

RnavGraph

RnavGraph and the tk canvas widget

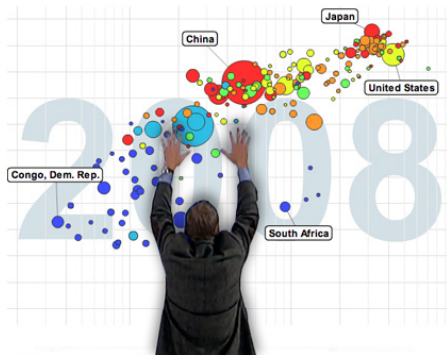
Adrian Waddell & R. Wayne Oldford

use  Conference in Warwick

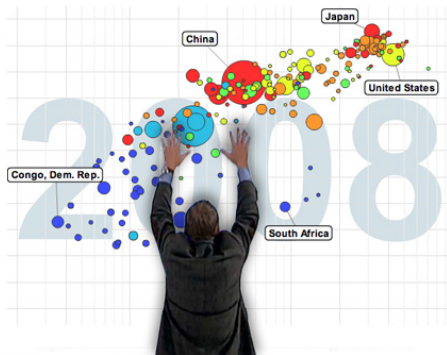
August 18, 2011

Overview

- ▶ R package on CRAN
- ▶ interactive data visualization
- ▶ data clustering



- ▶ Over 500 variables available
 - ▶ We look at 6 of them

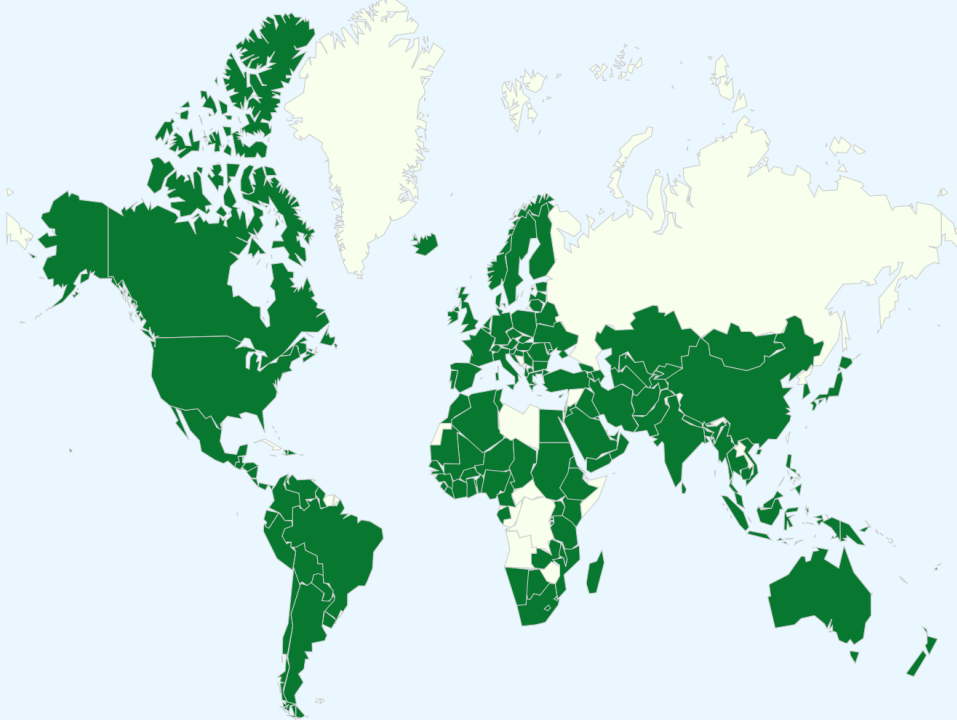


- ▶ Over 500 variables available
 - ▶ We look at 6 of them
 - ▶ RnavGraph could accommodate all 500 variables

Example Country Data

Shortname	Description
Murder	Murder, age adjusted, per 100,000.
Trdeath	Traffic mortality, age adjusted, per 100,000.
Frtly	Children per woman (total fertility).
Mort	Under-five mortality rate (per 1000 live births).
Age	Mean age at 1st marriage of women.
Suic	Suicide among women, per 100,000 standard population.

- ▶ average over years from 2004 to 2009
- ▶ 157 observations (countries)



Data Visualization



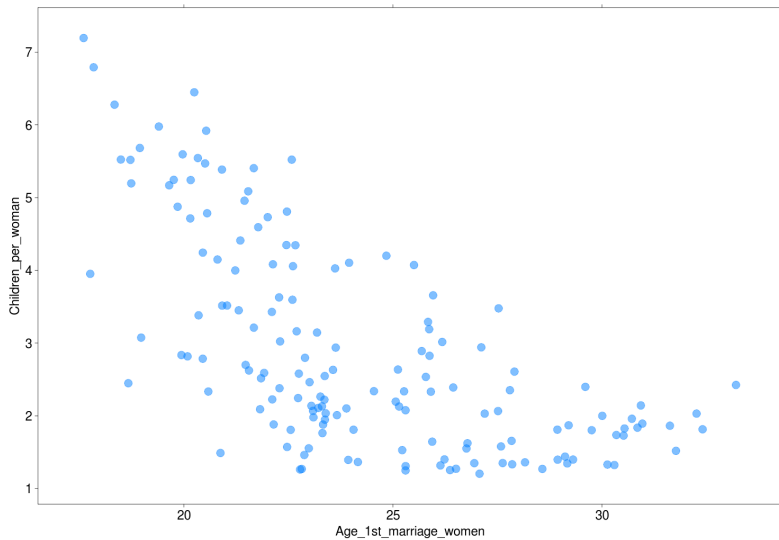
Scatter Plot Matrix

Data Visualization



Scatter Plot Matrix

Data Visualization



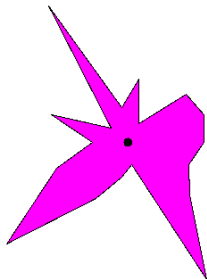
In-View Interactivity

01_tk2d.ogv

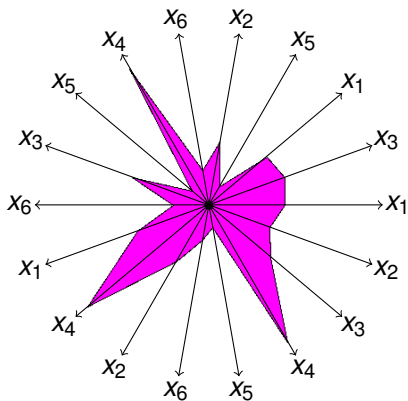
Star Glyphs

02_glyphs.ogv

Star Glyph for Iraq



Star Glyph for Iraq



- x_1 Age_1st_marriage_women
- x_2 Suicide_women
- x_3 Children_per_woman
- x_4 Traffic_deaths
- x_5 Murder
- x_6 Child_mortality

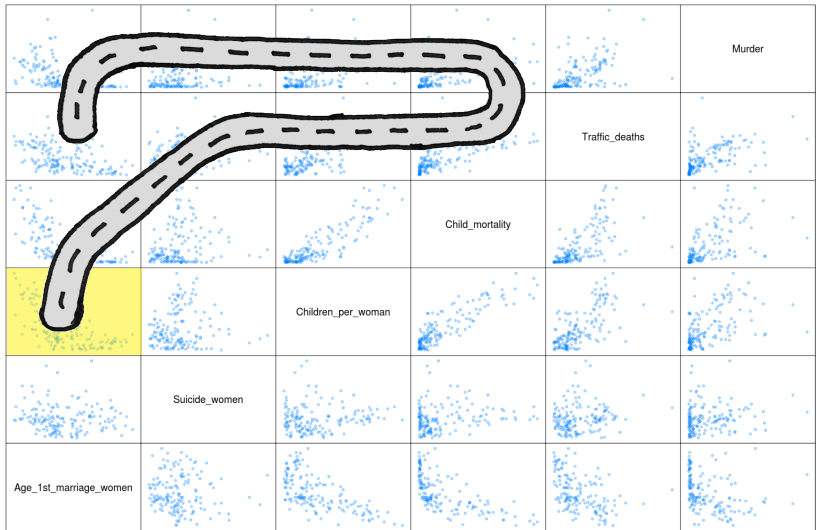
- every pair-wise comparison

Data Visualization



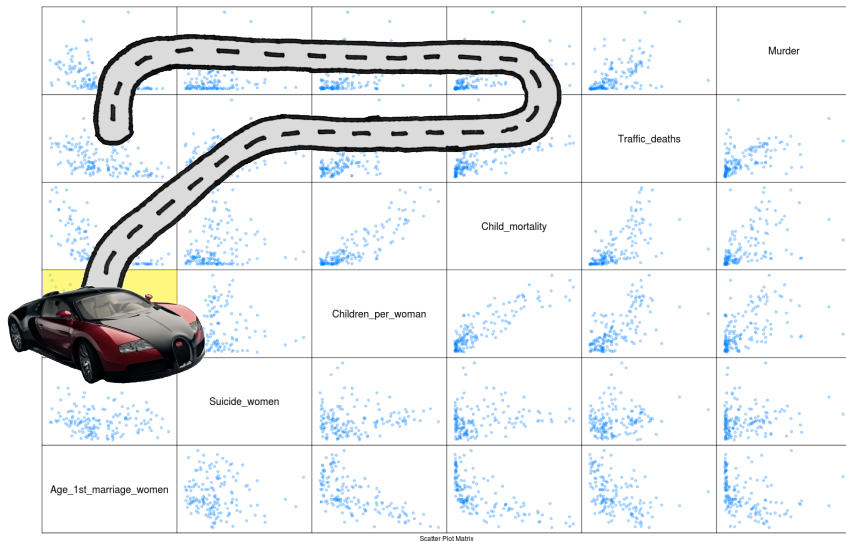
Scatter Plot Matrix

Data Visualization

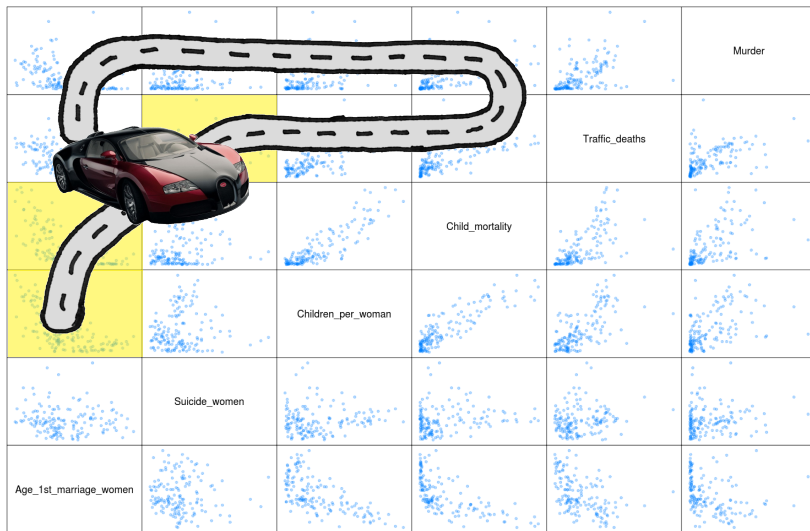


Scatter Plot Matrix

Data Visualization

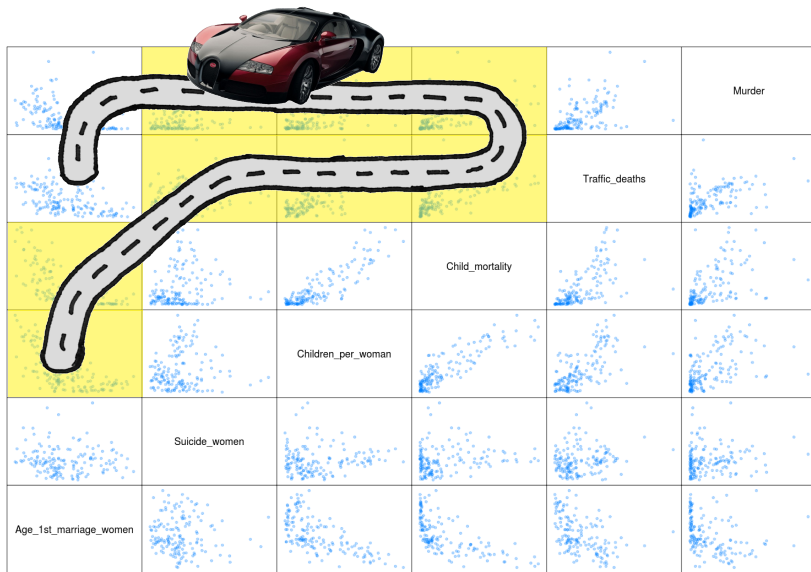


Data Visualization



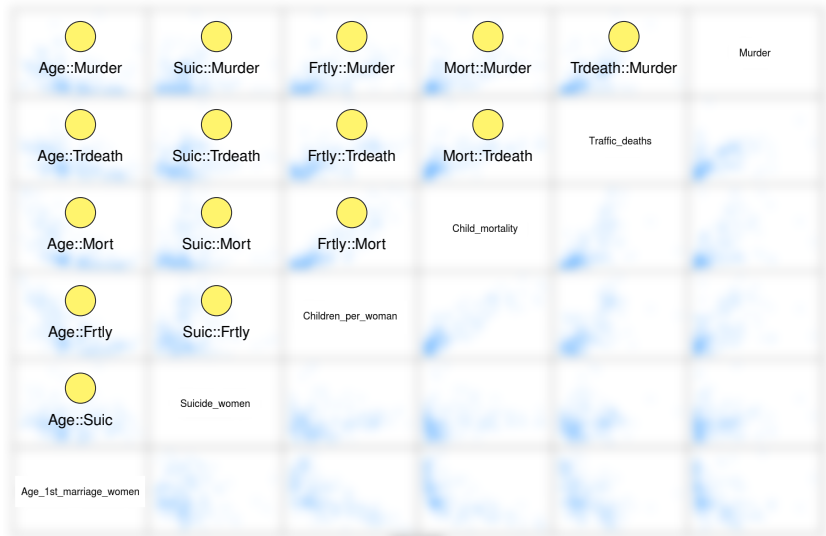
Scatter Plot Matrix

Data Visualization

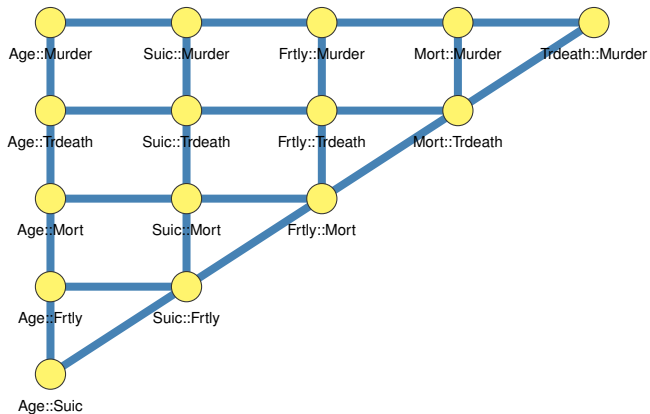


Scatter Plot Matrix

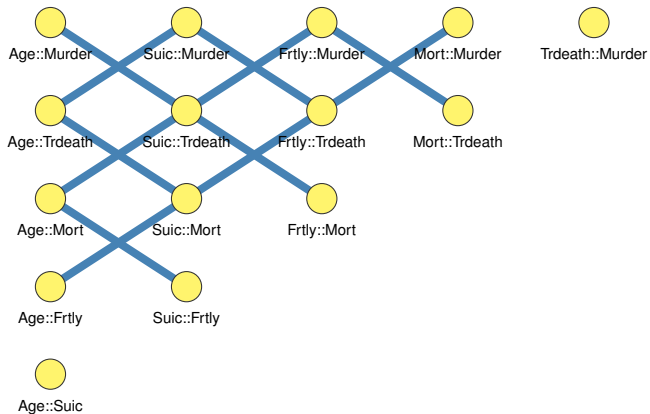
Data Visualization



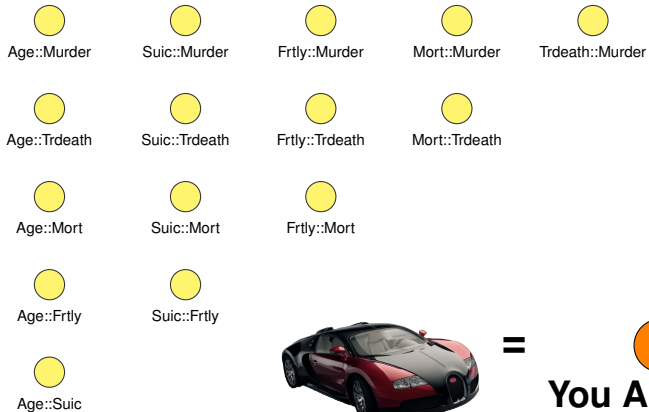
Map of some 3D Transitions



Map of some 4D Transitions

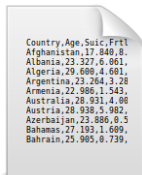


Data Visualization



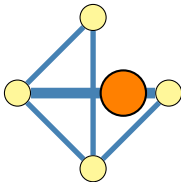
Recap

DATA

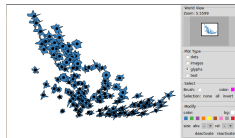


Country	Age	Suic	Frt1
Afghanistan	17.840	8.	
Albania	23.327	6.061	
Algeria	29.600	4.601	
Argentina	23.264	3.28	
Armenia	22.905	1.543	
Australia	28.031	4.00	
Austria	28.930	5.982	
Azerbaijan	23.886	0.5	
Bahamas	27.193	1.609	
Bahrain	25.905	0.739	

GRAPH

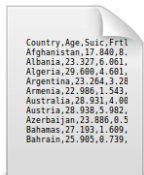


VIEW



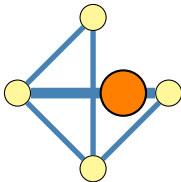
Recap

DATA

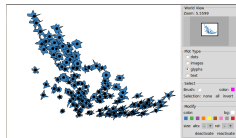


Country	Age	Suic	Frt1
Afghanistan	17.840	8.	
Albania	23.327	6.061	
Algeria	29.600	4.601	
Argentina	23.264	3.28	
Armenia	22.905	1.543	
Australia	28.031	4.00	
Austria	28.930	5.982	
Azerbaijan	23.886	0.5	
Bahamas	27.193	1.609	
Bahrain	25.905	0.739	

GRAPH



VIEW



RnavGraph

03_rnavgraoh.ogv

And now?

- ▶ Interactive Visual Clustering
- ▶ Data Mining
- ▶ 2D Image Scatterplots
- ▶ Look at structures resulting from dimensionality reduction algorithms
- ▶ Give new (visual) meaning to edges and nodes
 - ▶ RnavGraph can be easily extended
 - ▶ Read the vignette

Part II: Implementation

► The tk canvas widget. In R:

```
library(tcltk) # Thank you Peter Dalgaard

tt <- tktoplevel()
tktitle(tt) <- "How it all began!"
can <- tkcanvas(tt, bg = 'yellow')
tkpack(can, expand = TRUE, fill = 'both')

tkcreate(can, 'oval', 20, 20, 50, 50, fill = 'blue',
  tag = 'bullet')

tkitembind(can, 'bullet', '<B1-Motion>' ,
  function(x,y) {
    x <- as.numeric(x); y <- as.numeric(y)
    tkcoords(can, 'current', x-15,y-15,x+15,y+15)
  }
)
```

Tk canvas

03_canvas.ogv

- ▶ use pure tcl code (foreach etc.) where possible
 - ▶ easier debugging

```
library(tcltk) # Than
tt <- tktoplevel()
tttitle(tt) <- "How i
can <- tkcanvas(tt, b
tkpack(can, expand =
tkcreate(can, 'oval',
tag = 'bullet')
tkitembind(can, 'bull
function(x,y) {
x <- as.numeric(x
tkcoords(can, 'cur
```

demo.R

VS.

```
library(tcltk)
tt <- tktoplevel()
tcl('set', 'ttc', tt)
tcl("source", "demo2.
```

demo2.R

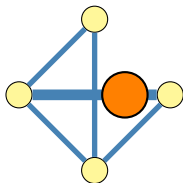
```
wm title $ttc "How it
set can [canvas $ttc.
pack $can -fill both
$can create oval 20 2
$can bind 'bullet' <B
%W coords current [
]
```

demo2.tcl

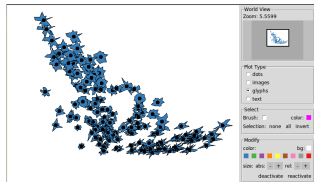
- ▶ use C extensions for tcl
- ▶ check whether re-drawing all elements on canvas is faster than modifying them

How RnavGraph was done

GRAPH



VIEW



- ▶ Mostly R Code
 - ▶ tcl wrapper functions in R
- ▶ plain tcl source
 - ▶ copying data is quick
- ▶ display redraw with C extension

- ▶ Oldford and Waddell
 - ▶ Visual clustering of high-dimensional data by navigating low-dimensional spaces (**ISI Dublin, 2011**)
 - ▶ RnavGraph: A visualization tool for navigating through high dimensional data (**ISI Dublin, 2011**)
 - ▶ RnavGraph R package, available on CRAN

- ▶ Hurley and Oldford
 - ▶ Graphs as a navigational infrastructure for high dimensional data spaces (**Comp Stats 2011**)
 - ▶ Pairwise display of high dimensional information via Eulerian tours and Hamiltonian decompositions (**JCGS, 2010**)
 - ▶ Eulerian tour algorithms for data visualization and the PairViz R package (**Comp Stats 2011**)
 - ▶ **PairViz** R package, available on CRAN

Murder, age adjusted, per 100,000

Combination of time series from WHO Violence and Injury Prevention (VIP); data from WHO Global Burden of Disease 2002 and 2004; police data from UNODC, Public Health data from UNODC and data from the Injury Mortality Data Collection of the GBD Injury Expert Group (GIMD).

Traffic mortality, age adjusted, per 100,000

Combination of time series from WHO Violence and Injury Prevention (VIP); data from WHO Global Burden of Disease 2002 and 2004; and WHO Global Road Safety Status Report.

Children per woman (total fertility)

Our “children per woman” indicator contains historic total fertility rate (TFR) data for 195 countries in the world, for the 1800-2008 period within their current borders. The main source of the data is the UN fertility dataset, which covers the period between 1950 and 2008. In addition, it has been complemented with data collected from statistical sources and various published articles.

... but users are warned that it contains both good quality observations and rough estimates. However, it is easy to distinguish between them.

Under-five mortality rate (per 1000 live births)

Source: CME www.childmortality.org

CME Info is a database containing the latest child mortality estimates based on the research of the Interagency Group for Child Mortality Estimation.

Mean age at 1st marriage of women

From the Gapminder Documentation:

Different sources.

... we have also included very rough estimates for countries and territories for which reliable data was not available.

Singulate mean age at first marriage (SMAM) ... is one of the most common measurements used. Hence, we try to use this measurement to the extent possible.

The SMAM measures the average number of years a woman would live as “never married”, if she has the same age specific “risks” of marriage as the population at large. Women that never married before their 50 year birthday are typically excluded. The measurement is typically calculated from a census or survey for a given year. The calculation use the proportion women that are “never married” in each age group.

It is also quite common to assume that no-one marries before the age of 15. In populations where child marriages are common this introduces a serious bias in the data. We tried to take that into account when we thought that such a bias were significant.

Suicide among women, per 100,000 standard population

Mortality due to self-inflicted injury among women, per 100 000 standard population, age adjusted.

WHO: Combination of time series from WHO Violence and Injury Prevention (VIP) and data from WHO Global Burden of Disease 2004.

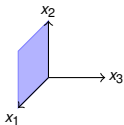
Age-Adjusted Rates

From www.health.state.ny.us/diseases/chronic/ageadj.htm

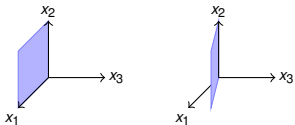
Age-adjustment is a statistical process applied to rates of disease, death, injuries or other health outcomes which allows communities with different age structures to be compared.

Age confounding occurs when the two populations being compared have different age distributions and the risk of the disease or outcome varies across the age groups. The process of age-adjustment by the direct method changes the amount that each age group contributes to the overall rate in each community, so that the overall rates are based on the same age structure. Rates that are based on the same age distribution can be compared to each other without the presence of confounding by age.

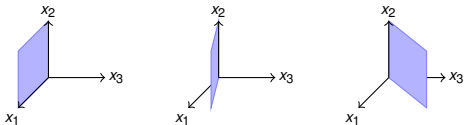
3d rigid rotation $(x_1 : x_2) \rightarrow (x_3 : x_2)$



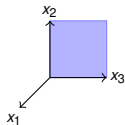
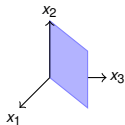
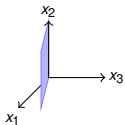
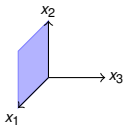
3d rigid rotation $(x_1 : x_2) \rightarrow (x_3 : x_2)$



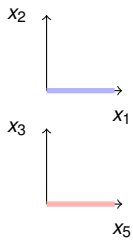
3d rigid rotation $(x_1 : x_2) \rightarrow (x_3 : x_2)$



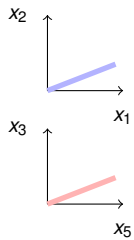
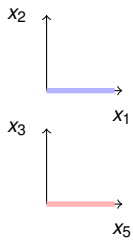
3d rigid rotation $(x_1 : x_2) \rightarrow (x_3 : x_2)$



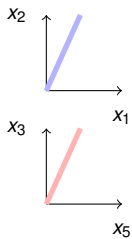
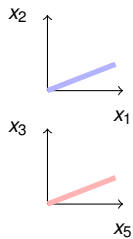
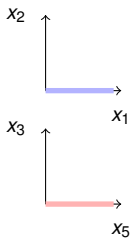
4d transition $(x_1 : x_5) \rightarrow (x_2 : x_3)$



4d transition $(x_1 : x_5) \rightarrow (x_2 : x_3)$



4d transition $(x_1 : x_5) \rightarrow (x_2 : x_3)$



4d transition $(x_1 : x_5) \rightarrow (x_2 : x_3)$

