The Textile Plot

A Key to Interaction Through Data

Kumasaka Natsuhiko
Fundamental Sci & Tech
Keio Univ

Ritei Shibata
Dept of Math
Keio Univ

Overview

- What is Textile Plot?
- Possible enhancements of Textile Plot
- Interactive environment for understanding data through Textile Plot

Textile Plot

Kumasaka and Shibata 2006, 2007

- High dimensional data visualisation according to parallel coordinate system
- Necessary and sufficient background information
- Any kind of data vector (variable) including missing values

Automobile Data

Consumer Reports (USA) 1990, S-PLUS

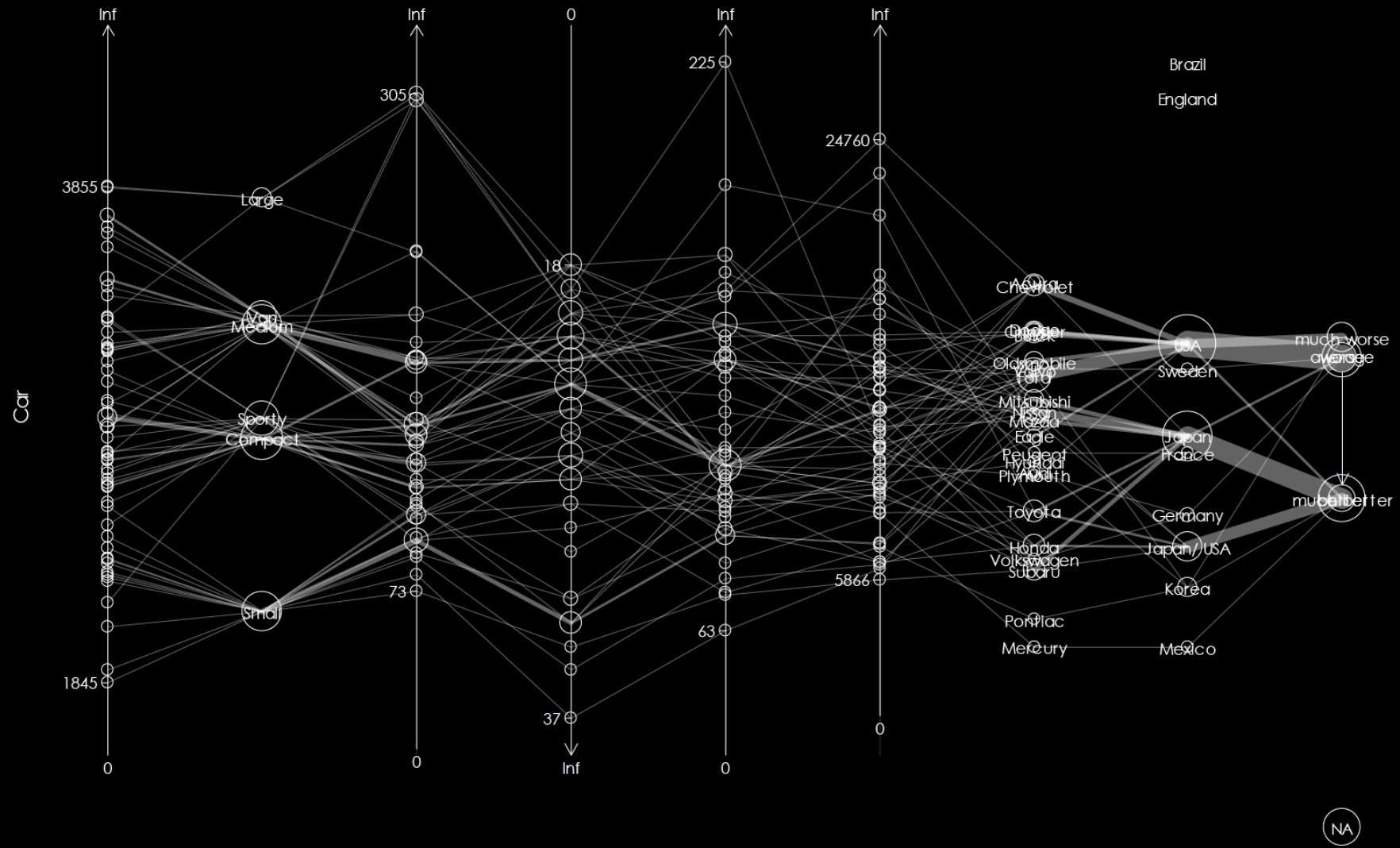
ID	price	country	reliability	type	weight	disp	hp	Manufacturer	mileage
Toyota Corolla	8748	Japan/USA	much better	Small	2390	97	102	Toyota	29
Peugeot 405	15930	France	NA	Compact	2575	116	120	Peugeot	24
Buick Le Sabre V6	16145	USA	average	Large	3325	231	165	Buick	23
Nissan Maxima V6	17899	Japan	much better	Medium	3200	180	160	Nissan	22
Mitsubishi Wagon	14929	Japan	NA	Van	3415	143	107	Mitsubishi	20
Plymouth Laser	10855	USA	NA	Sporty	2840	107	92	Plymouth	26
⋮									
Volvo 240	18450	Sweden	average	Compact	2985	141	114	Volvo	23
Volkswagen Jetta	9995	Germany	average	Small	2330	109	100	Volkswagen	26

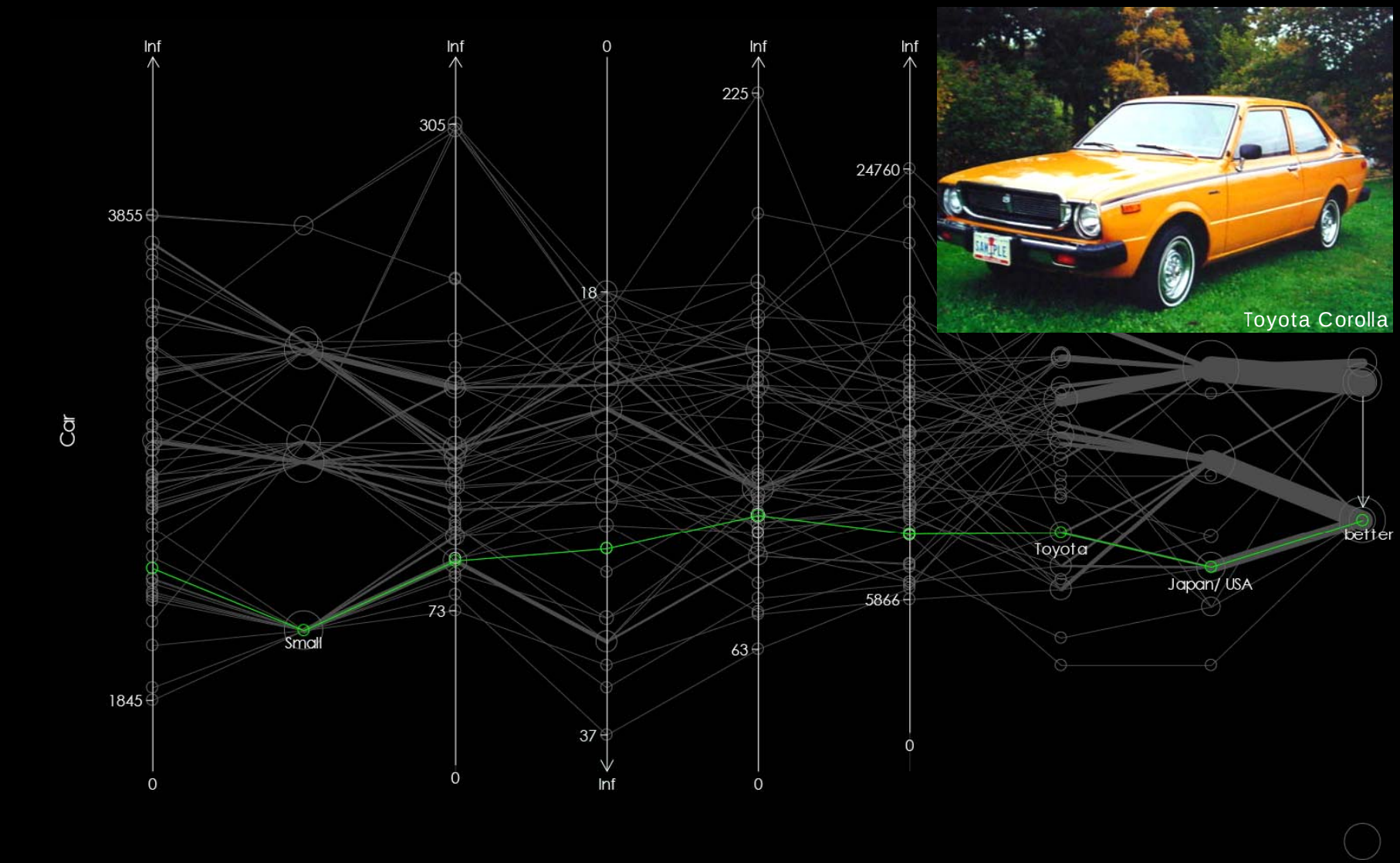
Records

60 cars

Data Vectors (9 dimensional)

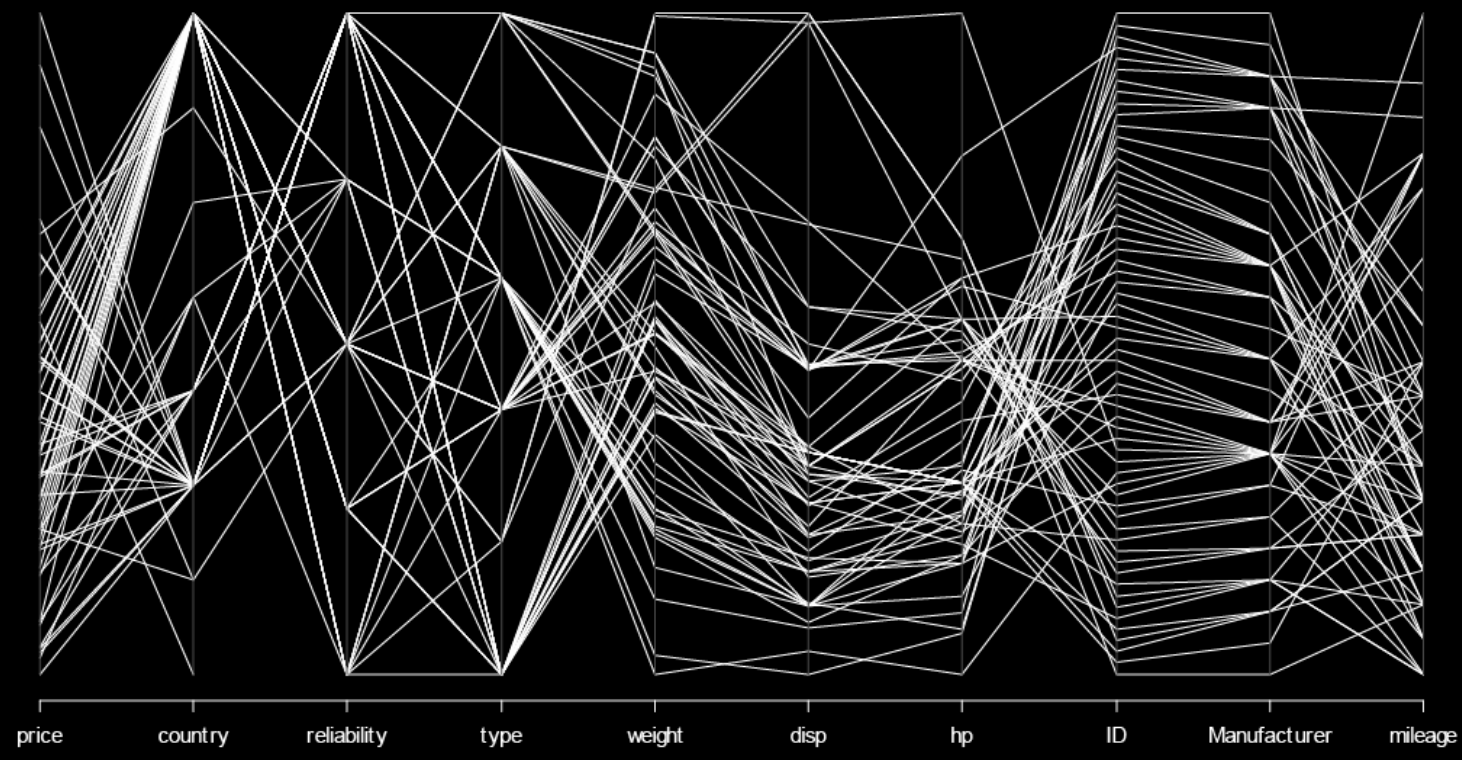
Price
Country
Reliability
Type
Weight
Displacement
Horsepower
Manufacturer
Mileage


Car
Inf
3855
1845
0
Large
Van
Medium
Sporty
Compact
Small
Inf
305
73
0
0
18
37
Inf
Inf
225
63
0
Inf
24760
5866
0
Brazil
England
Chevrolet
Oldsmobile
Mitsubishi
Mazda
Eagle
Peugeot
Plymouth
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
USA
Sweden
Japan
France
Germany
Japan/USA
Korea
Mexico
much worse
average
much better
NA
Weight (pound)
Type
Displacement (cubic inch)
Mileage (miles/ gallon)
Horsepower (hp)
Price (dollars)
Manufacturer
Country
Reliability


Inf
Inf
0
Inf
Inf
225
305
24760
3855
18
Car
Toyota Corolla
Toyota
better
Japan/ USA
5866
73
Small
63
1845
37
0
0
Inf
0
Weight (pound)
Type
Displacement (cubic inch)
Mileage (miles/ gallon)
Horsepower (hp)
Price (dollars)
Manufacturer
Country
Reliability

Parallel Coordinate Plots

Inselberg 1985, Wegman 1990



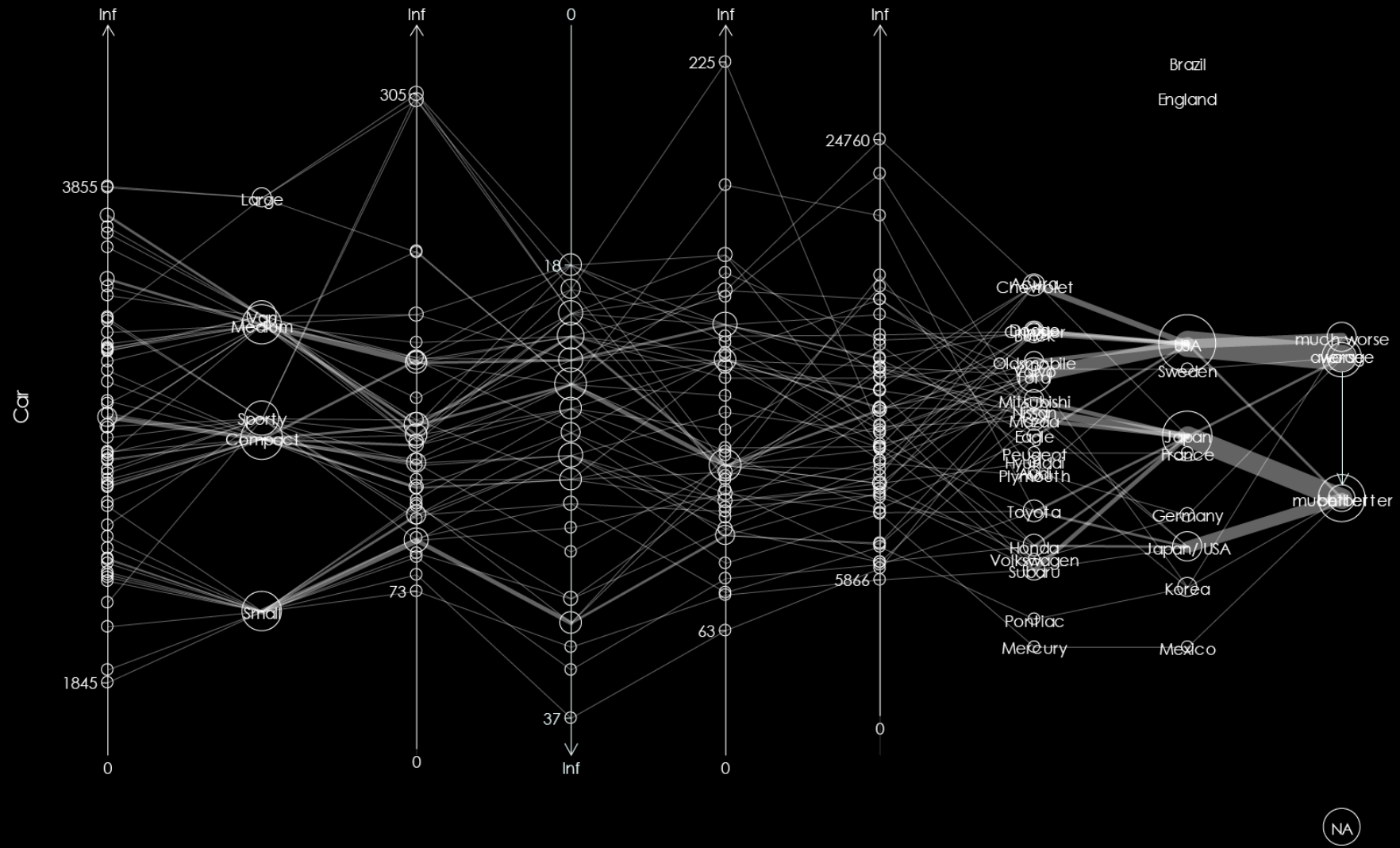
Location and scale of each axis

Independently chosen

Points fill up the range

Levels of categorical data

Equispaced on an axis


Car
Inf
3855
1845
0
Large
Van
Medium
Sporty
Compact
Small
Inf
305
73
0
0
18
37
Inf
Inf
225
63
0
Inf
24760
5866
0
Brazil
England
Chevrolet
Oldsmobile
Mitsubishi
Mazda
Eagle
Peugeot
Plymouth
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
USA
Sweden
Japan
France
Germany
Japan/USA
Korea
Mexico
much worse
average
much better
NA
Weight (pound)
Type
Displacement (cubic inch)
Mileage (miles/ gallon)
Horsepower (hp)
Price (dollars)
Manufacturer
Country
Reliability

Choice of Locations and Scales

Data Vector $\boldsymbol{x}_j$ (n-dimensional)

Categorical

Numerical

$\mathbf{X}_j : n \times (q_j - 1)$ matrix

contrasts $\mathbf{C} = \begin{pmatrix} 0 & \cdots & 0 \\ & \ddots & \vdots \\ 1 & & \\ \vdots & \ddots & 0 \\ 1 & \cdots & 1 \end{pmatrix}$

Unordered

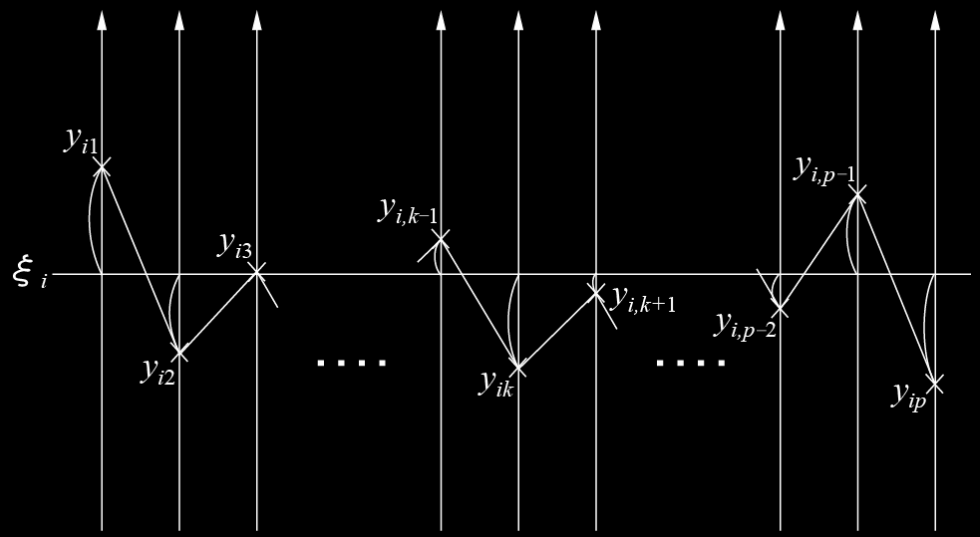
Ordered

$\boldsymbol{y}_j = \alpha_j \mathbf{1} + \beta_j \boldsymbol{x}_j$

$\boldsymbol{y}_j = \alpha_j \mathbf{1} + \mathbf{X}_j \boldsymbol{\beta}_j$

$\boldsymbol{y}_j = \alpha_j \mathbf{1} + \mathbf{X}_j \boldsymbol{\beta}_j$

$\boldsymbol{\beta}_j \leq \mathbf{0}$ or $\geq \mathbf{0}$



Deviations for the ith observation

The sum of squares between **coordinate vectors** and an **ideal coordinate vector** is minimised.

$$\sum_{j=1}^{p} \|\boldsymbol{y}_j - \boldsymbol{\xi}\|^2 \underset{\{\alpha_j, \boldsymbol{\beta}_j\}_{j=1}^{p}, \boldsymbol{\xi}}{\rightarrow} \min$$

SNP Data on Calpain-10

Iwasaki, et al. 2005

Data Vectors (14-dimensional)

Phenotype (affected or non-affected)

12 Genotypes on SNP loci

Genotype on INDEL locus

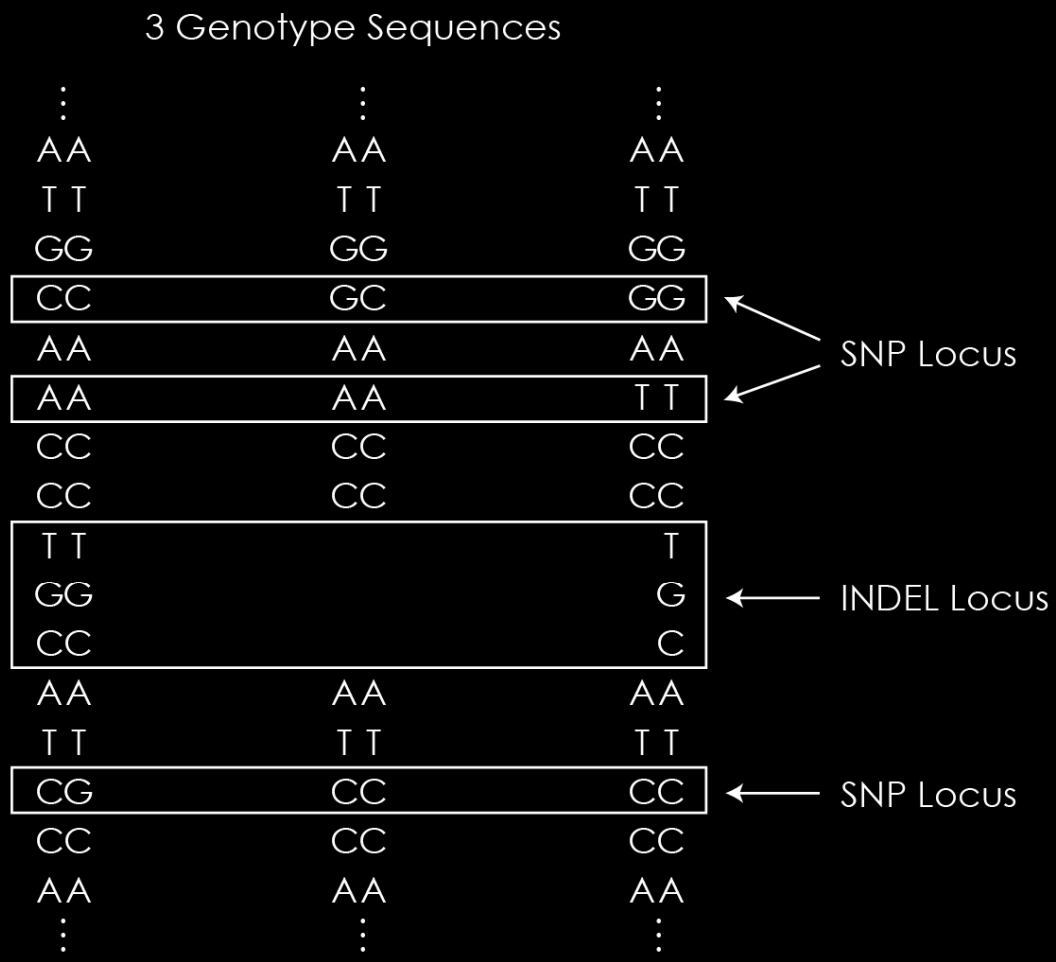
Joint work with

Dr Kamitsuji at Stagen (http://www.stagen.co.jp/)

Goal

To know the danger of a disease from genotypes

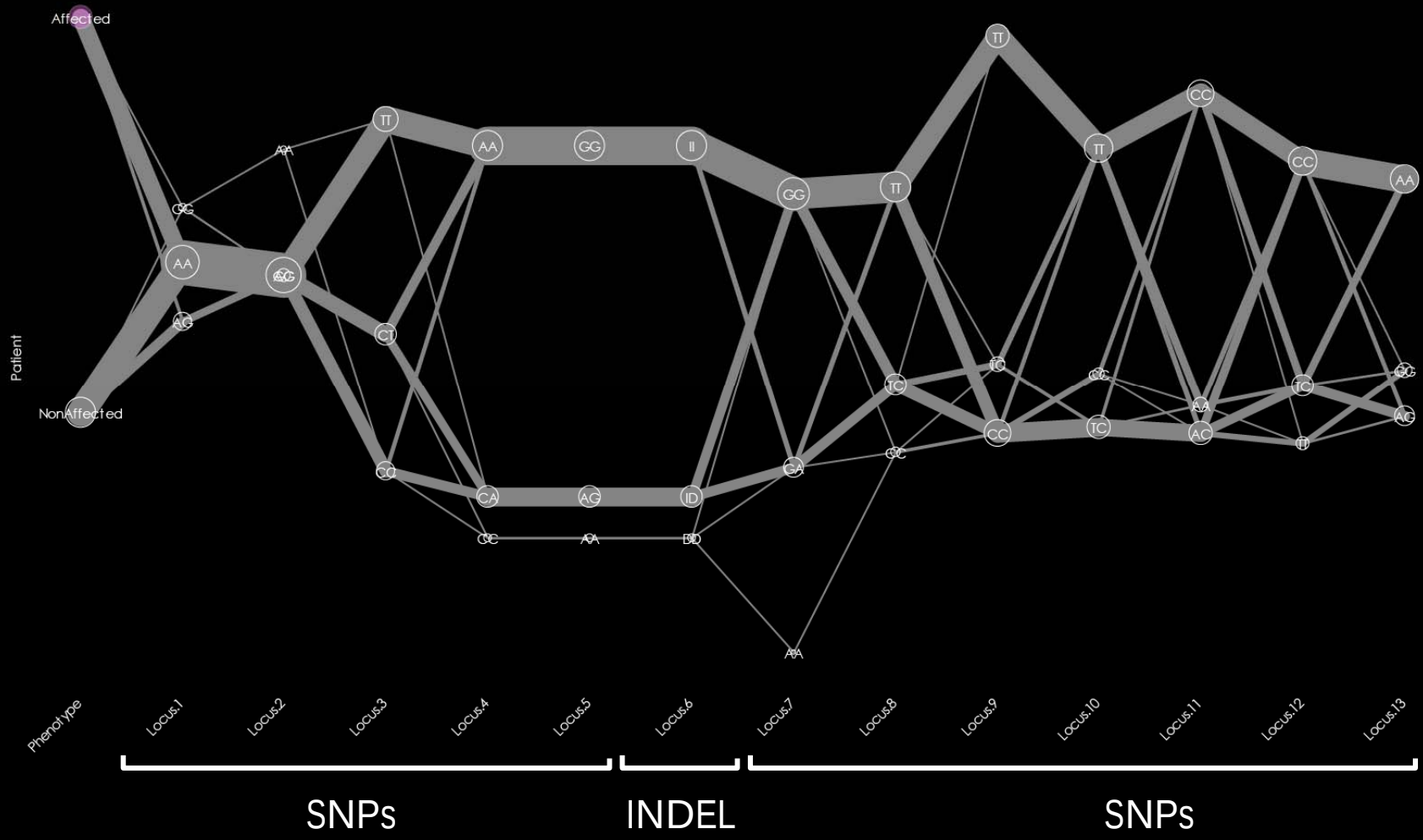
SNP / INDEL Locus



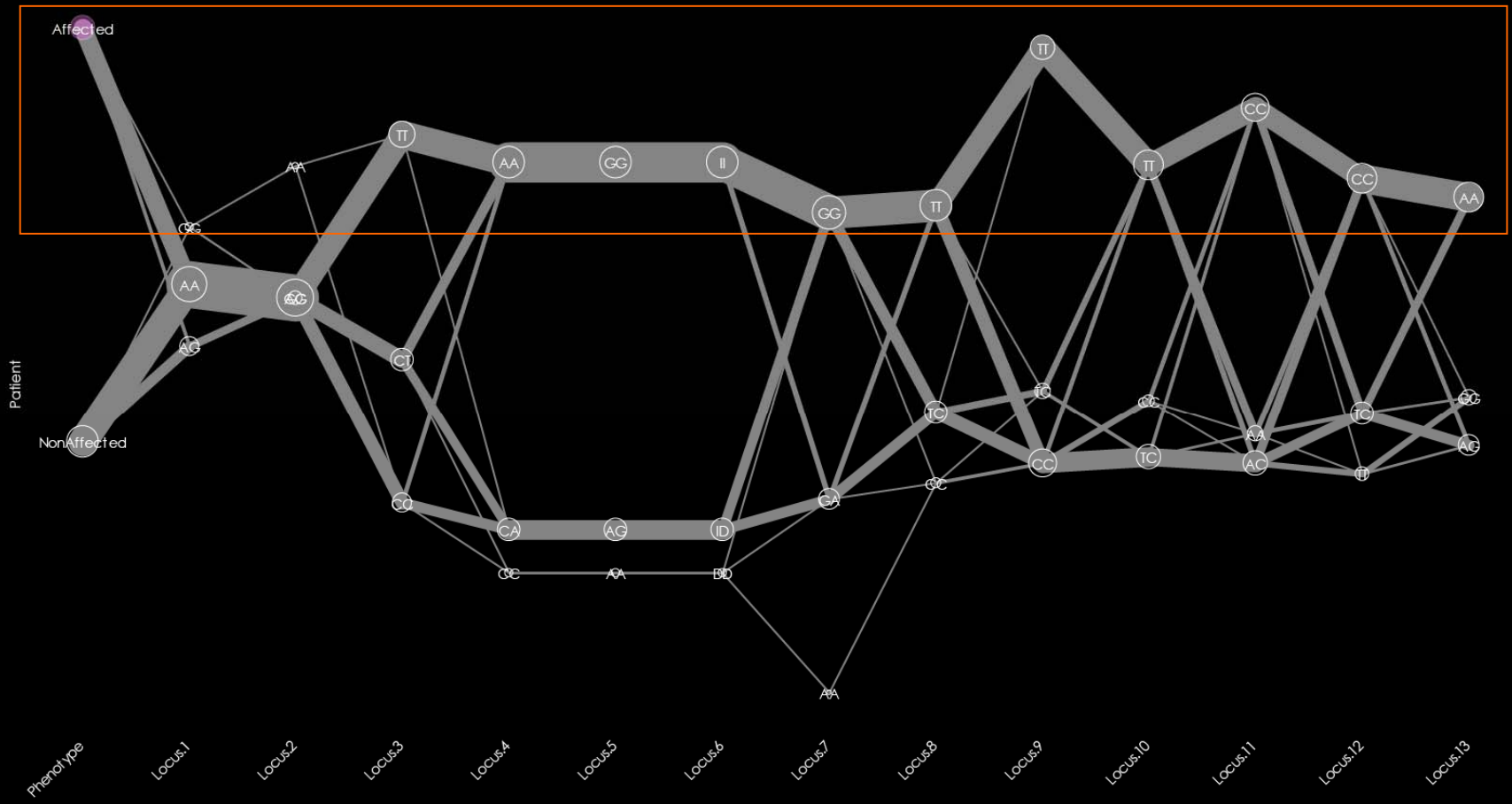


SNP = Single Nucleotide Polymorphism
INDEL = INsertion and DELetion

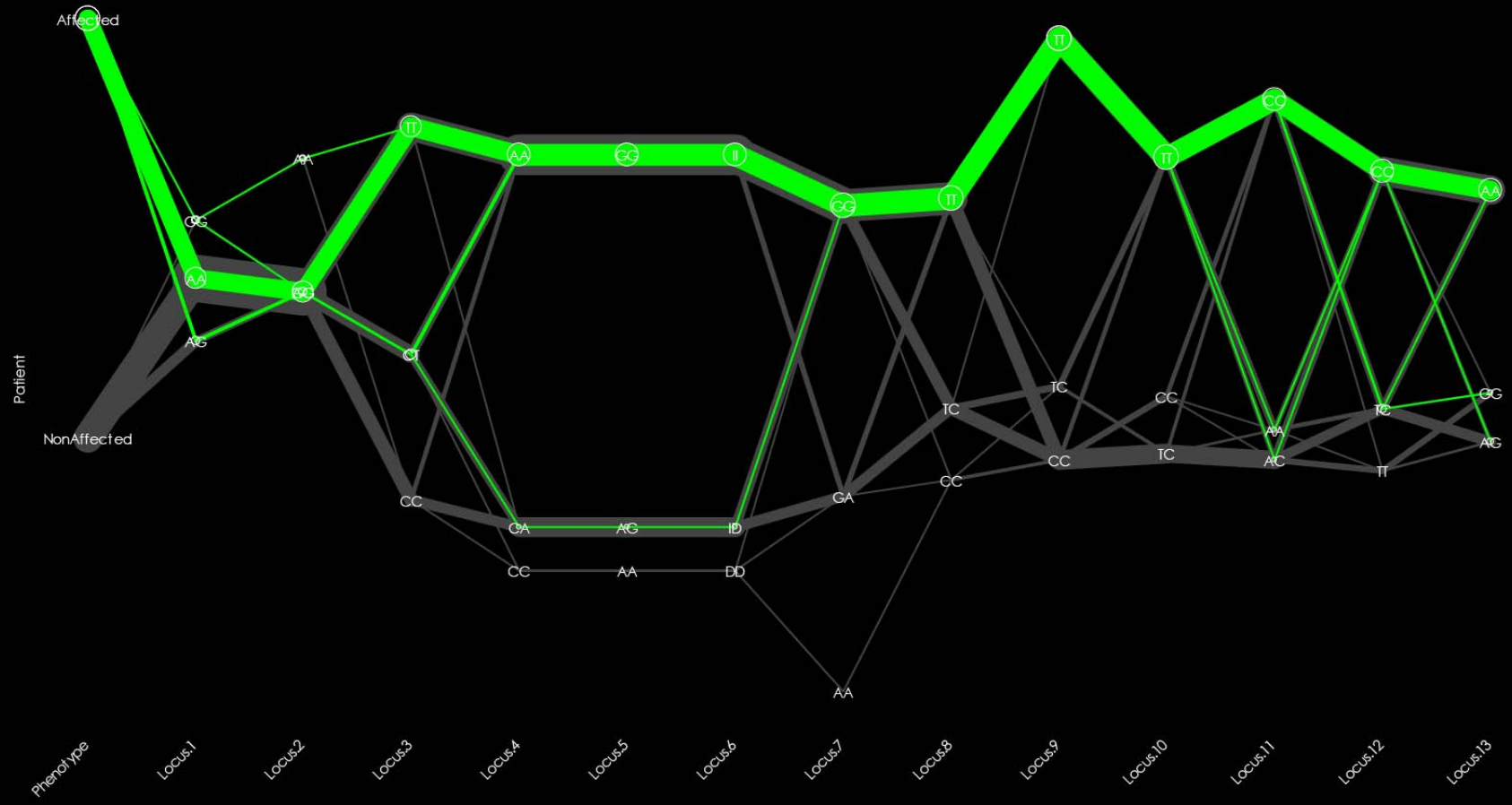
Textile Plot of SNP Data



Recessive Inheritance



Enhancements of Textile Plot



Enhancements of Textile Plot

Transform
- single data vector (log, inverse, etc)
- multiple data vectors (normalisation, etc)

Substitute Data Point

Extract/Remove
- records
- data vectors

Re-weight
- outliers

Order Data Vectors
- a priori
- distance to ideal coordinates
- distance between coordinate vectors

Highlight Connecting Lines

Select Graphical Parameters
- color, transparency, size, etc.

Enhancements of Textile Plot

Transform
- single data vector (log, inverse, etc)
- multiple data vectors (normalisation, etc)

Substitute Data Point

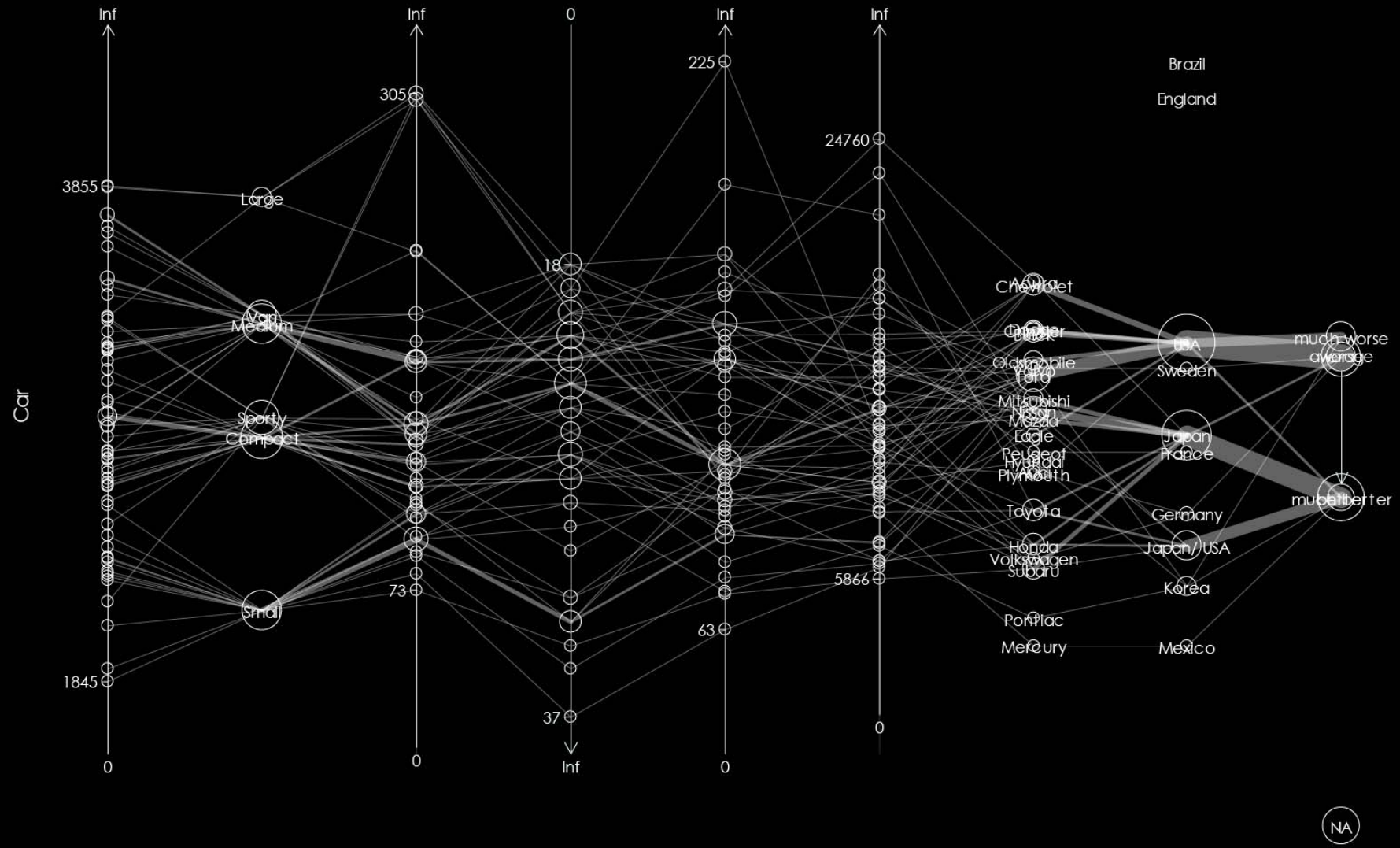
Extract/Remove
- records
- data vectors

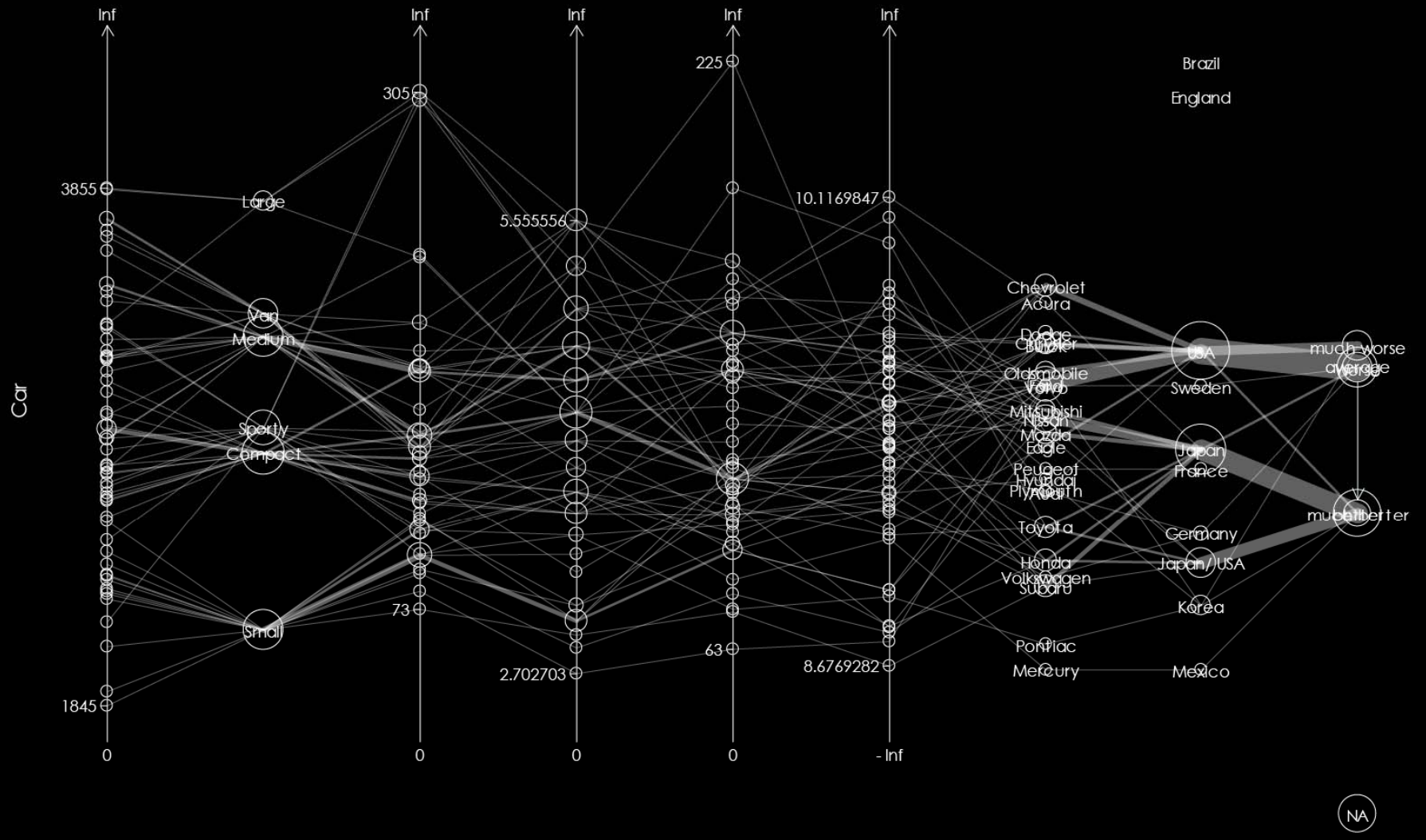
Re-weight
- outliers

Order Data Vectors
- a priori
- distance to ideal coordinates
- distance between coordinate vectors

Highlight Connecting Lines

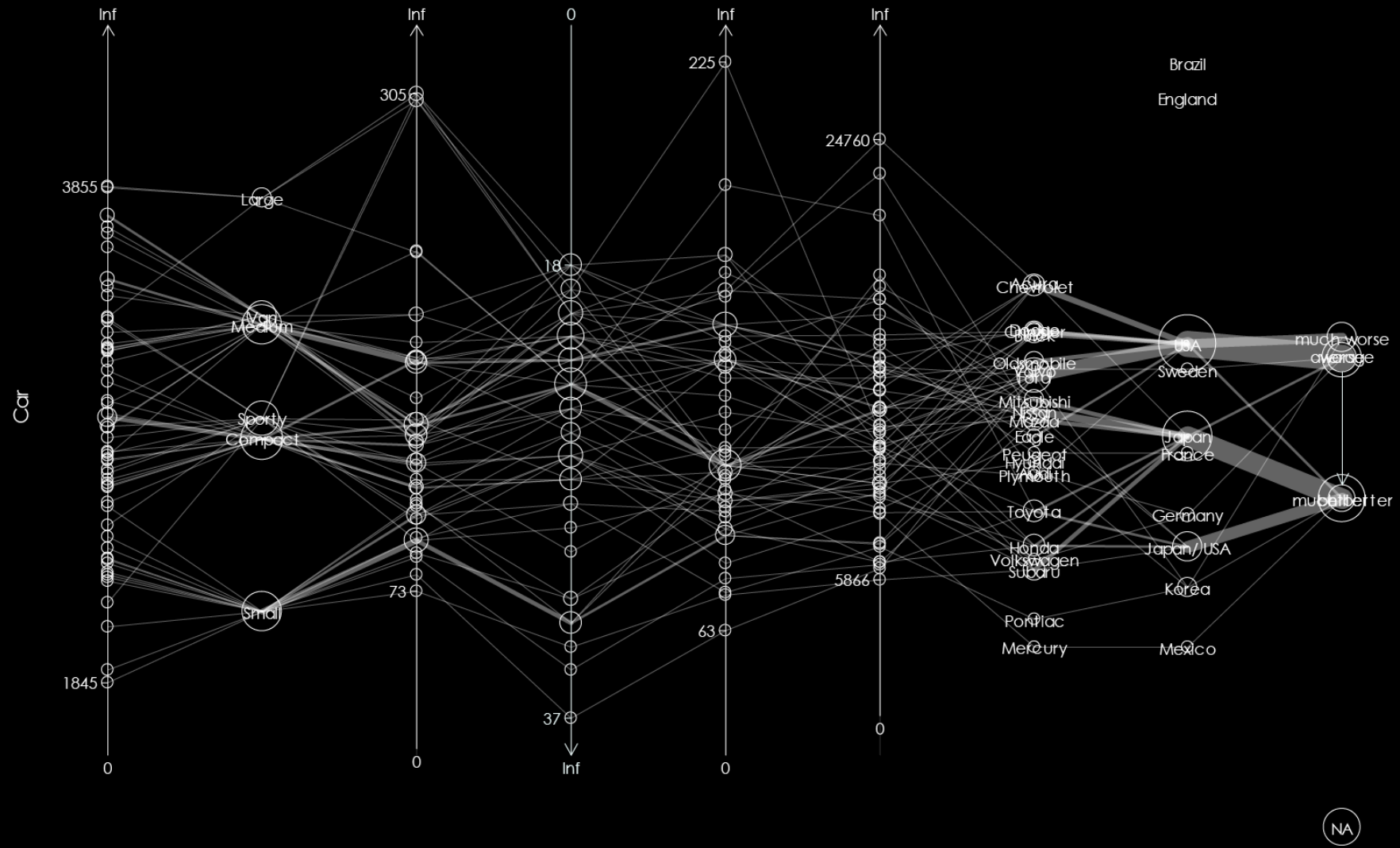
Select Graphical Parameters
- color, transparency, size, etc.

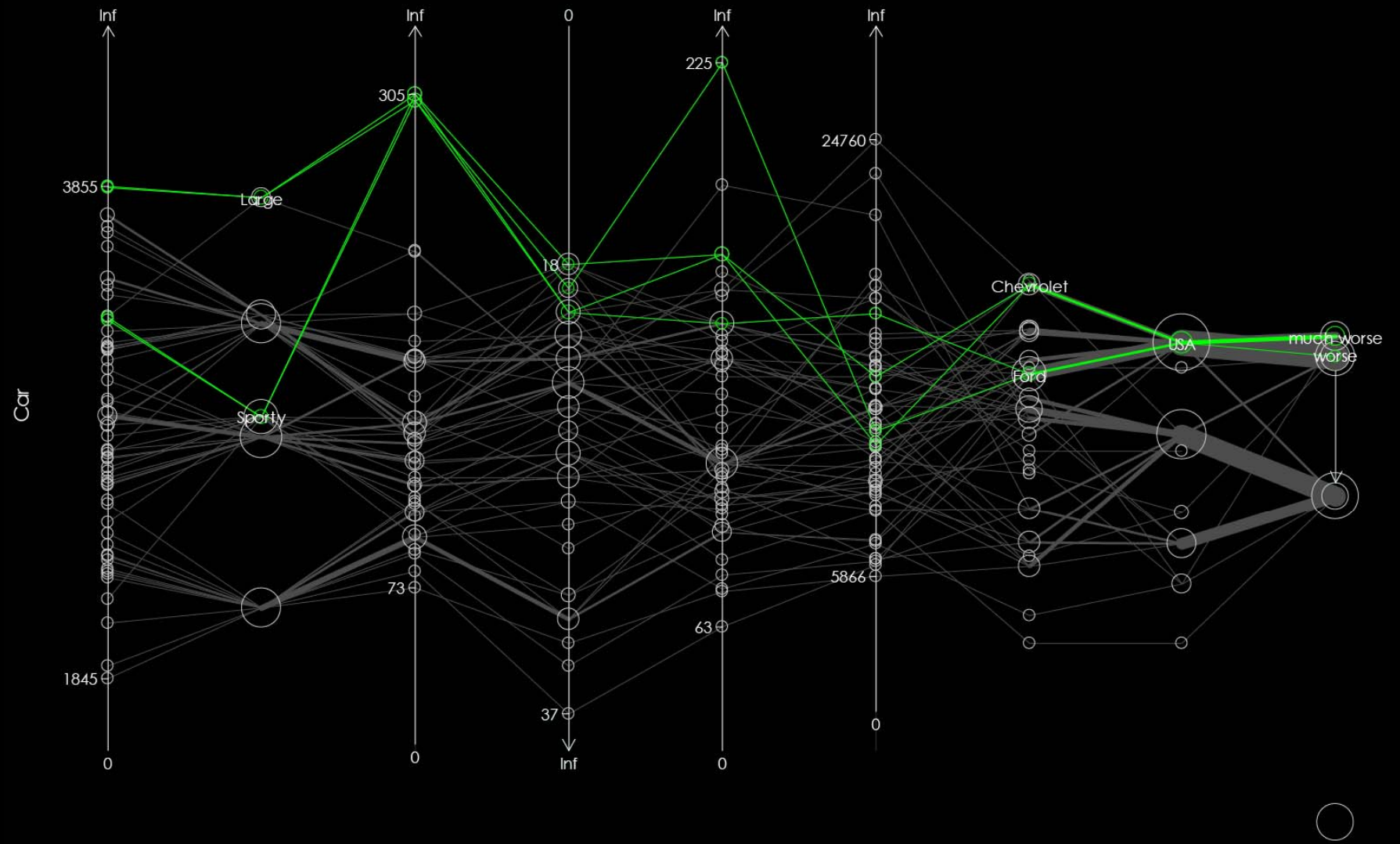

Car
Inf
3855
1845
0
Large
Van
Medium
Sporty
Compact
Small
Inf
305
73
0
0
18
37
Inf
Inf
225
63
0
Inf
24760
5866
0
Brazil
England
Chevrolet
Oldsmobile
Mitsubishi
Nissan
Mazda
Eagle
Peugeot
Hyundai
Plymouth
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
USA
Sweden
Japan
France
Germany
Japan/USA
Korea
Mexico
much worse
average
much better
NA
Weight (pound)
Type
Displacement (cubic inch)
Mileage (miles/ gallon)
Horsepower (hp)
Price (dollars)
Manufacturer
Country
Reliability

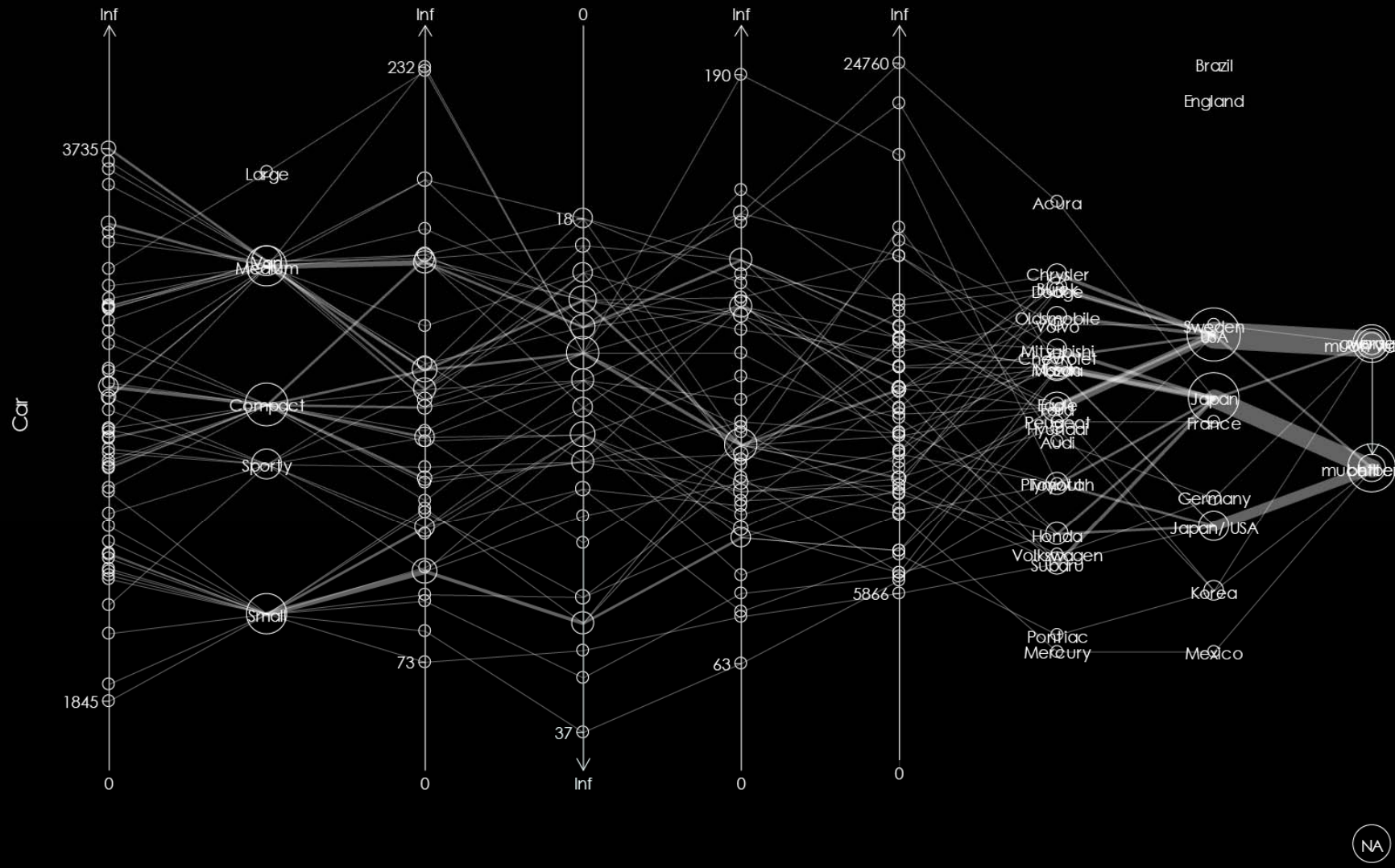

Car
Inf
3855
1845
0
Large
Van
Medium
Sporty
Compact
Small
Inf
305
73
0
Inf
5.555556
2.702703
0
Inf
225
63
0
Inf
10.1169847
8.6769282
- Inf
Brazil
England
Chevrolet
Acura
Dodge
Buick
Oldsmobile
Volvo
Mitsubishi
Nissan
Mazda
Eagle
Peugeot
Hyundai
Plymouth
Audi
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
USA
Sweden
Japan
France
Germany
Japan/USA
Korea
Mexico
much worse
average
much better
NA
Weight (pound)
Type
Displacement (cubic inch)
Fuel (100* gallon/ miles)
Horsepower (hp)
log(Price) (log(dollars))
Manufacturer
Country
Reliability

Enhancements of Textile Plot

- **Transform**
  - single data vector (log, inverse, etc)
  - multiple data vectors (normalisation, etc)
- **Substitute Data Point**
- **Extract/Remove**
  - records
  - data vectors
- **Re-weight**
  - outliers
- **Order Data Vectors**
  - a priori
  - distance to ideal coordinates
  - distance between coordinate vectors
- **Highlight Connecting Lines**
- **Select Graphical Parameters**
  - color, transparency, size, etc.

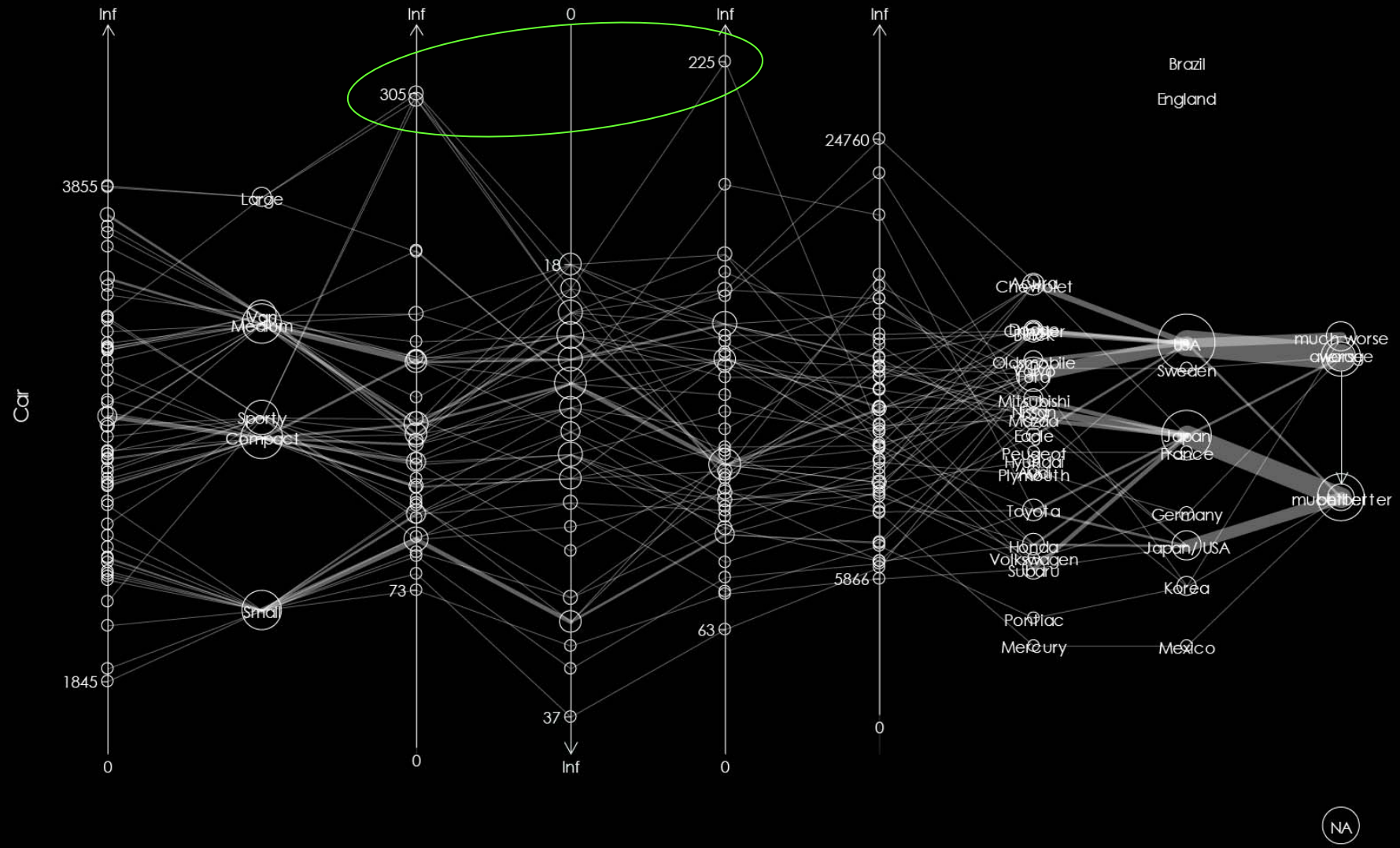

Car
Inf
3855
1845
0
Large
Van
Medium
Sporty
Compact
Small
Inf
305
73
0
0
18
37
Inf
Inf
225
63
0
Inf
24760
5866
0
Brazil
England
Chevrolet
Oldsmobile
Mitsubishi
Mazda
Eagle
Peugeot
Plymouth
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
USA
Sweden
Japan
France
Germany
Japan/USA
Korea
Mexico
much worse
average
much better
NA
Weight (pound)
Type
Displacement (cubic inch)
Mileage (miles/ gallon)
Horsepower (hp)
Price (dollars)
Manufacturer
Country
Reliability

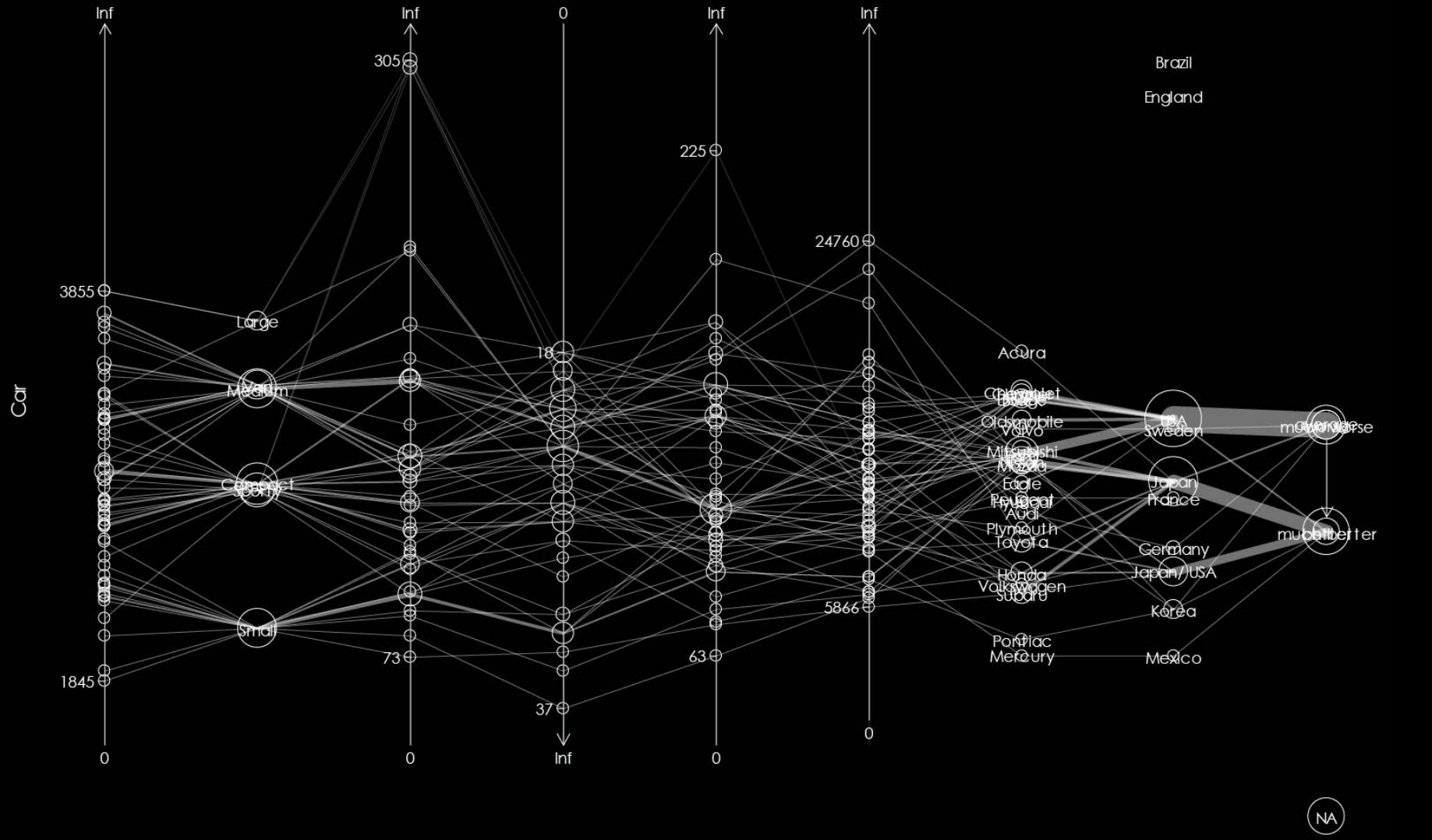

Inf
Inf
0
Inf
Inf
225
305
24760
3855
Large
18
Chevrolet
USA
much worse
worse
Ford
Car
Sporty
73
5866
63
1845
37
0
0
Inf
0
0
Weight (pound)
Type
Displacement (cubic inch)
Mileage (miles/ gallon)
Horsepower (hp)
Price (dollars)
Manufacturer
Country
Reliability


Inf
Inf
0
Inf
Inf
3735
1845
232
73
18
37
190
63
24760
5866
0
0
0
0
Inf
Car
Large
Medium
Compact
Sporty
Small
Brazil
England
Acura
Chrysler
Dodge
Oldsmobile
Volvo
Mitsubishi
Chevrolet
Mazda
Eagle
Peugeot
Hyundai
Audi
Plymouth
Honda
Volkswagen
Subaru
Pontiac
Mercury
Sweden
USA
Japan
France
Germany
Japan/USA
Korea
Mexico
NA
Weight (pound)
Type
Displacement (cubic inch)
Mileage (miles/ gallon)
Horsepower (hp)
Price (dollars)
Manufacturer
Country
Reliability

Enhancements of Textile Plot

- **Transform**
  - single data vector (log, inverse, etc)
  - multiple data vectors (normalisation, etc)
- **Substitute Data Point**
- **Extract/Remove**
  - records
  - data vectors
- **Re-weight**
  - outliers
- **Order Data Vectors**
  - a priori
  - distance to ideal coordinates
  - distance between coordinate vectors
- **Highlight Connecting Lines**
- **Select Graphical Parameters**
  - color, transparency, size, etc.


Car
Inf
3855
1845
0
Large
Van
Medium
Sporty
Compact
Small
Inf
305
73
0
0
18
37
Inf
Inf
225
63
0
Inf
24760
5866
0
Brazil
England
Chevrolet
Buick
Oldsmobile
Volvo
Mitsubishi
Nissan
Mazda
Eagle
Peugeot
Hyundai
Audi
Plymouth
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
USA
Sweden
Japan
France
Germany
Japan/USA
Korea
Mexico
much worse
average
much better
NA
Weight (pound)
Type
Displacement (cubic inch)
Mileage (miles/ gallon)
Horsepower (hp)
Price (dollars)
Manufacturer
Country
Reliability


Inf
Inf
0
Inf
Inf
305
225
24760
3855
1845
18
73
63
37
5866
0
Inf
0
0
Car
Large
Medium
Compact
Sporty
Small
Brazil
England
Acura
Volvo
Eagle
Audi
Plymouth
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
USA
Sweden
Japan
France
Germany
Japan/USA
Korea
Mexico
NA
Weight (pound)
Type
Displacement (cubic inch)
Mileage (miles/ gallon)
Horsepower (hp)
Price (dollars)
Manufacturer
Country
Reliability

Enhancements of Textile Plot

Transform
- single data vector (log, inverse, etc)
- multiple data vectors (normalisation, etc)

Substitute Data Point

Extract/Remove
- records
- data vectors

Re-weight
- outliers

Order Data Vectors
- a priori
- distance to ideal coordinates
- distance between coordinate vectors

Highlight Connecting Lines

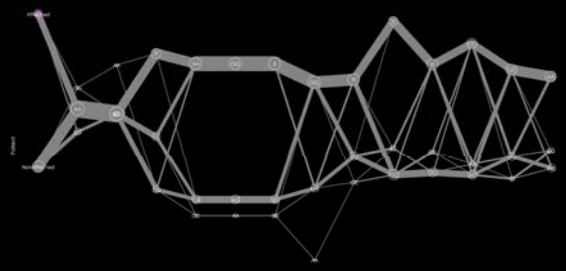
Select Graphical Parameters
- color, transparency, size, etc.

Order of Data Vectors

Data Vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$

A priori order

As it is



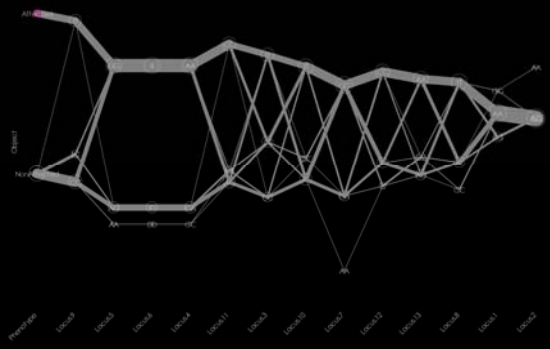
Distance to Ideal coord vec

The further left coordinate vector is closer to the ideal coordinate vector

$$\|\boldsymbol{y}_j - \hat{\boldsymbol{\xi}}\|^2 \leq \|\boldsymbol{y}_{j+1} - \hat{\boldsymbol{\xi}}\|^2$$

Ideal coordinate vector gives a set of ideal coordinates for each records

Classification of records



Distance between data vecs

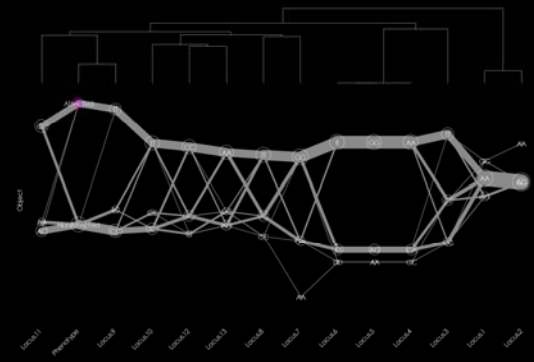
Sum of absolute values of slopes

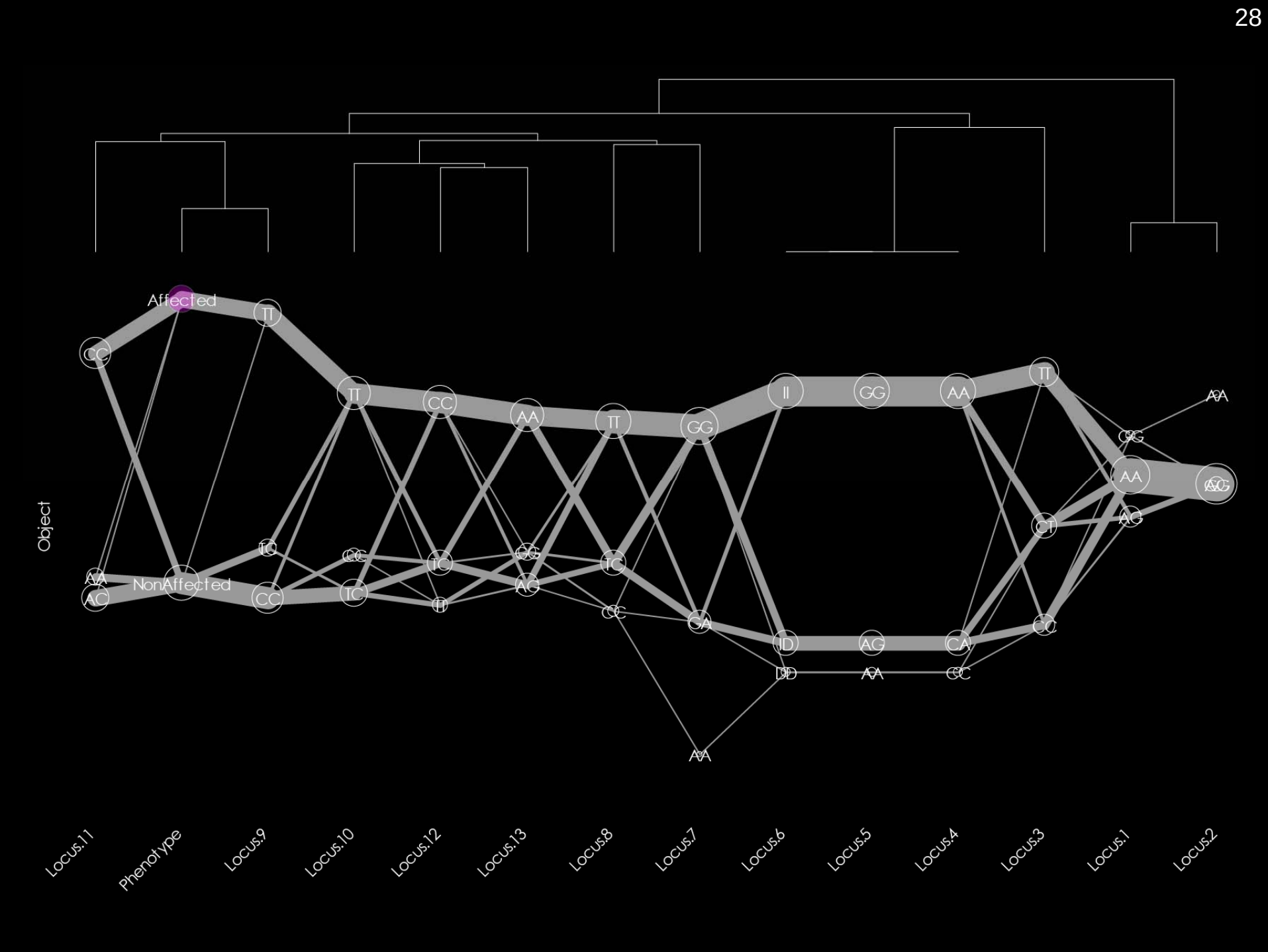
$$d(\boldsymbol{x}_j, \boldsymbol{x}_k) = \sum_{i=1}^{n} |y_{ij} - y_{ik}|$$

Linkage method

Ordered single end-linkage (Hurley 2004)

Classification of variables




Affected
NonAffected
Object
Locus.11
Phenotype
Locus.9
Locus.10
Locus.12
Locus.13
Locus.8
Locus.7
Locus.6
Locus.5
Locus.4
Locus.3
Locus.1
Locus.2

Enhancements of Textile Plot

Transform
- single data vector (log, inverse, etc)
- multiple data vectors (normalisation, etc)

Substitute Data Point

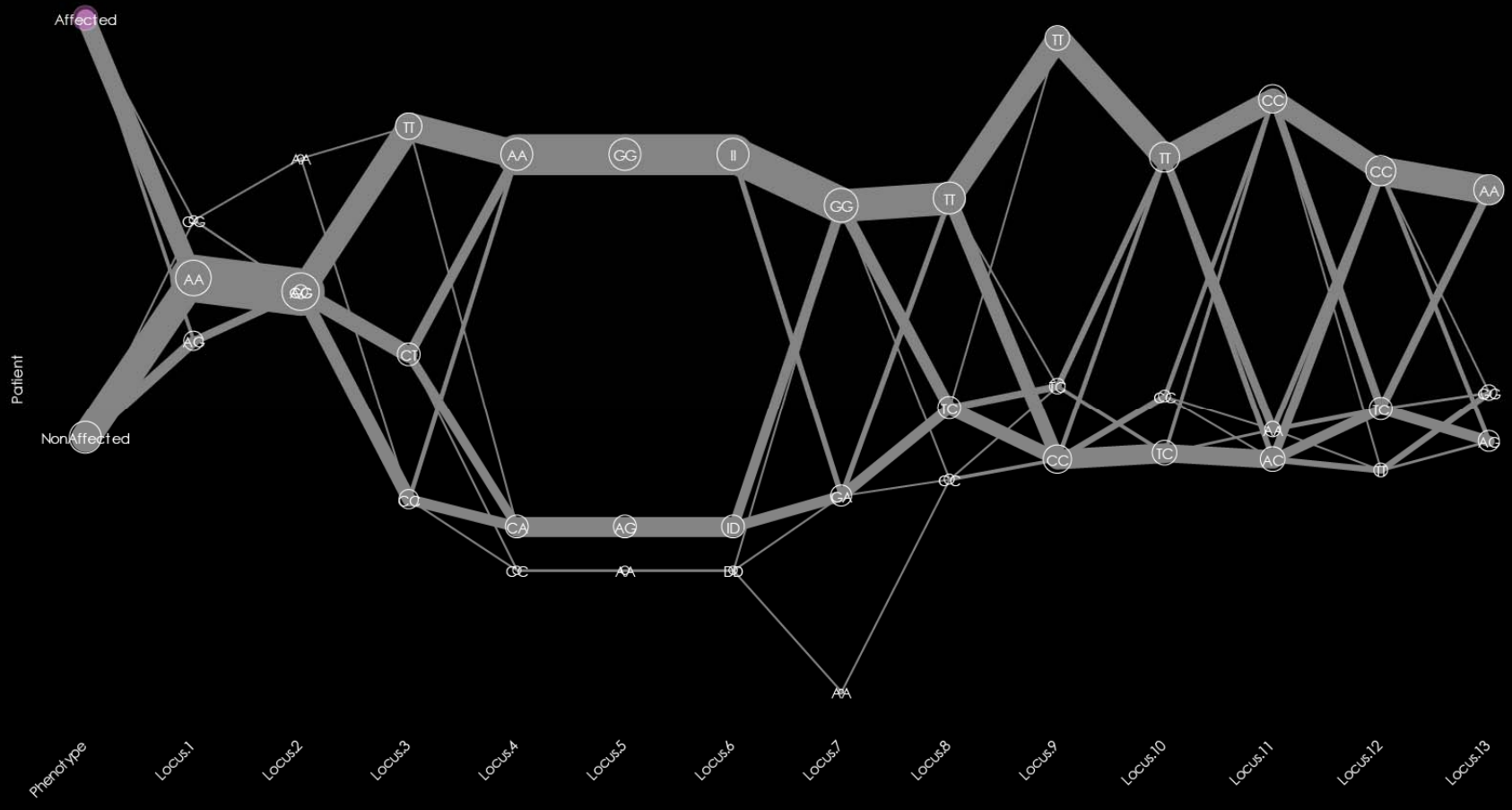
Extract/Remove
- records
- data vectors

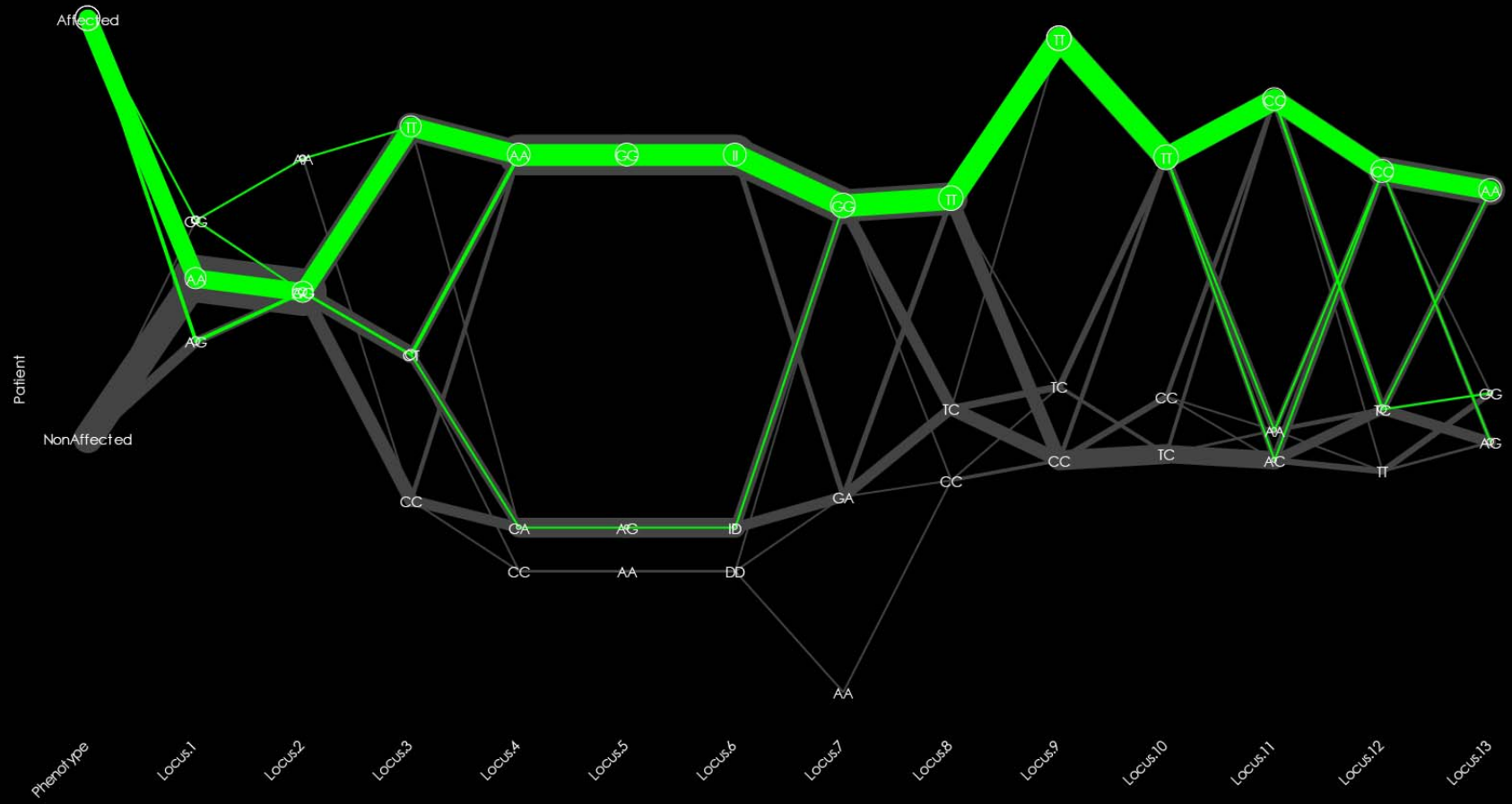
Re-weight
- outliers

Order Data Vectors
- a priori
- distance to ideal coordinates
- distance between coordinate vectors

Highlight Connecting Lines

Select Graphical Parameters
- color, transparency, size, etc.


Affected
NonAffected
Patient
Phenotype
Locus.1
Locus.2
Locus.3
Locus.4
Locus.5
Locus.6
Locus.7
Locus.8
Locus.9
Locus.10
Locus.11
Locus.12
Locus.13


Affected
NonAffected
Patient
Phenotype
Locus.1
Locus.2
Locus.3
Locus.4
Locus.5
Locus.6
Locus.7
Locus.8
Locus.9
Locus.10
Locus.11
Locus.12
Locus.13
AA
AG
GG
TT
CT
CC
GA
AG
ID
CC
AA
DD
AA
TC
CC
TC
CC
TC
AC
TT
CC
AA
TT
CC
AA
AG
GG

Enhancements of Textile Plot

Transform
- single data vector (log, inverse, etc)
- multiple data vectors (normalisation, etc)

Substitute Data Point

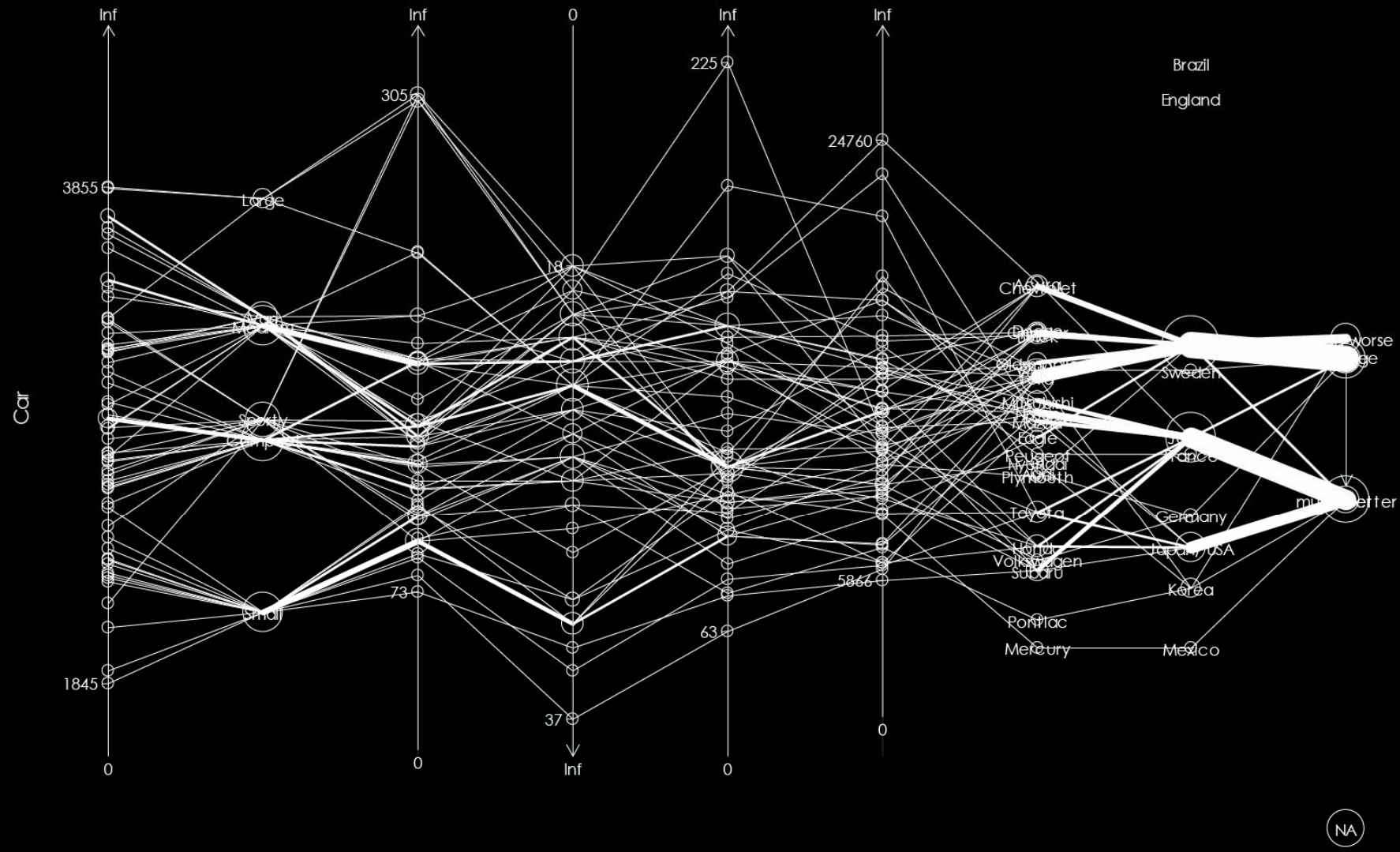
Extract/Remove
- records
- data vectors

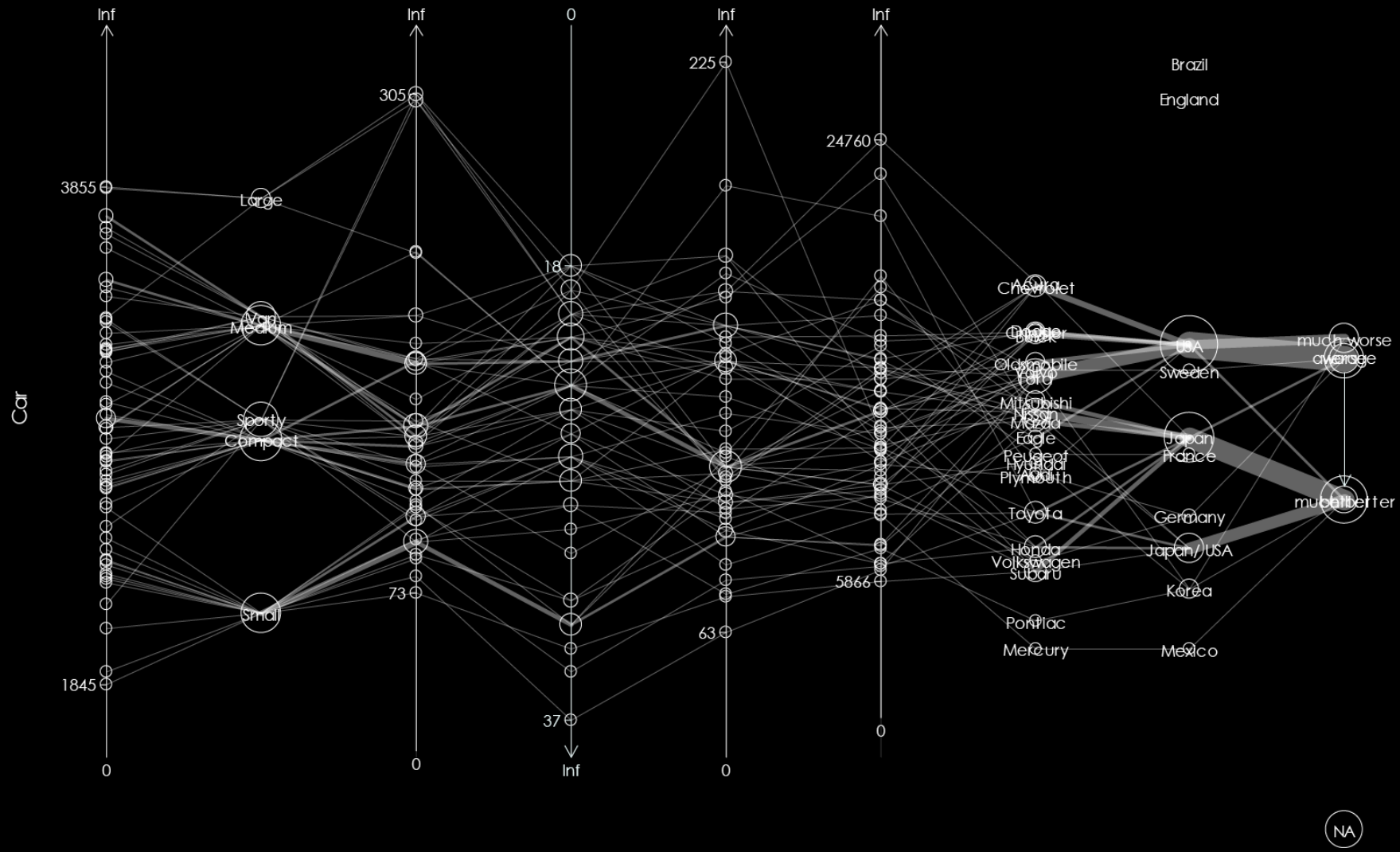
Re-weight
- outliers

Order Data Vectors
- a priori
- distance to ideal coordinates
- distance between coordinate vectors

Highlight Connecting Lines

Select Graphical Parameters
- color, transparency, size, etc.


Inf
Inf
0
Inf
Inf
225
305
24760
3855
1845
Large
Small
Car
Brazil
England
Chevrolet
Sweden
Germany
Japan/USA
Korea
Mexico
Toyota
Volkswagen
Subaru
Pontiac
Mercury
Eagle
Plymouth
worse
better
NA
73
63
5866
37
0
Inf
0
0
0
Weight (pound)
Type
Displacement (cubic inch)
Mileage (miles/ gallon)
Horsepower (hp)
Price (dollars)
Manufacturer
Country
Reliability


Car
Inf
Inf
0
Inf
Inf
225
305
24760
3855
Large
18
Van
Medium
Sporty
Compact
Small
73
63
5866
1845
37
0
0
Inf
0
Brazil
England
Chevrolet
Oldsmobile
Ford
Mitsubishi
Nissan
Mazda
Eagle
Peugeot
Hyundai
Plymouth
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
USA
Sweden
Japan
France
Germany
Japan/USA
Korea
Mexico
much worse
average
much better
NA
Weight (pound)
Type
Displacement (cubic inch)
Mileage (miles/ gallon)
Horsepower (hp)
Price (dollars)
Manufacturer
Country
Reliability

Enhancements of Textile Plot

Transform
- single data vector (log, inverse, etc)
- multiple data vectors (normalisation, etc)

Substitute Data Point

Extract/Remove
- records
- data vectors

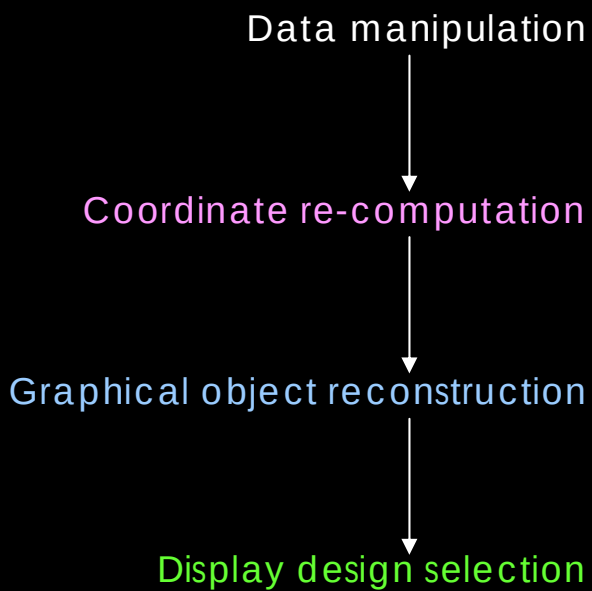
Re-weight
- outliers

Order Data Vectors
- a priori
- distance to ideal coordinates
- distance between coordinate vectors

Highlight Connecting Lines

Select Graphical Parameters
- color, transparency, size, etc.

Enhancements of Textile Plot



Transform
- single data vector (log, inverse, etc)
- multiple data vectors (normalisation, etc)

Substitute Data Point

Extract/Remove
- records
- data vectors

Re-weight
- outliers

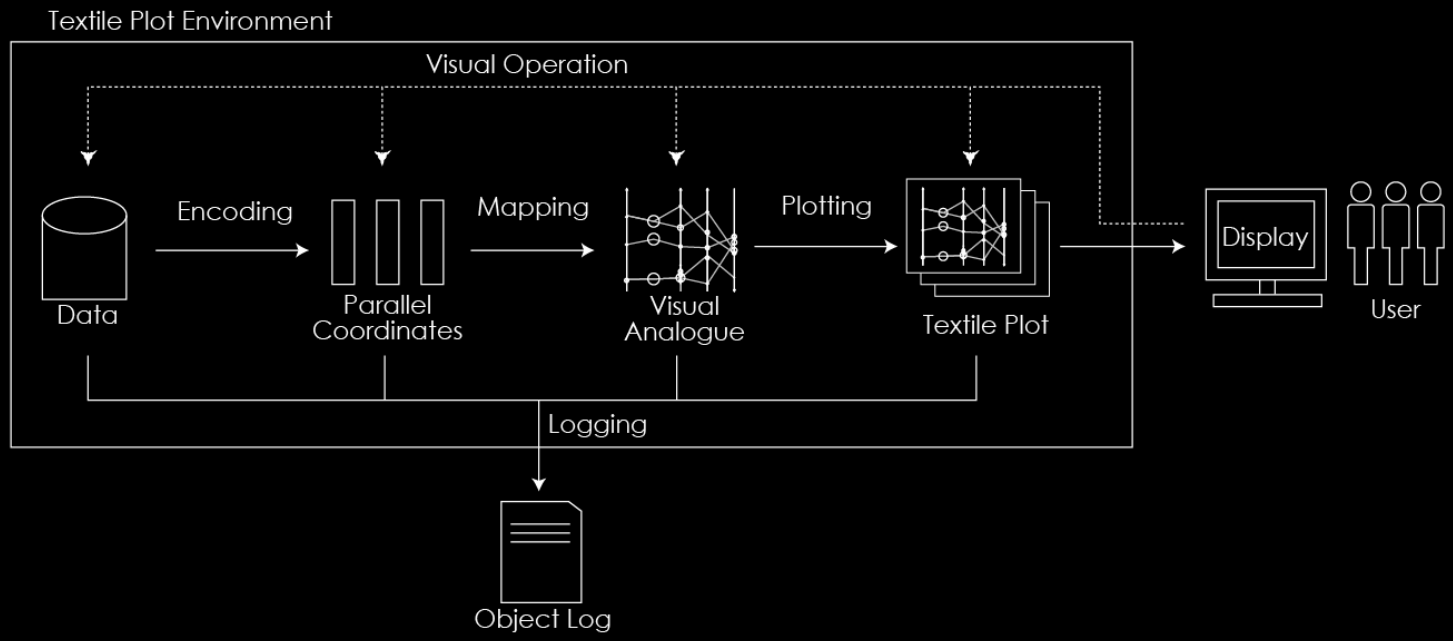
Order Data Vectors
- a priori
- distance to ideal coordinates
- distance between coordinate vectors

Highlight Connecting Lines

Select Graphical Parameters
- color, transparency, size, etc.

Reference Model

Kumasaka and Shibata 2007



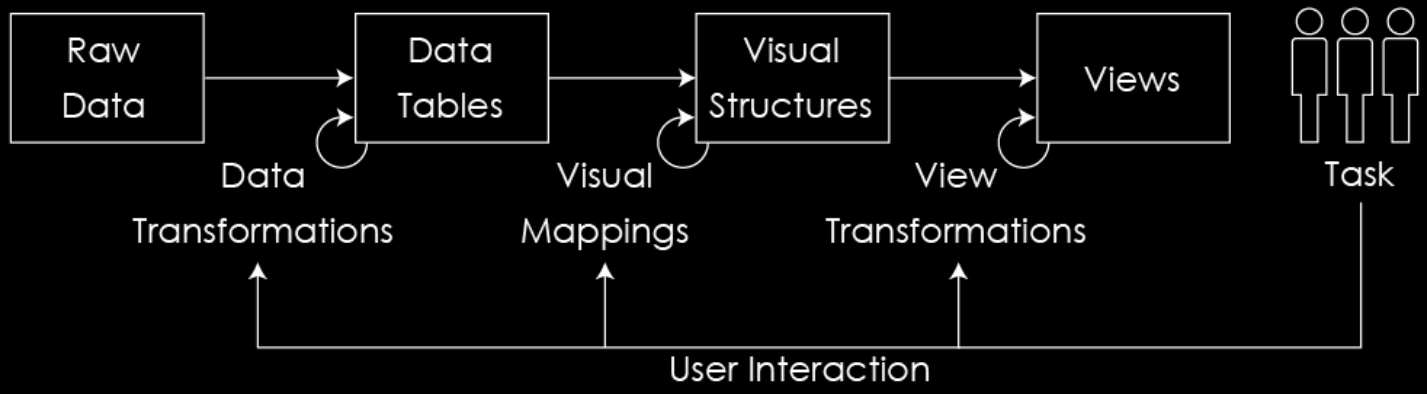
Sequence of objects

Visual operation

Object log

Reference Model for Information Visualization

Card, et al. 1999

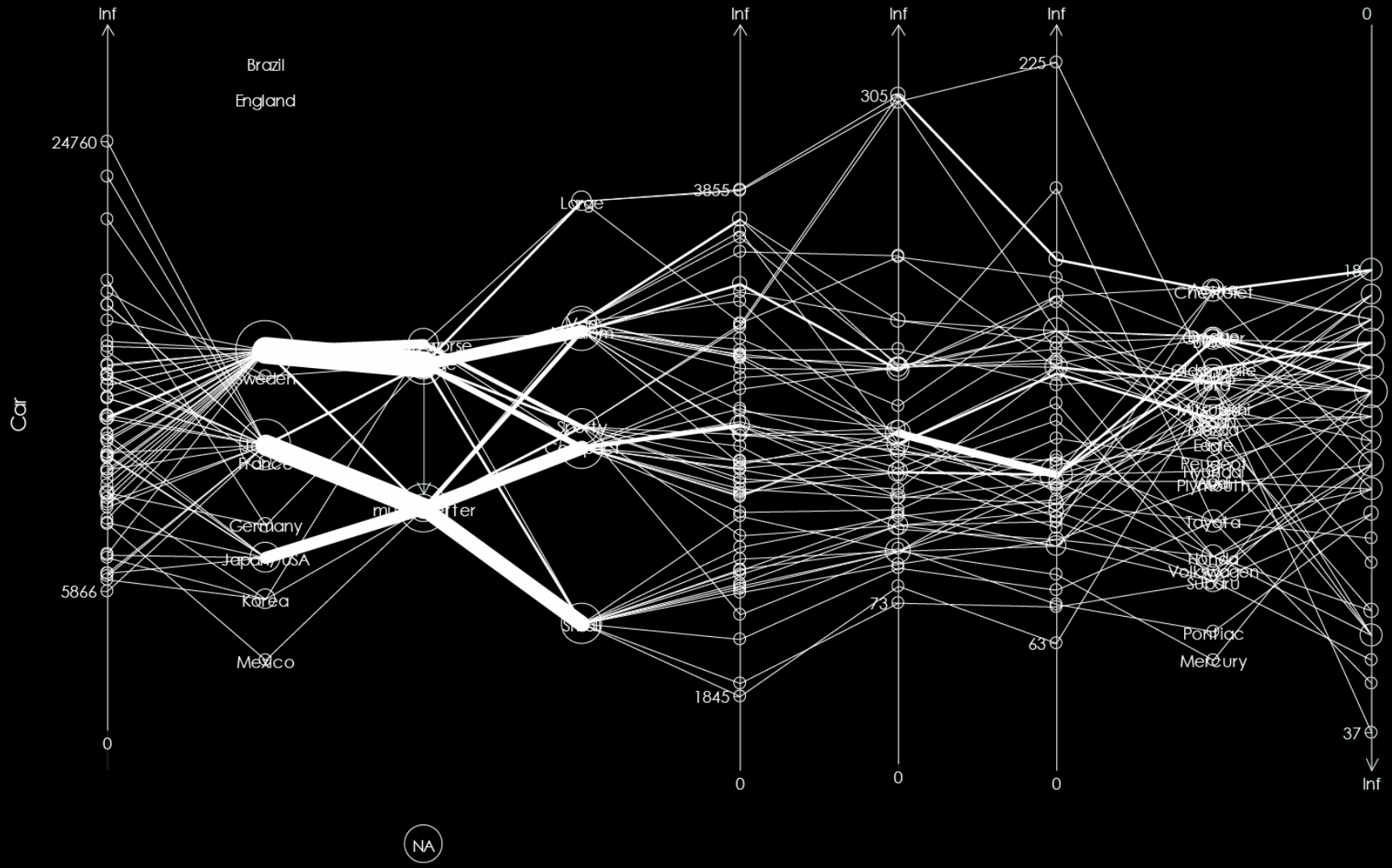


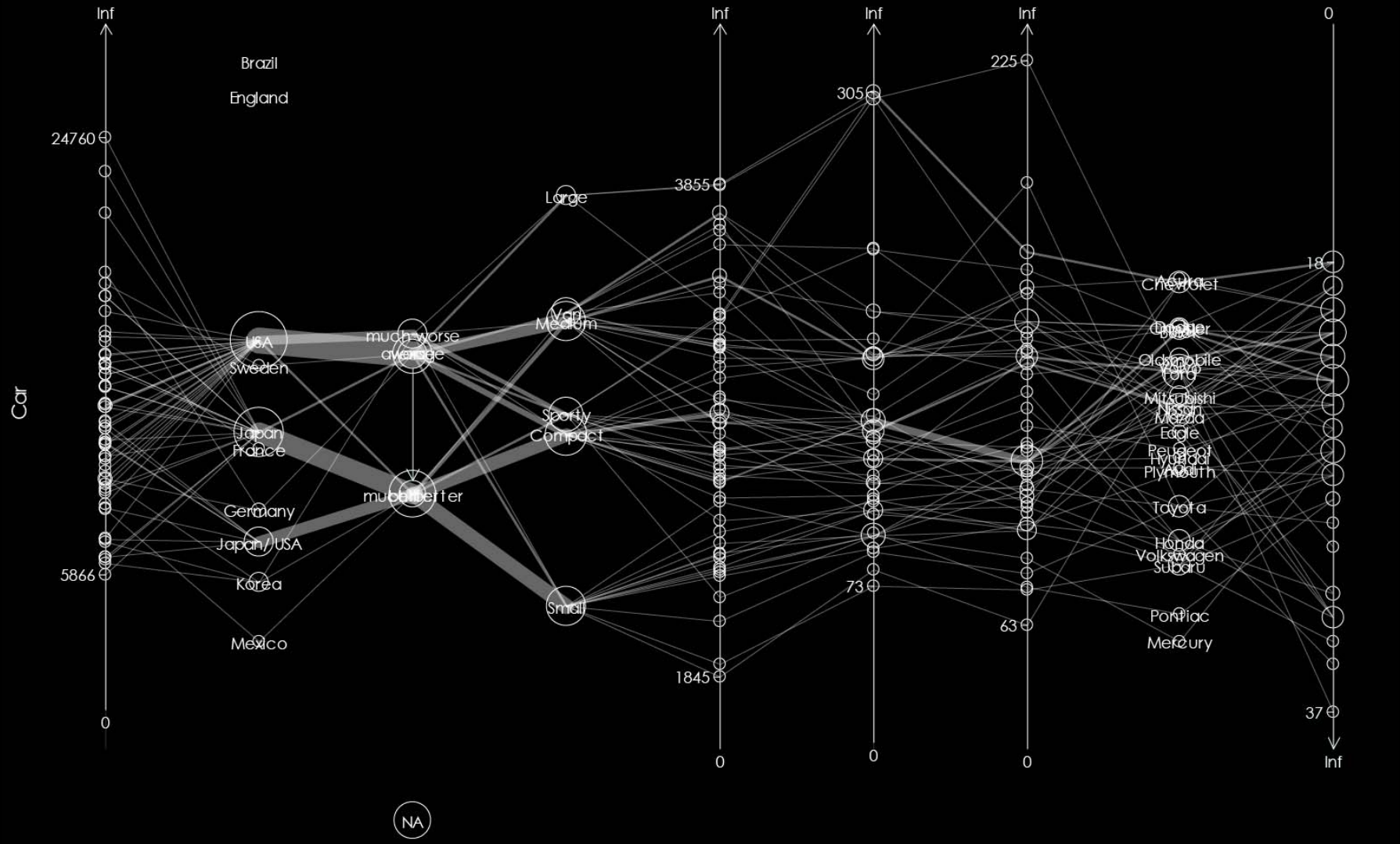
Raw Data idiosyncratic formats

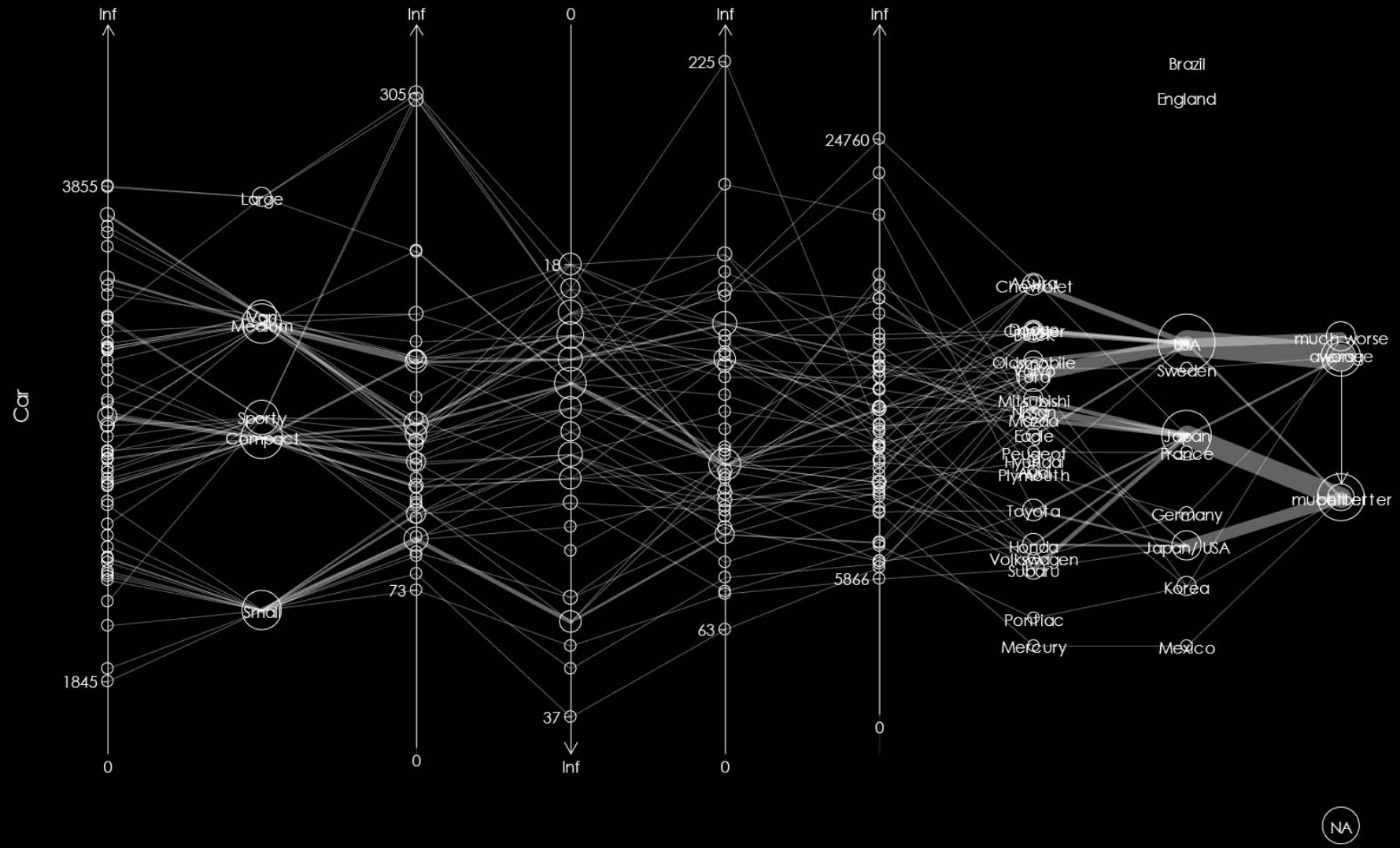
Data Tables relations (cases by variables) + metadata

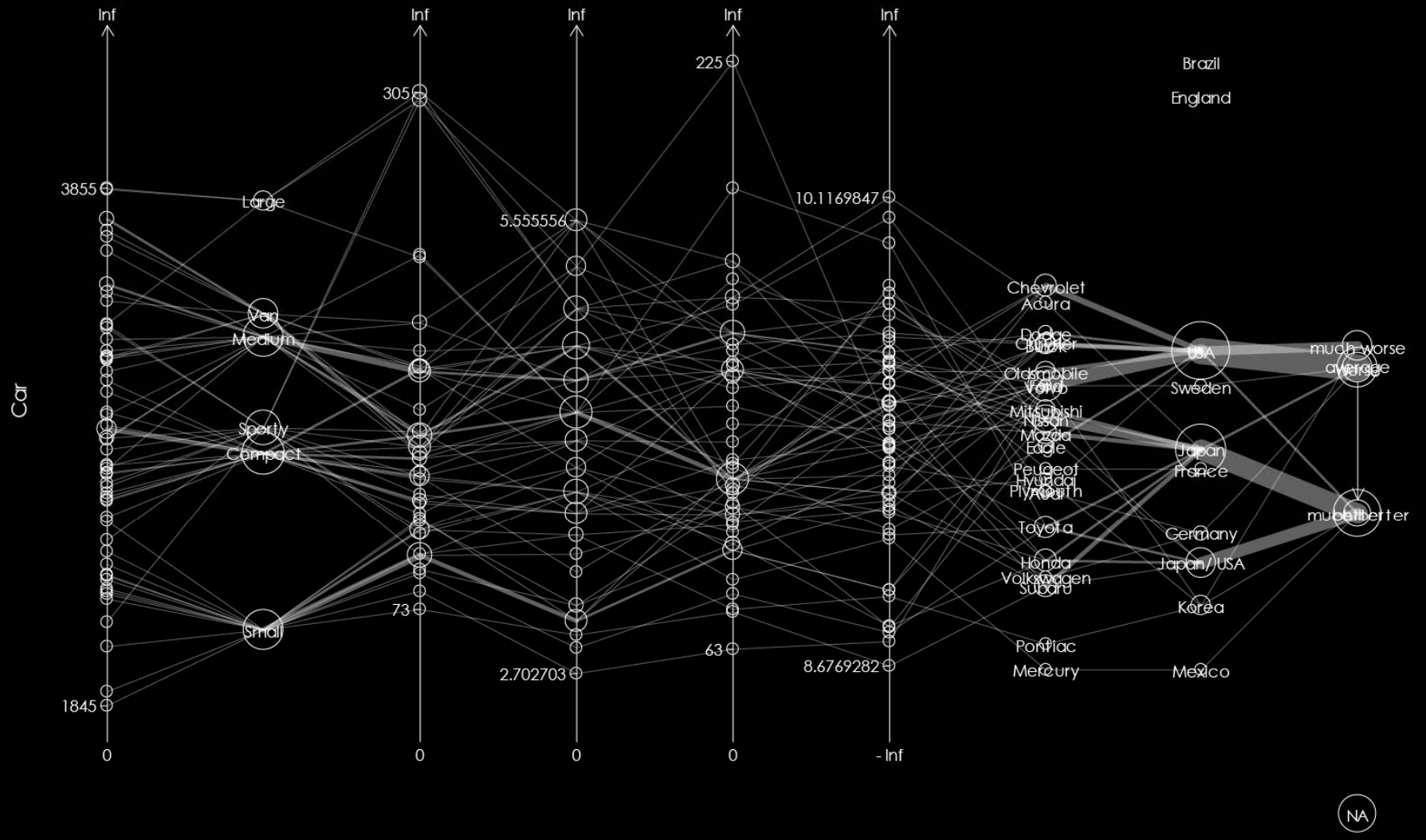
Visual Structures spatial substrates + marks + graphical properties

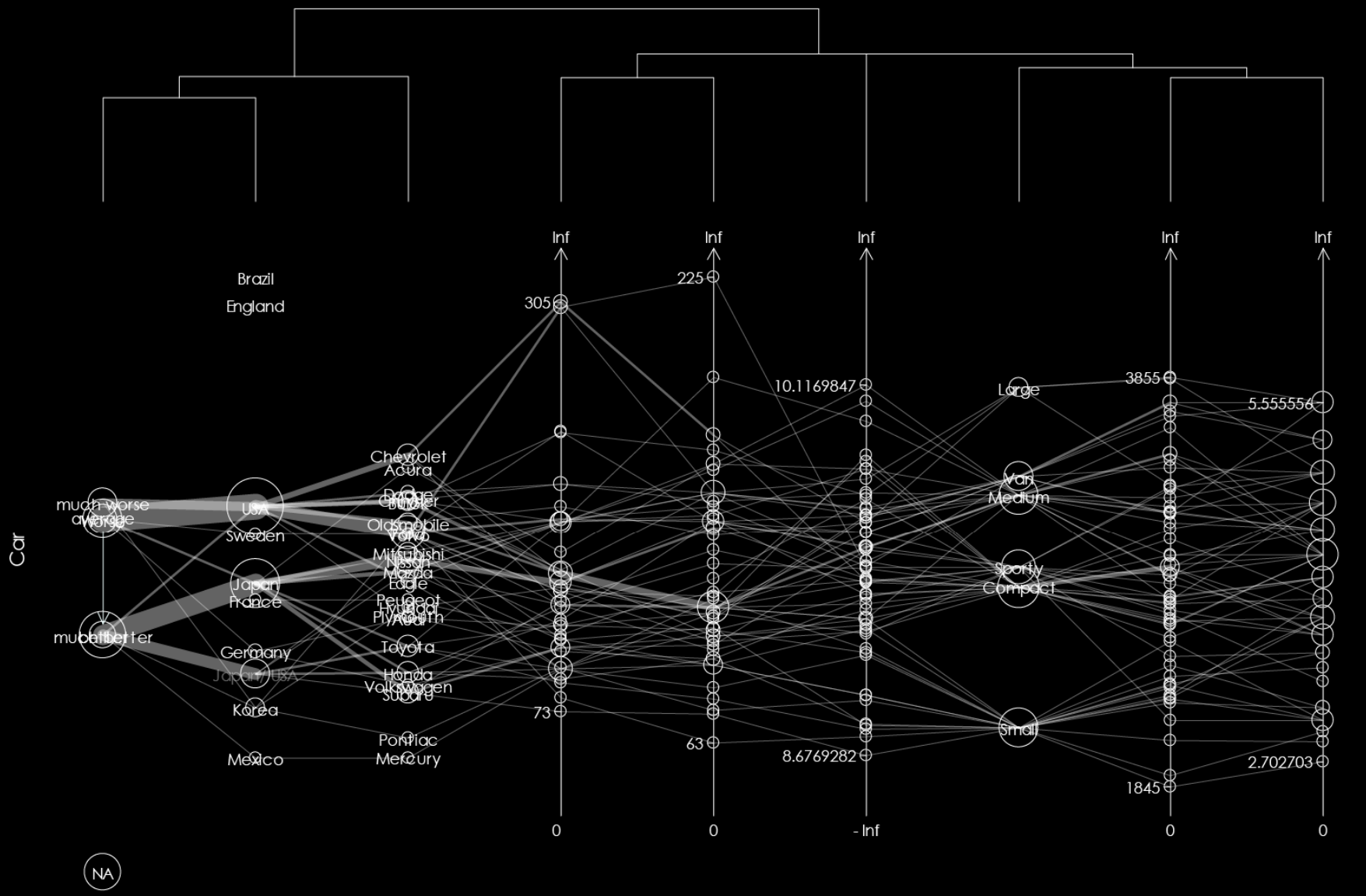
Views graphical parameters (positions, scaling, clipping,...)

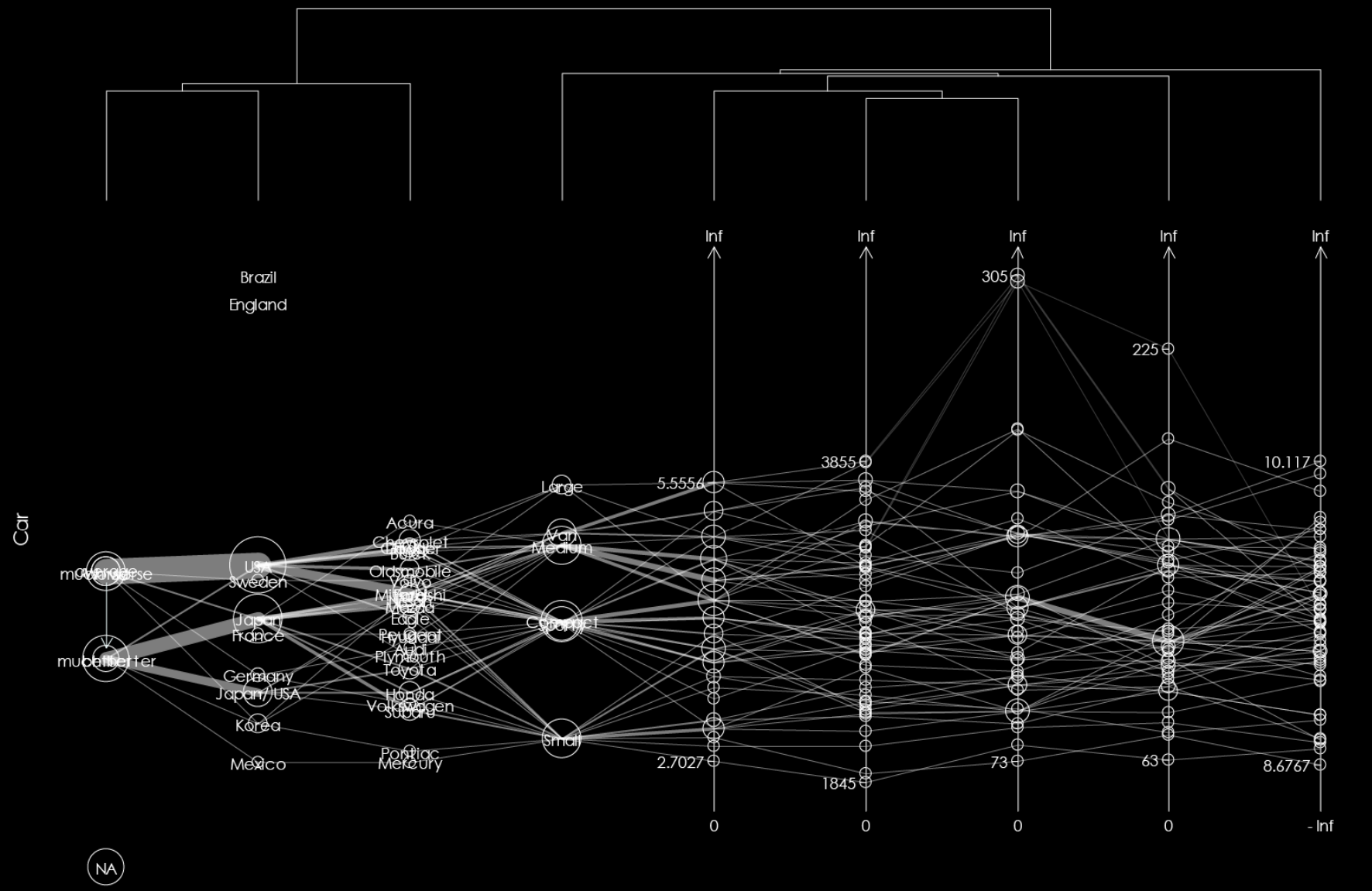

Inf
Brazil
England
24760
5866
0
Car
Large
Sweden
France
Germany
Japan/USA
Korea
Mexico
Small
NA
Inf
3855
1845
0
Inf
305
73
0
Inf
225
63
0
Chevrolet
Eagle
Plymouth
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
0
18
37
Inf
Price (dollars)
Country
Reliability
Type
Weight (pound)
Displacement (cubic inch)
Horsepower (hp)
Manufacturer
Mileage (miles/ gallon)

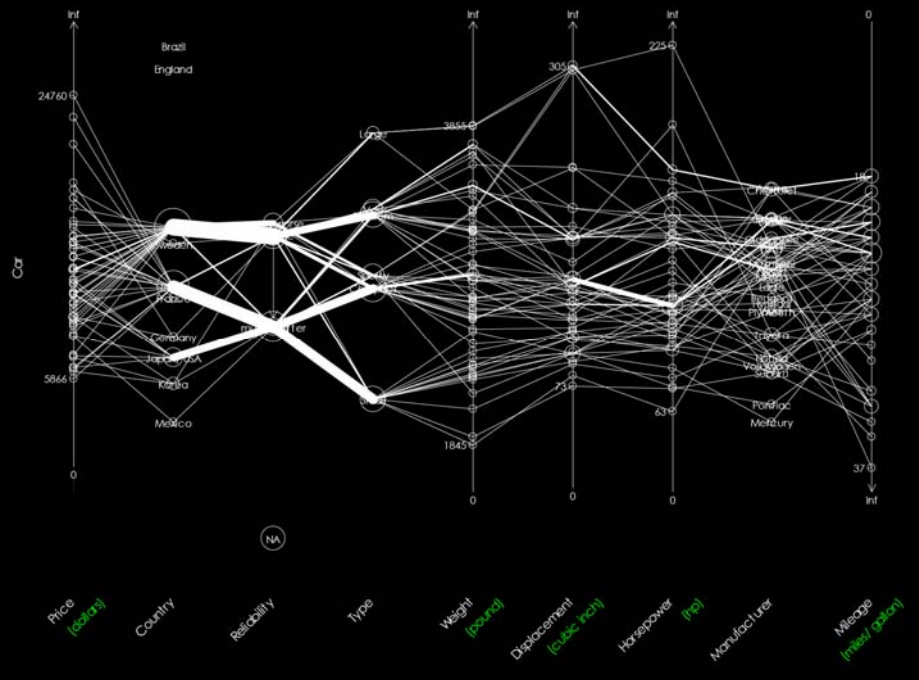

Car
Inf
24760
5866
0
Brazil
England
USA
Sweden
Japan
France
Germany
Japan/USA
Korea
Mexico
much worse
average
much better
NA
Large
Van
Medium
Sporty
Compact
Small
Inf
3855
1845
0
Inf
305
73
0
Inf
225
63
0
Chevrolet
Oldsmobile
Ford
Mitsubishi
Nissan
Mazda
Eagle
Peugeot
Hyundai
Plymouth
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
0
18
37
Inf
Price (dollars)
Country
Reliability
Type
Weight (pound)
Displacement (cubic inch)
Horsepower (hp)
Manufacturer
Mileage (miles/ gallon)

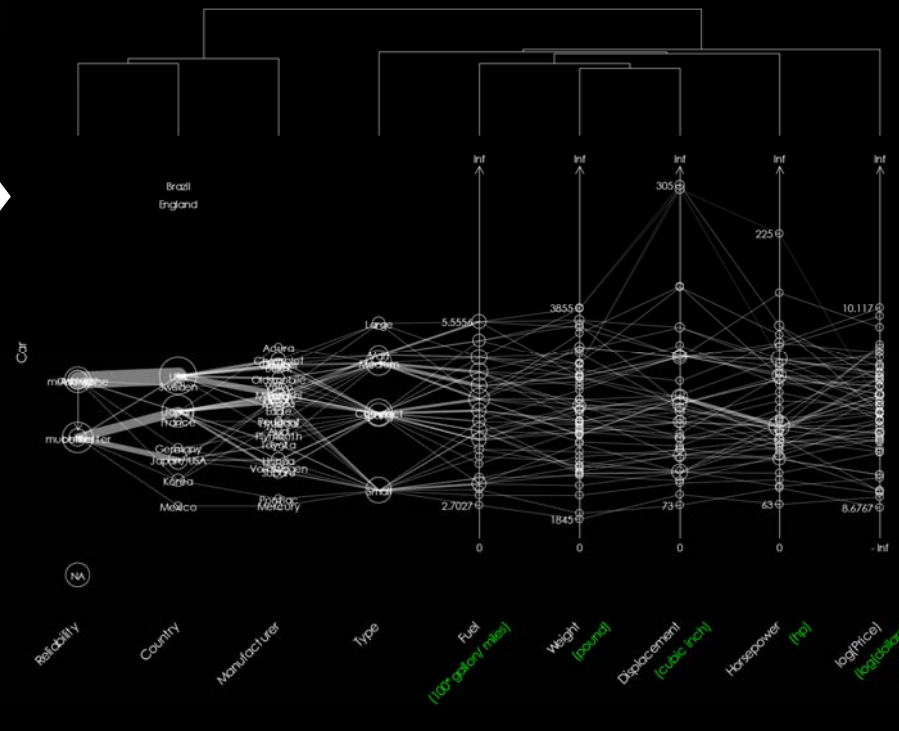

Car
Inf
3855
1845
0
Large
Van
Medium
Sporty
Compact
Small
Inf
305
73
0
0
Inf
18
37
Inf
225
63
0
Inf
24760
5866
0
Brazil
England
Chevrolet
Oldsmobile
Mitsubishi
Nissan
Mazda
Eagle
Peugeot
Hyundai
Plymouth
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
USA
Sweden
Japan
France
Germany
Japan/USA
Korea
Mexico
much worse
average
much better
NA
Weight (pound)
Type
Displacement (cubic inch)
Mileage (miles/ gallon)
Horsepower (hp)
Price (dollars)
Manufacturer
Country
Reliability


Car
Inf
3855
1845
0
Large
Van
Medium
Sporty
Compact
Small
Inf
305
73
0
Inf
5.555556
2.702703
0
Inf
225
63
0
Inf
10.1169847
8.6769282
- Inf
Brazil
England
Chevrolet
Acura
Dodge
Buick
Oldsmobile
Volvo
Mitsubishi
Nissan
Mazda
Eagle
Peugeot
Hyundai
Plymouth
Audi
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
USA
Sweden
Japan
France
Germany
Japan/USA
Korea
Mexico
much worse
average
much better
NA
Weight (pound)
Type
Displacement (cubic inch)
Fuel (100* gallon/ miles)
Horsepower (hp)
log(Price) (log(dollars))
Manufacturer
Country
Reliability

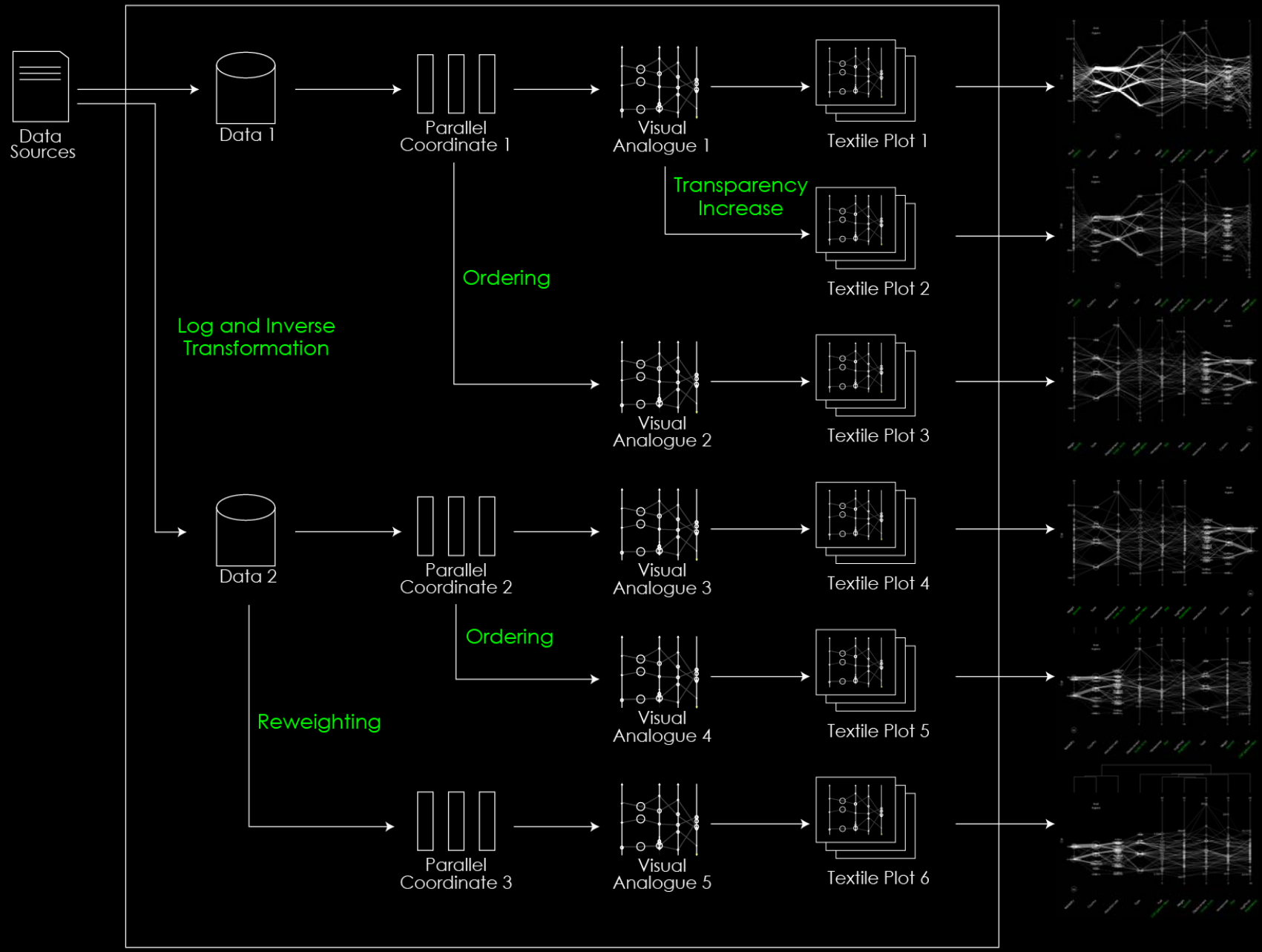

Car
much worse
worse
average
better
much better
NA
Brazil
England
USA
Sweden
Japan
France
Germany
Japan / USA
Korea
Mexico
Chevrolet
Acura
Dodge
Buick
Oldsmobile
Volvo
Mitsubishi
Nissan
Mazda
Eagle
Peugeot
Hyundai
Plymouth
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
Inf
305
73
0
225
63
0
10.1169847
8.6769282
- Inf
Large
Van
Medium
Sporty
Compact
Small
3855
1845
0
5.555556
2.702703
0
Reliability
Country
Manufacturer
Displacement (cubic inch)
Horsepower (hp)
log(Price) (log(dollars))
Type
Weight (pound)
Fuel (100* gallon/ miles)


Brazil
England
Car
Inf
Inf
Inf
Inf
Inf
305
225
3855
10.117
5.5556
Large
Van
Medium
Small
2.7027
73
63
8.6767
1845
0
0
0
0
- Inf
Acura
Oldsmobile
Volvo
Eagle
Peugeot
Audi
Plymouth
Toyota
Honda
Volkswagen
Subaru
Pontiac
Mercury
USA
Sweden
Japan
France
Germany
Japan/USA
Korea
Mexico
NA
Reliability
Country
Manufacturer
Type
Fuel
(100* gallon/ miles)
Weight
(pound)
Displacement
(cubic inch)
Horsepower
(hp)
log(Price)
(log(dollars))


Car
Brazil
England
Inf
24760
5866
0
3855
1845
305
225
63
37
Large
Small
Chevrolet
Pontiac
Mercury
Mexico
Germany
Korea
NA
Price (dollars)
Country
Reliability
Type
Weight (pound)
Displacement (cubic inch)
Horsepower (hp)
Manufacturer
Mileage (miles/ gallon)


Car
Brazil
England
Inf
305
225
3855
5.5554
10.117
2.7027
8.6767
1845
73
63
0
Large
Small
Mexico
Germany
Korea
NA
Reliability
Country
Manufacturer
Type
Fuel (100*gallon/ miles)
Weight (pound)
Displacement (cubic inch)
Horsepower (hp)
log(Price) (log(dollars))

Automobile Data Analysis Log



Bibliography 1

Card, S. K., et al. (1999), *Information Visualization*, Morgan Kaufmann Pub.

Hurley, Catherine B. (2004), Clustering Visualisations of Multidimensional Data, *Journal of Computational and Graphical Statistics*, **13** 788--806.

Inselberg, A. (1985), The plane with parallel coordinates, *The Visual Computer* **1** 69--91.

Iwasaki, N. et al. (2005), Genetic Variants in the calpain-10 gene and the development of type 2 diabetes in the Japanese population. *Journal of Human Genetics* **50** 92--98.

Kumasaka, N. and Shibata, R., High Dimensional Data Visualisation: the Textile Plot, *Computational Statistics & Data Analysis*, submitted.

Bibliography 2

Kumasaka, N. and Shibata, R. (2006), Implementation of Textile Plot, *Proceedings in COMPSTAT 2006* Roma.

Kumasaka, N. and Shibata, R. (2007), The Textile Plot Environment, 統計数理特集号, in Japanese.

Wegman, E. (1990), Hyperdimensional data analysis using parallel coordinates. *Journal of The American Statistical Association* **85** 664--675.