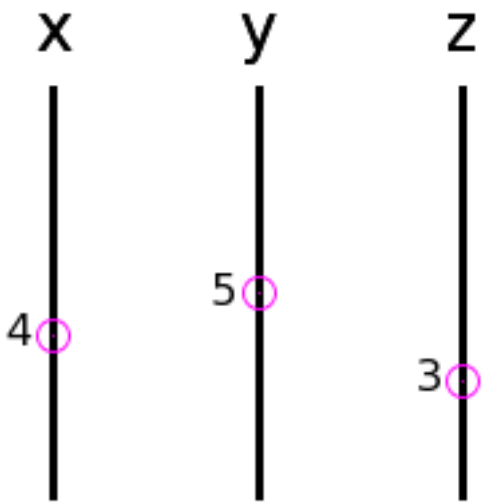




Parallel Coordinate Plots

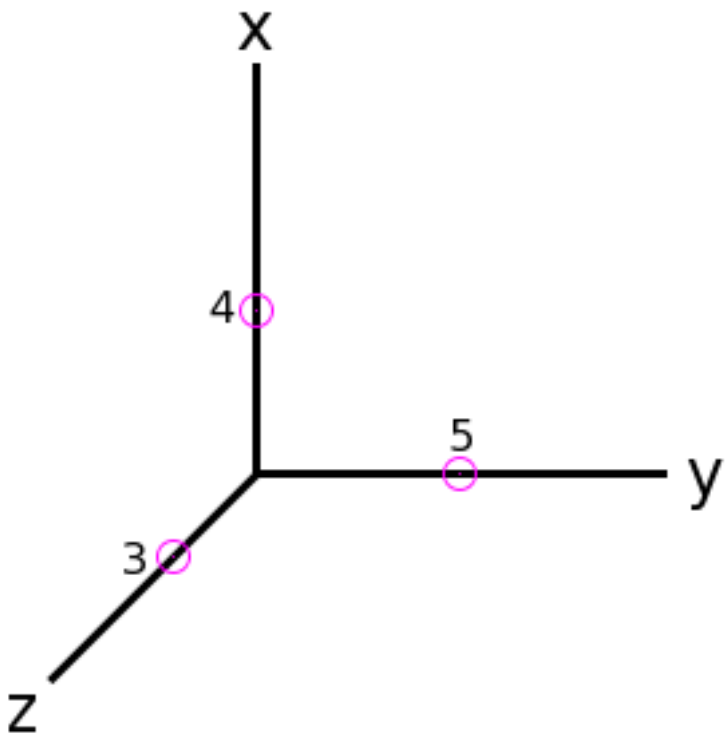
- Dimensions (axes) are laid out in parallel
- Use lines to join variable values

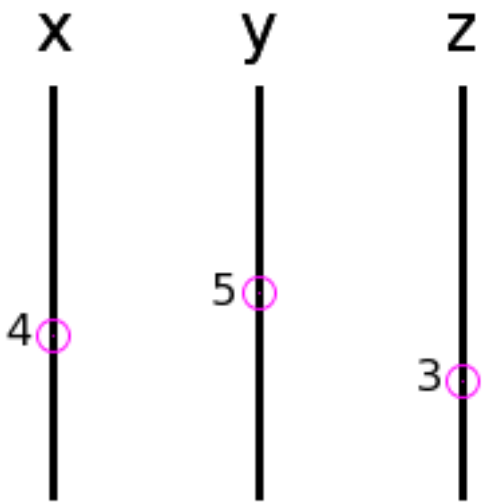




Parallel Coordinate Plots

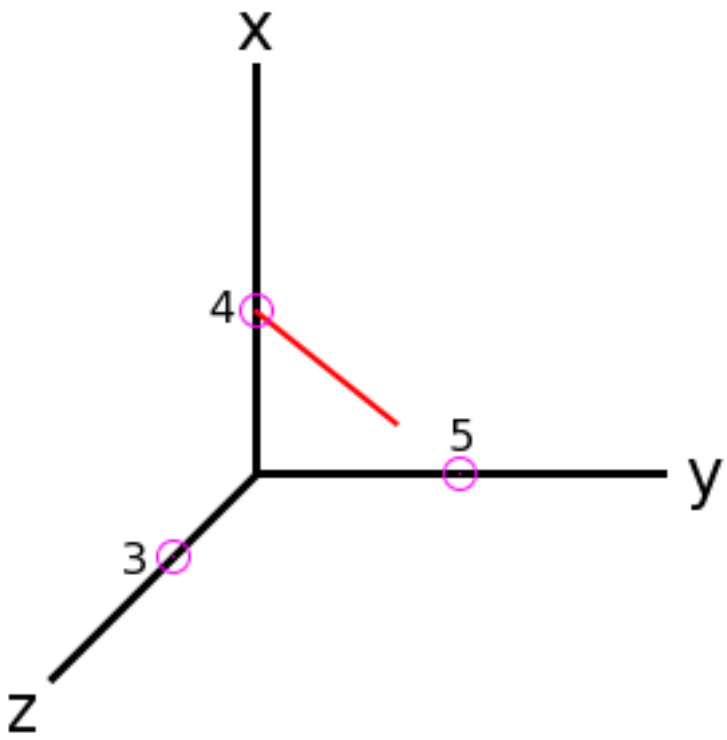
- Dimensions (axes) are laid out in parallel
- Use lines to join variable values

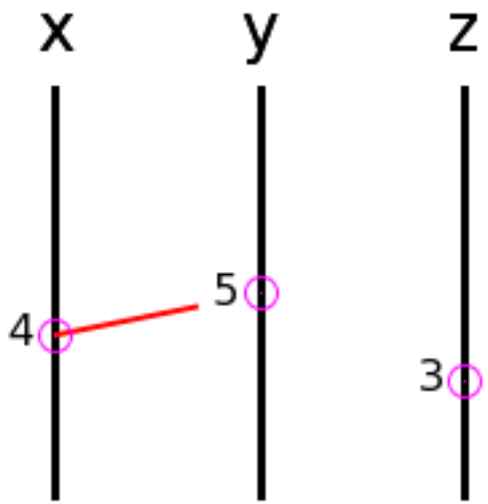




Parallel Coordinate Plots

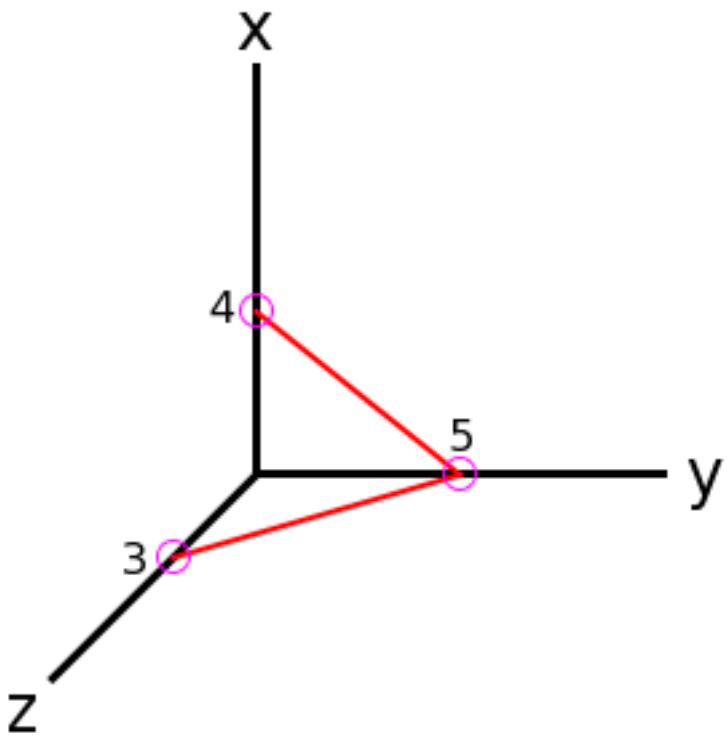
- Dimensions (axes) are laid out in parallel
- Use lines to join variable values

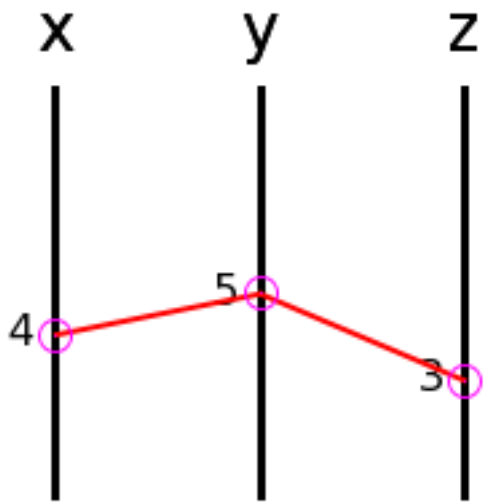




Parallel Coordinate Plots

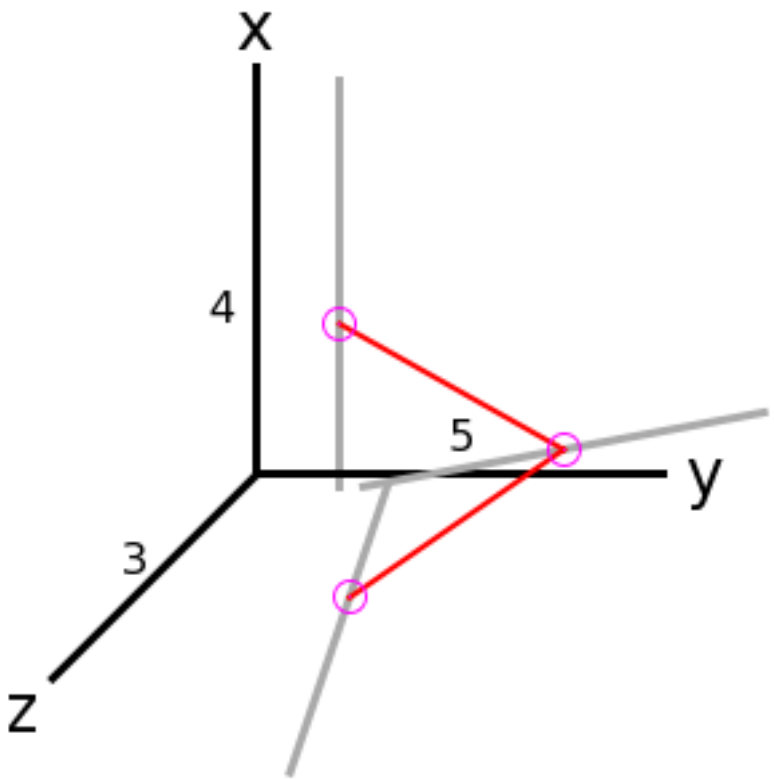
- Dimensions (axes) are laid out in parallel
- Use lines to join variable values

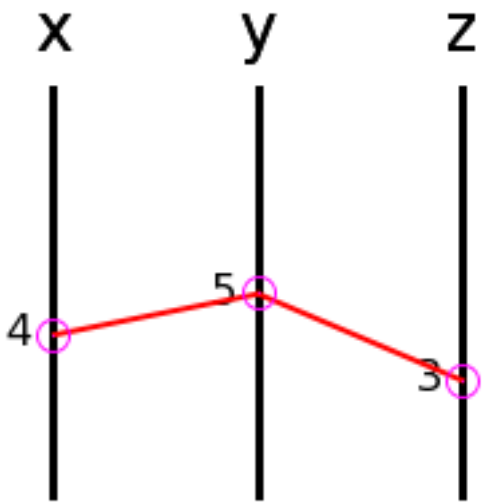




Parallel Coordinate Plots

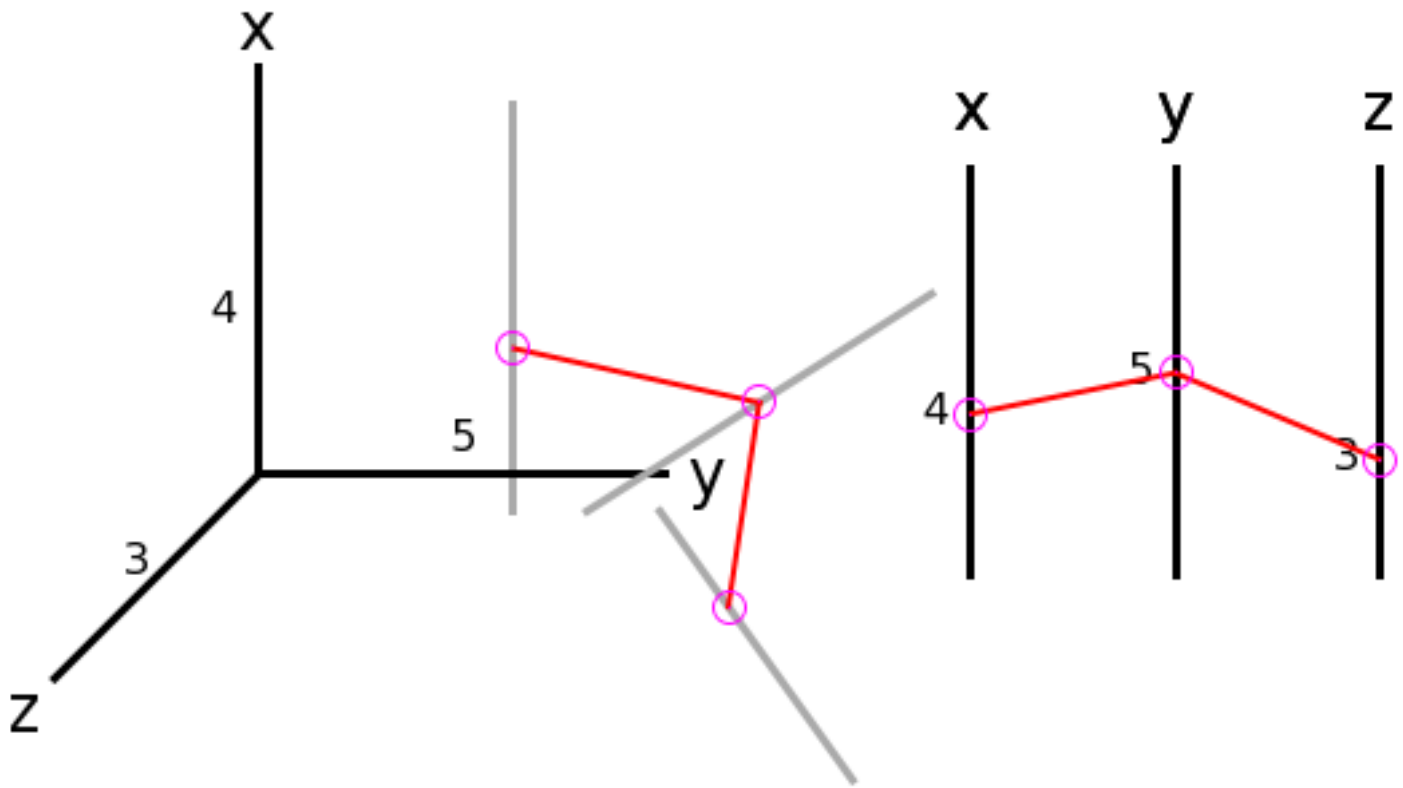
- Dimensions (axes) are laid out in parallel
- Use lines to join variable values





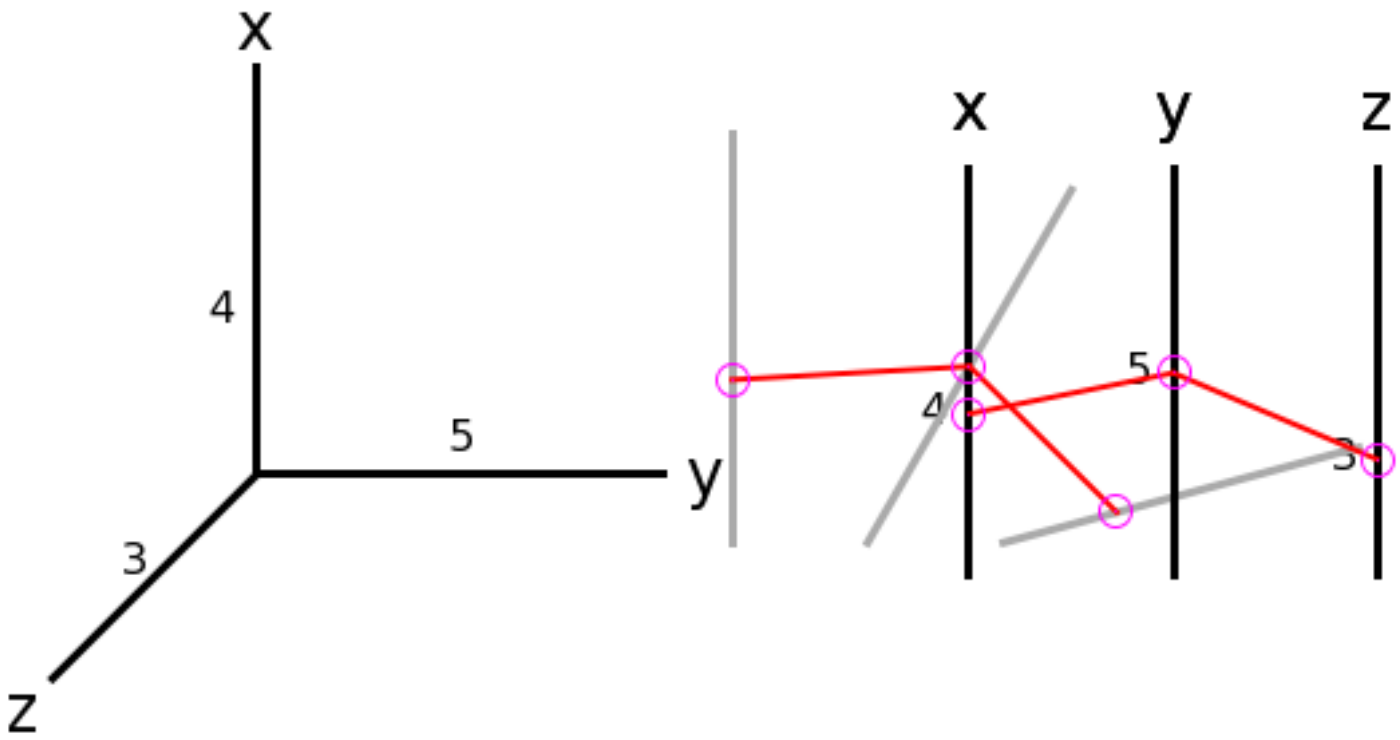
Parallel Coordinate Plots

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



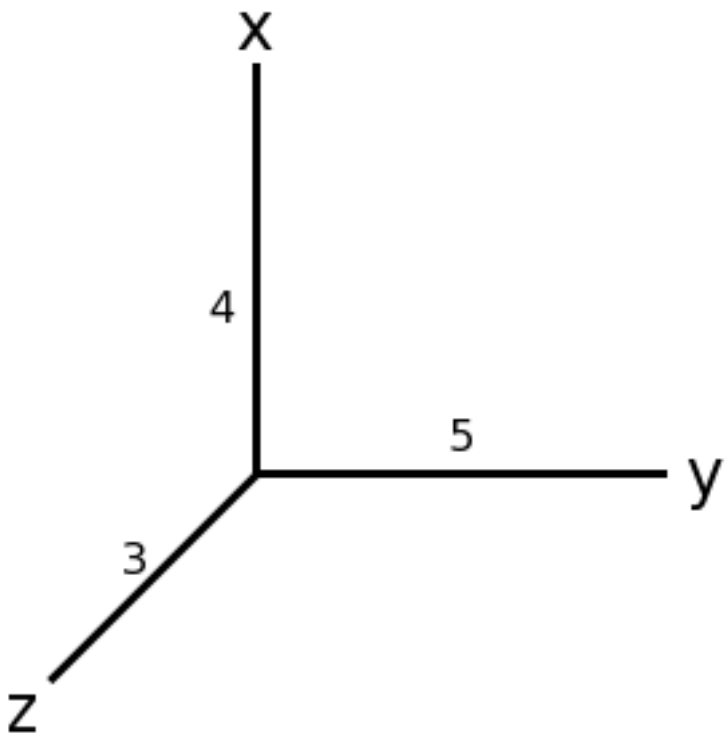
Parallel Coordinate Plots

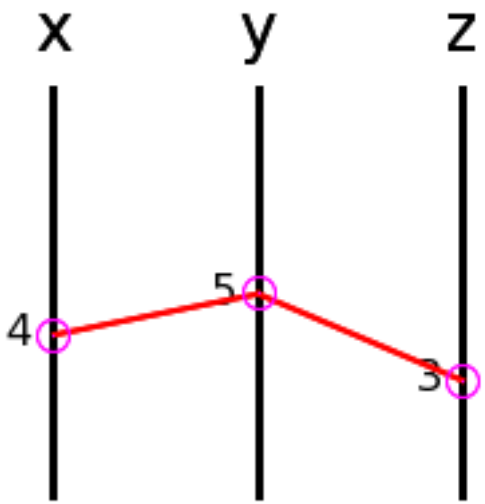
- Dimensions (axes) are laid out in parallel
- Use lines to join variable values



Parallel Coordinate Plots

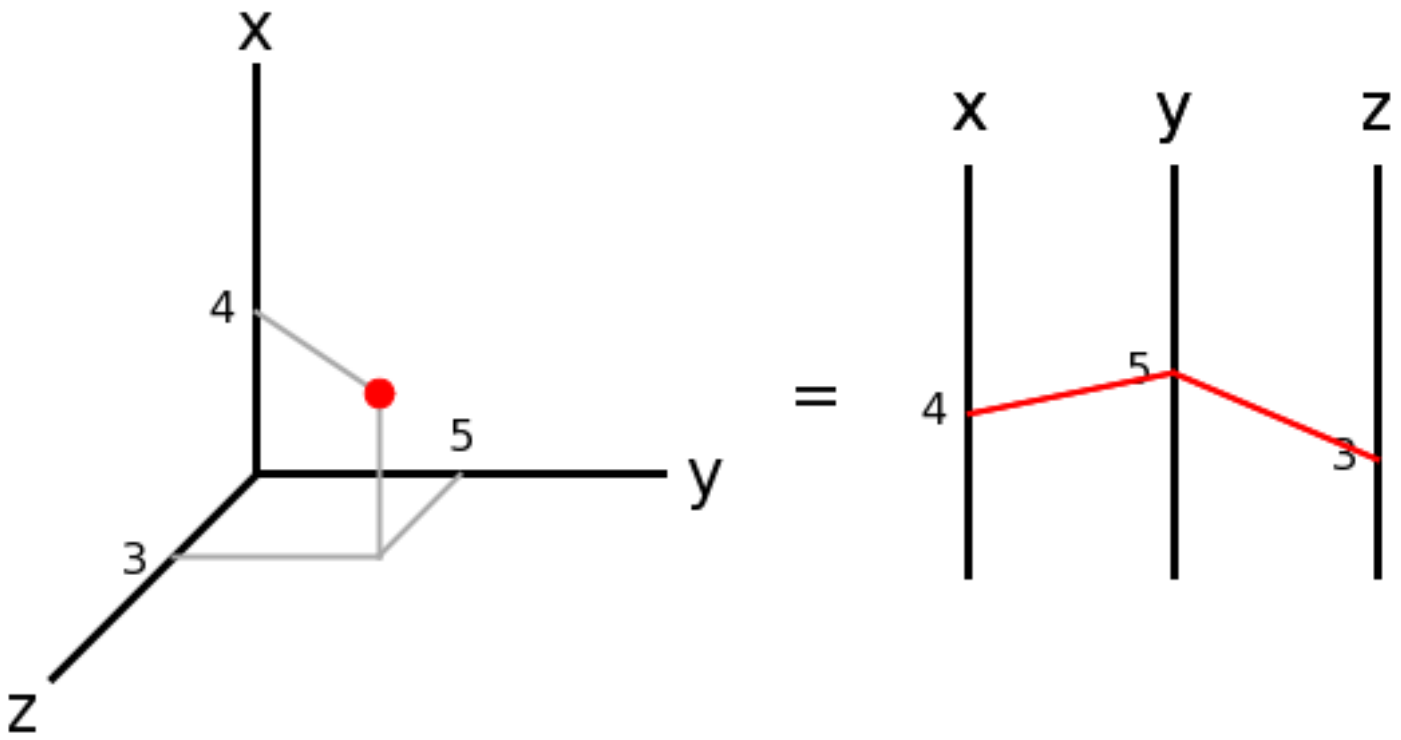
- Dimensions (axes) are laid out in parallel
- Use lines to join variable values





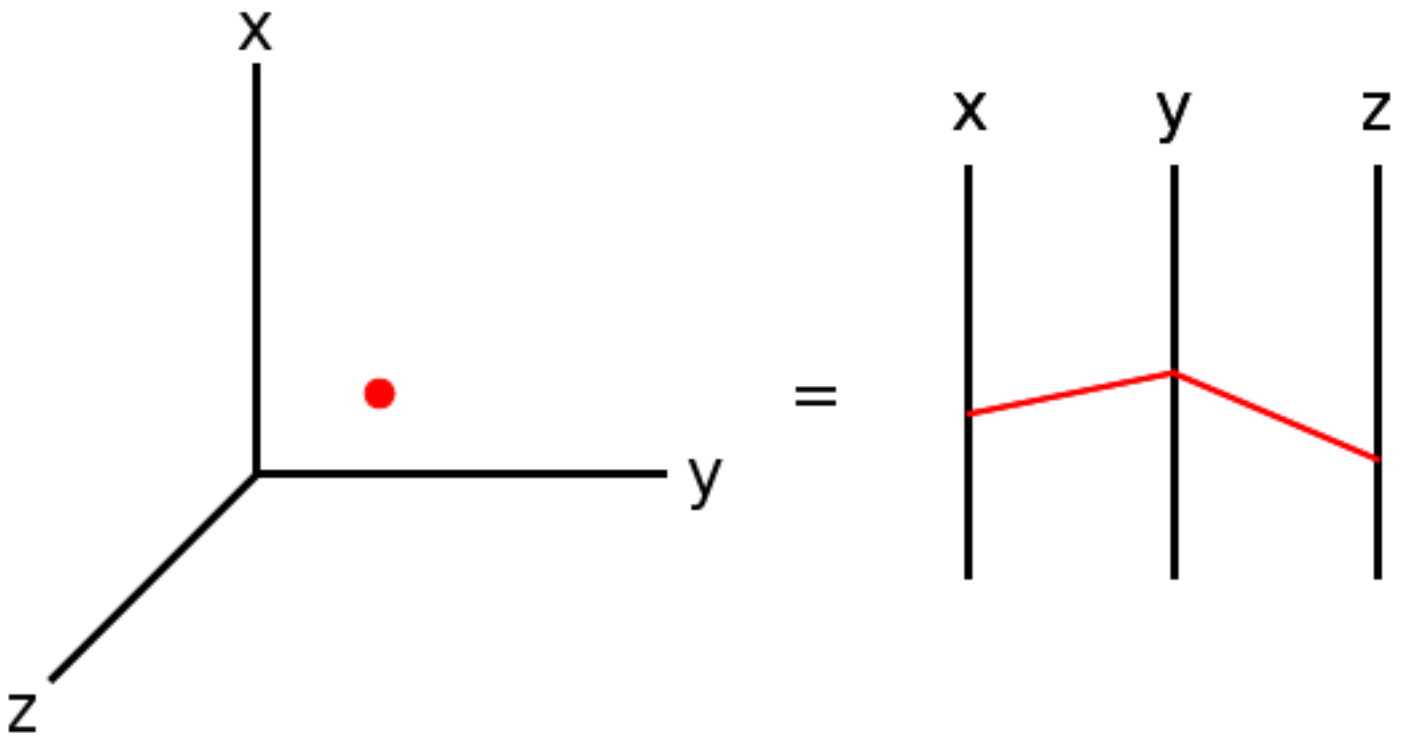
Parallel Coordinate Plots

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



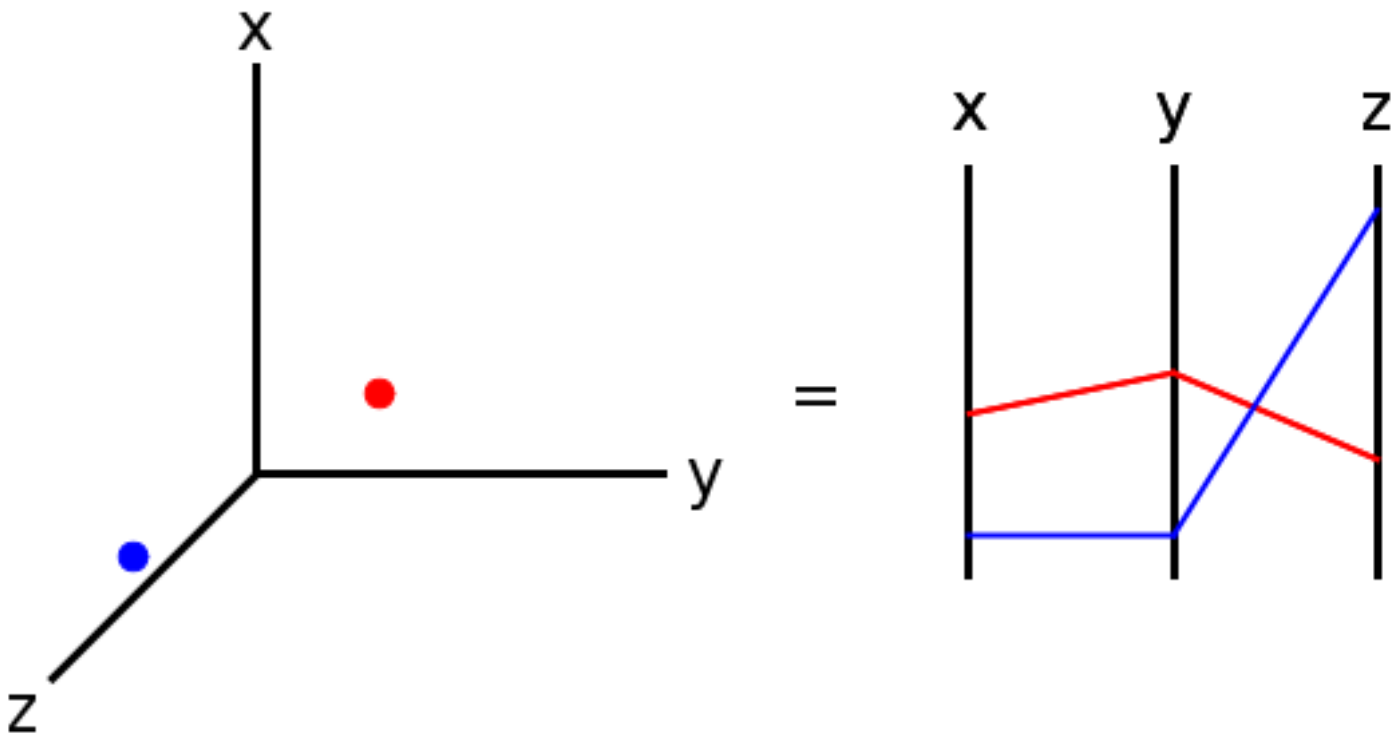
Parallel Coordinate Plots

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



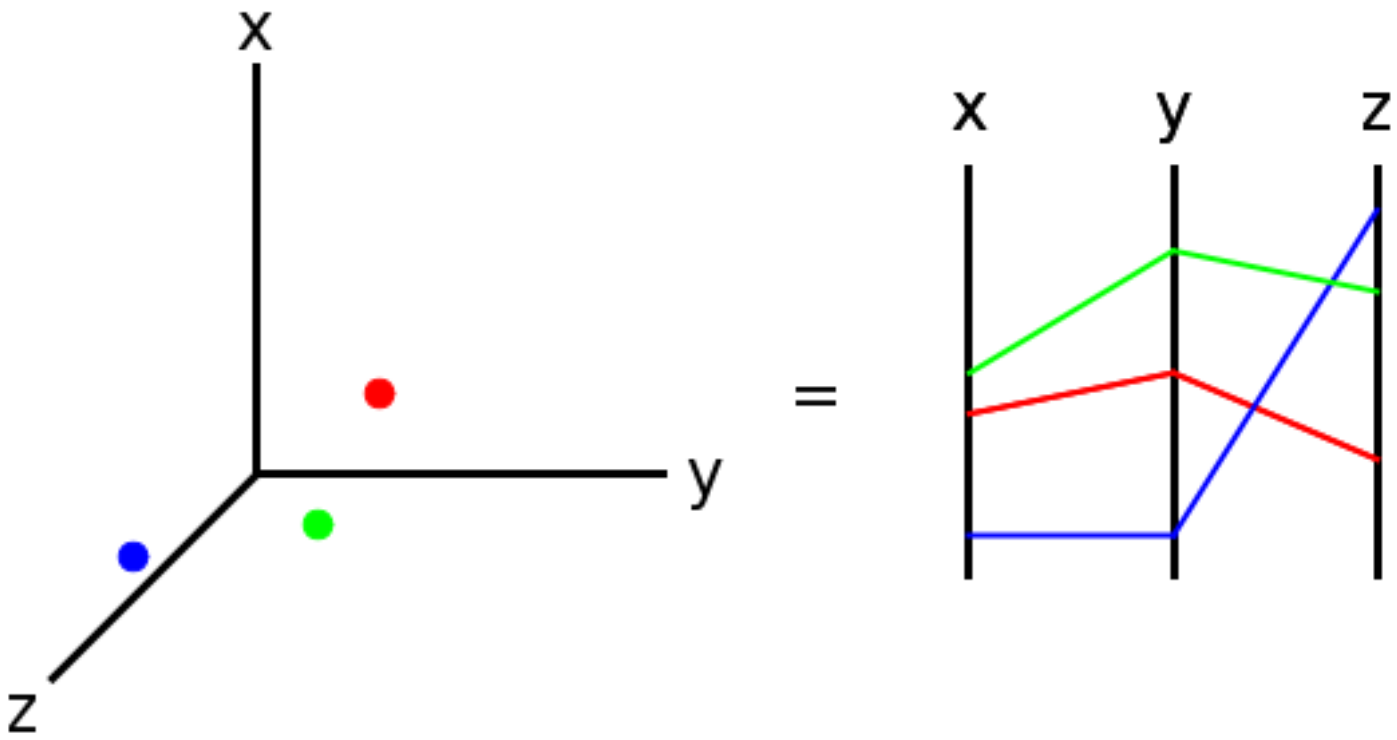
Parallel Coordinate Plots

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



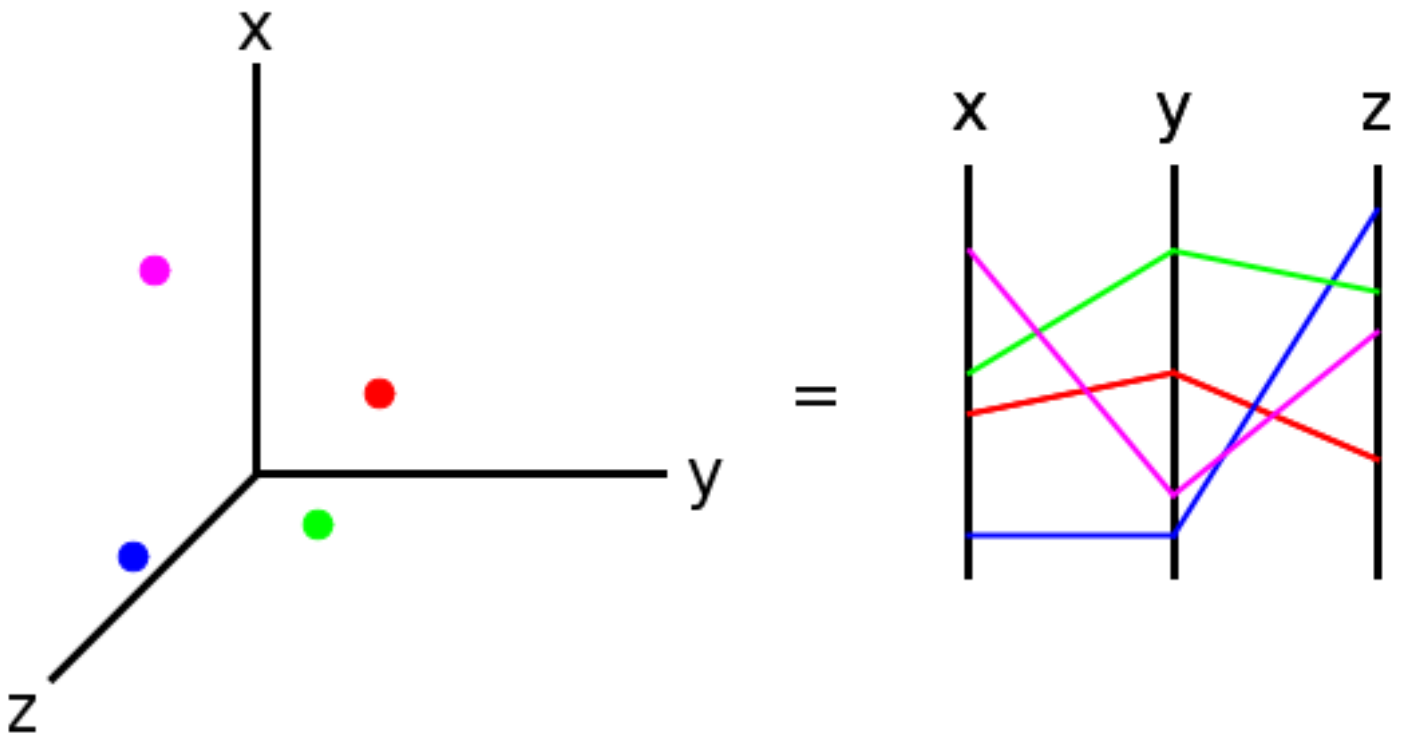
Parallel Coordinate Plots

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

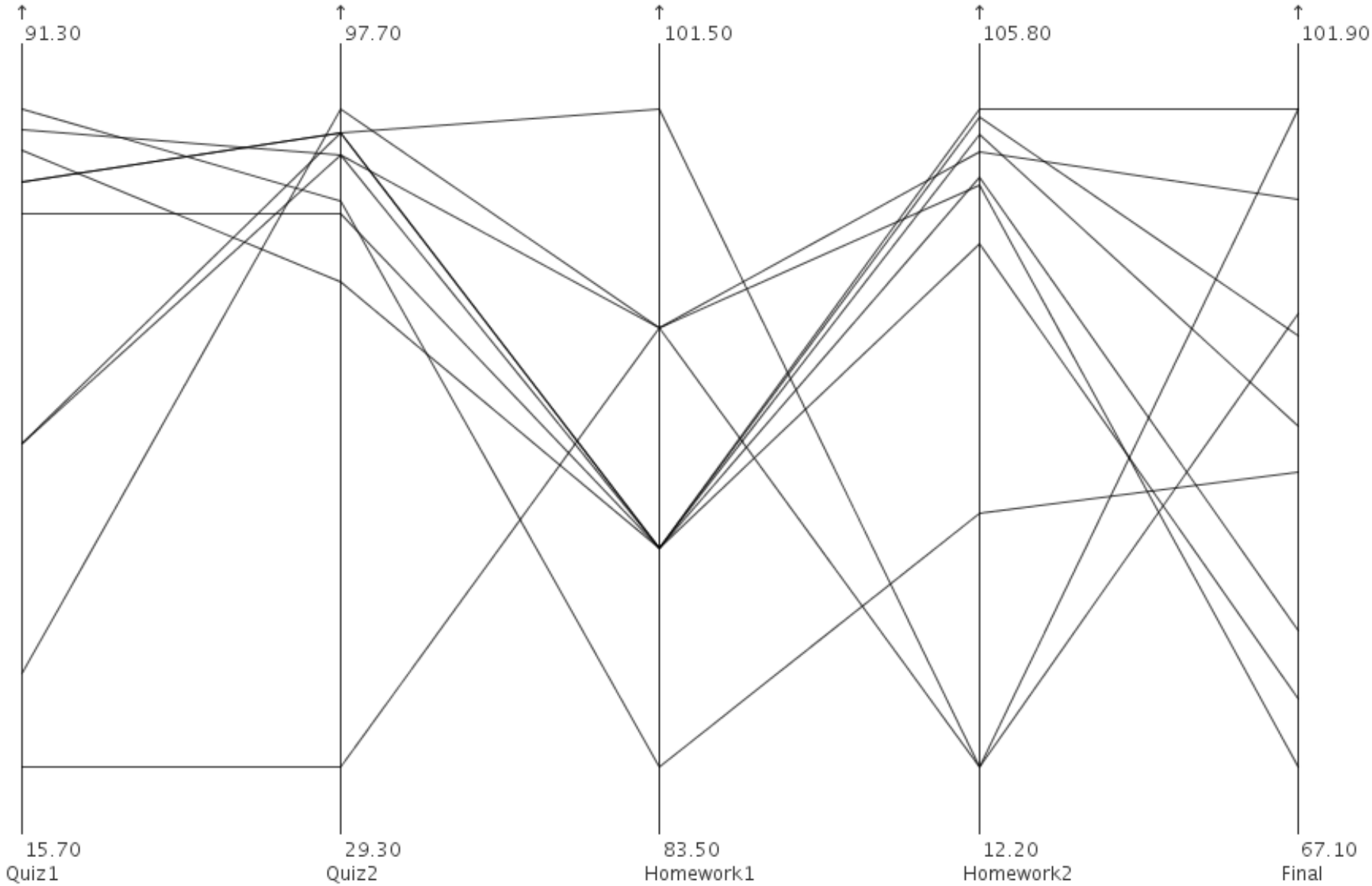


Parallel Coordinate Plots

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

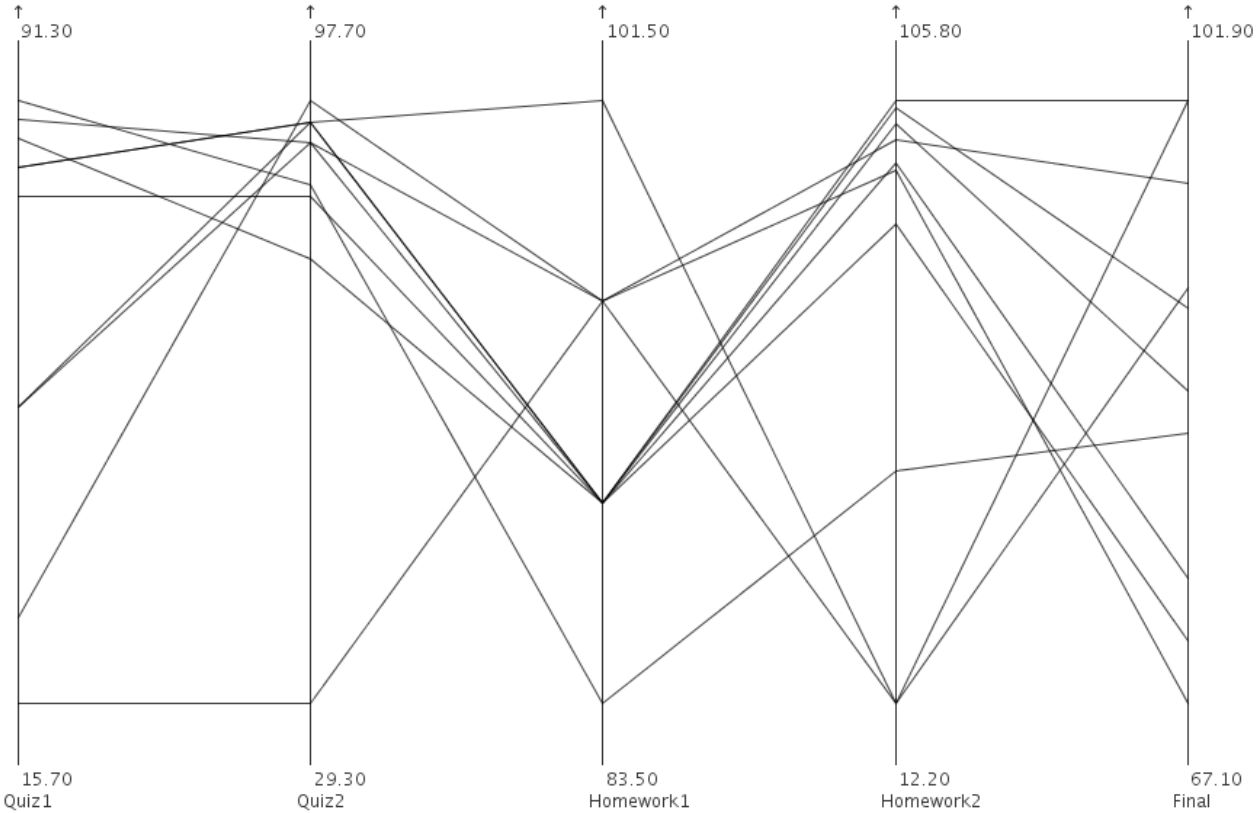


Example


91.30
97.70
101.50
105.80
101.90
15.70
Quiz1
29.30
Quiz2
83.50
Homework1
12.20
Homework2
67.10
Final

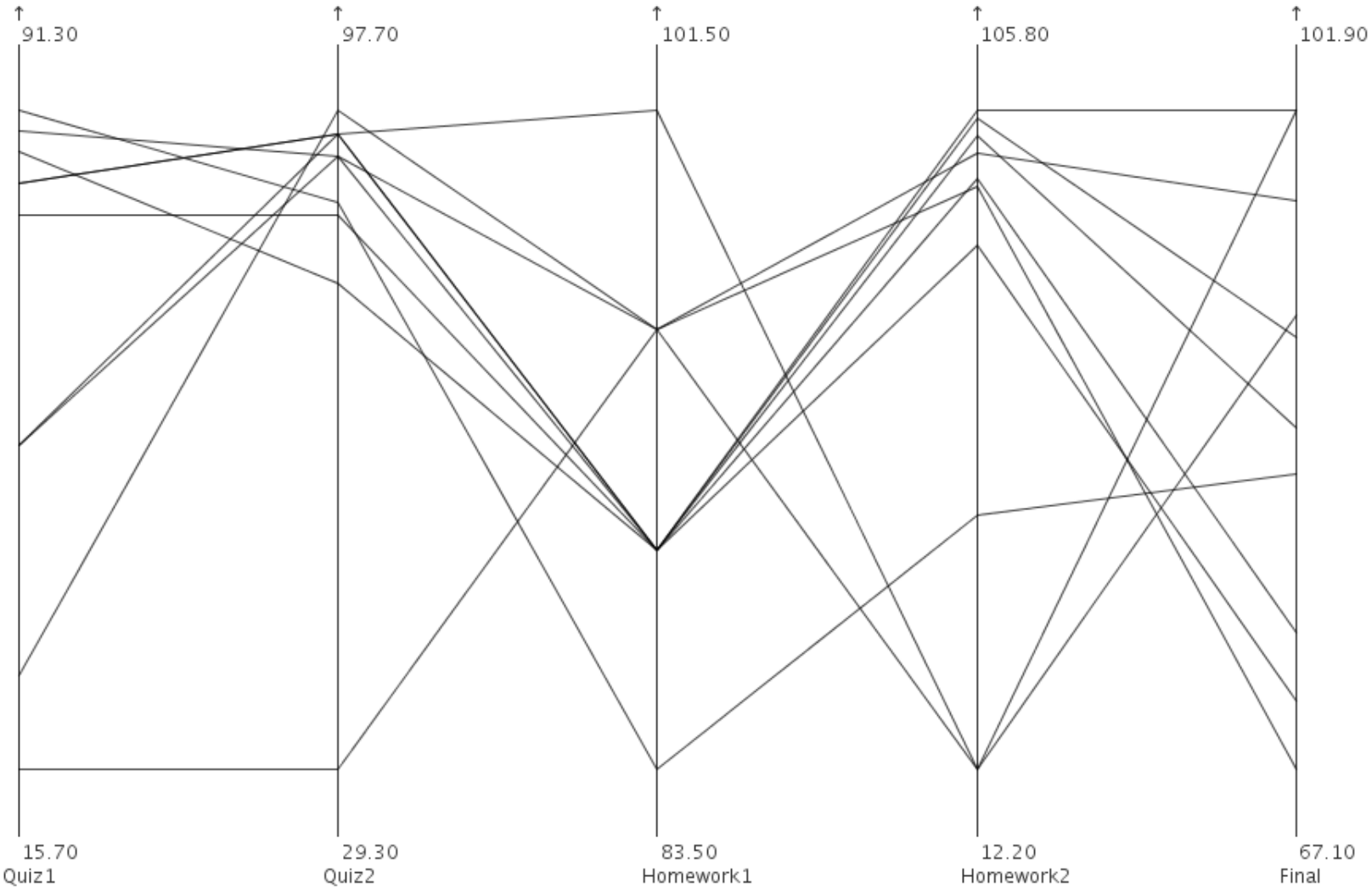
Advantages

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



Challenge: Ambiguity

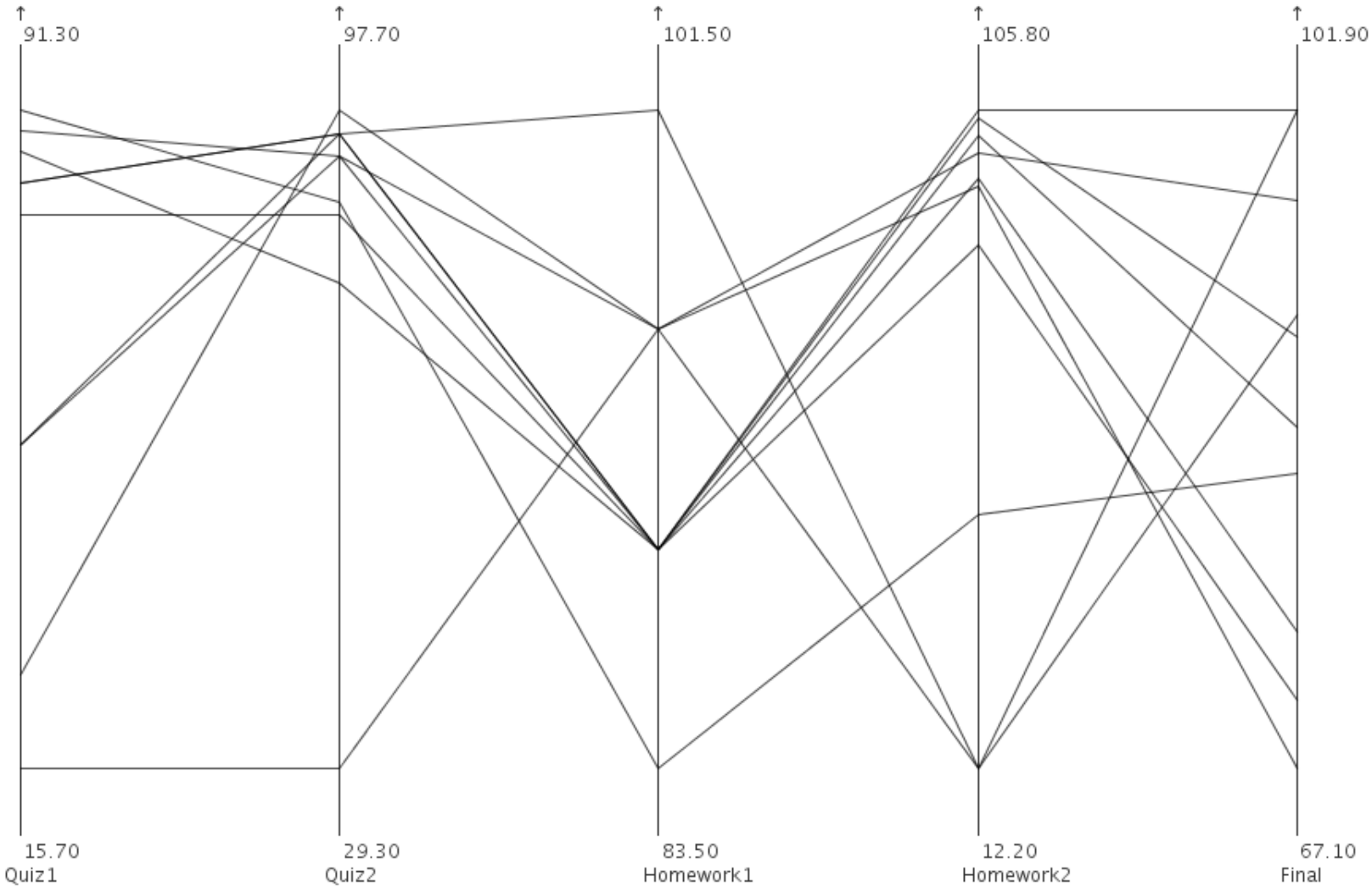
Common values cause ambiguity



Challenge: Ambiguity

Common values cause ambiguity

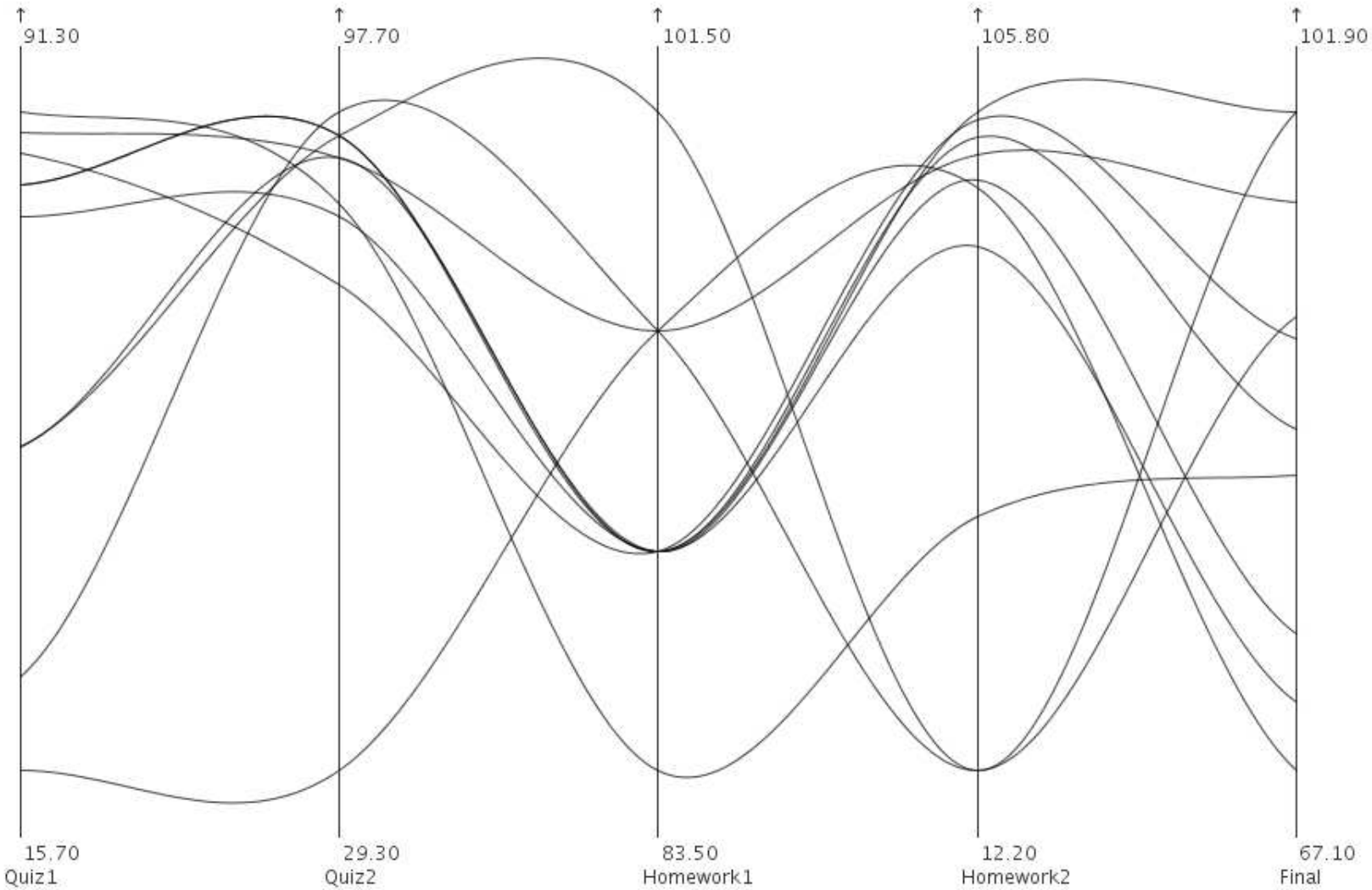
Solution: (a) Curves



Challenge: Ambiguity

Common values cause ambiguity

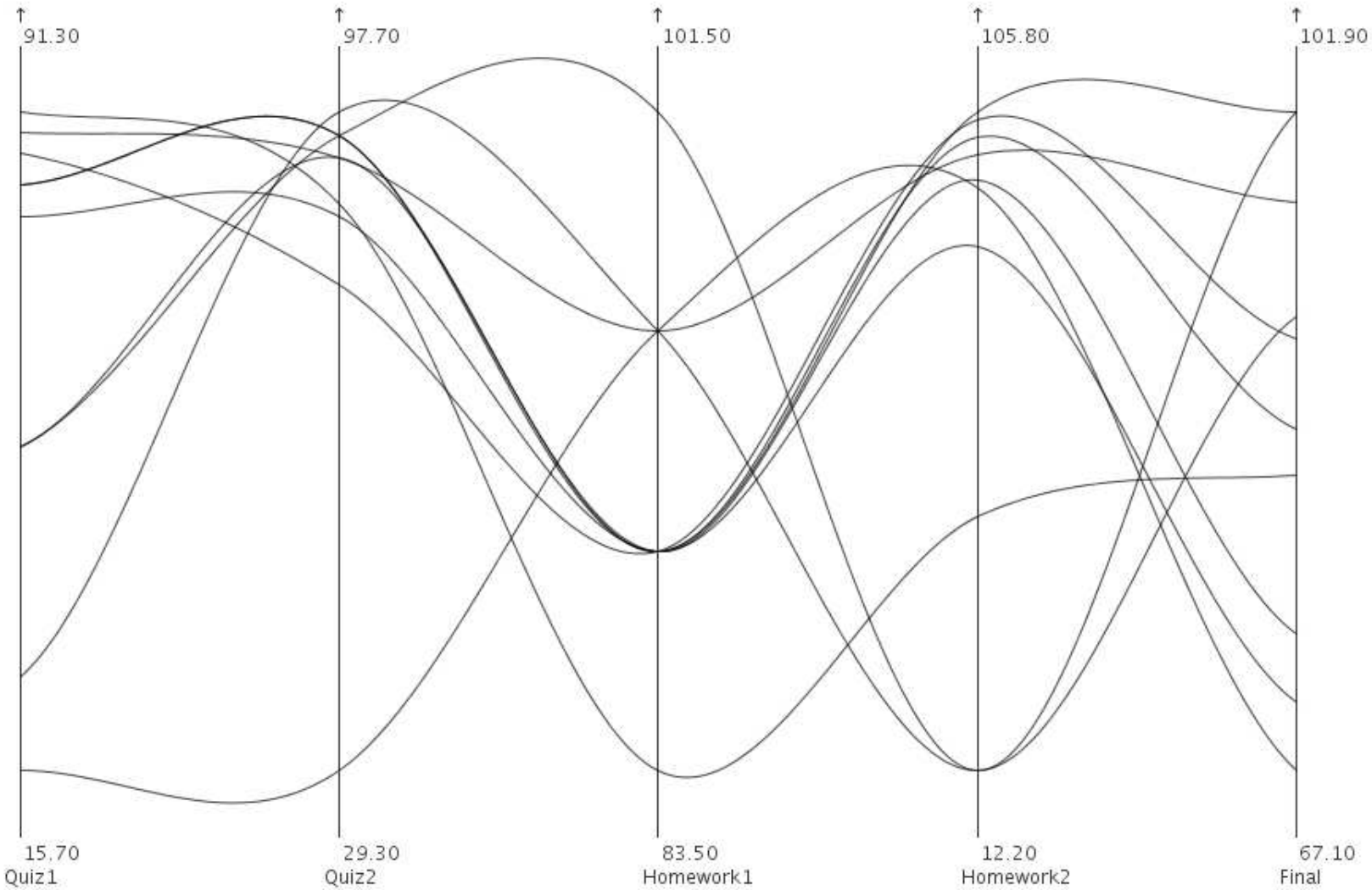
Solution: (a) Curves



Challenge: Ambiguity

Common values cause ambiguity

Solution: (a) Curves (b) Spreading

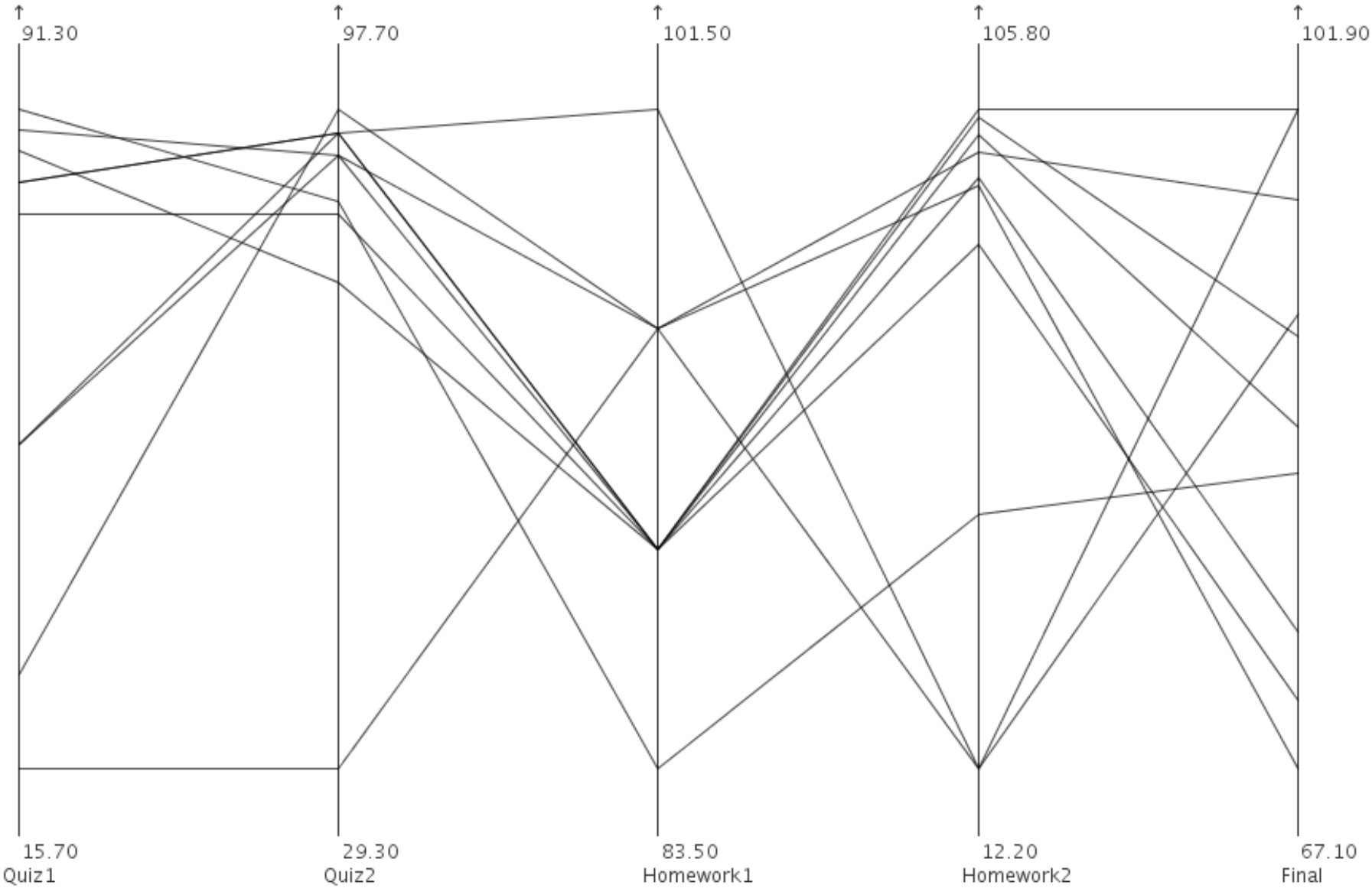


Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation

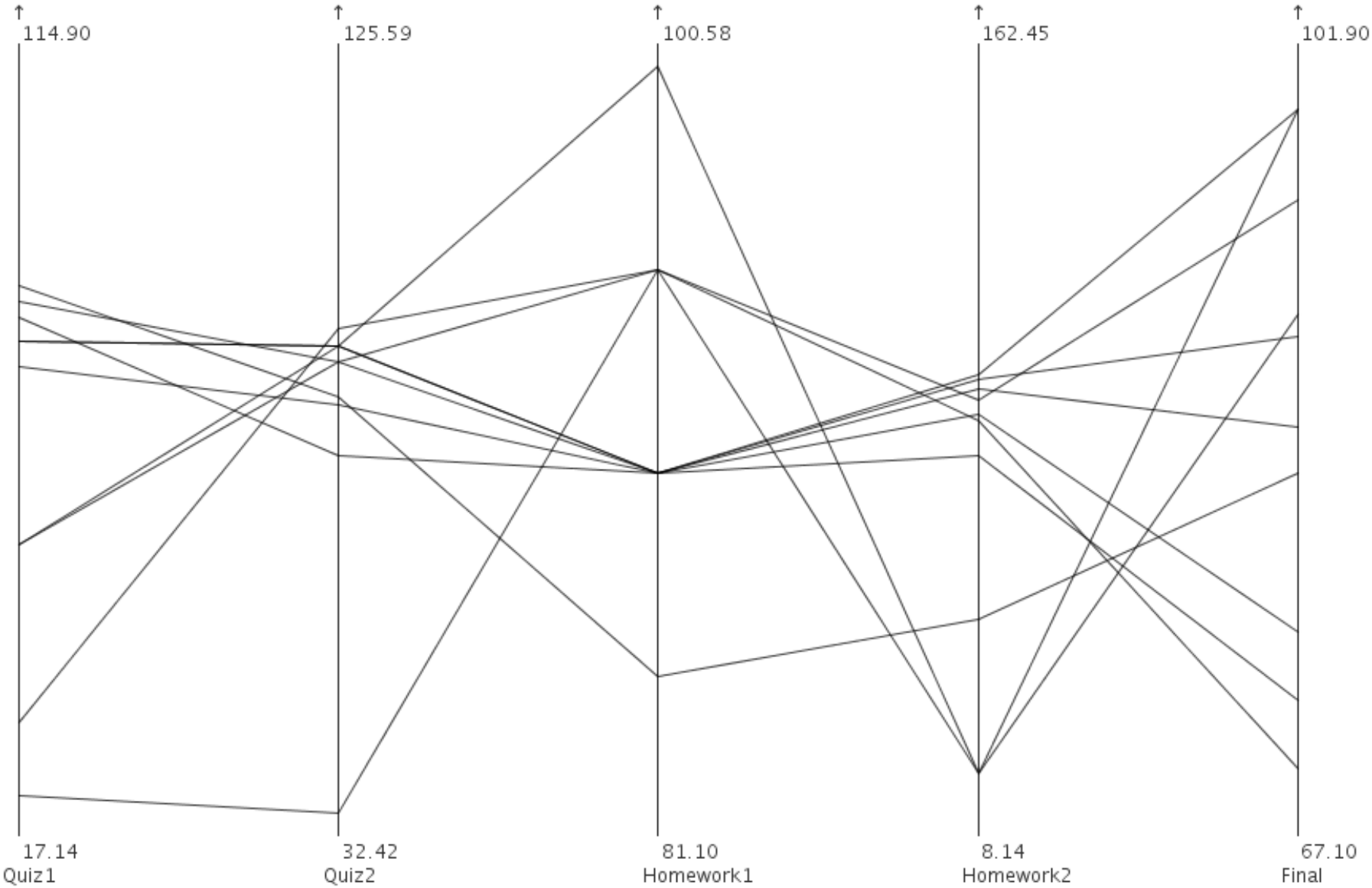
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation



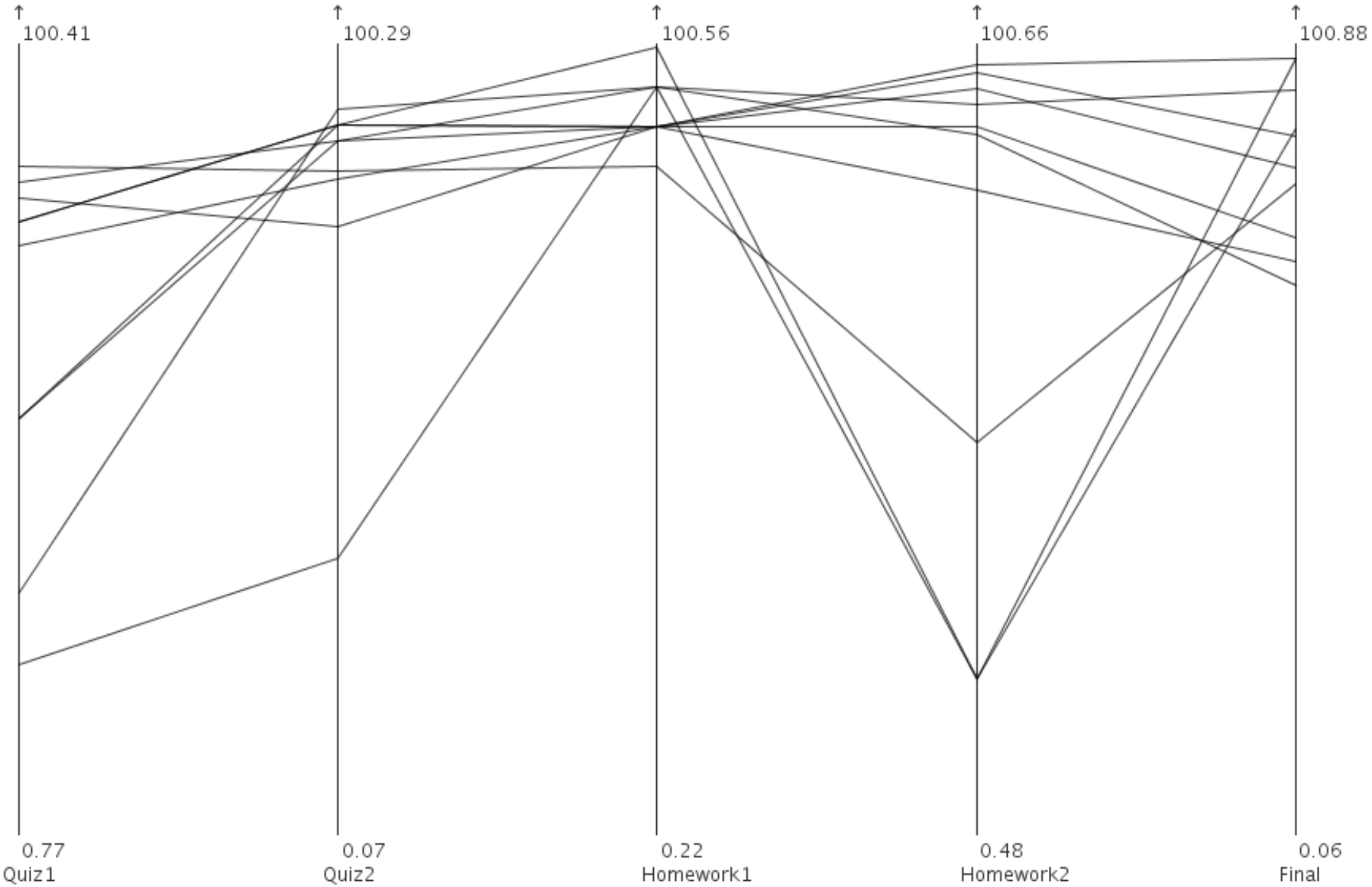
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation



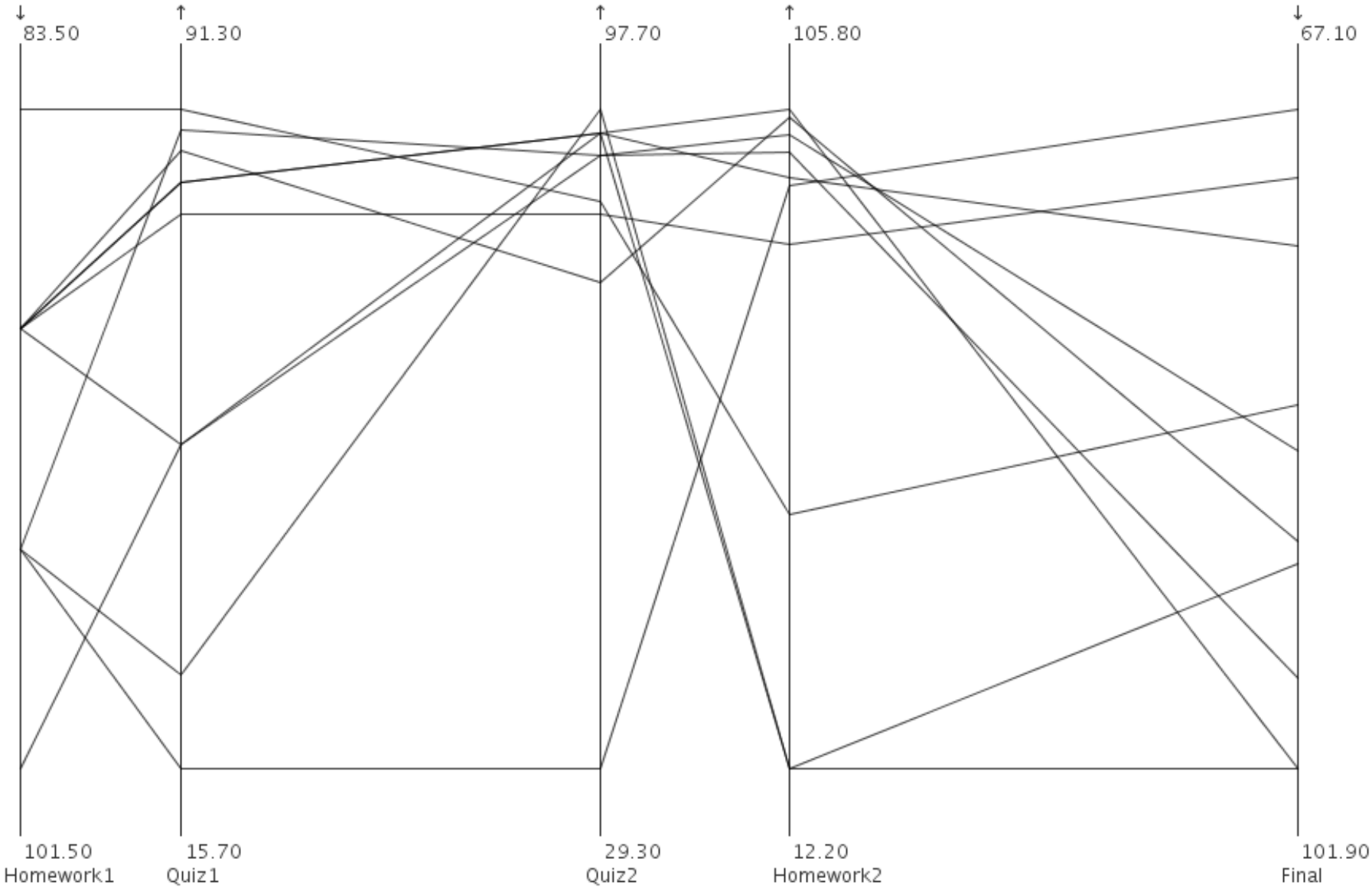
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation



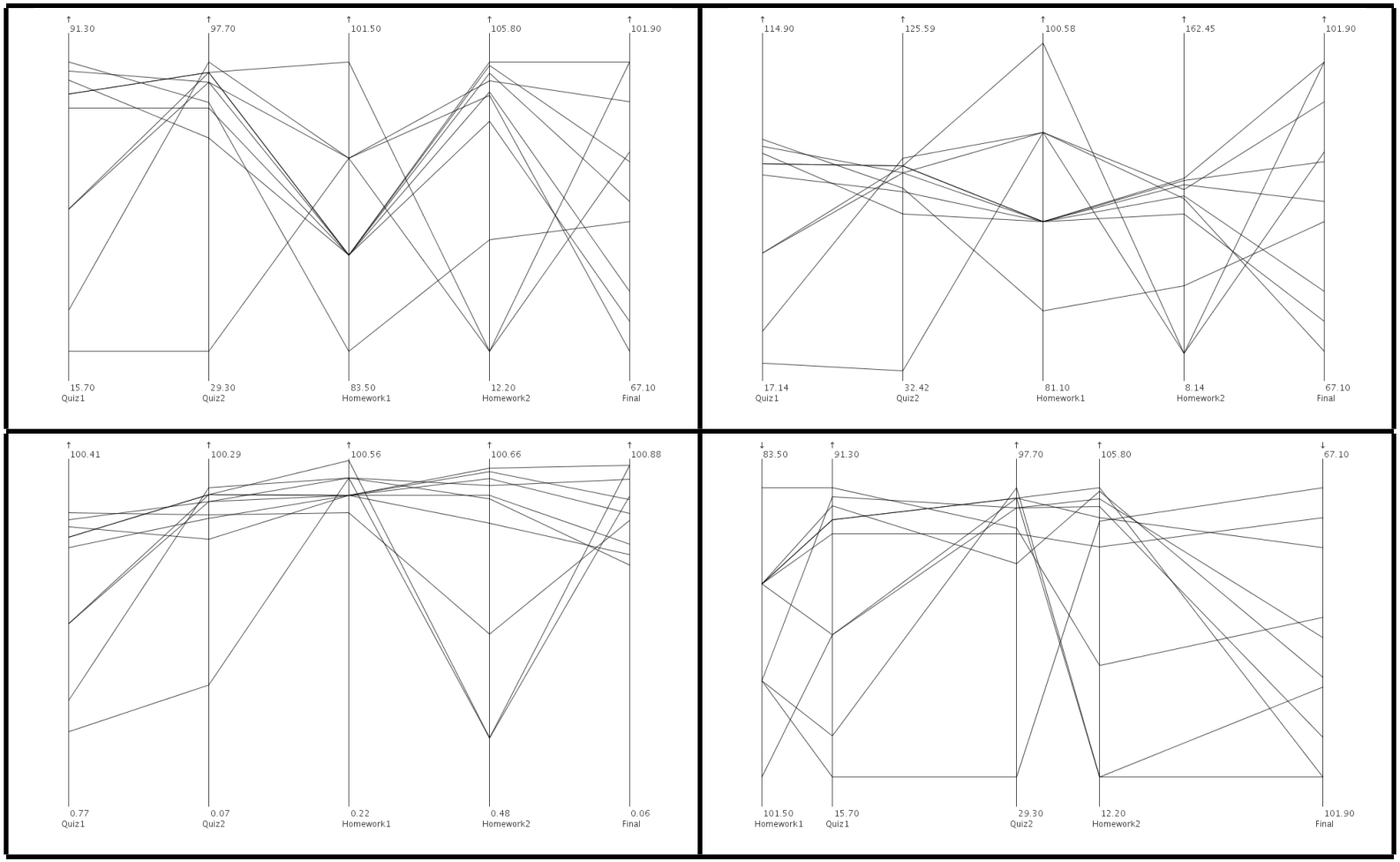
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation



Challenge: Axis parameters

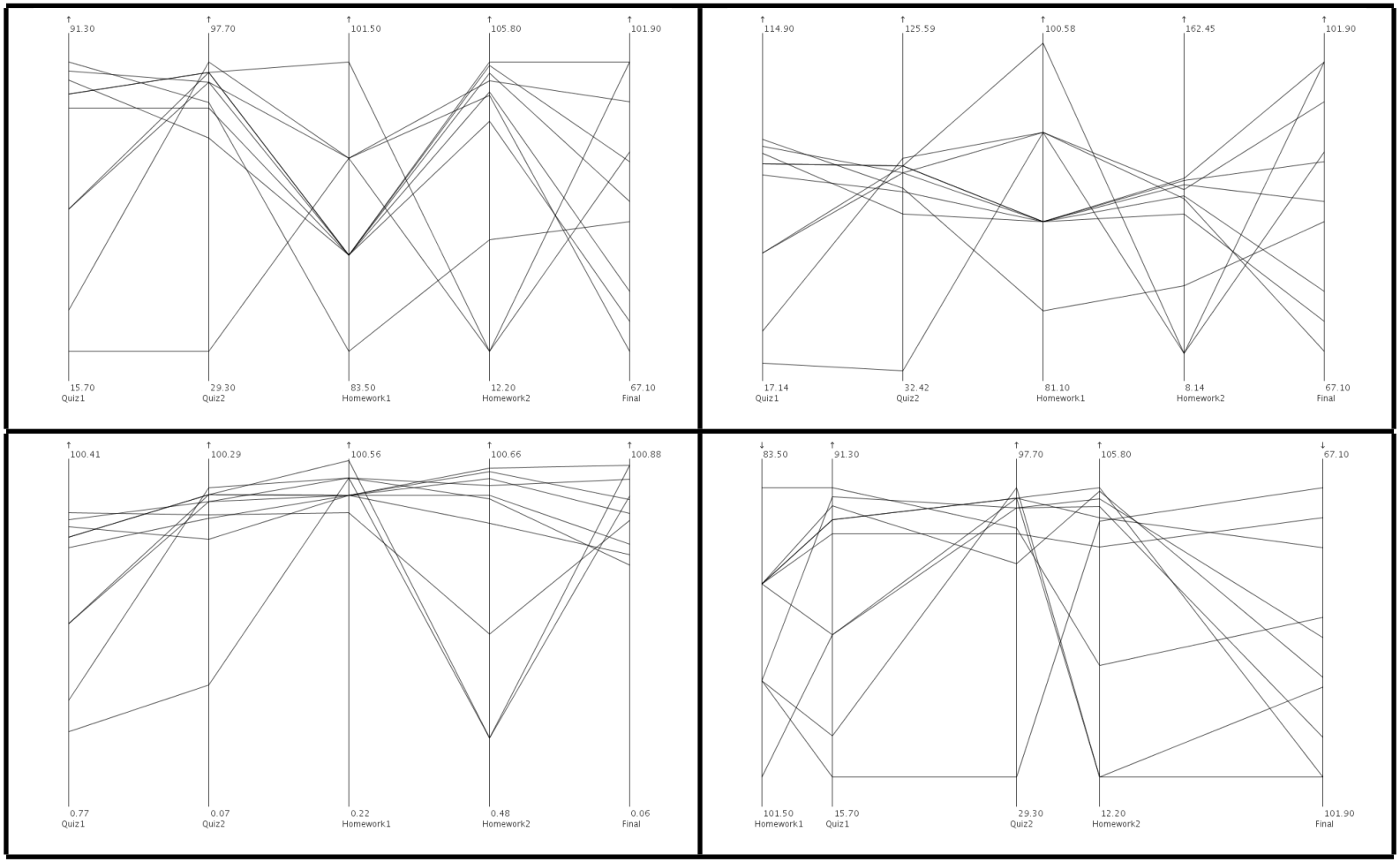
View affected by axis ordering, sign, scaling, translation



Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation

Solution: Interactive manipulation

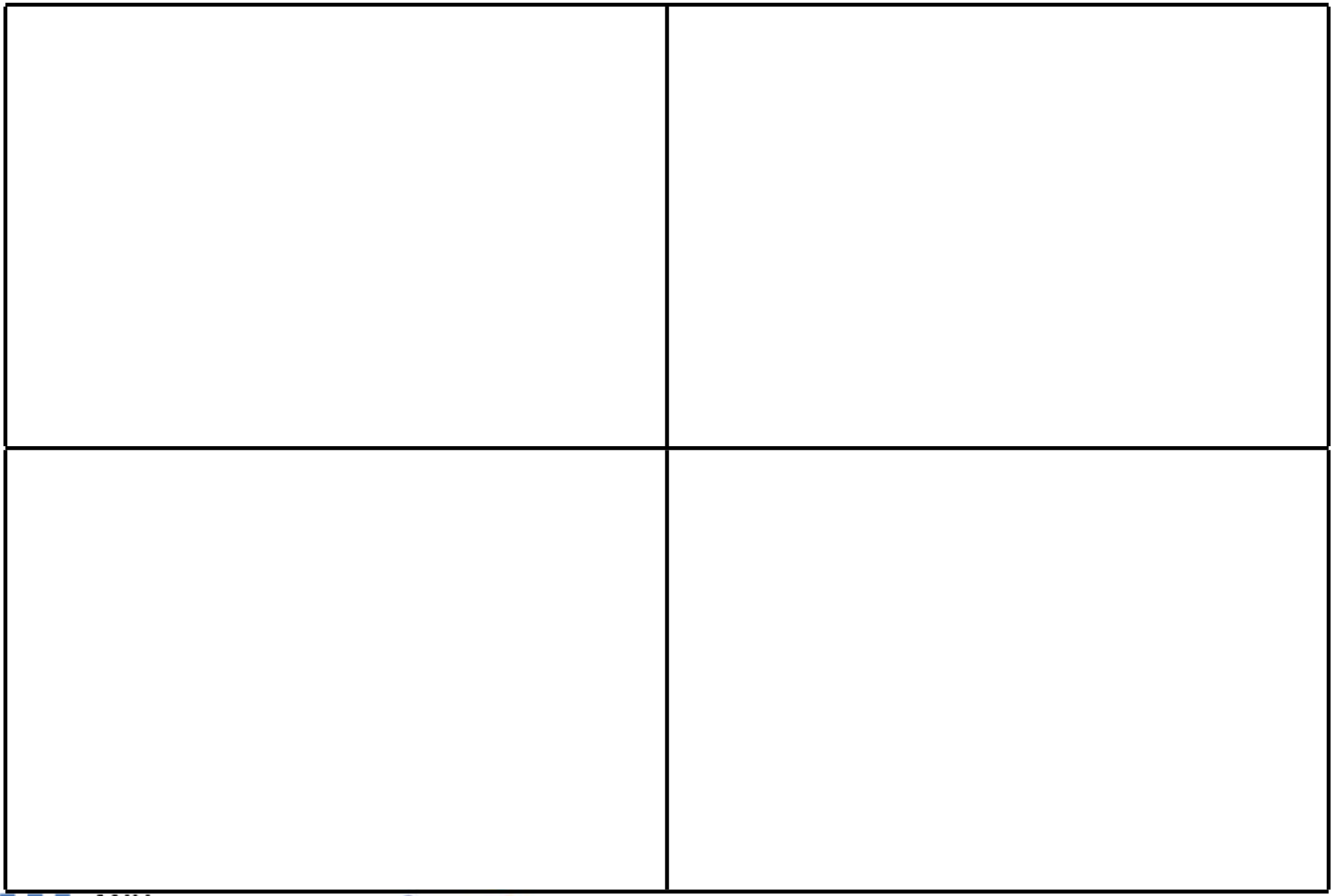


Social Science

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

Why Visualise? Ask Anscombe



Why Visualise? Ask Anscombe

$\overline{x_1} = 9.0$ $\mathrm{Var}(x_1) = 10.0$ $\overline{y_1} = 7.5$ $\mathrm{Var}(y_1) = 3.75$ $\mathrm{Corr}(x_1, y_1) = 0.816$ Regression $y_1 = 3 + 0.5x_1$	

Why Visualise? Ask Anscombe

$\overline{x_1} = 9.0$ $\mathrm{Var}(x_1) = 10.0$ $\overline{y_1} = 7.5$ $\mathrm{Var}(y_1) = 3.75$ $\mathrm{Corr}(x_1, y_1) = 0.816$ Regression $y_1 = 3 + 0.5x_1$	$\overline{x_2} = 9.0$ $\mathrm{Var}(x_2) = 10.0$ $\overline{y_2} = 7.5$ $\mathrm{Var}(y_2) = 3.75$ $\mathrm{Corr}(x_2, y_2) = 0.816$ Regression $y_2 = 3 + 0.5x_2$

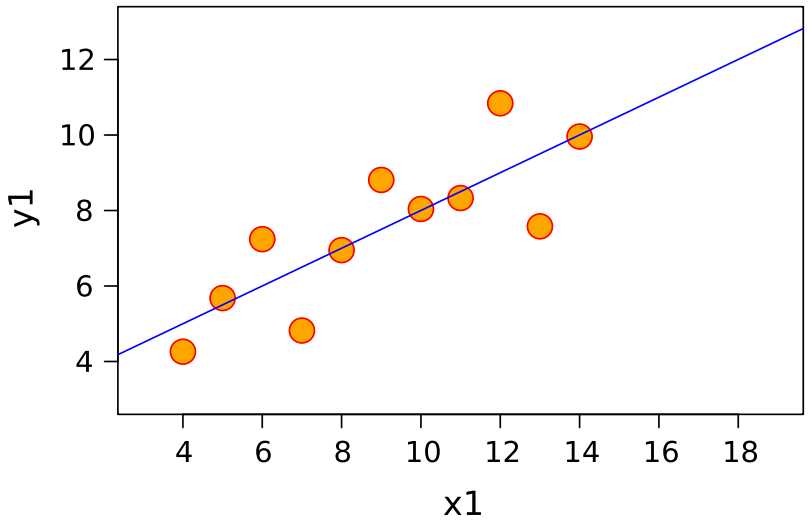
Why Visualise? Ask Anscombe

$\overline{x_1} = 9.0$ $\mathrm{Var}(x_1) = 10.0$ $\overline{y_1} = 7.5$ $\mathrm{Var}(y_1) = 3.75$ $\mathrm{Corr}(x_1, y_1) = 0.816$ Regression $y_1 = 3 + 0.5x_1$	$\overline{x_2} = 9.0$ $\mathrm{Var}(x_2) = 10.0$ $\overline{y_2} = 7.5$ $\mathrm{Var}(y_2) = 3.75$ $\mathrm{Corr}(x_2, y_2) = 0.816$ Regression $y_2 = 3 + 0.5x_2$
$\overline{x_3} = 9.0$ $\mathrm{Var}(x_3) = 10.0$ $\overline{y_3} = 7.5$ $\mathrm{Var}(y_3) = 3.75$ $\mathrm{Corr}(x_3, y_3) = 0.816$ Regression $y_3 = 3 + 0.5x_3$	

Why Visualise? Ask Anscombe

$\overline{x_1} = 9.0$ $\mathrm{Var}(x_1) = 10.0$ $\overline{y_1} = 7.5$ $\mathrm{Var}(y_1) = 3.75$ $\mathrm{Corr}(x_1, y_1) = 0.816$ Regression $y_1 = 3 + 0.5x_1$	$\overline{x_2} = 9.0$ $\mathrm{Var}(x_2) = 10.0$ $\overline{y_2} = 7.5$ $\mathrm{Var}(y_2) = 3.75$ $\mathrm{Corr}(x_2, y_2) = 0.816$ Regression $y_2 = 3 + 0.5x_2$
$\overline{x_3} = 9.0$ $\mathrm{Var}(x_3) = 10.0$ $\overline{y_3} = 7.5$ $\mathrm{Var}(y_3) = 3.75$ $\mathrm{Corr}(x_3, y_3) = 0.816$ Regression $y_3 = 3 + 0.5x_3$	$\overline{x_4} = 9.0$ $\mathrm{Var}(x_4) = 10.0$ $\overline{y_4} = 7.5$ $\mathrm{Var}(y_4) = 3.75$ $\mathrm{Corr}(x_4, y_4) = 0.816$ Regression $y_4 = 3 + 0.5x_4$

Why Visualise? Ask Anscombe

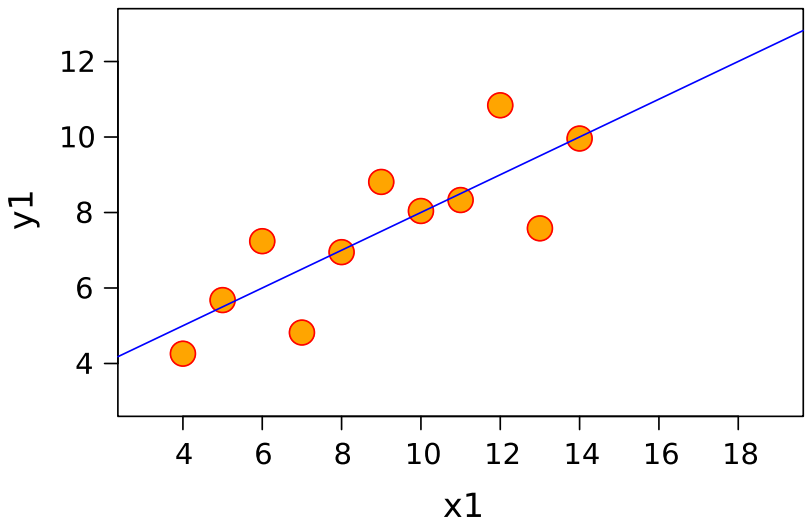


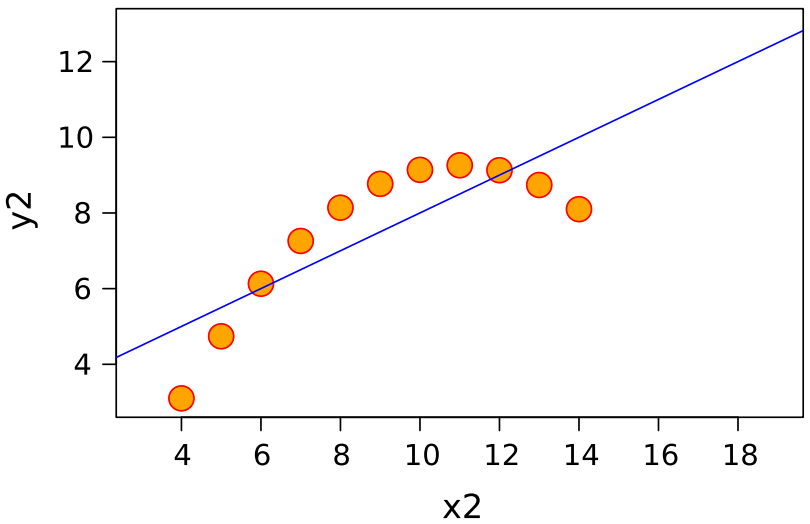
$$\overline{x_2} = 9.0$$
$$\mathrm{Var}(x_2) = 10.0$$
$$\overline{y_2} = 7.5$$
$$\mathrm{Var}(y_2) = 3.75$$
$$\mathrm{Corr}(x_2, y_2) = 0.816$$
$$\text{Regression } y_2 = 3 + 0.5x_2$$

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

$$\overline{x_4} = 9.0$$
$$\mathrm{Var}(x_4) = 10.0$$
$$\overline{y_4} = 7.5$$
$$\mathrm{Var}(y_4) = 3.75$$
$$\mathrm{Corr}(x_4, y_4) = 0.816$$
$$\text{Regression } y_4 = 3 + 0.5x_4$$

Why Visualise? Ask Anscombe

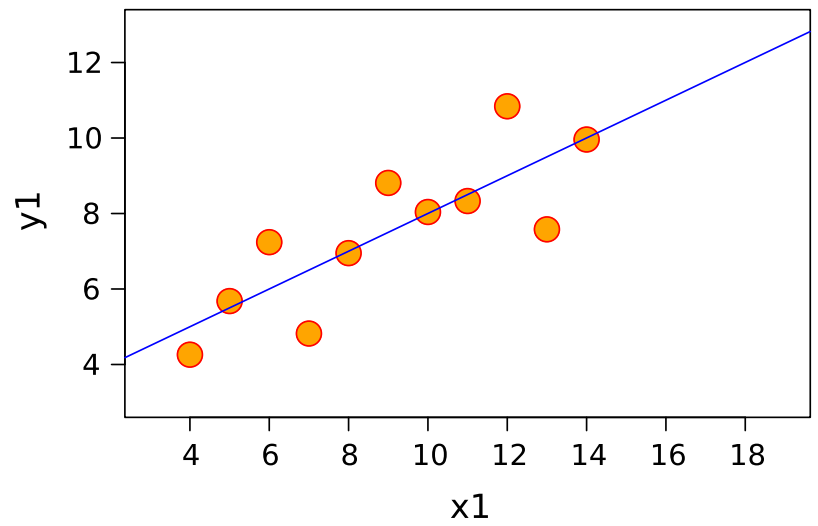


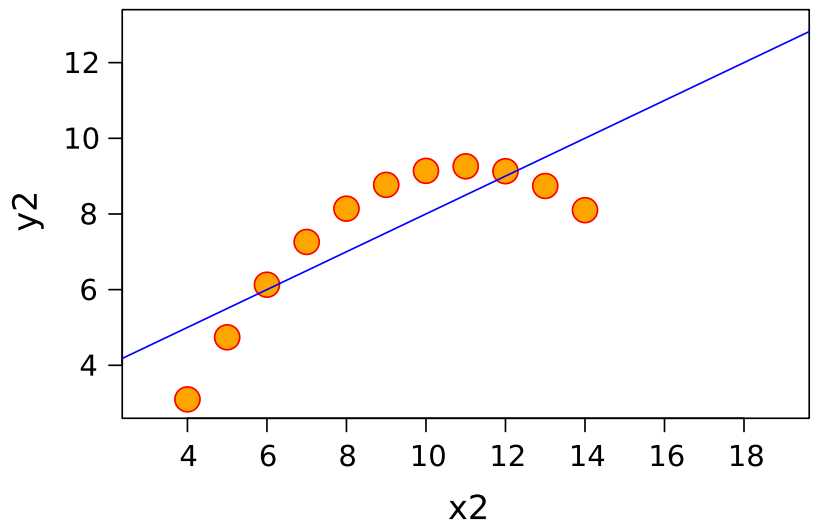


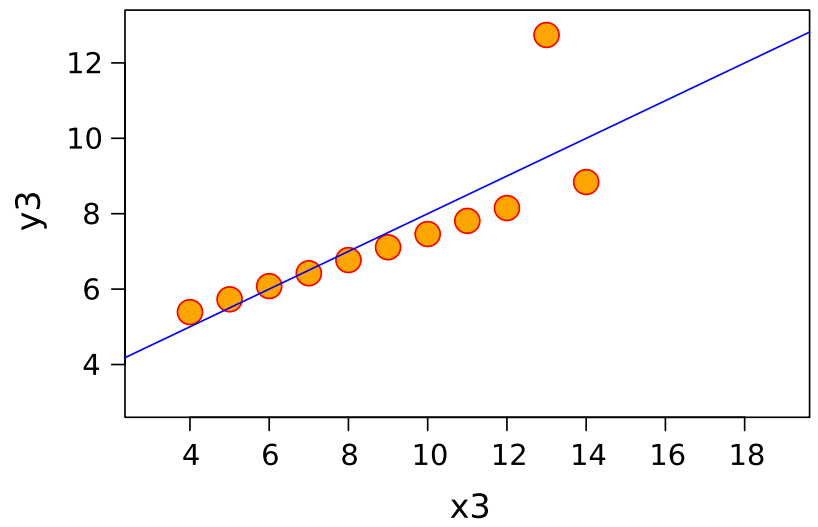
$$\overline{x_3} = 9.0$$
$$\mathrm{Var}(x_3) = 10.0$$
$$\overline{y_3} = 7.5$$
$$\mathrm{Var}(y_3) = 3.75$$
$$\mathrm{Corr}(x_3, y_3) = 0.816$$
$$\text{Regression } y_3 = 3 + 0.5x_3$$

$$\overline{x_4} = 9.0$$
$$\mathrm{Var}(x_4) = 10.0$$
$$\overline{y_4} = 7.5$$
$$\mathrm{Var}(y_4) = 3.75$$
$$\mathrm{Corr}(x_4, y_4) = 0.816$$
$$\text{Regression } y_4 = 3 + 0.5x_4$$

Why Visualise? Ask Anscombe

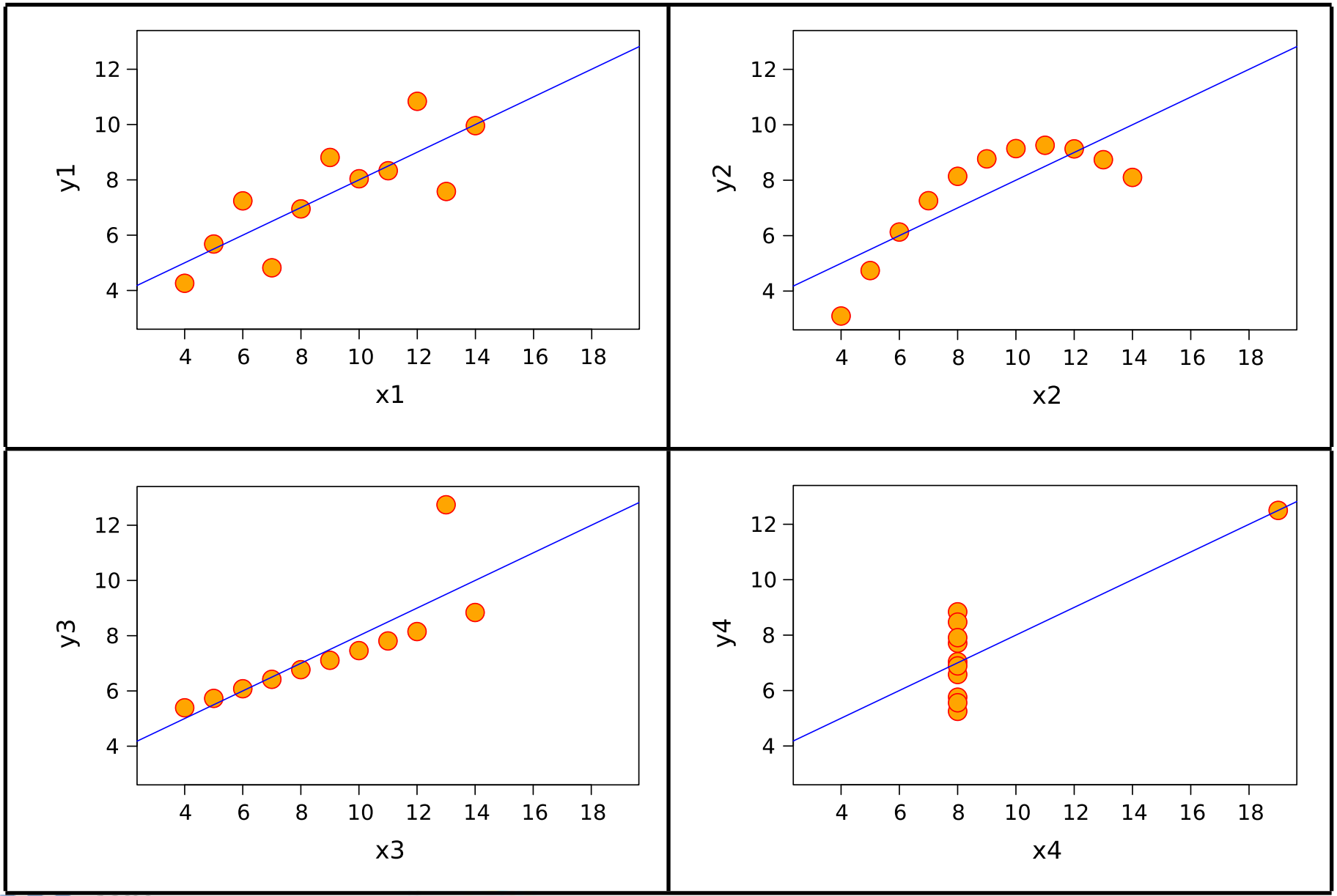






$$\overline{x_4} = 9.0$$
$$\mathrm{Var}(x_4) = 10.0$$
$$\overline{y_4} = 7.5$$
$$\mathrm{Var}(y_4) = 3.75$$
$$\mathrm{Corr}(x_4, y_4) = 0.816$$
$$\text{Regression } y_4 = 3 + 0.5x_4$$

Why Visualise? Ask Anscombe



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")

person	year	employment	study	relationship
1	2008	none	full-time	single
1	2009	part-time	part-time	cohabitating
1	2010	full-time	none	married
2	2008	full-time	part-time	cohabitating
2	2009	full-time	none	cohabitating
2	2010	full-time	none	single
⋮	⋮	⋮	⋮	⋮

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")

	person		
year	**1**	**2**	...
2008	n,f,s	f,p,c	...
2009	p,p,c	f,n,c	...
2010	f,n,m	f,n,s	...
⋮	⋮	⋮	⋱

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

Application to longitudinal dataset

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 respondants
  - Wave 2 (2000): 1768 respondants
  - Wave 3 (2003): 1192 respondants
  - Wave 4 (2006): 1138 respondants + 2000 new
- Noise added to address confidentiality
- http://lifecourse.anu.edu.au/

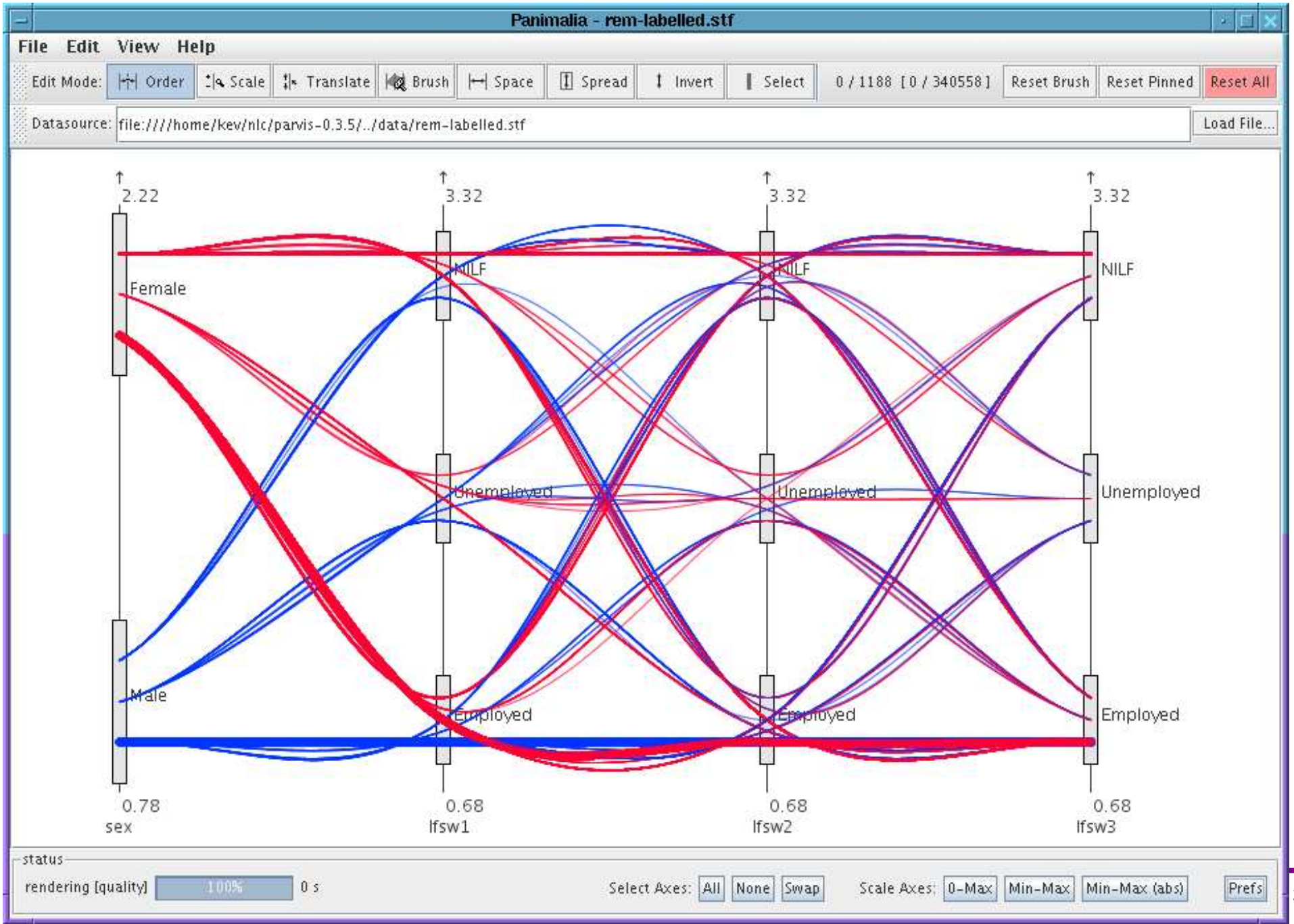
Pilot software tool

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

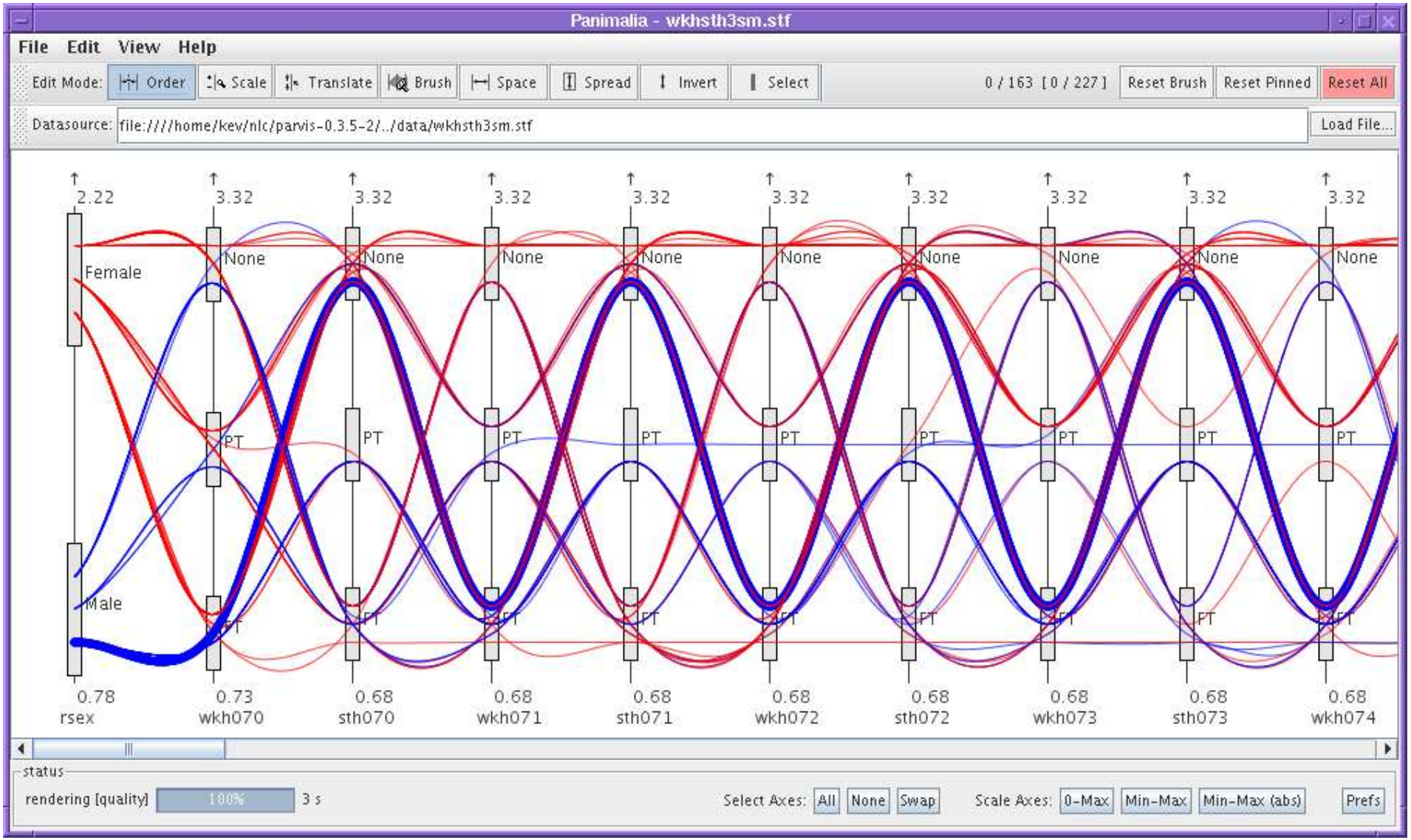
Labour force status

Labour force status

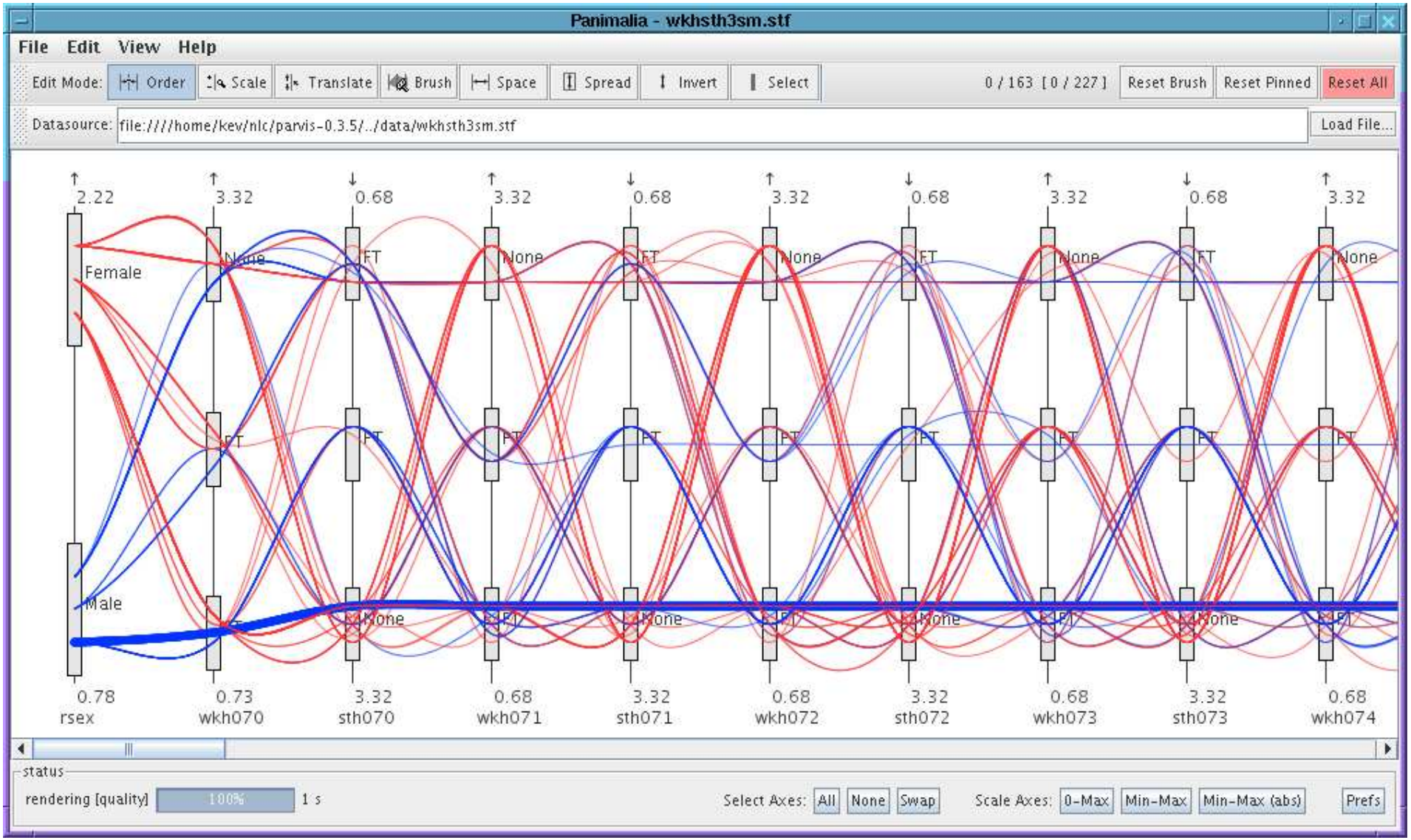


Work/study: interleaved by year

Work/study: interleaved by year

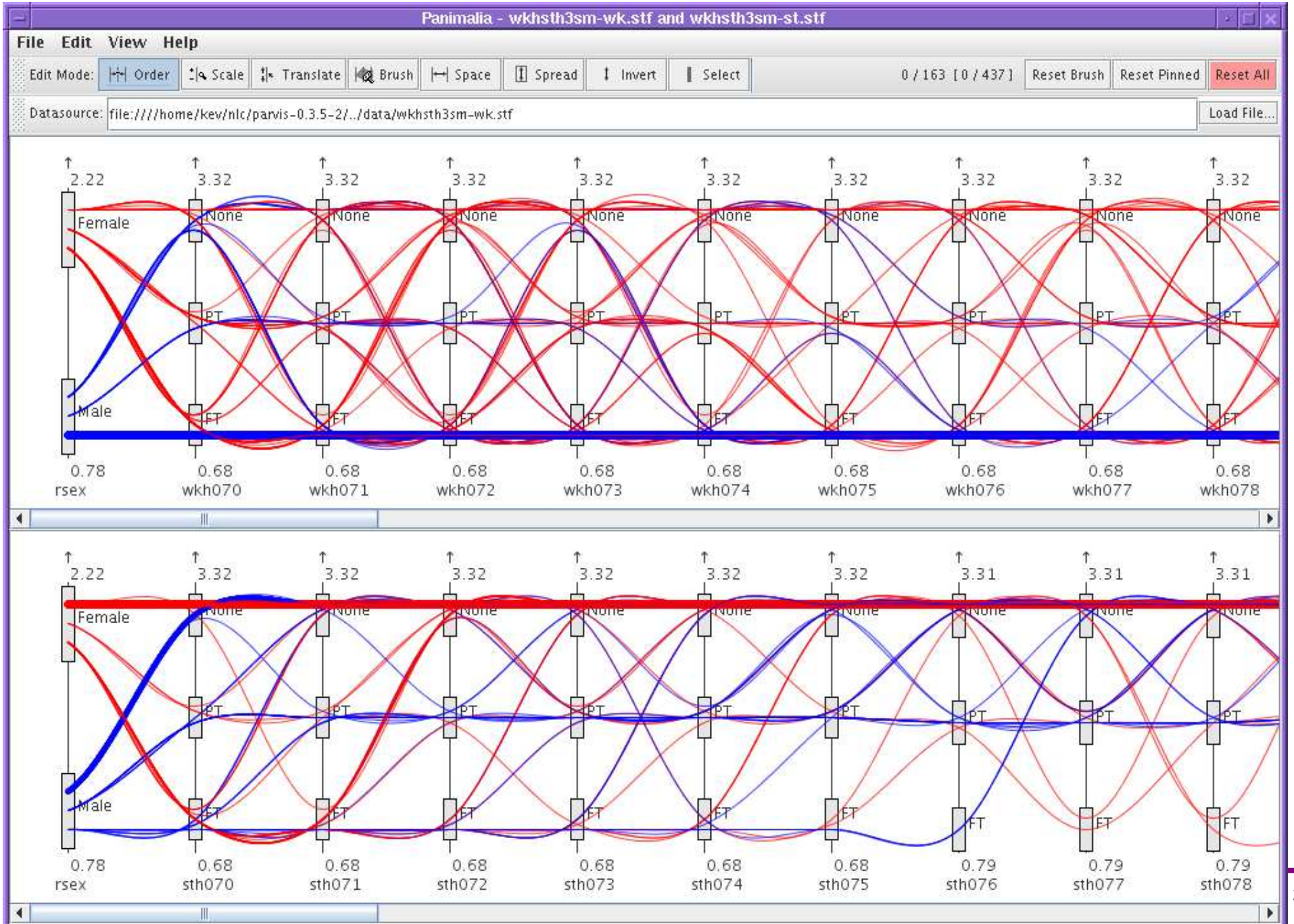


Work/study: interleaved by year



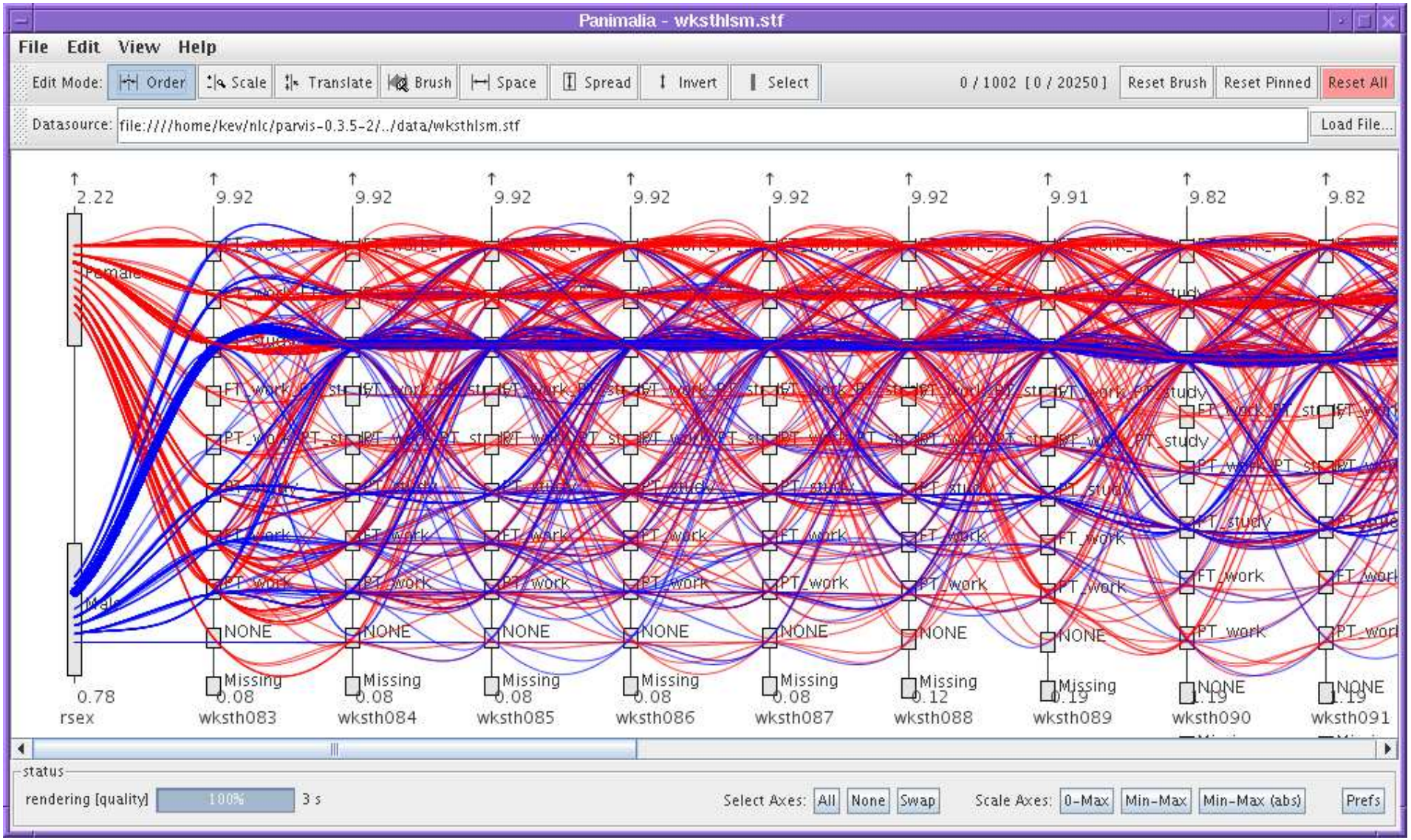
Work/study: separately by year

Work/study: separately by year



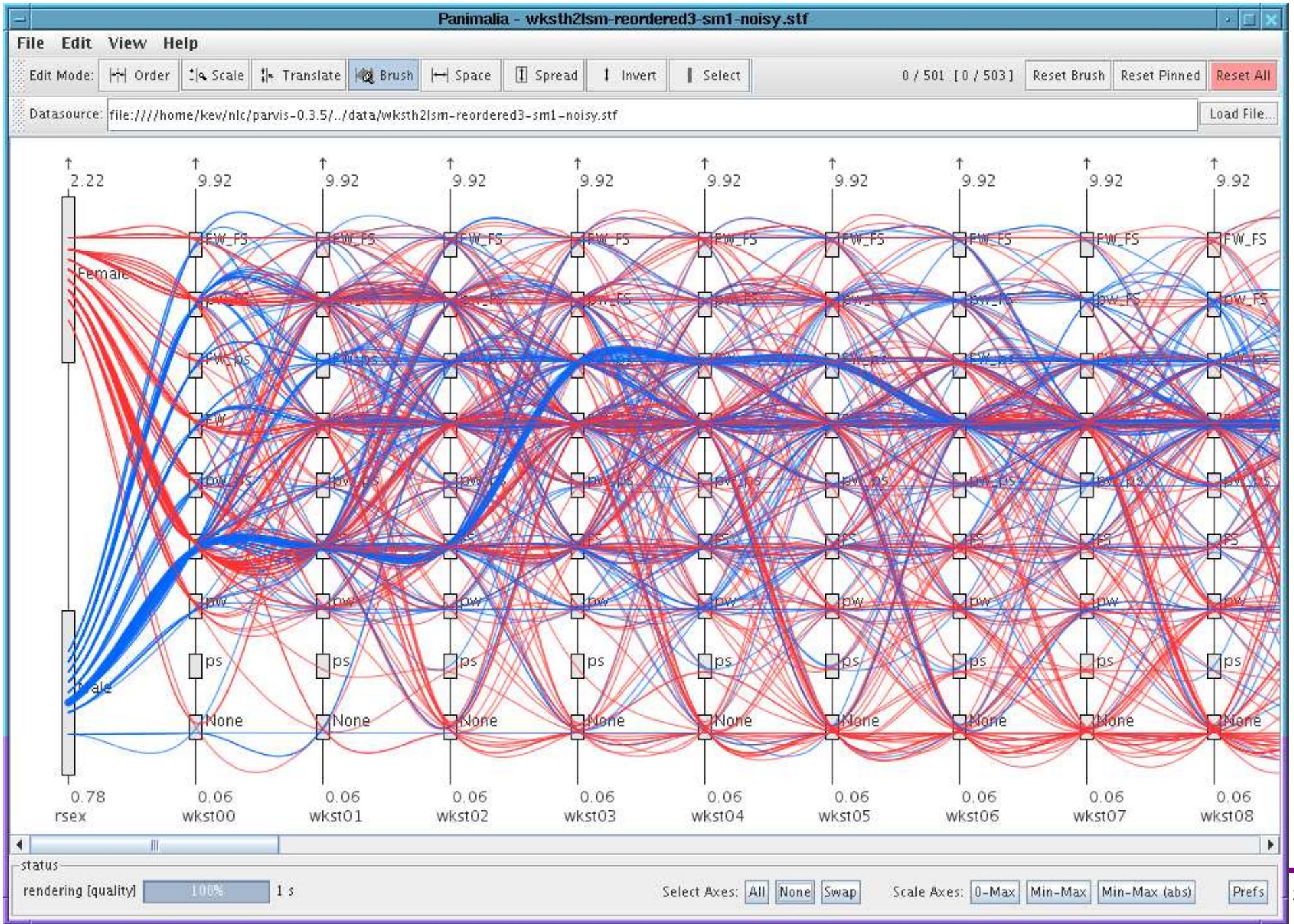
Work/study: combined by year

Work/study: combined by year

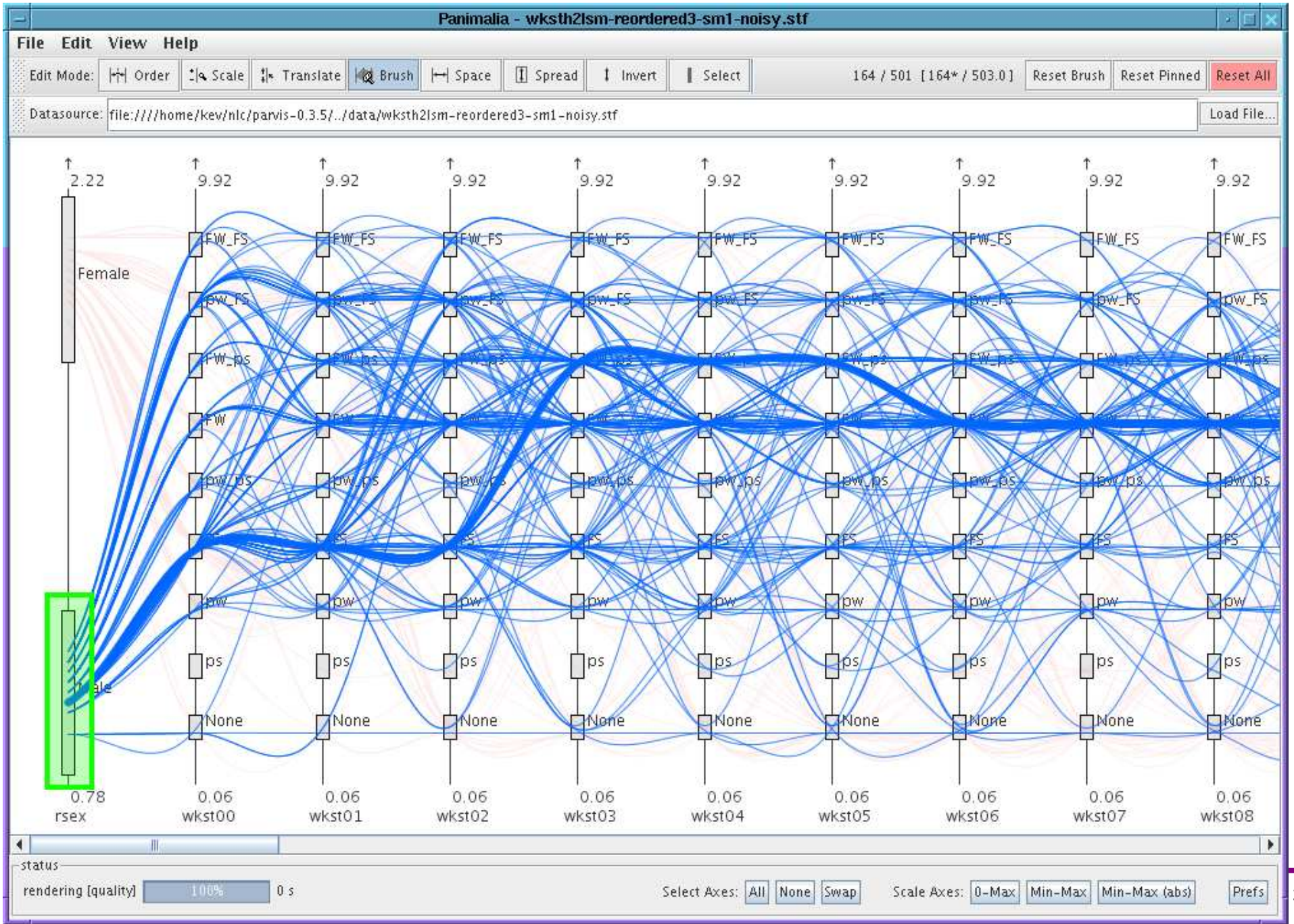


Work/study: combined by age

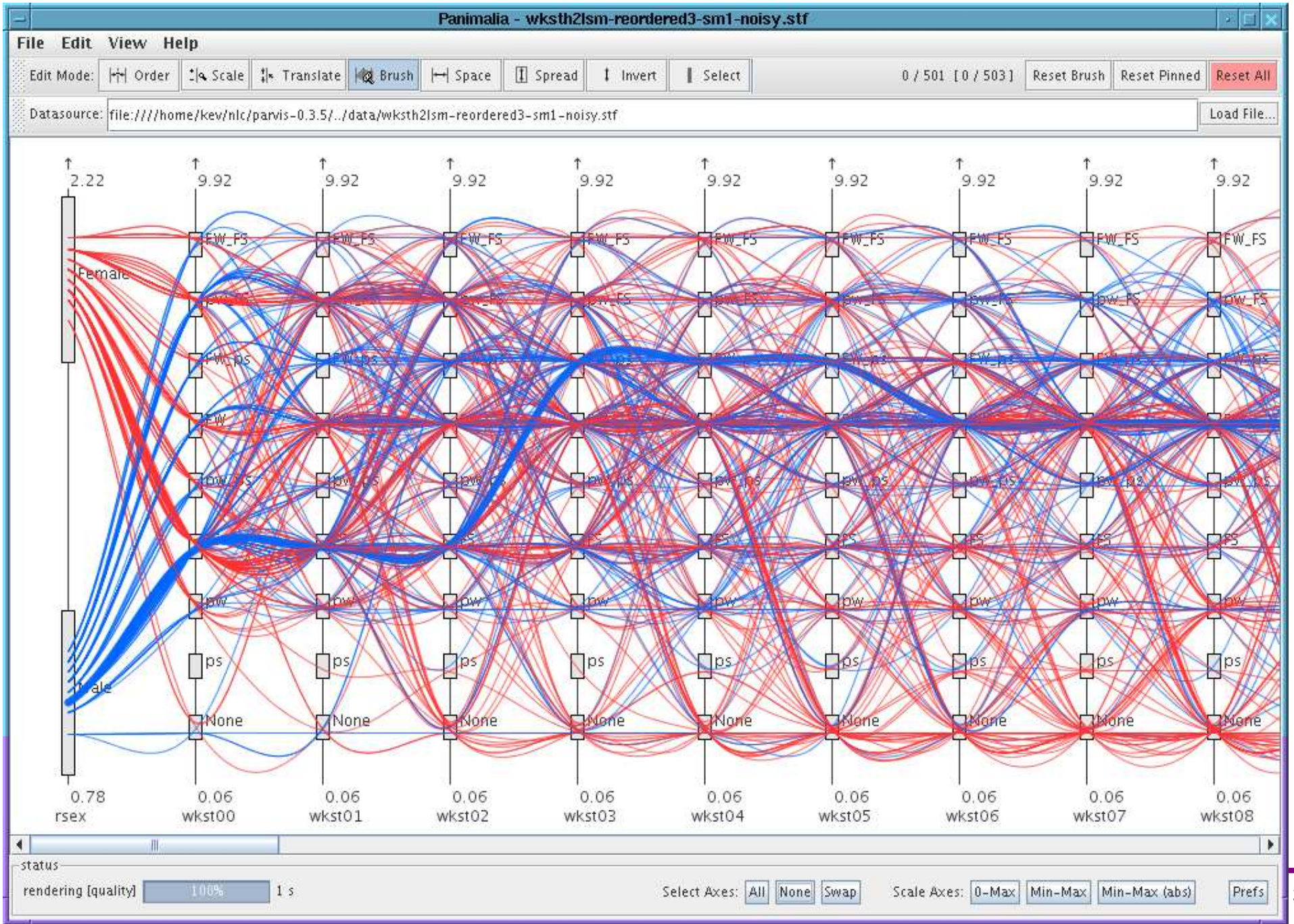
Work/study: combined by age



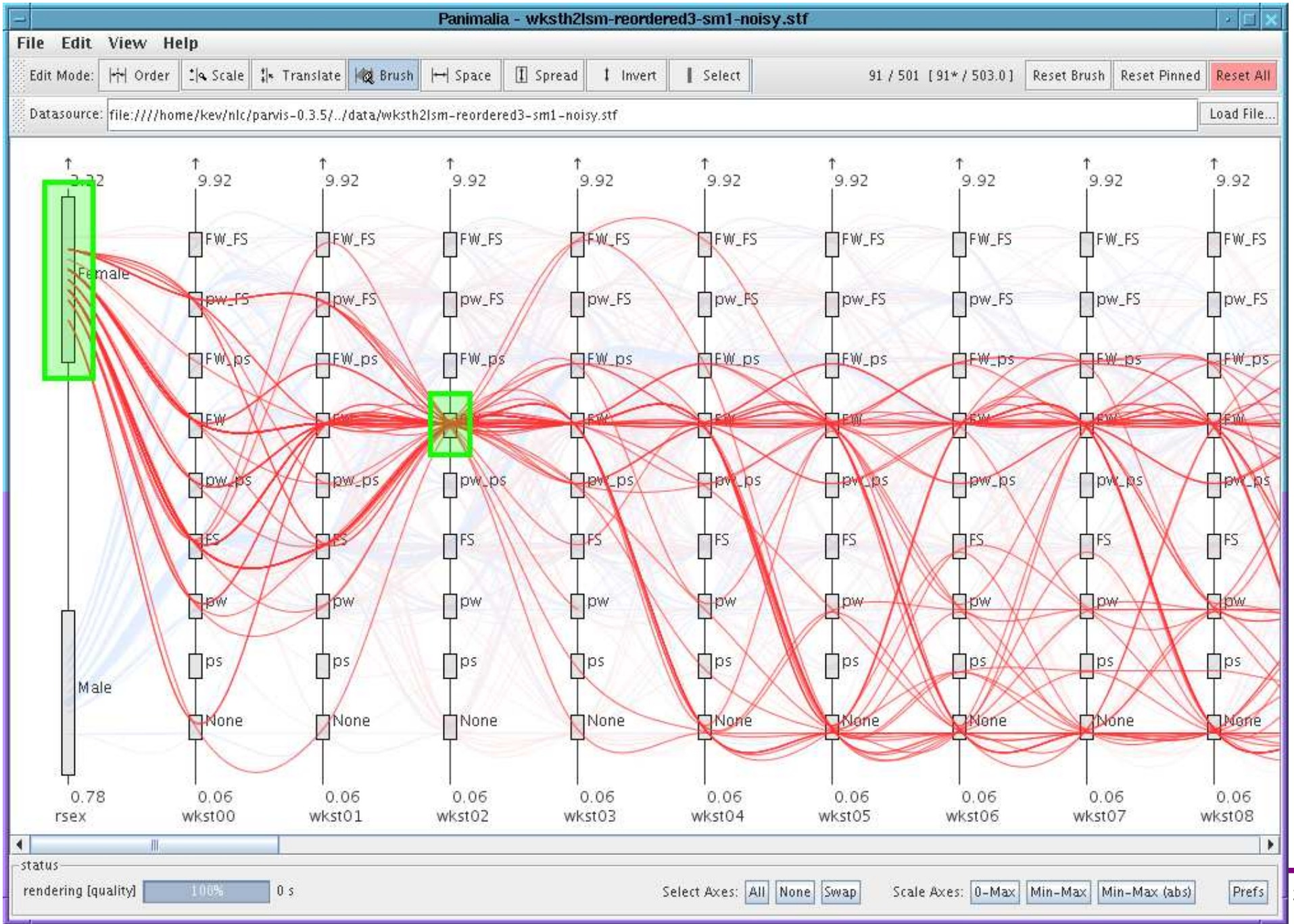
Work/study: combined by age



Work/study: combined by age

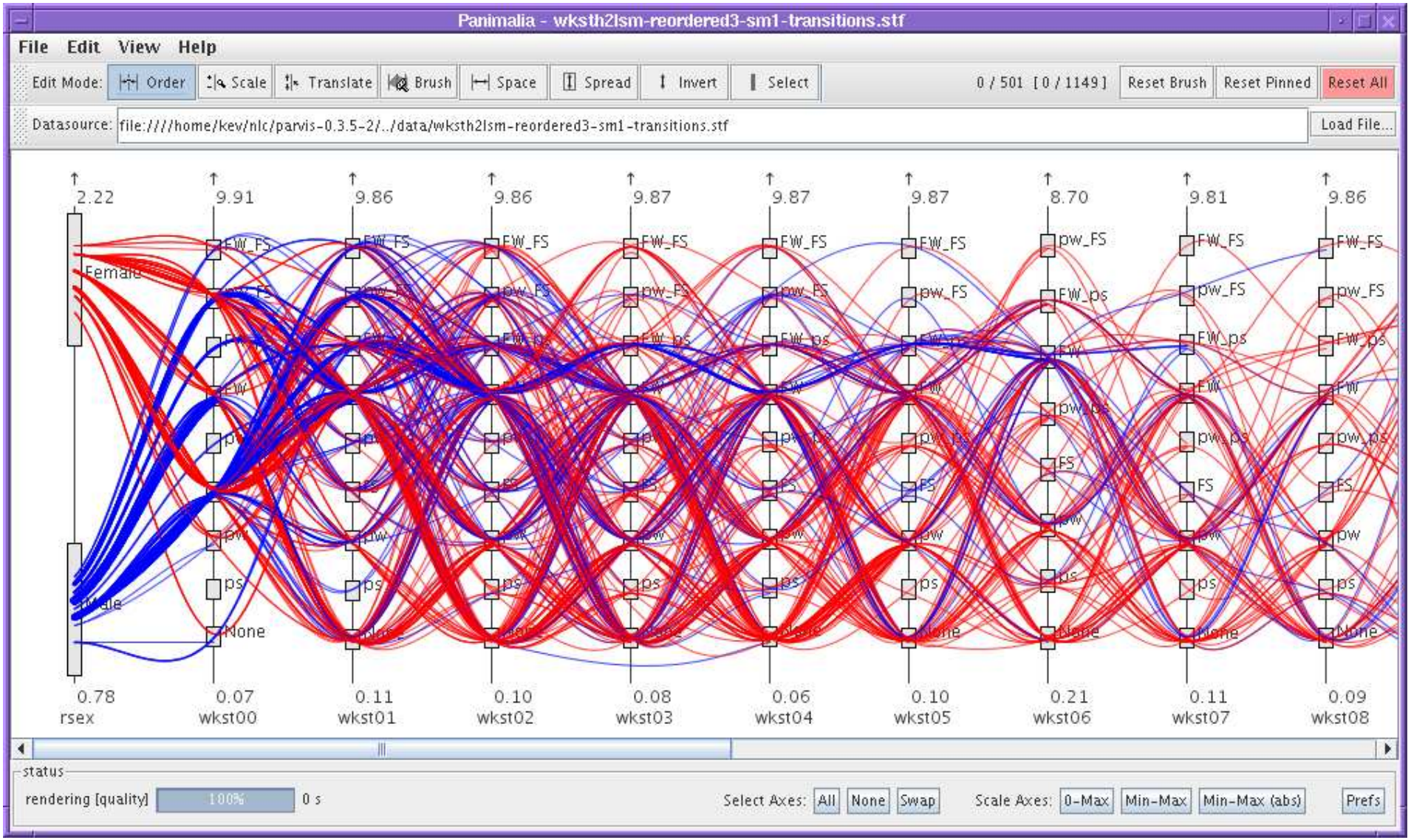


Work/study: combined by age



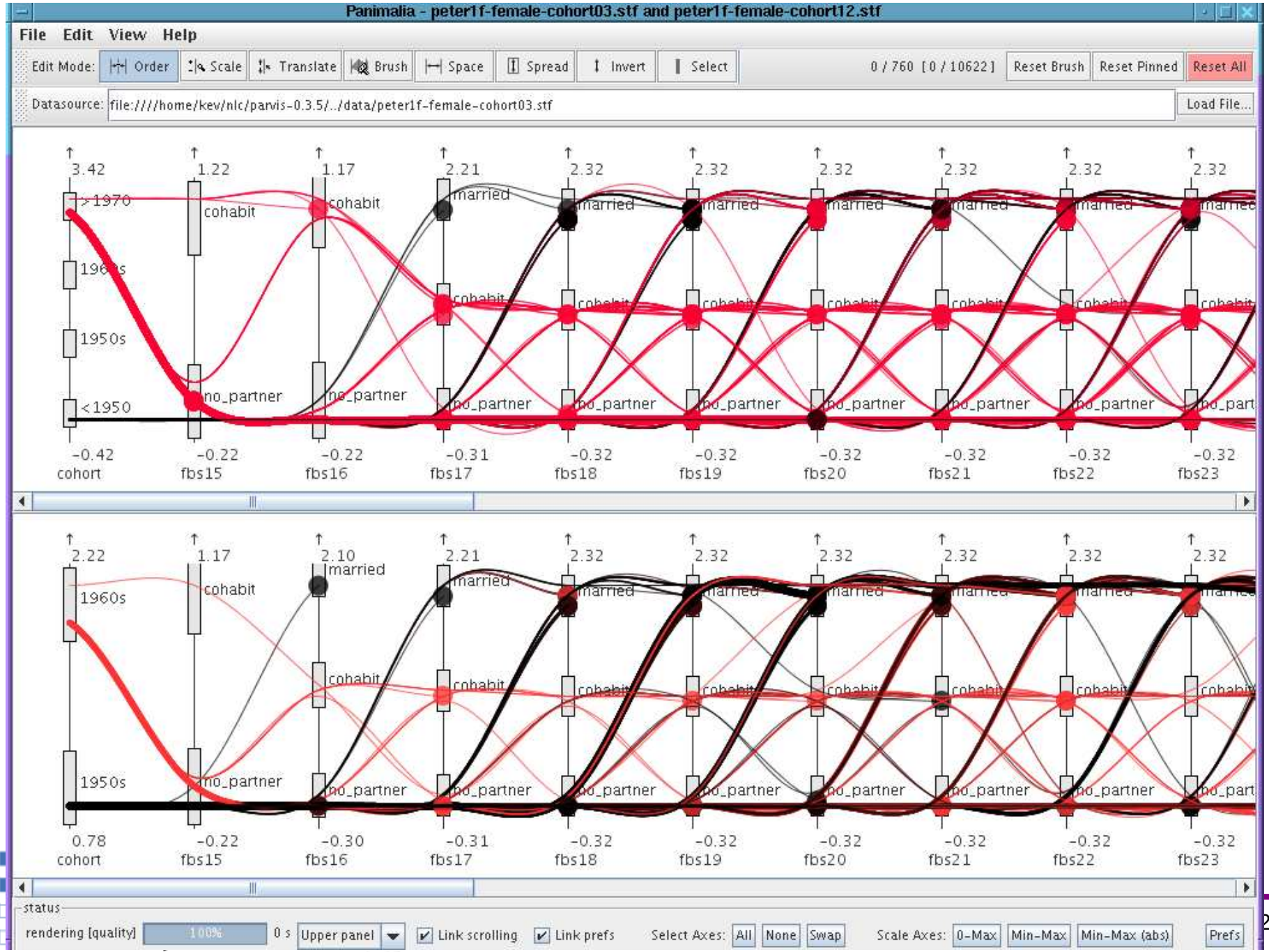
Work/study: combined transitions

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

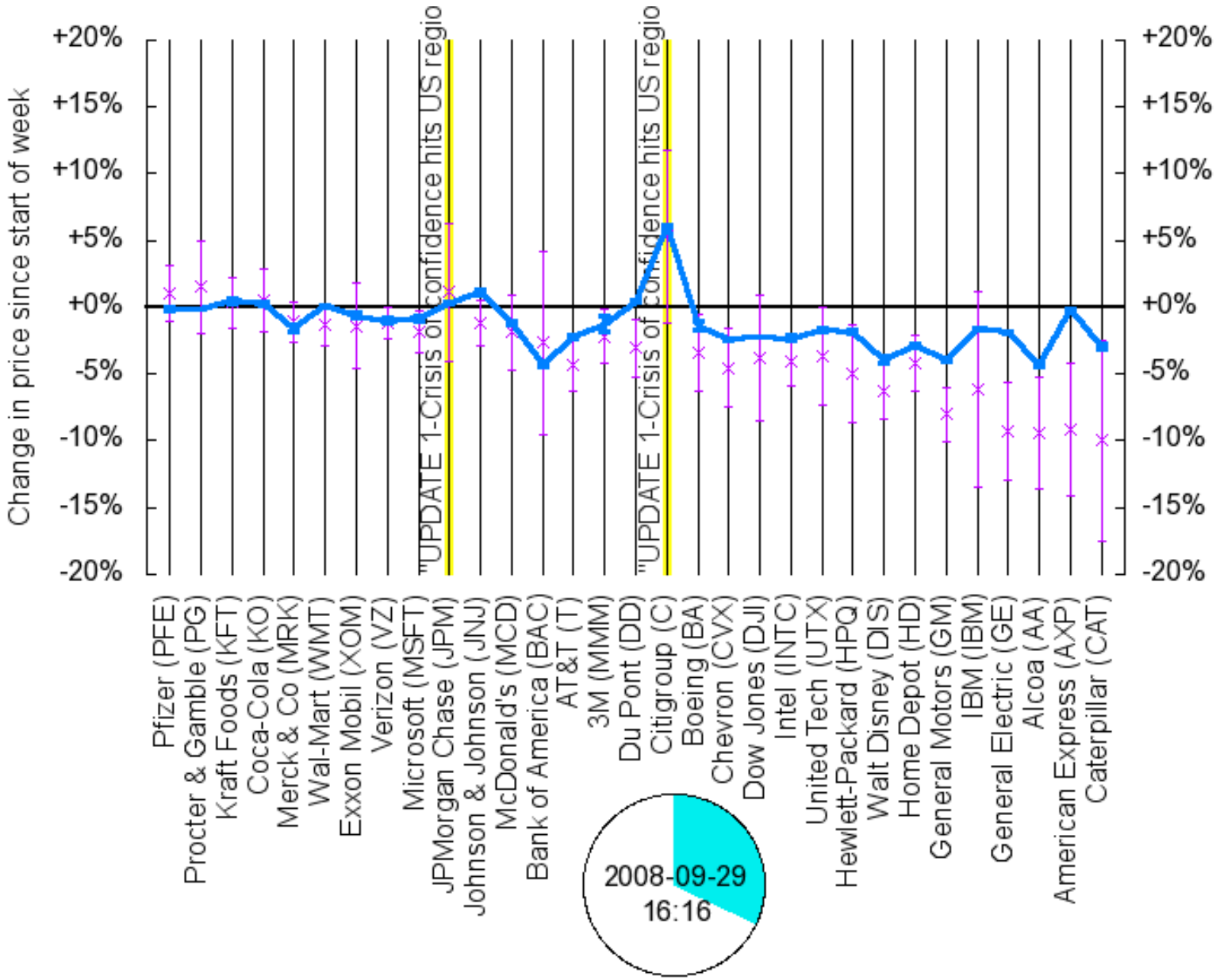
- 1Gb CSV file, sorted by stock/time
- 19,050,304 records (3 types):
  - Trades: 3,634,444 records
  - Quotes: 15,413,586 records
  - News: 2,274 records
- Each record has:
  - Stock name, date/time (milliseconds)
  - Trades: price/volume
  - Quotes: bid/ask, price/volume
  - News: headline (free text)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

Sample data

```
AA.N,29-SEP-2008,18:33:10.103,-4,Quote,,,21.1,21.14,
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-2008,18:33:10.985,-4,Quote,,,21.11,21.14,
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,
AA.N,29-SEP-2008,18:33:11.319,-4,News,,,,,"STOCKS NEWS→
    → US-Wall St routed as House rejects bailout bill"
AA.N,29-SEP-2008,18:33:11.585,-4,Quote,,,21.1,21.14,
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,
AA.N,29-SEP-2008,18:33:14.003,-4,Trade,21.1,800,,,
AA.N,29-SEP-2008,18:33:14.025,-4,Quote,,,21.09,21.1,
AA.N,29-SEP-2008,18:33:14.114,-4,Quote,,,21.09,21.1,
```

Dow Jones Animated Multiverse

Dow Jones Animated Multiverse



Result: Second place

- *[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!