
Direct visualization of longitudinal data

Kevin Pulo

kevin.pulo@anu.edu.au

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

2010-06-02



Social Science and Visualisation

- Social science tends to be “undervisualised”
 - Analytic 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

$\bar{x} = 9.0$ $\text{Var}(x) = 10.0$ $\bar{y} = 7.5$ $\text{Var}(y) = 3.75$ $\text{Corr}(x, y) = 0.816$ $\text{Regression : } y = 3 + 0.5x$	

Why Visualise? Ask Anscombe

$\bar{x} = 9.0$ $\text{Var}(x) = 10.0$ $\bar{y} = 7.5$ $\text{Var}(y) = 3.75$ $\text{Corr}(x, y) = 0.816$ $\text{Regression : } y = 3 + 0.5x$	$\bar{x} = 9.0$ $\text{Var}(x) = 10.0$ $\bar{y} = 7.5$ $\text{Var}(y) = 3.75$ $\text{Corr}(x, y) = 0.816$ $\text{Regression : } y = 3 + 0.5x$

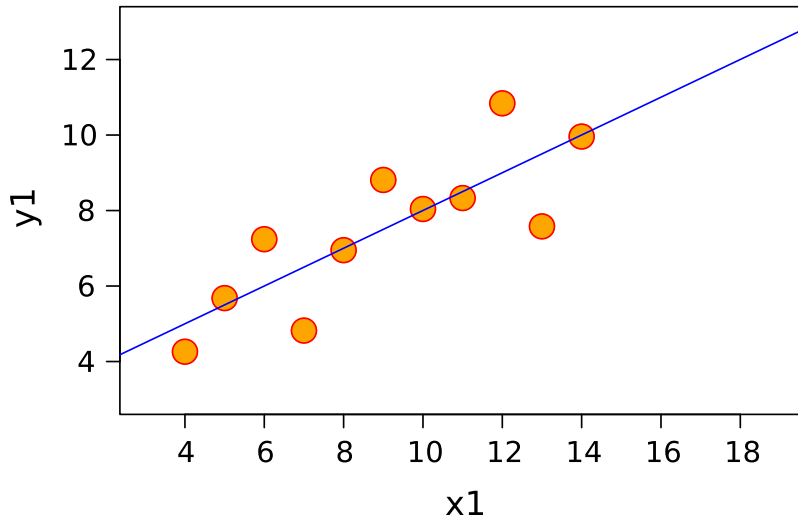
Why Visualise? Ask Anscombe

$\bar{x} = 9.0$ $\text{Var}(x) = 10.0$ $\bar{y} = 7.5$ $\text{Var}(y) = 3.75$ $\text{Corr}(x, y) = 0.816$ $\text{Regression : } y = 3 + 0.5x$	$\bar{x} = 9.0$ $\text{Var}(x) = 10.0$ $\bar{y} = 7.5$ $\text{Var}(y) = 3.75$ $\text{Corr}(x, y) = 0.816$ $\text{Regression : } y = 3 + 0.5x$
$\bar{x} = 9.0$ $\text{Var}(x) = 10.0$ $\bar{y} = 7.5$ $\text{Var}(y) = 3.75$ $\text{Corr}(x, y) = 0.816$ $\text{Regression : } y = 3 + 0.5x$	

Why Visualise? Ask Anscombe

$\bar{x} = 9.0$ $\text{Var}(x) = 10.0$ $\bar{y} = 7.5$ $\text{Var}(y) = 3.75$ $\text{Corr}(x, y) = 0.816$ $\text{Regression : } y = 3 + 0.5x$	$\bar{x} = 9.0$ $\text{Var}(x) = 10.0$ $\bar{y} = 7.5$ $\text{Var}(y) = 3.75$ $\text{Corr}(x, y) = 0.816$ $\text{Regression : } y = 3 + 0.5x$
$\bar{x} = 9.0$ $\text{Var}(x) = 10.0$ $\bar{y} = 7.5$ $\text{Var}(y) = 3.75$ $\text{Corr}(x, y) = 0.816$ $\text{Regression : } y = 3 + 0.5x$	$\bar{x} = 9.0$ $\text{Var}(x) = 10.0$ $\bar{y} = 7.5$ $\text{Var}(y) = 3.75$ $\text{Corr}(x, y) = 0.816$ $\text{Regression : } y = 3 + 0.5x$

Why Visualise? Ask Anscombe

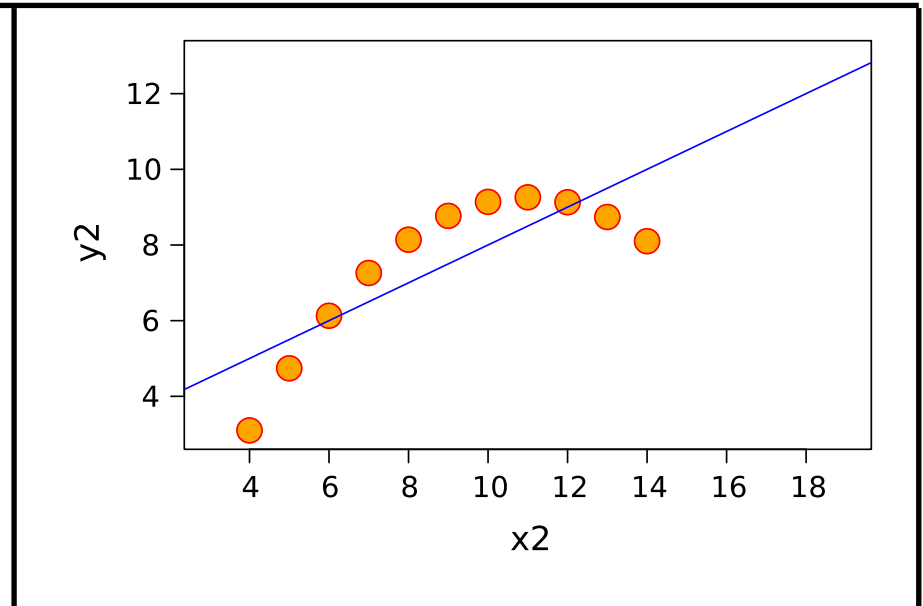
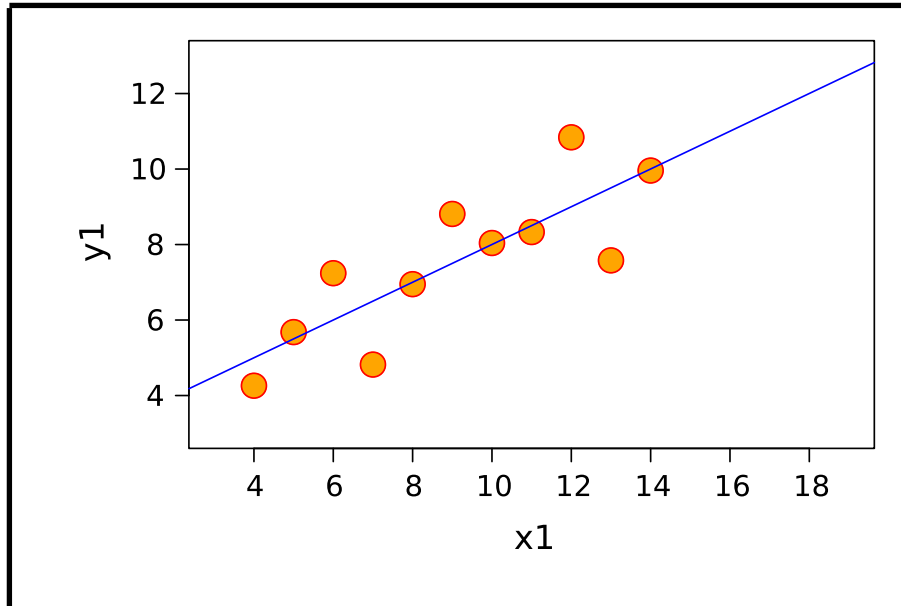


$$\begin{aligned}\bar{x} &= 9.0 \\ \text{Var}(x) &= 10.0 \\ \bar{y} &= 7.5 \\ \text{Var}(y) &= 3.75 \\ \text{Corr}(x, y) &= 0.816 \\ \text{Regression : } y &= 3 + 0.5x\end{aligned}$$

$$\begin{aligned}\bar{x} &= 9.0 \\ \text{Var}(x) &= 10.0 \\ \bar{y} &= 7.5 \\ \text{Var}(y) &= 3.75 \\ \text{Corr}(x, y) &= 0.816 \\ \text{Regression : } y &= 3 + 0.5x\end{aligned}$$

$$\begin{aligned}\bar{x} &= 9.0 \\ \text{Var}(x) &= 10.0 \\ \bar{y} &= 7.5 \\ \text{Var}(y) &= 3.75 \\ \text{Corr}(x, y) &= 0.816 \\ \text{Regression : } y &= 3 + 0.5x\end{aligned}$$

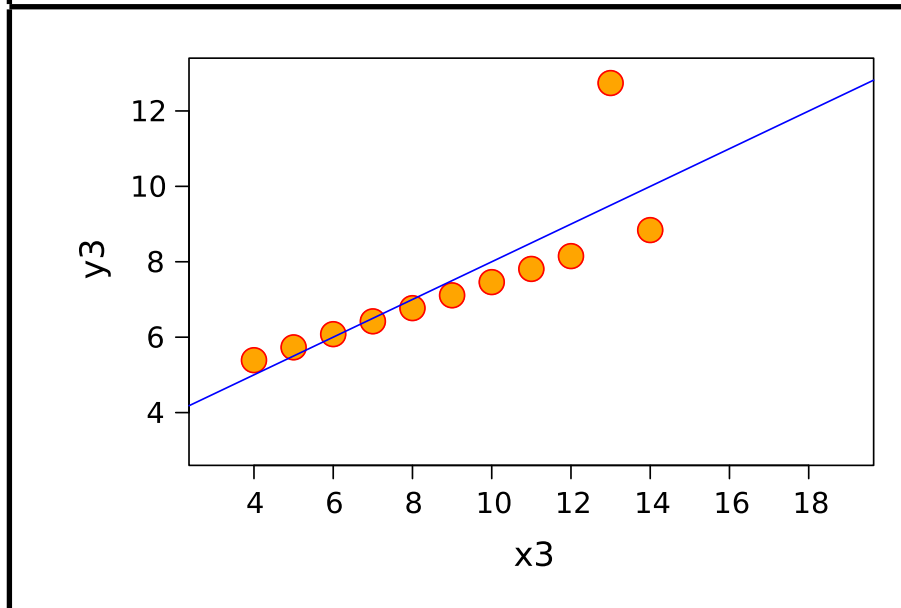
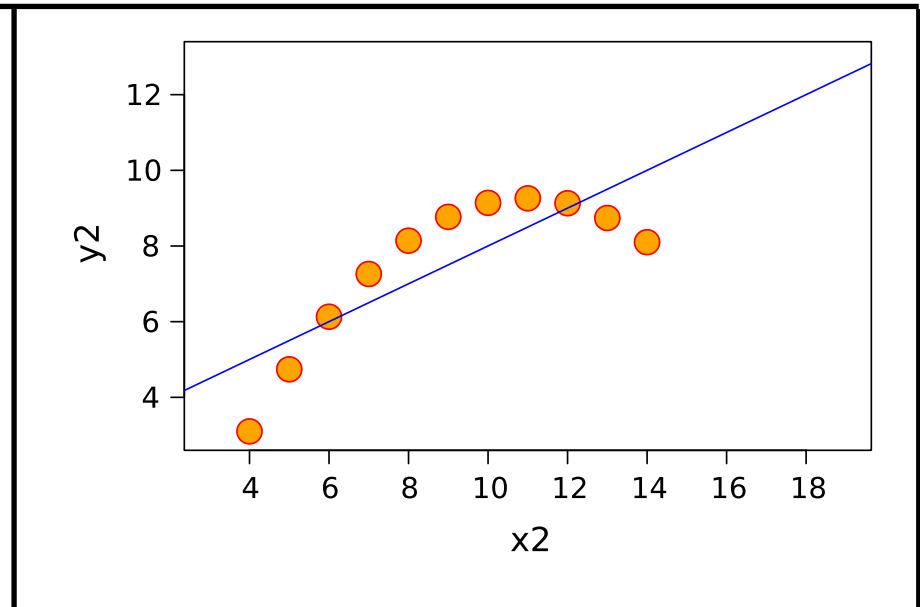
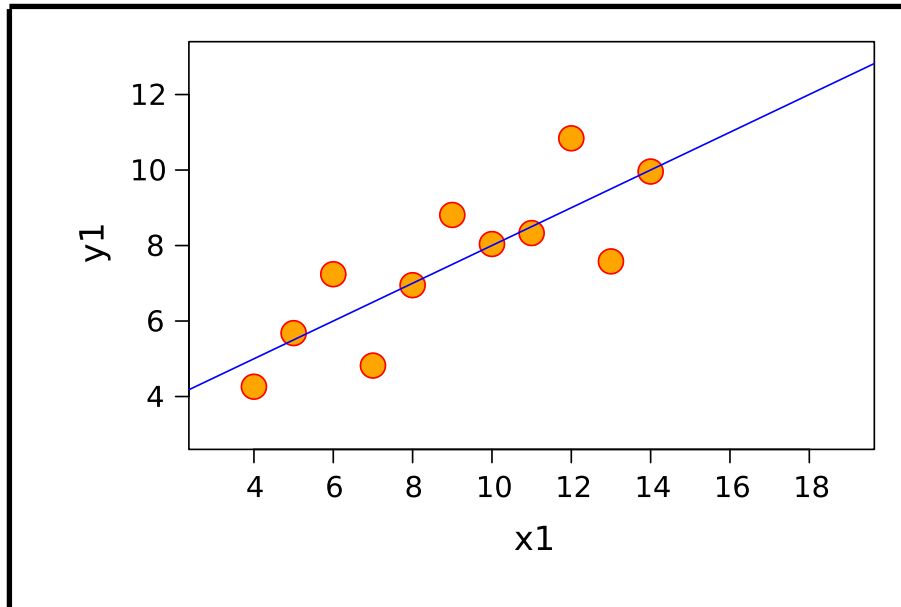
Why Visualise? Ask Anscombe



$\bar{x} = 9.0$
 $\text{Var}(x) = 10.0$
 $\bar{y} = 7.5$
 $\text{Var}(y) = 3.75$
 $\text{Corr}(x, y) = 0.816$
Regression : $y = 3 + 0.5x$

$\bar{x} = 9.0$
 $\text{Var}(x) = 10.0$
 $\bar{y} = 7.5$
 $\text{Var}(y) = 3.75$
 $\text{Corr}(x, y) = 0.816$
Regression : $y = 3 + 0.5x$

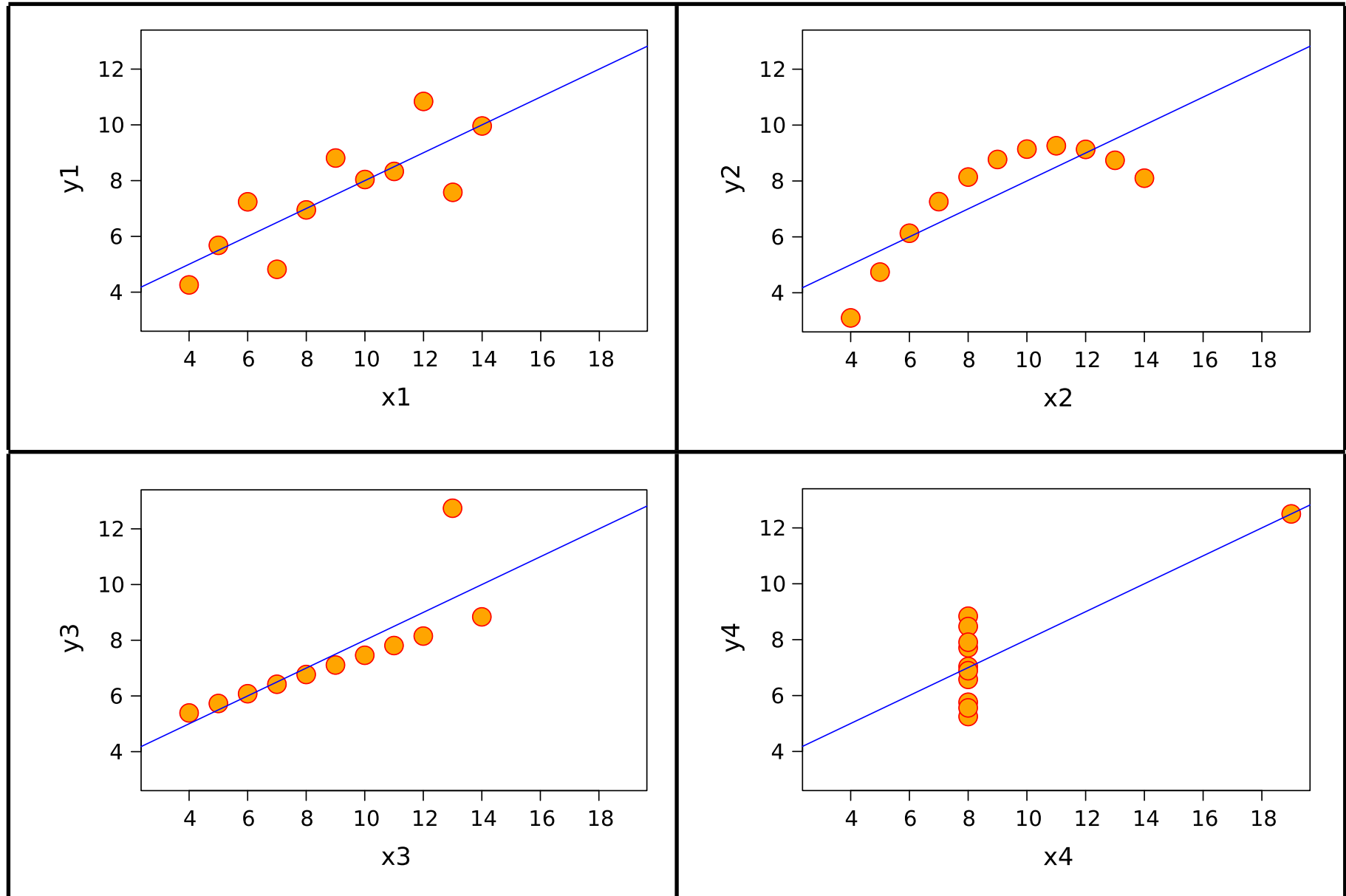
Why Visualise? Ask Anscombe



$\bar{x} = 9.0$
 $\text{Var}(x) = 10.0$
 $\bar{y} = 7.5$
 $\text{Var}(y) = 3.75$
 $\text{Corr}(x, y) = 0.816$
Regression : $y = 3 + 0.5x$



Why Visualise? Ask Anscombe



Project goals

- Longitudinal/panel survey data
- Direct visualisation
- Support interactive exploration
- Main use cases:
 - Initial familiarisation
 - Researchers looking for interesting features
 - ▷ via integration with Australian Social Science Data Archive (ASSDA) website
 - Archivists performing data cleaning
 - ▷ via desktop application

Longitudinal visualisation

- But the data contains:
 - Thousands of respondents, answering
 - Hundreds of questions, over
 - Multiple waves

Longitudinal visualisation

- But the data contains:
 - Thousands of respondents, answering
 - Hundreds of questions, over
 - Multiple waves
- The challenge:

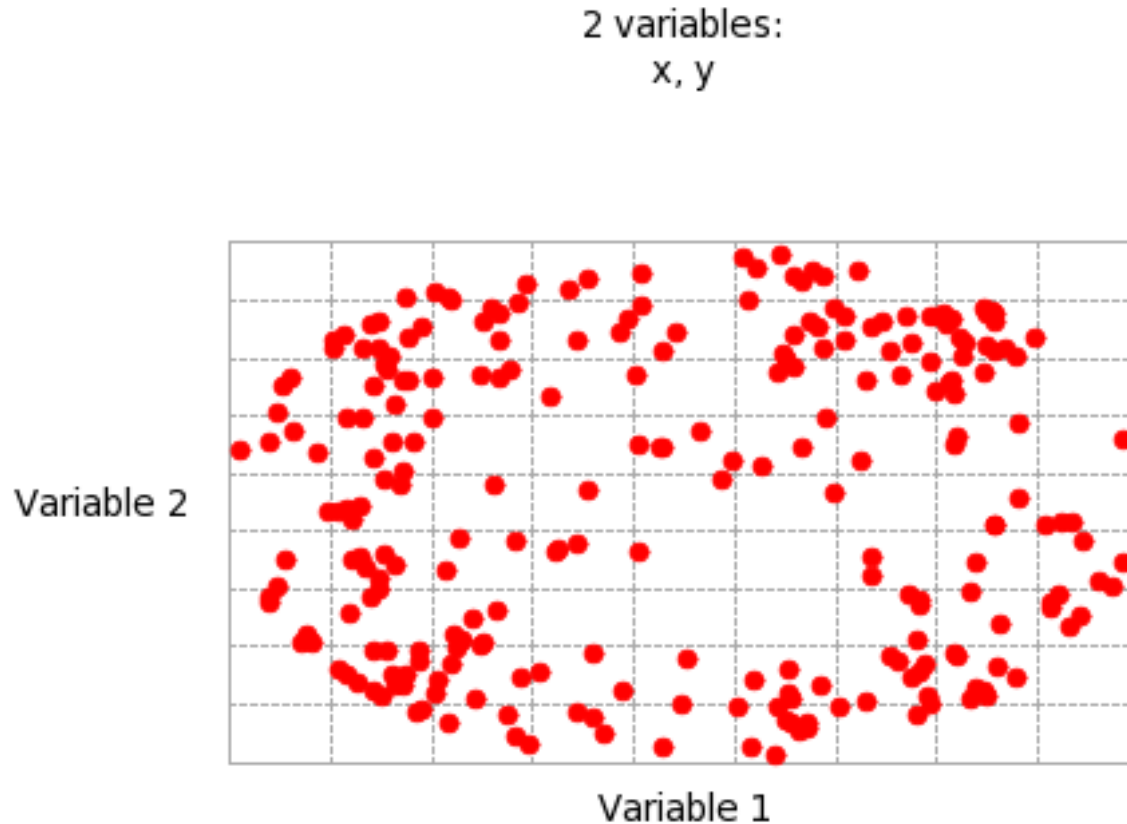
How to visualise this data in a useful way?

Mapping data to space

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

Mapping data to space

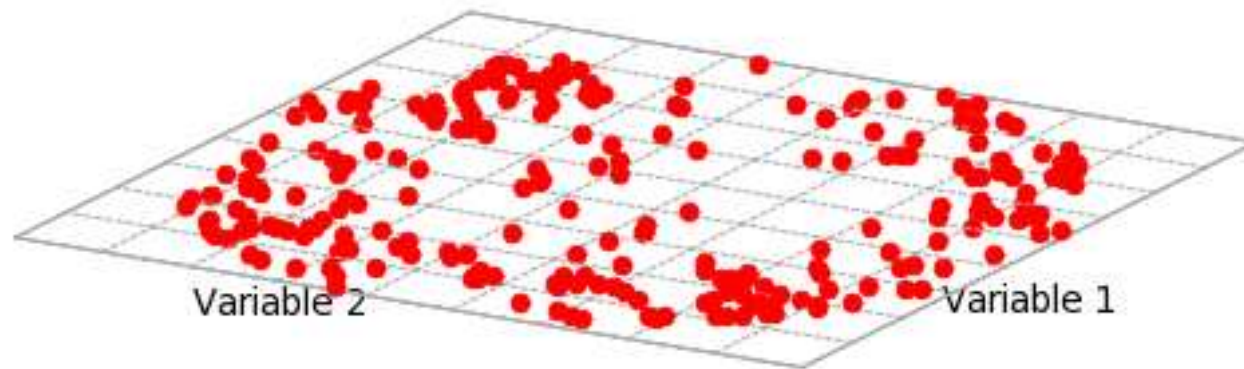
- Graphs map variables *directly* to 2D/3D space



Mapping data to space

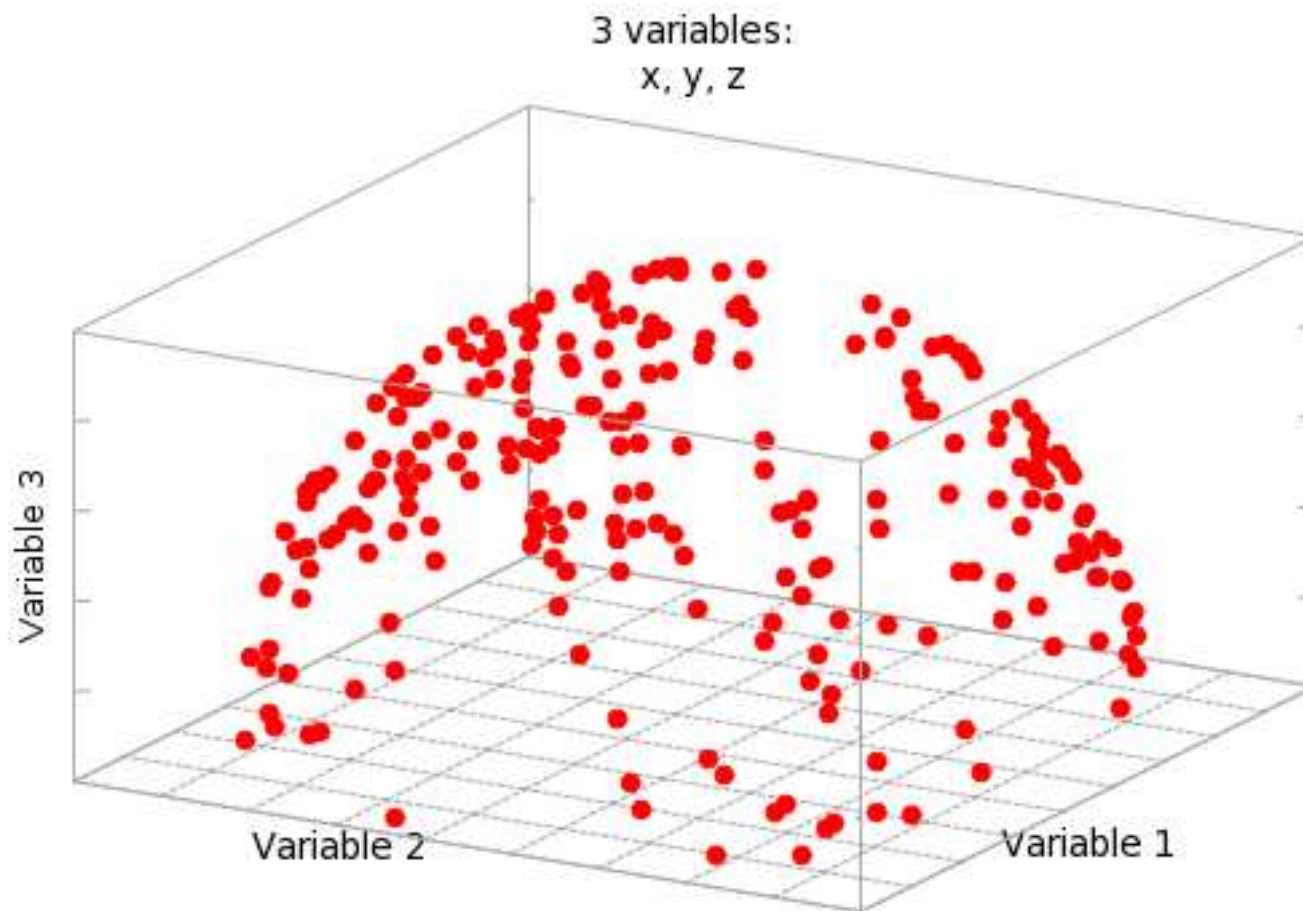
- Graphs map variables *directly* to 2D/3D space

2 variables:
x, y



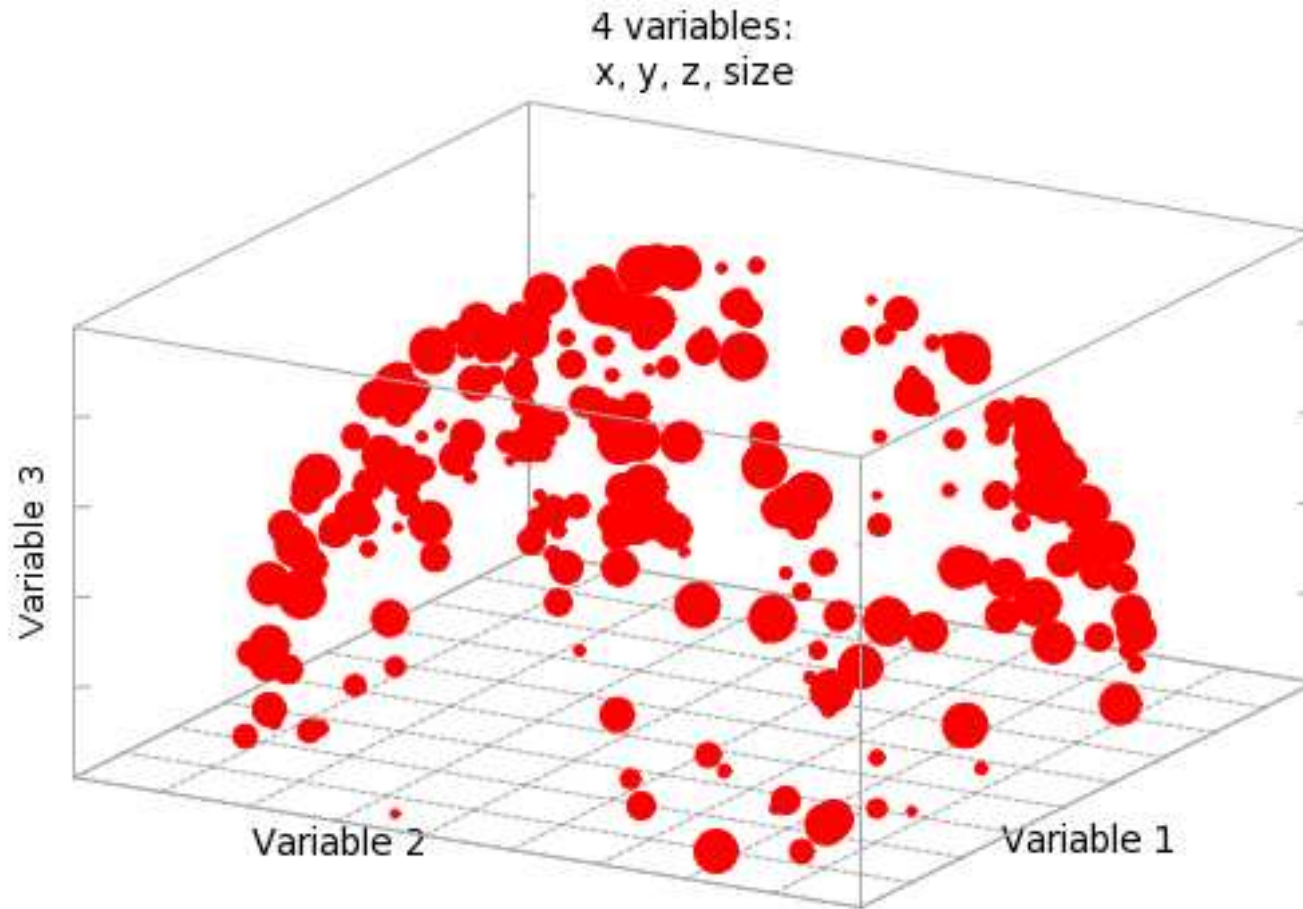
Mapping data to space

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



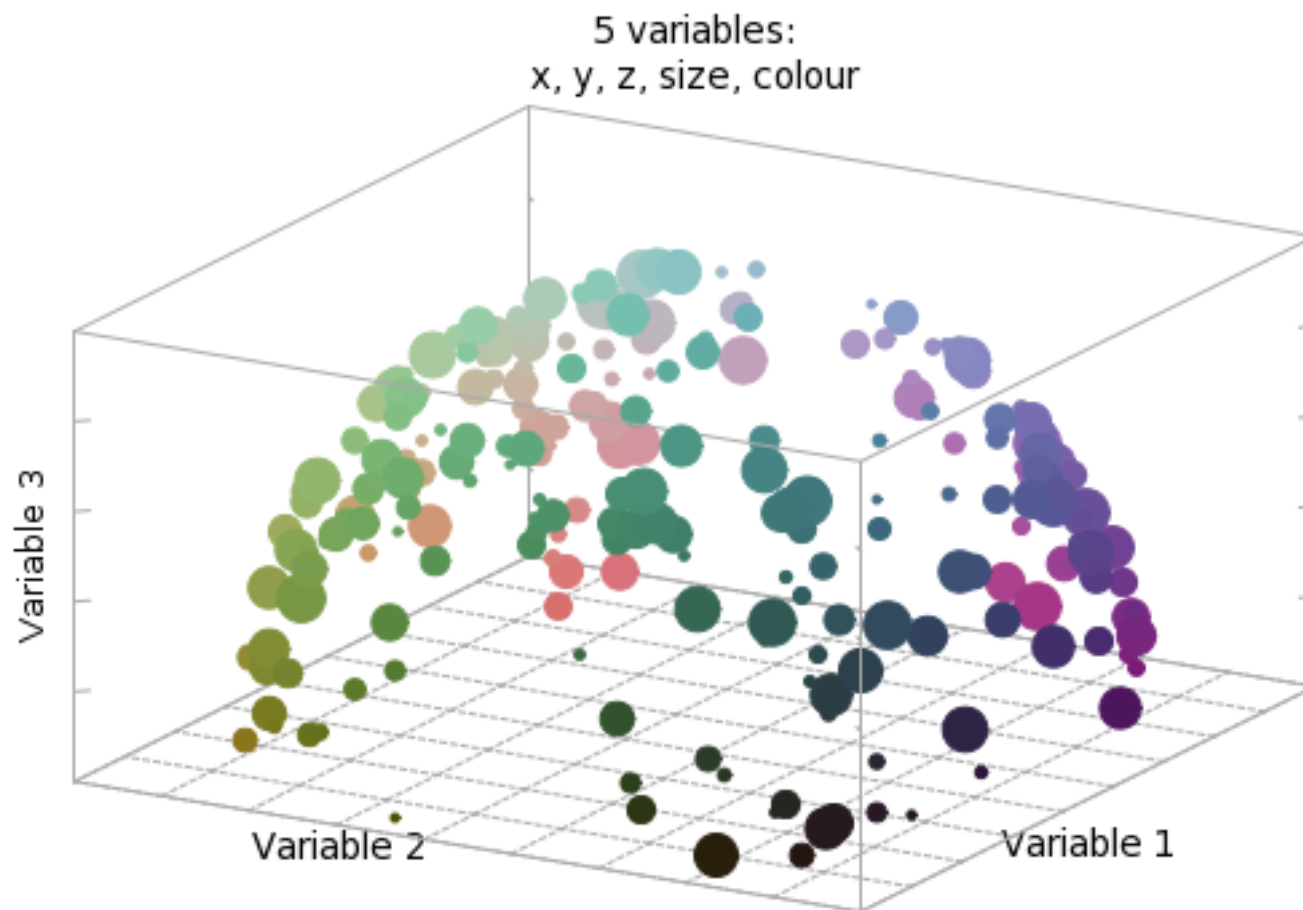
Mapping data to space

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



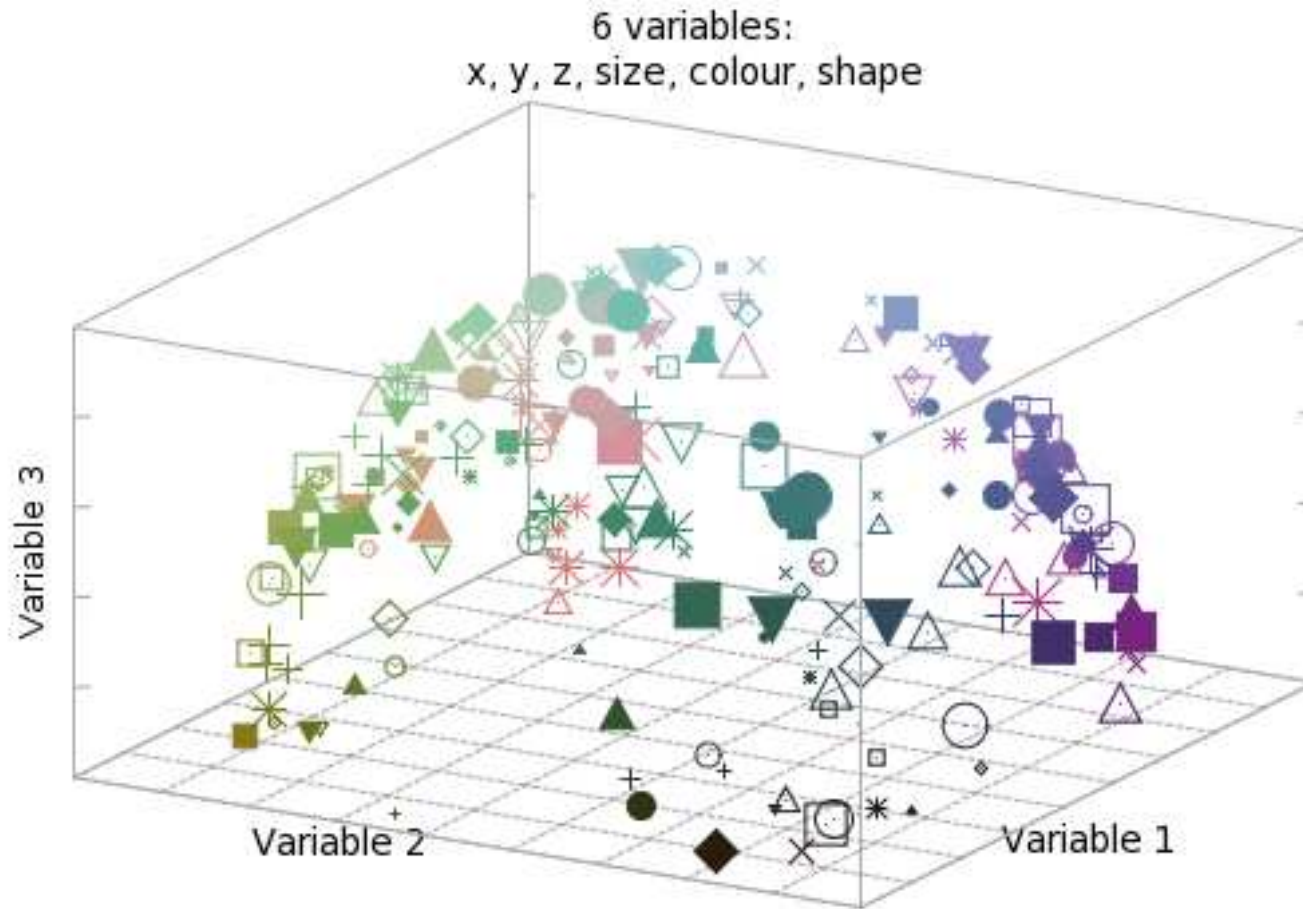
Mapping data to space

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



Mapping data to space

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



Visualising multi-dimensional data

- But longitudinal surveys have hundreds of questions (ie. hundreds of variables)

Visualising multi-dimensional data

- But longitudinal surveys have hundreds of questions (ie. hundreds of variables)
- Cannot possibly map this directly to hundred-dimensional space

Visualising multi-dimensional data

- But longitudinal surveys have hundreds of questions (ie. hundreds of variables)
- Cannot possibly map this directly to hundred-dimensional space
- Need some way of mapping this high-dimensional data to low-dimensional space

Visualising multi-dimensional data

- But longitudinal surveys have hundreds of questions (ie. hundreds of variables)
- Cannot possibly map this directly to hundred-dimensional space
- Need some way of mapping this high-dimensional data to low-dimensional space

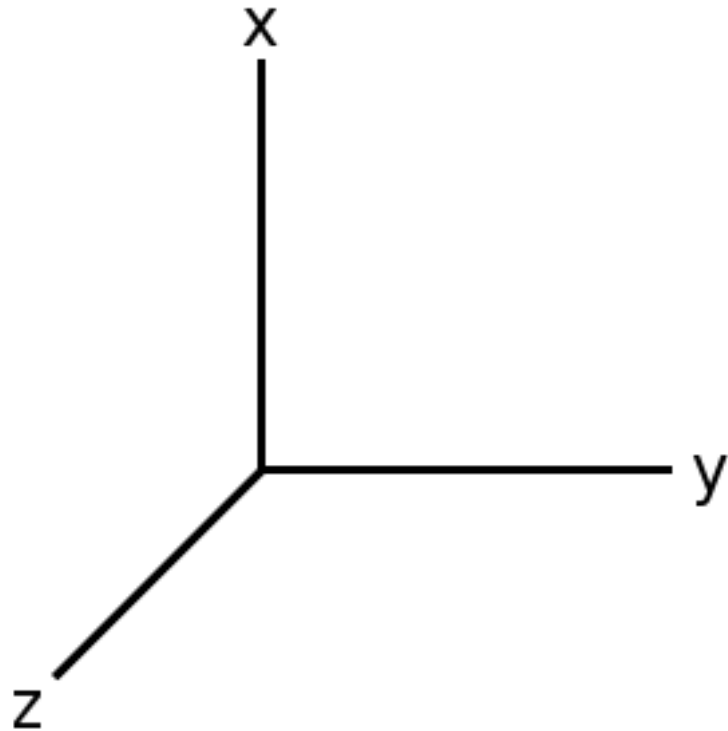
Parallel Coordinate Plots

Parallel Coordinate Plots

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

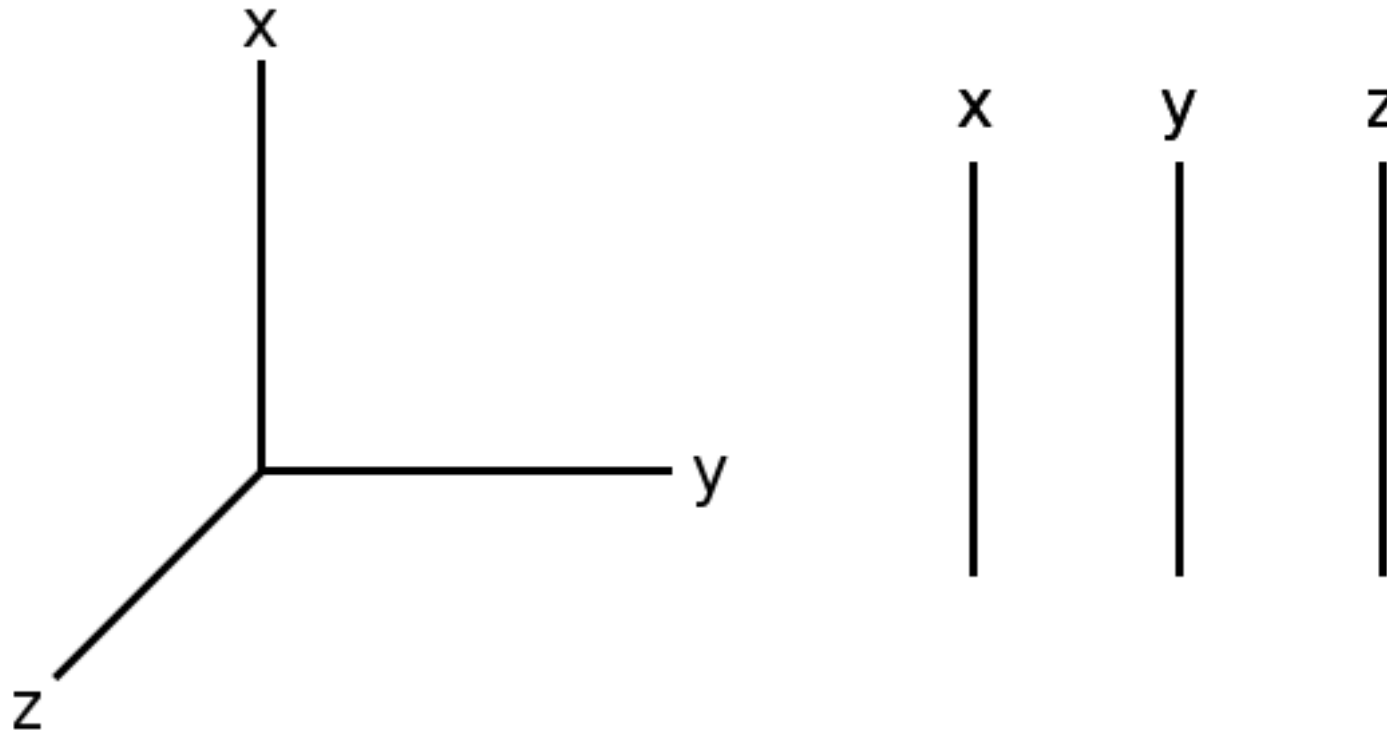
Parallel Coordinate Plots

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



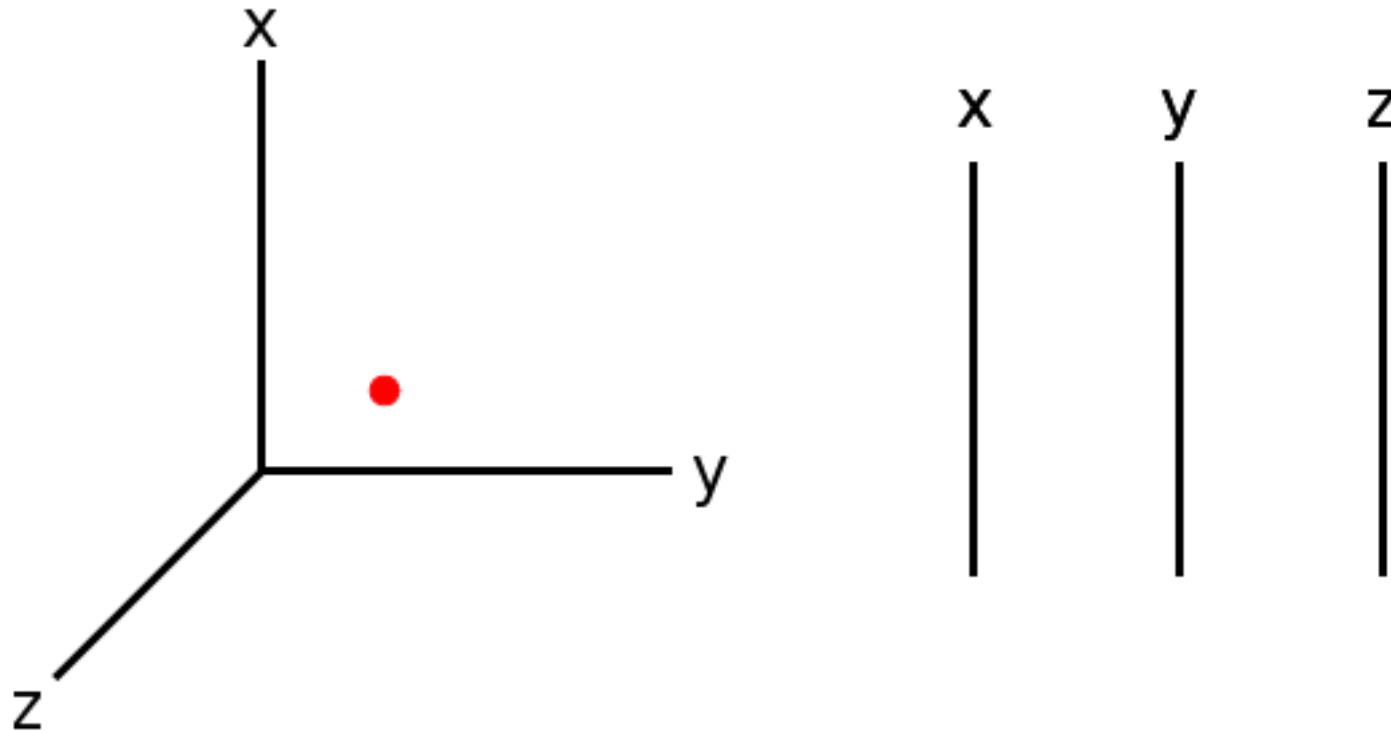
Parallel Coordinate Plots

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



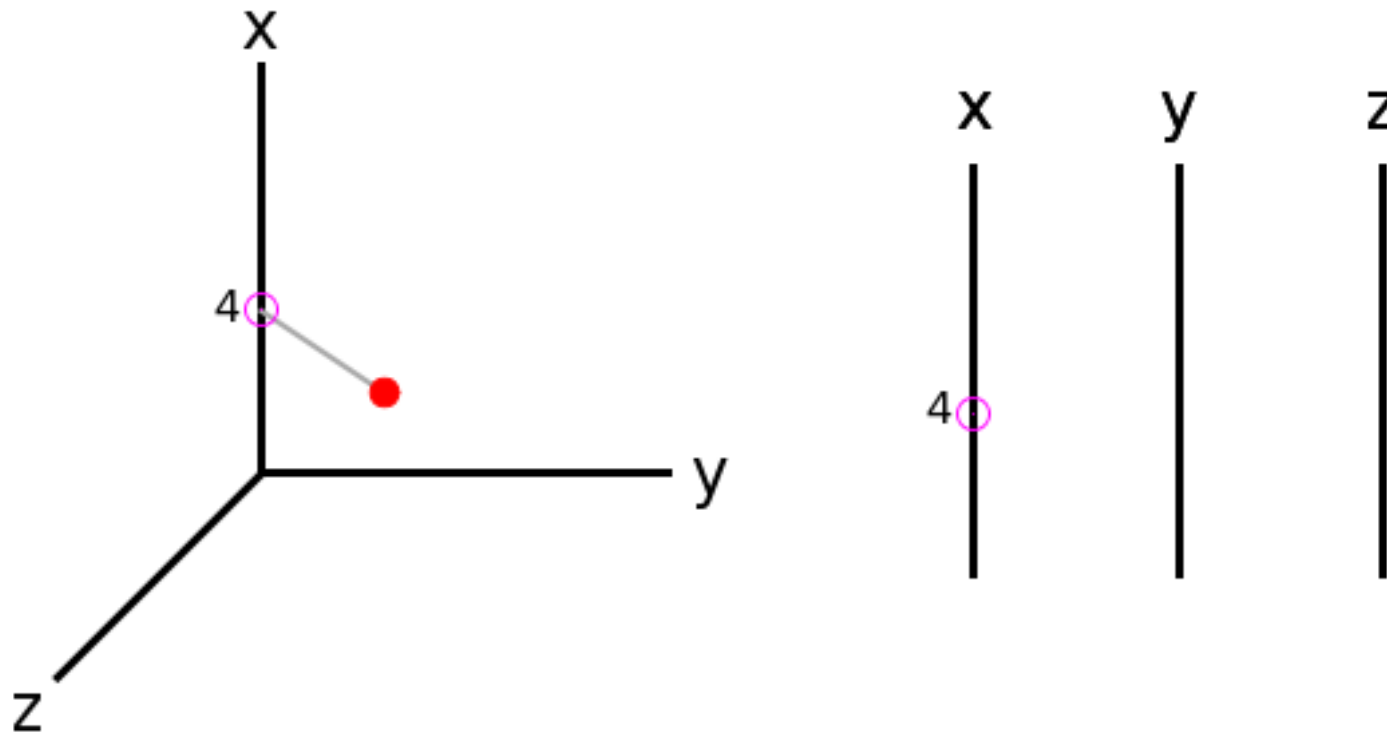
Parallel Coordinate Plots

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



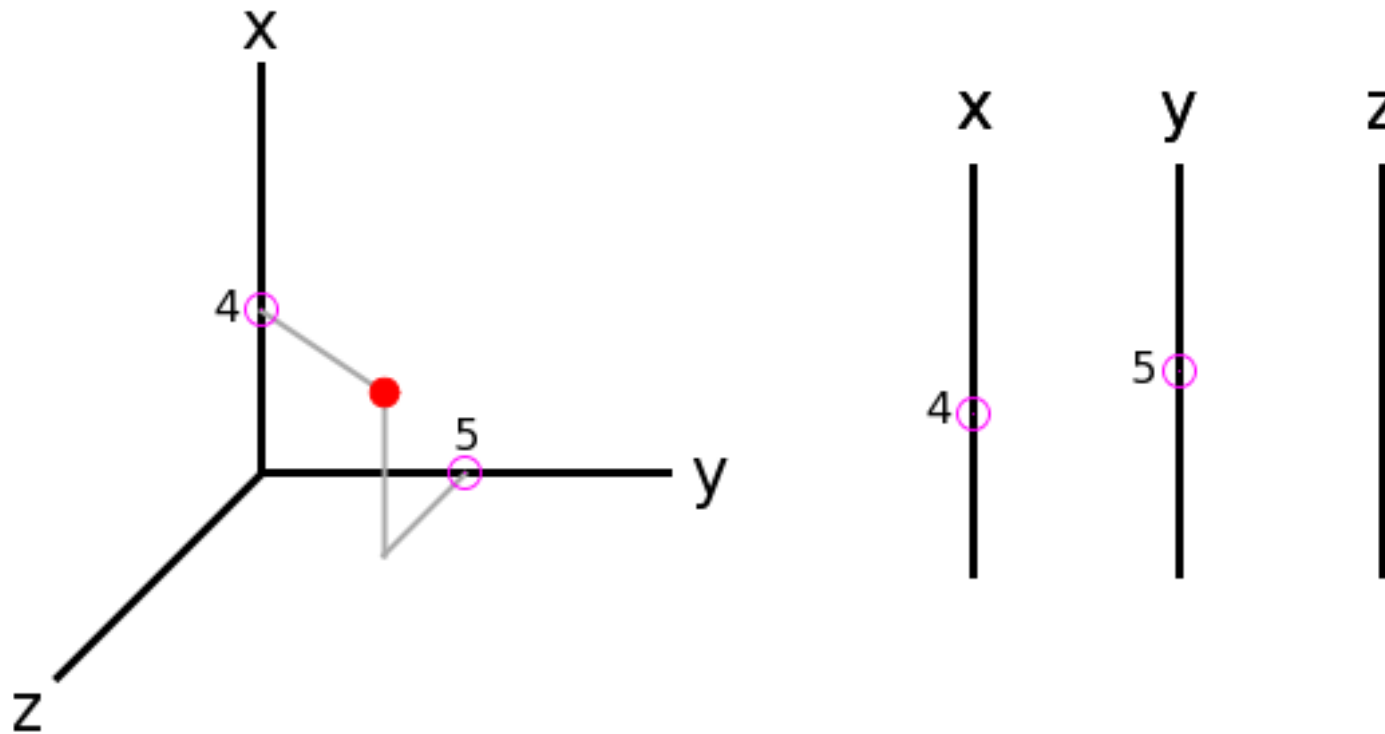
Parallel Coordinate Plots

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



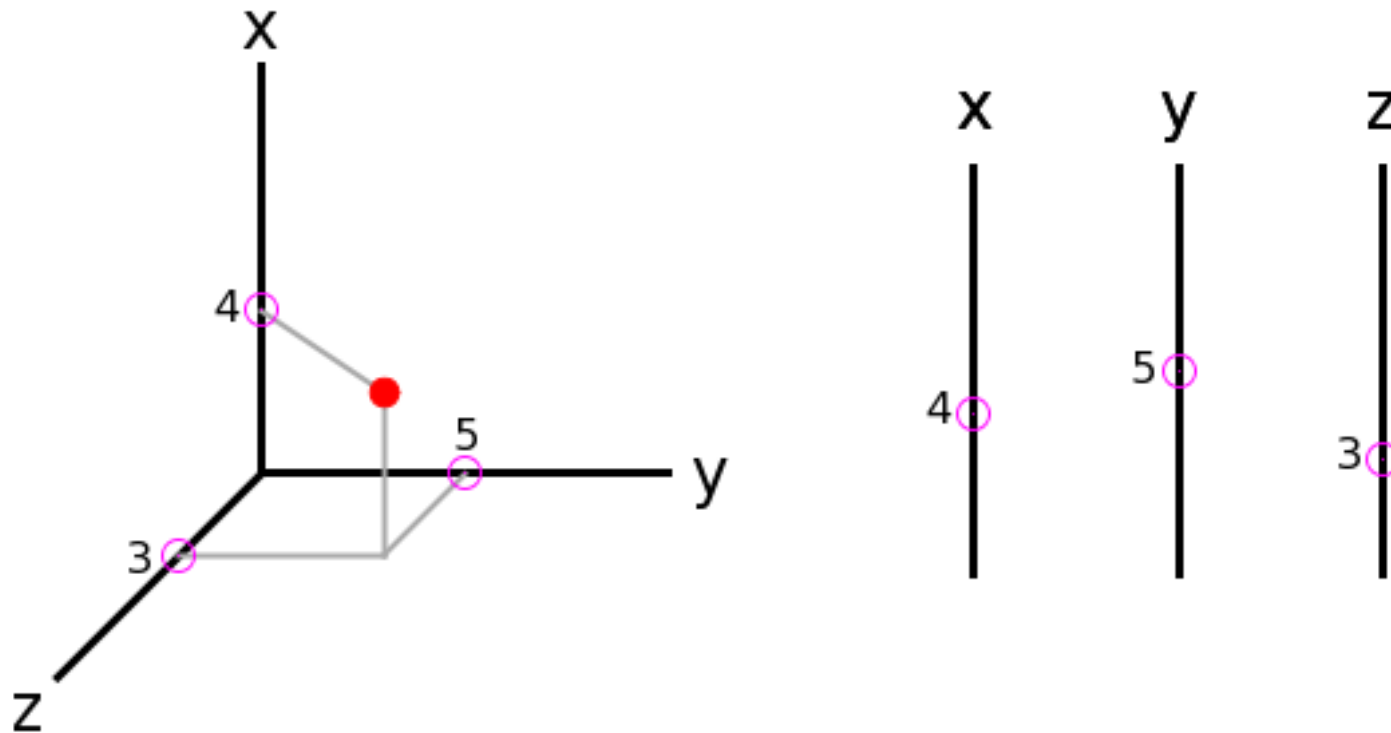
Parallel Coordinate Plots

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



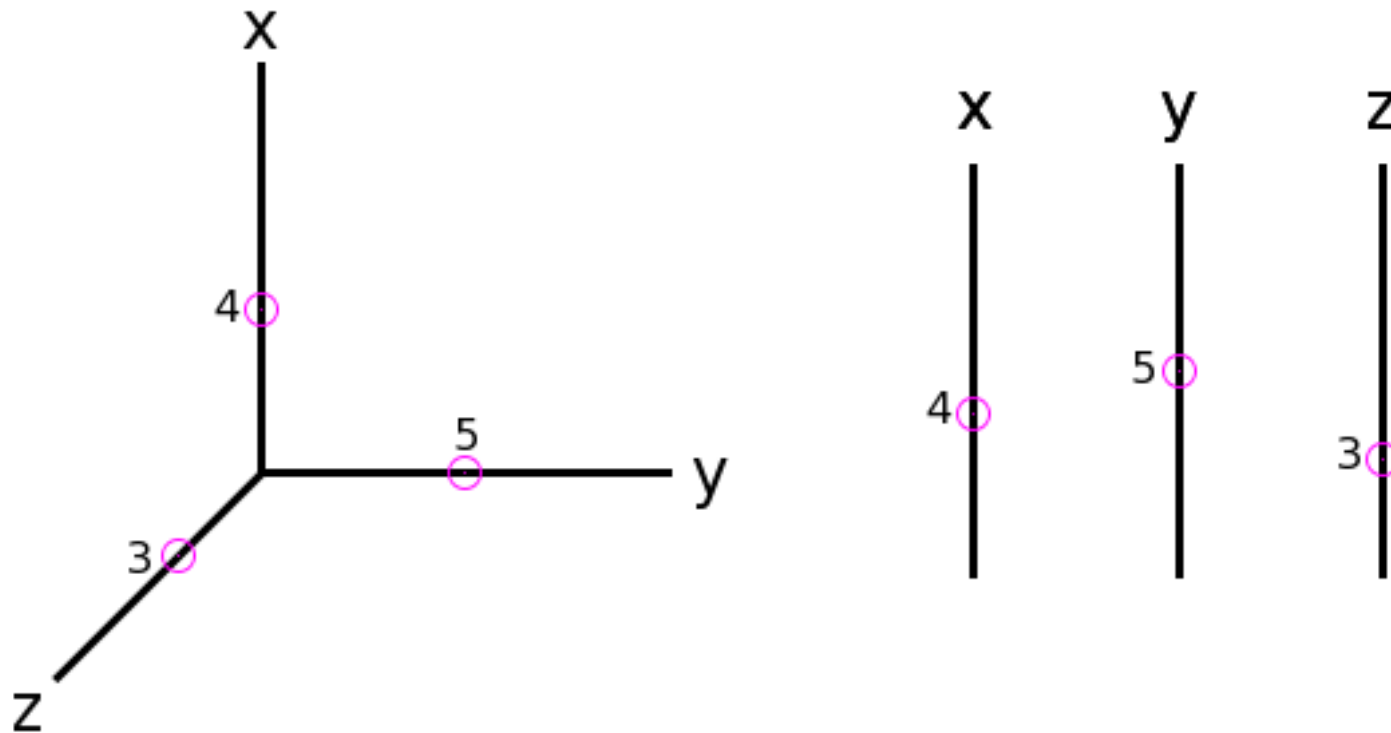
Parallel Coordinate Plots

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



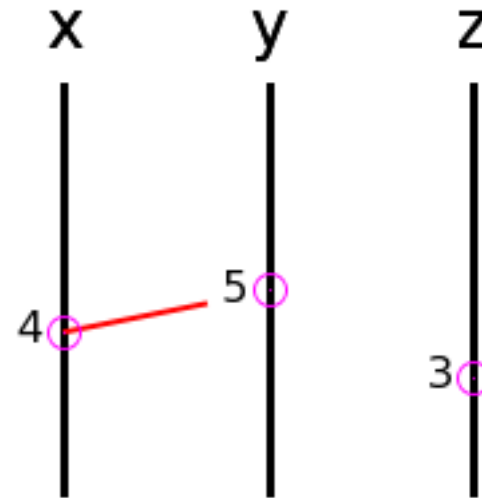
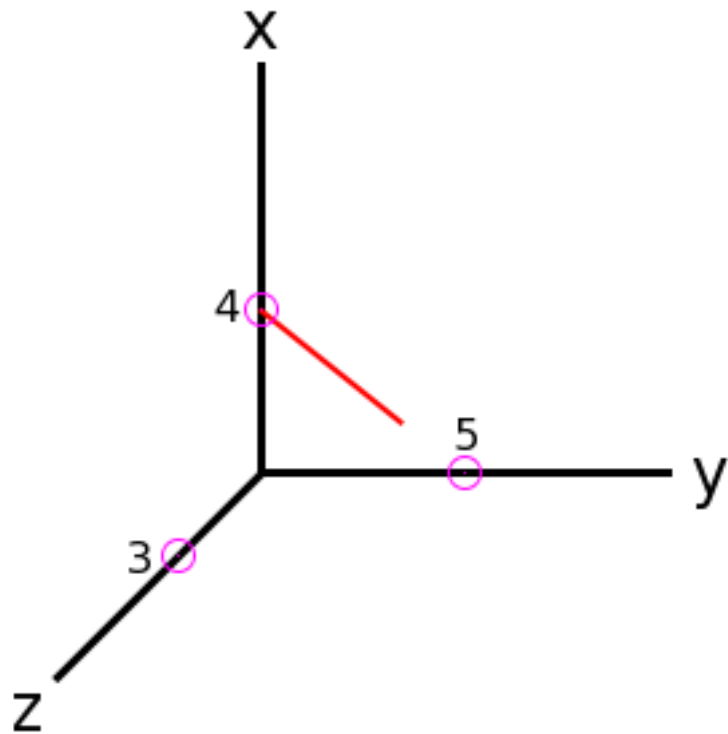
Parallel Coordinate Plots

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



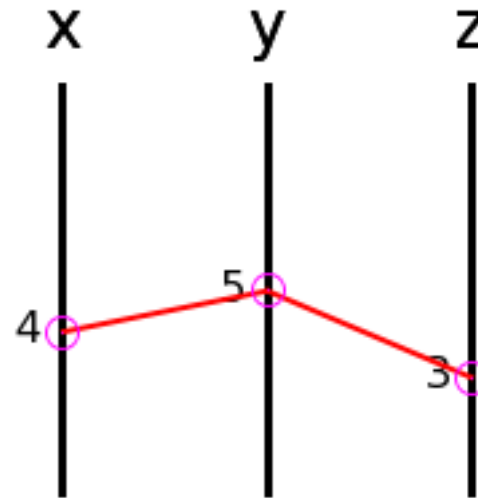
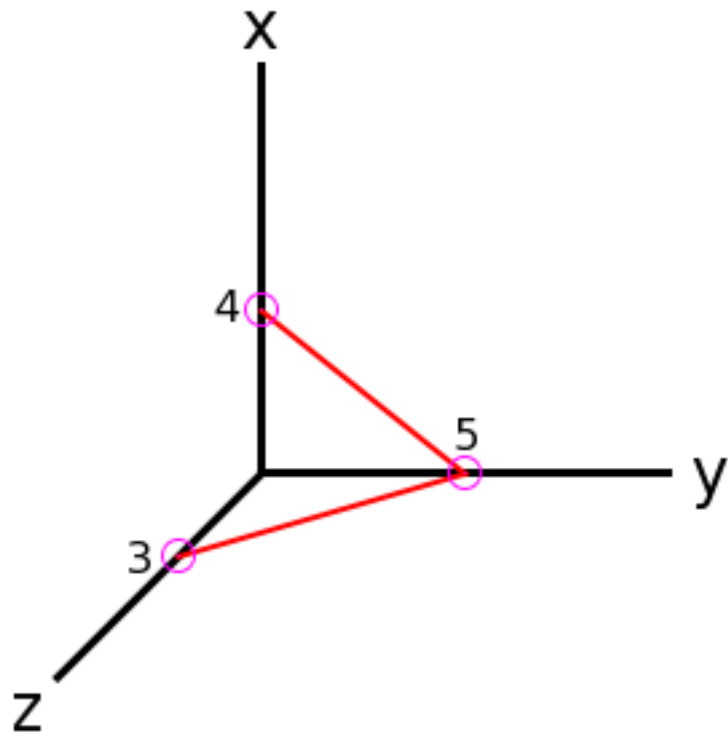
Parallel Coordinate Plots

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



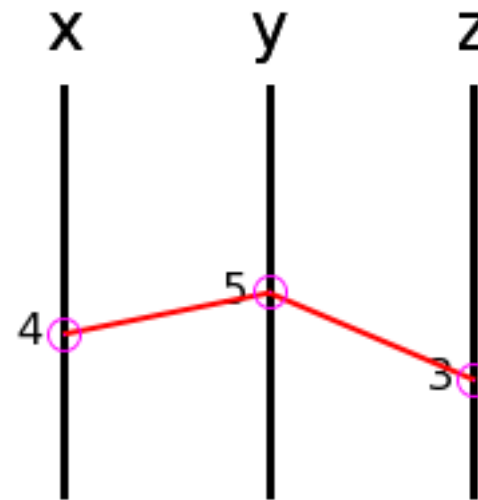
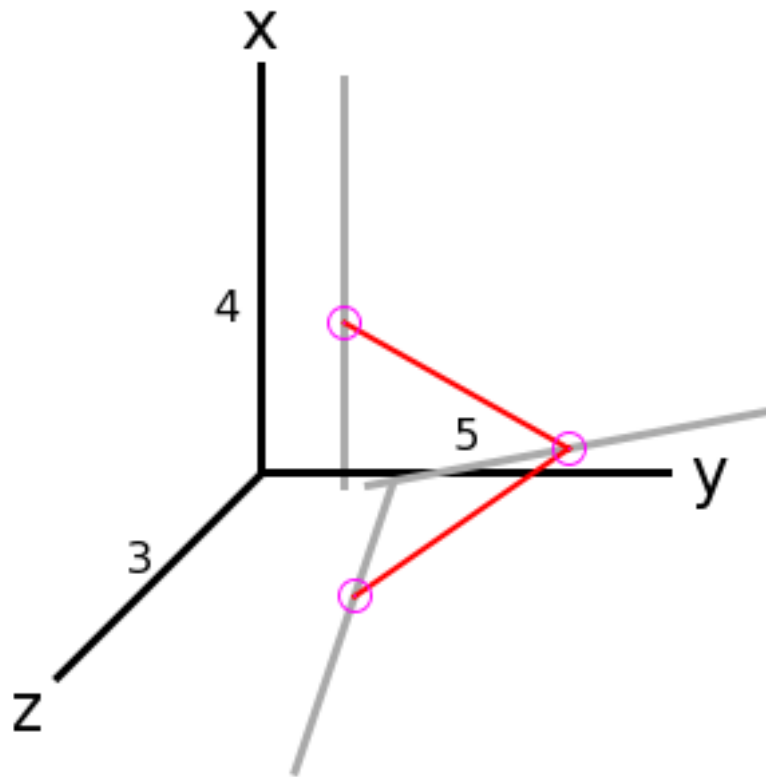
Parallel Coordinate Plots

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



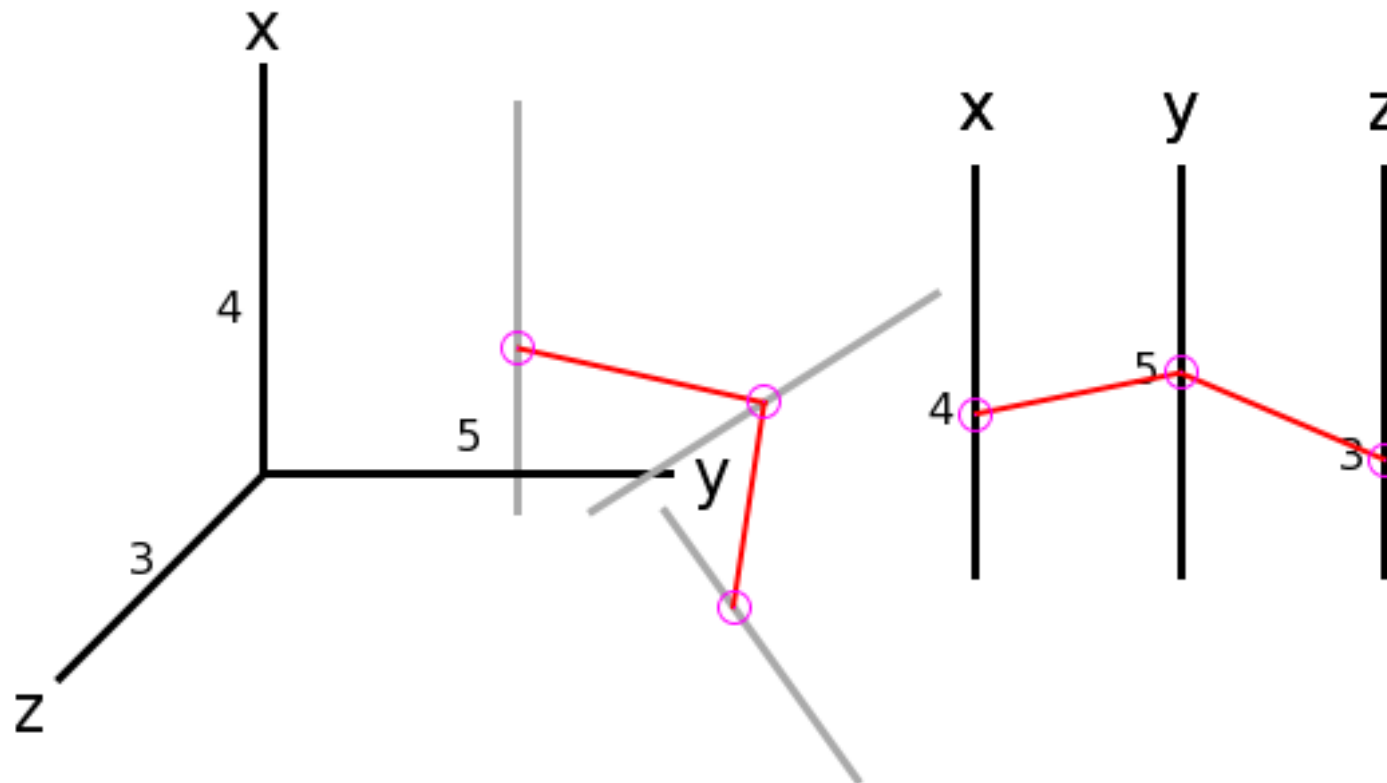
Parallel Coordinate Plots

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



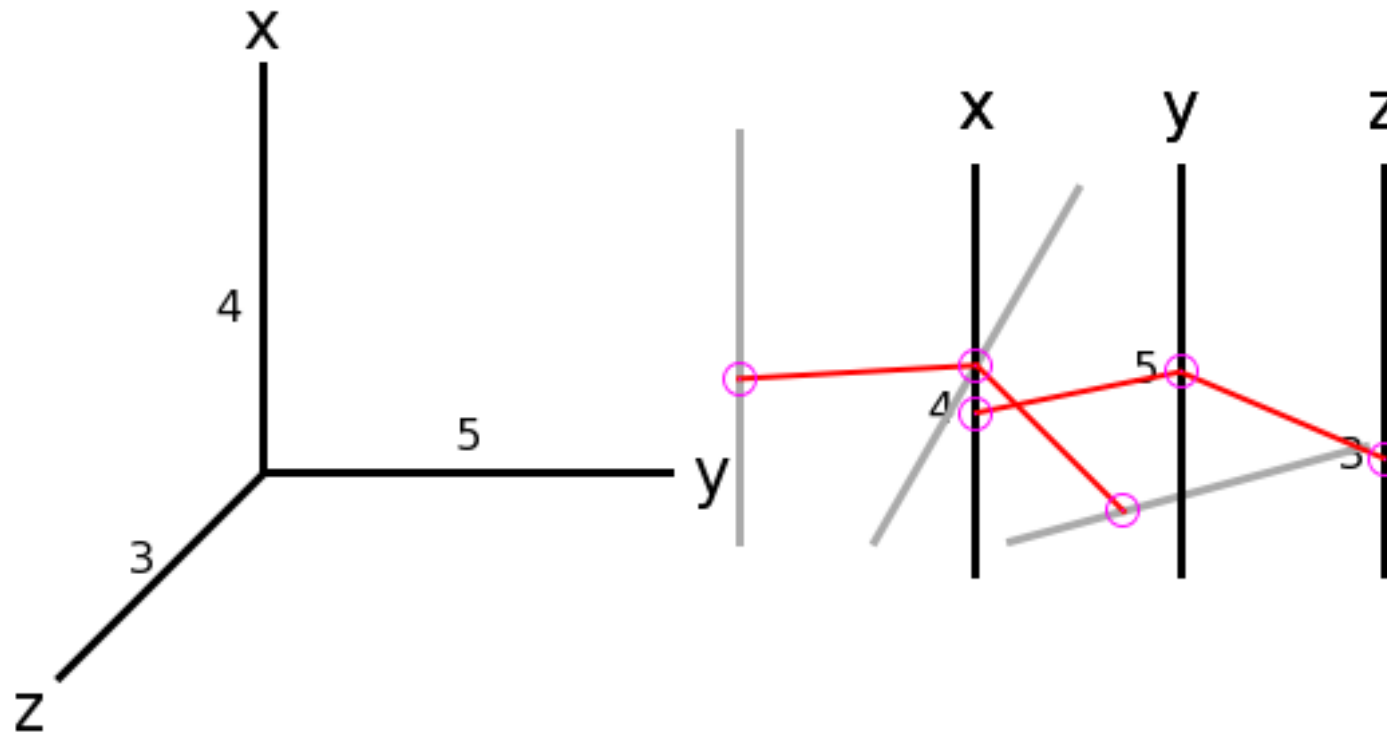
Parallel Coordinate Plots

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



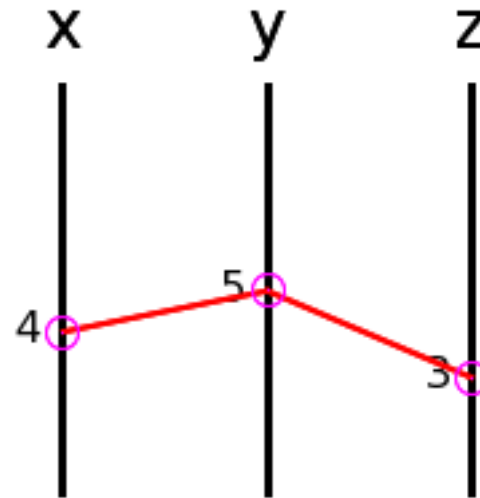
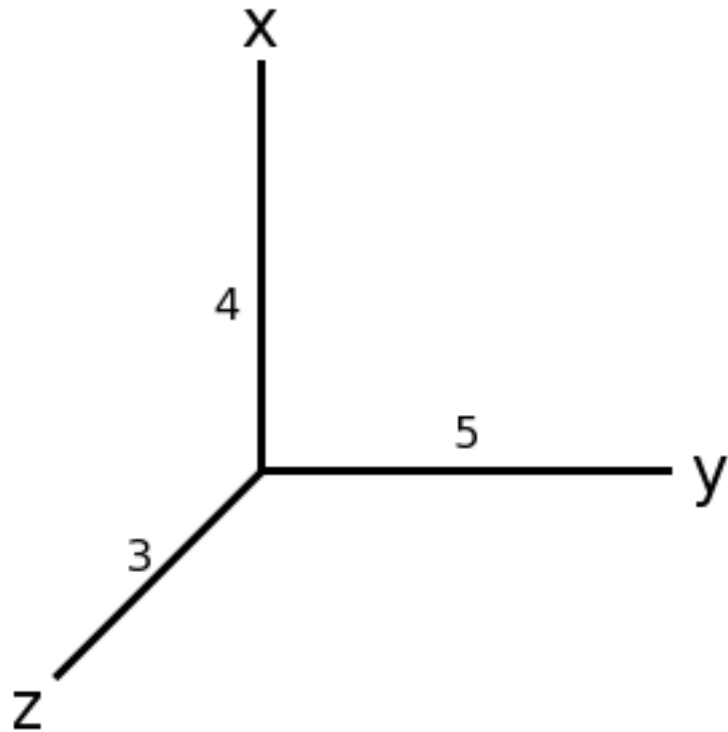
Parallel Coordinate Plots

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



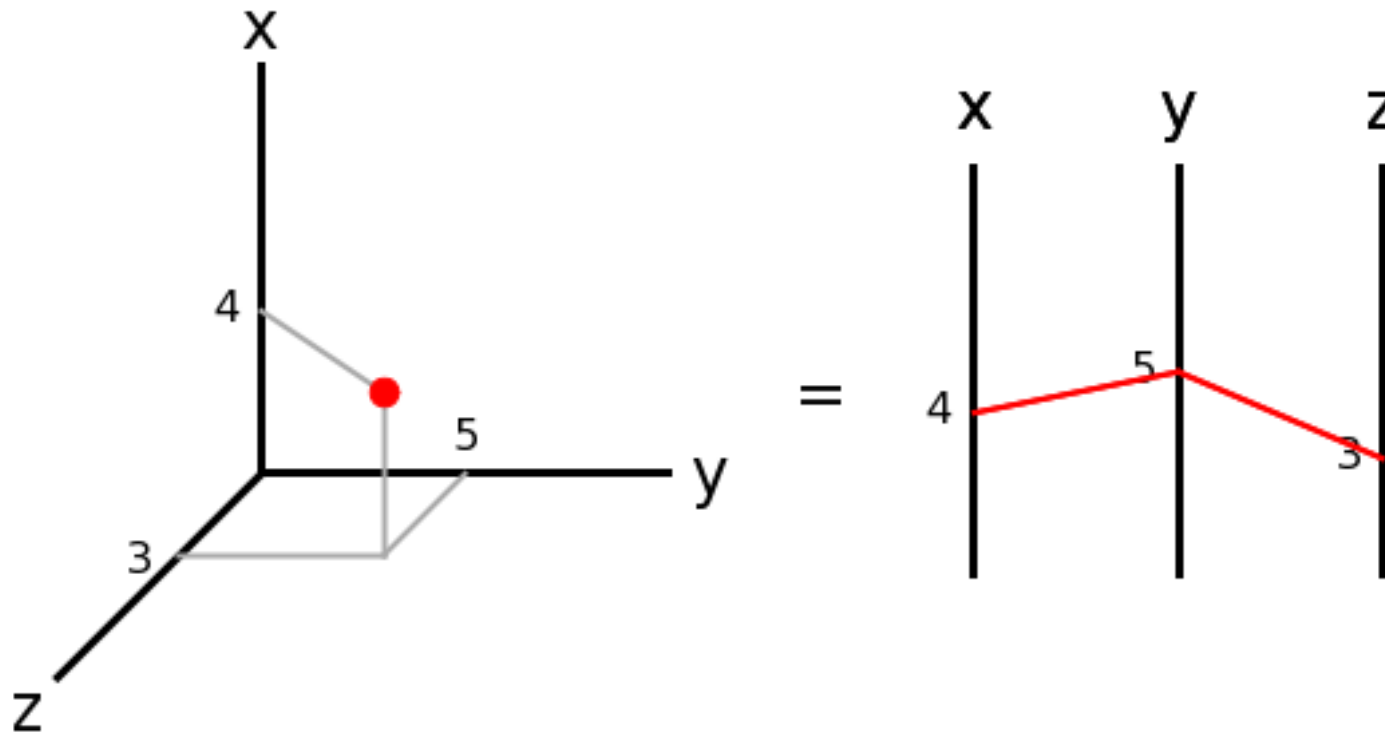
Parallel Coordinate Plots

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



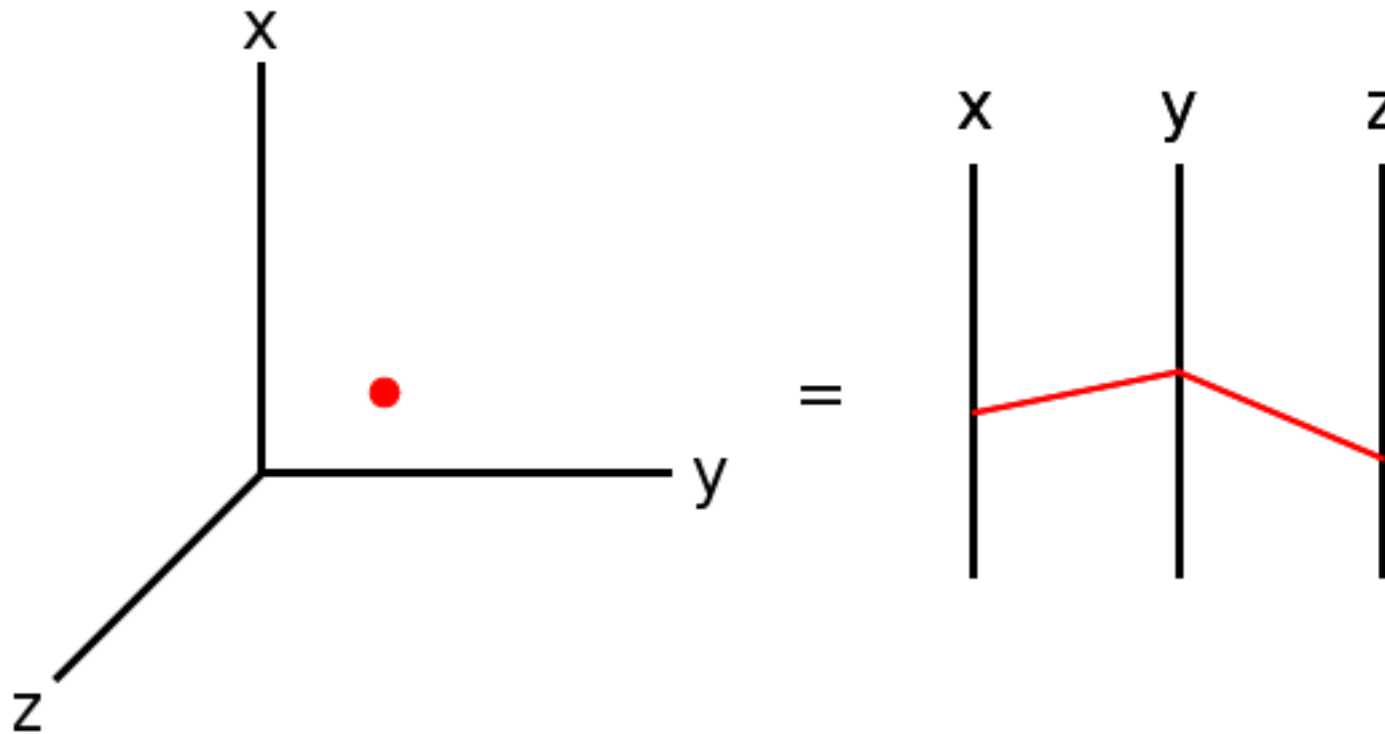
Parallel Coordinate Plots

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



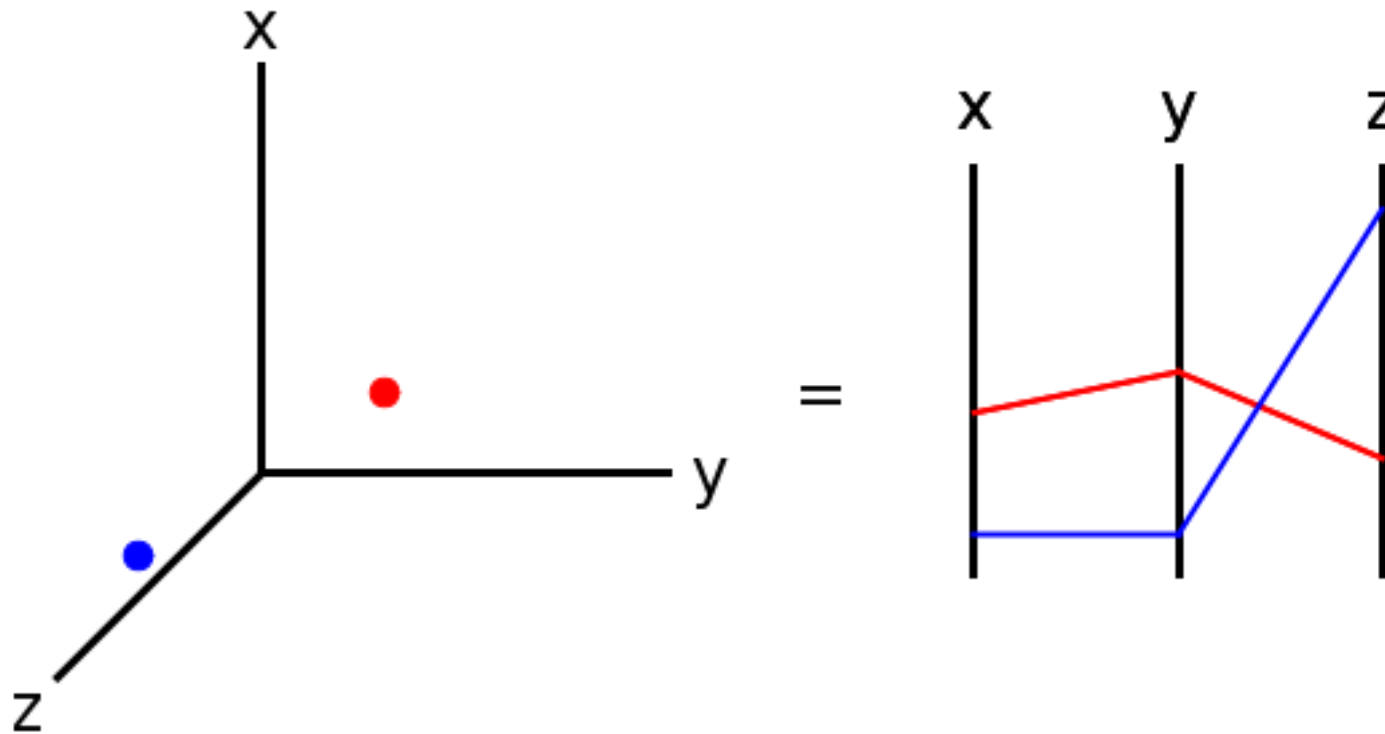
Parallel Coordinate Plots

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



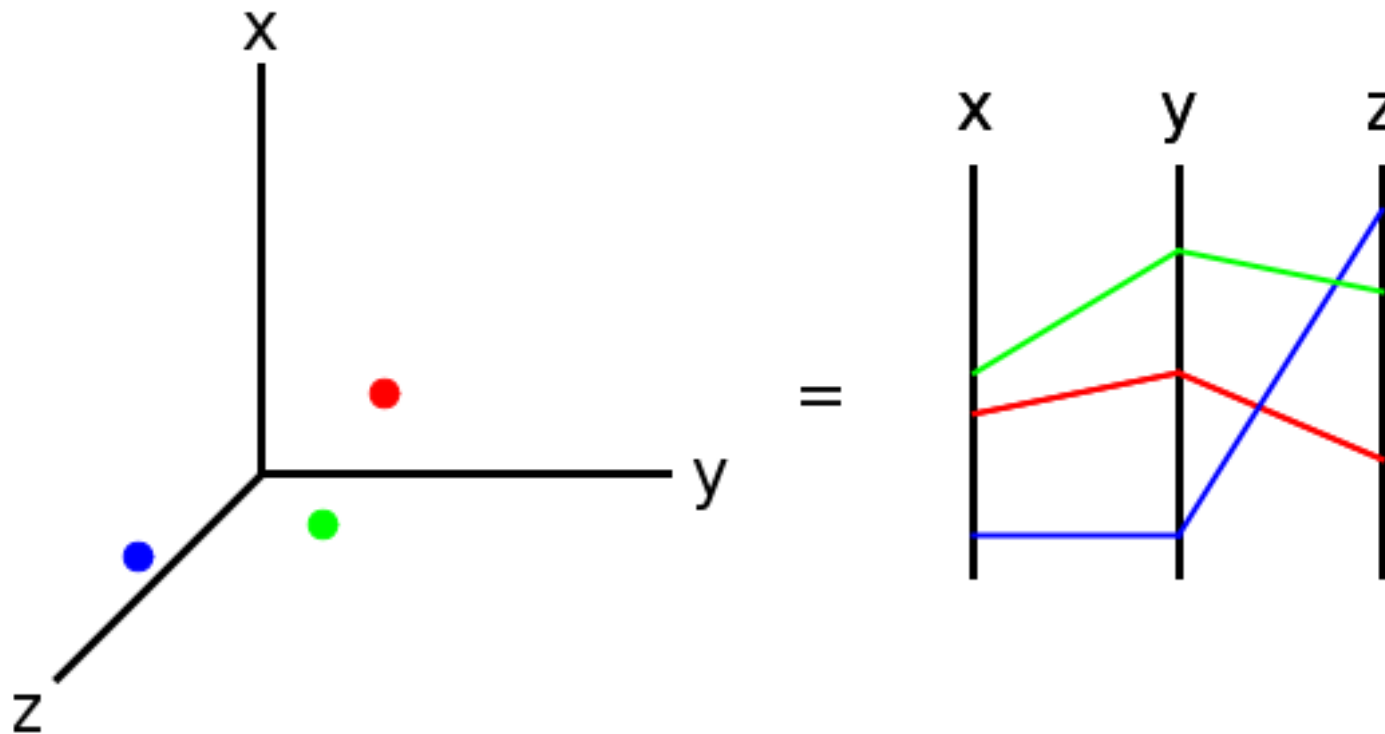
Parallel Coordinate Plots

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



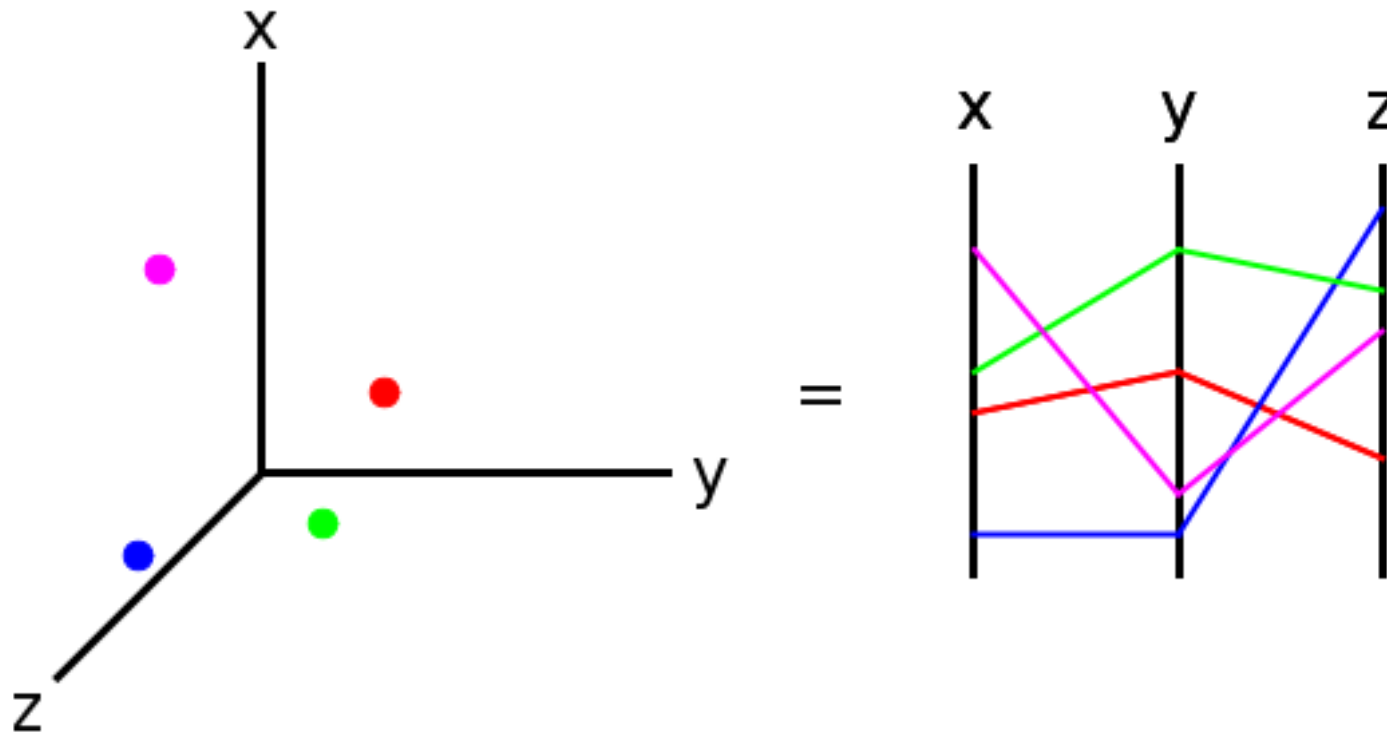
Parallel Coordinate Plots

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

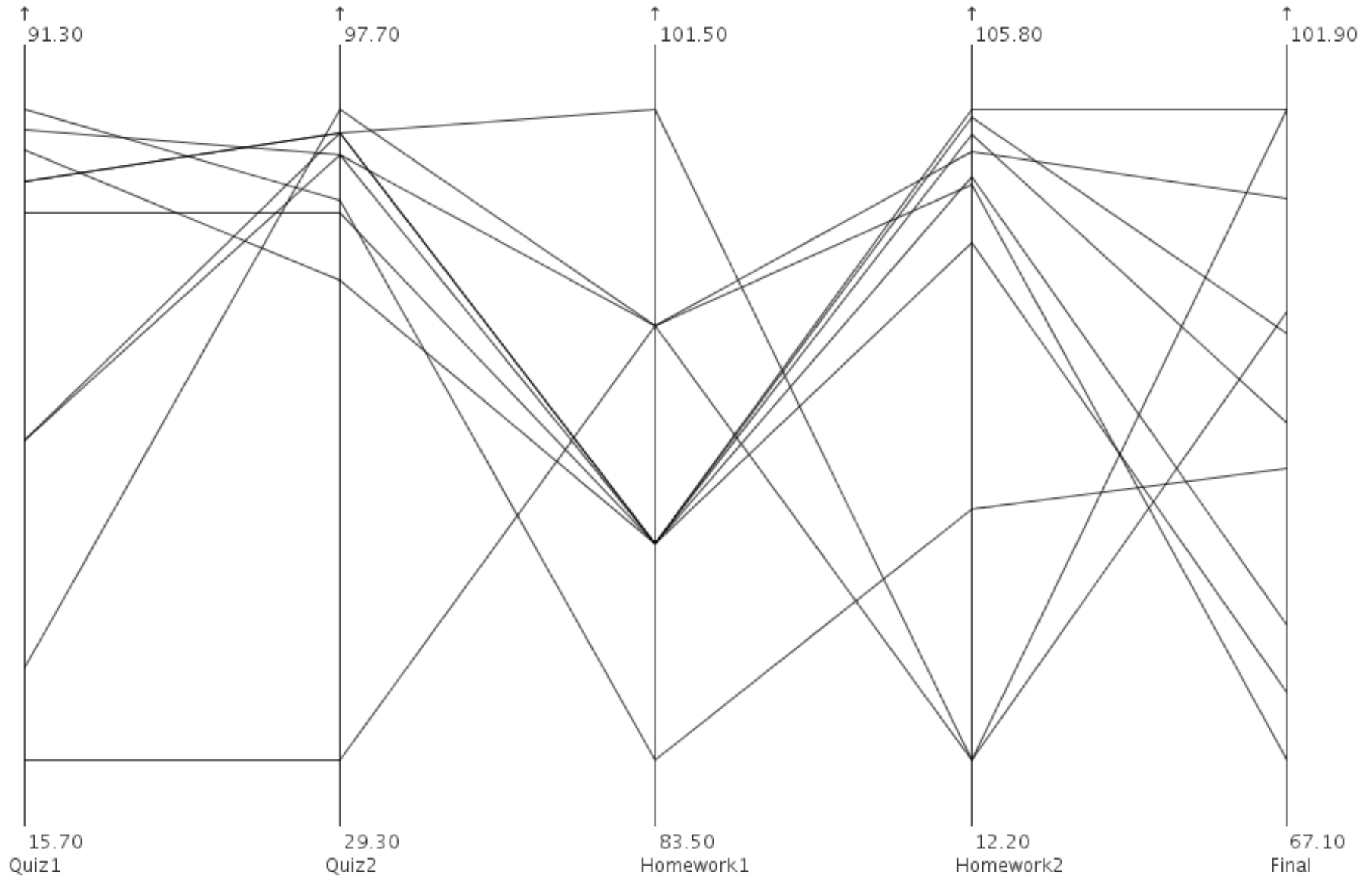


Parallel Coordinate Plots

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

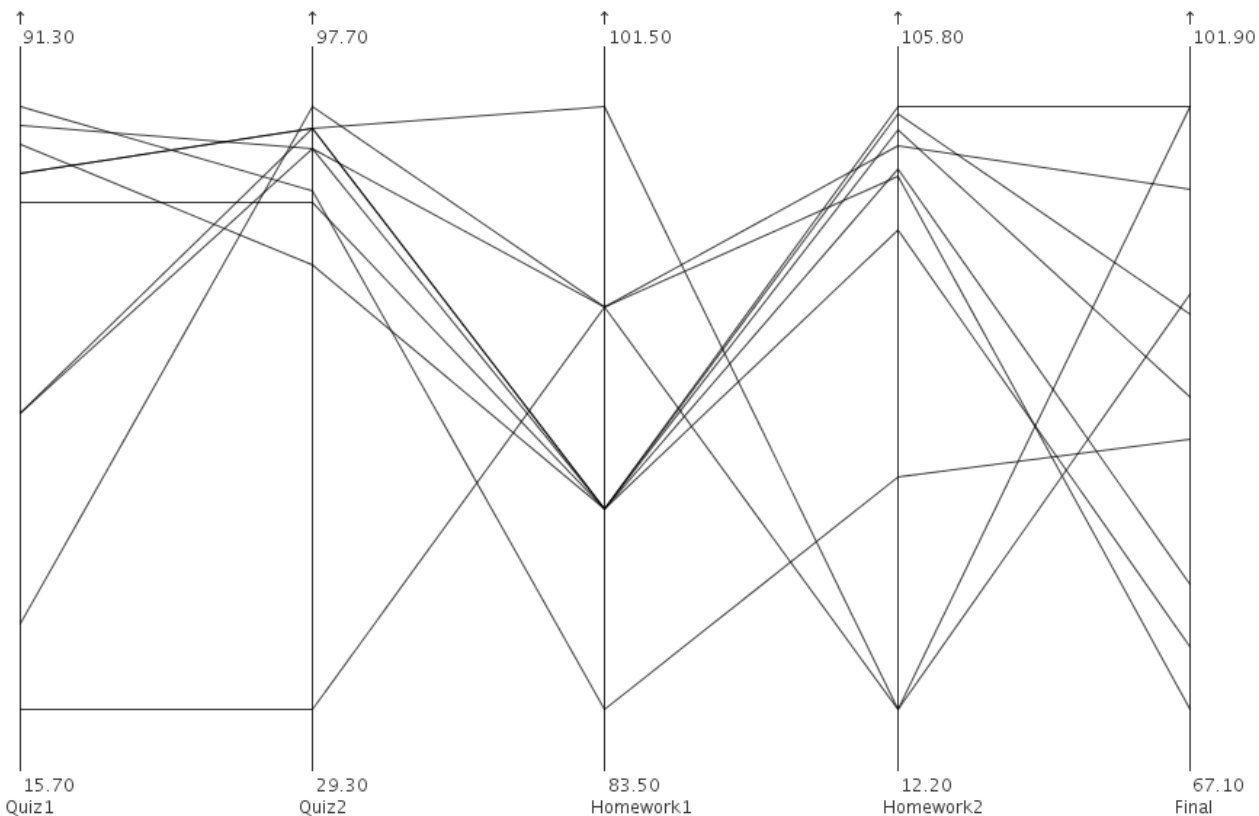


Example



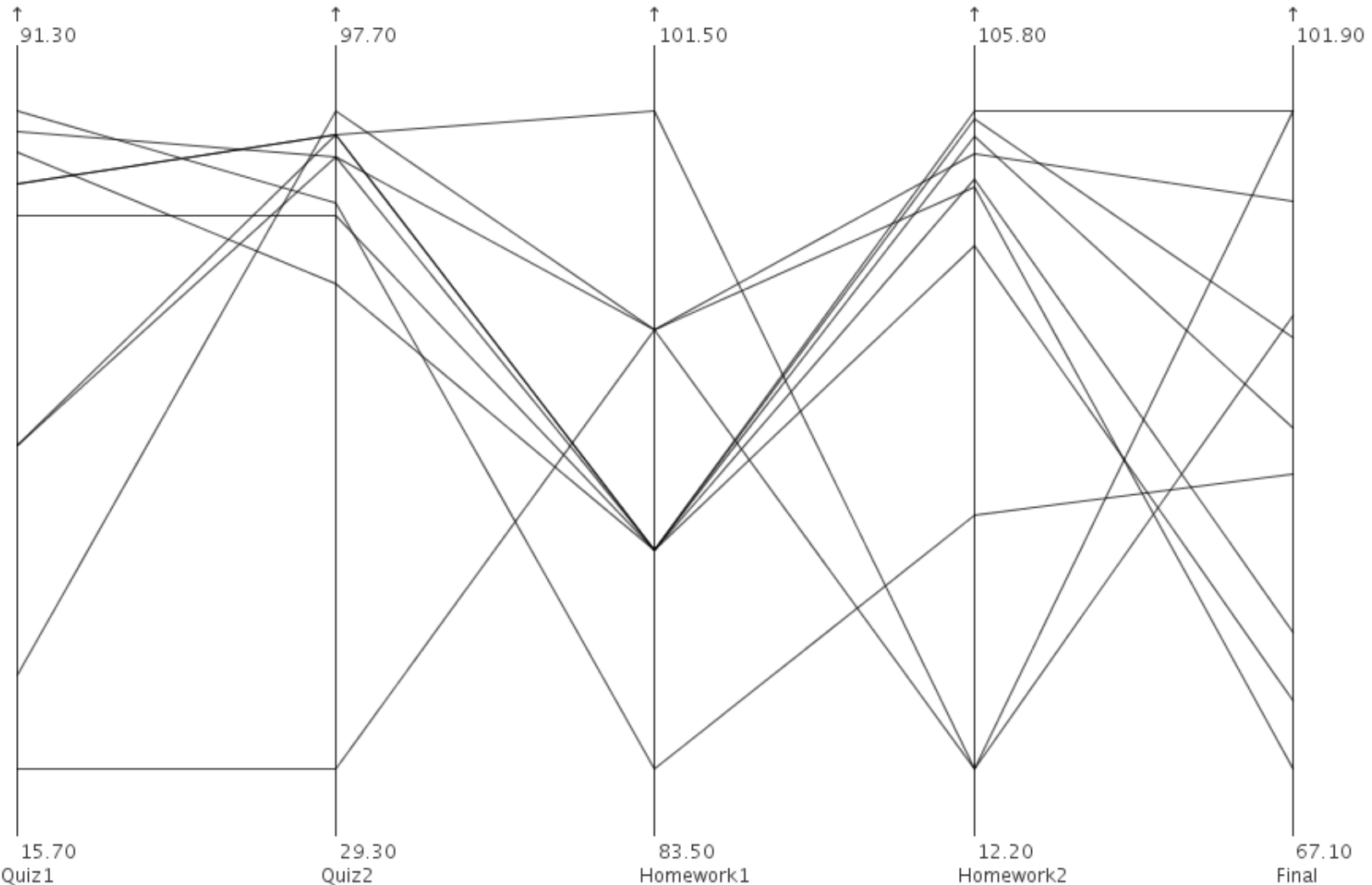
Advantages

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



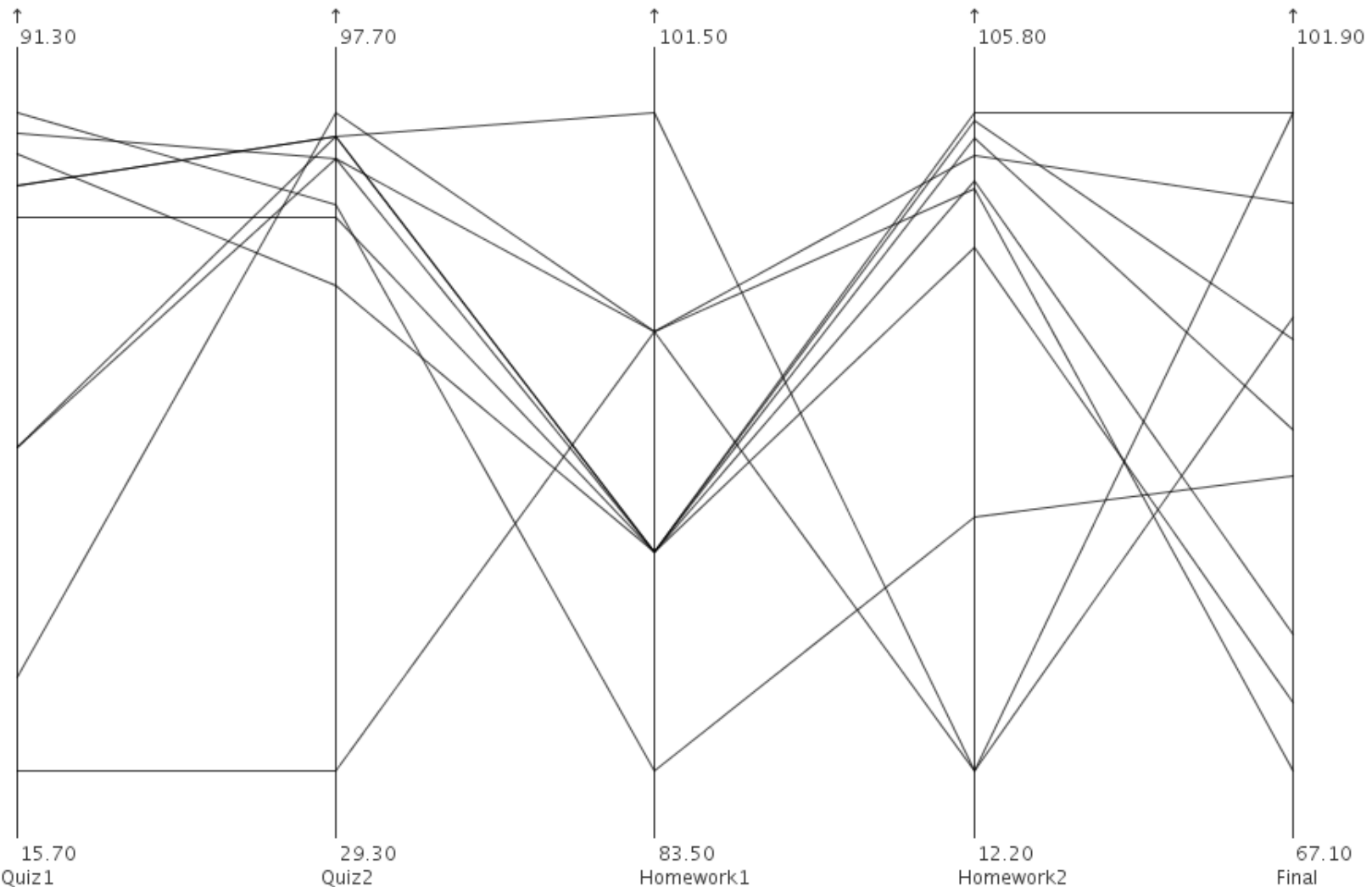
Challenge: Ambiguity

Common values cause ambiguity



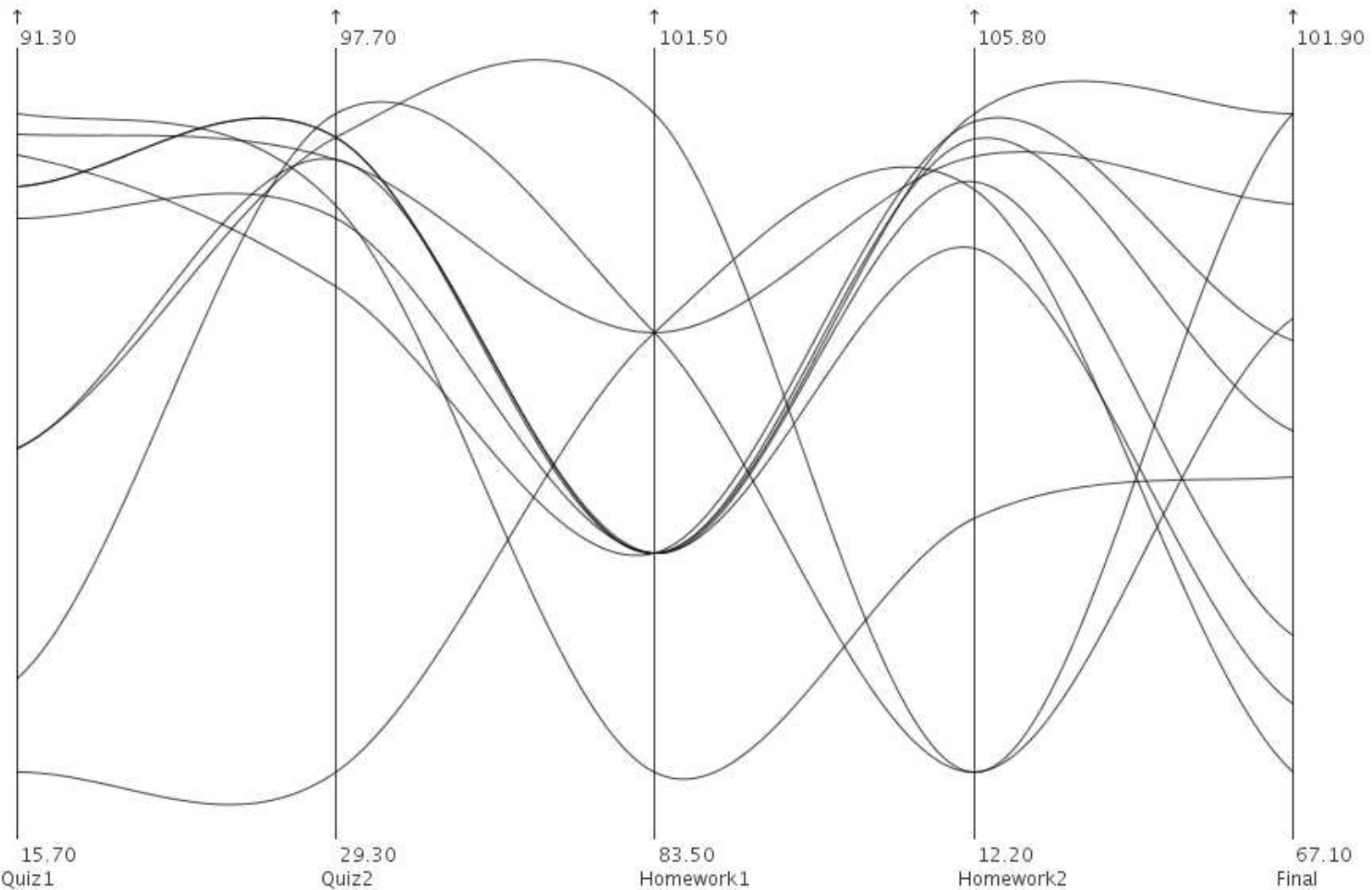
Challenge: Ambiguity

Common values cause ambiguity
Solution: (a) Curves



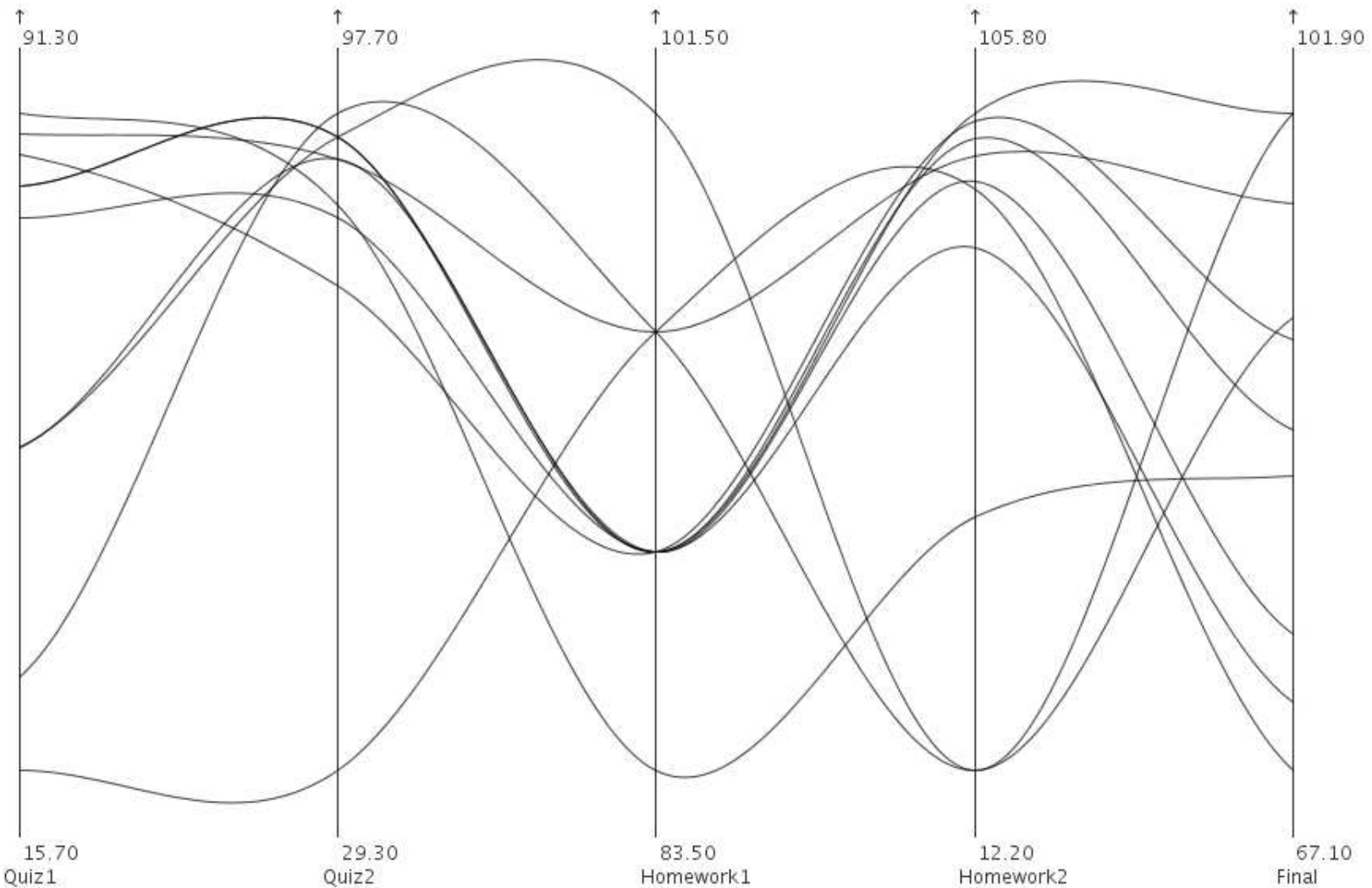
Challenge: Ambiguity

Common values cause ambiguity
Solution: (a) Curves



Challenge: Ambiguity

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

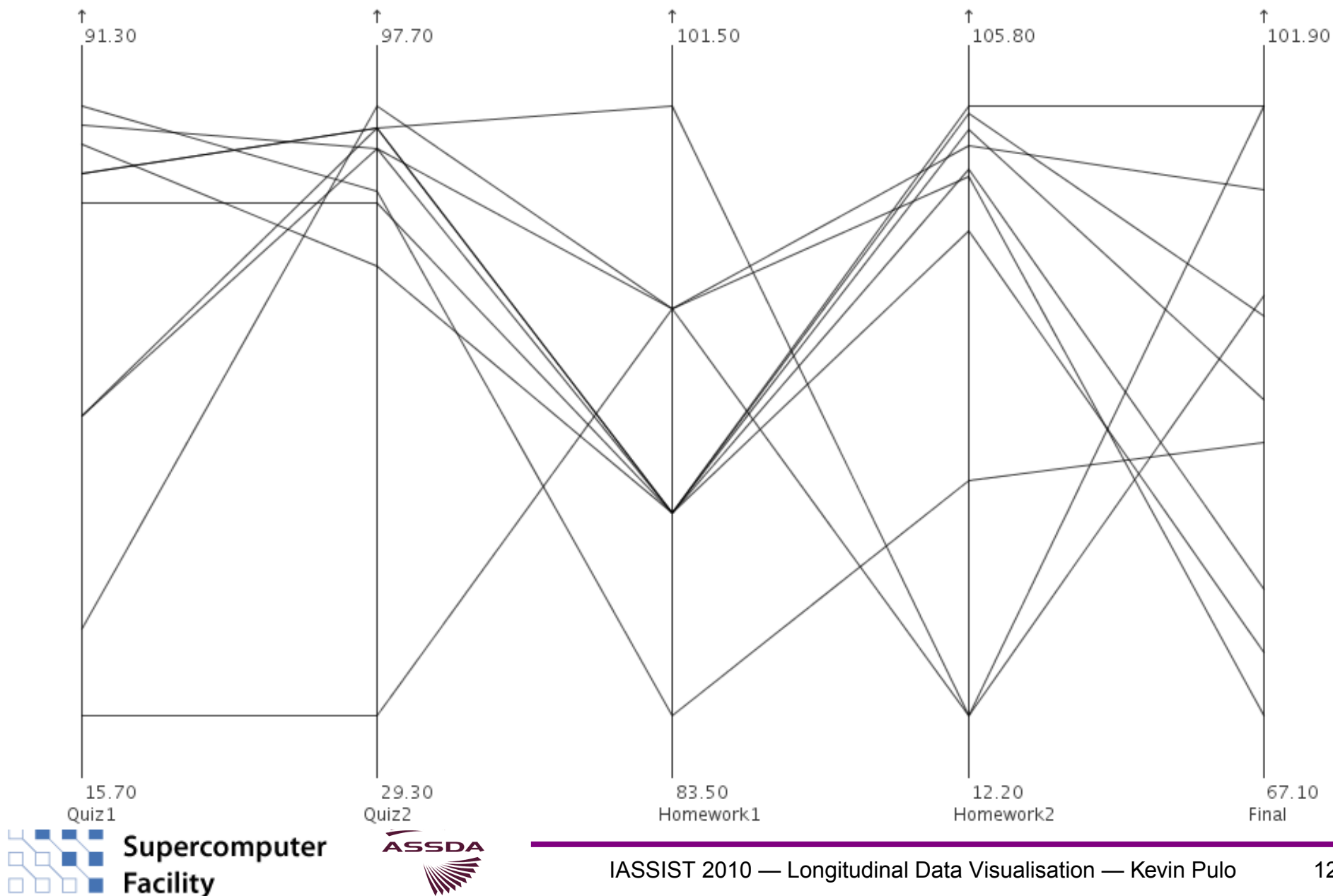


Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation

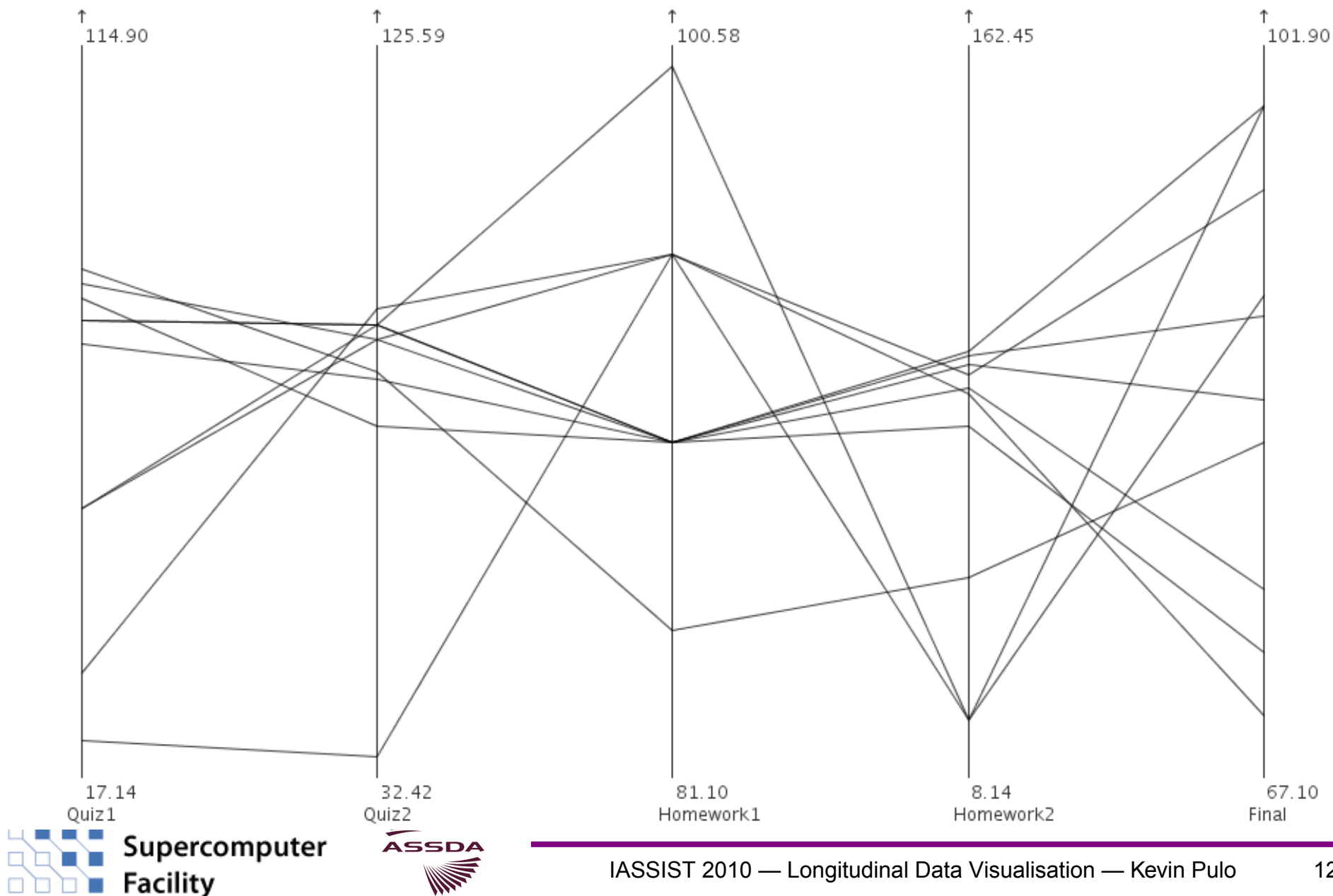
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation



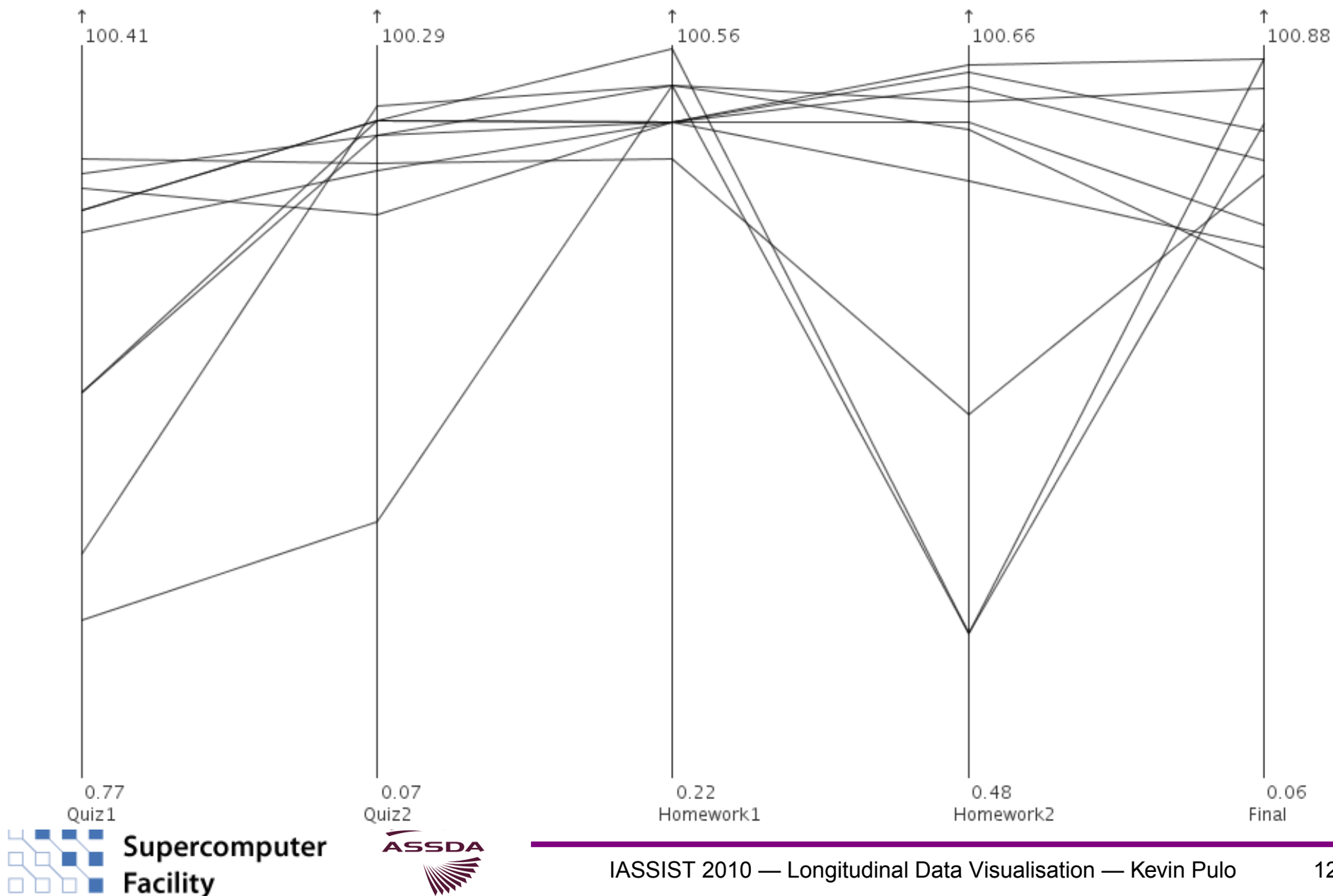
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation



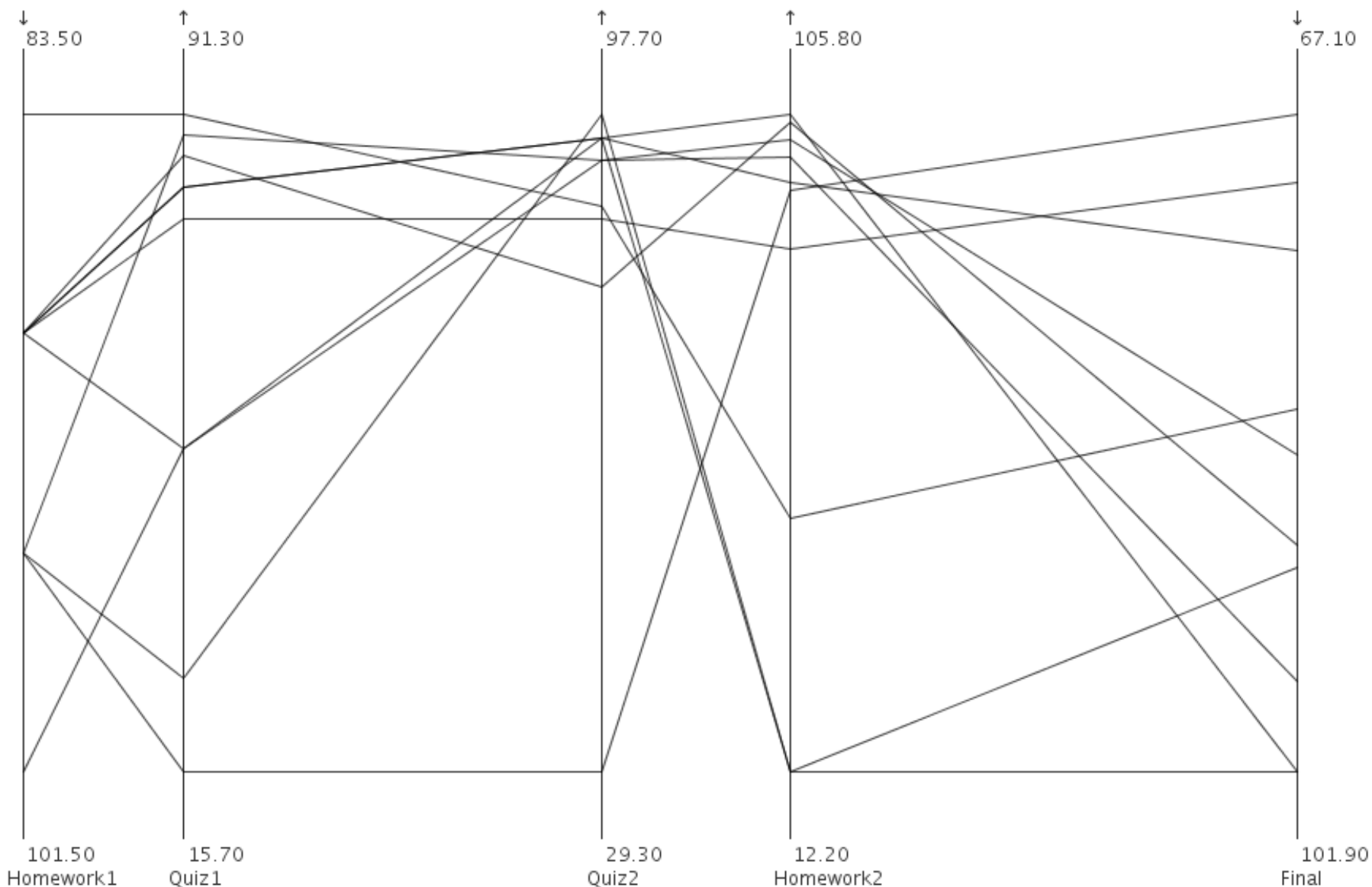
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation



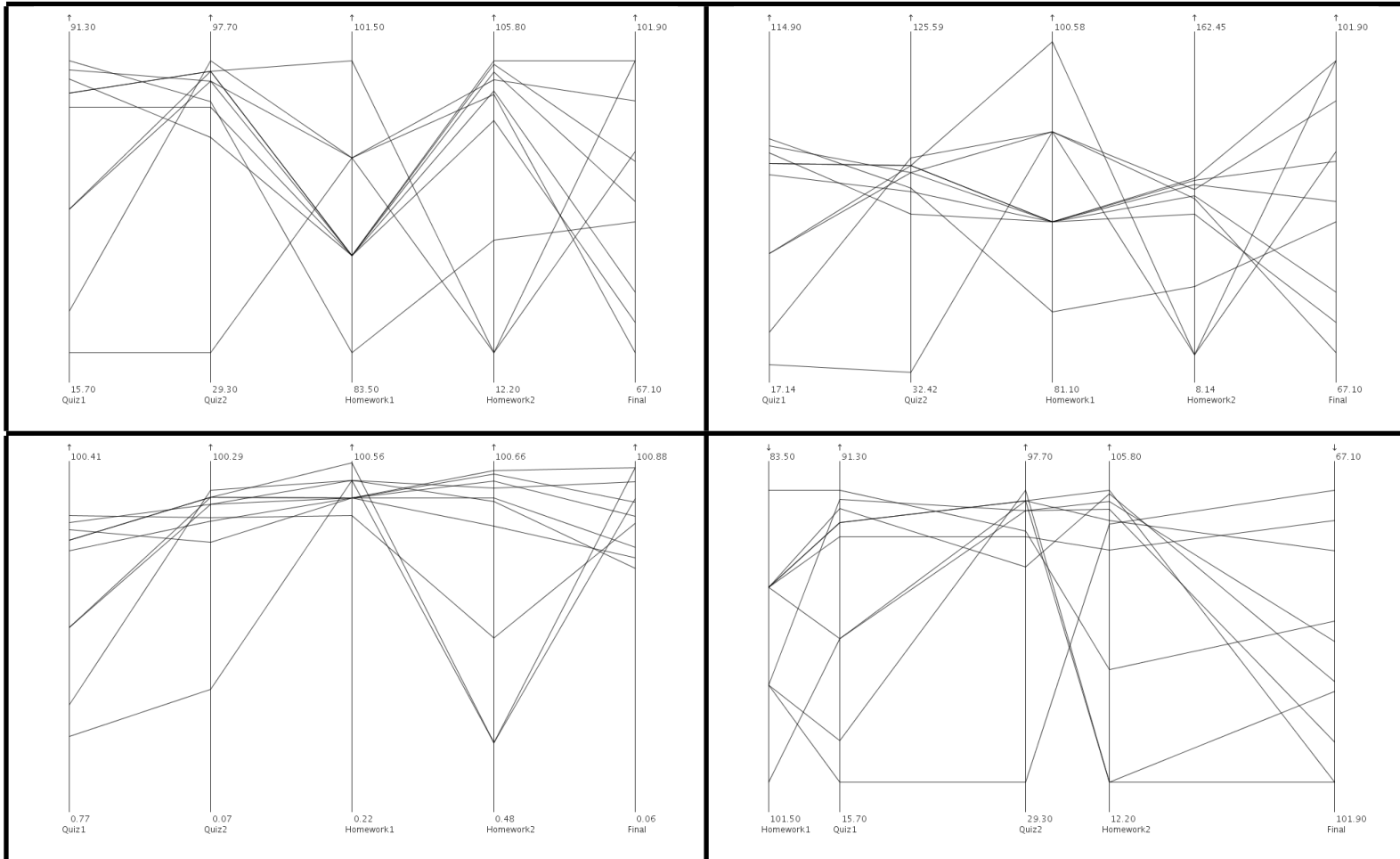
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation



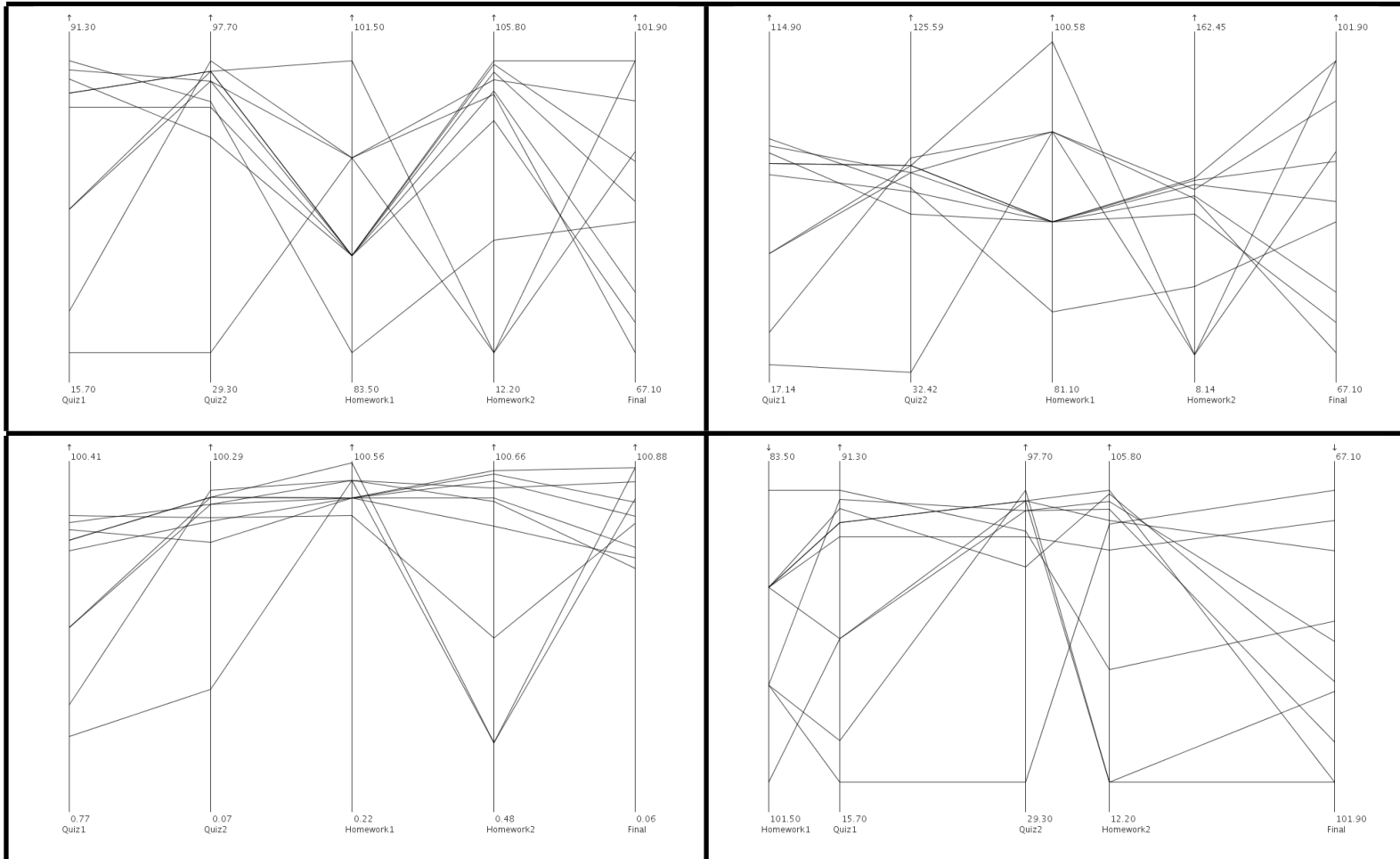
Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation



Challenge: Axis parameters

View affected by axis ordering, sign, scaling, translation
Solution: Interactive manipulation



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

Aspects examined

- Labour force status
 - At each wave
 - Employed, unemployed, NILF

Aspects examined

- Labour force status
 - At each wave
 - Employed, unemployed, NILF
- Work and study history
 - Retrospectively from age 15 (up to age 62)
 - Full-time, part-time, none

Aspects examined

- Labour force status
 - At each wave
 - Employed, unemployed, NILF
- Work and study history
 - Retrospectively from age 15 (up to age 62)
 - Full-time, part-time, none
- Relationship path to first birth
 - Retrospectively from age 15 (up to age 62)
 - Single, co-habiting, married

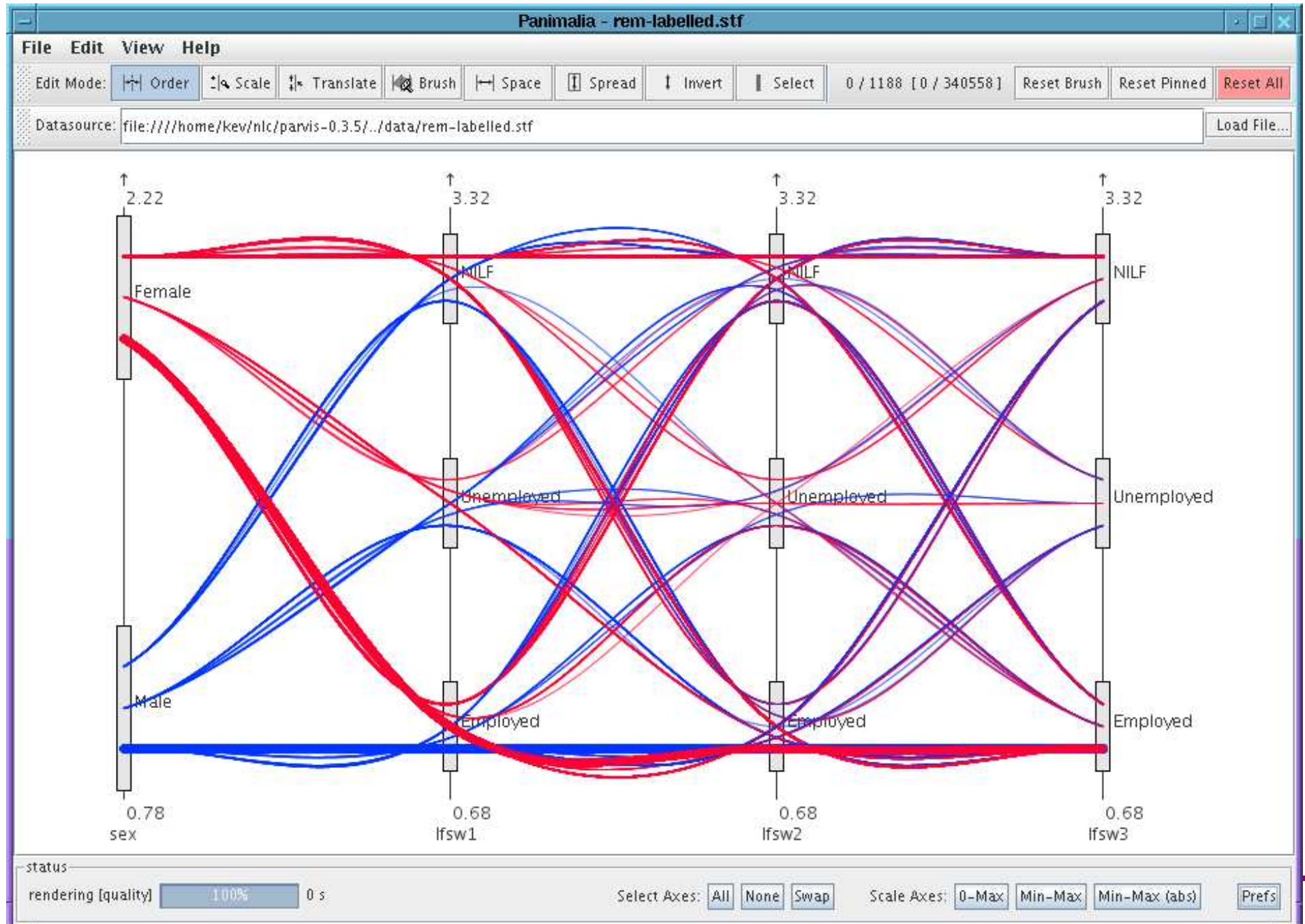
Pilot software tool

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)

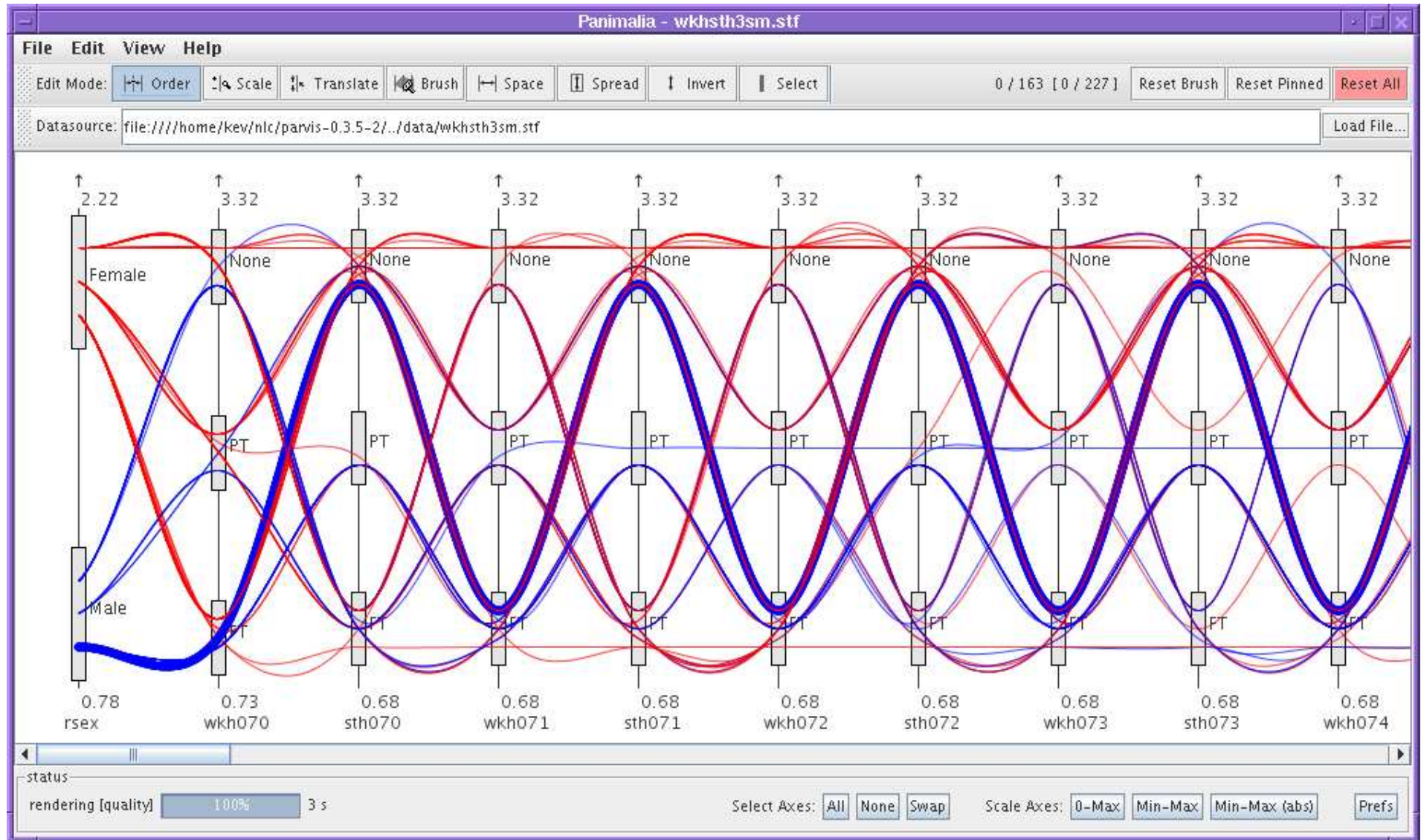
Labour force status

Labour force status

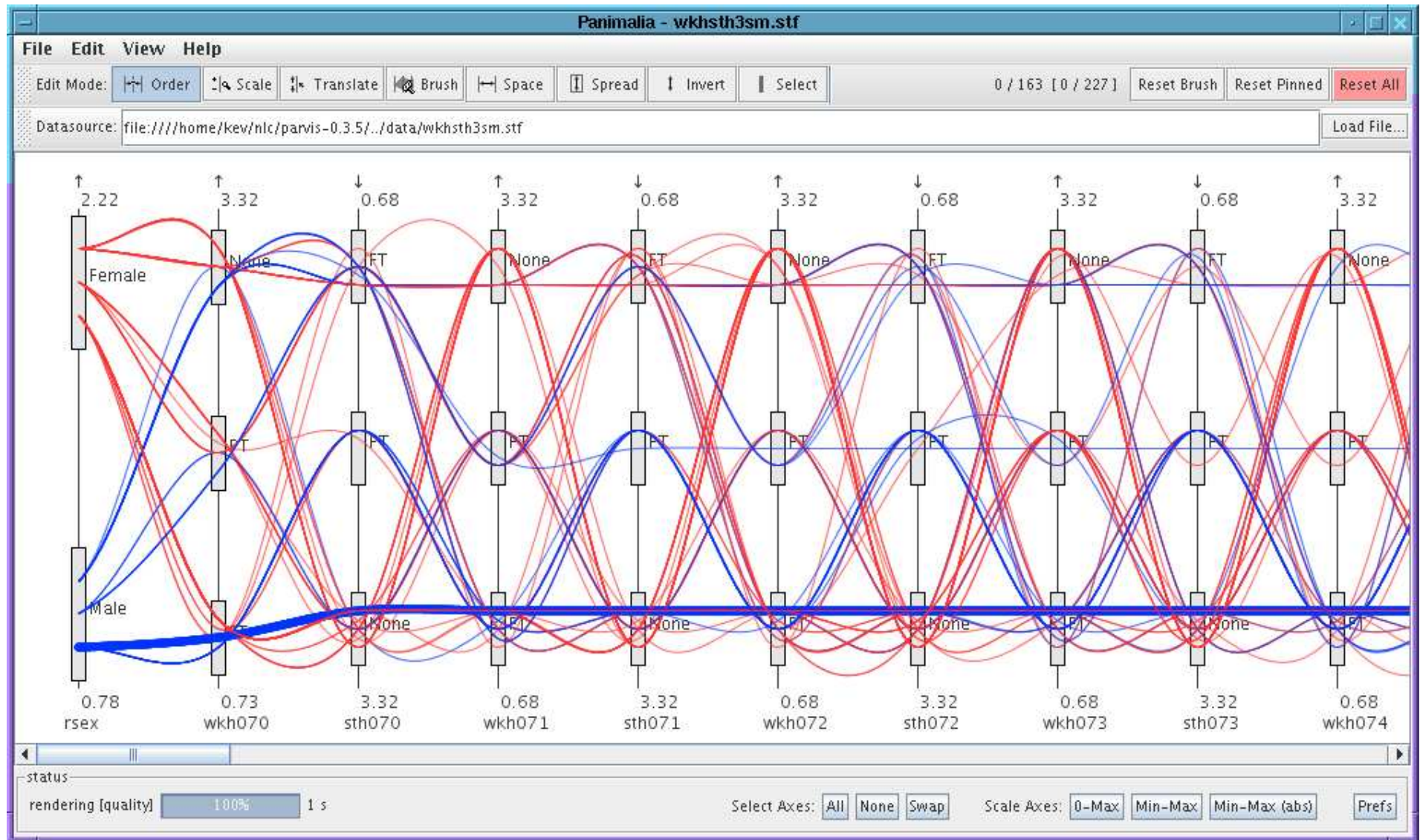


Work/study: interleaved by year

Work/study: interleaved by year

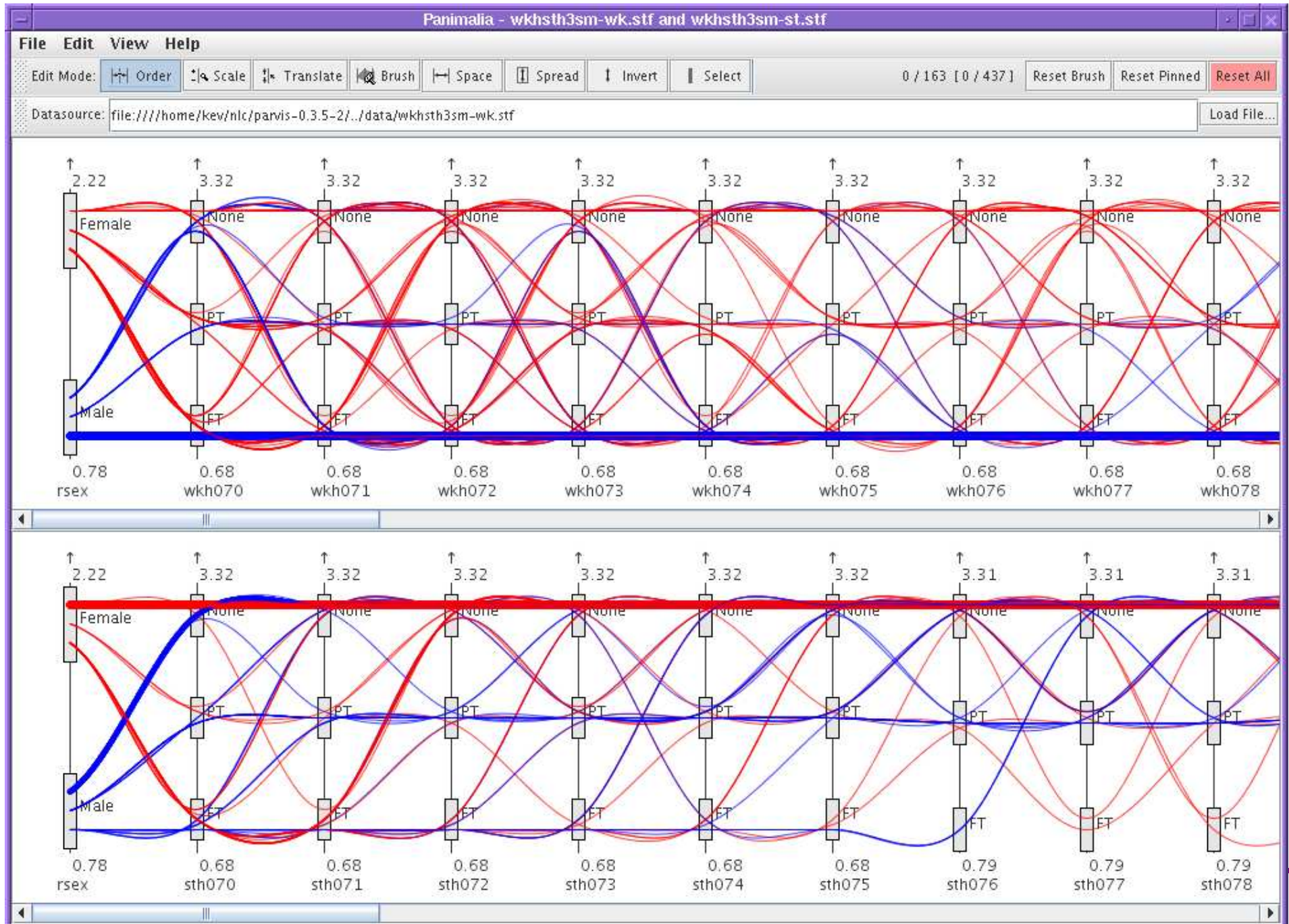


Work/study: interleaved by year



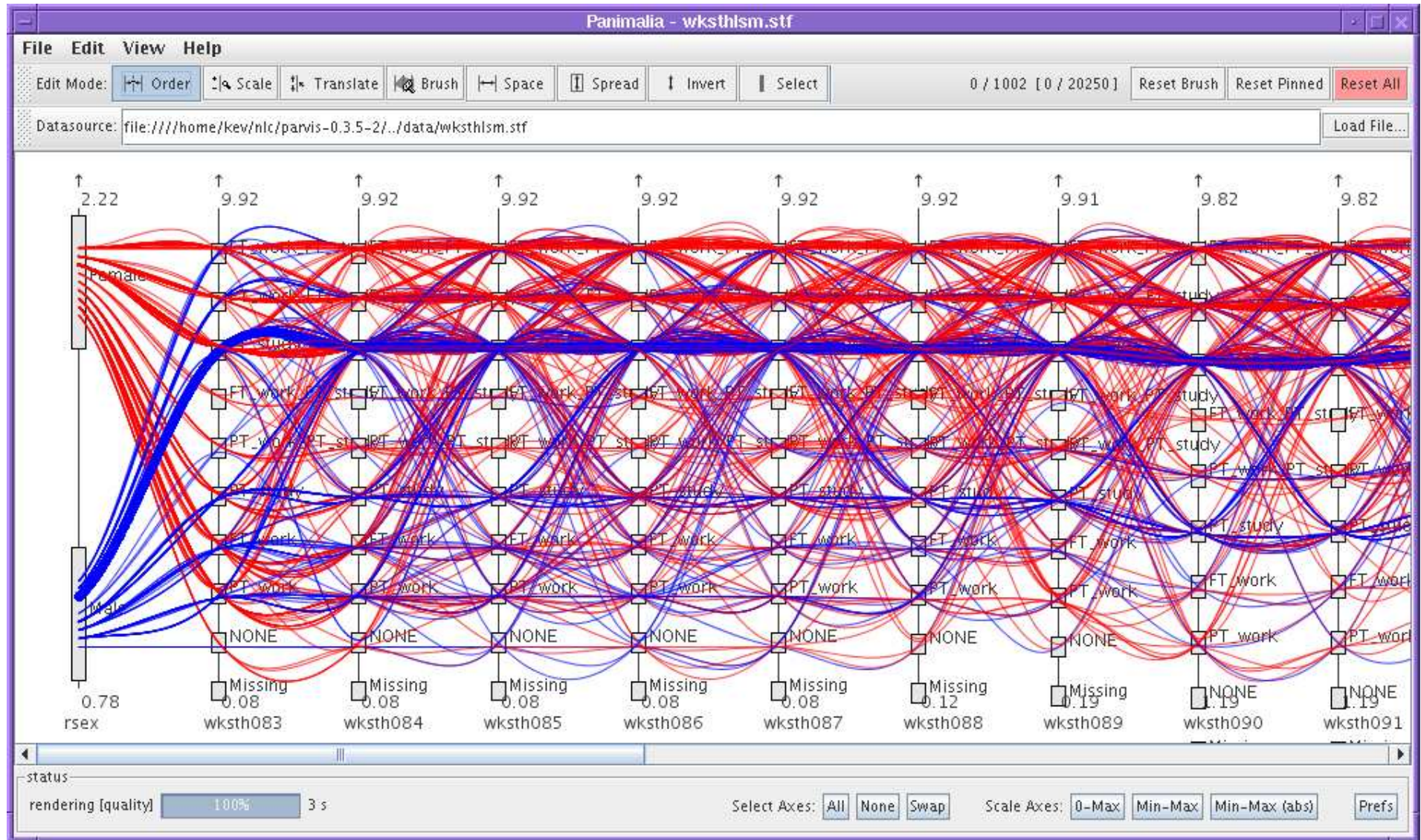
Work/study: separately by year

Work/study: separately by year



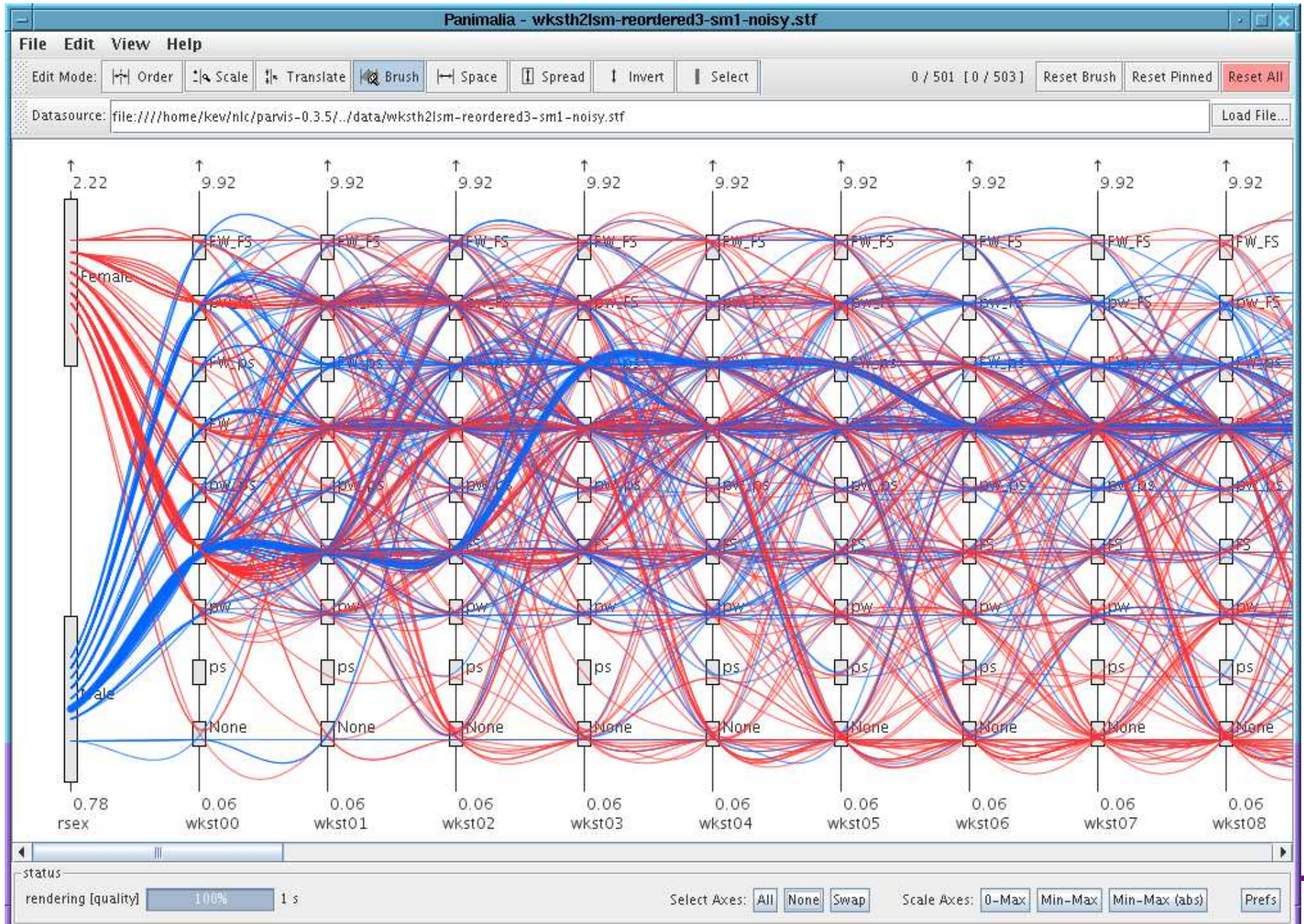
Work/study: combined by year

Work/study: combined by year

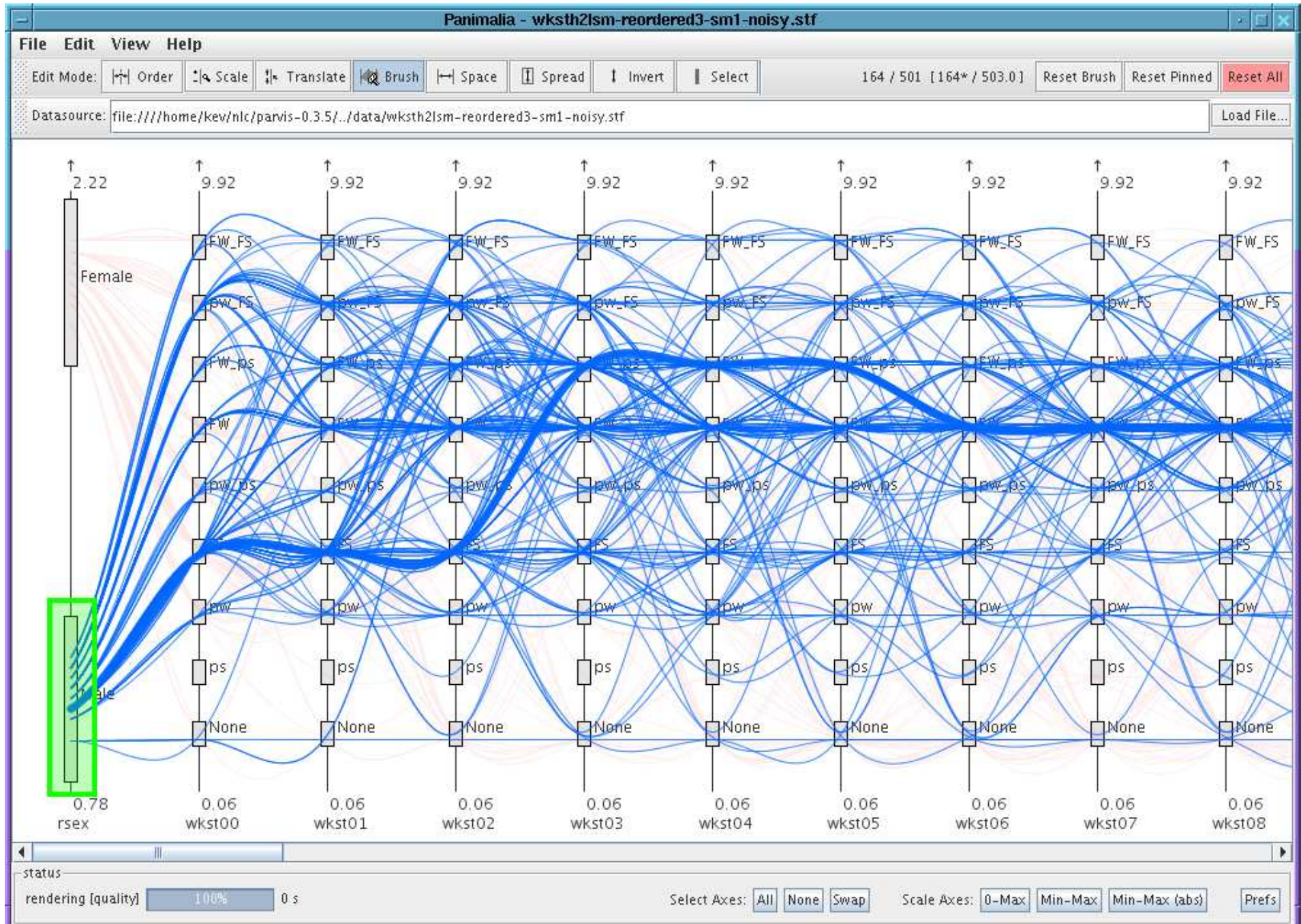


Work/study: combined by age

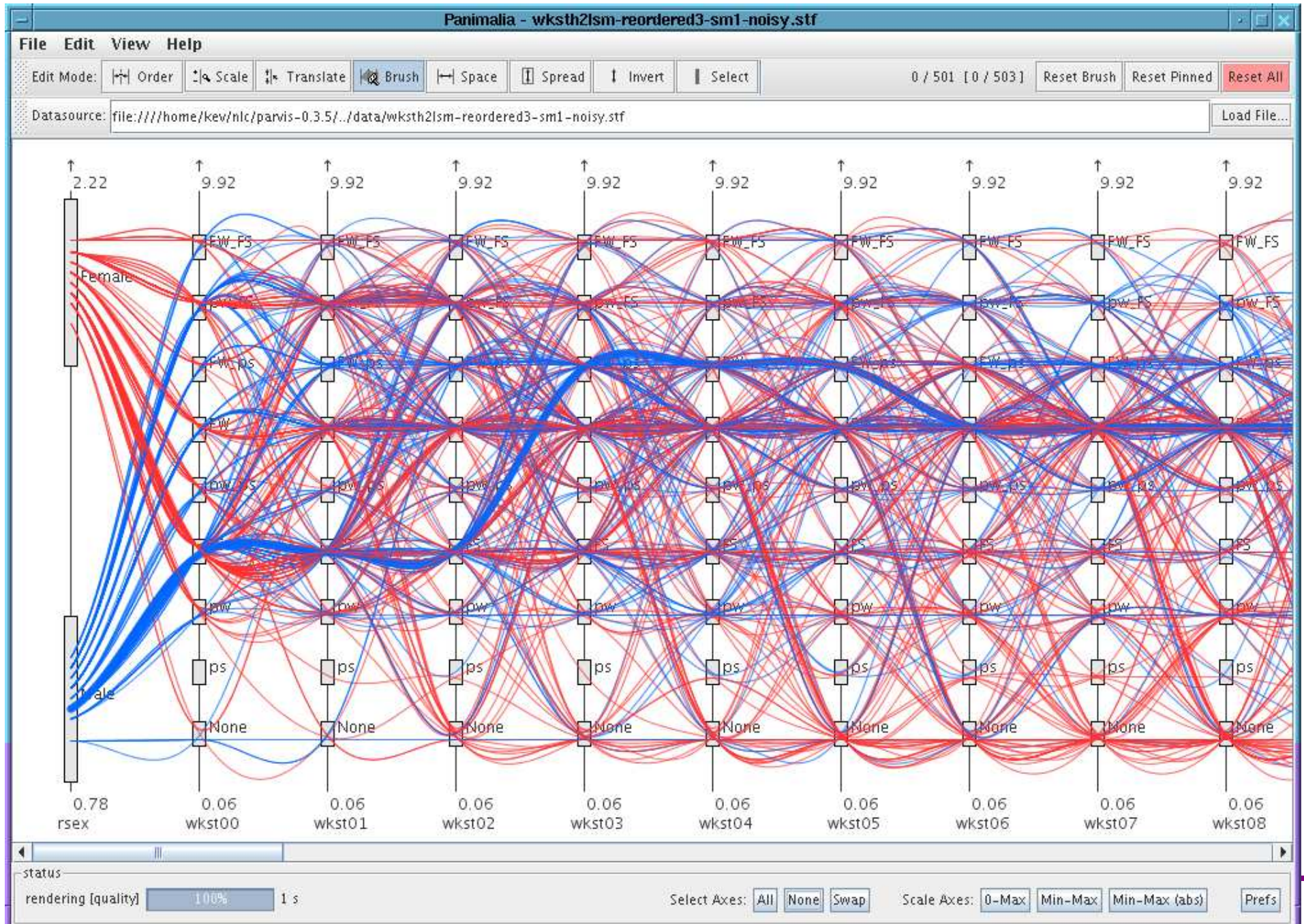
Work/study: combined by age



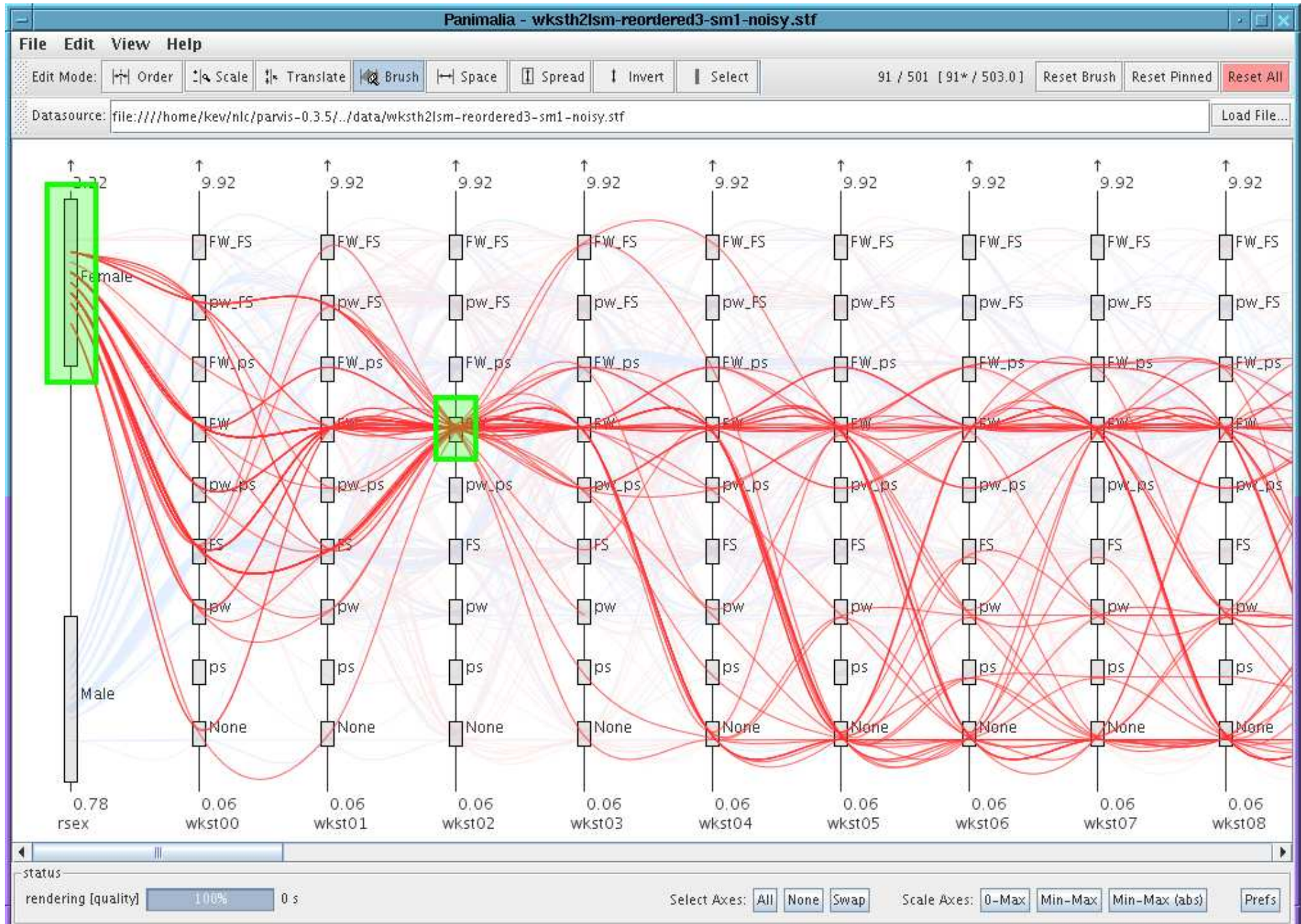
Work/study: combined by age



Work/study: combined by age

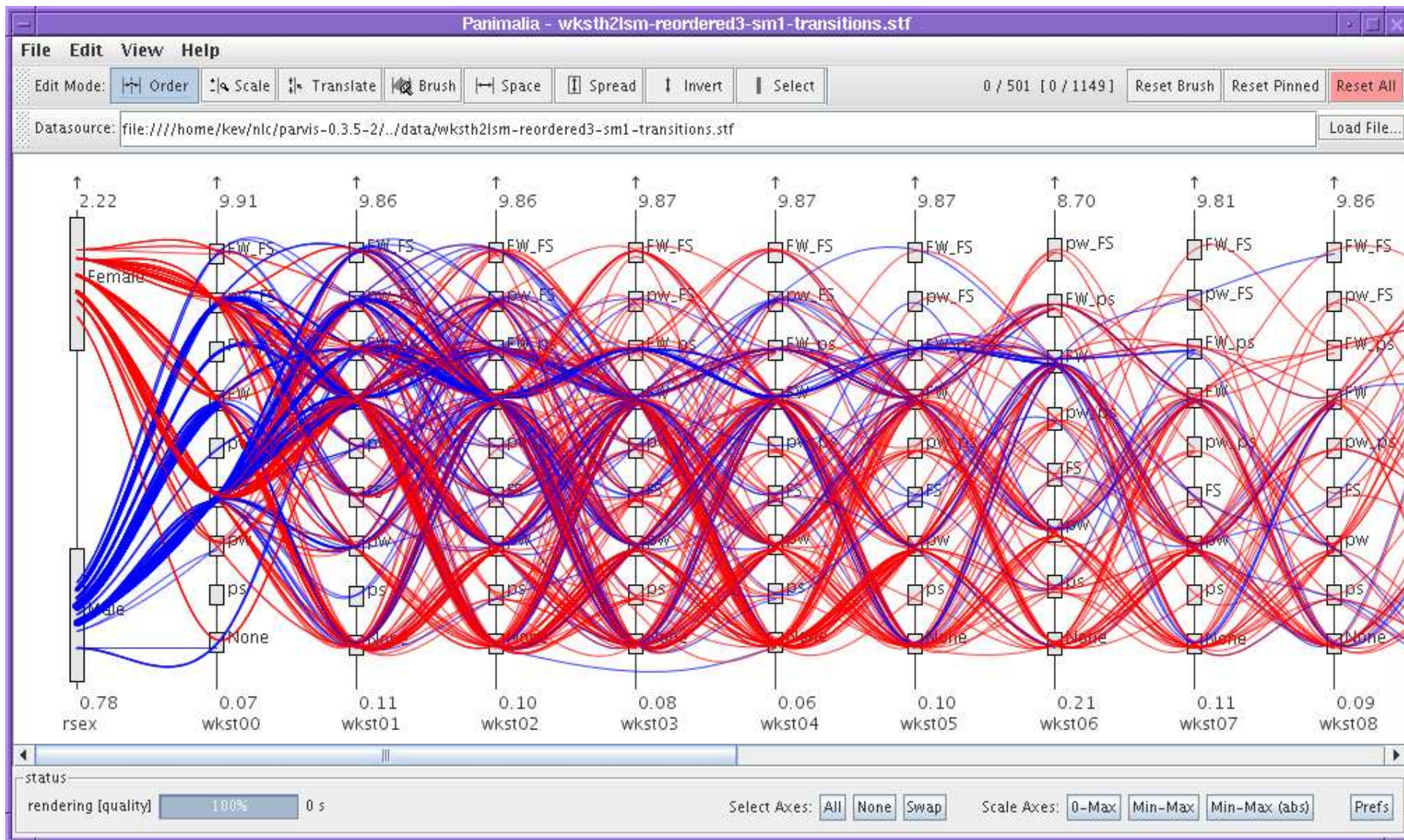


Work/study: combined by age



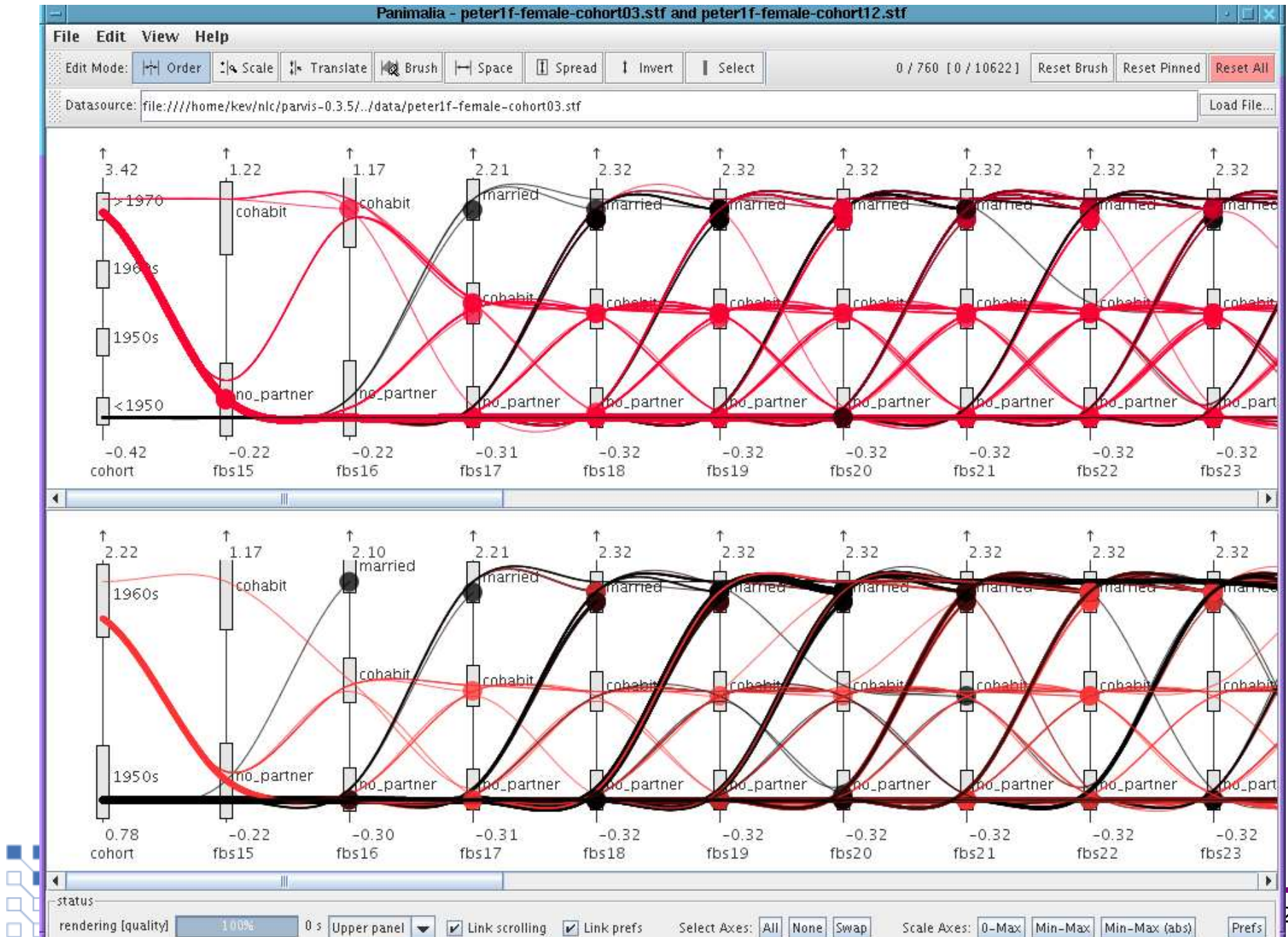
Work/study: combined transitions

Work/study: combined transitions



Relationship path to first birth

Relationship path to first birth



Conclusion

- Advanced visualisation is a fertile area
- Parallel Coordinate Plots
 - High dimensional data (esp. longitudinal)
 - Direct
 - Interactive
- Supports data exploration and familiarisation
- Useful for both researchers and data archivists
- Sample application to Negotiating the Life Course

Questions? Comments?
