

# 12

## DADOS MULTIVARIADOS

*Profa. Raquel C. de Melo Minardi*

- O objetivo principal por trás de análises multivariadas são comparações complexas entre múltiplas instâncias sob o ponto de vista de múltiplas variáveis

# EXEMPLO

---

- Comparação entre automóveis
  - Preço
  - Consumo
  - Velocidade
  - Capacidade
  - Custo do seguro
  - Custo de manutenção
  - Taxa de satisfação dos clientes

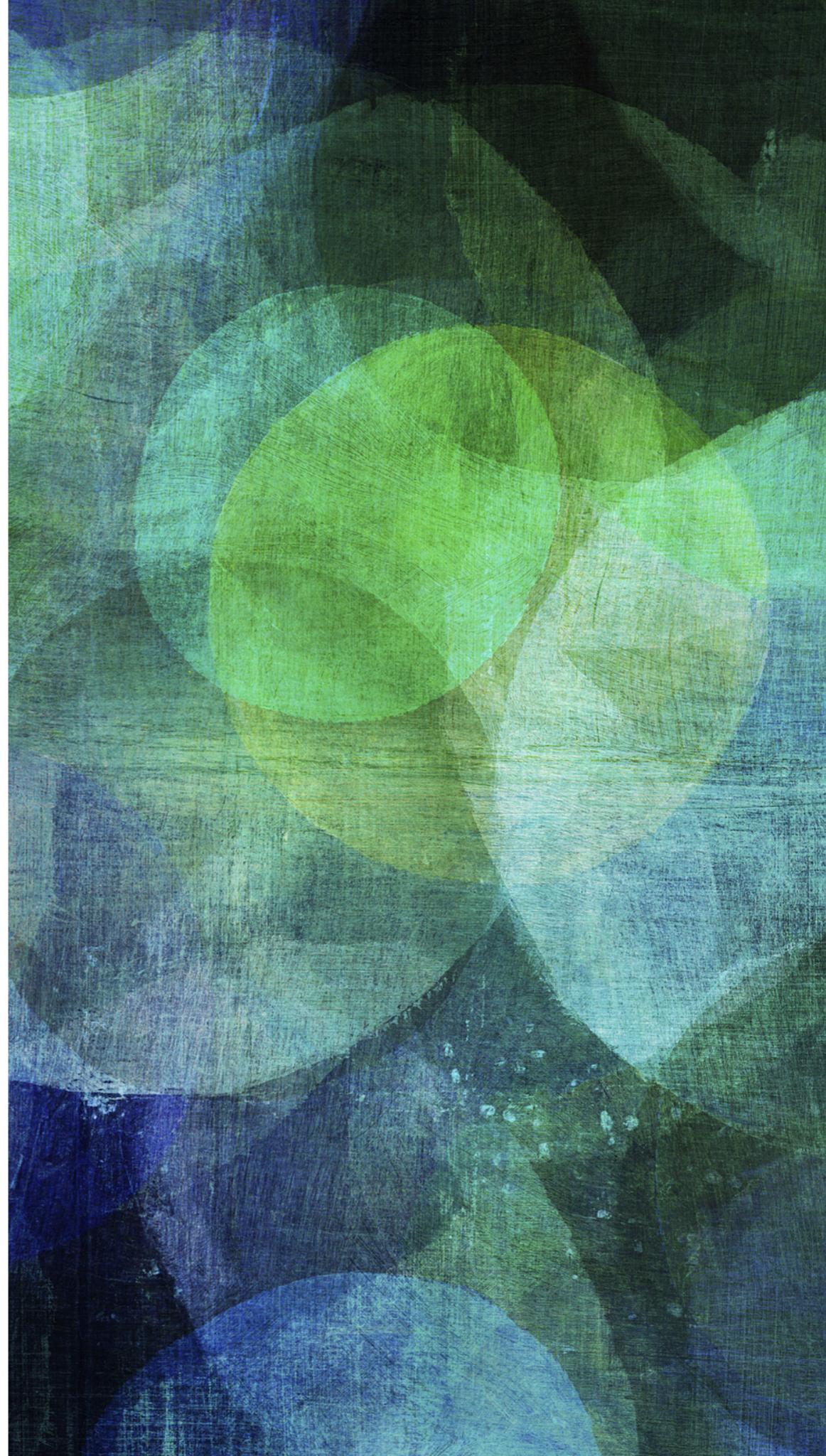
# COMPARAÇÃO ENTRE PERFIS MULTIVARIADOS

---

- Que tipo de veículo é mais apropriado para cada tipo de consumidor
- Que itens são mais parecidos?
- Que itens são excepcionais?
- Como os itens podem ser combinados de acordo com sua similaridade?

# REPRESENTAÇÕES VISUAIS

---



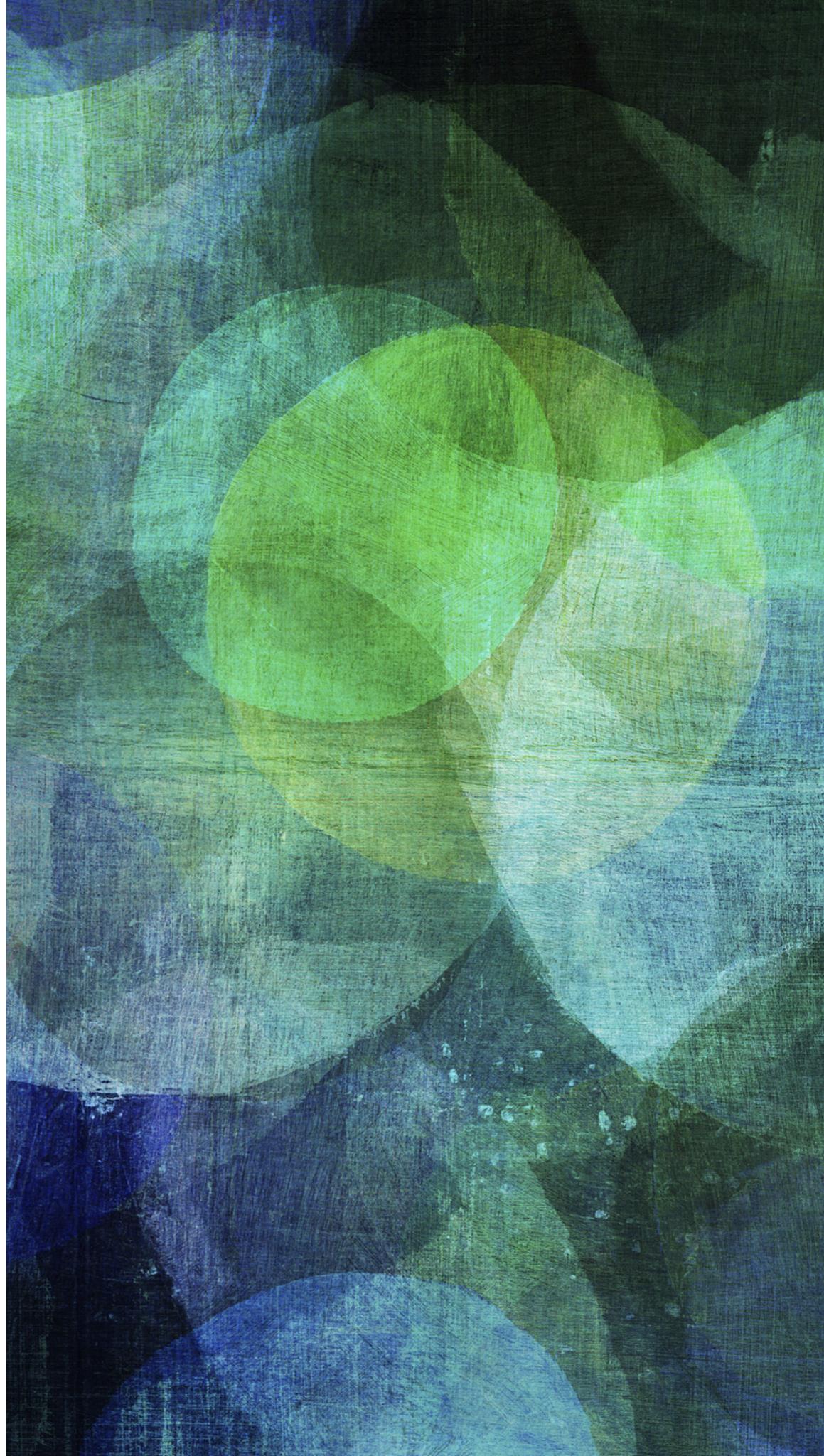
# TIPOS REPRESENTAÇÕES VISUAIS

---

- Projeções geométricas
- Representações hierárquicas
- Representações iconográficas
- Técnicas orientadas a pixels

# PROJEÇÕES GEOMÉTRICAS

---



# PROJEÇÃO GEOMÉTRICA

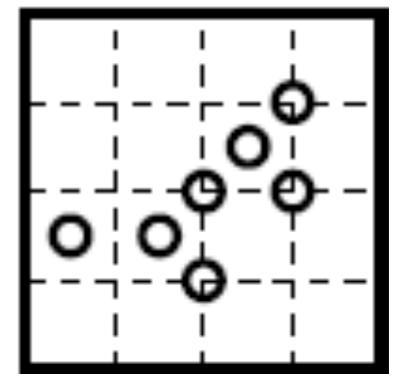
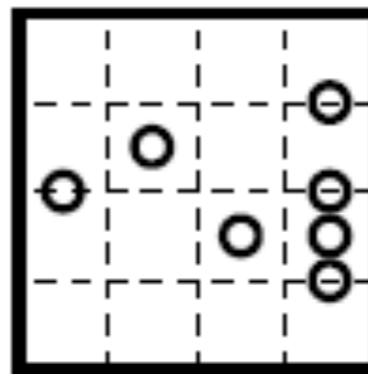
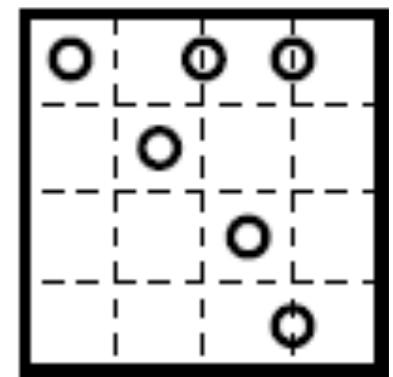
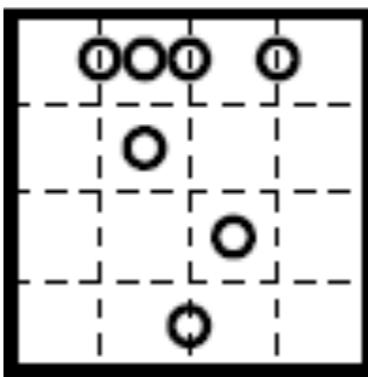
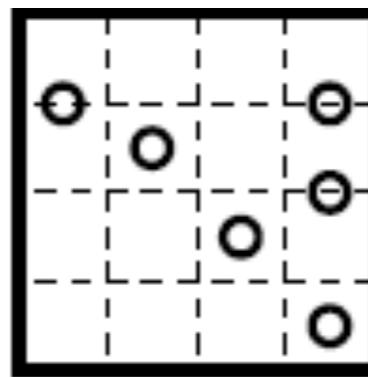
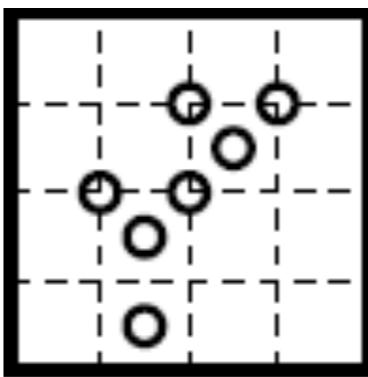
---

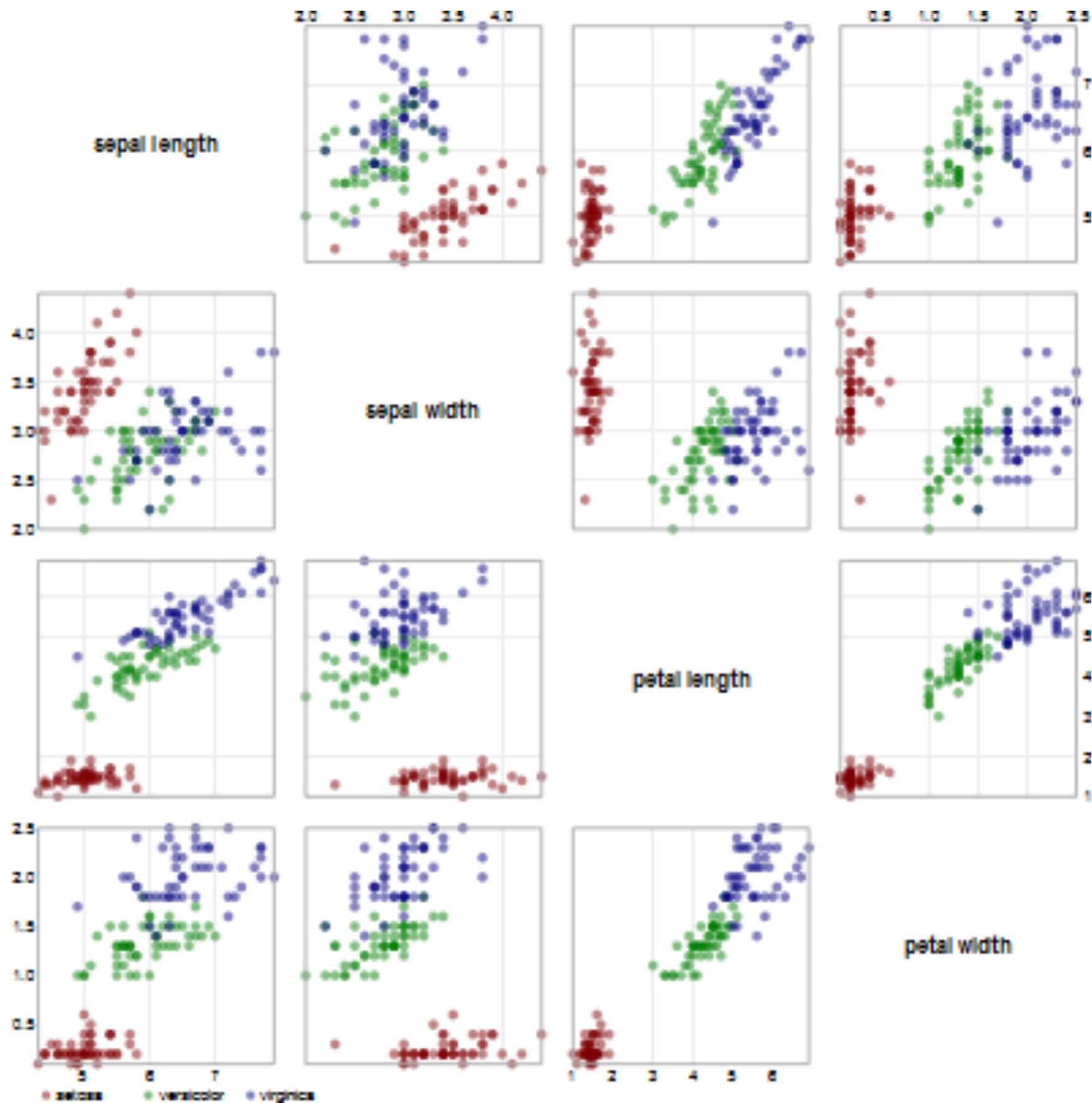
- Busca de projeções 2D para dados multidimensionais
- Úteis na detecção de exceções e ainda de correlações entre diferentes dimensões

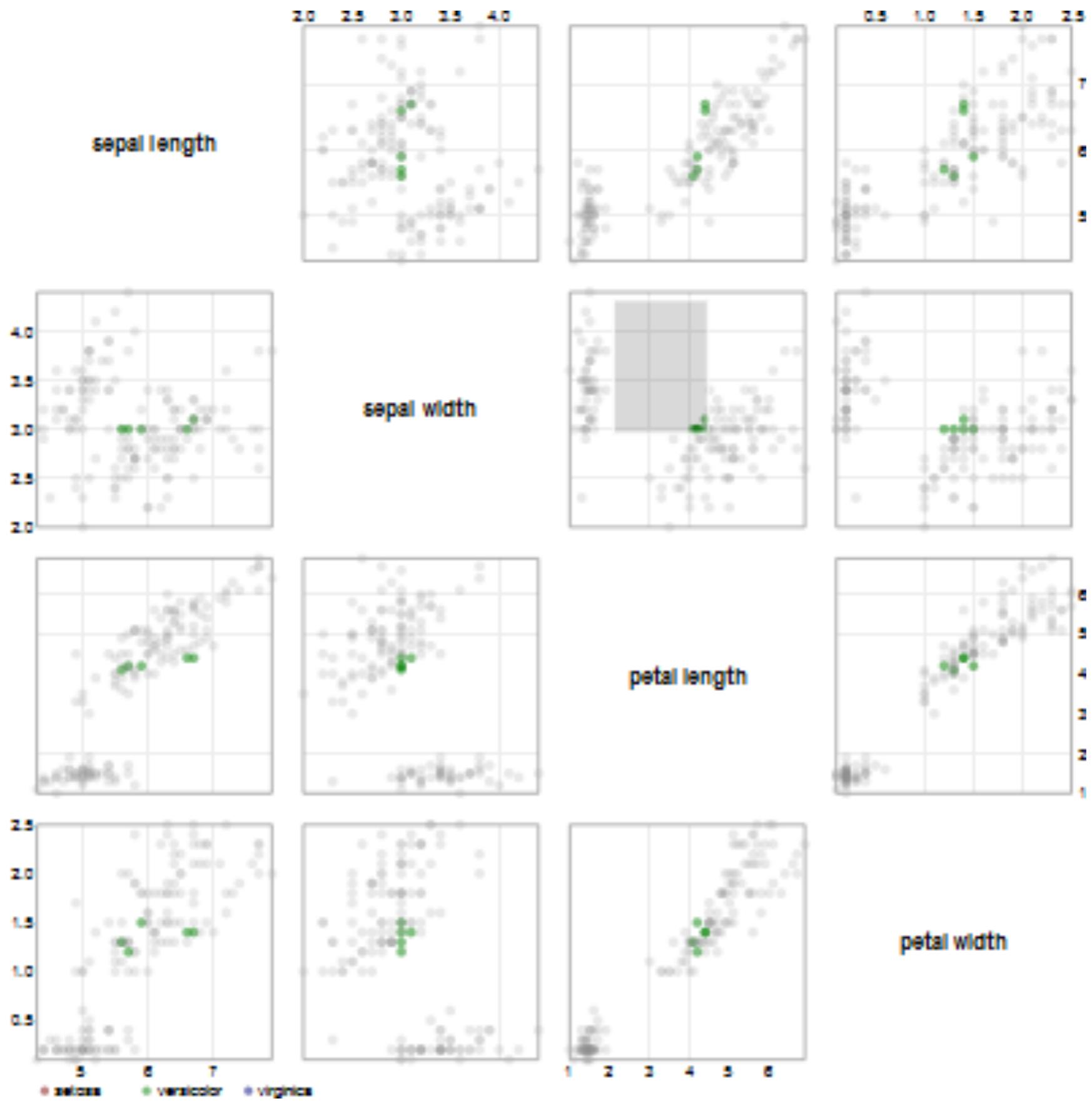
# MATRIZES DE GRÁFICOS DE DISPERSÃO

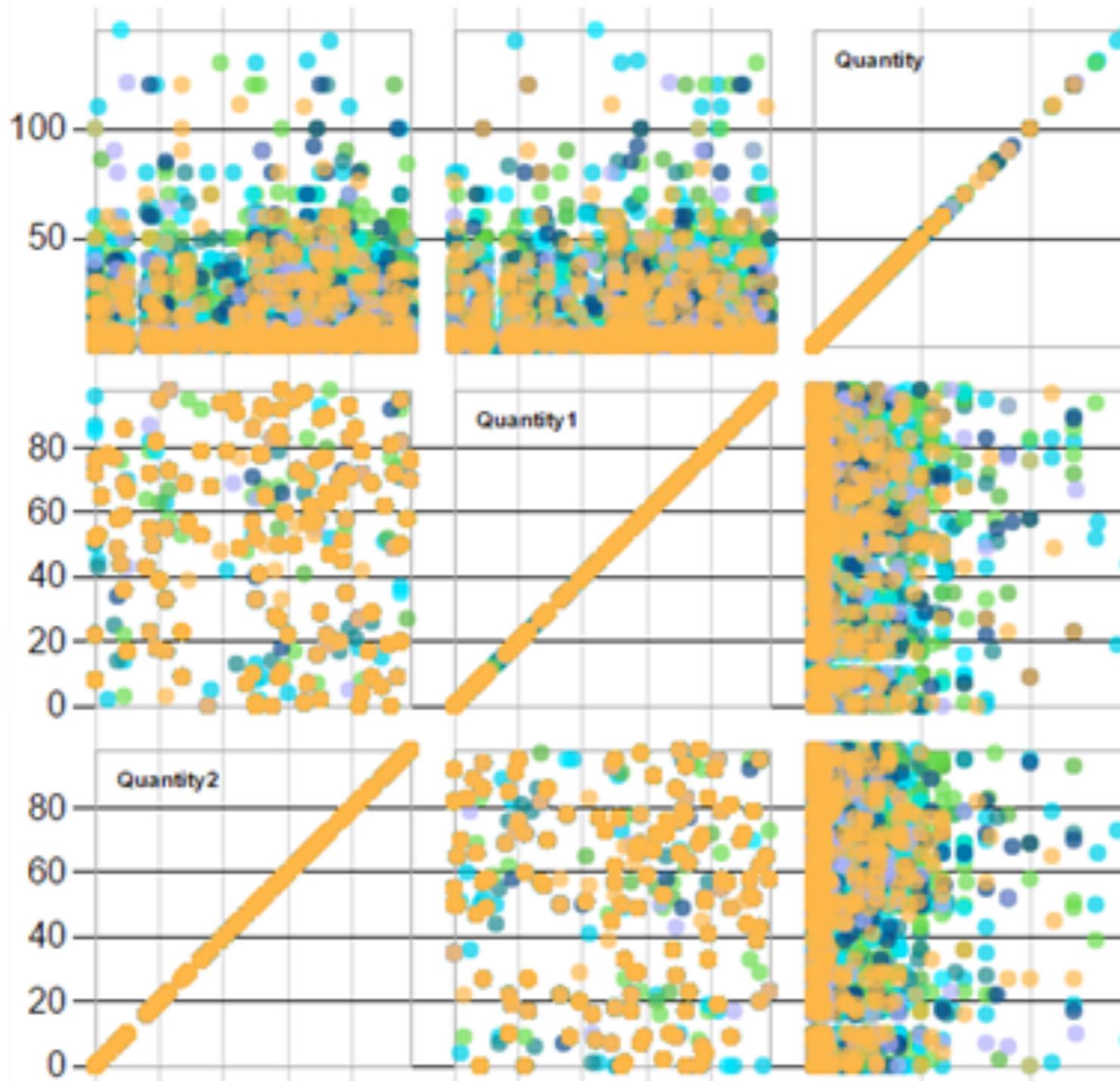
---

- A matriz de scatterplots é útil para mostrar correlações entre pares de variáveis
- O uso das técnicas de brushing e linking são muitos úteis para a identificação dos itens nos diversos plots

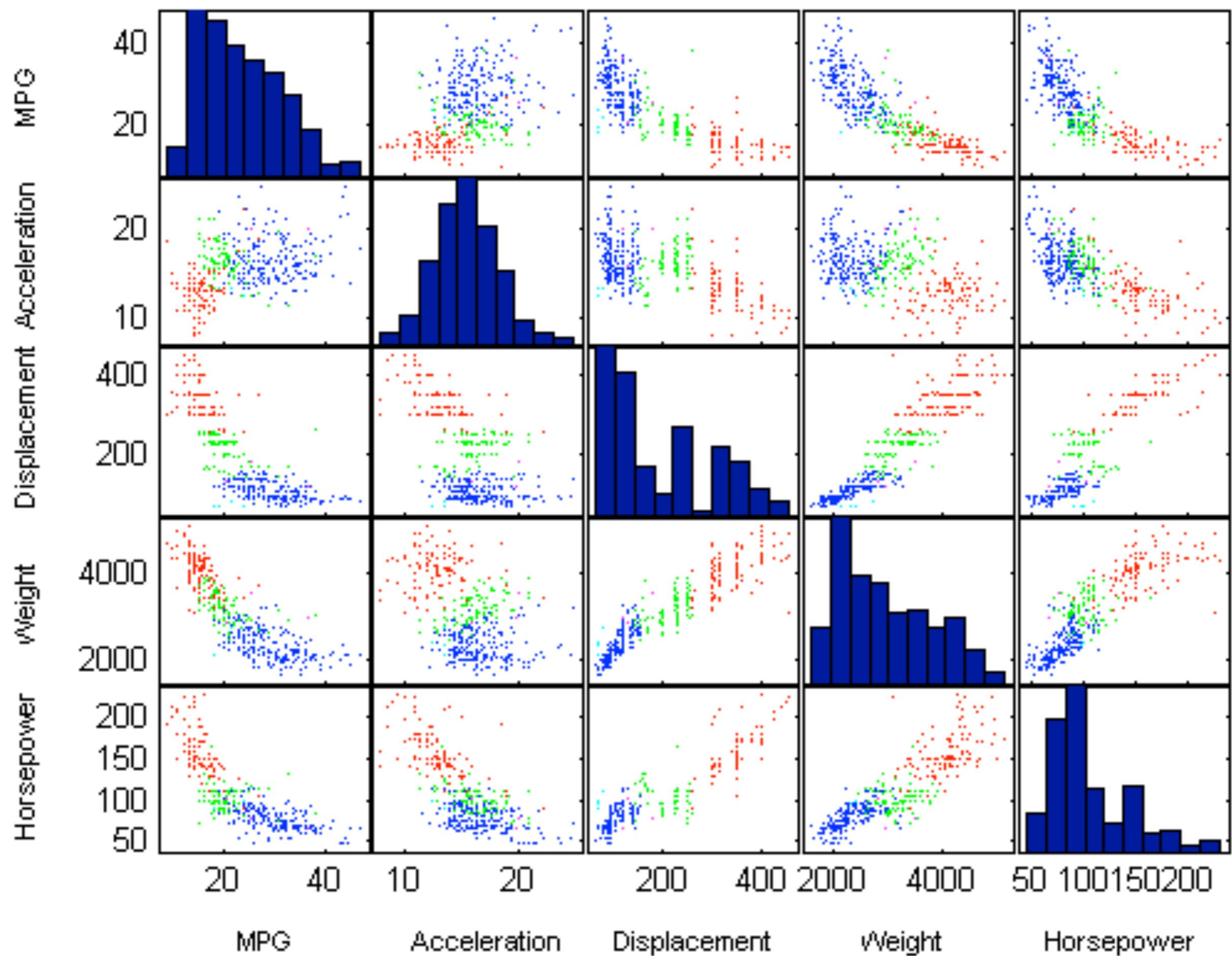




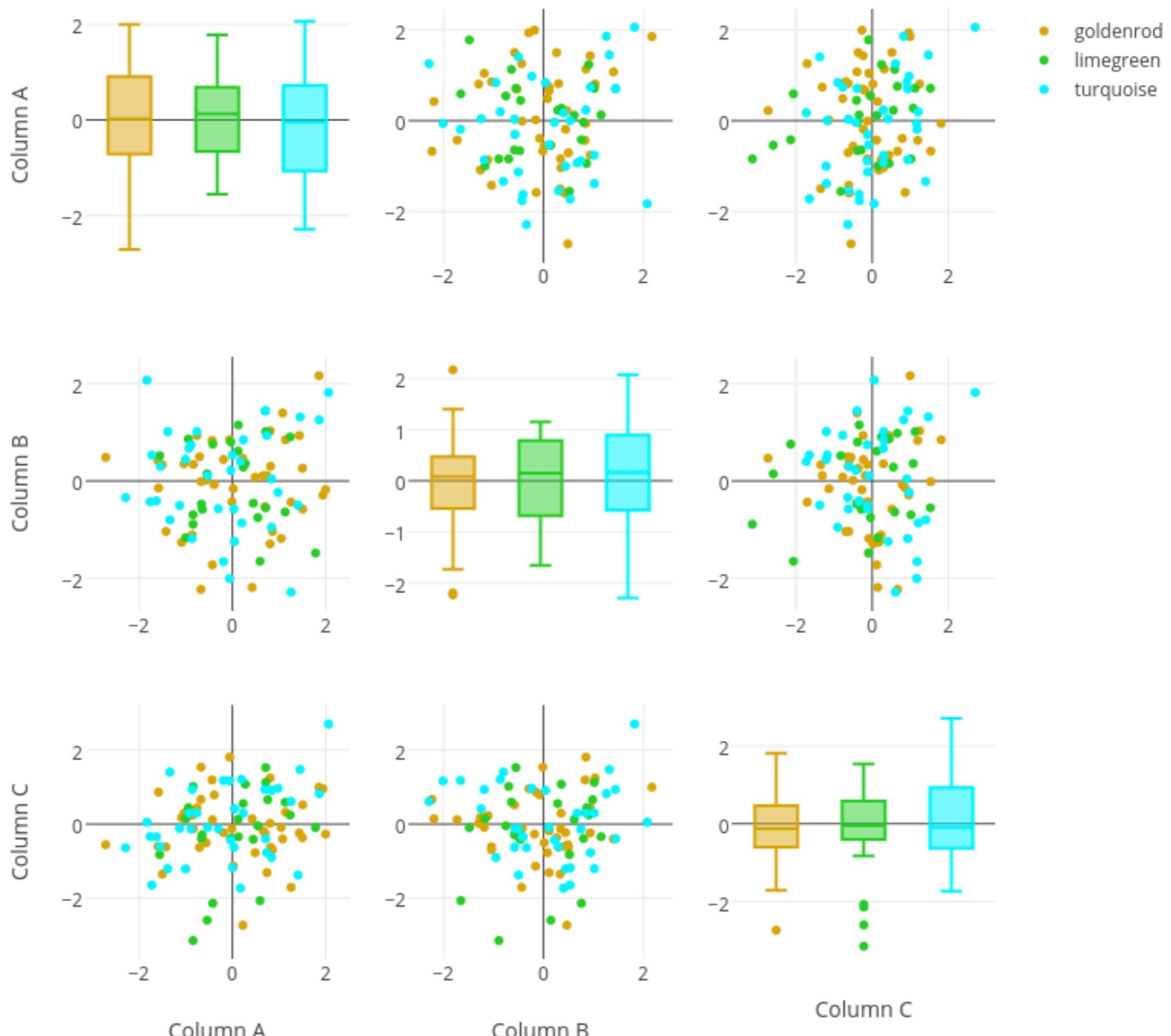


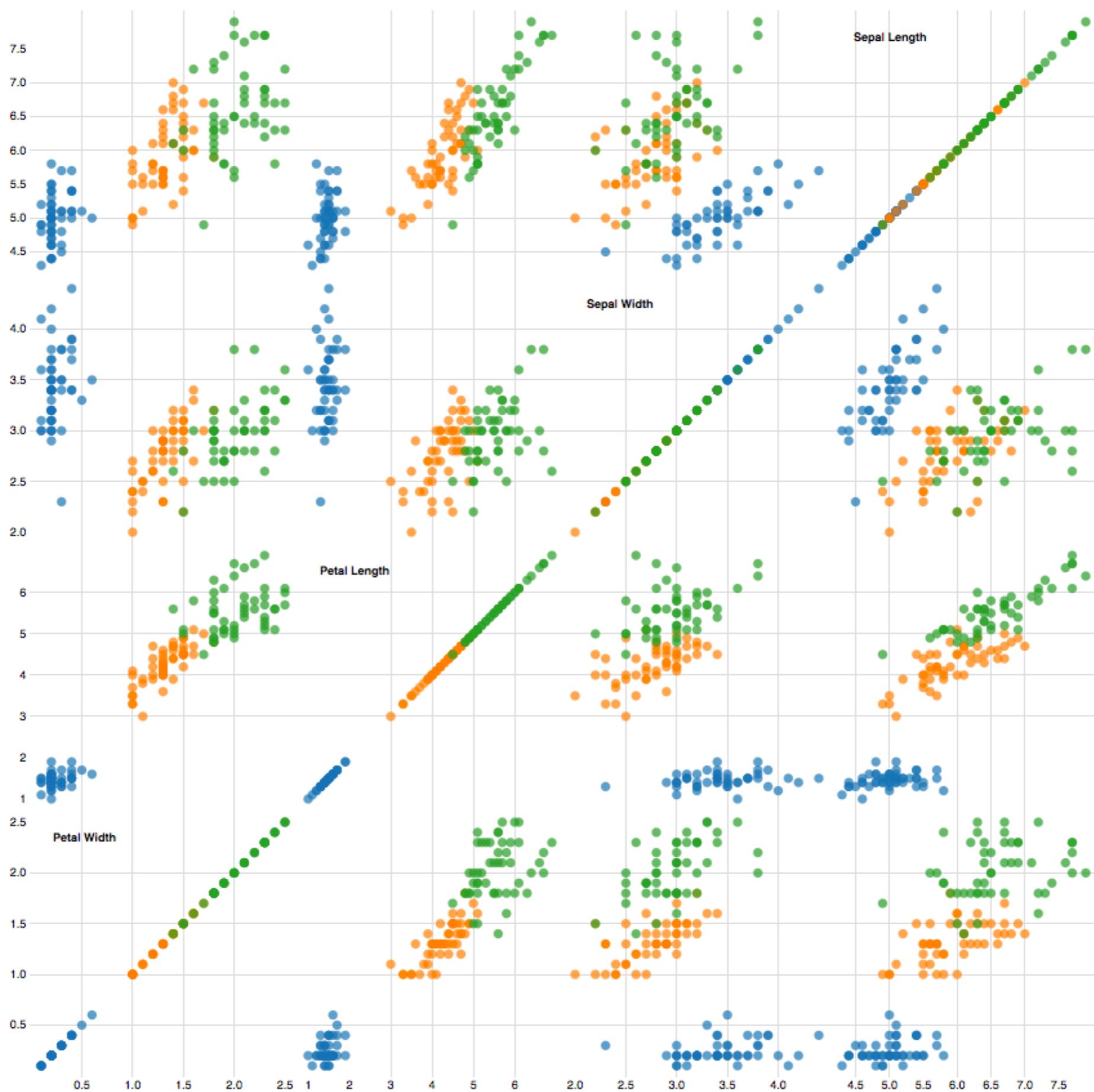


O que colocar na diagonal para aproveitar melhor o espaço com dados?  
(princípio Data-ink de Tufte)



Scatterplot Matrix

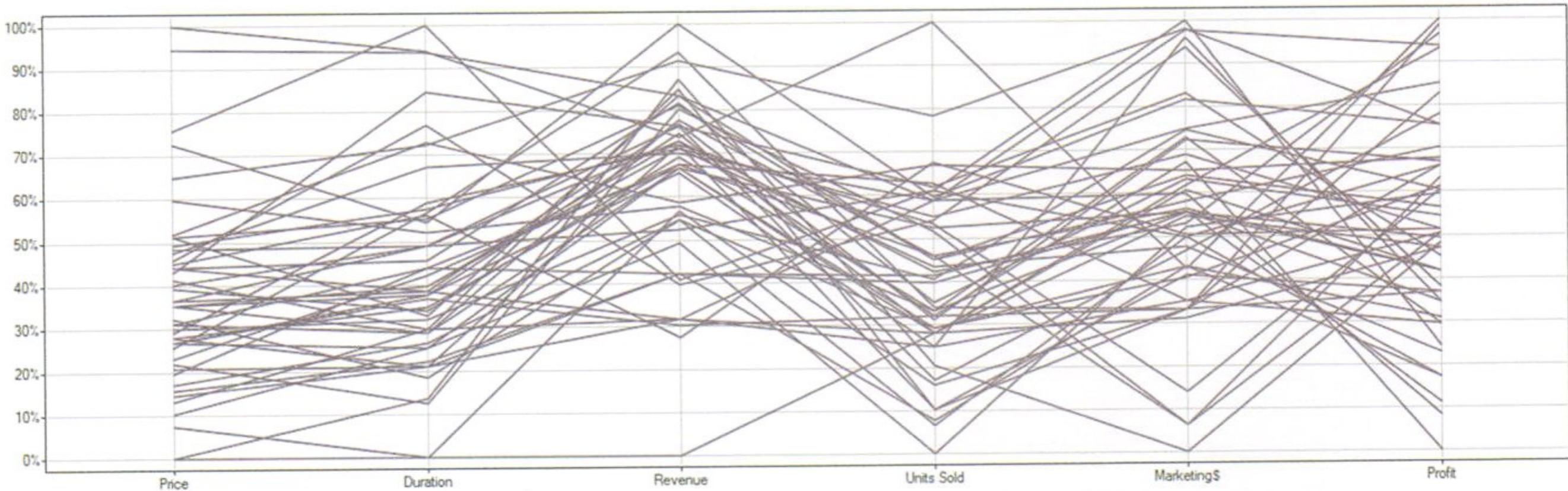




# COORDENADAS PARALELAS

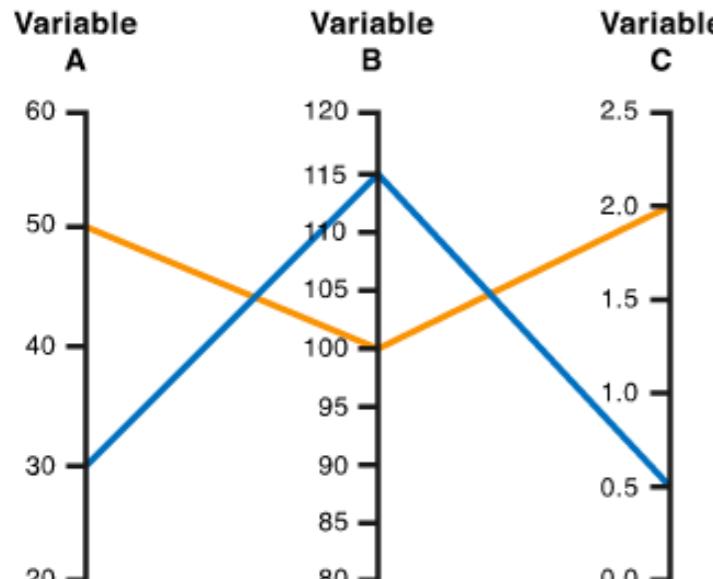
---

- Para representar um conjunto de pontos em um espaço n-dimensional, um plano é desenhado e sobre ele n linhas verticais paralelas e igualmente espaçadas.
- Um ponto no espaço n-dimensional é representado como uma polilinha com vértices nos eixos paralelos
- A posição do vértice no i-ésimo eixo corresponde à i-ésima coordenada do ponto.

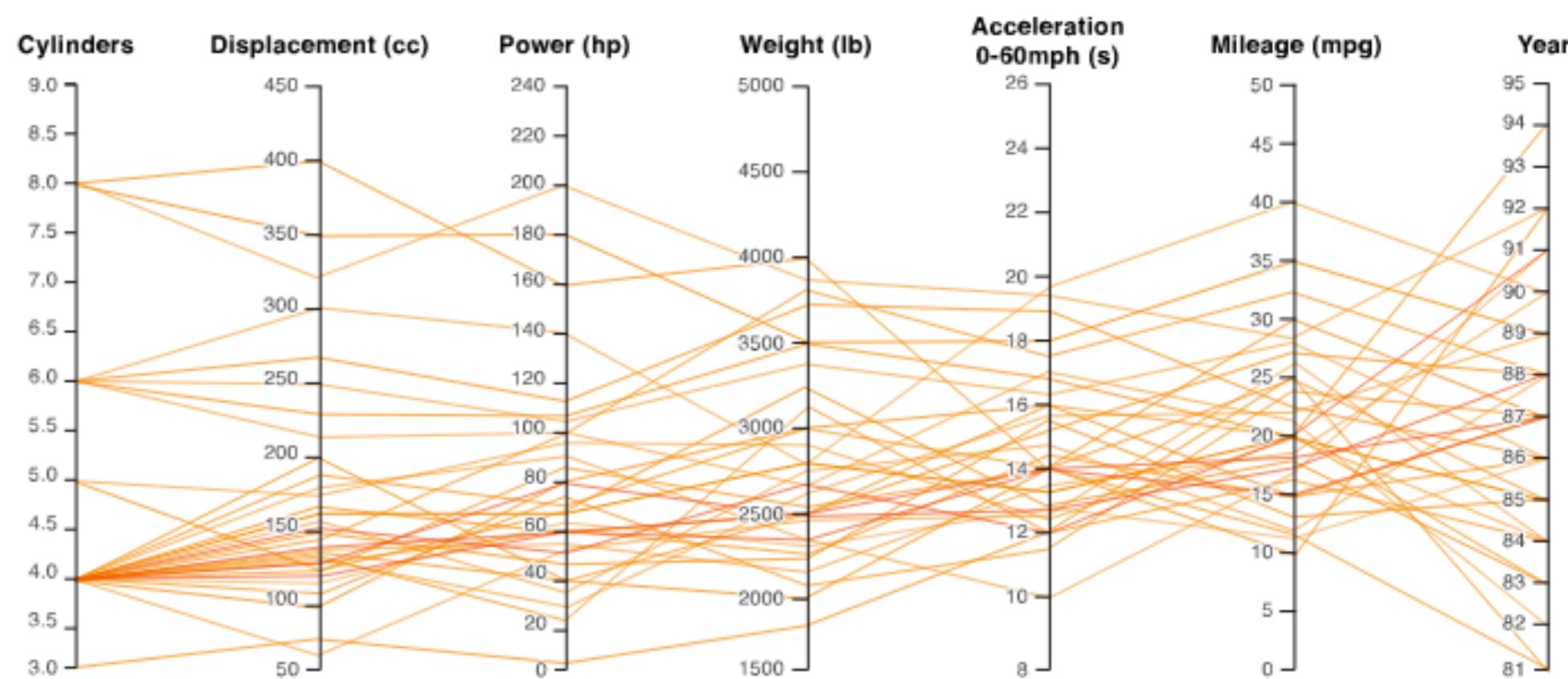


# COORDENADAS PARALELAS

- Para representar um conjunto de pontos em um espaço n-dimensional, um plano é desenhado e sobre ele n linhas verticais paralelas e igualmente espaçadas.
- Um ponto no espaço n-dimensional é representado como uma polilinha com vértices nos eixos paralelos
- A posição do vértice no i-ésimo eixo corresponde à i-ésima coordenada do ponto.



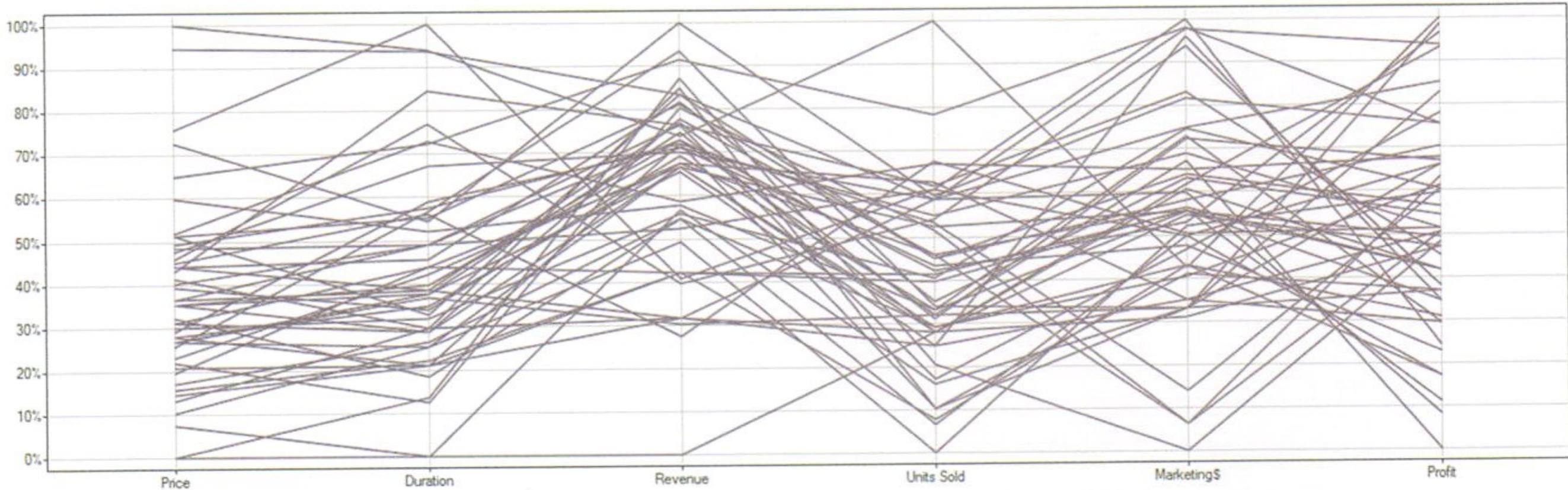
Data			
	Variable A	Variable B	Variable C
Item 1	50	100	2.0
Item 2	30	115	0.5

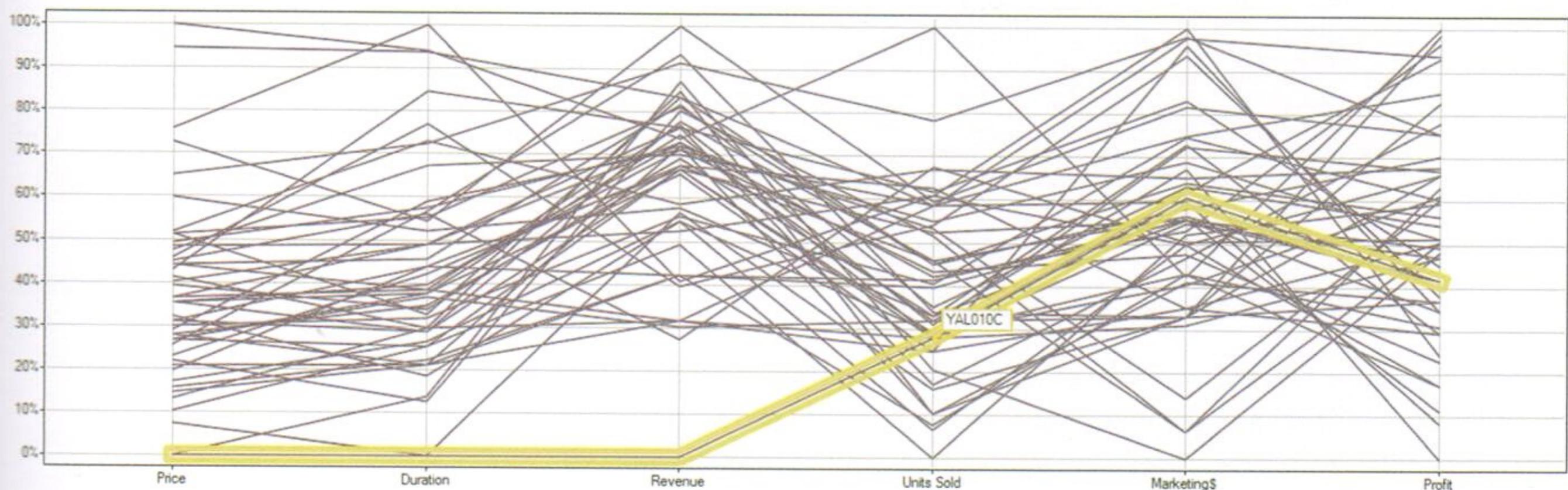


# COORDENADAS PARALELAS

---

- São gráficos extremamente complexos
- São capazes de mostrar alta densidade de dados
- Apesar de cada coluna representar uma variável diferente, a conexão entre elas revela o padrão multivariado





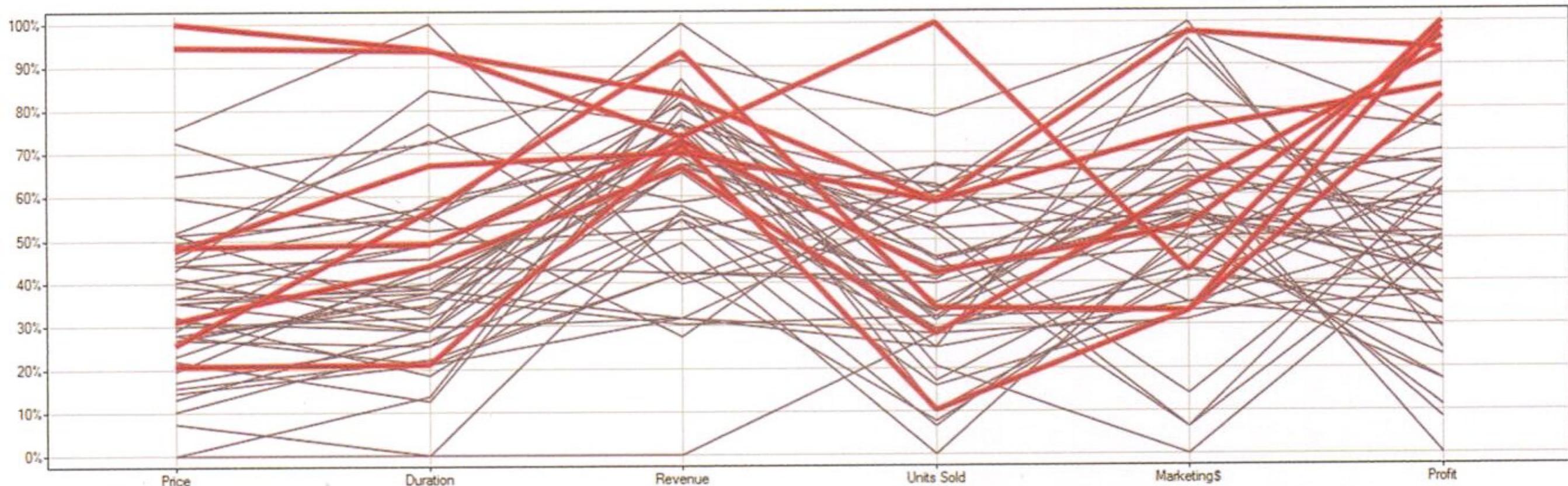
- O produto em destaque tem a menor renda que todos os outros (talvez devido ao baixo preço e ciclo de vida curto)
- Um produto é muito mais vendido que o seu competidor mais próximo
- A maioria dos produtos está no mercado entre 20% e 60% do tempo total avaliado
- Há uma concentração das rendas entre 65% e 85%

- Como se pode perceber, mesmo com o grande volume de curvas, padrões, tendências e exceções podem ser identificados
- Análises mais profundas dependem de interações com a visualização

# EXEMPLO

---

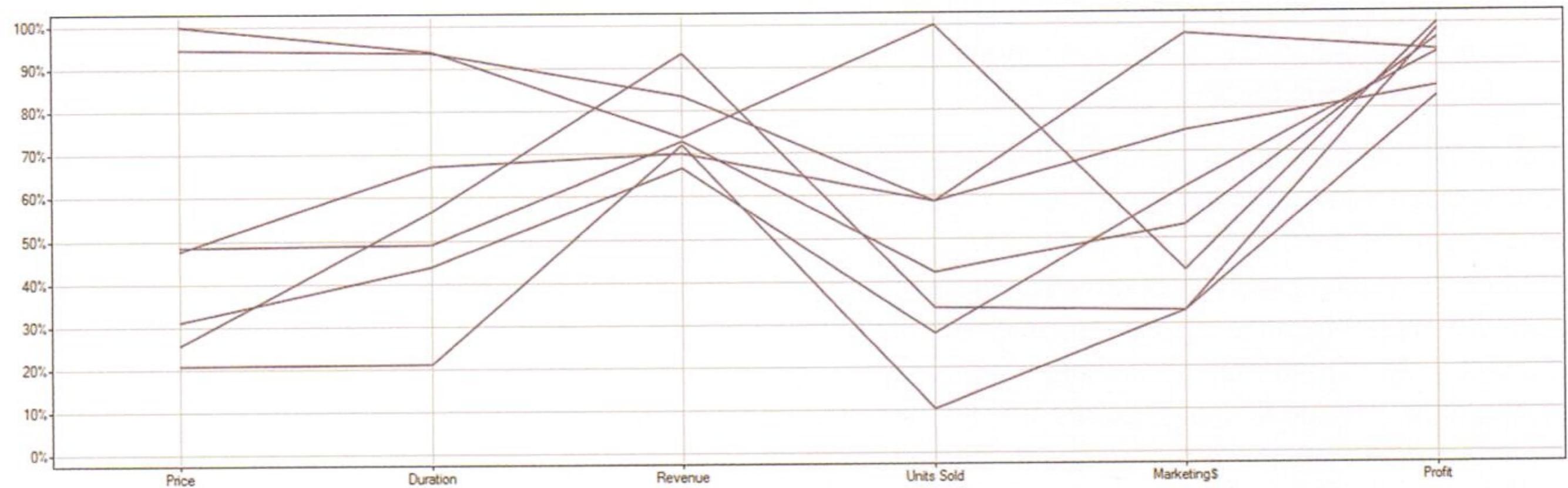
- Selecionar todos os produtos com lucros acima de 80%

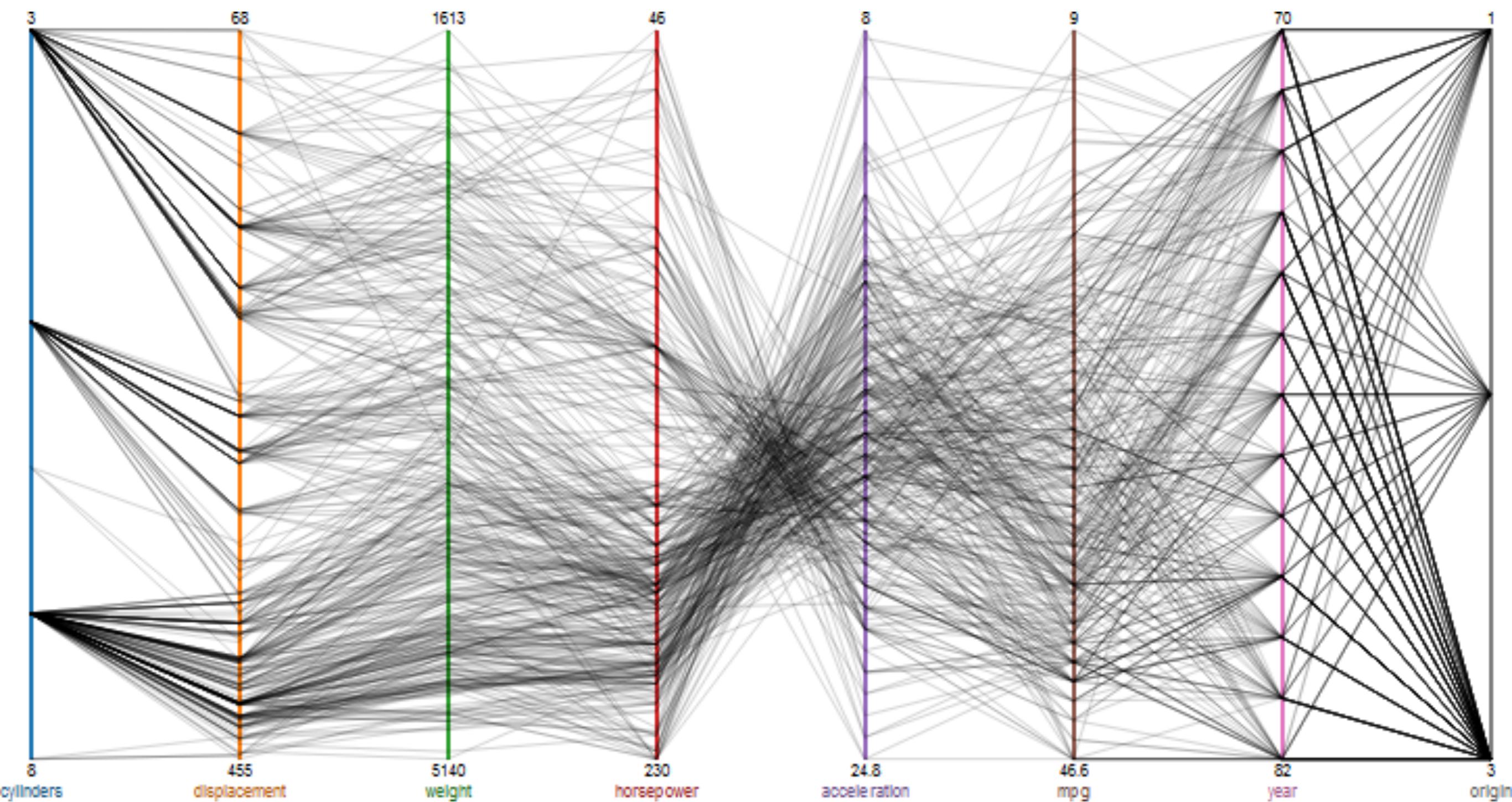


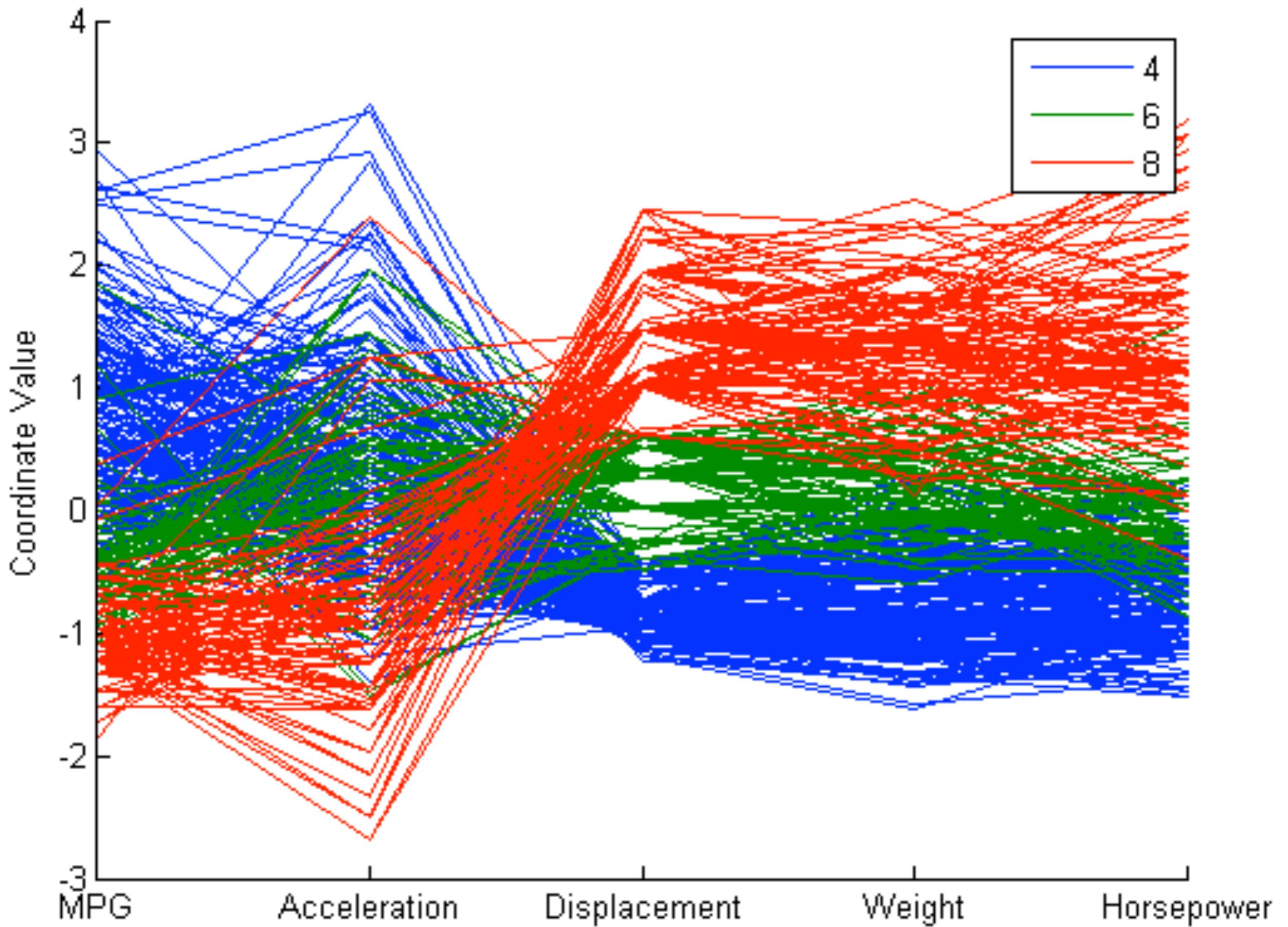
# EXEMPLO

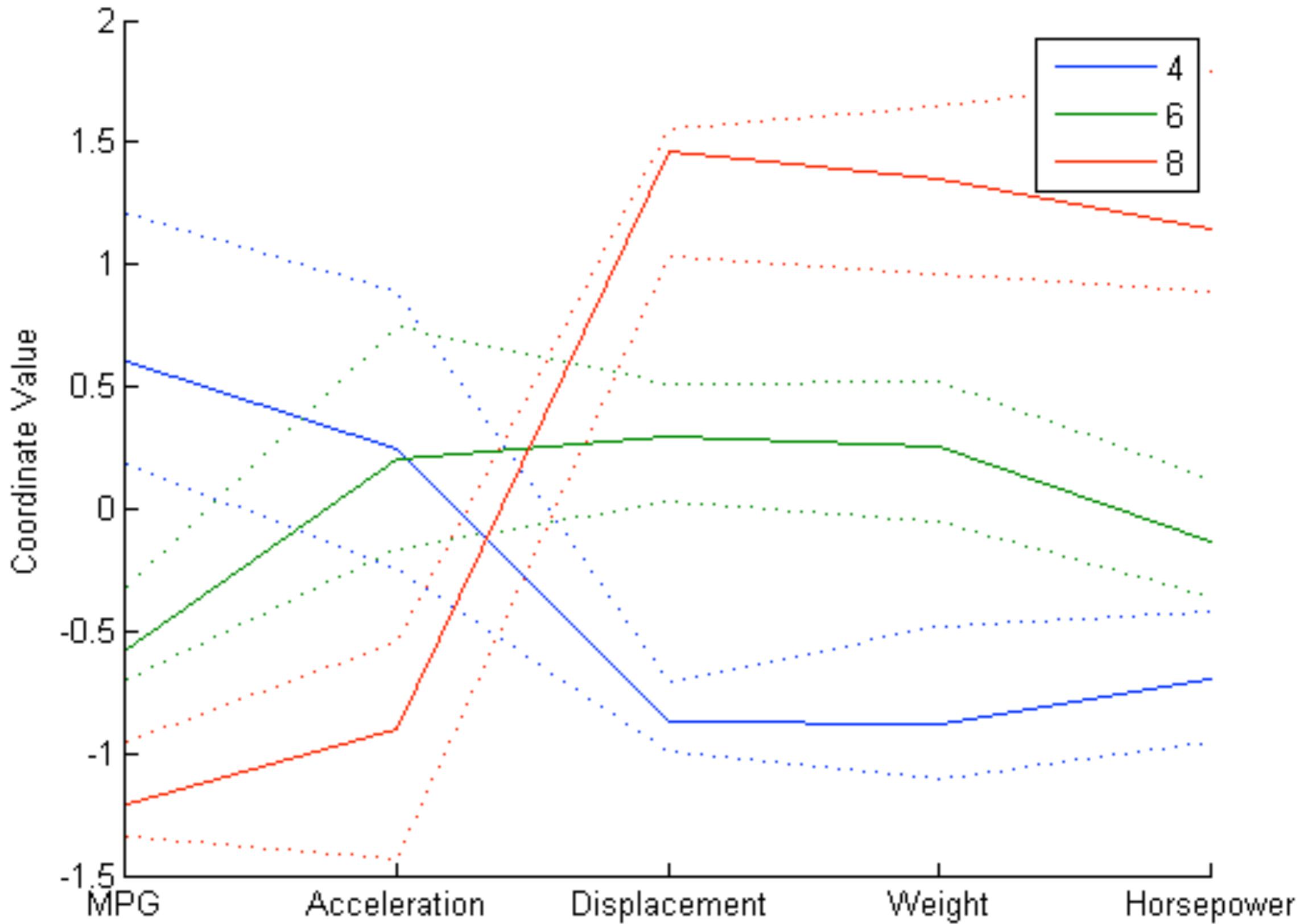
---

- Todos os produtos com alto lucro geram alta renda
- Estes produtos apresentam sempre despesas com marketing acima de 30%

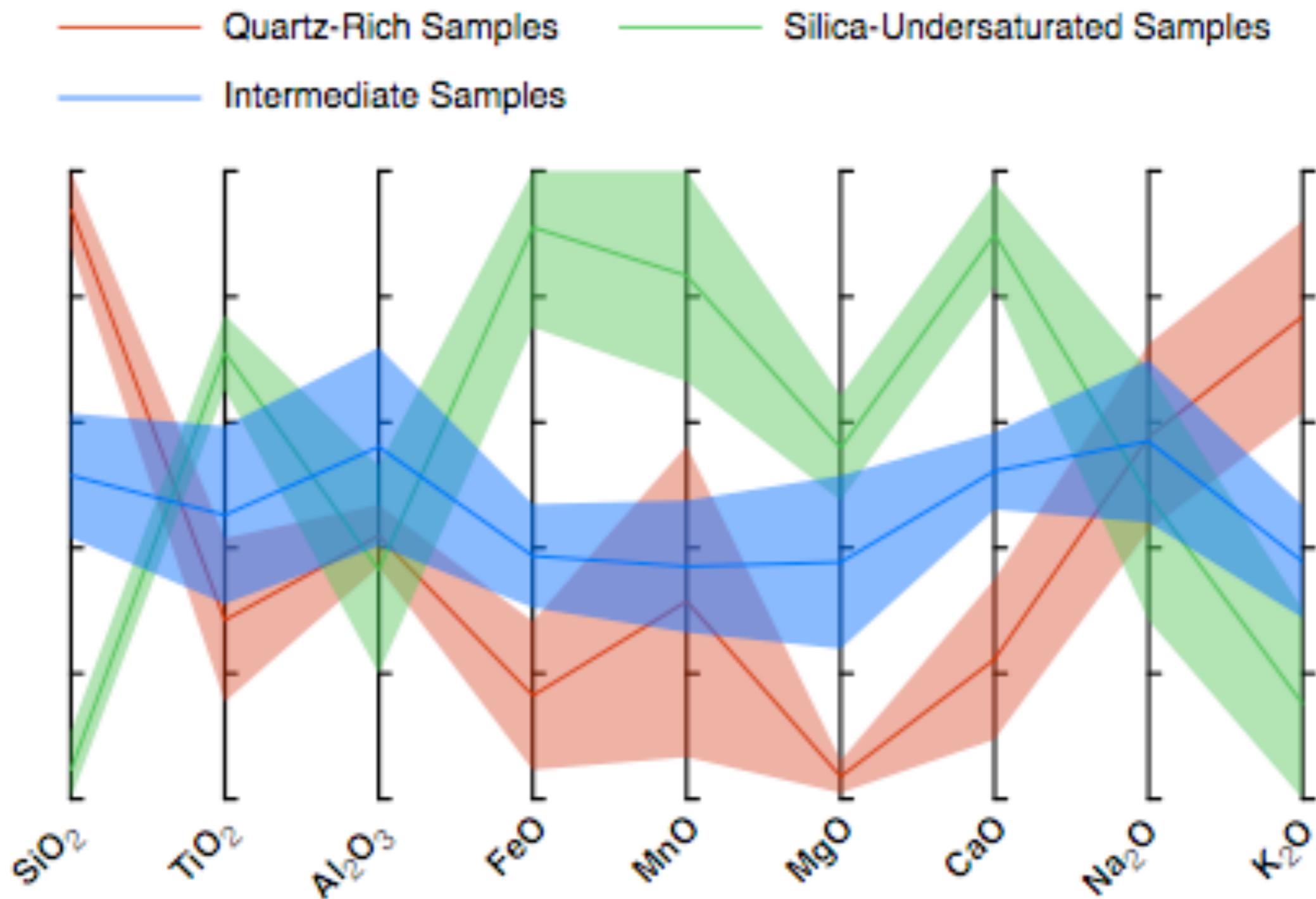






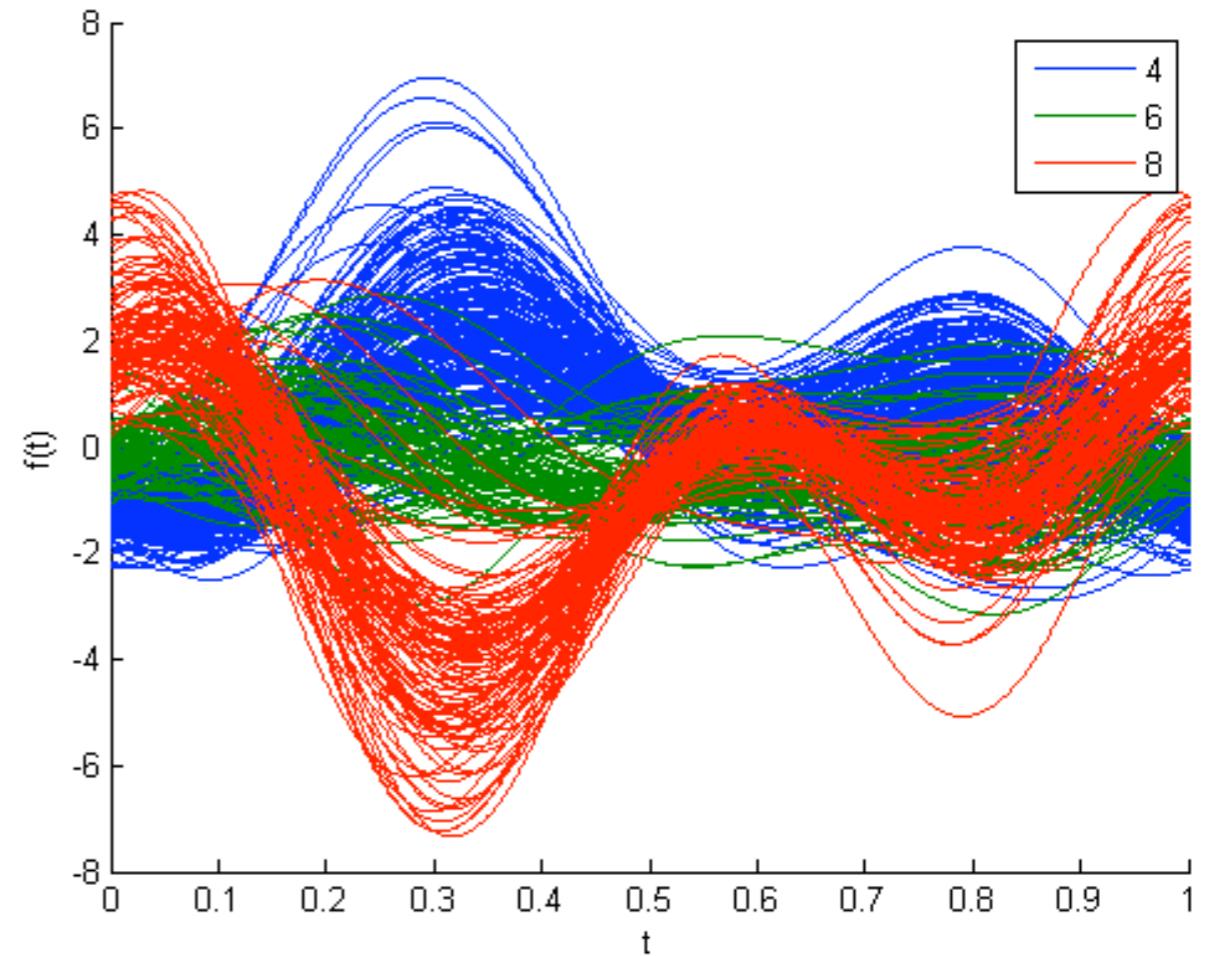


A linha sólida indica a mediana e as linhas pontilhadas, o 25º e o 75º percentil

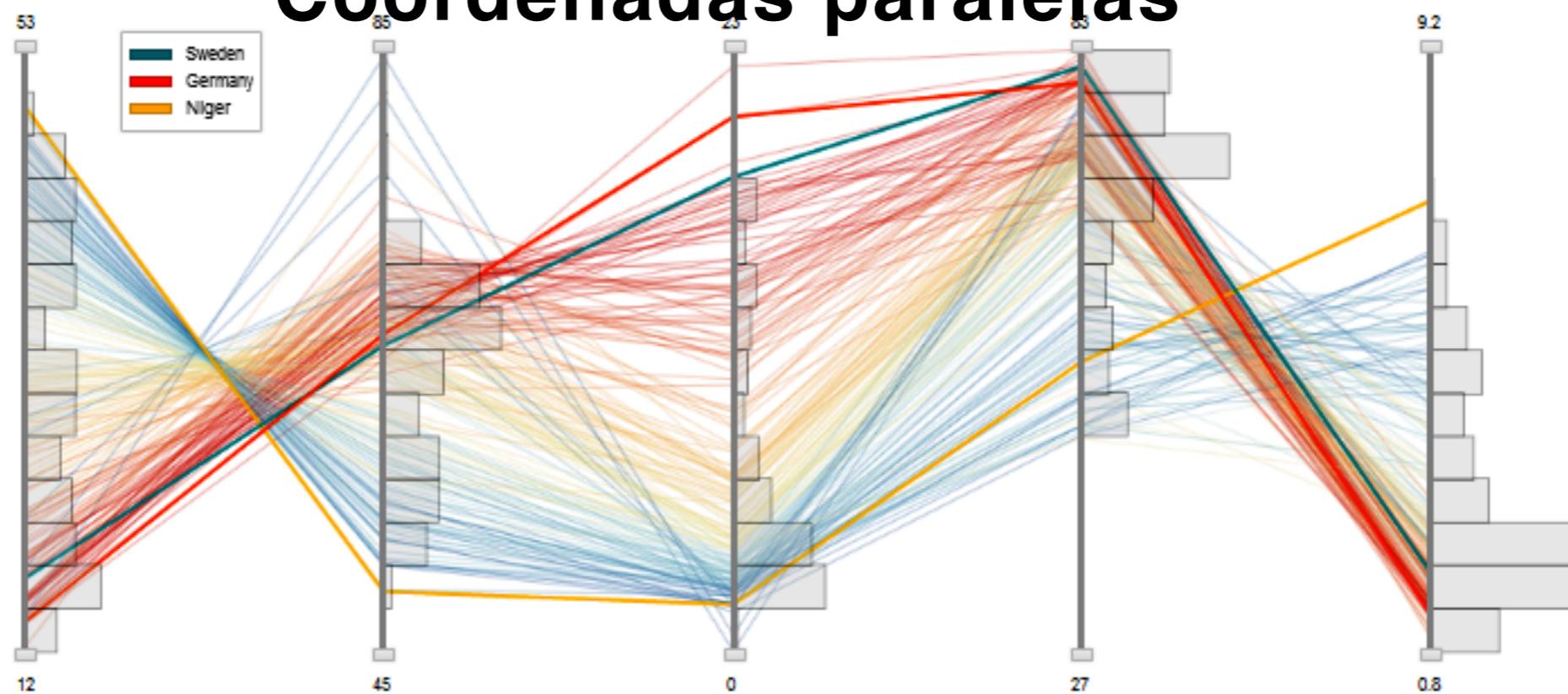


# Andrew curves

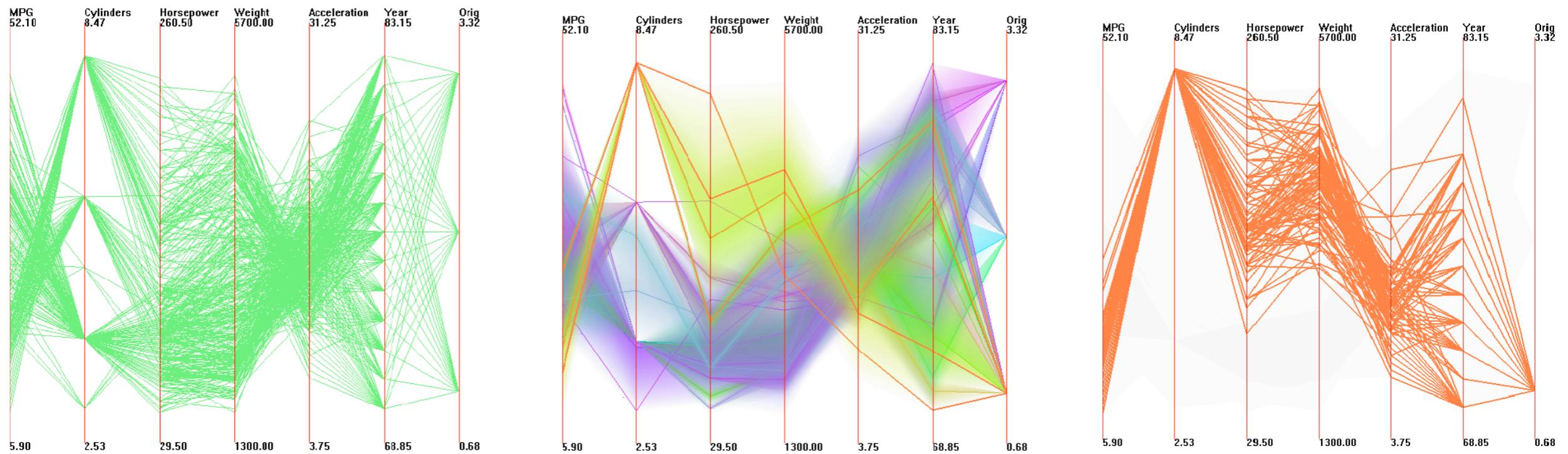
Cada ponto é representado por uma curva obtida por uma transformada de Fourier



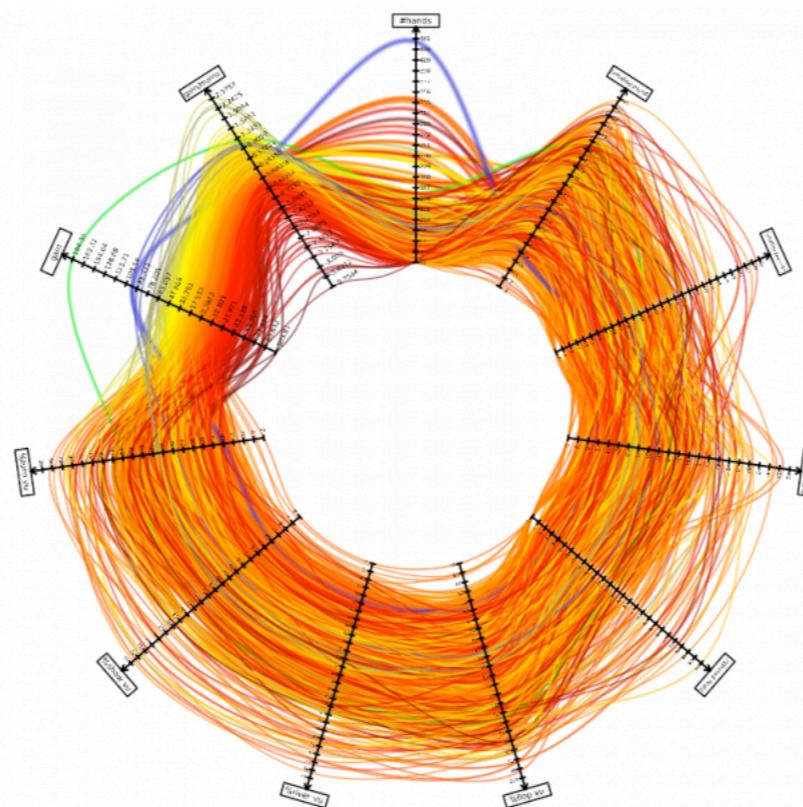
## Coordenadas paralelas



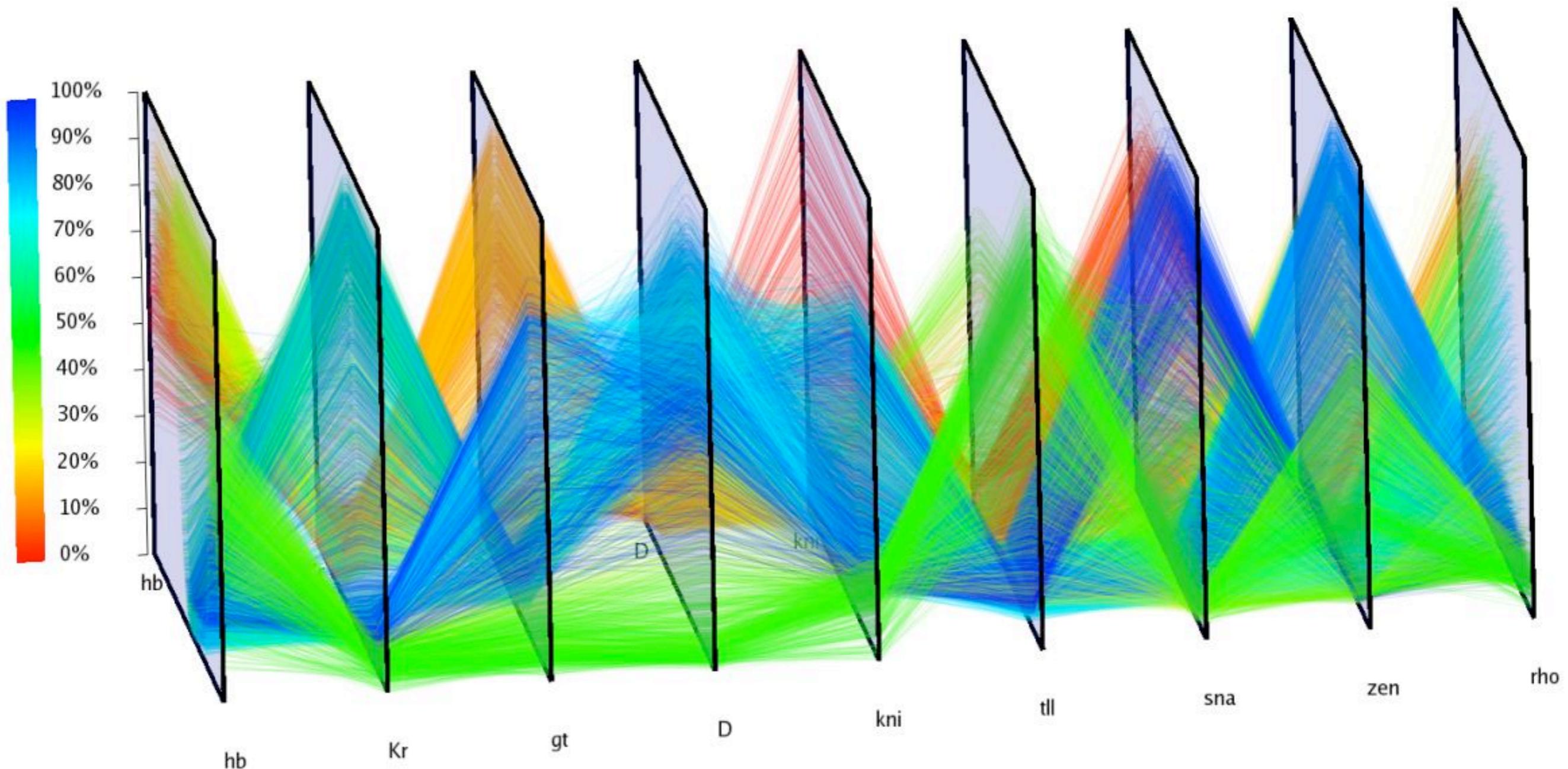
## Coordenadas paralelas



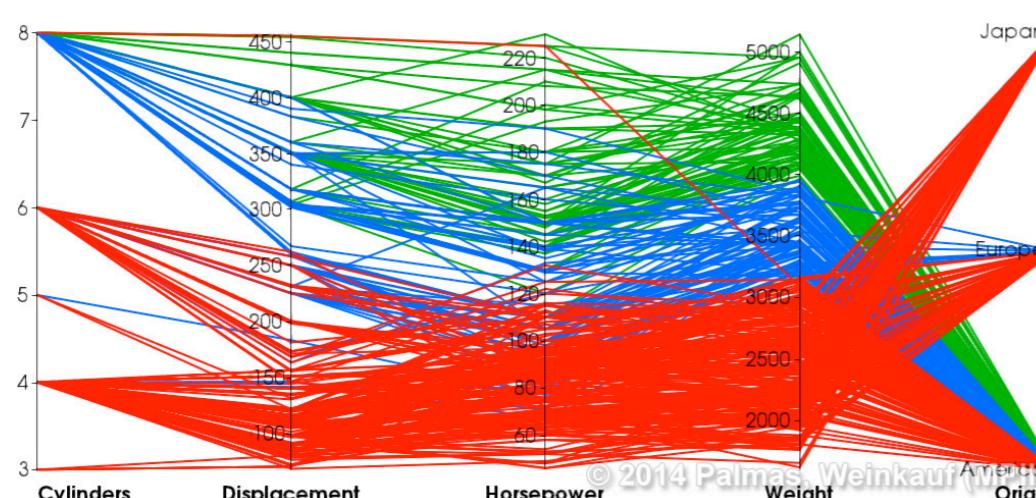
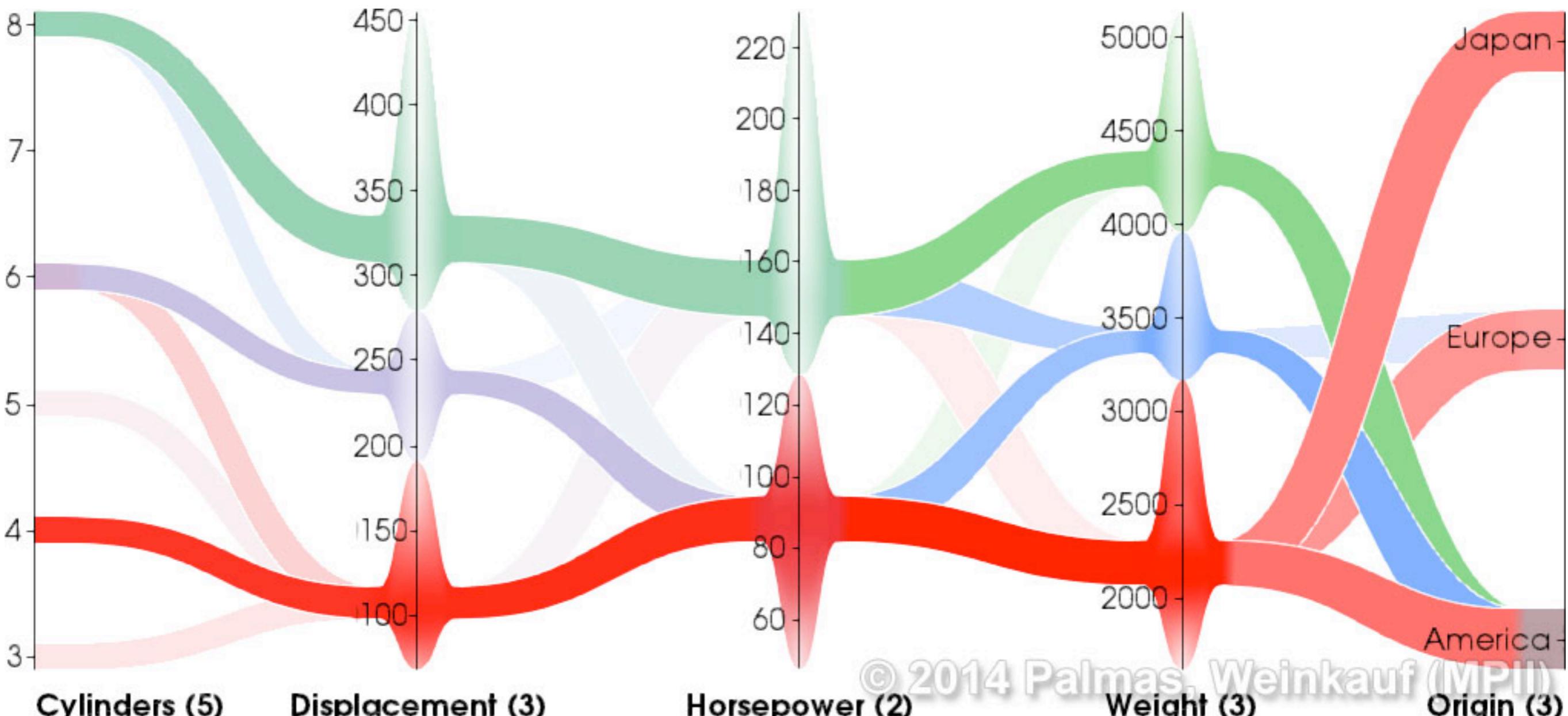
## Coordenadas paralelas



## Coordenadas paralelas



# Coordenadas paralelas com



G. Palmas, M. Bachynskyi, A. Oulasvirta, H.-P. Seidel, T. Weinkauf  
An Edge-Bundling Layout for Interactive Parallel Coordinates  
Proc. IEEE PacificVis, Yokohama, Japan, March 4 - 7, 2014

# Nutrient Explorer

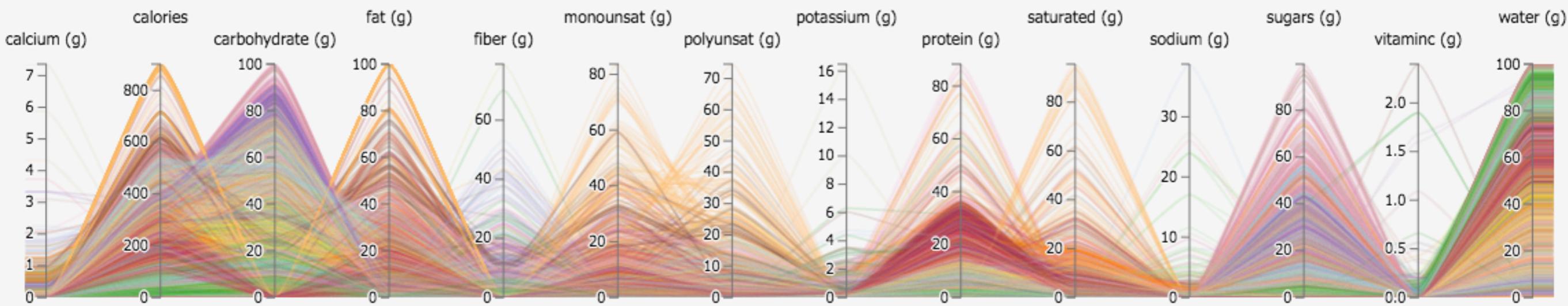
Keep Exclude Export

7637/7637

Lines at 13.6% opacity.

Hide Ticks

Dark



## What is this?

A multidimensional explorer of [nutrient data](#) from the [USDA](#).

The parallel coordinates displays the nutrient content of foods in the database across 14 dimensions, colored by food group.

Never heard of parallel coordinates? [Read this tutorial](#).

[Brush](#) the visualization to update other charts on this page.

Let me know what you think on [Reddit](#).

## Controls

**Brush:** Drag vertically along an axis.

**Remove Brush:** Tap the axis background.

**Reorder Axes:** Drag a label horizontally.

**Invert Axis:** Tap an axis label.

**Remove Axis:** Drag axis label to the left edge.

## Credits & License

Adapted from examples by

[Mike Bostock](#) and [Jason Davies](#)

Copyright © 2012, Kai Chang

All rights reserved. Released under the [BSD](#)

[License](#).

## Food Groups

327 Baby Foods
496 Baked Products
619 Beef Products
278 Beverages
403 Breakfast Cereals
183 Cereal Grains and Pasta
239 Dairy and Egg Products
165 Ethnic Foods
365 Fast Foods
215 Fats and Oils
255 Fintech and Shellfish Products
328 Fruits and Fruit Juices
345 Lamb, Veal, and Game Products
365 Legumes and Legume Products
57 Meals, Entrees, and Sidedishes
128 Nut and Seed Products
328 Pork Products
381 Poultry Products
51 Restaurant Foods
234 Sausages and Luncheon Meats
162 Snacks
499 Soups, Sauces, and Gravies
61 Spices and Herbs
341 Sweets

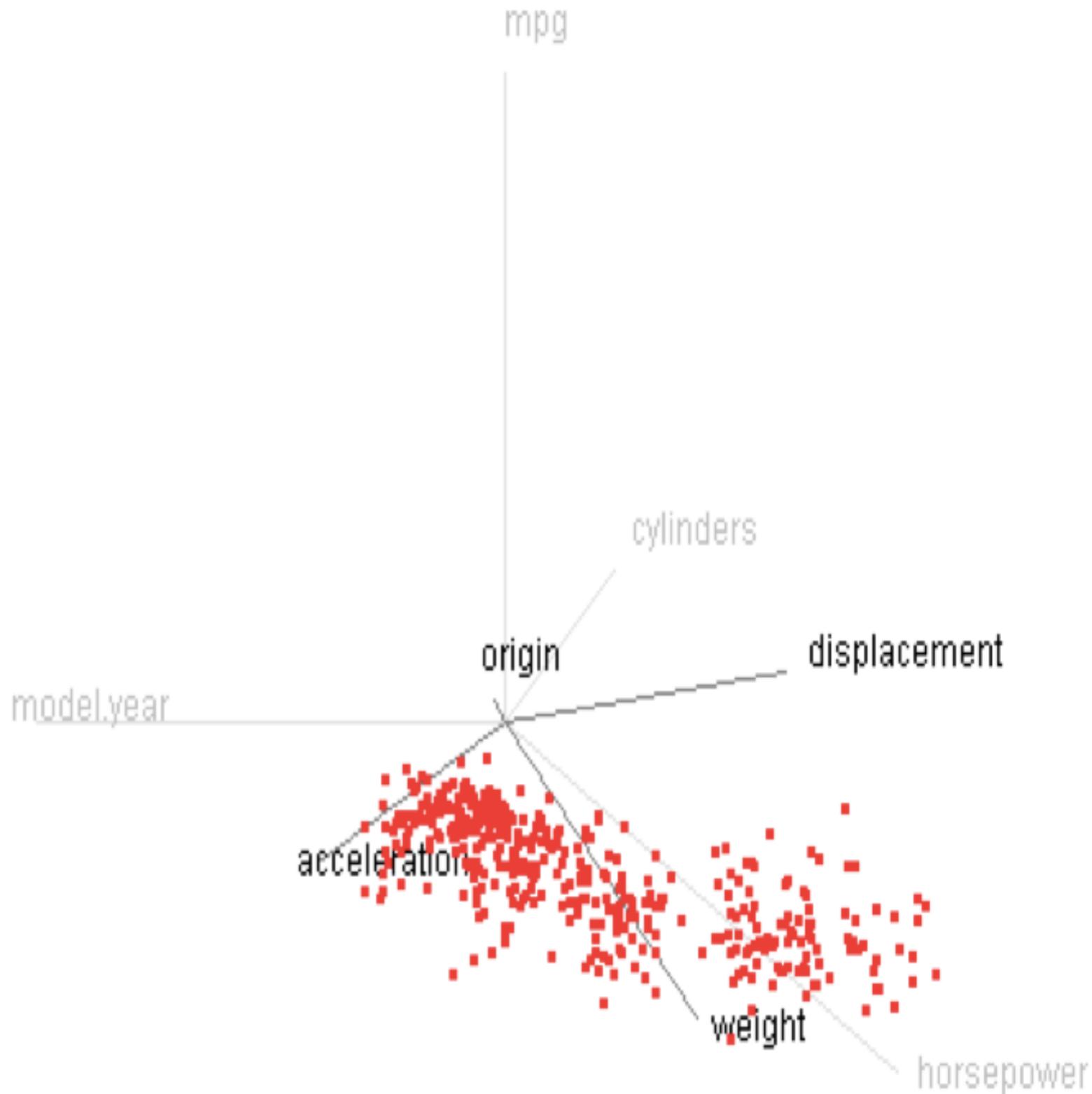
## Sample of 25 entries

- ARCHWAY Home Style Cookies, Ruth's Oatmeal
- Babyfood, dessert, peach cobbler, strained
- Bacon and beef sticks
- Beans, adzuki, mature seeds, canned, sweetened
- Beef, chuck, under blade center steak, boneless, Denver Cut, c
- Beef, flank, steak, separable lean and fat, trimmed to 0" fat, c
- Beef, round, eye of round, roast, separable lean only, trimme
- Candies, MARS SNACKFOOD US, TWIX chocolate fudge cookie
- Cereals ready-to-eat, POST, Honeycomb Cereal
- Cereals, QUAKER, Creamy Wheat, farina, enriched, prepared i
- Cookies, oatmeal, prepared from recipe, without raisins
- Cowpeas, catjang, mature seeds, cooked, boiled, with salt
- Cranberry juice cocktail, frozen concentrate, prepared with wa
- Fish, butterfish, raw
- Fish, mackerel, Atlantic, raw
- Game meat, elk, tenderloin, separable lean only, cooked, broil
- Granola bar, soft, milk chocolate coated, peanut butter
- KENTUCKY FRIED CHICKEN, Popcorn Chicken
- Lamb, domestic, leg, sirloin half, separable lean and fat, trimm
- Lamb, variety meats and by-products, spleen, cooked, braised
- Nuts, almonds, oil roasted, without salt added
- Seeds, sesame seeds, whole, roasted and toasted
- Soup, scotch broth, canned, prepared with equal volume wate
- Soup, stock, beef, home-prepared

# COORDENADAS ESTRELA

---

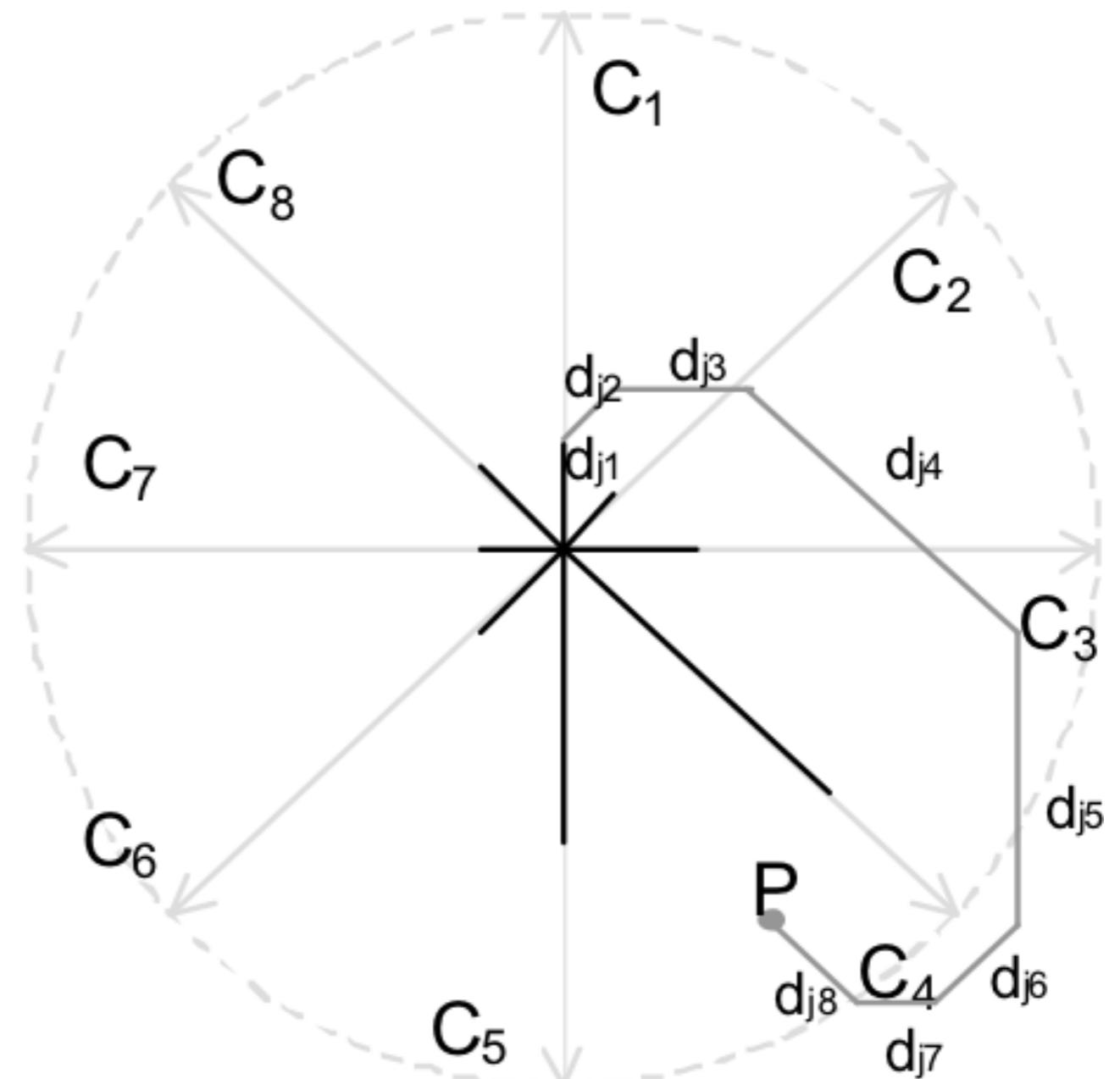
- Os eixos são distribuídos radialmente em um círculo separados por ângulos iguais inicialmente
- Pode-se aplicar deformações aos comprimentos dos eixos atribuindo maior ou menor ênfase a estes



# COORDENADAS ESTRELA

---

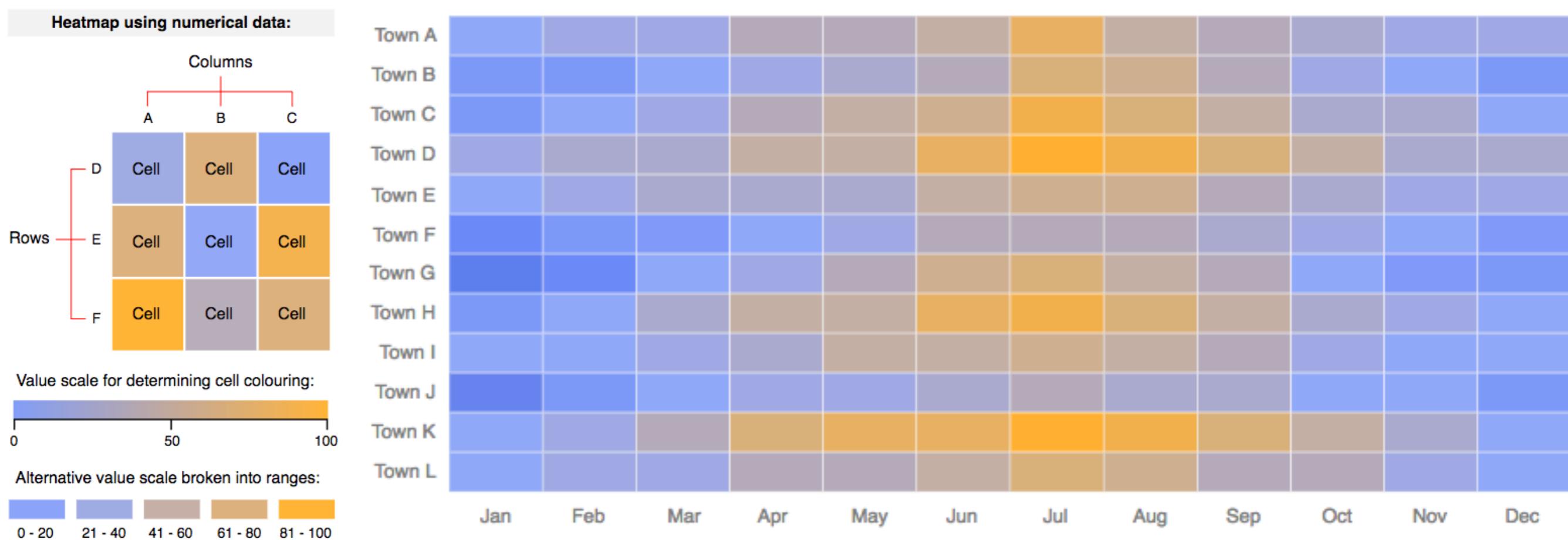
- Algoritmo para cálculo das posições dos pontos no novo sistema de coordenadas



# MAPAS DE CALOR

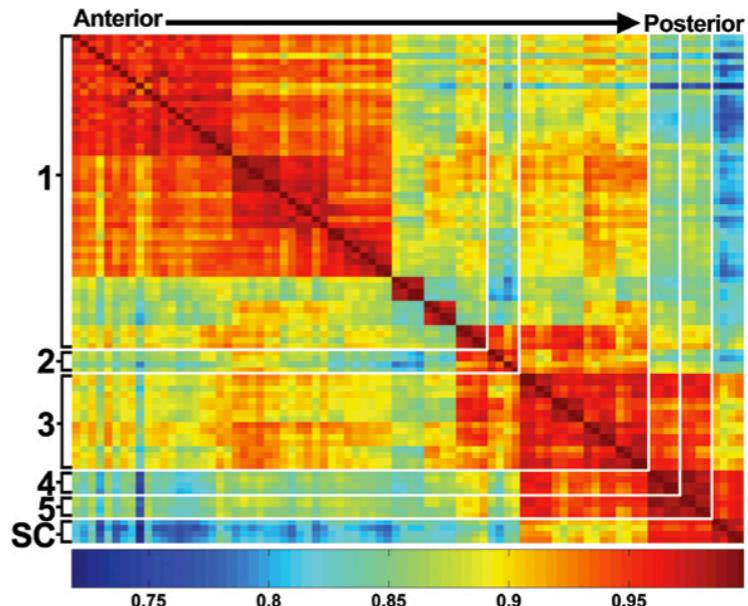
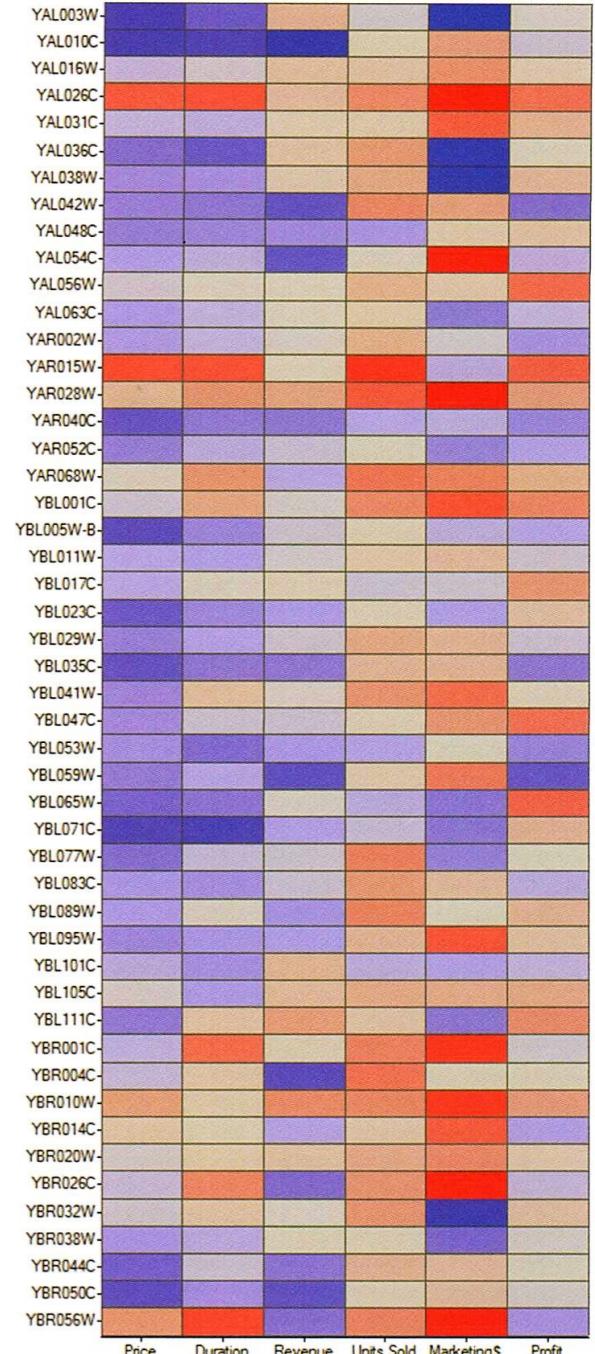
---

- Usam variações de cores para indicar quantidades
- Diferentes instâncias são representadas em linhas
- Cada coluna representa uma variável
- Células verdes indicam valores acima da média e vermelhos abaixo



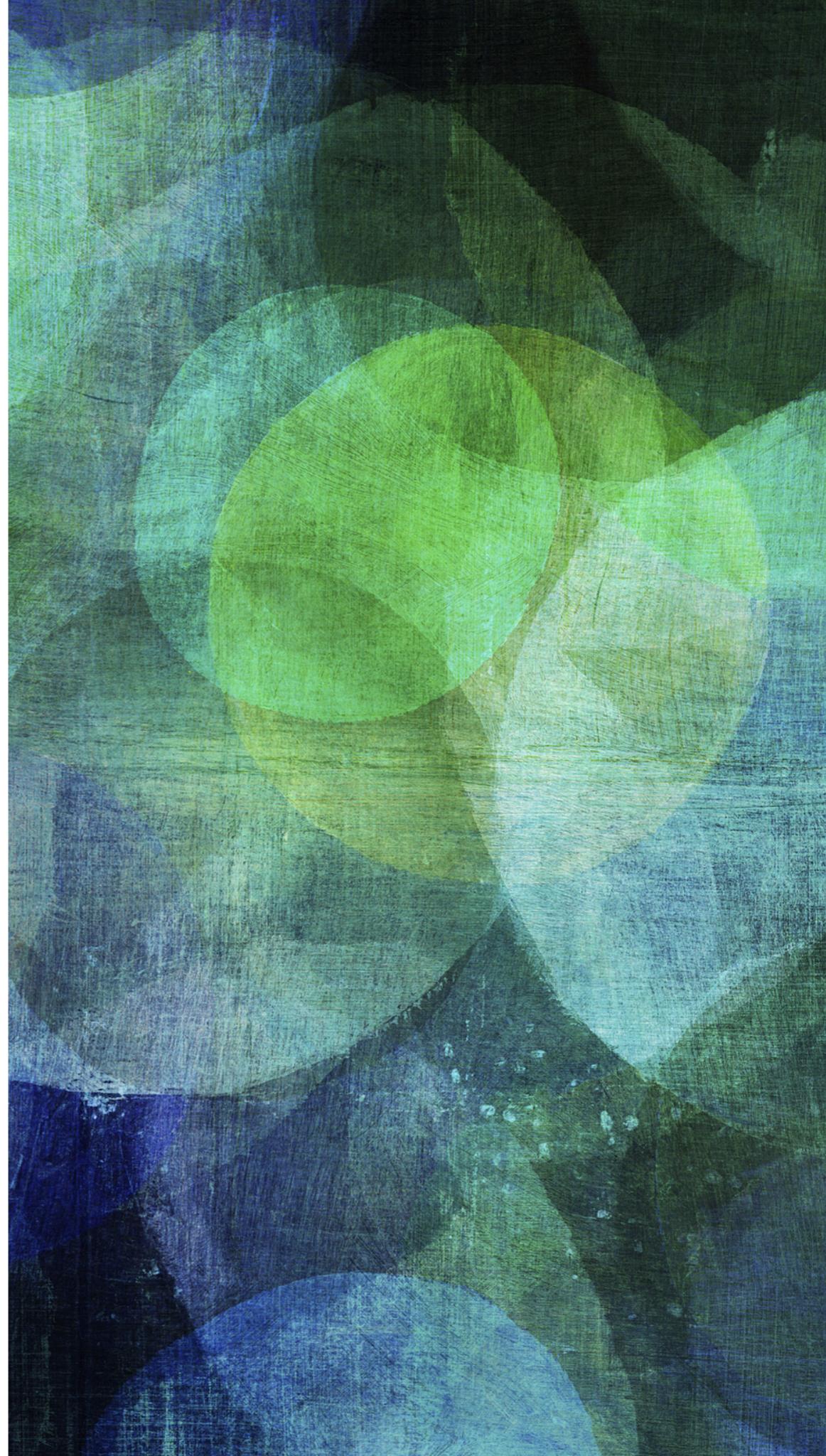
# PROBLEMA: CORES

- Cerca de 10% dos homens e 1% das mulheres não diferenciam bem entre verde e vermelho
- Preto não é uma boa cor para representar médias
- É difícil distinguir uma grande variedade de cores
- Nossa percepção de quantidades através de cores é imprecisa



# REPRESENTAÇÕES HIERÁRQUICAS

---



# REPRESENTAÇÕES HIERÁRQUICAS

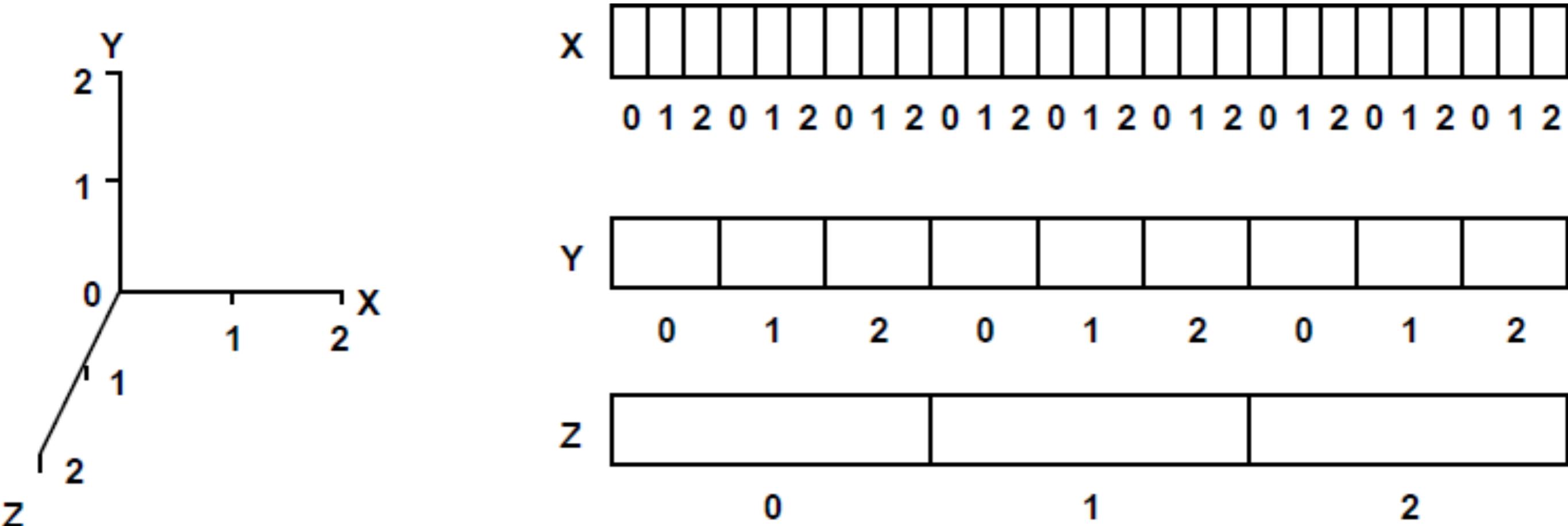
---

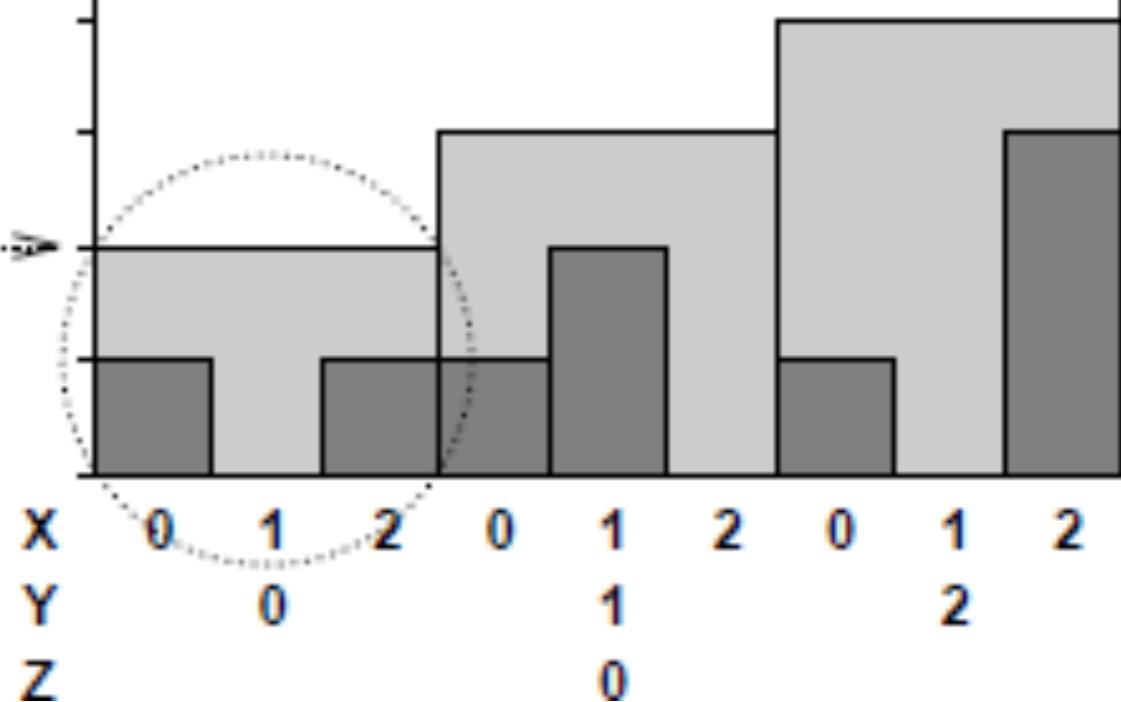
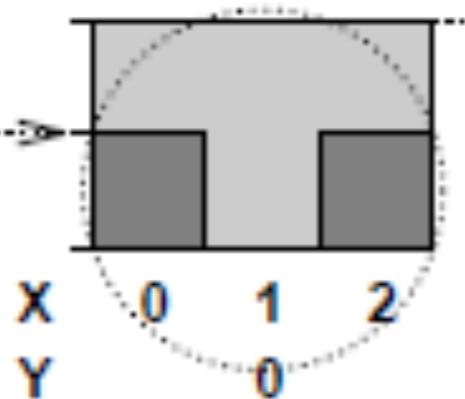
- Sub dividem o espaço e apresentam sub-espacos de modo hierárquico
- Atributos são tratados de forma diferente, portanto geralmente são úteis quando se trata de atributos hierárquicos
- Normalmente, treinamento é necessário para compreensão desse tipo de abordagem

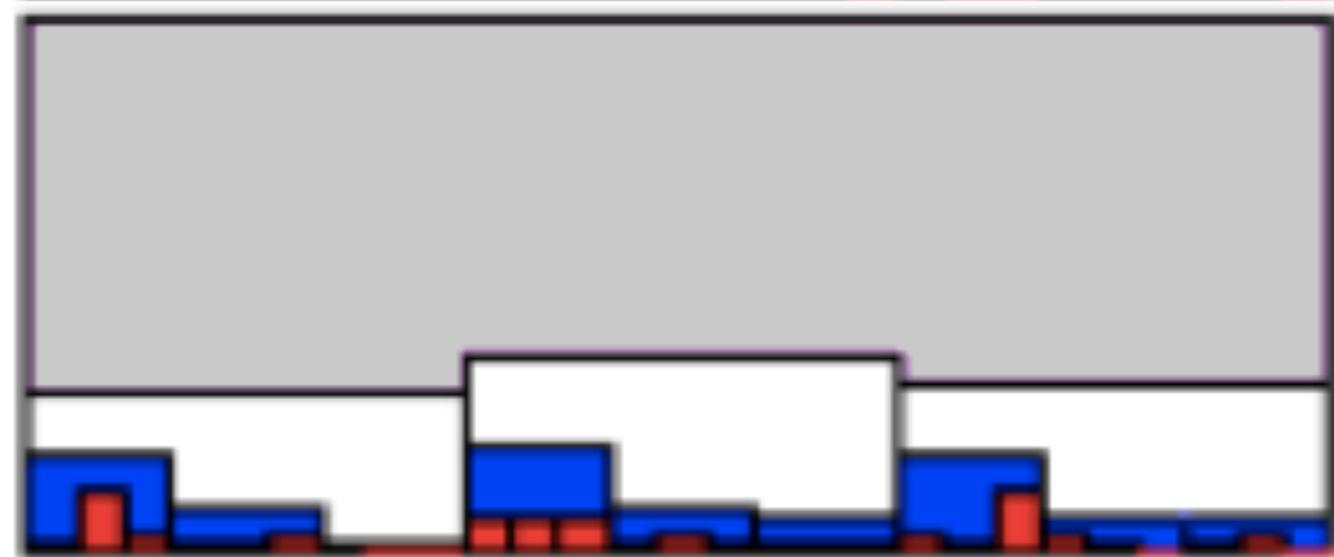
# EIXOS HIERÁRQUICOS

---

- Nos eixos hierárquicos, eixos ortogonais são representados horizontalmente de maneira hierárquica

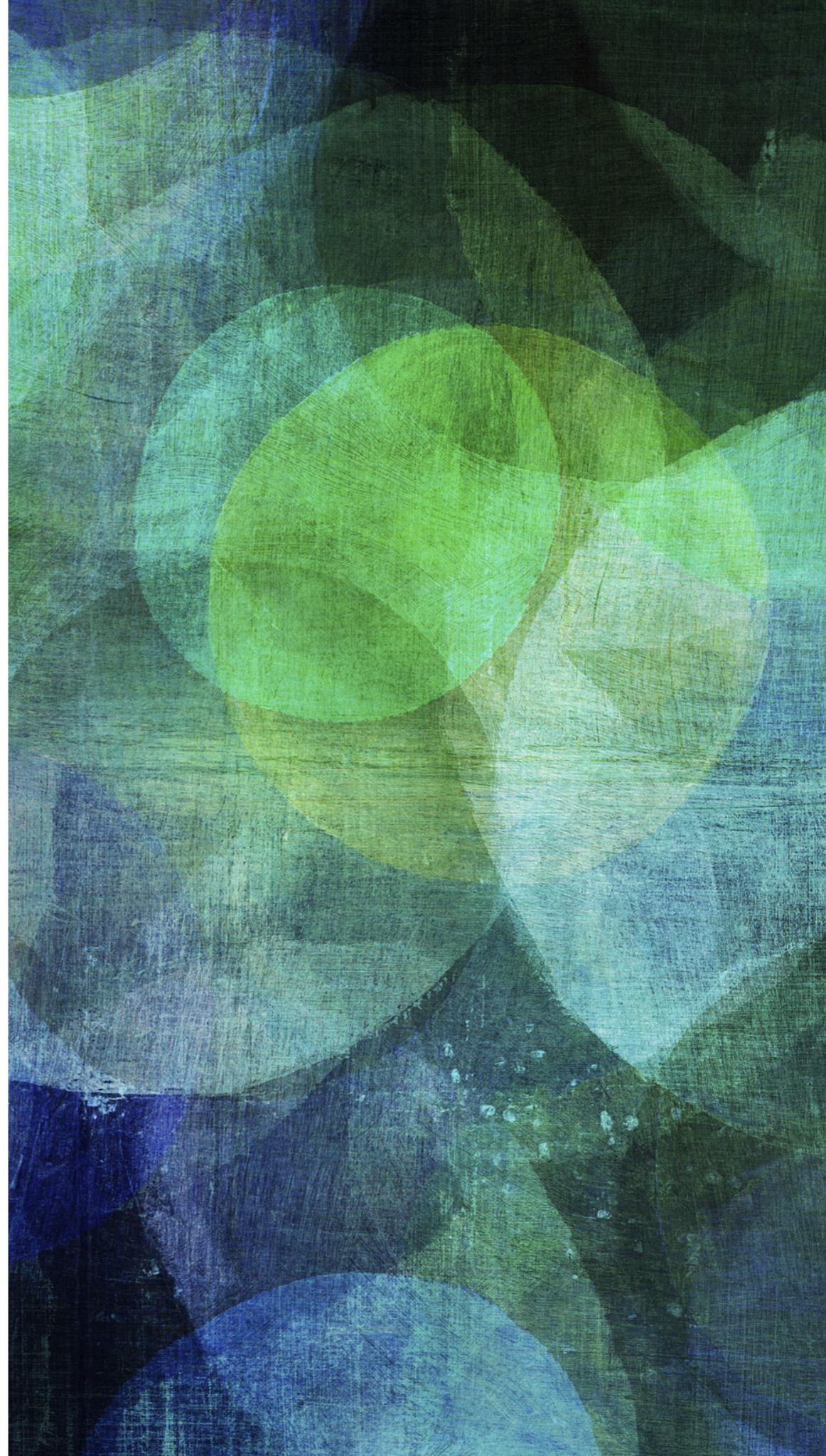






# REPRESENTAÇÕES ICONOGRÁFICAS

---



# REPRESENTAÇÕES ICONOGRÁFICAS

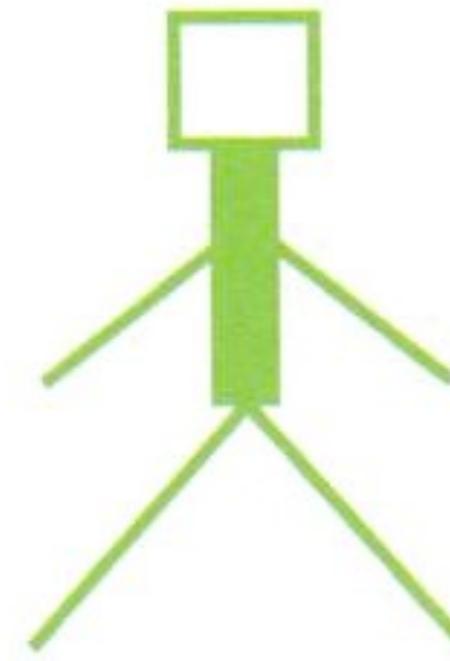
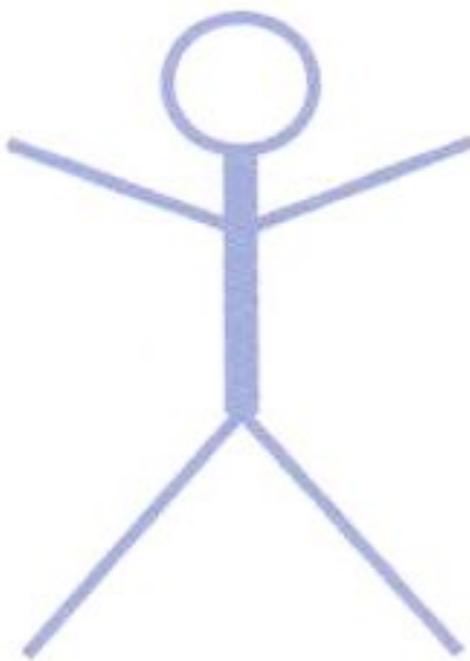
---

- Consistem no mapeamento de objetos multidimensionais em ícones ou glifos
- As características visuais destes ícones dependem dos atributos

# GLIFOS

---

- São objetos gráficos compostos por um conjunto de atributos visuais



## Atributo visual

## Variável

Cor

Temperatura corporal

Formato da cabeça

Tipo sanguíneo

Largura do tronco

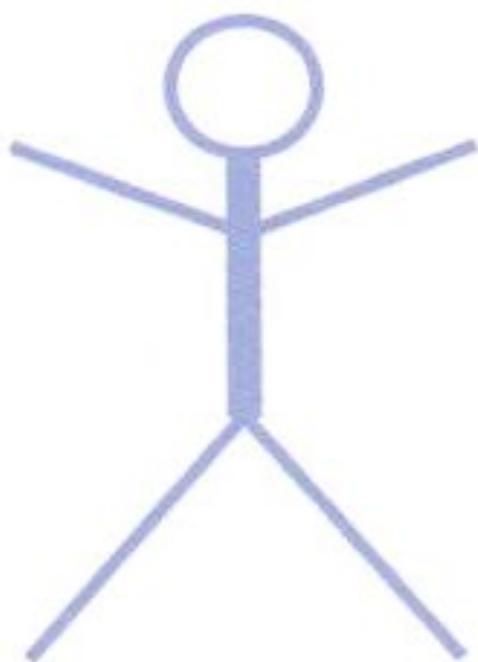
IMC

Posição dos braços

Batimentos cardíacos

Posição das pernas

Taxa de açúcar

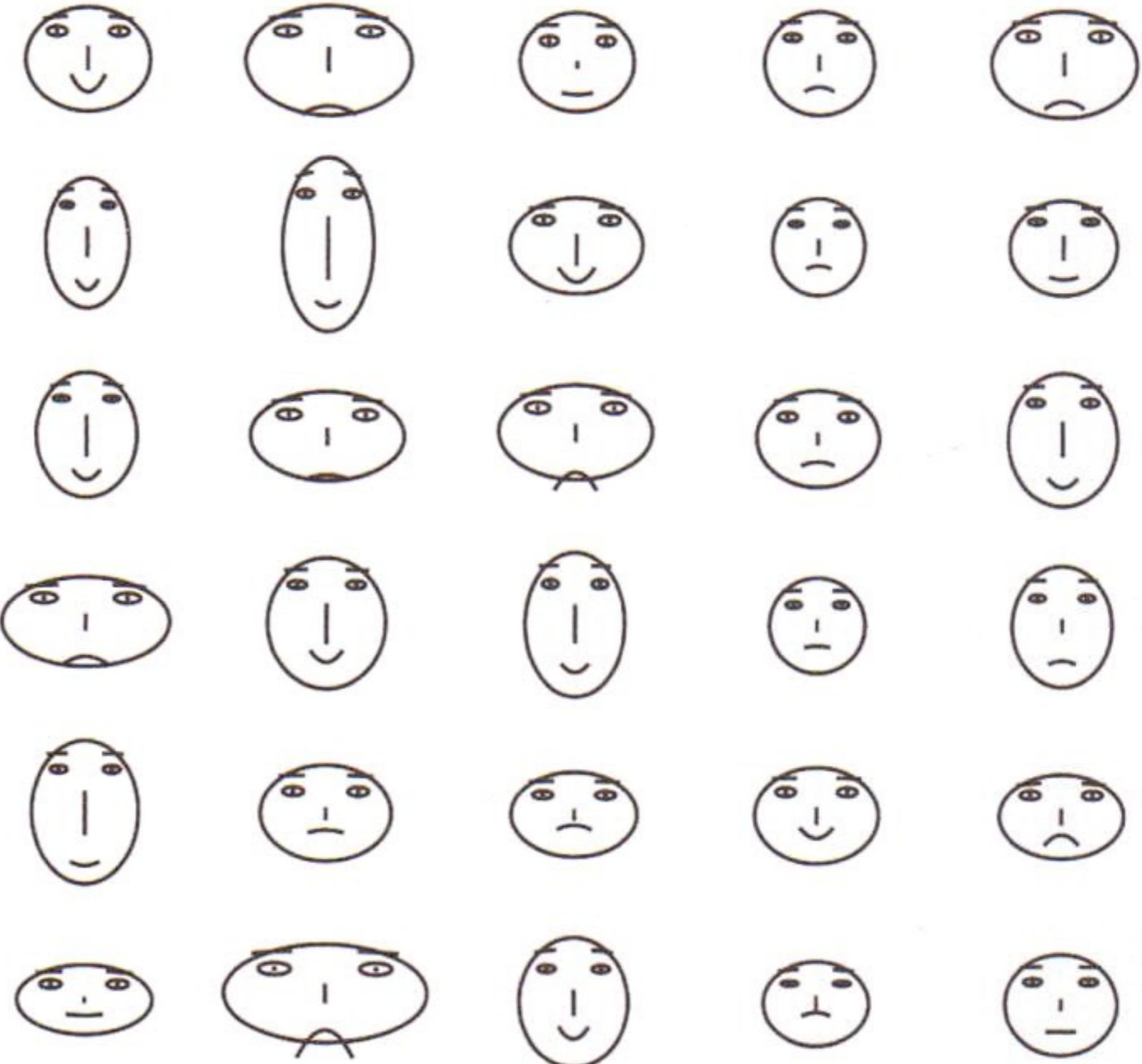


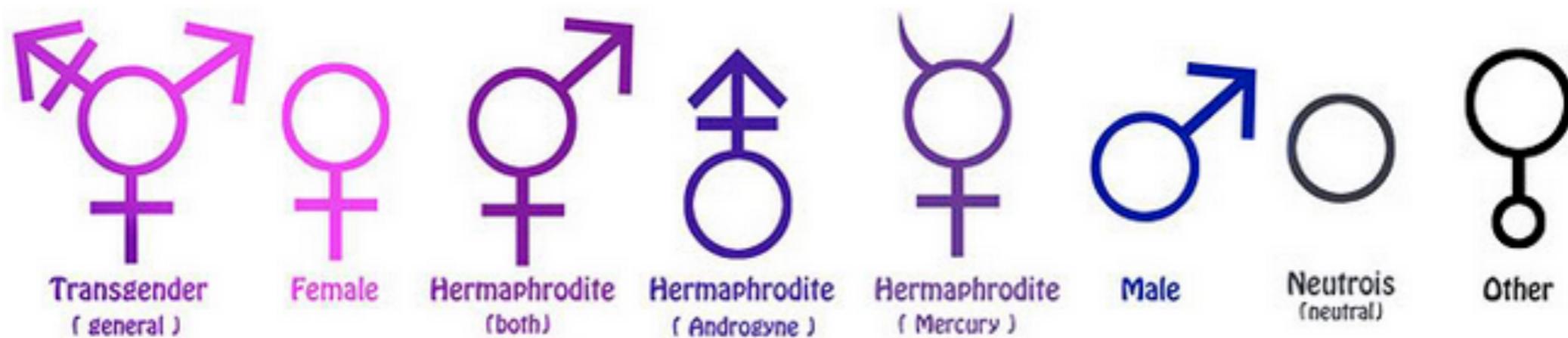


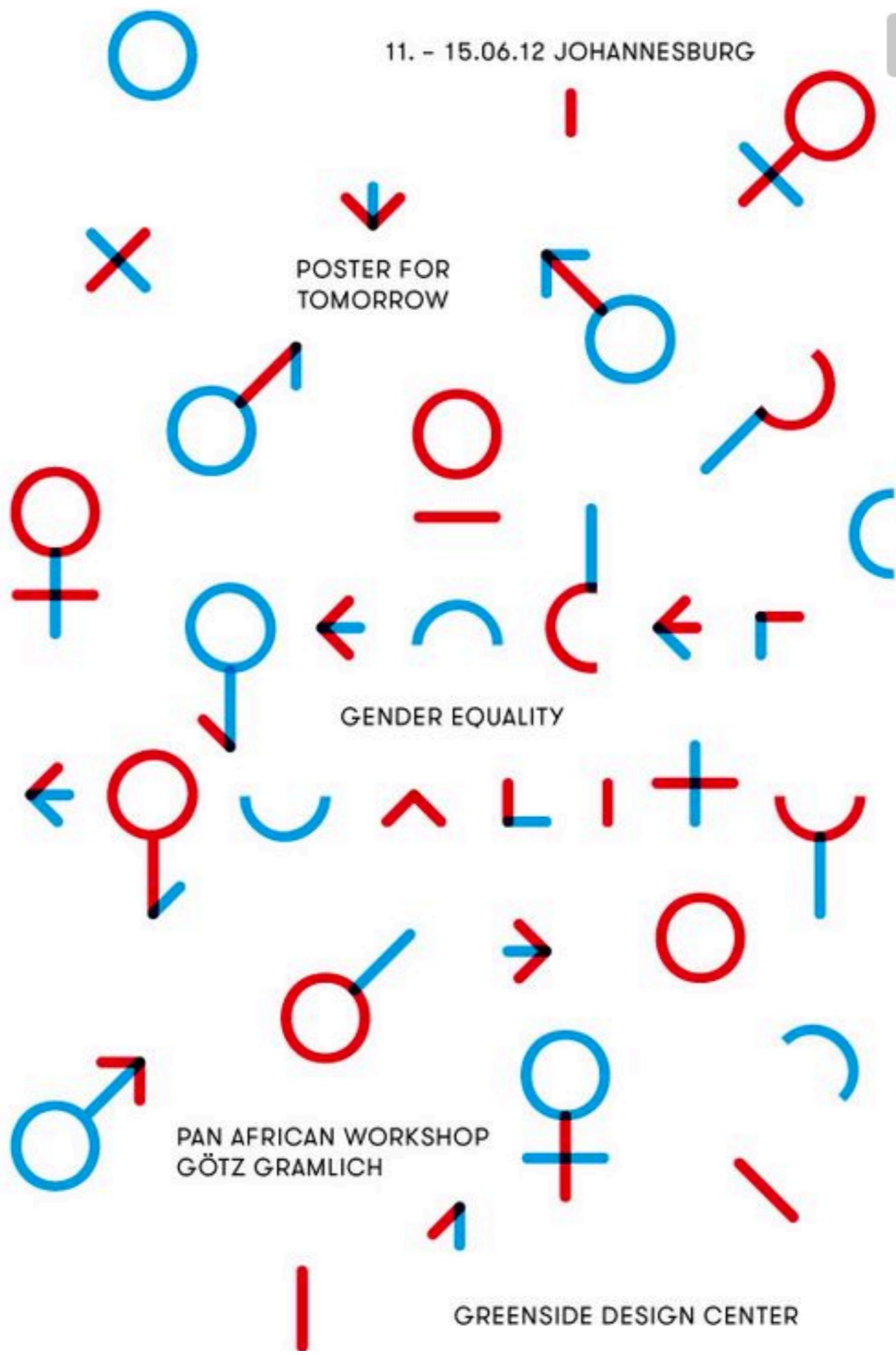
# FACES DE CHERNOFF

---

- Faces de Chernoff foram propostas na década de 70 para representação de dados multivariados
- Baseada no fato de que somos capazes de reconhecer similaridades entre faces



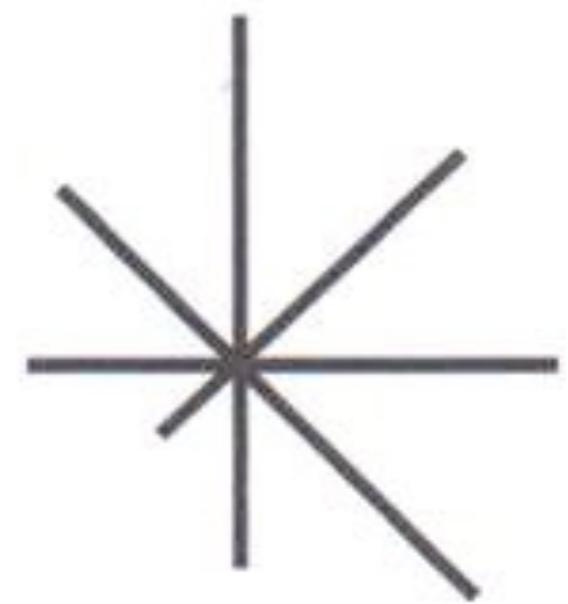
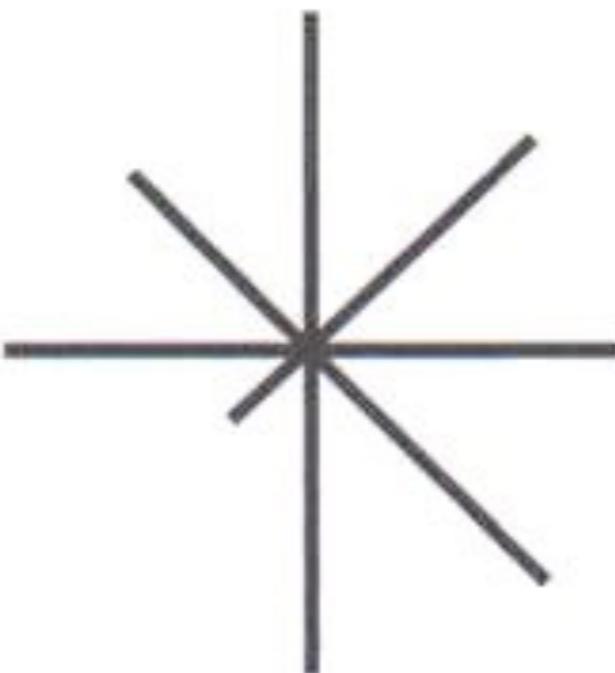
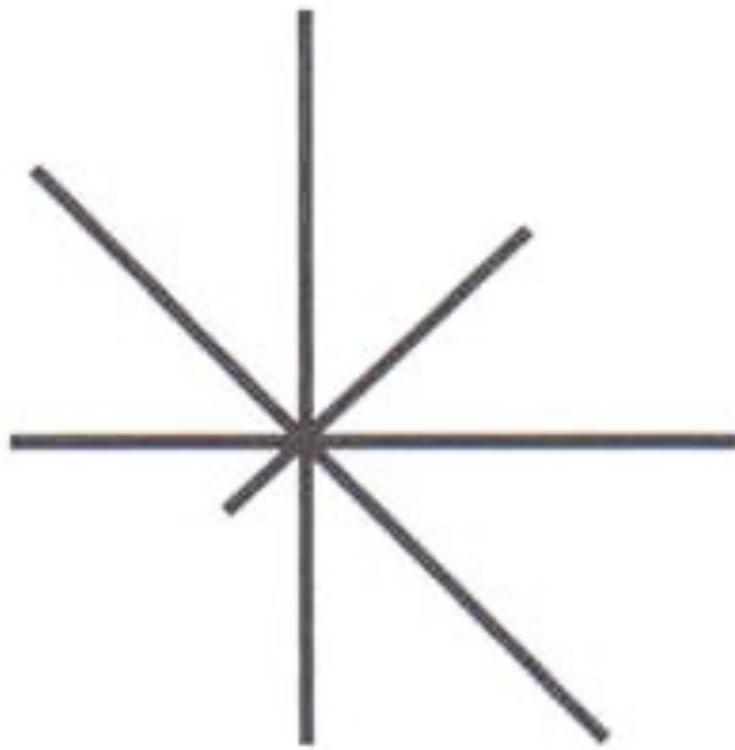




# WHISKERS

---

- Whiskers consistem em múltiplas linhas irradiando de um ponto central cada qual representando o valor de uma variável através do comprimento das linhas



# Create Your Better Life Index

How do you define a better life? Use our interactive tool to see how your country performs on the topics you feel make for a better life.

✓ Start with all topics rated equally

or set your own preferences [here](#).



## Create Your Better Life Index

Rate the topics according to their importance to you:

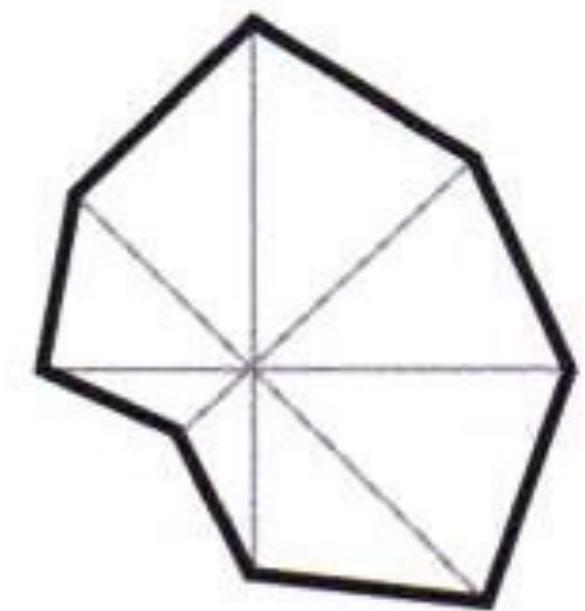
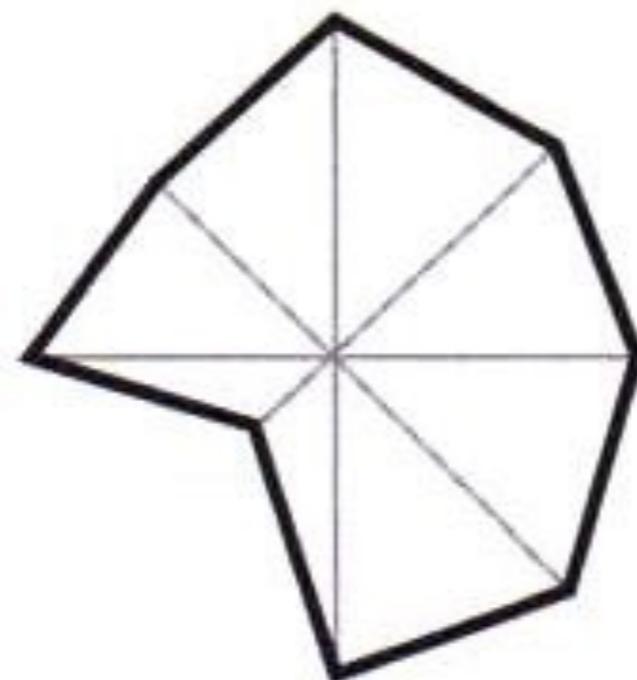
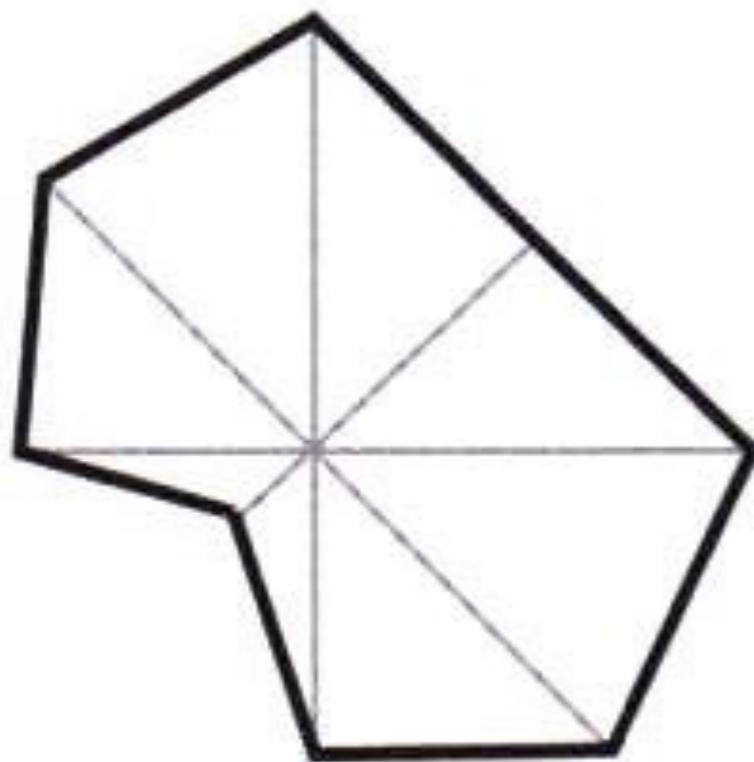
	Housing	<input type="range" value="0"/>
	Income	<input type="range" value="0"/>
	Jobs	<input type="range" value="0"/>
	Community	<input type="range" value="0"/>
	Education	<input type="range" value="0"/>
	Environment	<input type="range" value="0"/>
	Governance	<input type="range" value="0"/>
	Health	<input type="range" value="0"/>
	Life Satisfaction	<input type="range" value="0"/>
	Safety	<input type="range" value="0"/>
	Work-Life Balance	<input type="range" value="0"/>

[+ Share this index](#)

# GLIFOS ESTRELA

---

- Glifos estrela ou gráficos de radar são semelhantes aos whiskers

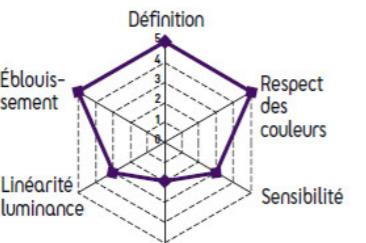


## PANASONIC TM-300

### Mémoire flash



Tests laboratoire



32 Go



L

Optique  
Stabilisé

Zoomx12

10,6 Mpix

3MOS

HDMI

USB

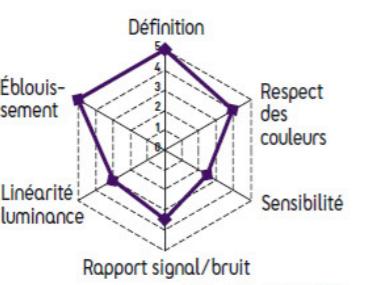
Autonomie\*: 119 mn,  
avec écran LCD: 110 mn

## CANON HF S10

### Mémoire flash



Tests laboratoire



32 Go



L

Optique  
Stabilisé

Zoomx10

8 Mpix

HDMI

USB

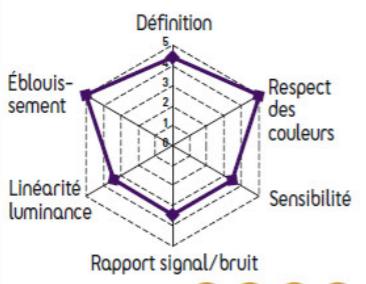
Autonomie\*: 119 mn,  
avec écran LCD: 110 mn

## PANASONIC HDC-HS300

### Disque Dur



Tests laboratoire



120 Go



L

Optique  
Stabilisé

Zoomx12

10,6 Mpix

3MOS

HDMI

USB

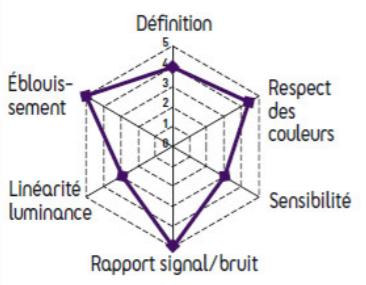
Autonomie\*: 118 mn,  
avec écran LCD: 103 mn

## SONY HDR-XR 520

### Disque dur



Tests laboratoire



240 Go



L

Optique  
Stabilisé

Zoomx12

12 Mpix

HDMI

USB

Autonomie\*: 113 mn,  
avec écran LCD: 103 mn

**Le jugement de la Fnac:** autonomie et mémoire d'éléphant pour le TM300 et ses 32 Go de mémoire flash intégrée. Avec son viseur électronique et sa bague multifonction autour de l'objectif, ce caméscope joue la carte de l'ergonomie. Son originalité se cache du côté des capteurs, puisqu'il est doté d'un capteur triMOS, permettant un meilleur rendu des couleurs. Côté optique, le zoom 12x stabilisé est signé Leica Dicomar. Côté son, on appréciera l'enregistrement en 5.1, avec deux micros, l'un suivant les mouvements de l'optique tandis que l'autre capte les sons d'ambiance. Côté labo : effet de halo parfait.

**Le jugement de la Fnac:** support avec 12 heures d'enregistrement en Full HD sur mémoire interne de 32 Go sont au programme de la HFS 10. Reprenant la même plateforme technique que le HFS 100 dont l'enregistrement en 24 Mbps, le rythme le plus élevé du marché, pour une captation des mouvements plus fluide et une reproduction des détails et des couleurs précise. Le HFS 10 est doté d'un autofocus rapide et d'une optique stabilisée (zoom 10x). Côté connectiques, mini griffe porte-accessoires, entrée micro et prises casque sont au rendez-vous ! Un vrai concentré de technologie.

**Le jugement de la Fnac:** frère jumeau du Panasonic HDC-TM 300 mais cette fois-ci sur disque dur de 120 Go. Il met l'accent sur une nouvelle gestion des couleurs grâce à trois capteurs triMOS Full HD afin de doubler la quantité de lumière reçue, pour une meilleure qualité d'image. Par ailleurs, afin d'assurer un meilleur rendu des films en Full HD (1080i), le caméscope est doté de la stabilisation optique OIS avancée (4000x/s). Un programme que les amateurs de Panasonic connaissent bien : la fonction IA détermine elle-même le mode scène approprié.

**Le jugement de la Fnac:** ce modèle HD équipé d'un disque dur de 240 Go est le produit « Star » de la gamme HD Sony. Doté de la fonction hybride pour filmer au format AVCHD sur disque dur ou sur carte, il enregistre jusqu'à 101 heures de film en 1920 x 1080 (full HD), mais aussi en 1440 x 1080 au choix. Reprenant les caractéristiques de la XR 200, il dispose en plus d'une molette de réglages manuels (Mise au point, exposition...). Ses plus : Capteur Exmor R, optique Sony G, un écran LCD de 8,1 cm, viseur en prime ! Existe en version 120 Go sous la référence XR 500.



chevrolet chevelle malibu



buick skylark 320



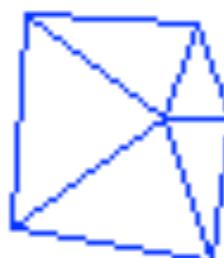
plymouth satellite



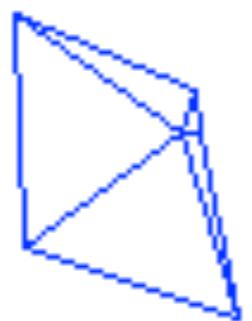
amc rebel sst



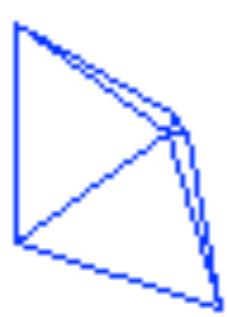
ford torino



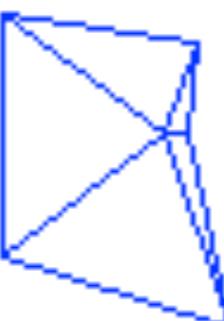
ford galaxie 500



chevrolet impala



plymouth fury iii

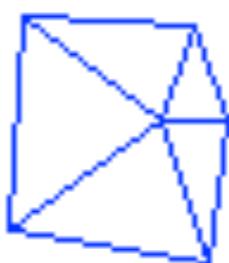


pontiac catalina



chevrolet chevelle malibu buick skylark 320

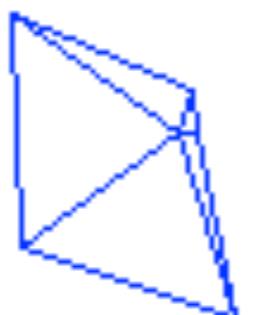
plymouth satellite



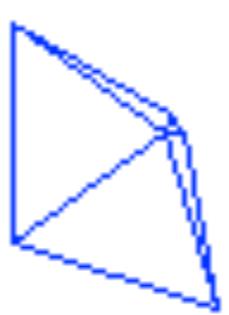
amc rebel sst

ford torino

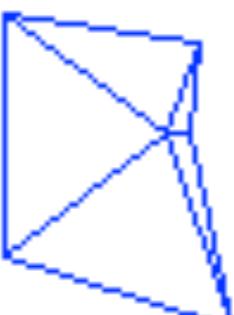
ford galaxie 500



chevrolet impala

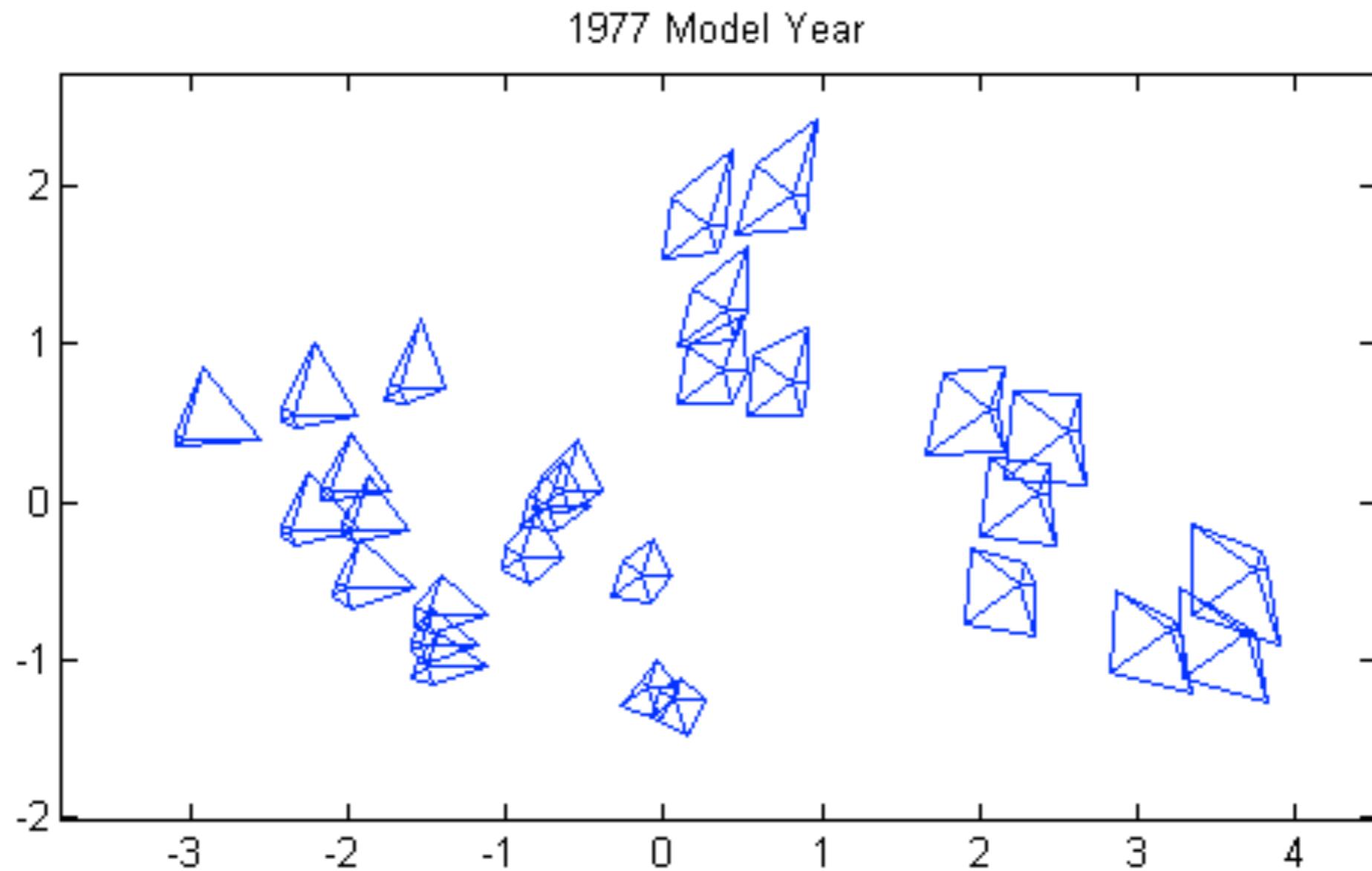


plymouth fury iii



pontiac catalina





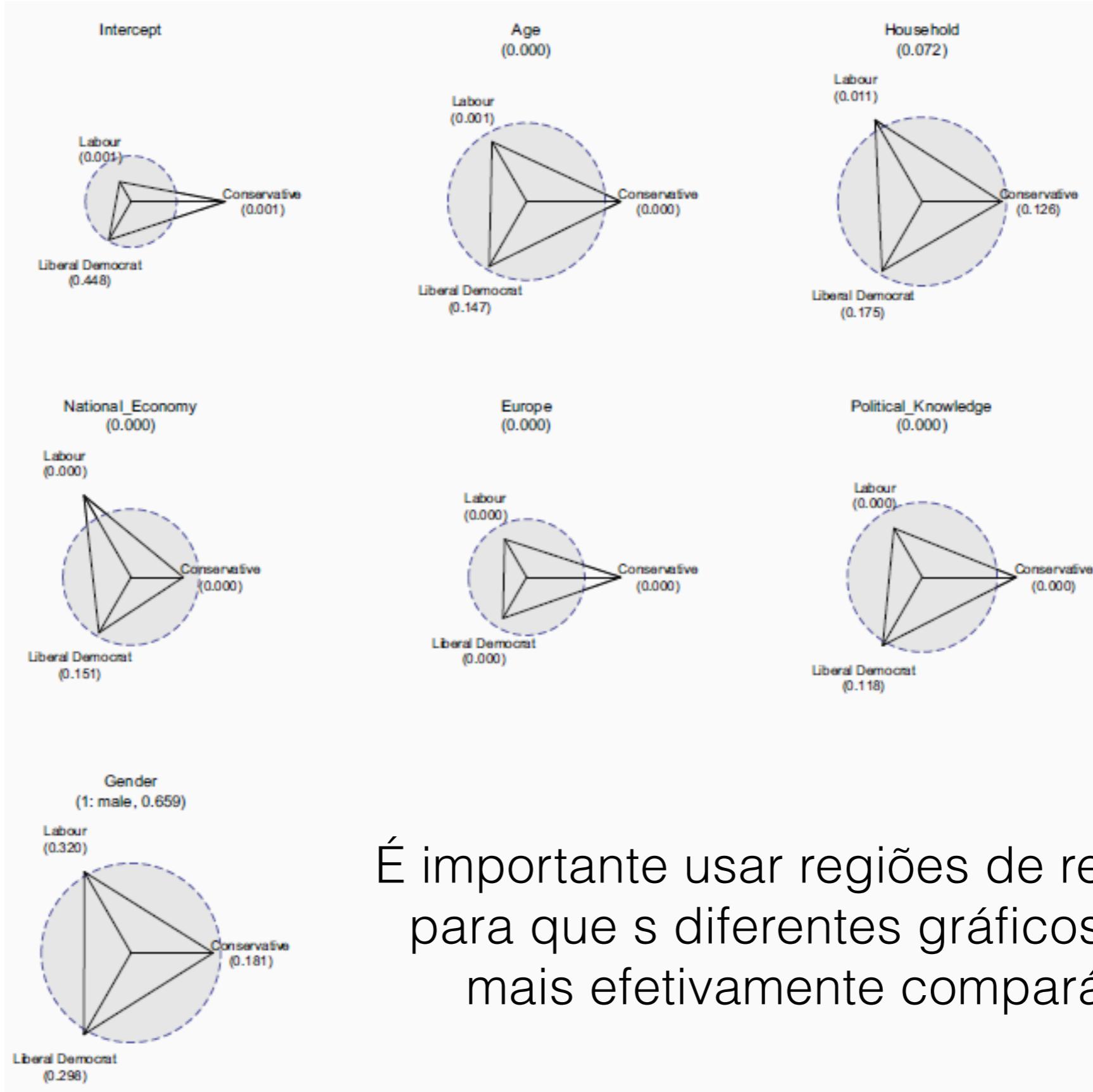
Técnicas de redução de dimensionalidade como:

**MDS** (Dimensionamento multidimensional)

**SVD** (Decomposição por valor singular) ou

**PCA** (Análise de componentes principais)

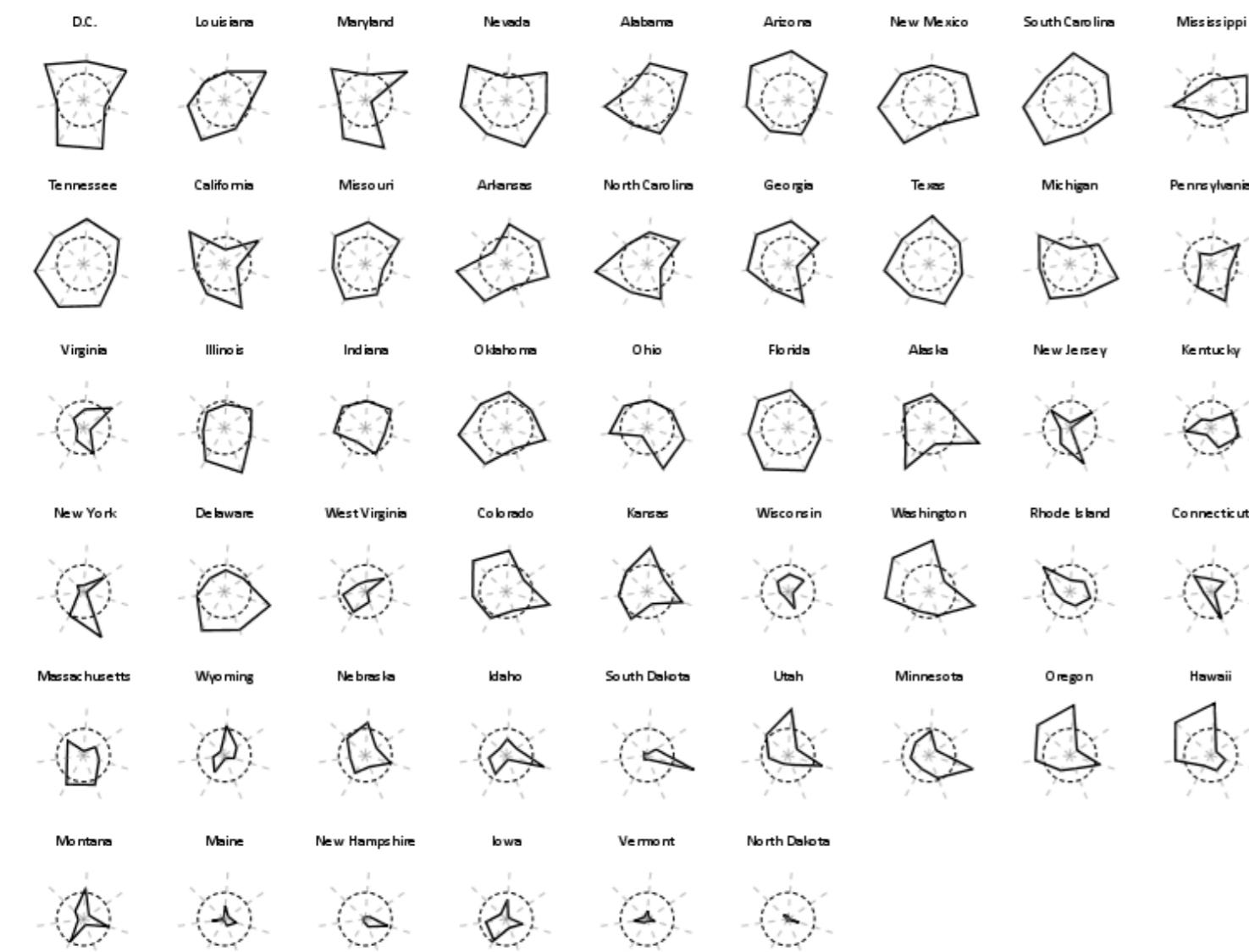
são úteis para desenhar os objetos multivariados em um espaço 2D



É importante usar regiões de referência para que os diferentes gráficos sejam mais efetivamente comparáveis

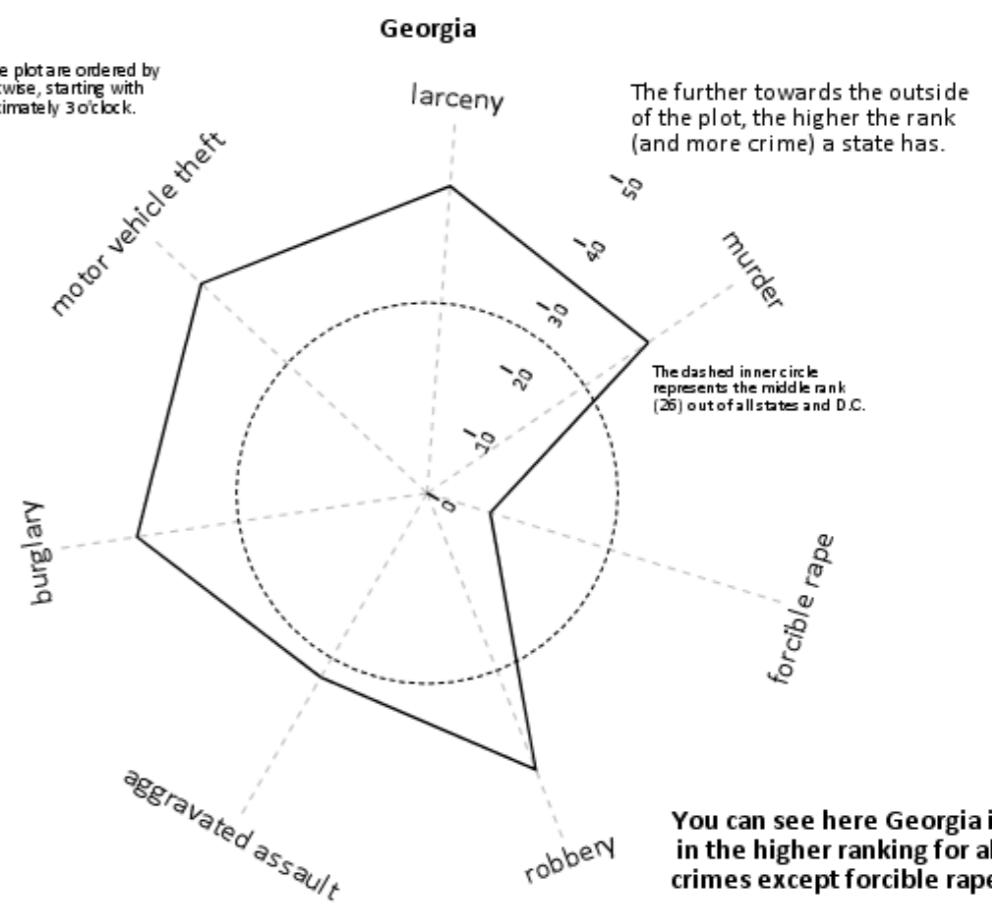
## Rankings of Crime Rates for 50 States (and D.C.) Star Plot

States ordered by homicide ranking in plot (left to right)



### Annotated Legend for Star Plots

Crimes around the plot are ordered by seriousness clockwise, starting with murder at approximately 3 o'clock.



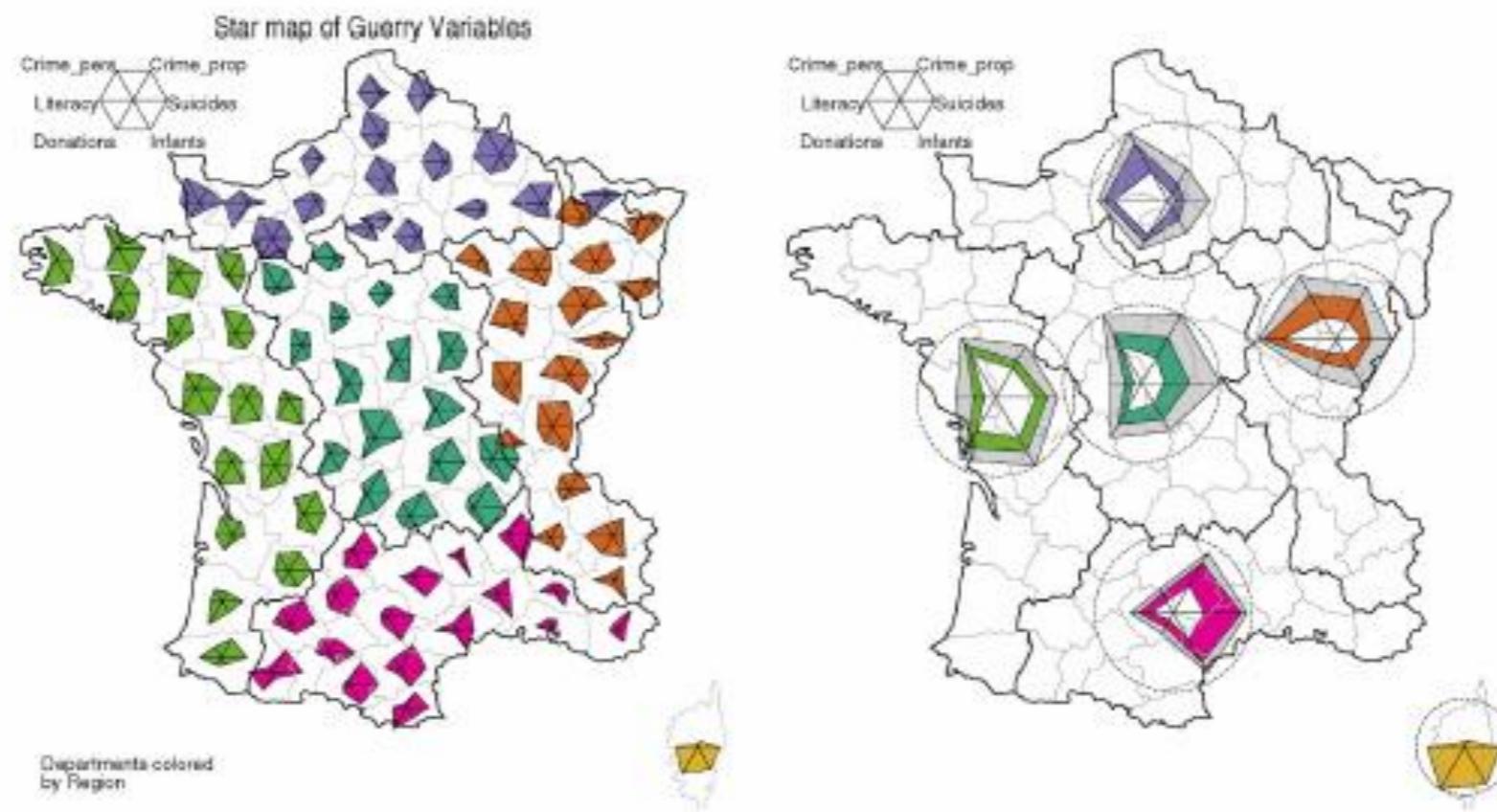
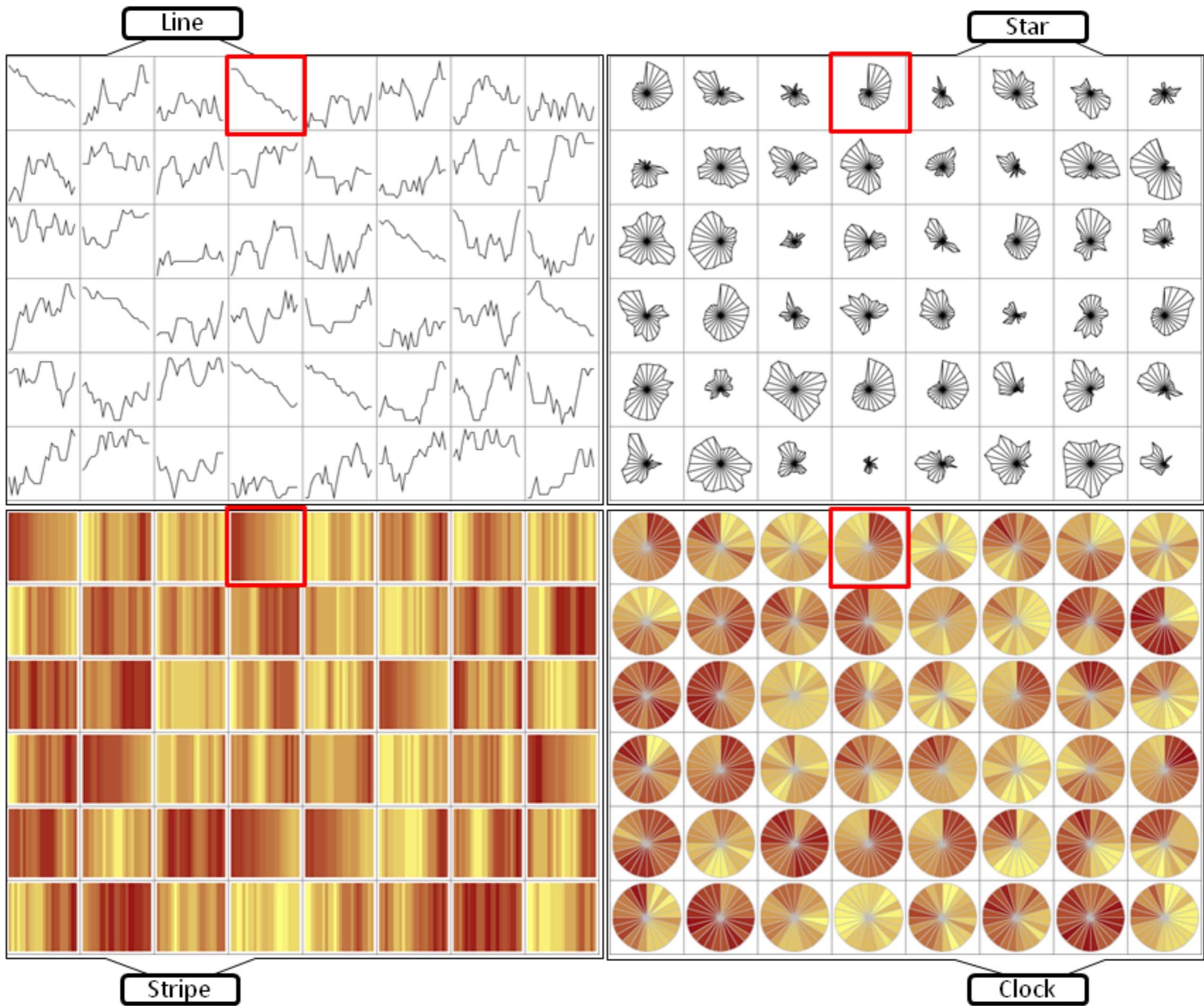
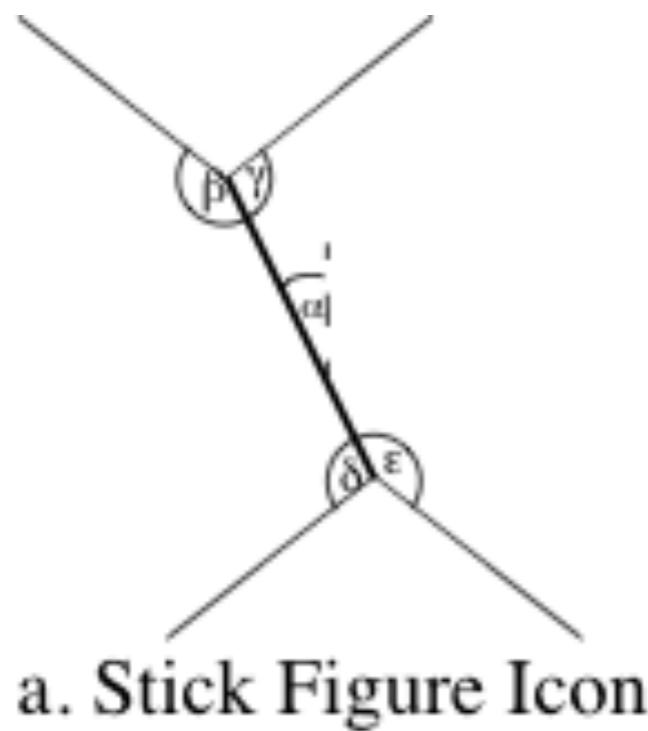


FIG. 17. Star maps of Guerry's data, using rays proportional to the rank of each variable (longer = better). Variables have been ordered according to their angular positions in the biplot (Figure 14). Left: Glyphs for individual départements. Right: Multivariate boxplot glyphs for the medians and quartiles across départements in each region of France.

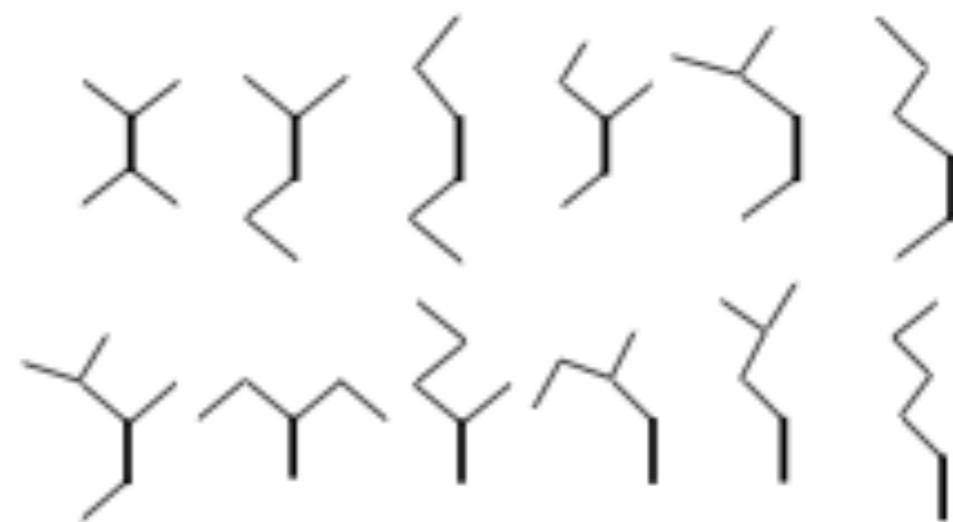


# REPRESENTAÇÕES BASEADAS EM VARETAS

---

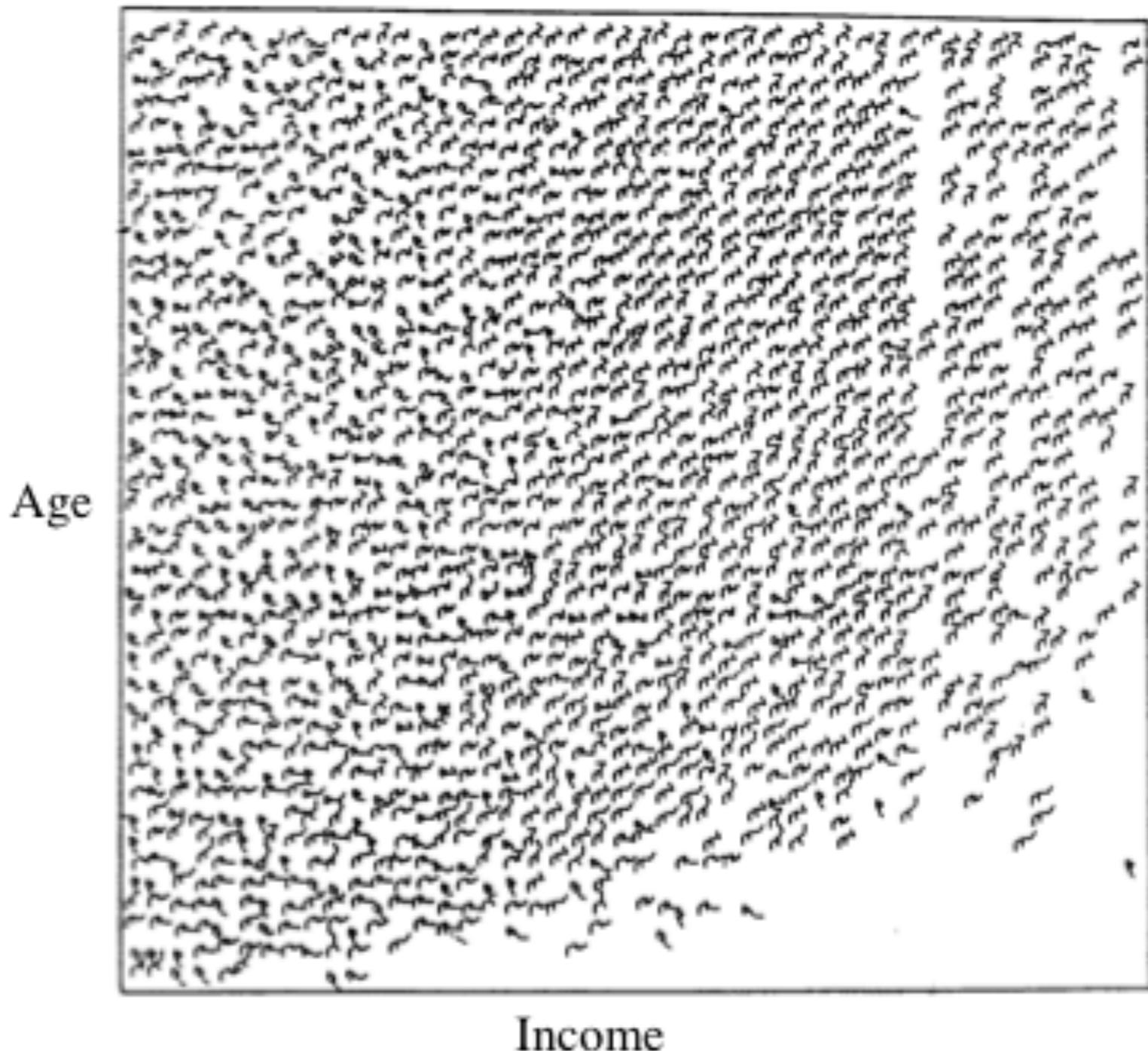


a. Stick Figure Icon



b. A Family of Stick Figures

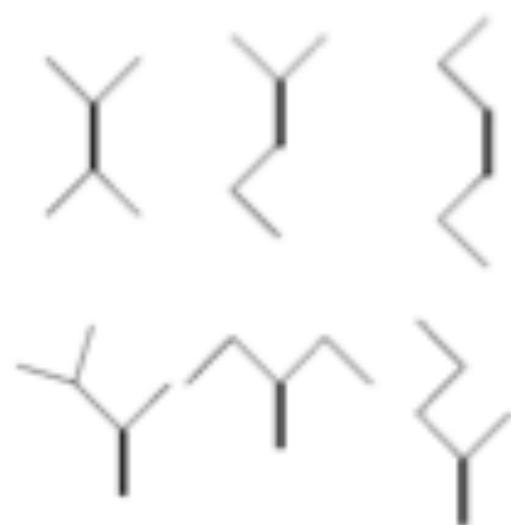
**Figure 5: Stick Figure Visualization Technique**



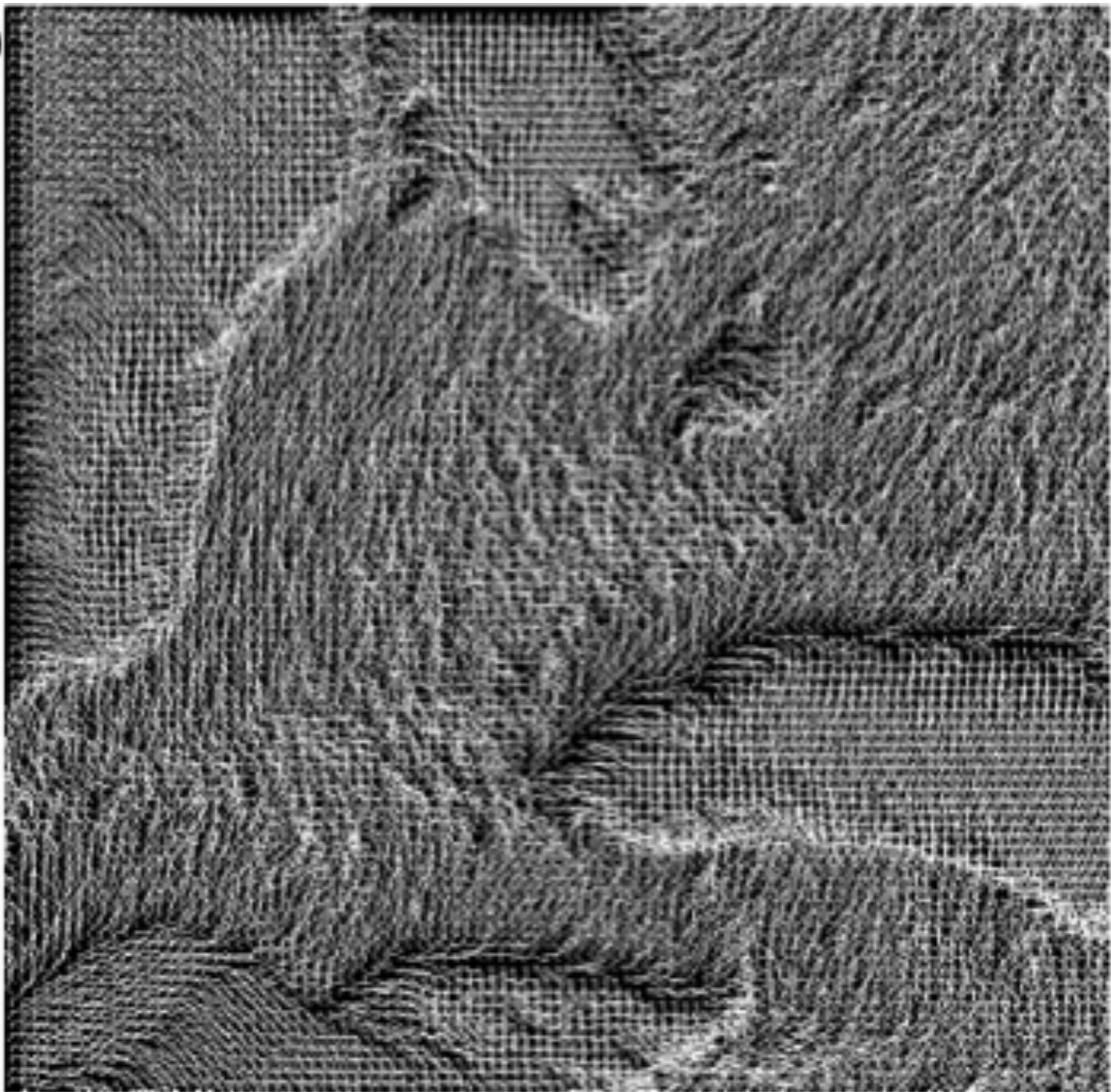
**Figure 6: Stick Figure Visualization of Census Data**

(used by permission of G. Grinstein, Institute of Visualization and Perception Research,  
University of Massachusetts at Lowell; cf. [GPW 89])

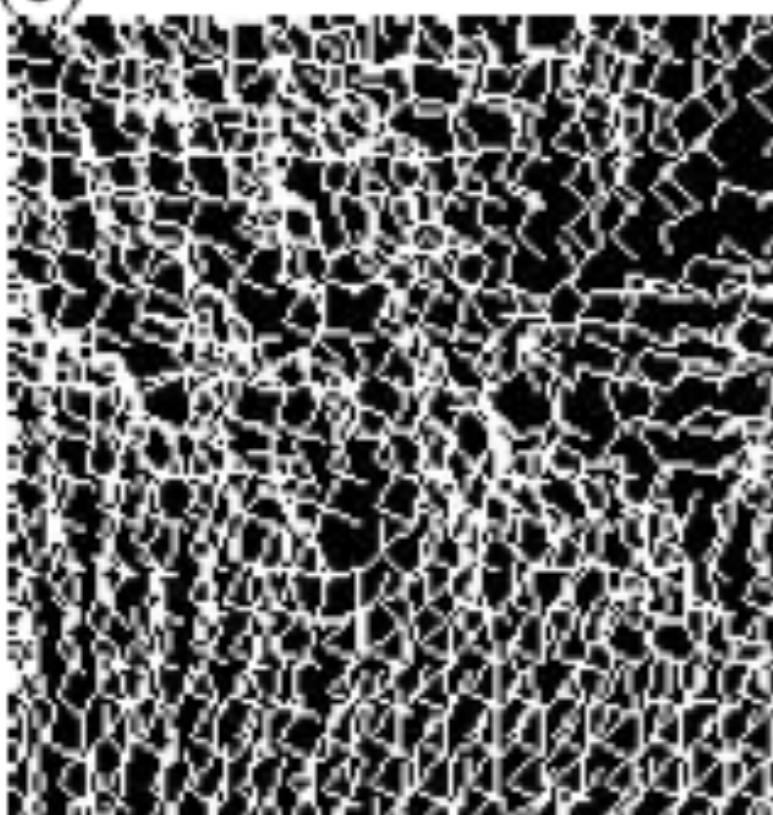
(a)



(b)



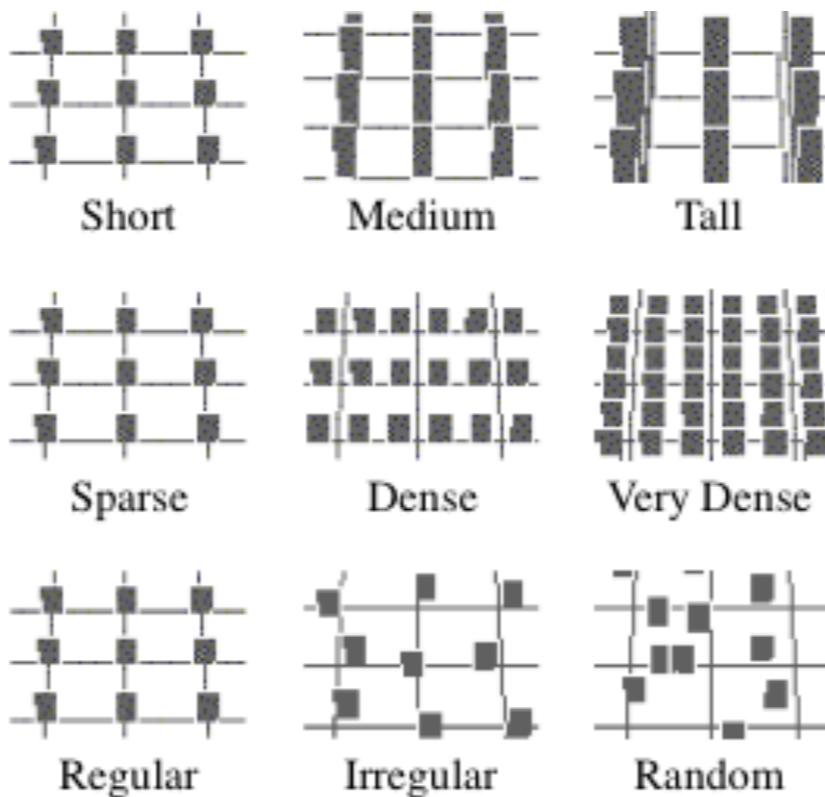
(c)



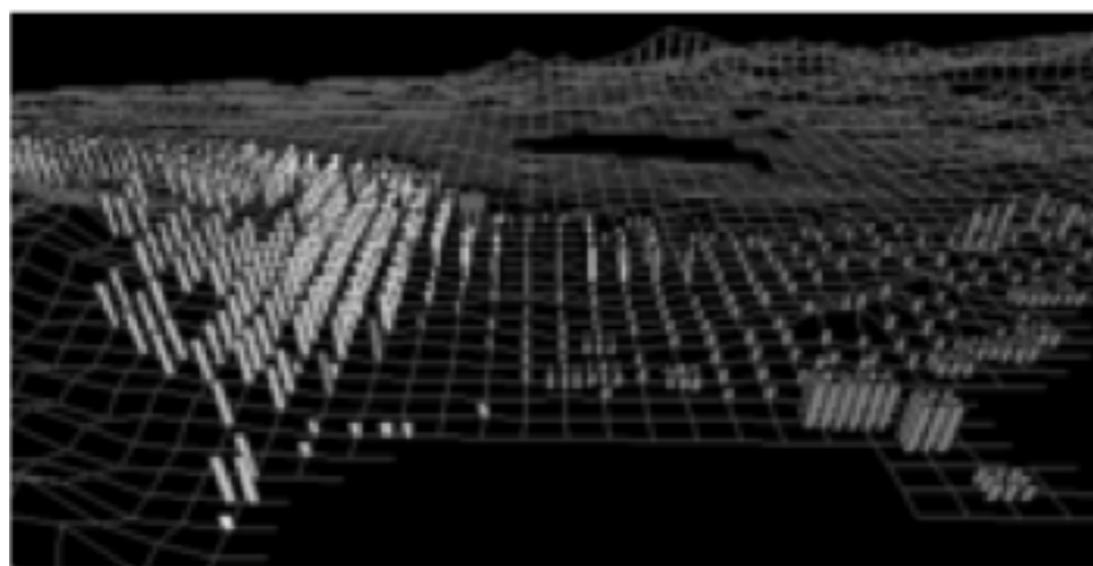
**Figure 3.23:** (a) Stick figure family [23], (b) 5D image data using stick figures, (c) Part of (b) in original size [60].

# REPRESENTAÇÕES BASEADAS EM CARTÕES

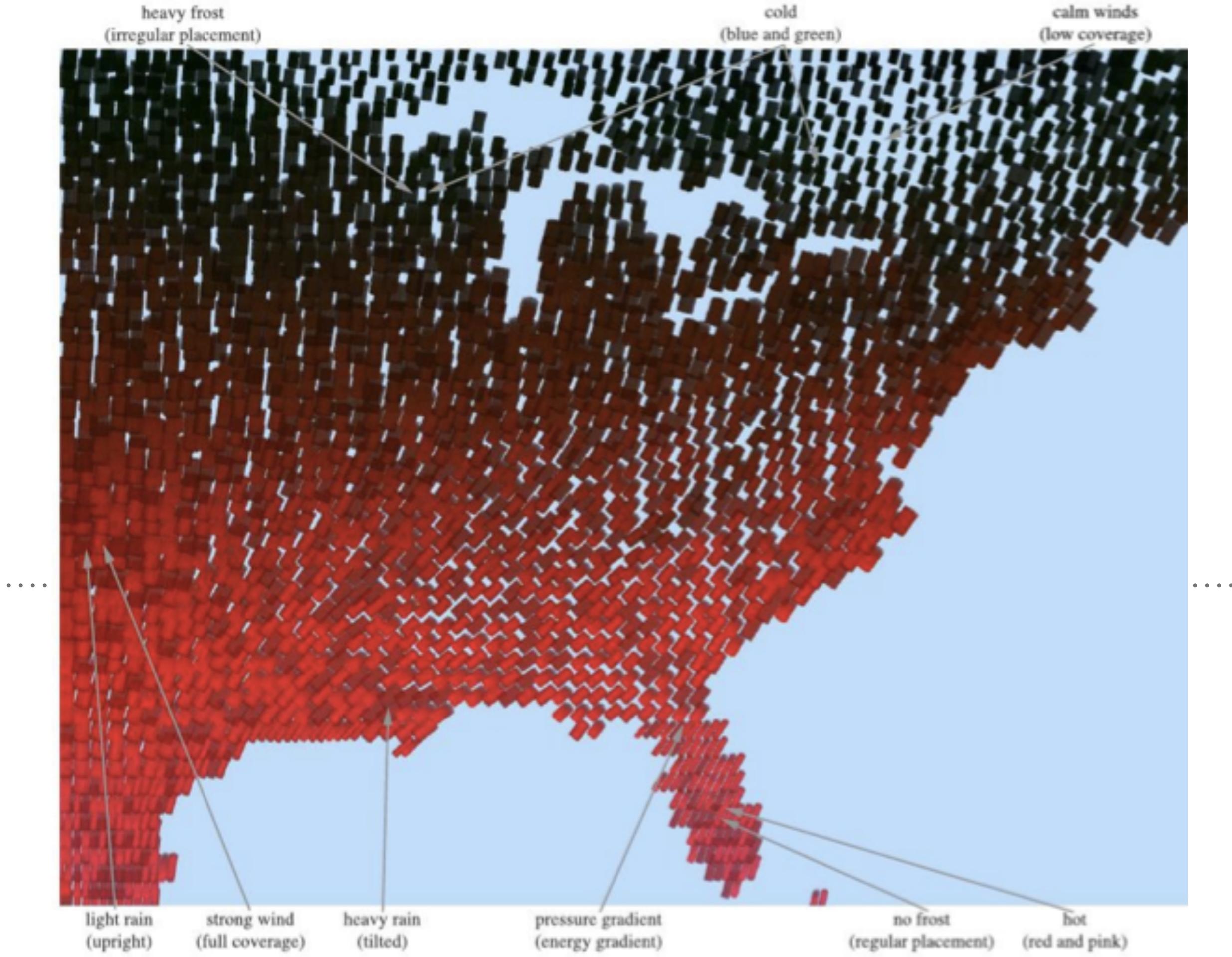
---



(a)



(b)



**Figure 3.28:** Nonphotorealistic visualization of weather conditions [68].

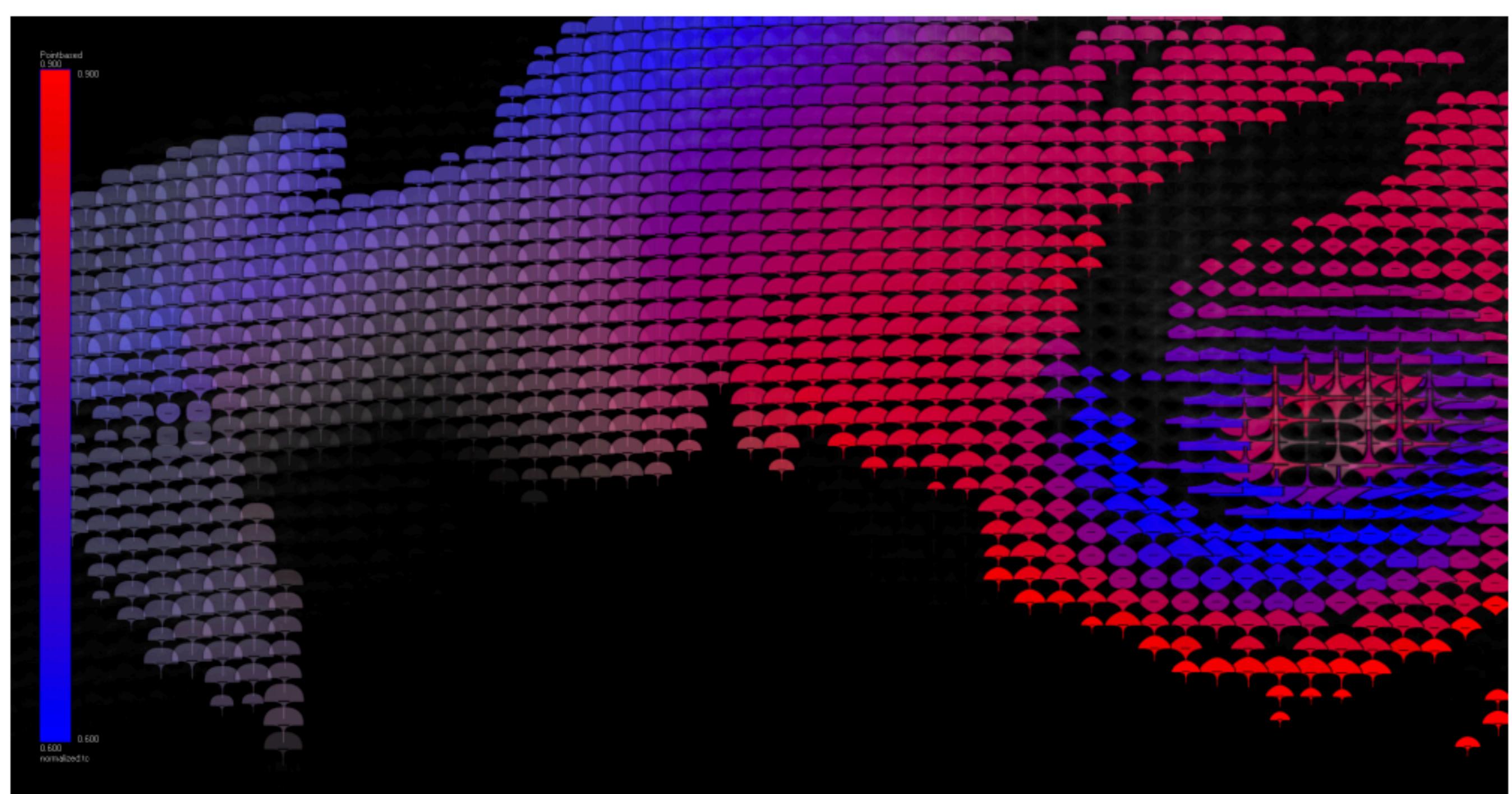


Figure 10: The hurricane Isabel dataset, timestep 12. Color represents temperature, and the amount of clouds is mapped to glyph size. The upper shape represents pressure, and the lower shape precipitation. The visualization depicts the fast moving clouds (specified via brushing), and the eye of the hurricane is visible in the lower right, surrounded by high amounts of precipitation and cold airflow.

Lie, Andreas E., Johannes Kehrer, and Helwig Hauser. "Critical design and realization aspects of glyph-based 3D data visualization." Proceedings of the 25th Spring Conference on Computer Graphics. ACM, 2009.

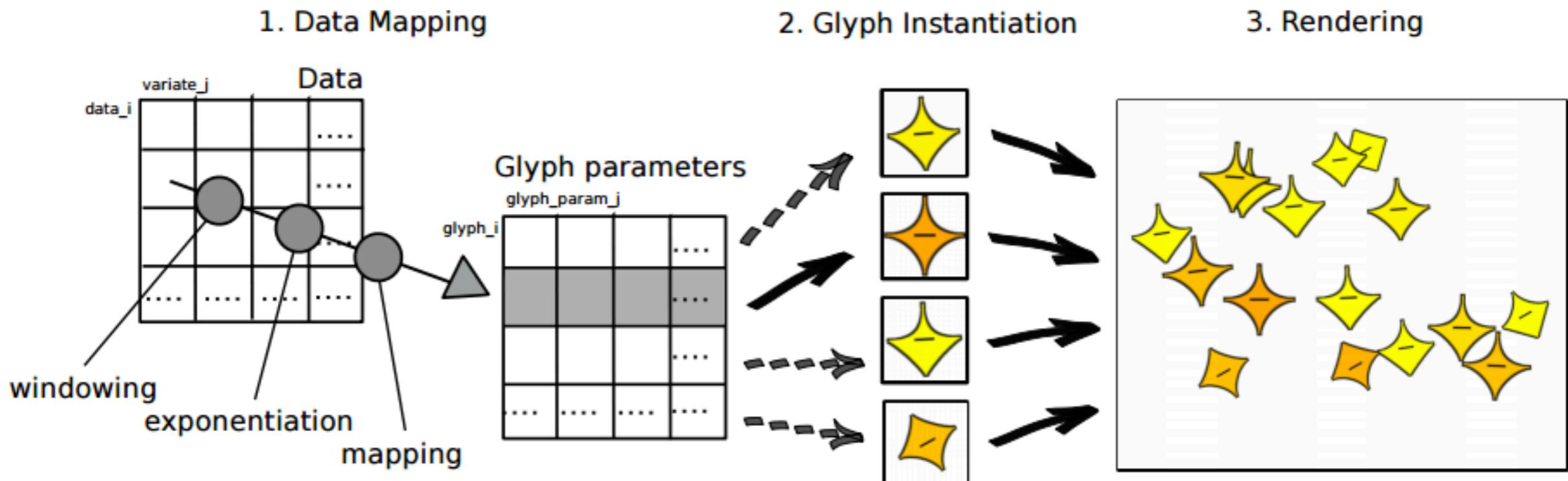


Figure 2: Data variates undergo data mapping stages; windowing, exponentiation and mapping. These values are then used to instantiate the corresponding glyphs, (e.g, determining shape, size) and finally the glyphs are rendered into the context.

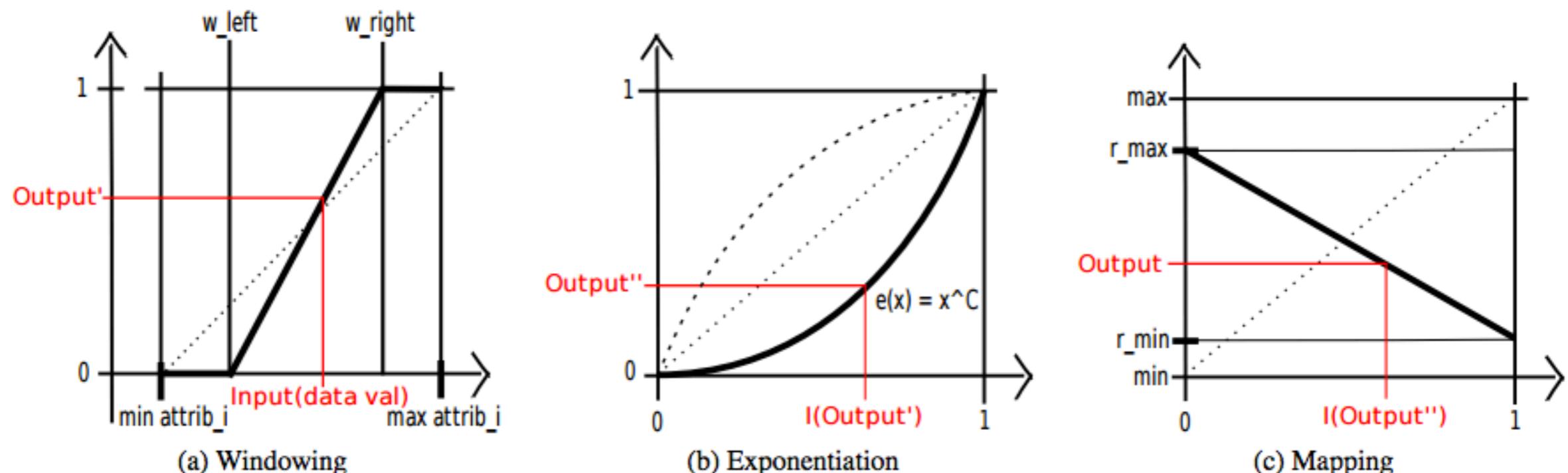
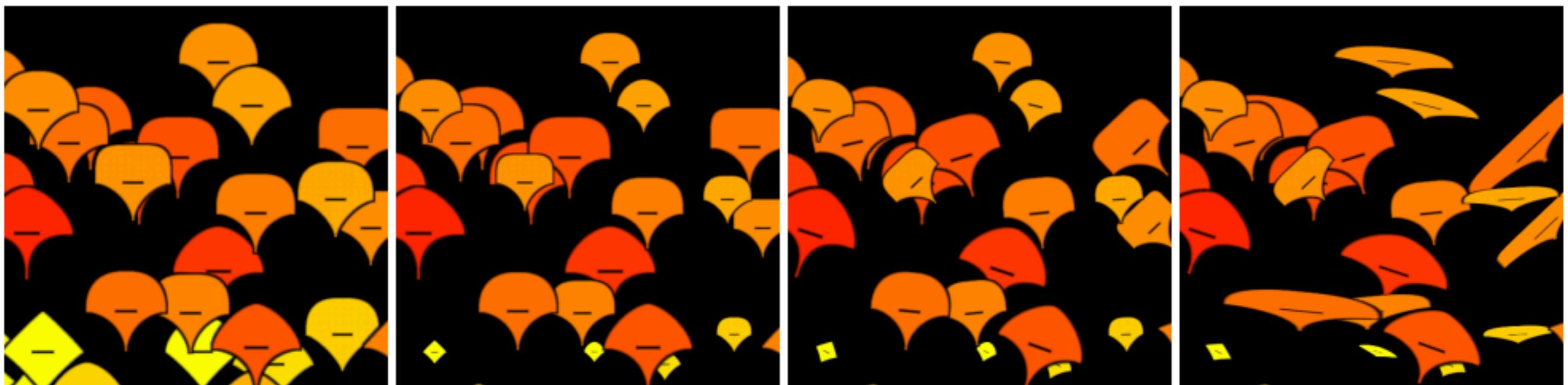


Figure 3: (a) User selects  $w_{left}$  and  $w_{right}$  which adjusts the data mapping ramp for input datavalues. The dotted line represents the default value for the ramp. (b) User can control the data mapping curve by adjusting the C term in  $x^C$ . The middle dotted line represents the default curve, the curve can be bent upwards or downwards. (c) The output range of the datamapping can be adjusted to fit the datavariates. User selects  $r_{min}$  and  $r_{max}$  to clamp the output range.



(a) Two data attributes are represented as the upper / lower glyph shape

(b) Added data attribute to overall glyph size

(c) Glyph rotation has been assigned a data attribute as well

(d) A data attribute has been assigned to glyph aspect ratio

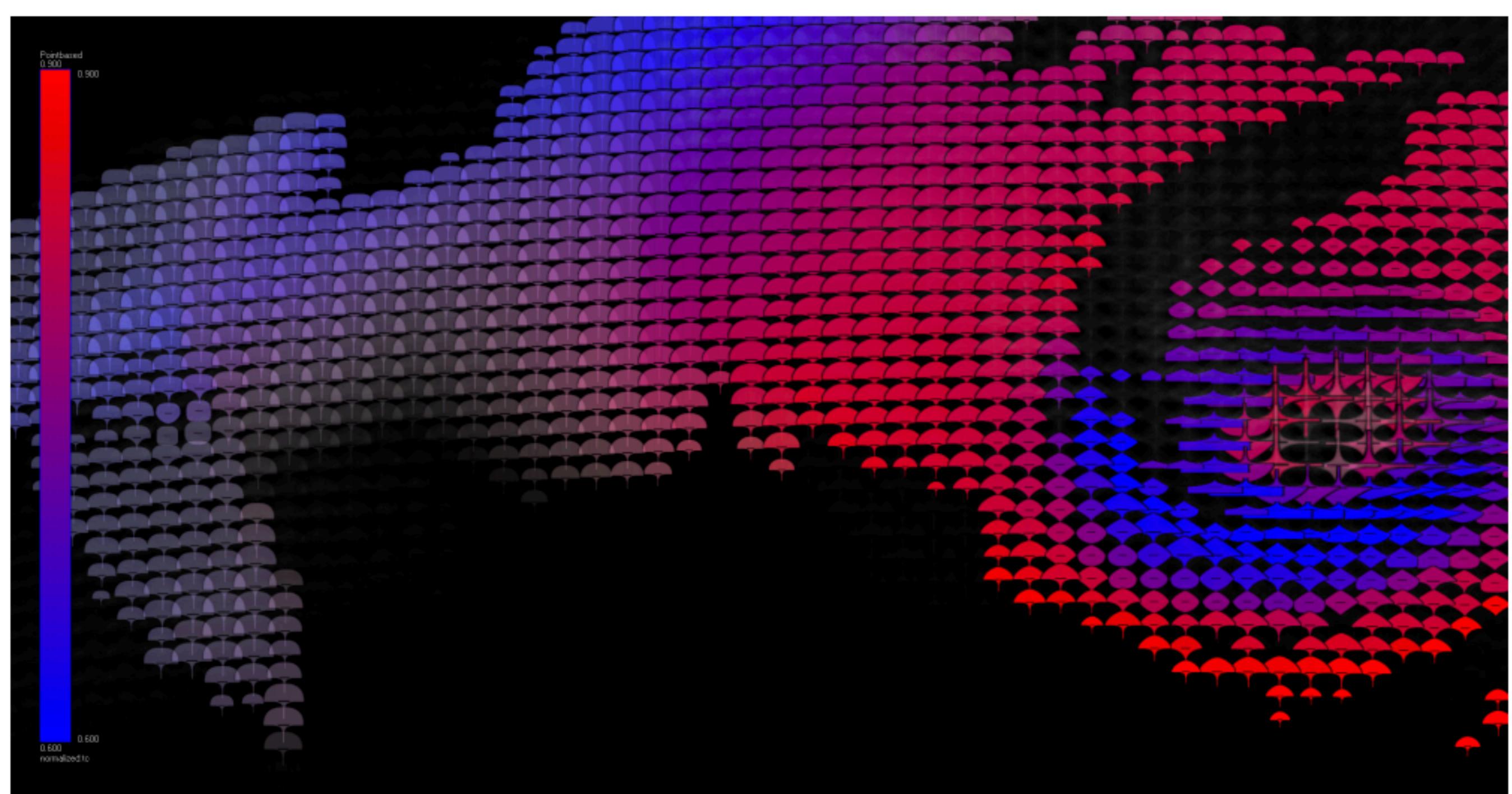


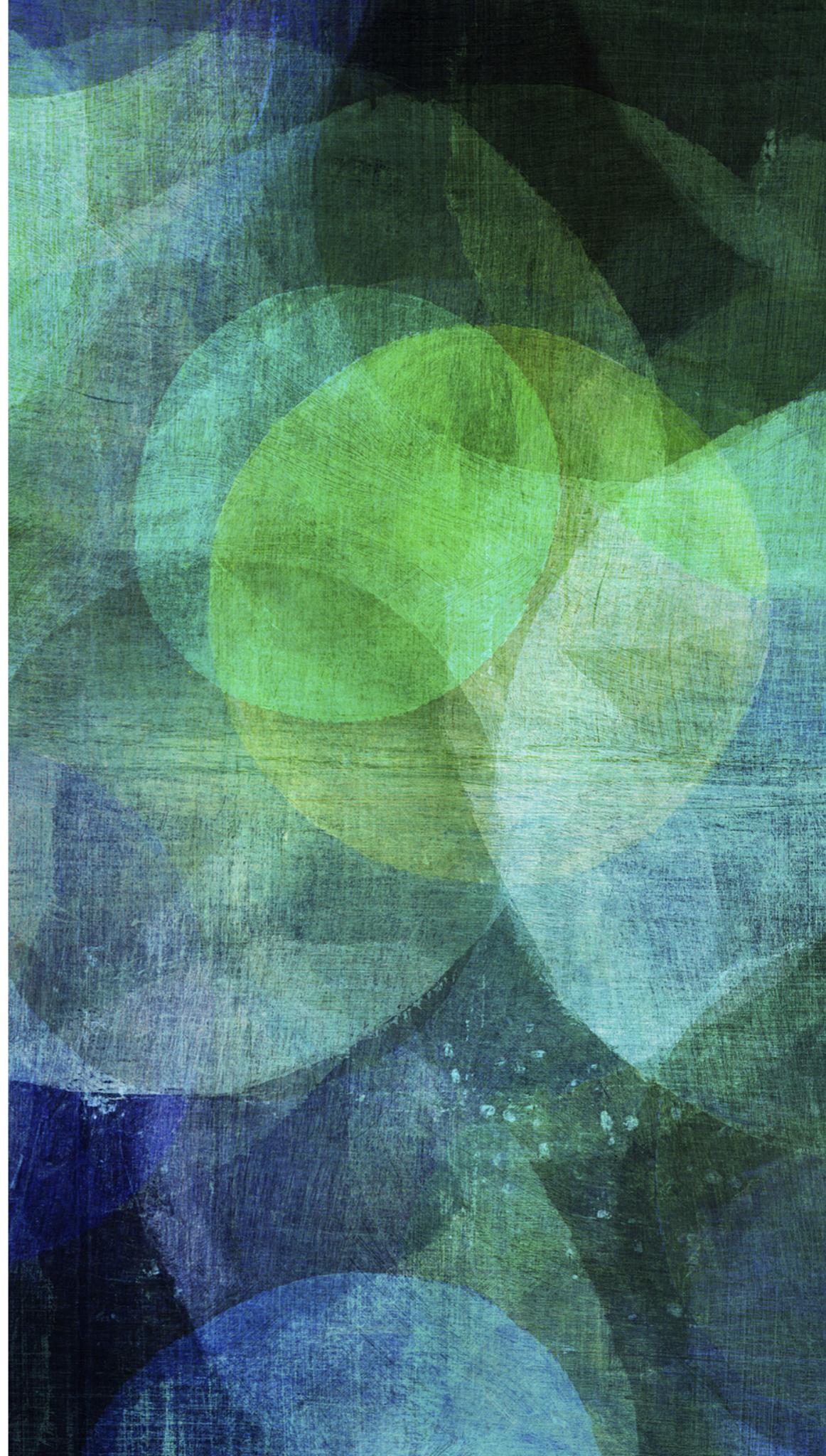
Figure 10: The hurricane Isabel dataset, timestep 12. Color represents temperature, and the amount of clouds is mapped to glyph size. The upper shape represents pressure, and the lower shape precipitation. The visualization depicts the fast moving clouds (specified via brushing), and the eye of the hurricane is visible in the lower right, surrounded by high amounts of precipitation and cold airflow.

Color	Temperature
Glyph Upper	Pressure
Glyph Lower	Precipitation
Glyph Size	Clouds

Table 2: Glyph property mapping for Hurricane Isabel dataset

# TÉCNICAS ORIENTADAS A PIXELS

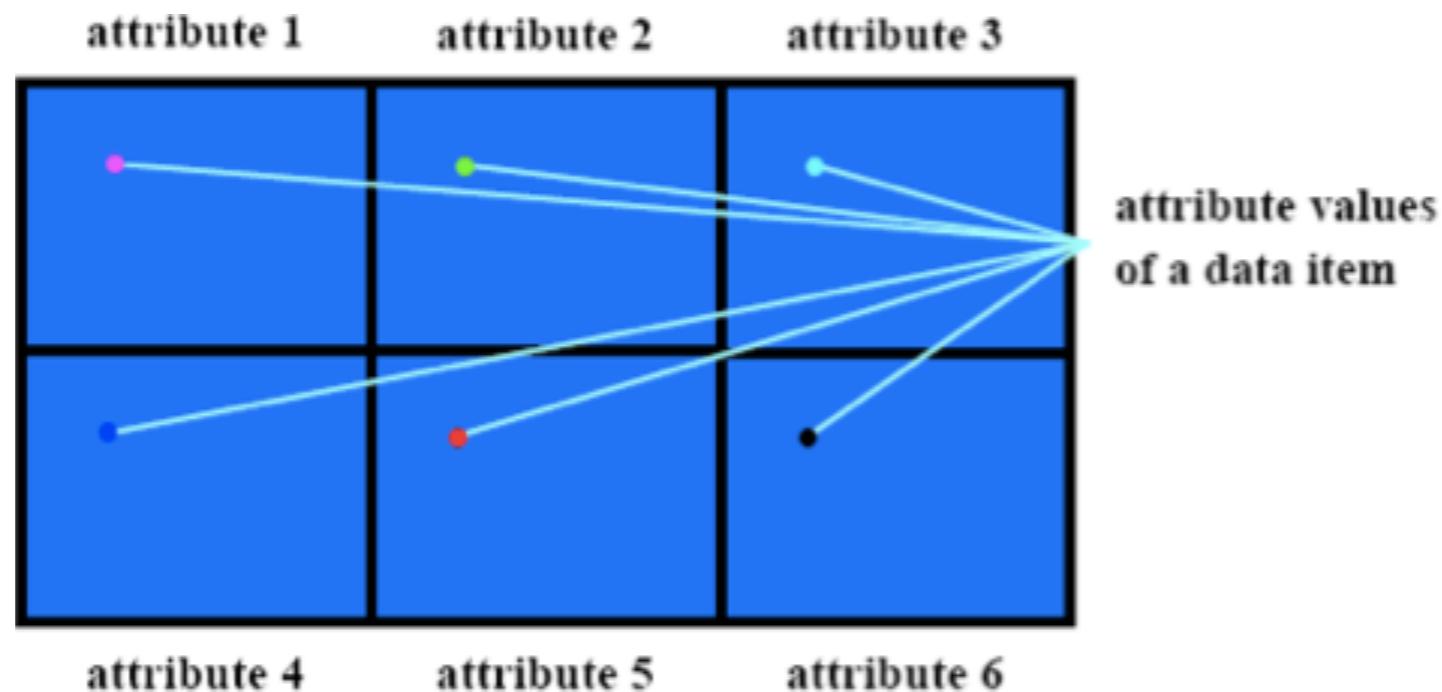
---

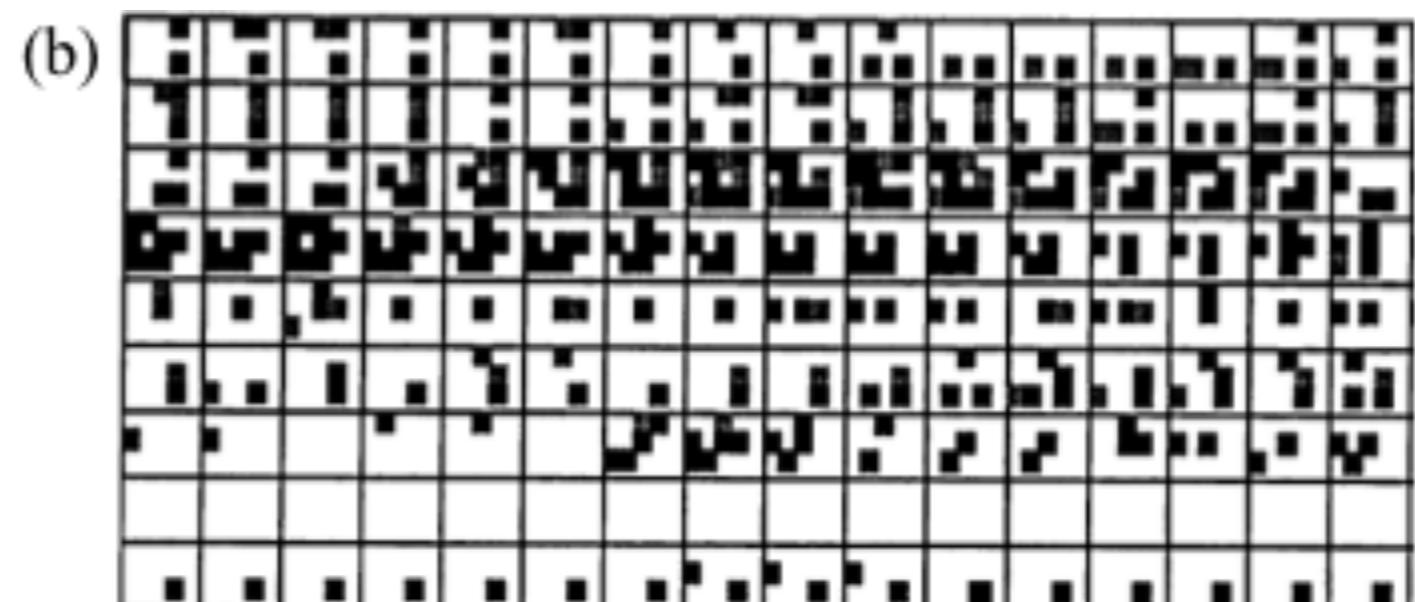
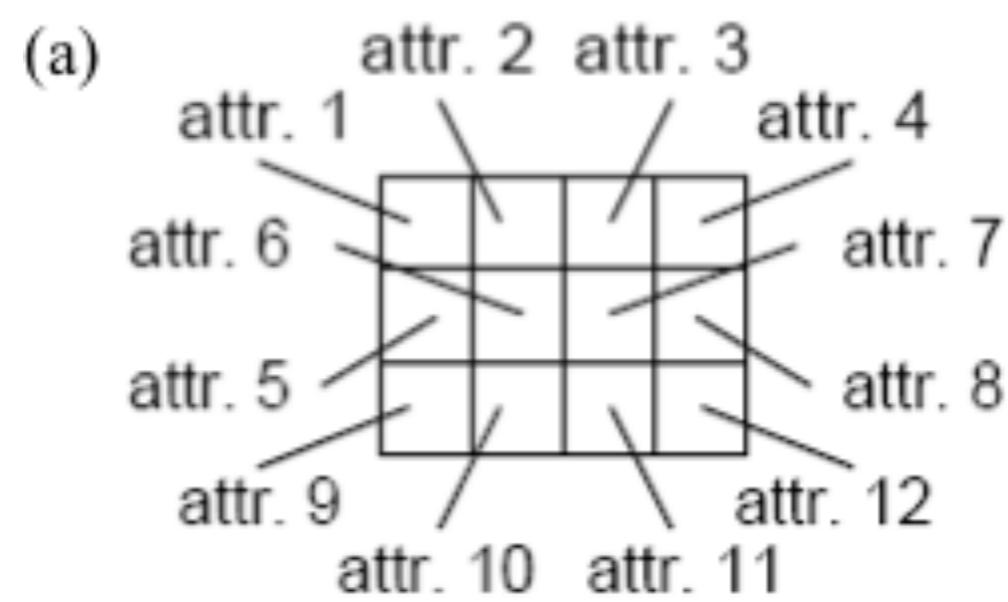


# TÉCNICAS ORIENTADAS A PIXELS

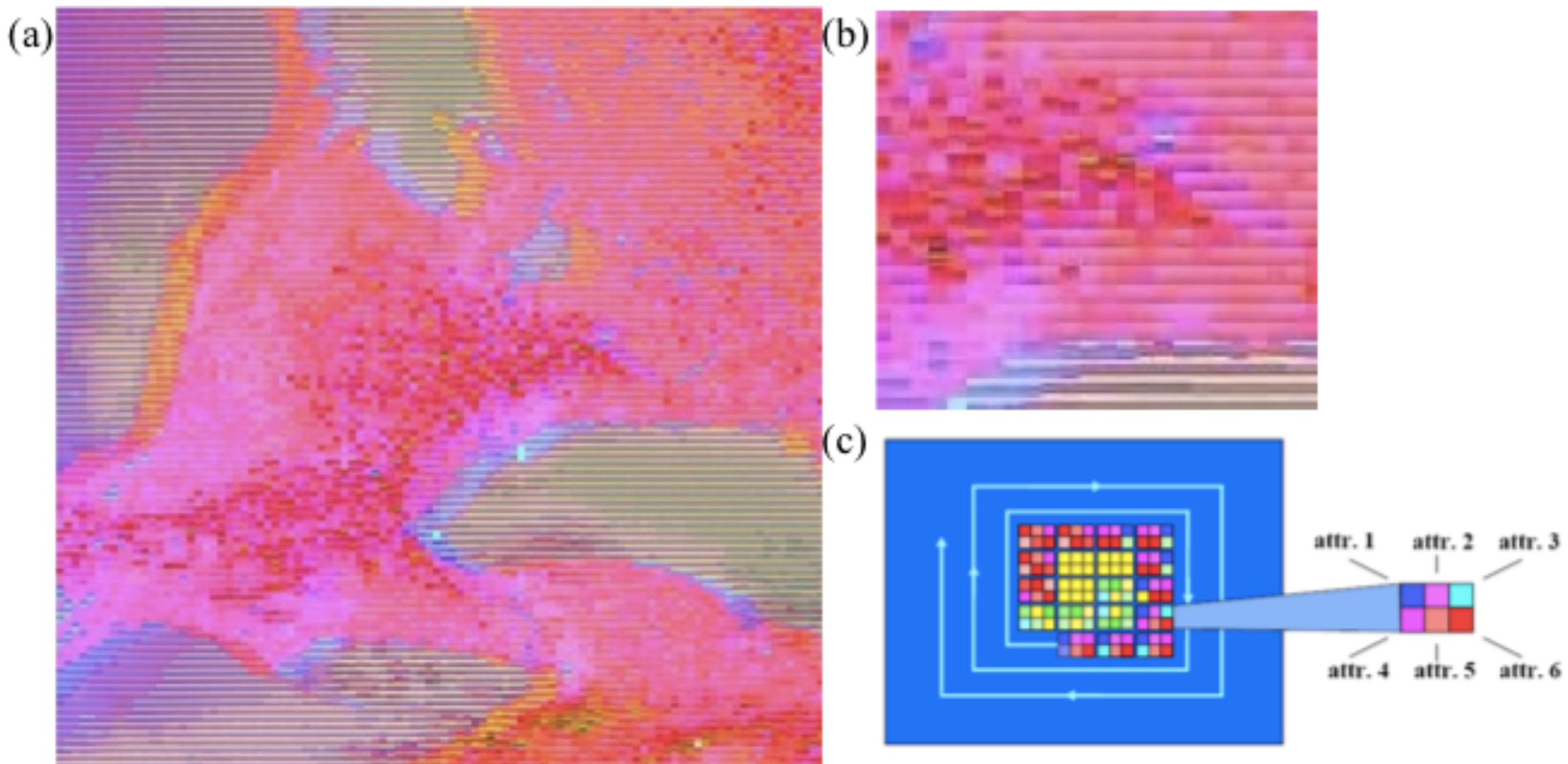
---

- Consiste na representação de um atributo por um pixel baseado em alguma escala de cores: cada cor indica um valor possível
- Para um conjunto de dados n-dimensional, n pixels coloridos são necessários para representar cada item





**Figure 3.24:** (a) Shape coding array [15], (b) An example [61].



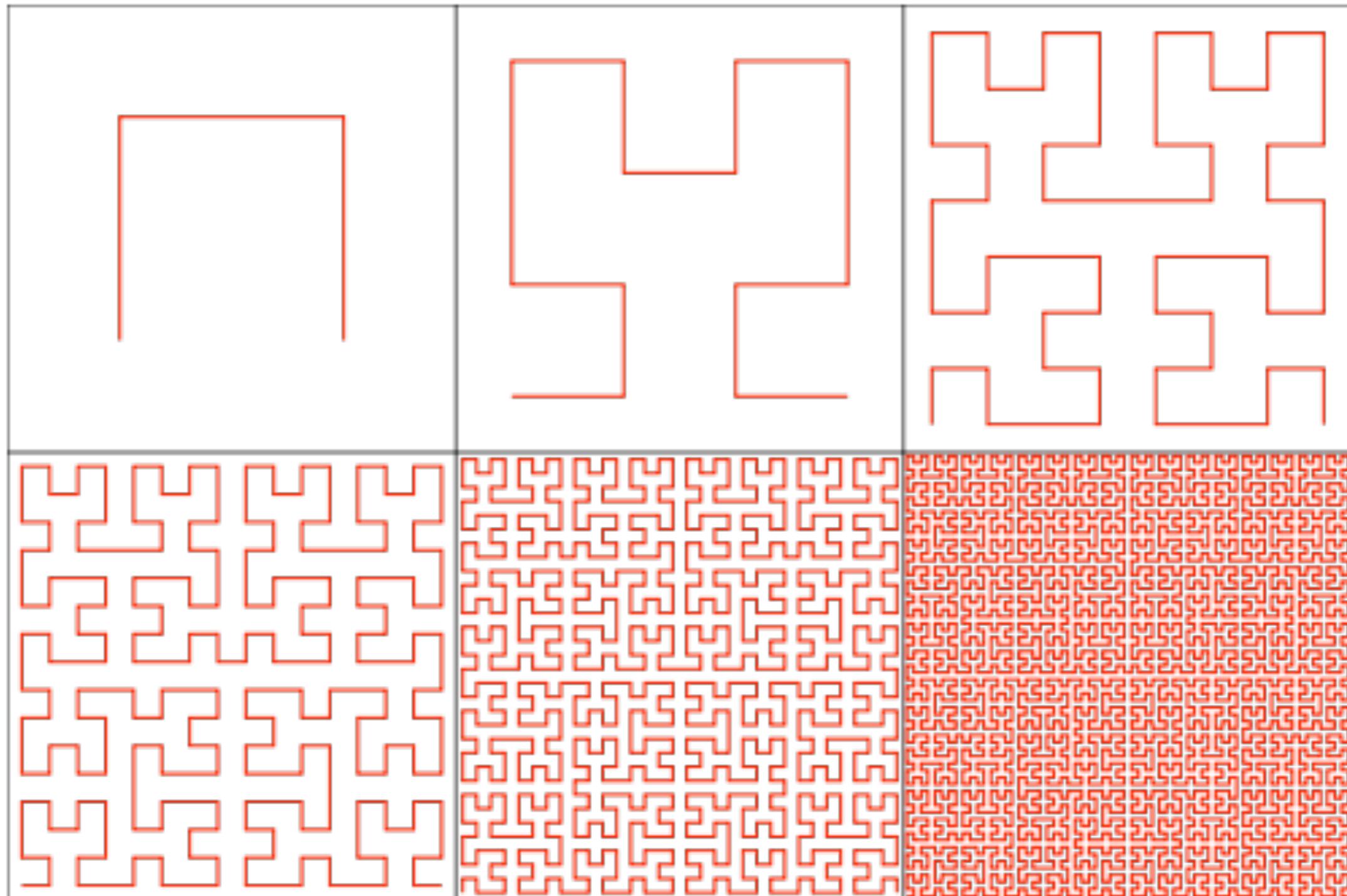
**Figure 3.25:** (a) 5D image data using color icons, (b) Part of (a) in original size [60],  
(c) Color icon scheme [15].

# TIPOS DE ABORDAGENS BASEADAS EM PIXELS

---

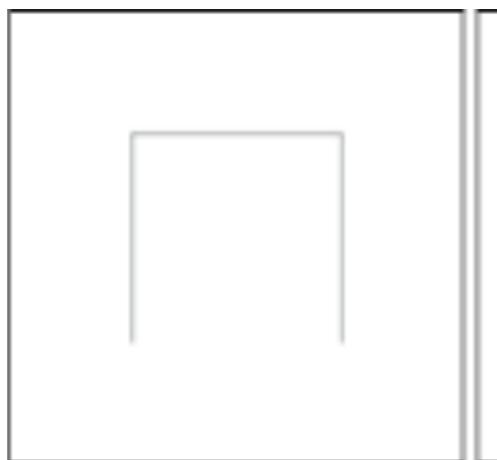
- Independentes de consulta:
  - os pixels são coloridos de acordo com os atributos reais de cada item
  - normalmente são interessantes quando algum atributo propicia uma ordenação natural dos itens
- Dependentes de consulta
  - as cores representam a diferença entre os valores da consulta e de cada item representado

# EXEMPLO DE ABORDAGEM INDEPENDENTE DE CONSULTA



# EXEMPLO DE ABORDAGEM INDEPENDENTE DE CONSULTA

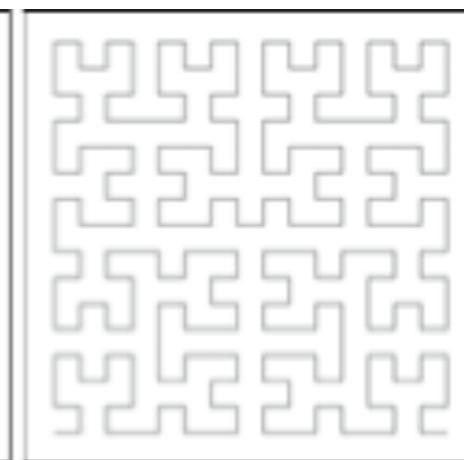
---



(a) Hilbert order 1



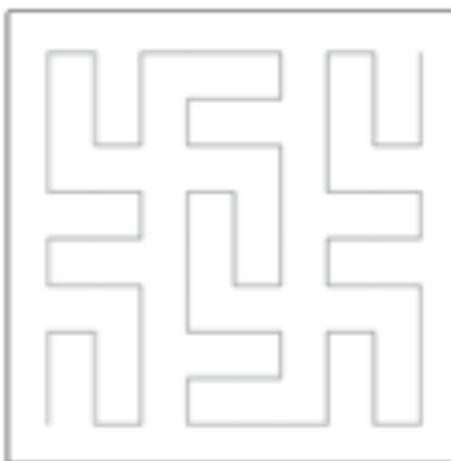
(b) Hilbert order 2



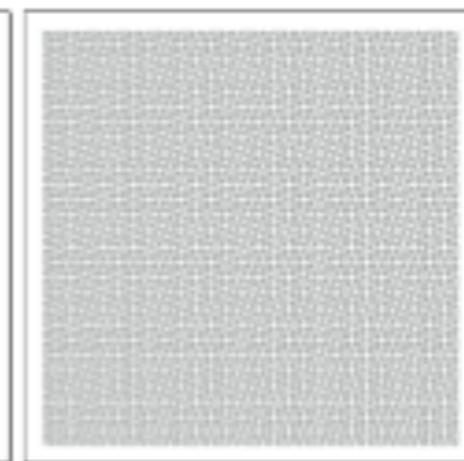
(c) Hilbert order 4



(d) Peano order 1



(e) Peano order 2



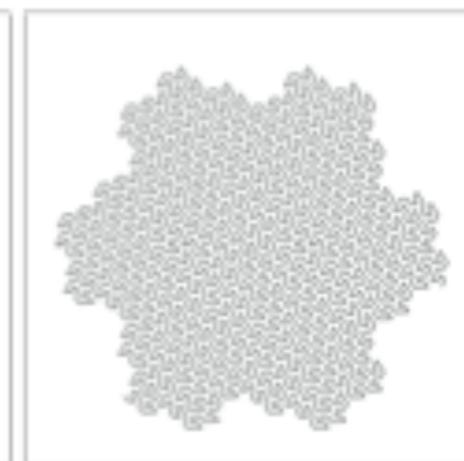
(f) Peano order 4



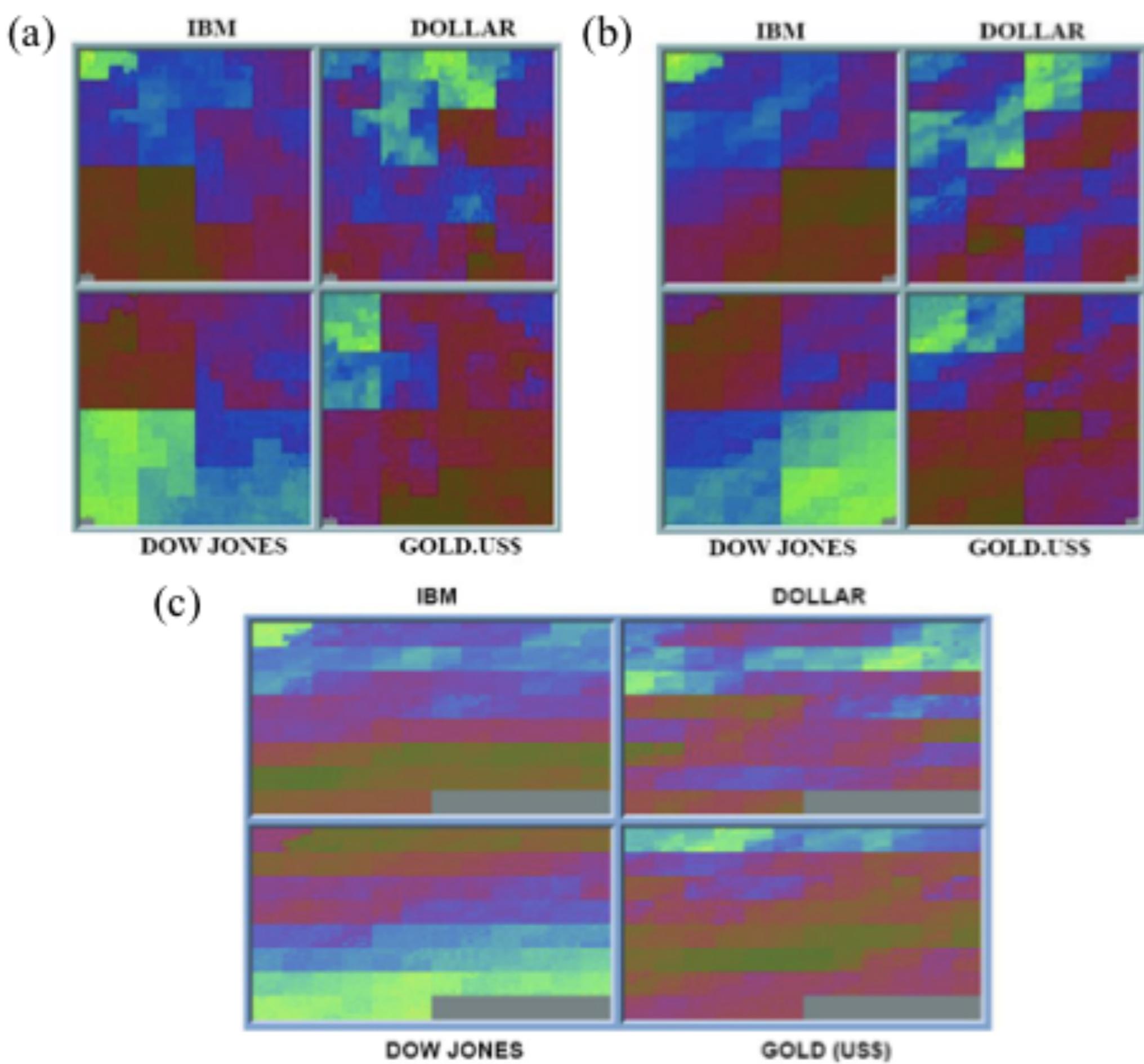
(g) Gosper order 1



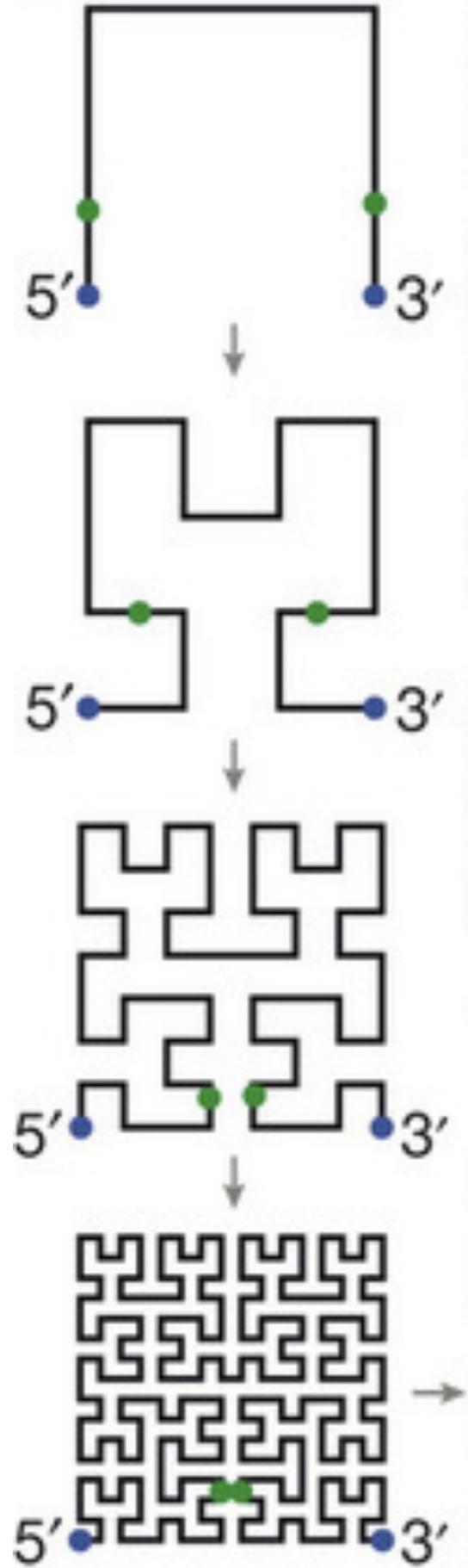
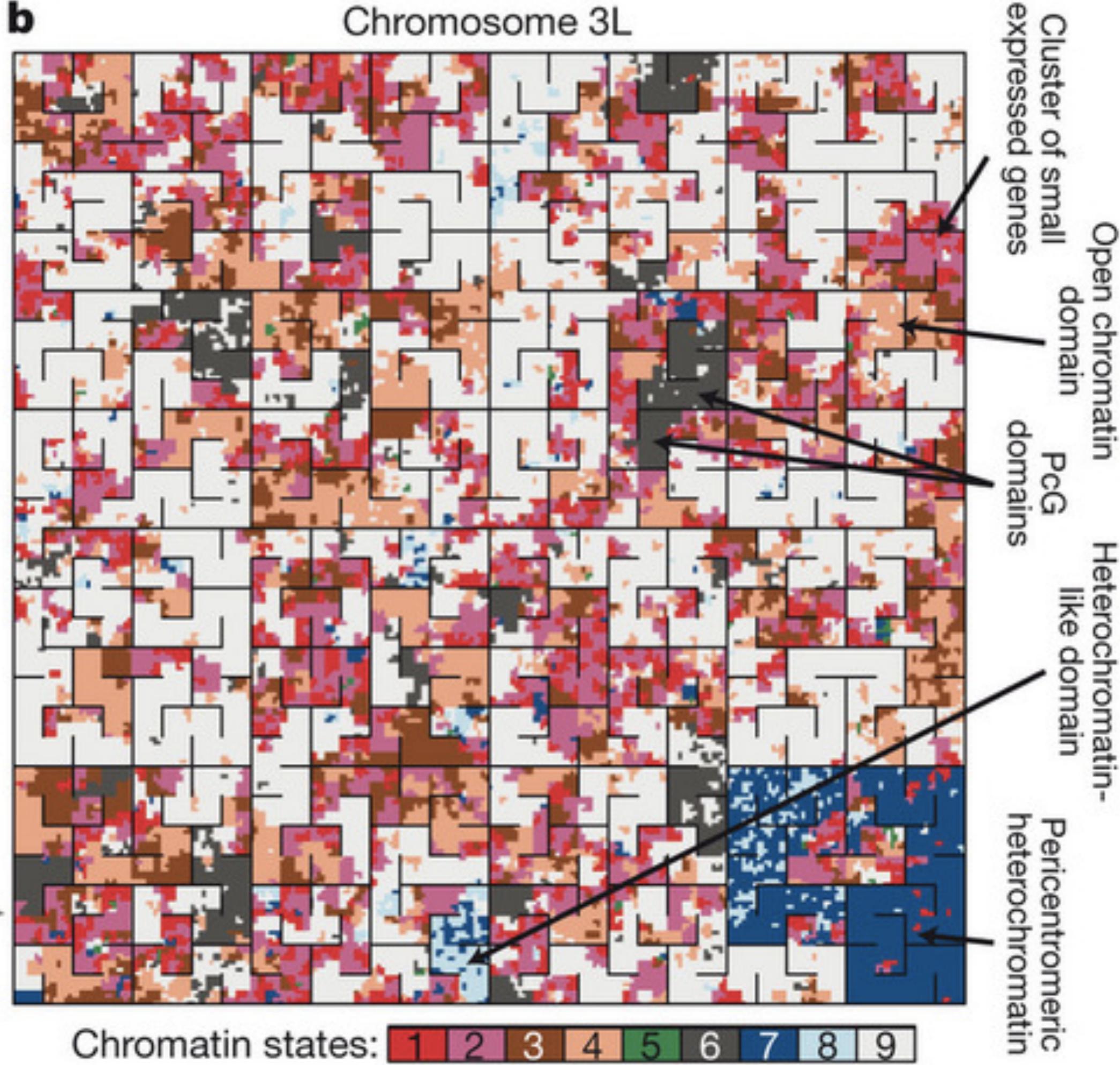
(h) Gosper order 2



(i) Gosper order 4



**Figure 3.14:** (a) Peano-Hibert, (b) Morton or Z-Curve, (c) Recursive Pattern [15].

**a****b**

Companion website for Kharchenko et al

BG3 - CHR2L



1 2 3 4 5 6 7 8 9



<b>BG3</b>	CHR2L	CHR3R
	CHR2R	CHR4
	CHR3L	CHRX

<b>S2</b>	CHR2L	CHR3R
	CHR2R	CHR4
	CHR3L	CHRX

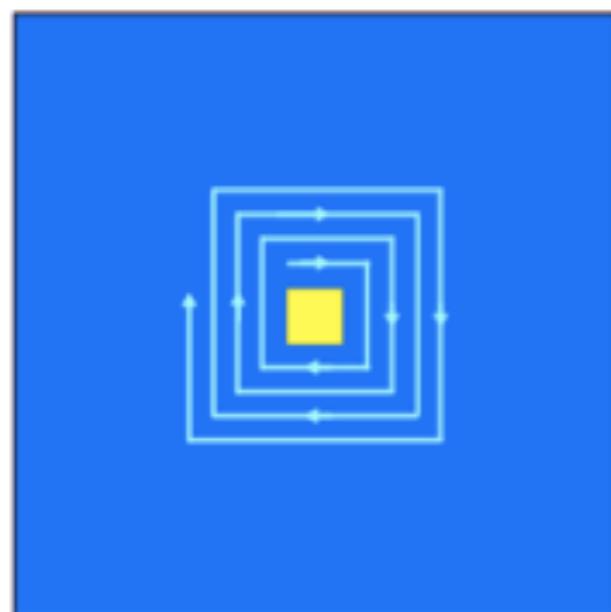
SEARCH: ELF GO

<b>VIEWS:</b>	BROWSER	RNA-SEQ
	DHS	TSS
	REPTIMING	

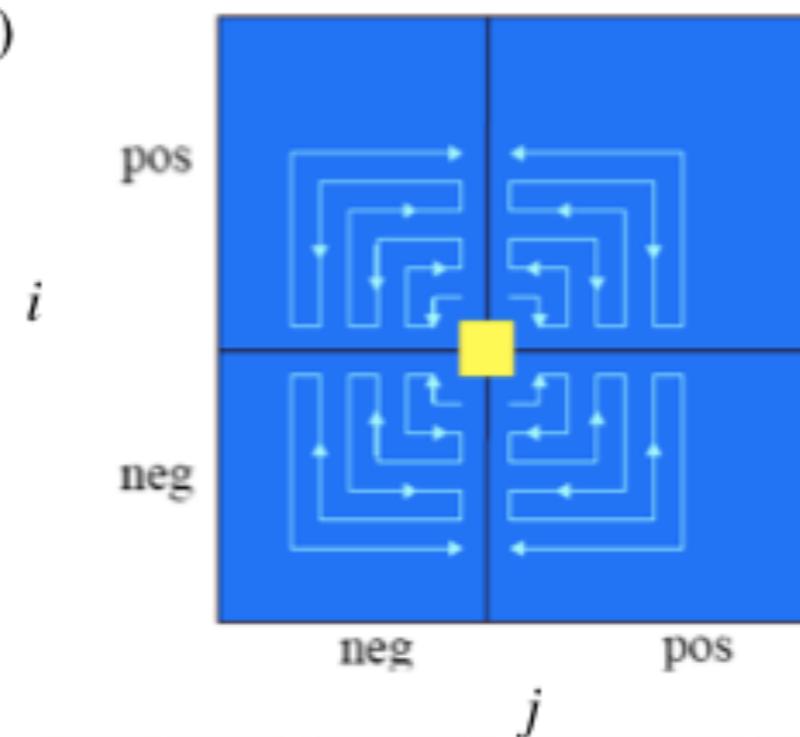
# EXEMPLO DE ABORDAGEM DEPENDENTE DE CONSULTA

---

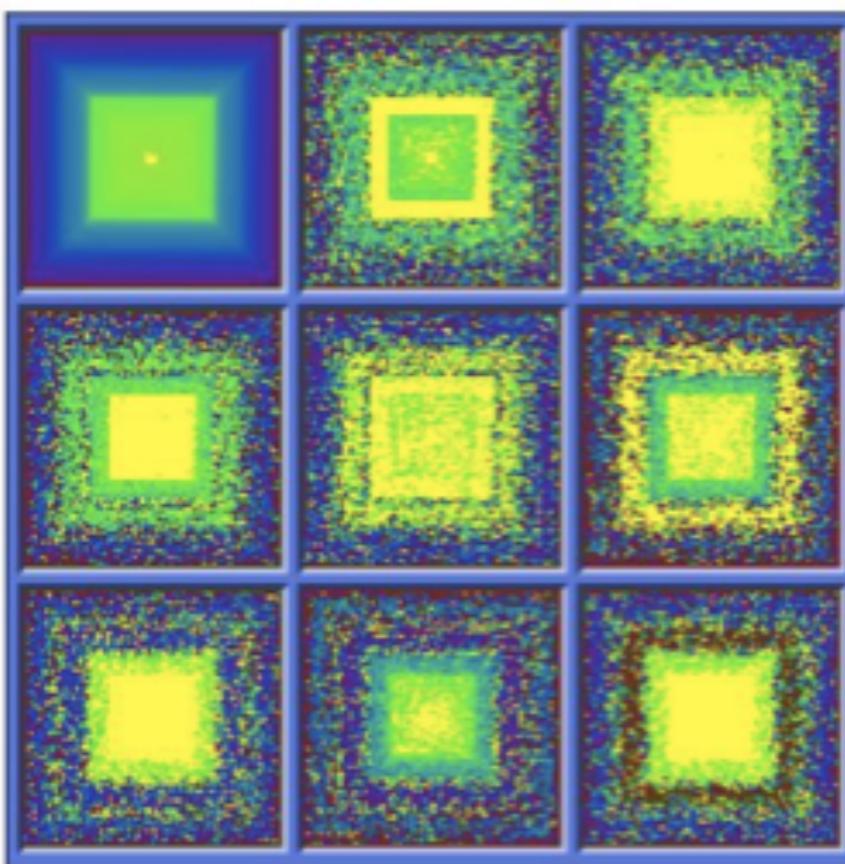
(a)



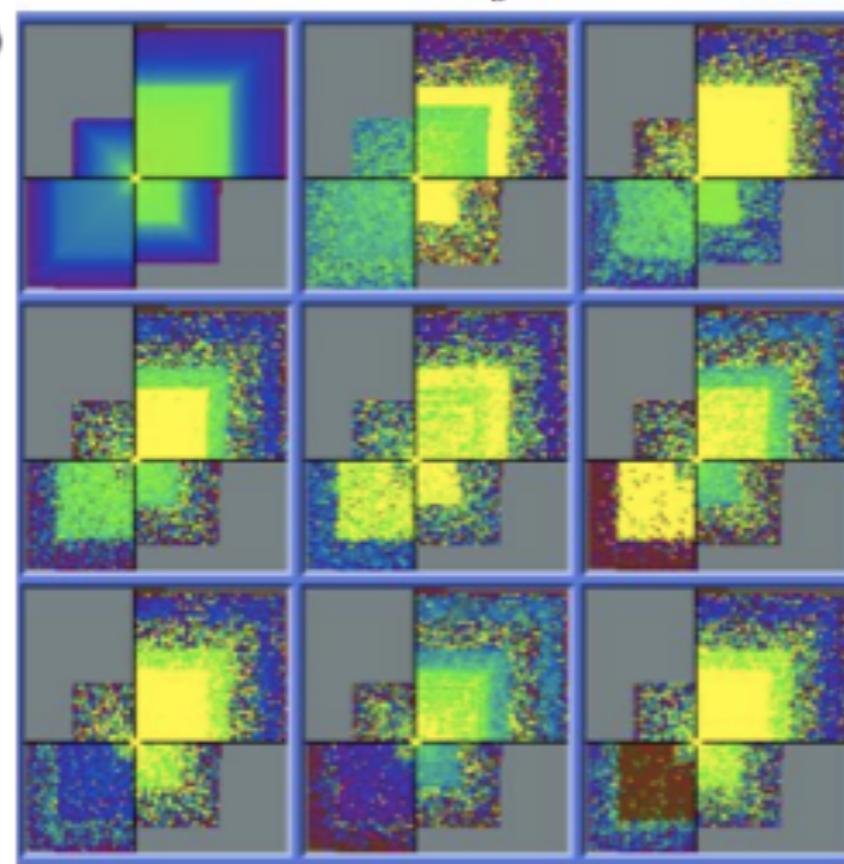
(b)



(c)

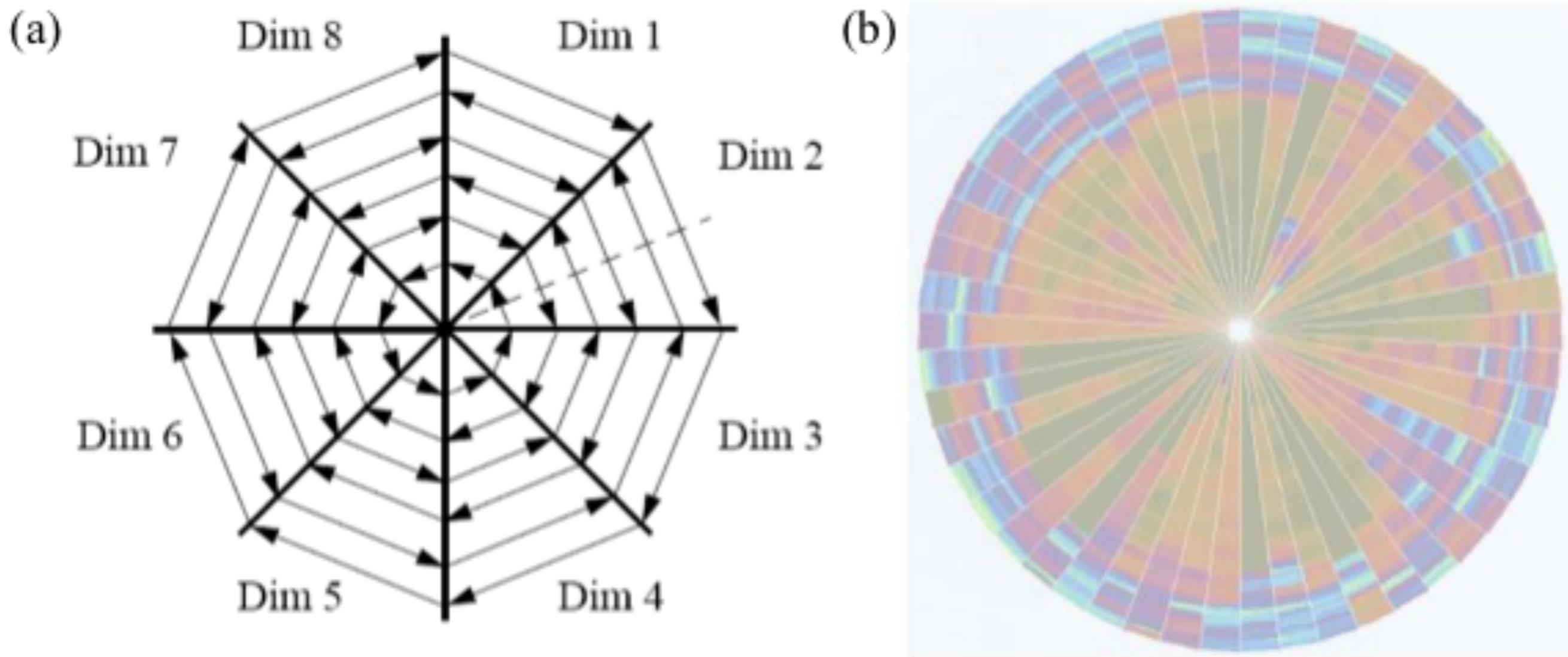


(d)



# ESPIRAL

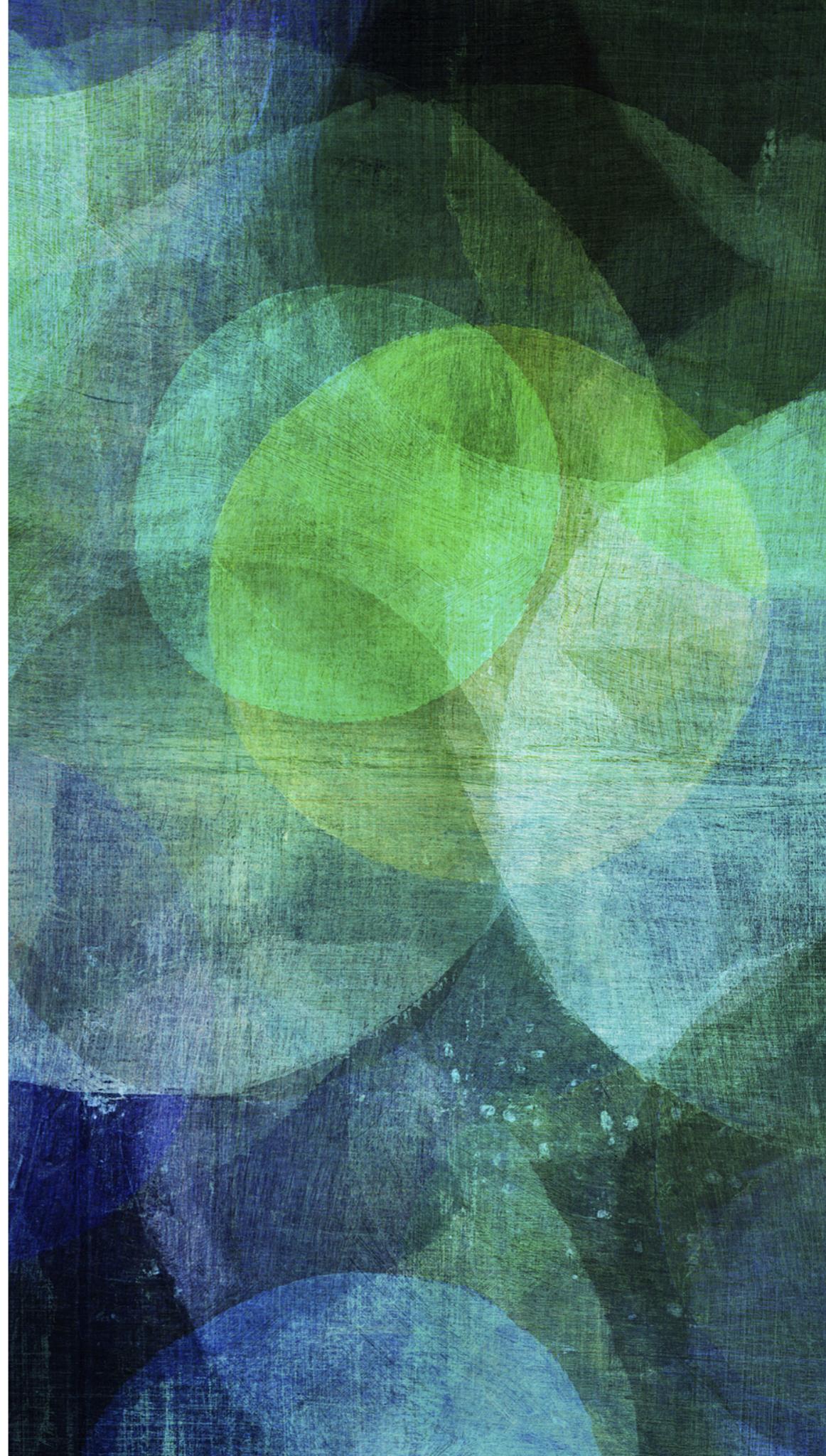
---



**Figure 3.16:** (a) Circle segment arrangement for 8-diemsional data [15],  
(b) An example of circle segments [7].

# BOAS PRÁTICAS

---

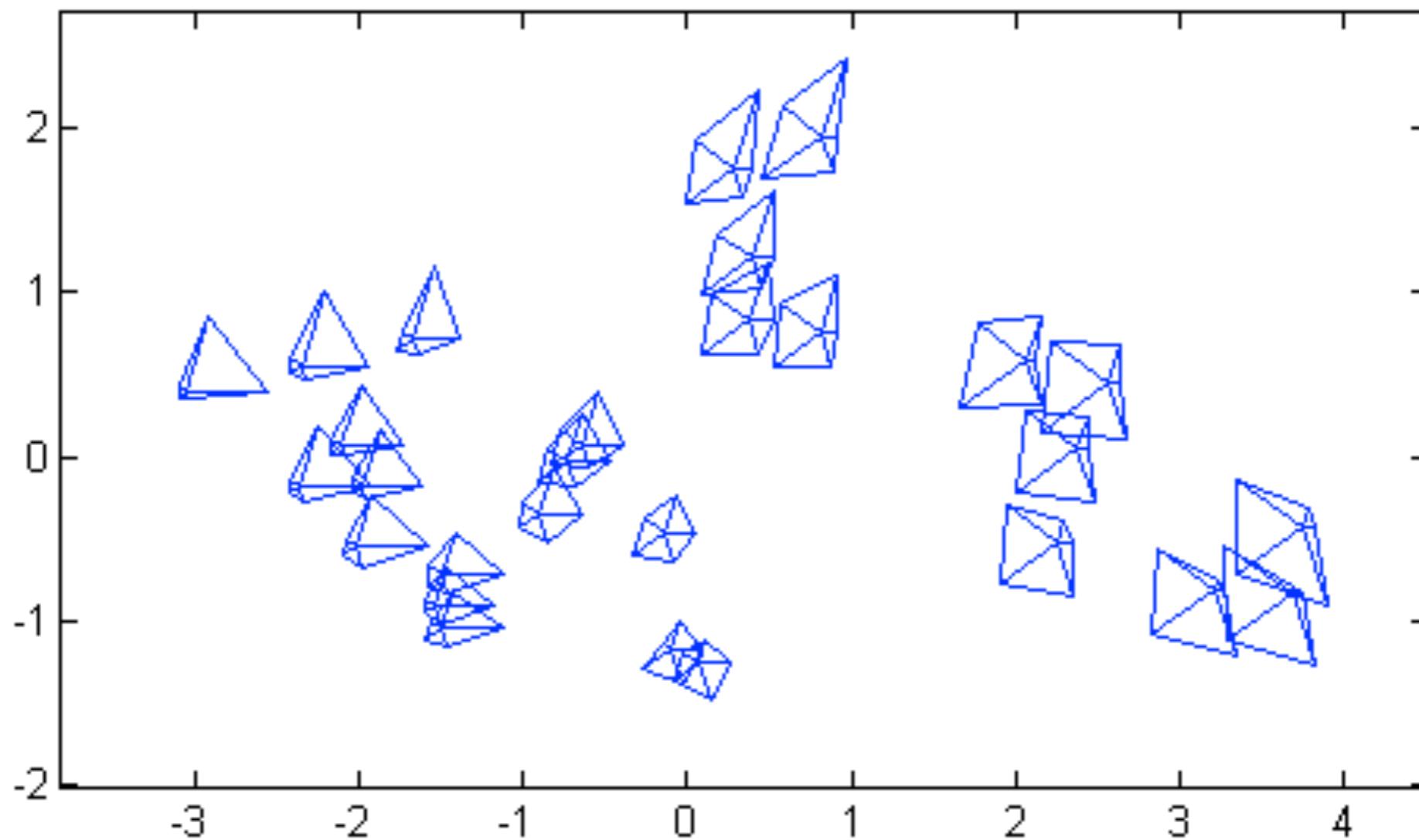


O uso de técnicas de ordenação e agrupamento dos itens por similaridade pode auxiliar na descoberta de padrões interessantes nos dados

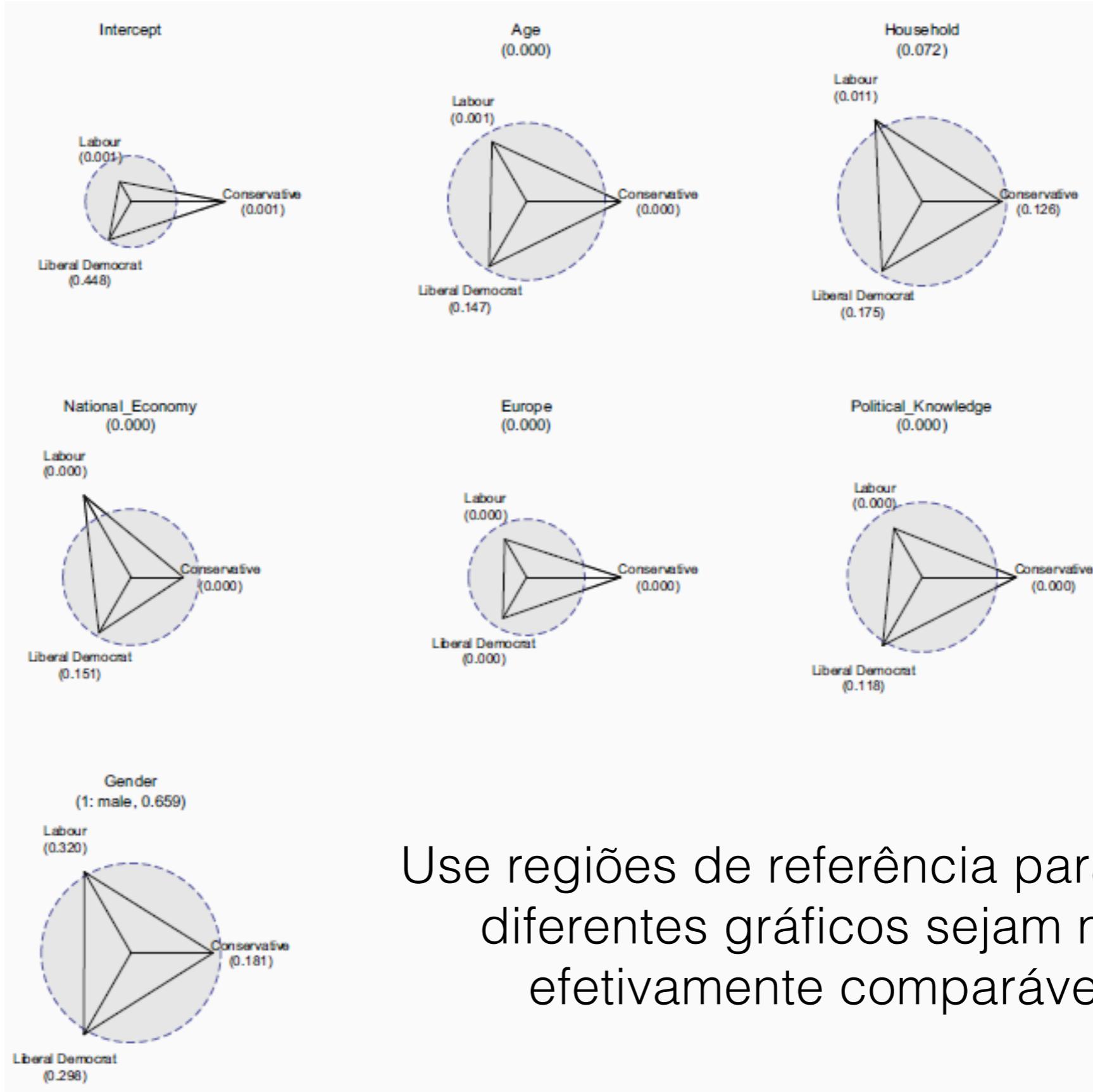
# Moral Matrix



1977 Model Year



Técnicas de redução de dimensionalidade como MDS, SVD ou PCA devem ser usadas para posicionar os objetos em espaço 2D preferencialmente



Use regiões de referência para que os diferentes gráficos sejam mais efetivamente comparáveis