

# Wizualizacja danych

*Prezentowanie wielowymiarowych danych*

*Część 2*



**Jerzy Stefanowski**  
*Instytut Informatyki*  
*Politechnika Poznańska*

Wykład TWO - grudzień 2015

# Plan wykładu

---

Dane wielowymiarowe

Prezentacje danych liczbowych

(..)

Inne rozwiązania (glyphs)

Redukcja wymiarowości i wizualizacja danych

Dane mieszane – kolejny wykład

1. Trellis display / Wykresy kratownicowe
2. Mapy ciepła

Dane jakościowe – kolejny wykład

1. Tablice logiczne
2. Mosaic plots

Dzisiaj



The greatest value of a picture is  
when it forces us to notice what we  
never expected to see.

— *John Tukey* —

# Nie spać – warto robić notatki

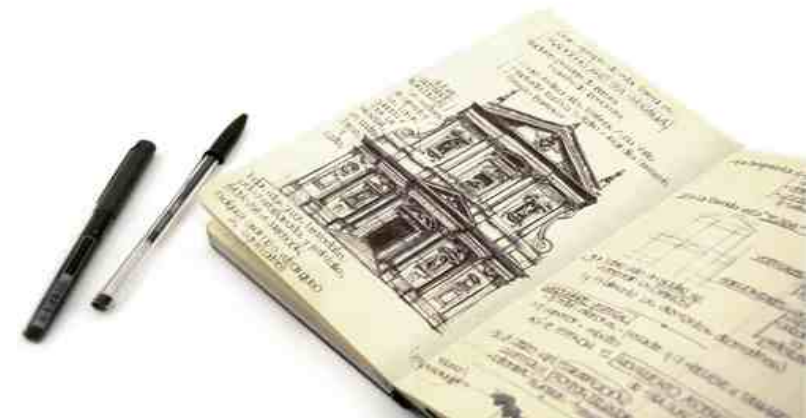


## Classic Architecture Studies

### Learning abilities College Assignment

These are some drawings and sketches from my degree of Architecture, in Spain. Indeed, they were done during the classes of the best teacher I've ever had: Mr. Joaquín Lorda.

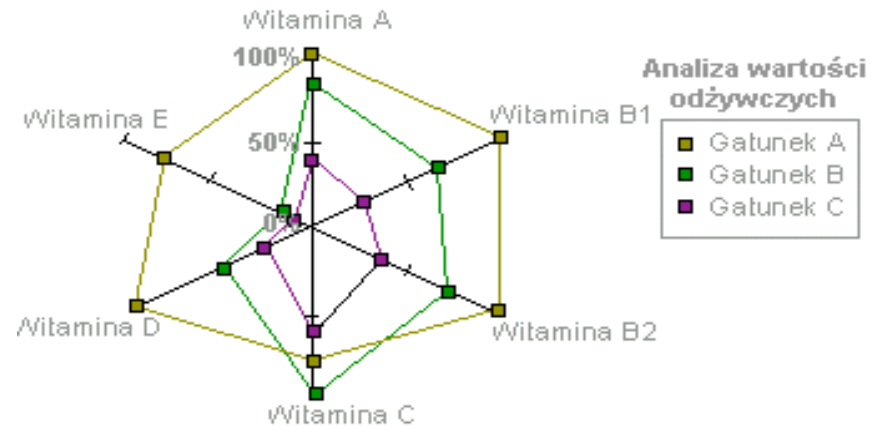
He thoroughly explained to us every single aspect of classical buildings and traditional construction, and encouraged us never to give up learning, for which I am much obliged to him!



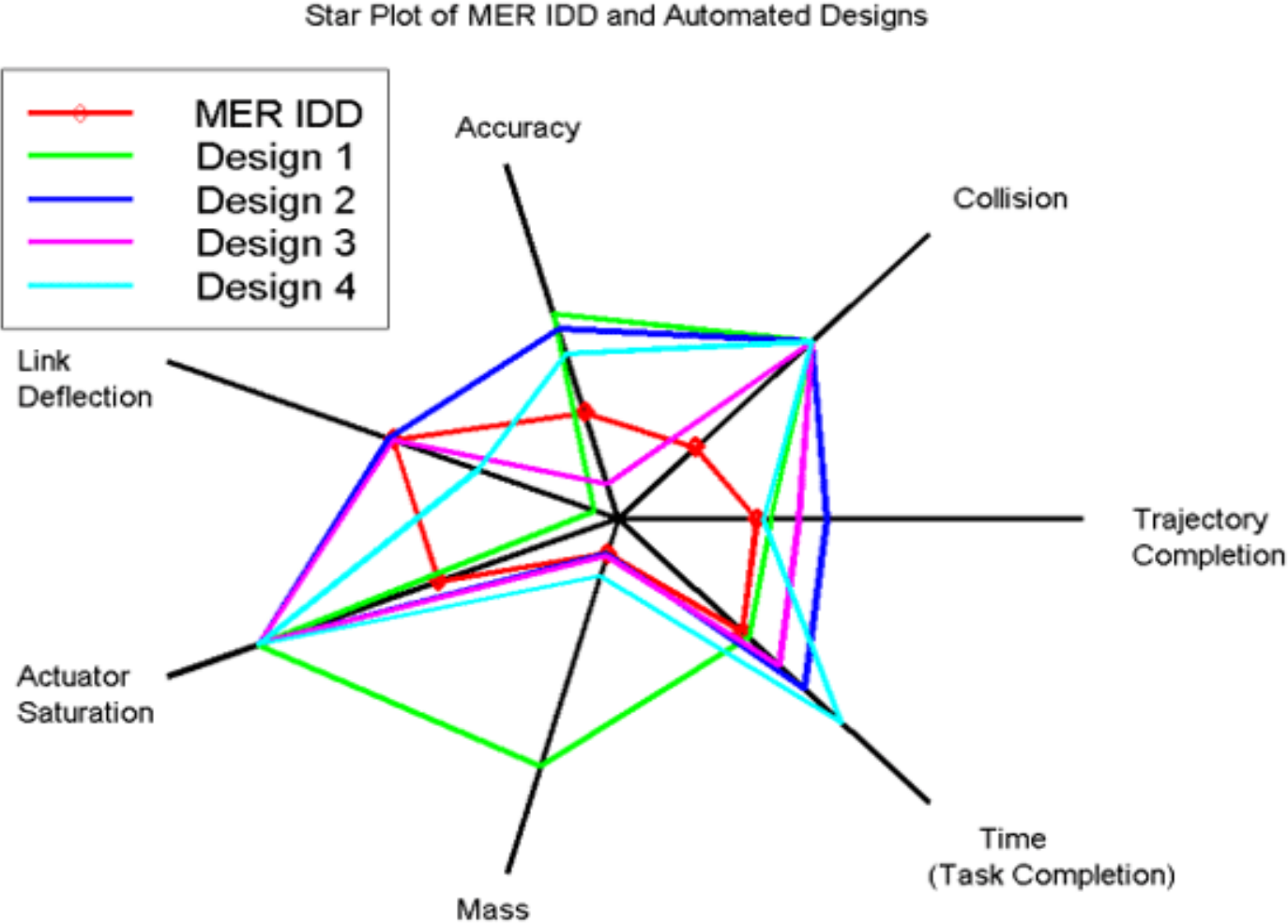
# Wykresy radarowe

---

- Porównuje się zagregowane wartości kilku serii danych (obserwacji wielowymiarowych)
- Każda seria danych na wykresie ma unikatowy kolor lub wzór i jest reprezentowana w legendzie wykresu
- Na wykresie można wykreślić jedną lub kilka serii danych



# Rader plot from NASA, with some of the most desirable design results



# Glyph Techniques

---

- Odwzoruj wartości danych to prostych elementów geometrycznych, tzw. glif
- Sprawdź definicje “a glyph”
- Zaproponowano wiele podstawowych glifów
  - Star glyphs
  - Faces
  - Arrows
  - Sticks
  - Shape coding

▼ ————— *Angielski* —————

**glyph** | glif |

noun

1 a hieroglyphic character or symbol. *flanges painted with esoteric glyphs.*

- a sculptured symbol (e.g. as forming the ancient Mayan writing system). *these glyphs refer to an ancient Olmec ruler.*

- Computing a small graphic symbol.

2 Architecture an ornamental carved groove or channel, as on a Greek frieze.

# Glyph Layouts

- How do we place the glyphs on a chart?
- Sometimes there will be a natural location – for example?
- If not... two of the varieties can be allocated to spatial position, and the remainder to the attributes of the glyph

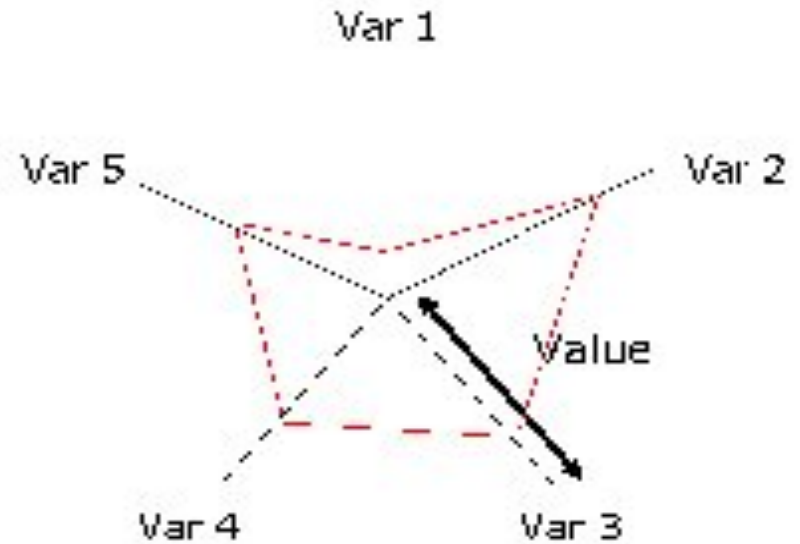




# Glyph Techniques – Star Plots

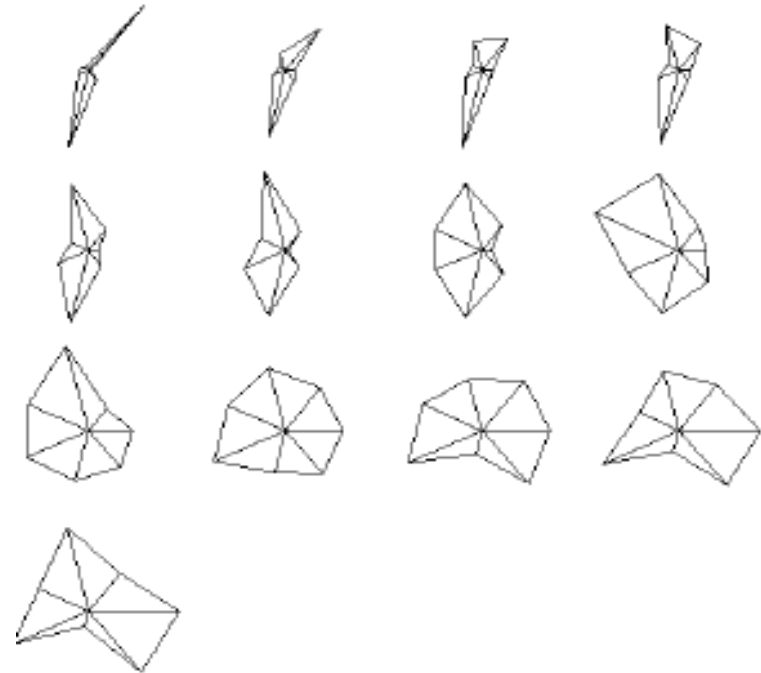
---

- Each observation represented as a 'star'
- Each spike represents a variable
- Length of spike indicates the value



## Glyph Techniques – Star Plots

- Each observation represented as a 'star'
- Each spike represents a variable
- Length of spike indicates the value



Crime in  
Detroit

### Key for Glyphs:

```
ft_police; 0 degrees  
unemp; 51 degrees  
manu_wrkr; 102 degrees  
handgun_lcs; 154 degrees  
gov_wrkr; 205 degrees  
cleared; 257 degrees  
homicides; 308 degrees
```

# Cars dataset – porównywanie wielu samochodów

---

The variable list for the sample star plot is:

Price

Mileage (MPG)

1978 Repair Record (1 = Worst, 5 = Best)

1977 Repair Record (1 = Worst, 5 = Best)

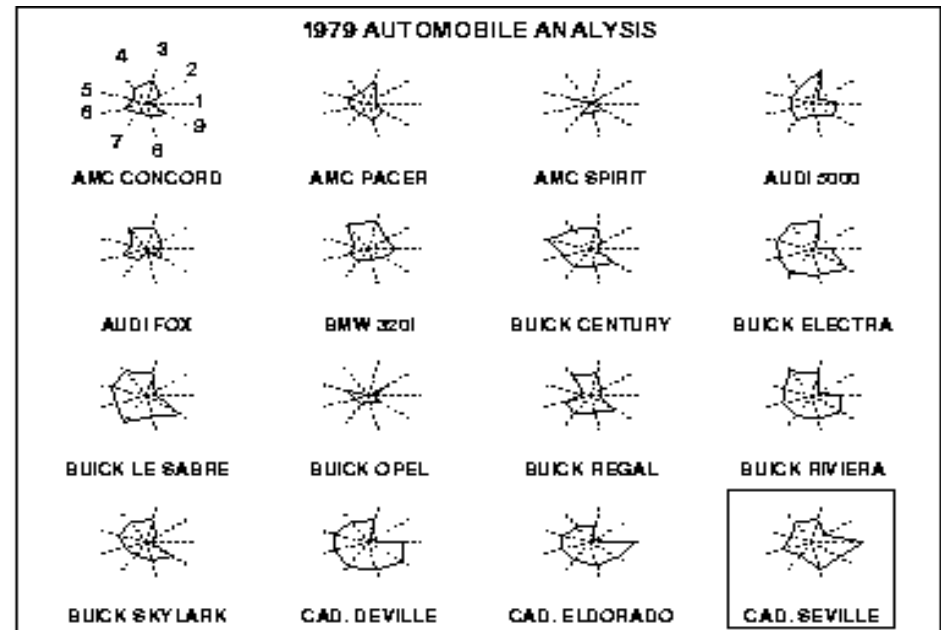
Headroom

Rear Seat Room

Trunk Space

Weight

Length



## Cars dataset – porównywanie wielu samochodów

---

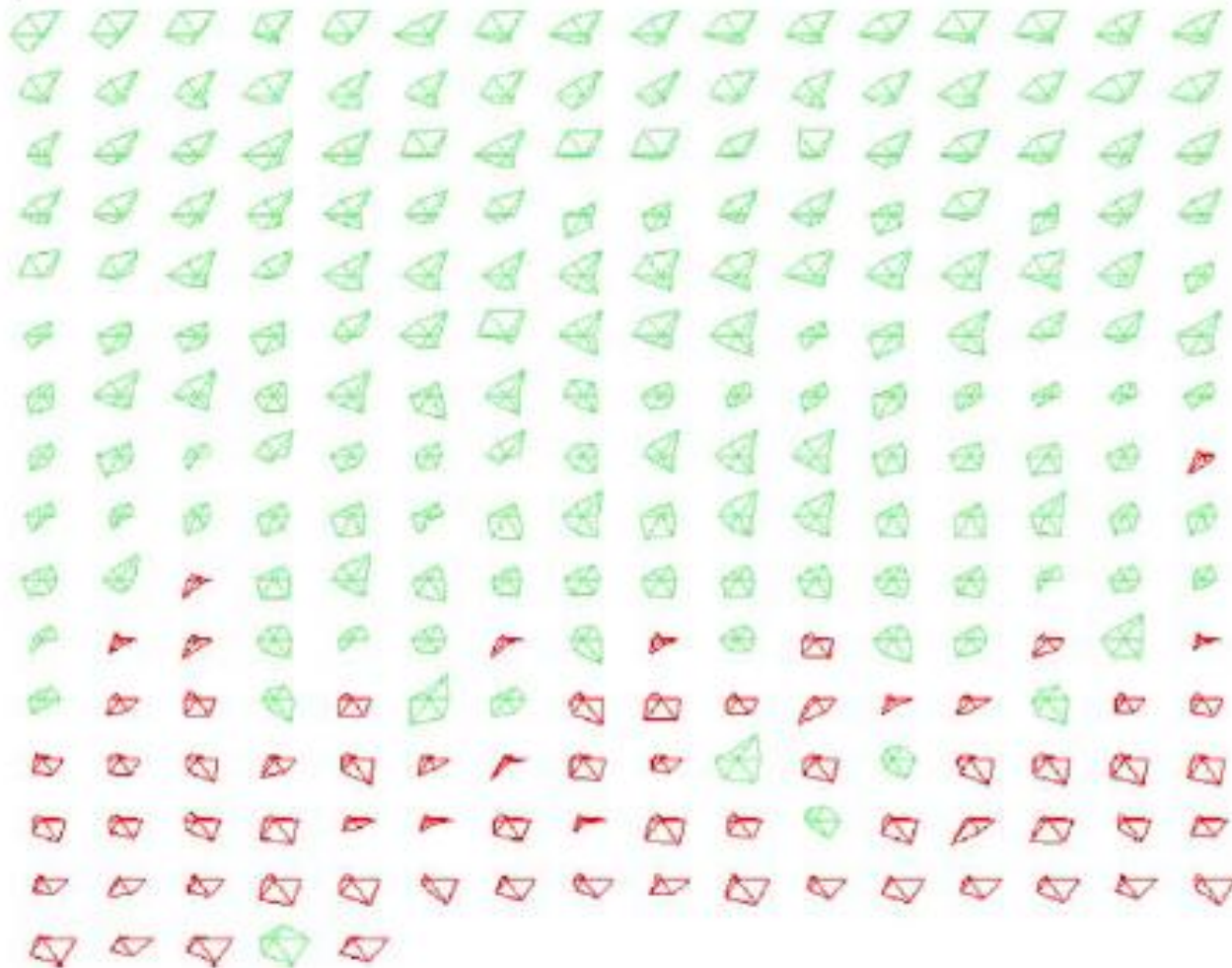
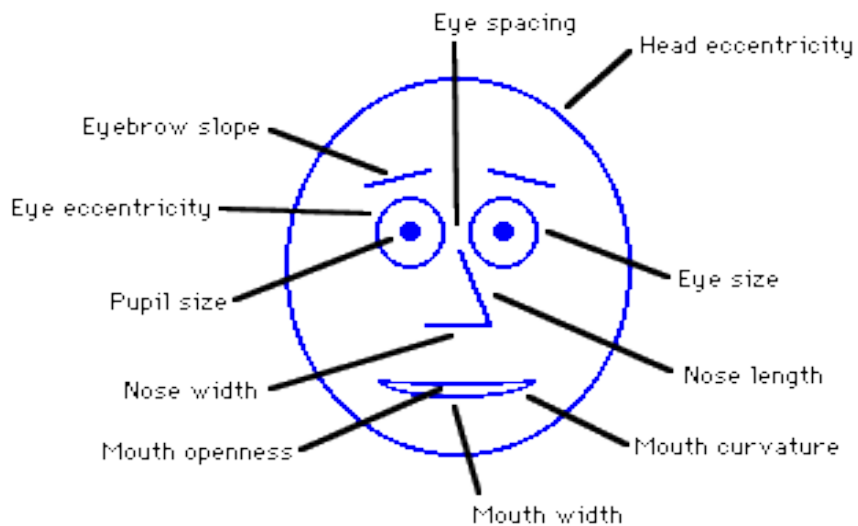


Figure 7.7. Example of ordering glyphs according to one dimension. In this case, the data set is a subset of the cars data, sorted by the MPG variable (*top to bottom* corresponds to low to high MPG). The *highlighted (dark) glyphs* represent four-cylinder cars. A clear grouping of shapes is visible. A few outliers can be seen near the *bottom* of the figure, which represent six-cylinder cars with good MPG

# Chernoff Faces

Zakoduj zmienne jako charakterystyczne elementy ludzkiej twarzy



Kia Cee'd



Skoda Octavia



Ford Focus



Volkswagen Golf

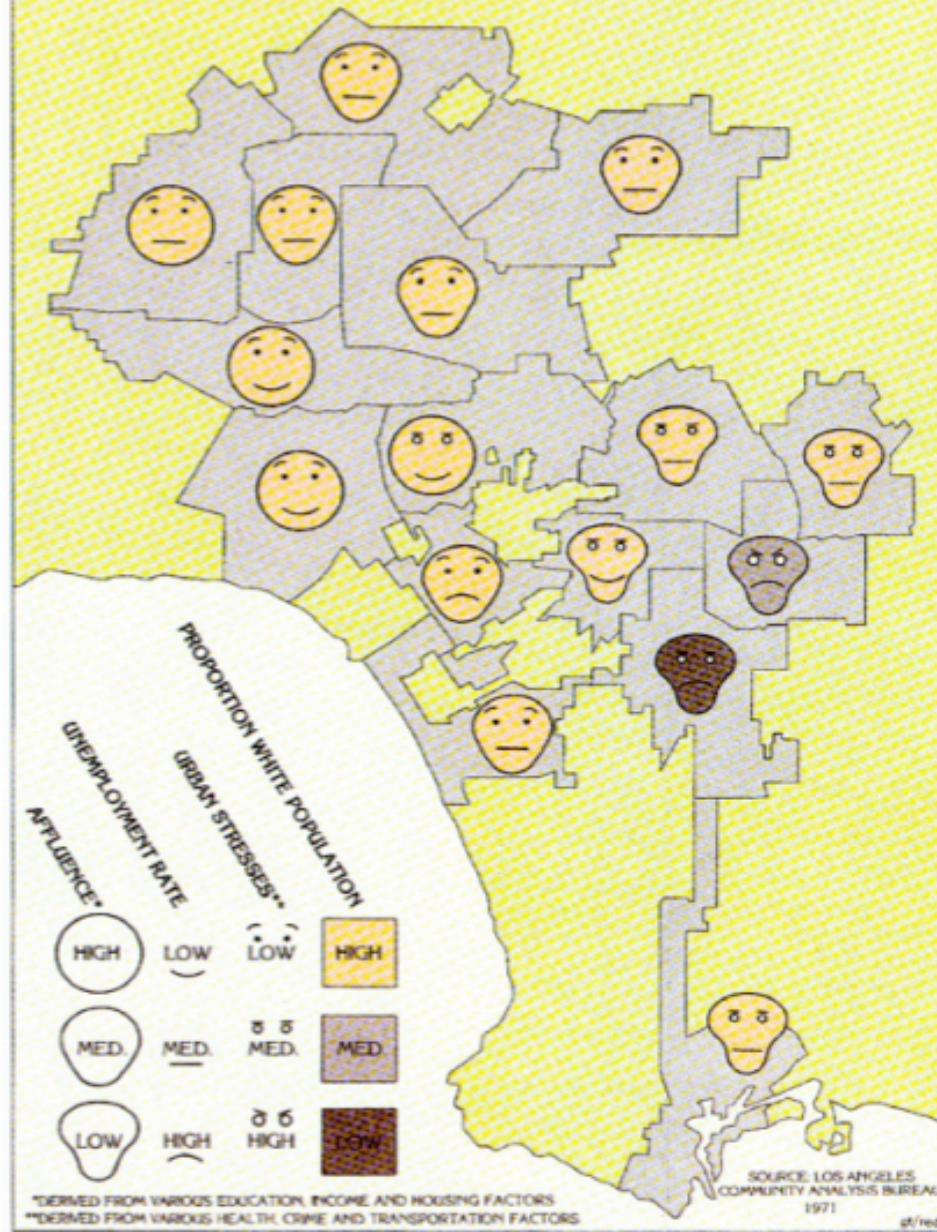


Cute applets:

<http://www.cs.uchicago.edu/~wiseman/chernoff/>

<http://hesketh.com/schampeo/projects/Faces/chernoff.html>

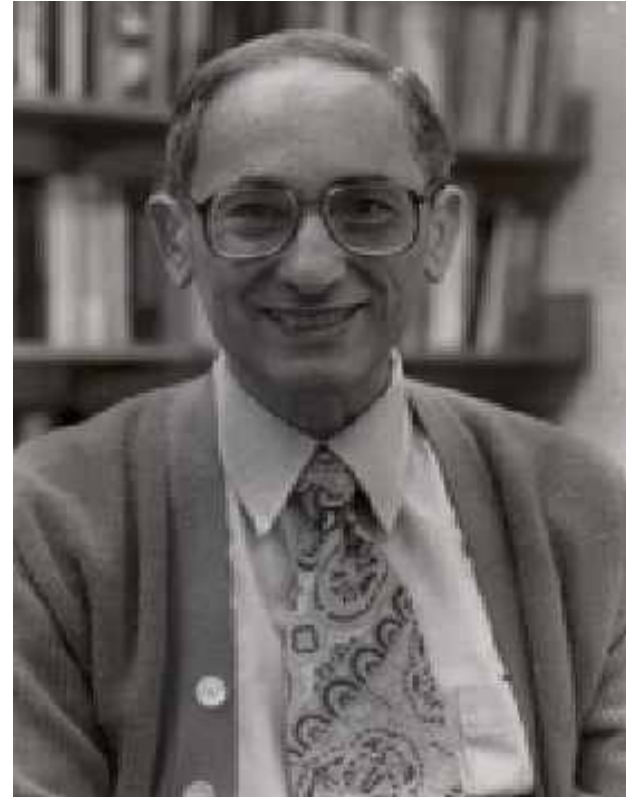
# Life in Los Angeles



## Chernoff's Face

---

- .. And here is Chernoff's face 😊



# Redukcja wymiarowości i prezentacja danych

Trochę statystycznej analizy danych oraz metod  
sztucznej inteligencji

Wykorzystamy dodatkowe slajdy



# Prezentacja danych mieszanych

---

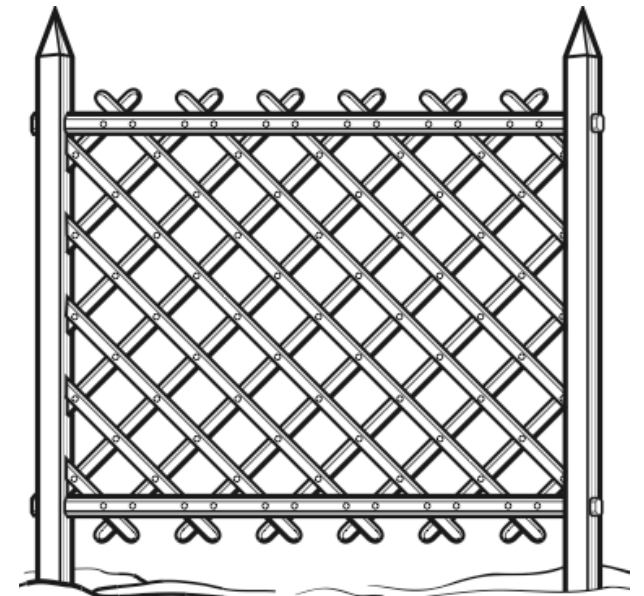
- Wiele zbiorów danych wykorzystuje atrybuty mieszanego typów – **ilościowe** oraz **jakościowe**
- Prezentacja może wykorzystać poznane wcześniej wykresy
- Odmiany wykresów rozrzutu
  - Przykłady z poprzedniego wykładu
- **Specjalizowane rozwiązania**
  - Kratownica wykresów – Trellis displays
  - Mapy ciepła
  - oraz inne ...

# Trellis Displays – Kratownica wykresów

---

- Rodzaj zbioru wielu wykresów umieszczone we wspólnej strukturze kratownicy (tzw. grid – siatka kartograficzna)
  - Uwarunkowanie wykresów względem wybranej zmiennej (tzw. kategoryzacja)
  - Wykorzystanie wspólnej skali dla wszystkich wykresów – ułatwia porównywania
  - Możliwość zbudowania wielu układów „kratownicy”

Słownik: a framework of light wooden or metal bars used as a support for fruit trees or creepers, typically fastened against a wall.



# Kratownica wykorzystująca boxplots

---

Najprostsza wersje – wykorzystanie wiele wykresów pudełkowych

- Minimum para zmiennych
  - zmienna liczbowa oś Y;
  - zmienna jakościowa na osi X – kategoryzuje wykresy podstawowe
- Przykład
  - cars data

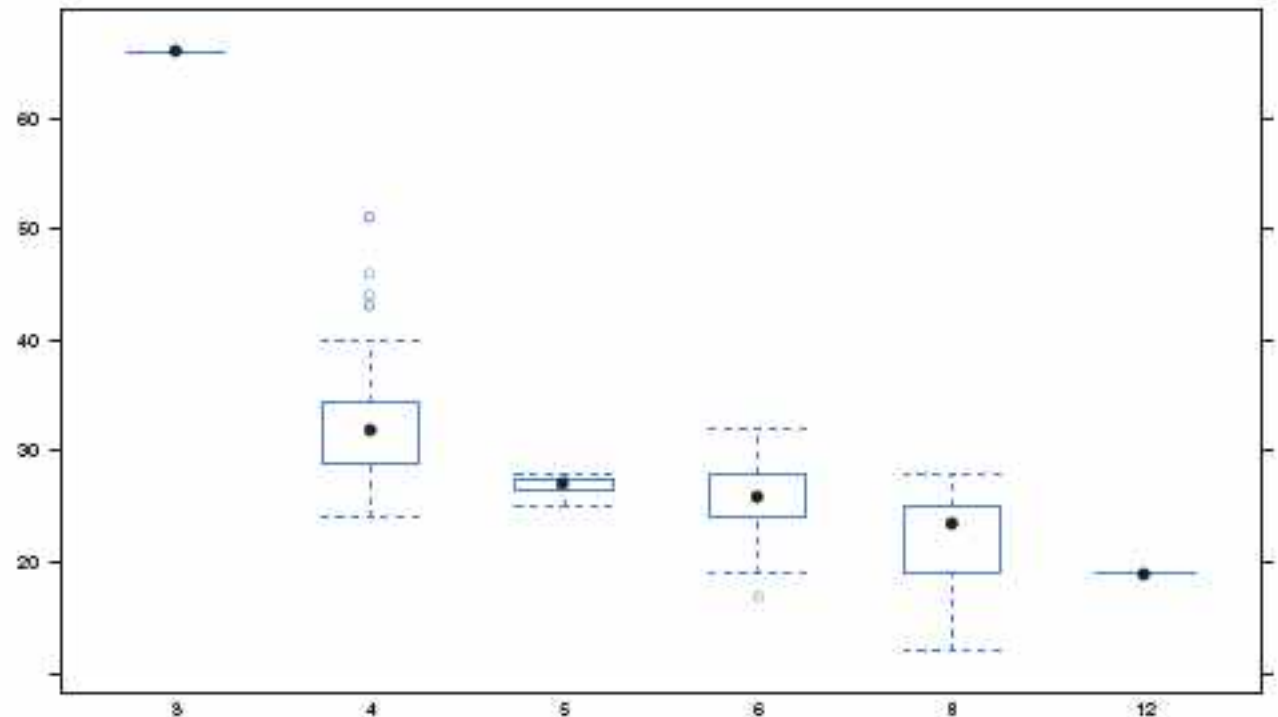
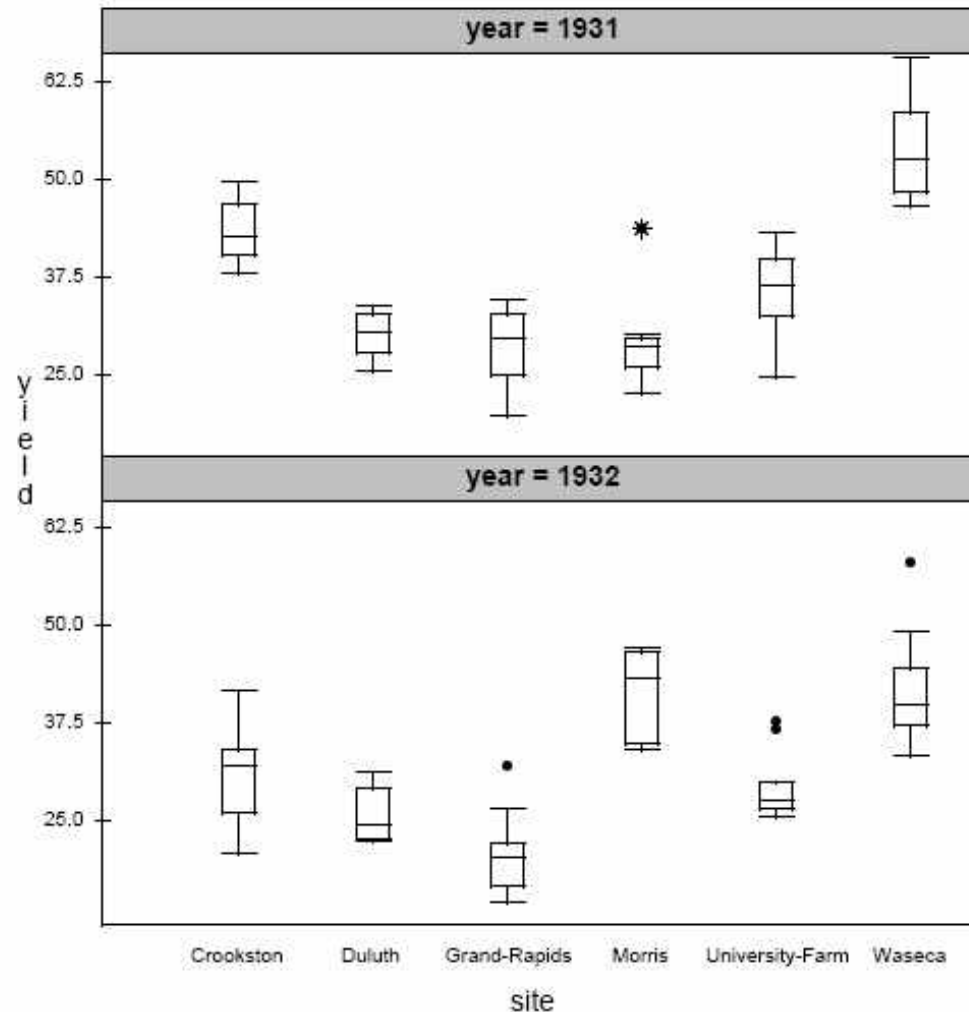


Fig. 2.18. Boxplot MPG by Cylinder as simple form of a trellis display.

# Kategoryzacja w kracie wykresów pudełkowych



Inne dane;  
Rok jako dodatkowa  
kategoryzacja

Figure 3: Parallel boxplots serve for judging ANOVA-models

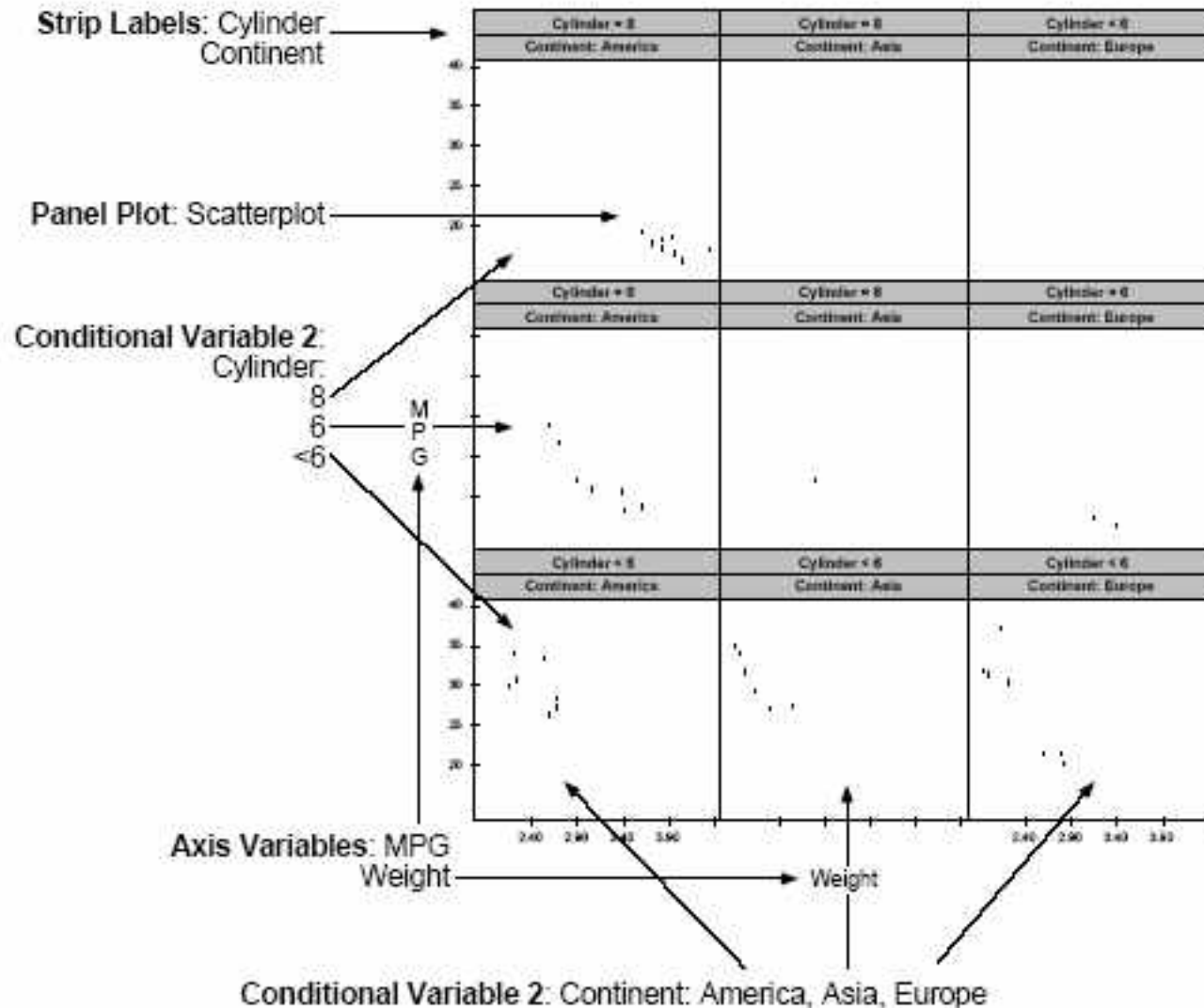
# Trellis Displays – Kratownica wykresów rozrzutu

---

Wykresy kratownicowe mają większą moc jeśli stosujemy:

- Uwarunkowanie (tzw. kategoryzacja) większą liczbą zmiennych
- Wykorzystanie bardziej złożonych wykresów podstawowych, chociażby XY scatterplots
- Przykład dane cars

# Trellis Displays – Kratownica wykresów rozrzutu



# Trellis Displays – Cars - poziomy

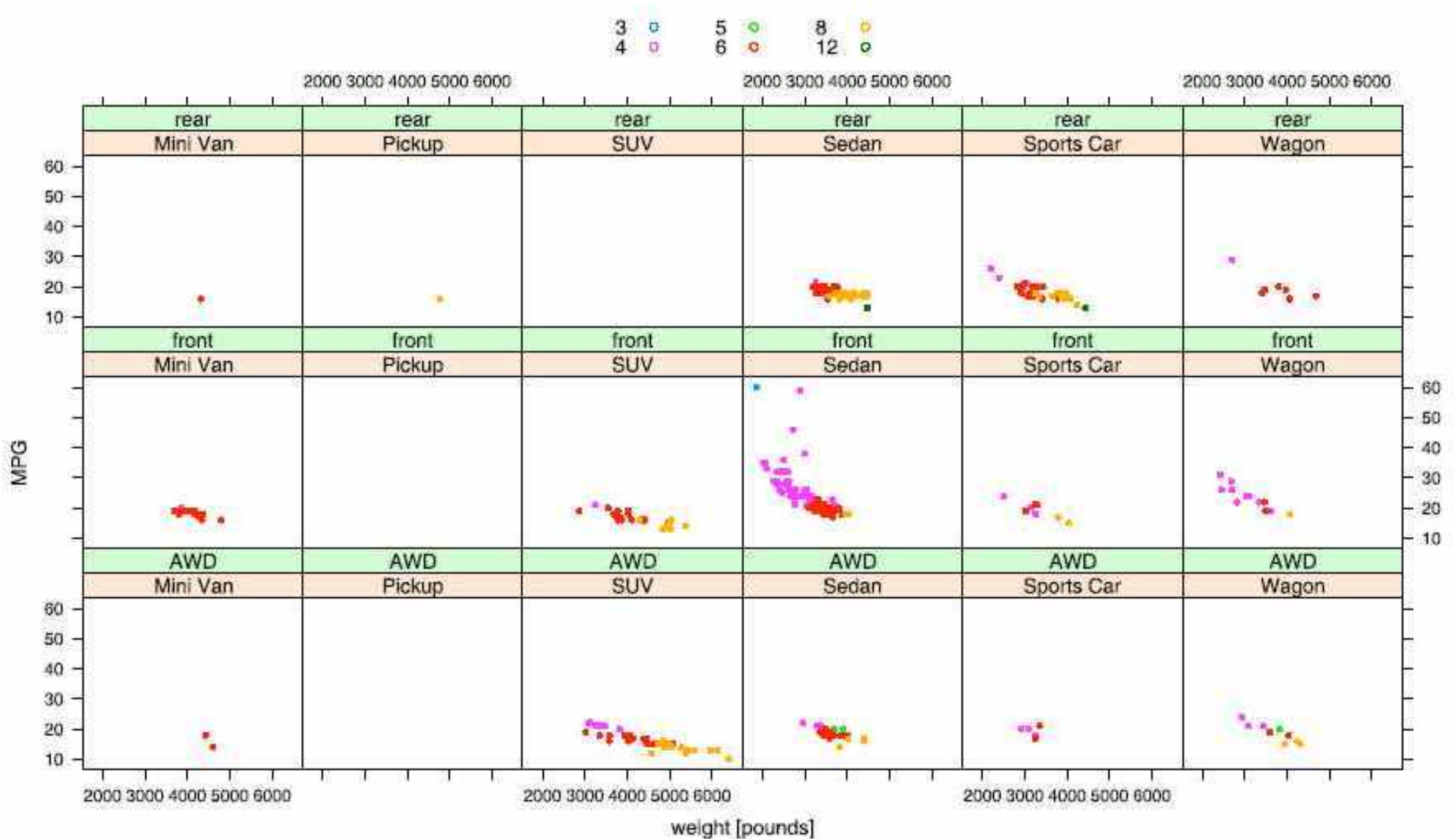


Figure 6.5. [This figure also appears in the color insert.] A trellis display incorporating five variables of the cars data set

# Trellis Displays – Cars – zaznaczenie modelu

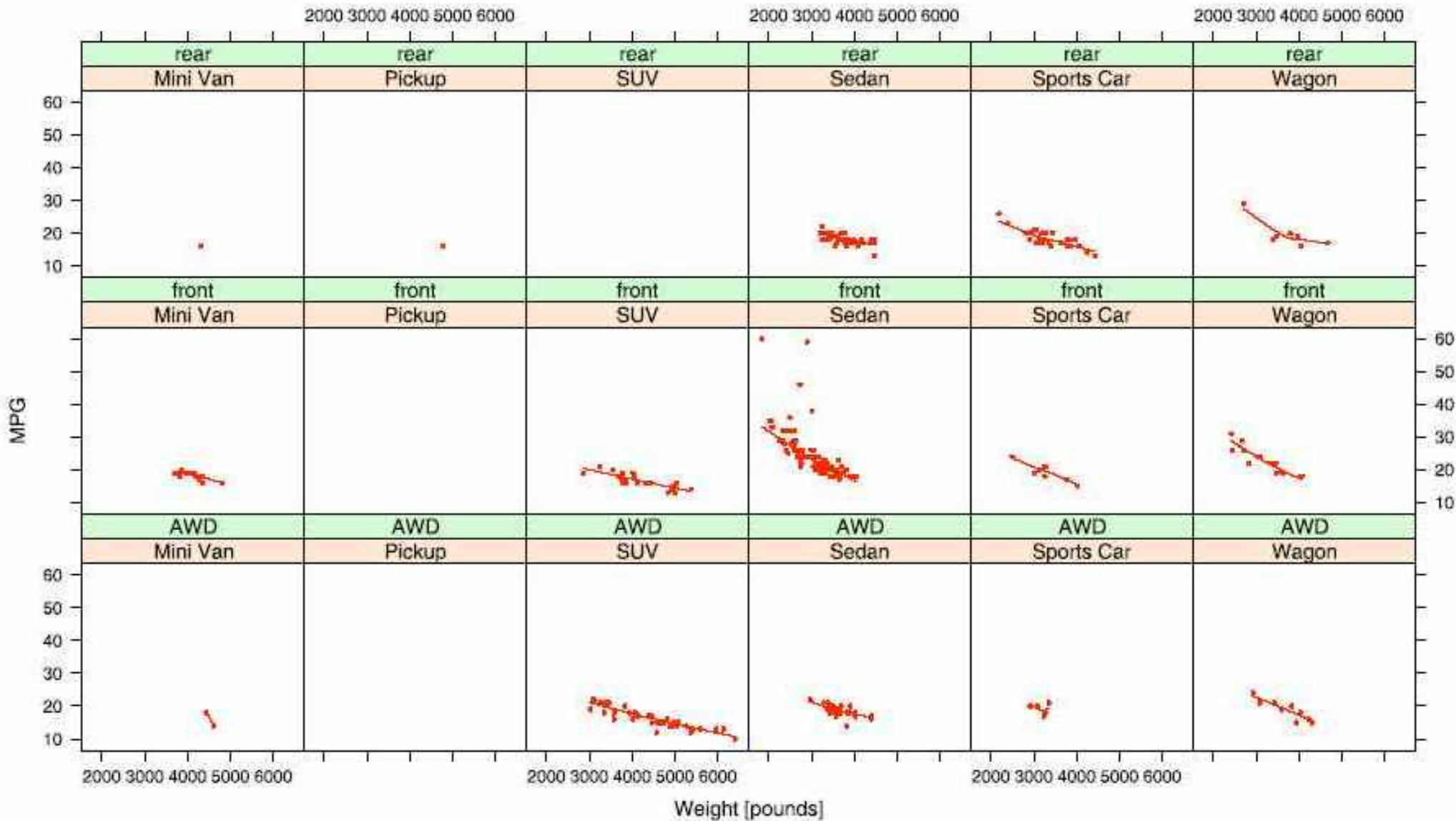
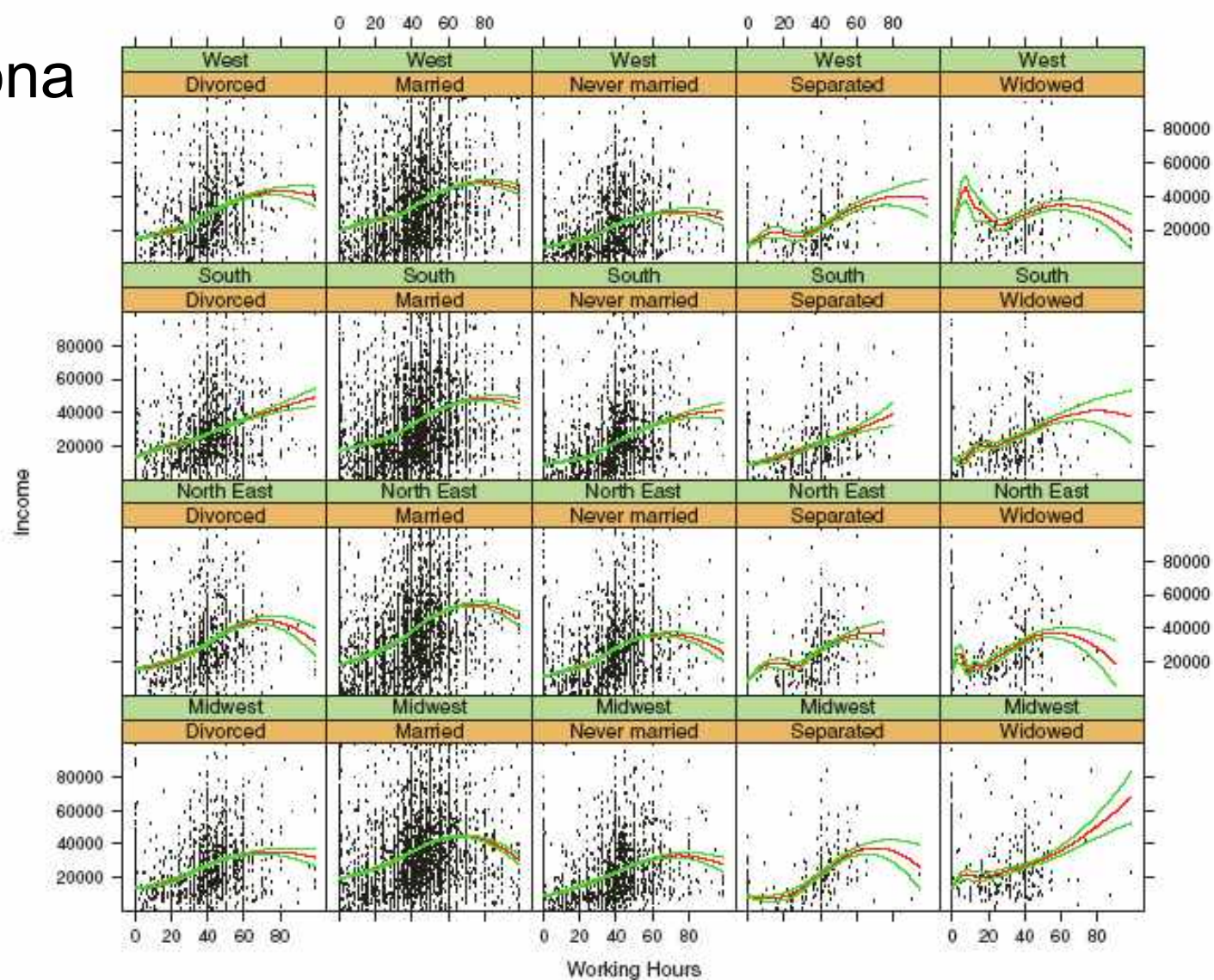


Figure 6.10. The same trellis display as in Fig. 6.5 with an additional lowess smoother superimposed



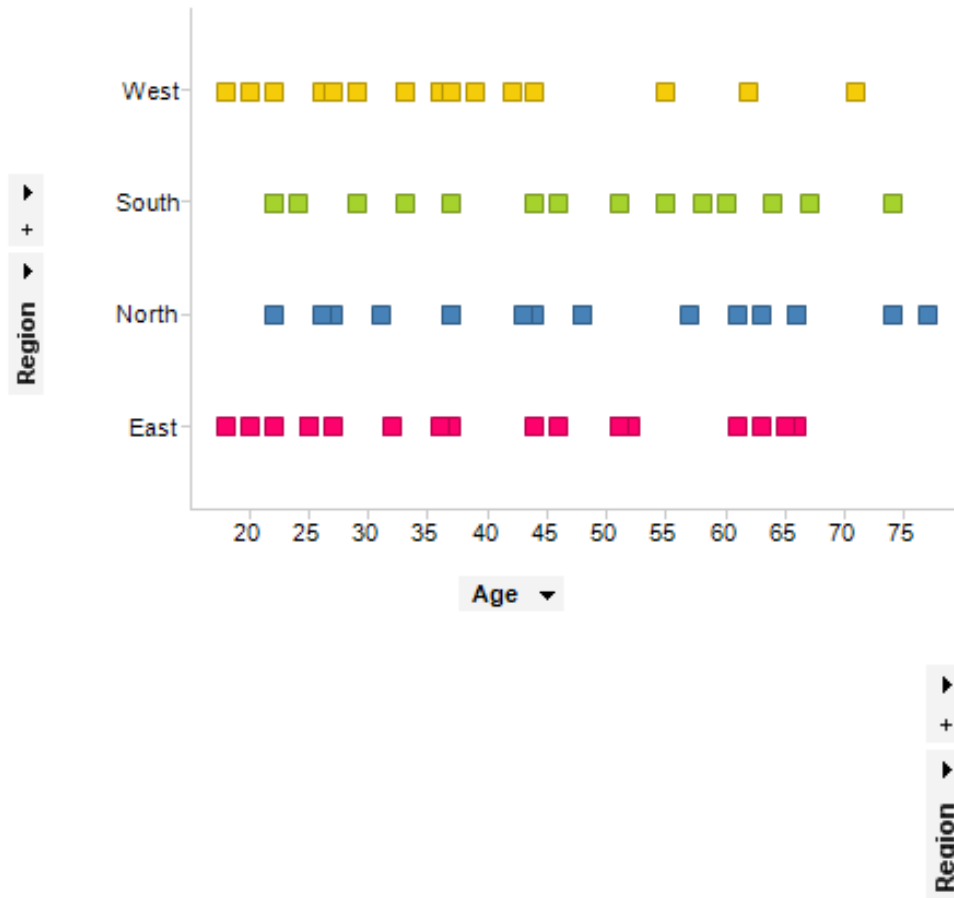
# Bardziej złożona kratownica



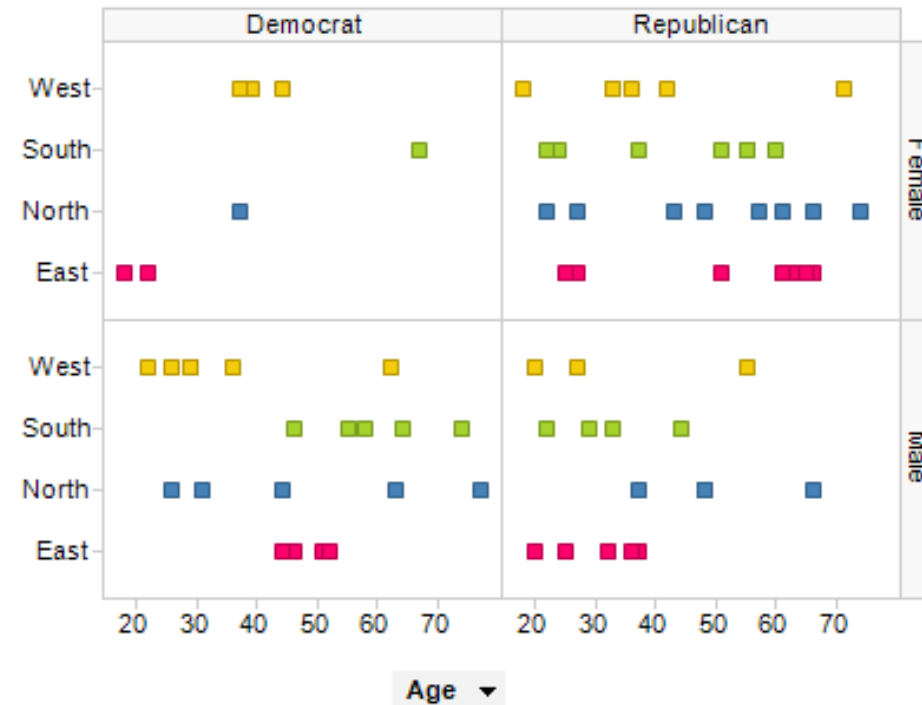
Census data – większa liczba obserwacji

**Fig. 2.19.** A trellis display showing scatterplots for Income vs. Working Hours conditioned by Region and Marital Status for the Census data. Each plot panel has a local regression smoother superimposed. Confidence bands have been added to illustrate the variability of the estimate.

# Trellis Displays – Inny rodzaj danych



## Dane wyborcze USA



# Tableau cross mini tabs

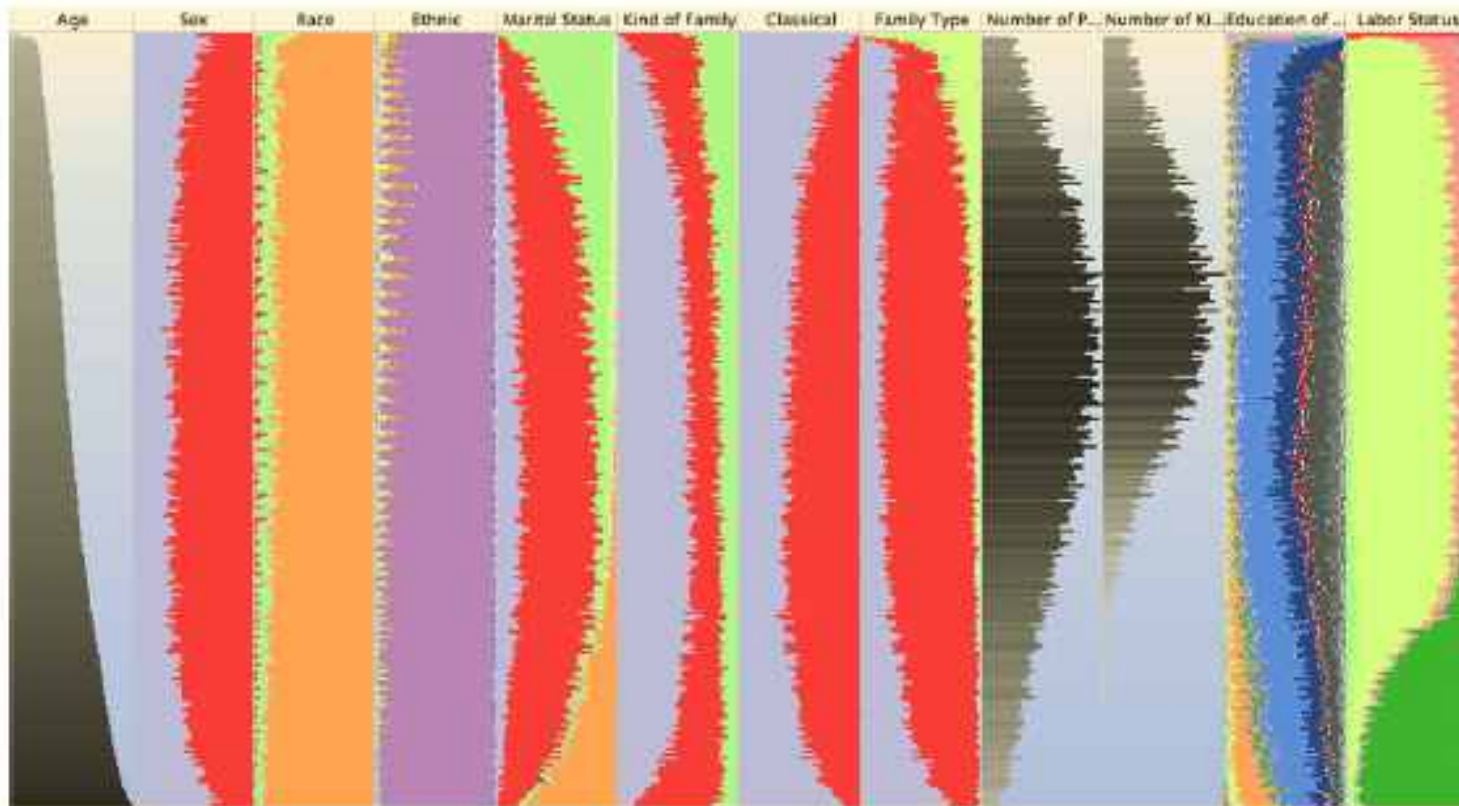
---



Figure 1: Example of a crosstab arrangement of small multiples, created with Tableau Software.

# Obrazowanie wielowymiarowej struktury - Tableplots

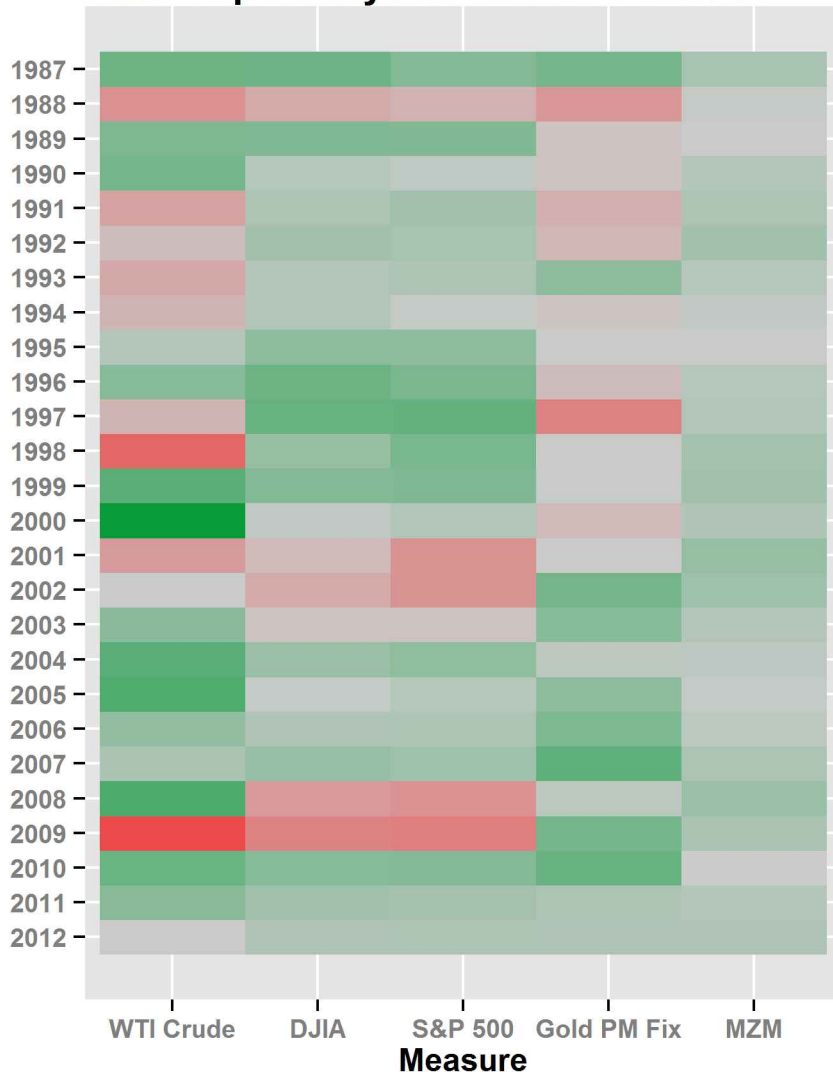
- Cel: zwizualizować złożoną informację danych wielowymiarowych w postaci pojedynczego wykresu odpowiadającego pomysłowi “arkusza”
- Przykład UC Census (12 zmiennych, wiersze 250 zagregowanych wartości uporządkowanych względem wieku, zawierających pasek o proporcjonalnej wielkości



**Fig. 2.25.** A tableplot of the US Census dataset for the first 12 variables.

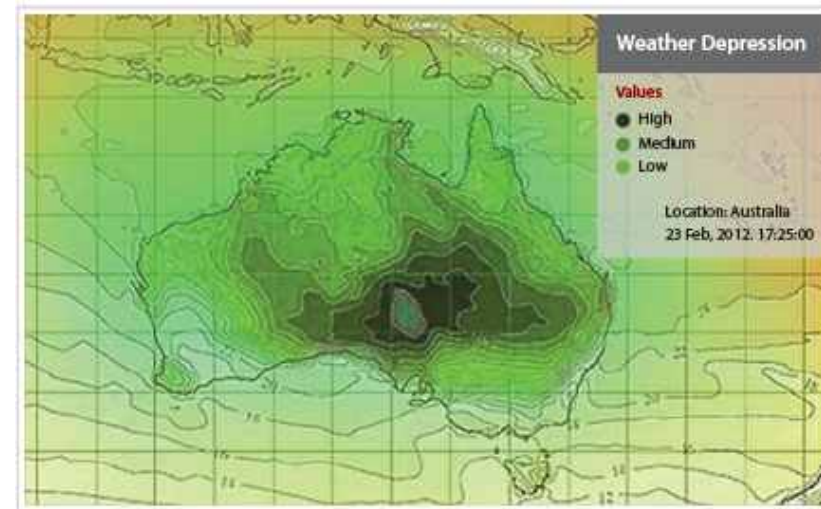
# Mapy ciepła - heatmaps

## Heat Map of Key Economic Measures



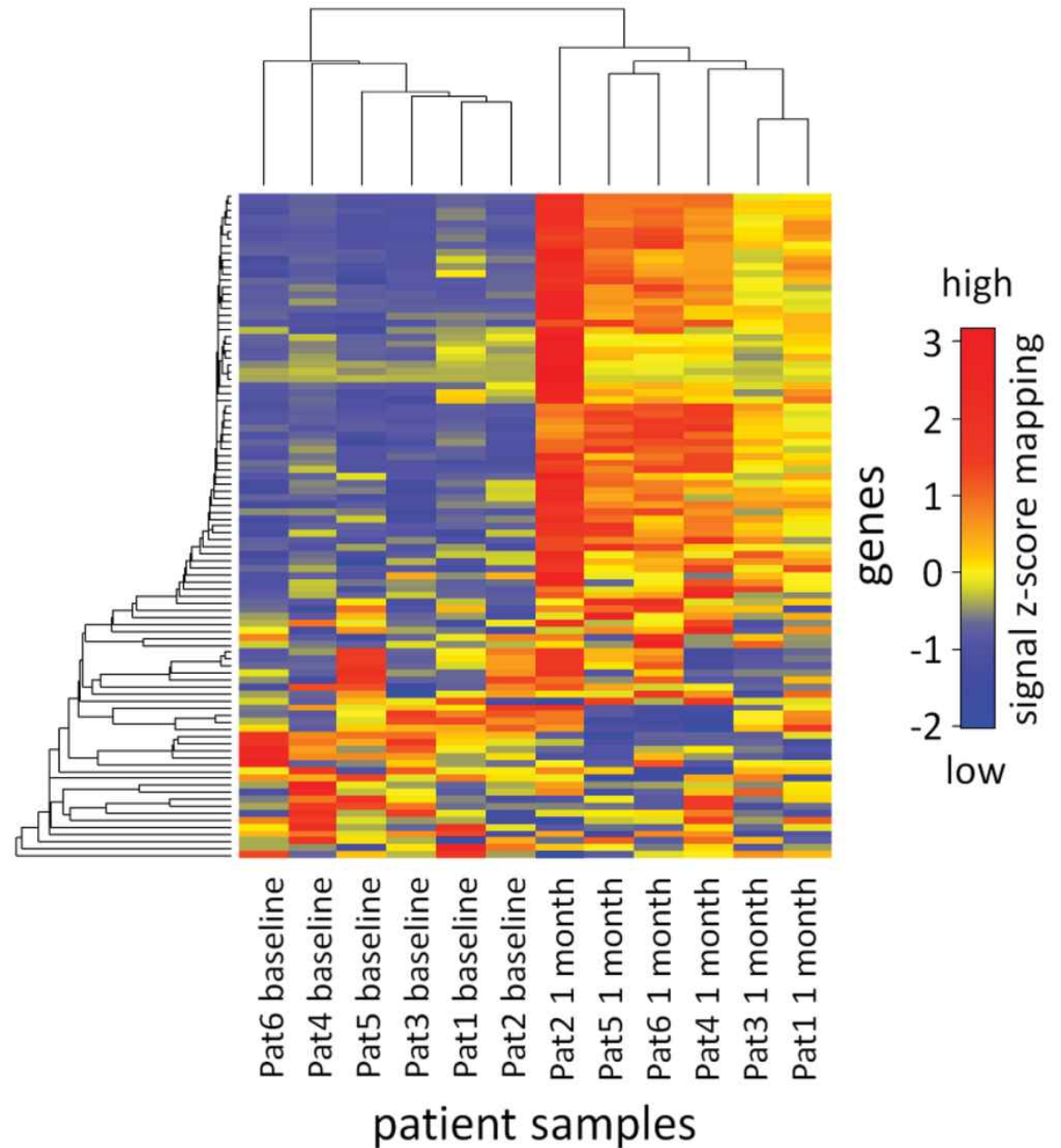
“A heat map is a two-dimensional representation of data in which values are represented by colors”  
Inspiracje – kartografia i mapy

Change in index or price



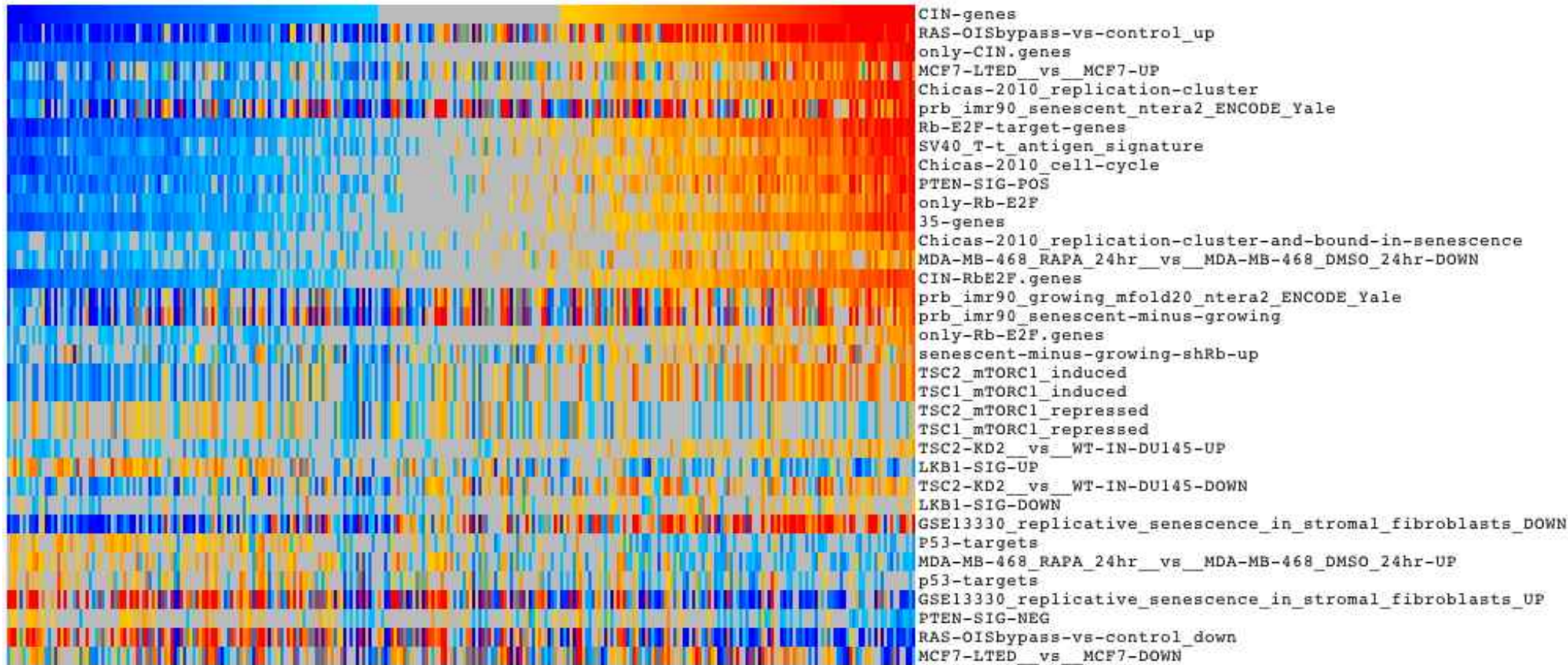
## Mapy ciepła - heatmaps

Przykład analizy danych  
biomedycznych –  
mikromacierze ekspresji  
genów

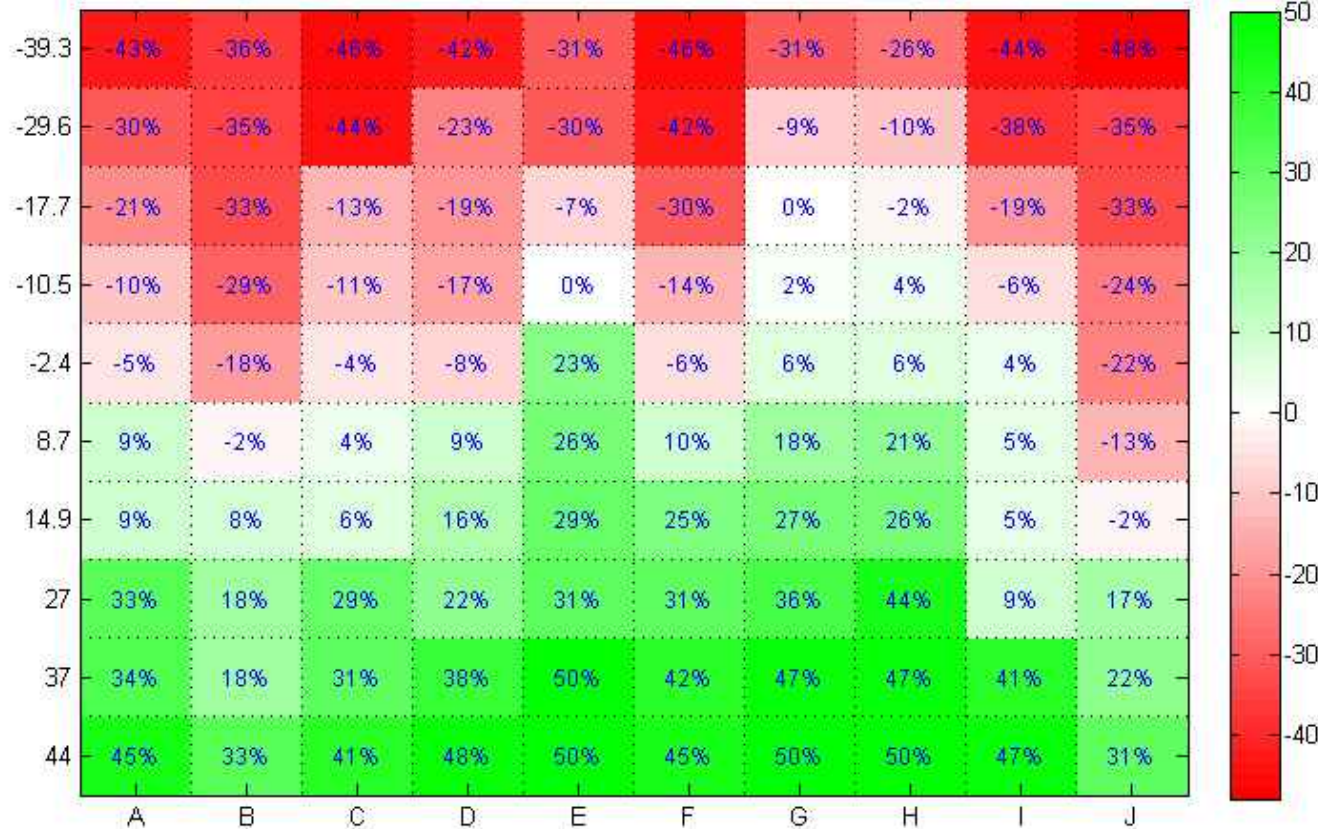


# Genetyczne zastosowania

---



# Mapy ciepła - heatmaps



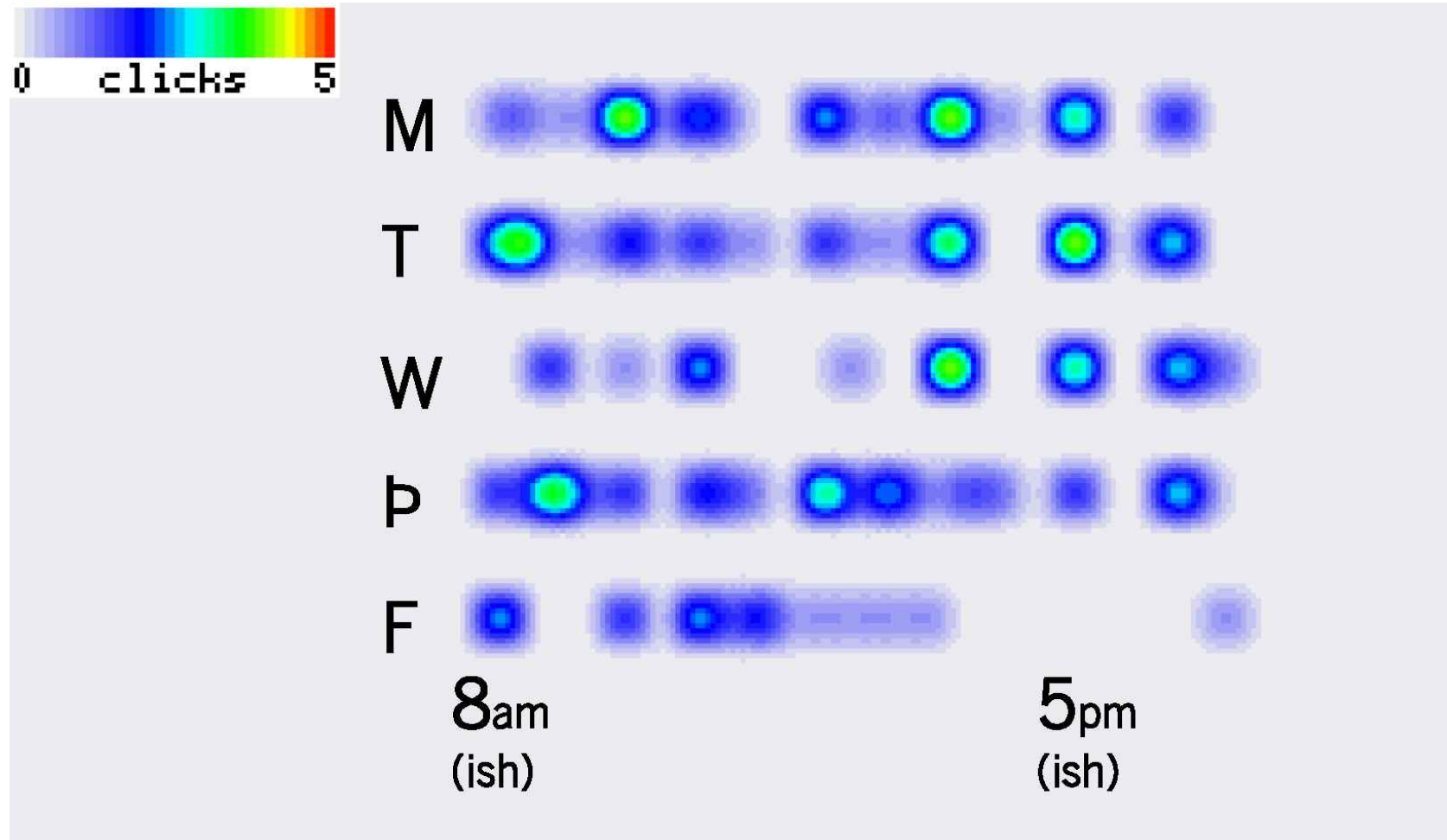
Team	WAR	Batting	Fielding	BsR
BOS	29.4	68.8	22.5	-6.7
LAD	28.0	62.6	22.6	-3.6
STL	27.5	48.5	31.0	-2.8
WAS	27.3	60.4	12.7	1.8
LAA	27.2	74.7	-8.9	-1.8
TOR	27.0	50.0	12.8	-0.2
BAL	25.4	17.3	31.4	-1.2
DET	24.8	46.8	-1.7	-2.7
PIT	24.8	53.4	-5.3	4.1
CLE	24.0	40.6	-8.4	2.6
OAK	24.0	18.0	14.1	2.3
SEA	23.3	31.2	-0.7	-1.9
KC	23.2	-18.8	38.6	7.3
MIL	22.8	29.6	7.8	-2.9
NYN	22.8	-6.3	29.7	0.2
TB	22.2	6.3	12.0	0.4
HOU	21.3	25.9	-16.8	1.5
CHC	21.1	36.2	-15.0	-2.5
NYM	20.7	28.1	-13.2	0.3
SF	20.7	13.6	3.8	-2.4
COL	20.3	5.7	8.4	-2.9
MIA	19.8	5.6	0.0	0.9
CIN	19.4	-18.6	21.3	0.1
TEX	18.7	-9.9	-1.2	-2.2
ARZ	17.8	-18.5	6.8	0.2
CWS	17.8	-0.7	-18.7	-2.3
MIN	17.0	-5.3	-23.4	-0.4
SD	16.8	4.8	-25.8	0.4
ATL	13.4	-49.0	0.1	-3.2
PHI	11.4	-55.2	-14.0	-0.3

Ostrożność w powiązaniu z liczbami – pilnuj czytelności



# Nietypowe heatmaps

---



# Nietypowe heatmaps – Tableau software

---

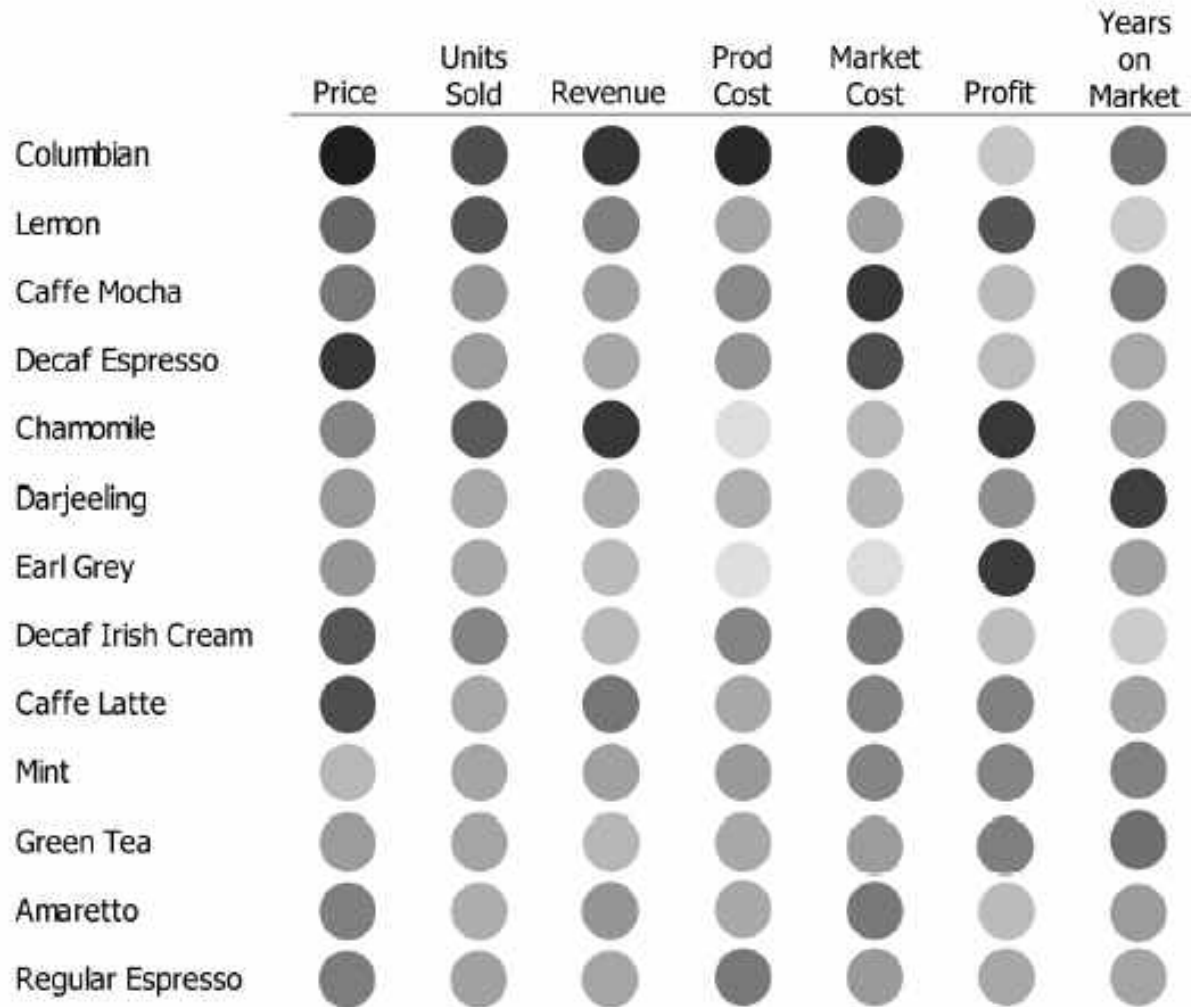


Figure 4: Example of a multivariate heatmap matrix.

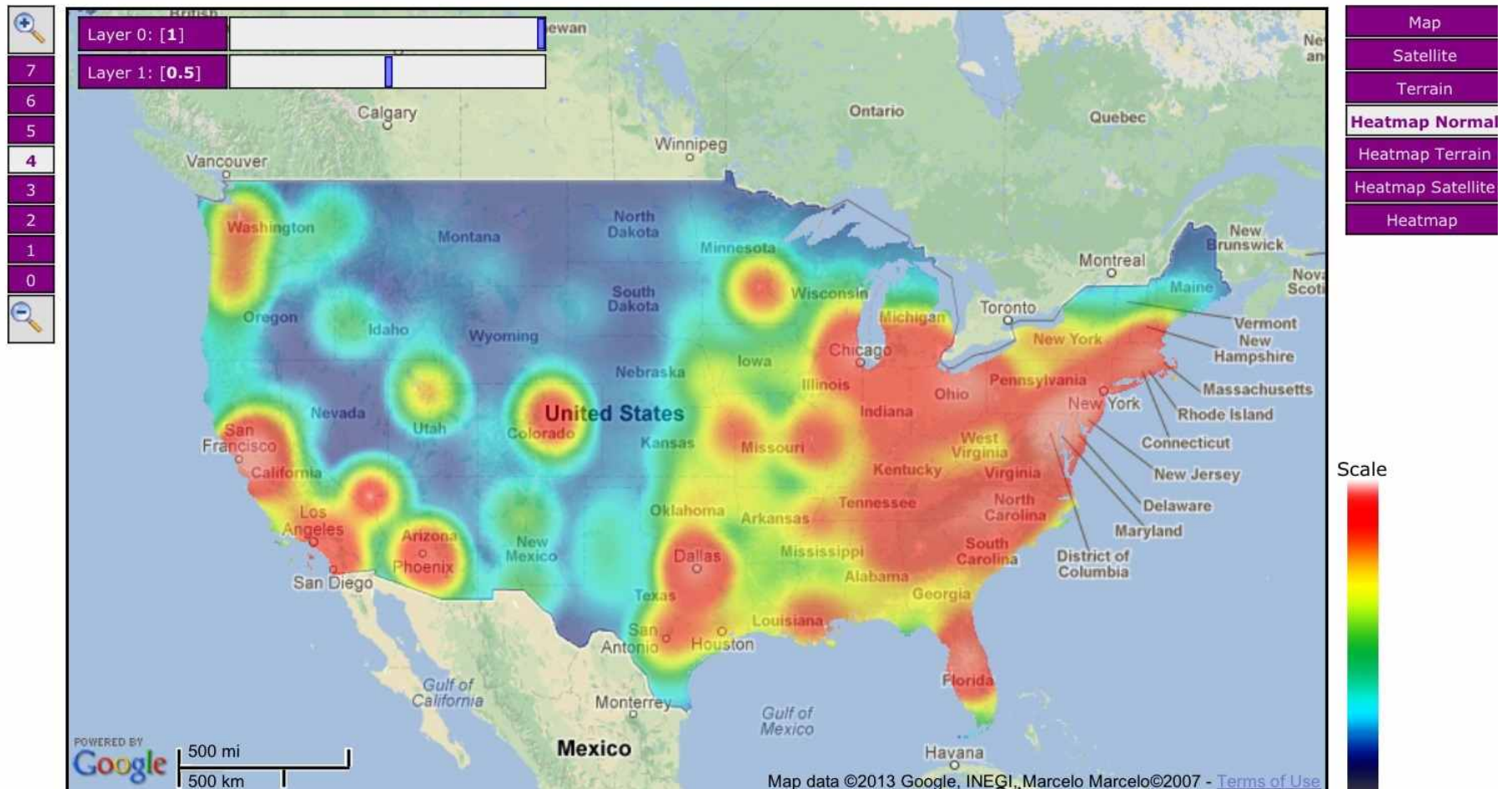
# Powrót do inspiracji kartograficznych

[Home \(More maps\)](#) ::: [Send me a message](#)

Google Maps - Heatmap - Density of fast food restaurants in the 48 contiguous states

(made with my own Perl script - Only zoom 4 to 7 for now.)

The same data can be seen as markers in the [Server Side Clusterer demo](#)



---

Przechodzimy do wizualizacji danych jakościowych

wraz z zastosowaniami do eksploracji danych  
oraz uczenia maszynowego

# Polish contribution – prof. Ryszard Michalski

## Father of Machine Learning and rule induction



[Interests](#)

[Biosketch](#)

[Publications](#)

[Teaching](#)

[Research](#)

[Solving  
problems](#)

[Machine Learning  
and Inference  
Laboratory](#)

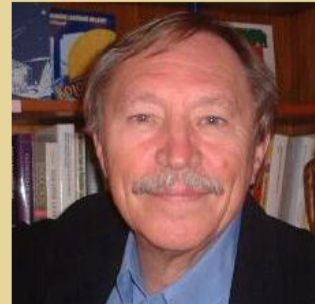
[School of  
Computational  
Sciences](#)

[George Mason  
University](#)

### Ryszard S. Michalski (1937 - 2007)

PRC Chaired Professor of Computational Sciences  
and Health Informatics  
Director of the Center for Discovery Science and Health  
Informatics

George Mason University



This page has been visited **15691** since January 1, 1999

6/27/06 [R.S. Michalski gives a banquet address at the International Conference on Machine Learning, to celebrate the return of the conference to Carnegie-Mell after 26 years since the very first conference was organized there by Carbonell, Michalski and Mitchell](#)

#### Articles in *Mason Gazette*:

7/31/07 [New Center to Help Investigators Discover New Knowledge in Medical Databases](#)

3/12/03 [University Wins 10th Patent for Machine Learning Invention](#)

11/19/02 [Spotlight on Research: Grants Support Machine Learning and Inference Research](#)

7/27/00 [Michalski Receives Prestigious Science Honor](#)

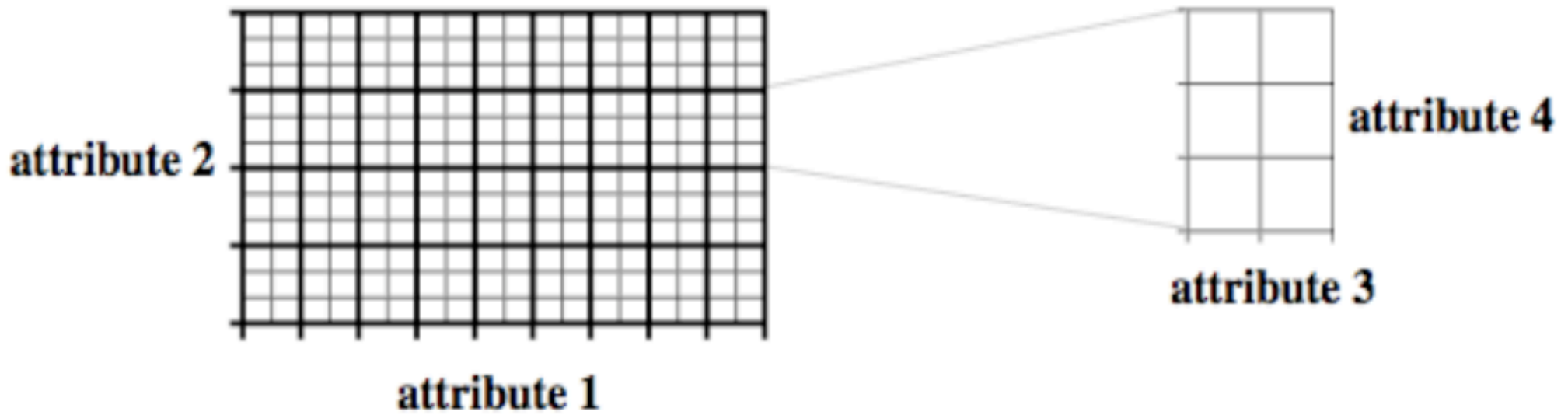
### *Interests*

#### Research areas:

[Machine Learning](#), [Data Mining and Knowledge Discovery](#), [Inductive Databases and Knowledge Scouts](#), [Non-Darwinian Evolutionary Computation](#) and [Plausibility applications of these areas to Bioinformatics, Medicine, User Modeling, Intrusion Detection, and Very Complex System Design.](#)

# Multidimensional Stacking

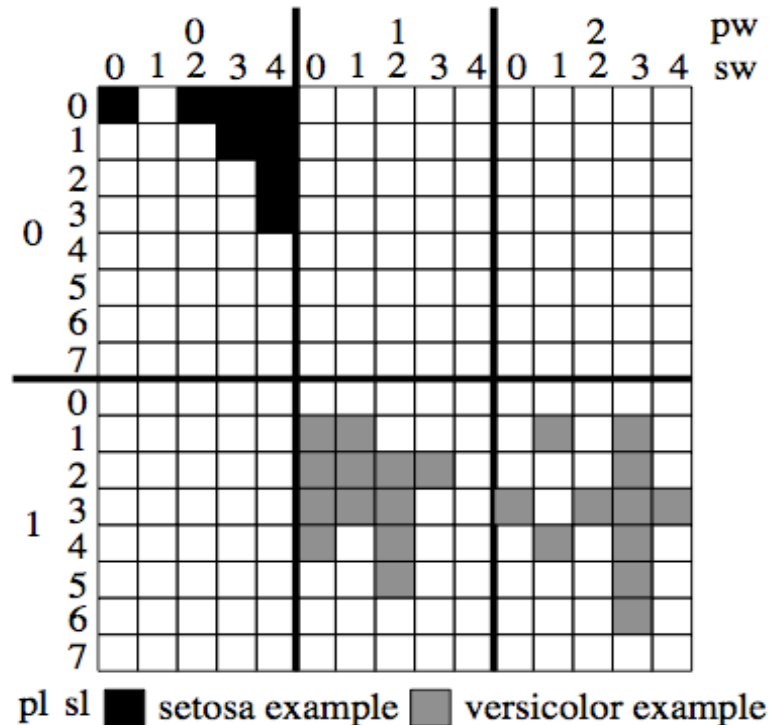
---



# VL – Generalized Logic Diagrams

VL1 diagrams (Michalski 70) for machine learning

## Visualization of Training Examples: Discrete Version of the Iris Data Set



		Size		
		S	M	L
Green	T			
	C			
	R			
Blue	T			
	C			
	R			
Red	T			
	C			
	R			

Color Shape  
Target  
concept,  
 $t = 1 \dots 40.$

		Size		
		S	M	L
Green	T			
	C			
	R			
Blue	T			
	C			
	R			
Red	T			
	C			
	R			

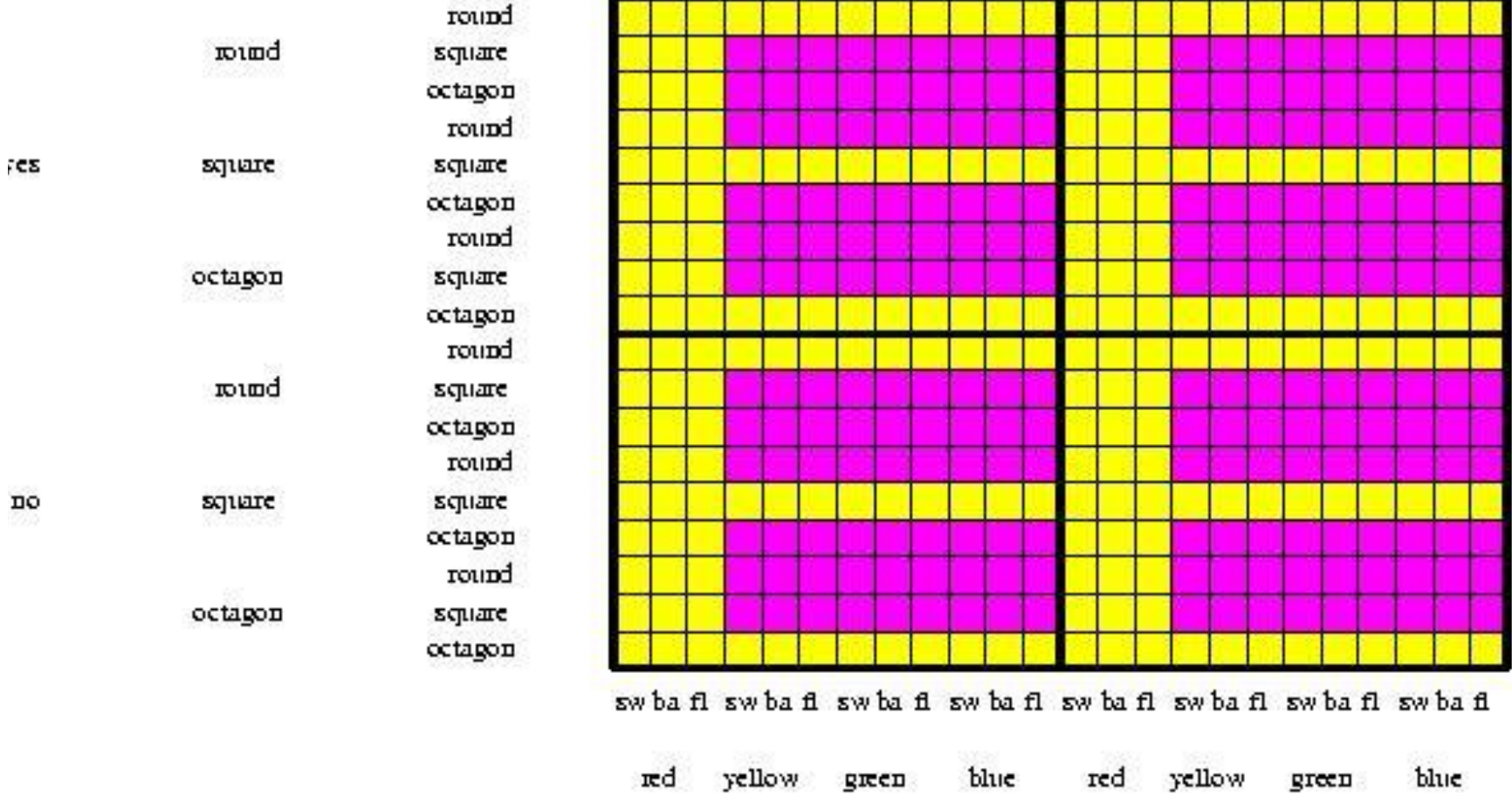
Color Shape  
Target  
concept,  
 $t = 41 \dots 80.$

		Size		
		S	M	L
Green	T			
	C			
	R			
Blue	T			
	C			
	R			
Red	T			
	C			
	R			

Color Shape  
Target  
concept,  $t =$   
 $81 \dots 120.$

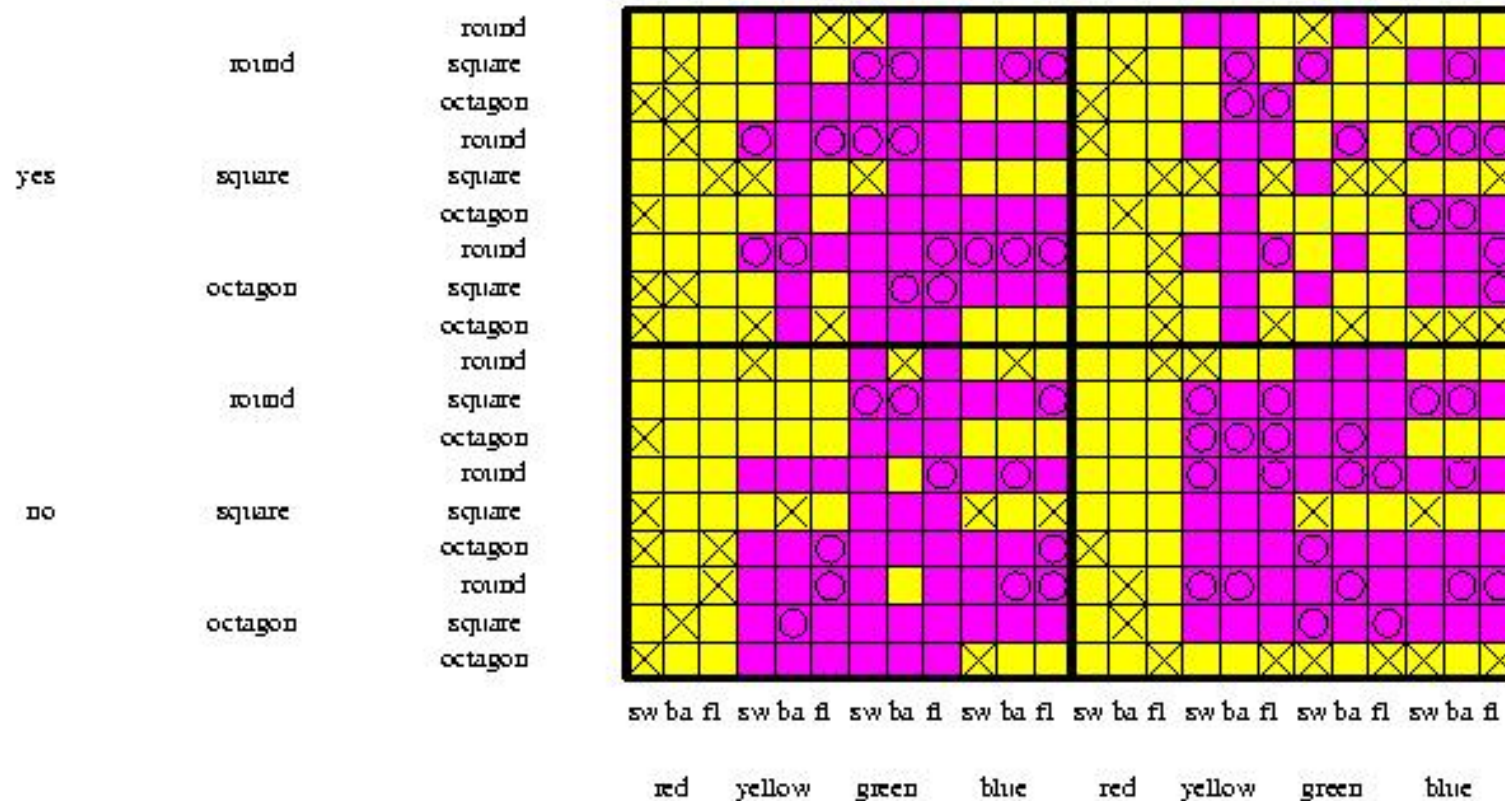
STAGGER and concept drift

# Przykład problemu MONK 1 – oznaczniki pokryć





# VL – oznaczenie przykładów testowych



Poprawne i błędne decyzje klasyfikacji przykładów  
wybrany algorytmem klasyfikacyjnym

# Reguły i pokrywane przykłady

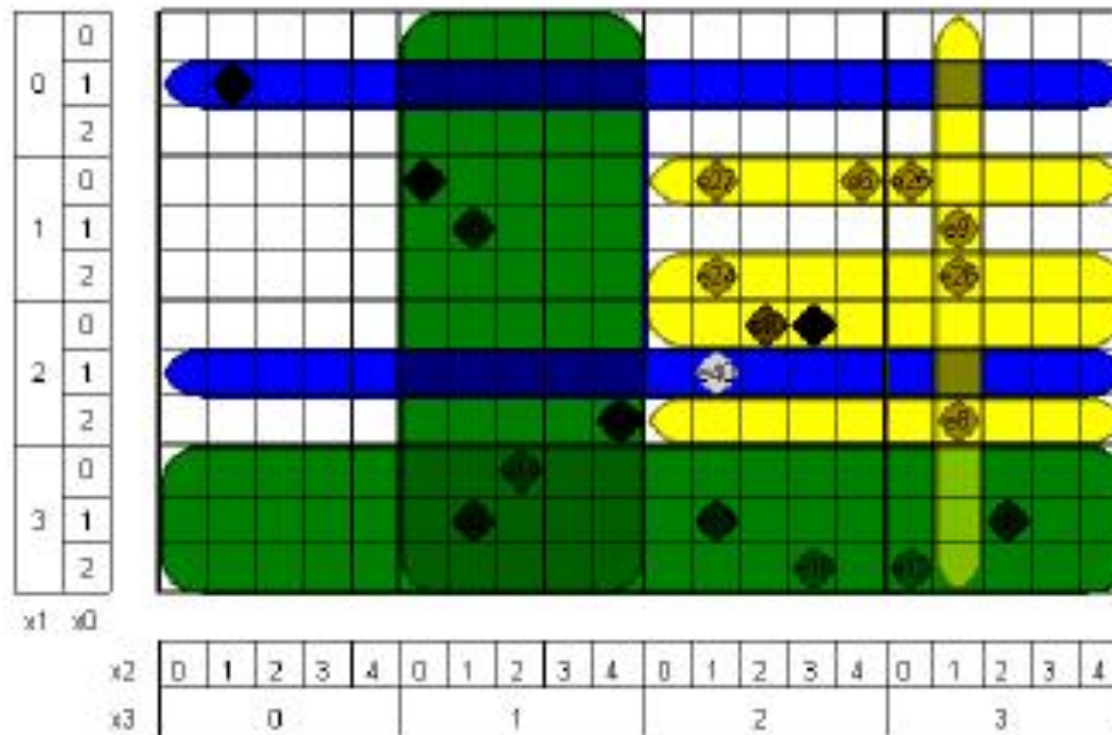
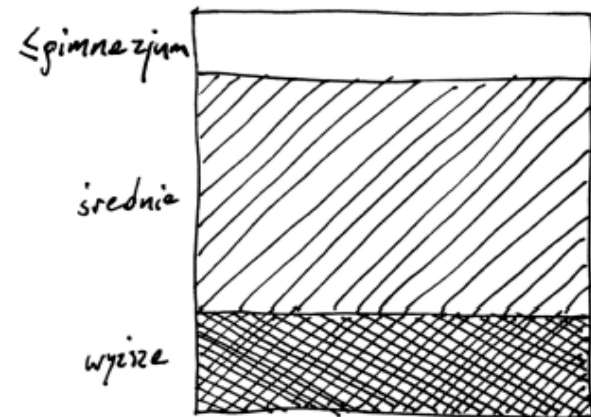


Fig. 2. GLD for rules generated from robots data

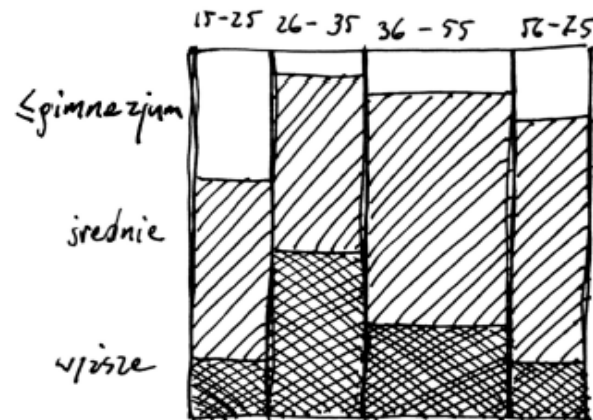
Za artykułem B.Snieżyński I R,Michalski

# Od wykresów słupkowych do kafelkowych

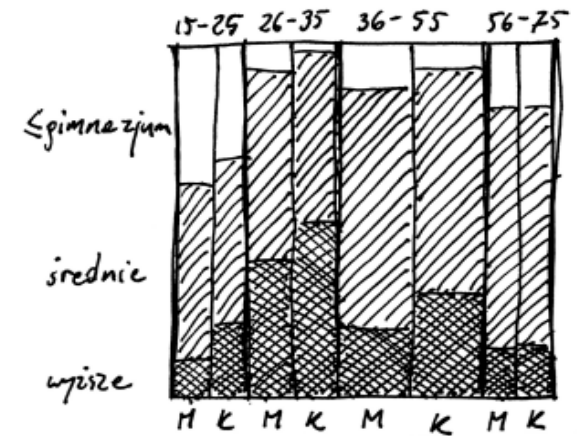
- Wykresy kafelkowe (mozaikowe) - zwiększenie precyzji elementów większej całości (także hierarchii)
- Hierarchia elementów → rozwinięcie do TreeMaps



Udział osób z wykształceniem wśród mieszkańców Polski Mających powyżej 15 lat



Udział osób z wykształceniem z rozbiciem na grupy wiekowe



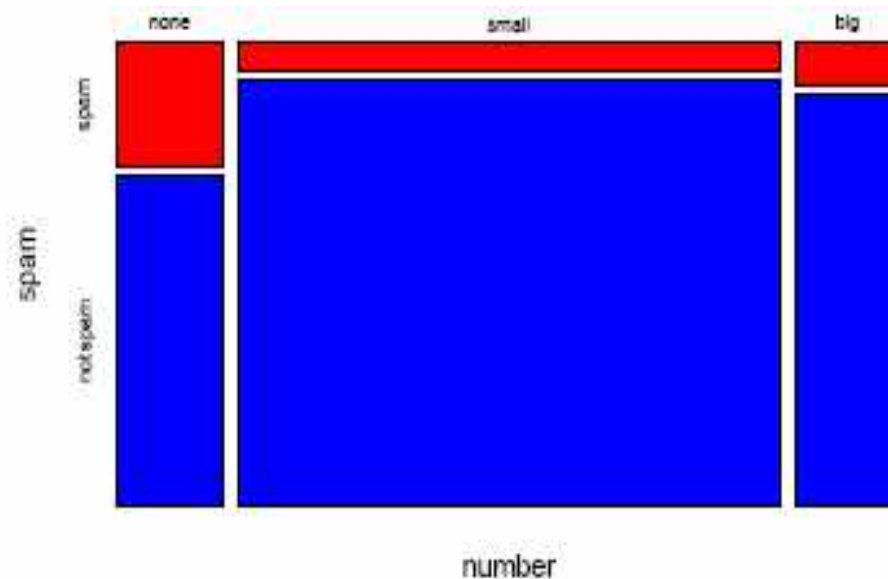
Udział osób z wykształc. z rozbiciem na grupy wieku oraz płeć

Wykresy kafelkowe → **Mosaic plots** --- przedyskutujemy dokładniej

# Mosaic plots

---

- Mozaika czy kafelek? – wywodzą się od Hartigan & Kleiner 1981 ;  
Później wiele różnych odmian
  - Dostępne w wielu programach (np. R – mosaic plot)
- Pokazanie współzależności zmiennych jakościowych
  - “In statistical graphics, the mosaic display is a graphical method to show values (cell frequency) in a contingency table cross-classified in one or more factors” (Michael Friendly 2001)
- Wymagają dużego doświadczenia od analityka



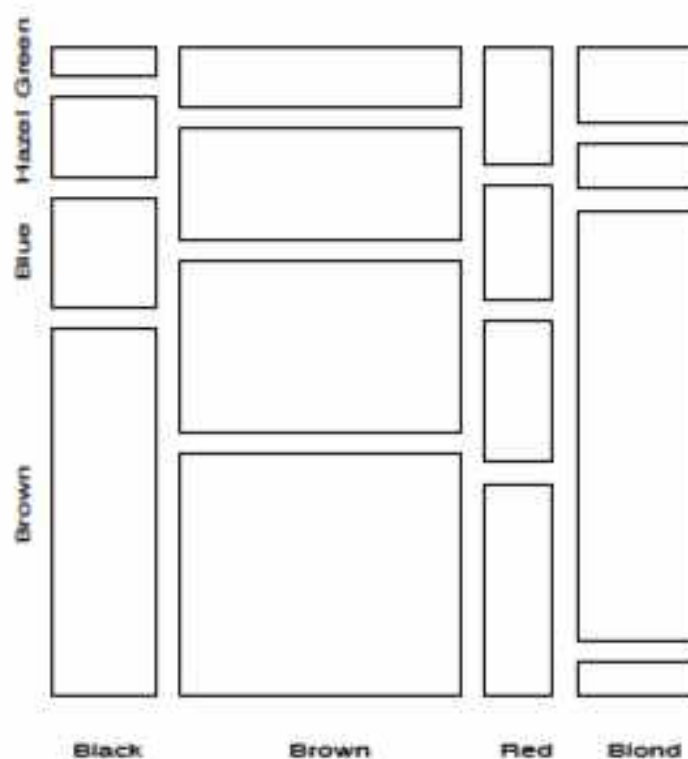
# Od tablicy dwudzielczej do wykresu mozaikowego

---

Przykład ilustracyjny

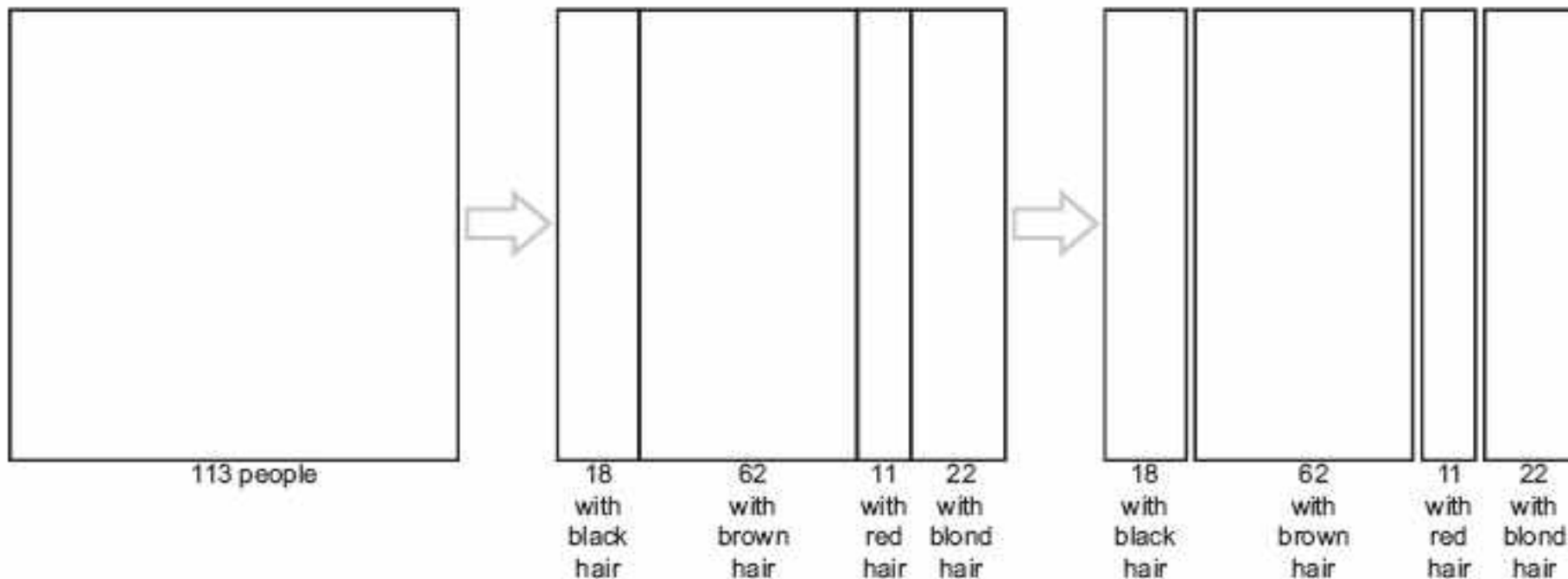
Eye Color	Hair Color				Total
	Black	Brown	Red	Blond	
Brown	10	25	4	2	41
Blue	4	18	3	15	40
Hazel	3	13	2	2	20
Green	1	6	2	3	12
Total	18	62	11	22	113

Wykres



# Etapy tworzenia wykresu mozaikowego

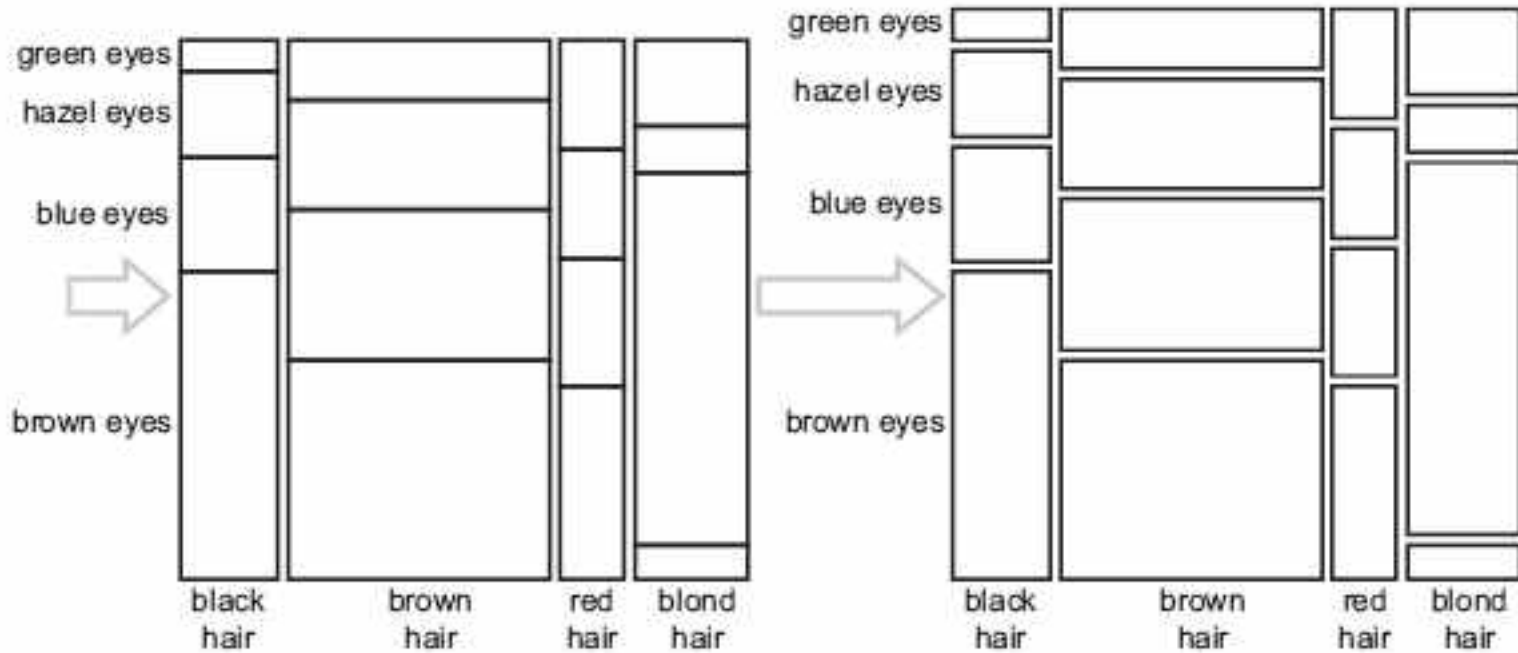
---



Wykorzystaj zmienną kolor włosów

# Etapy tworzenia wykresu mozaikowego

---

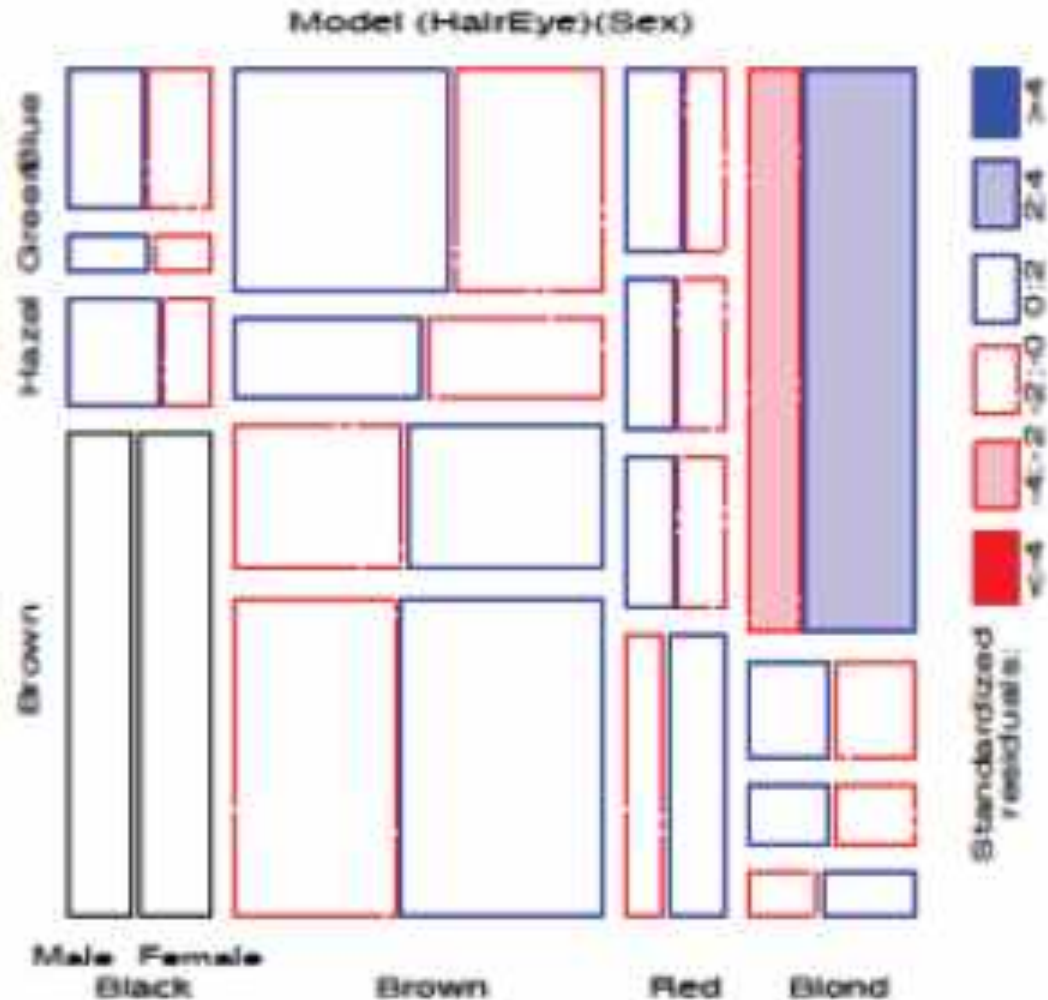


Uzłędnij zmianę kolor oczu i podziel słupki

# Możliwości wprowadzenia dodatkowych zmiennych

---

3 zmienna – płeć osób -> dodatkowa kategoryzacja obszarów



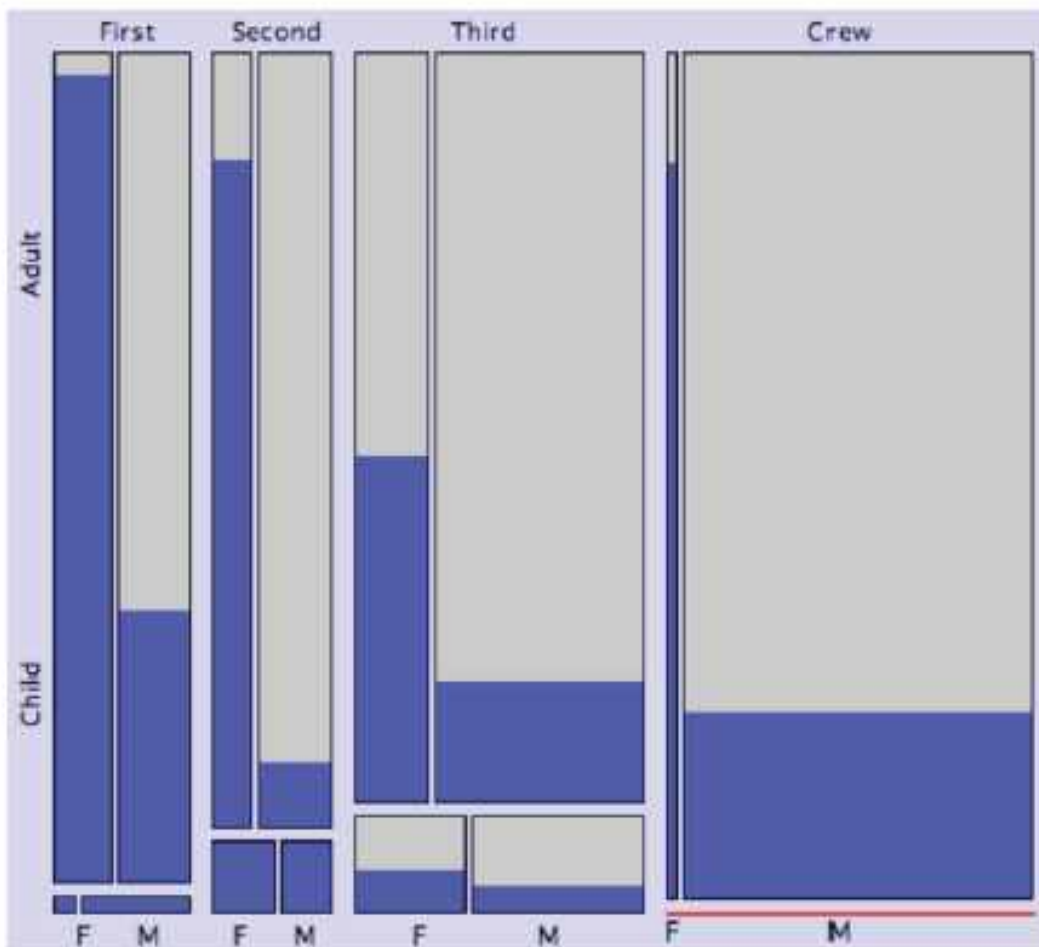


# Zalety i ograniczenia

---

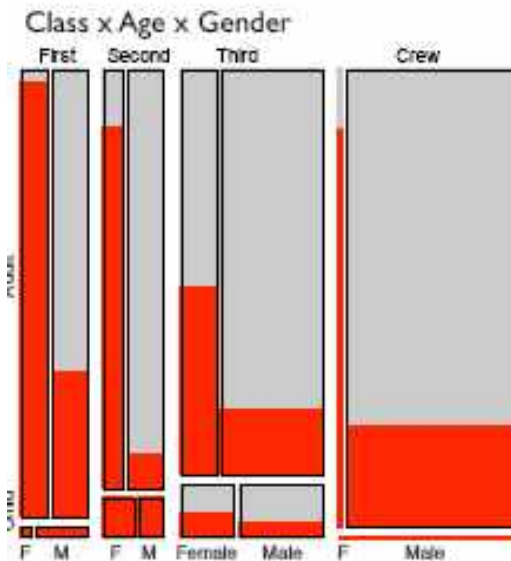
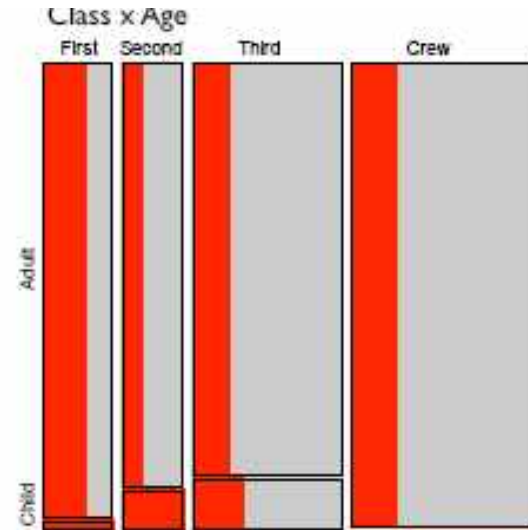
- “Mosaic plots and their variants open up the powerful visualization of multivariate categorical data” Martin Theus 2011
- Porównywanie prostokątów (i ich rozmiarów) może być trudne
- Problemy percepcji, gdy zmieniają się równocześnie obie zmienne
- Lecz (Richards)
  - The strength of this display is that we do not shift our perception. We compare proportions within each group, we study height only. When we compare counts between groups, we study only width.”
  - It is difficult to compare dimensions that are not arranged side by side along a common baseline (Few).
- Nie widzimy bezpośrednio wartości liczbowych
- Dobór kolejności zmiennych ma znaczenie dla czytelności

# Titanic – przykład dokładniejszej analizy



Przeżycie katastrofy w zależności od wielu czynników

# Kolejność tworzenia wykresu



# Titanic survivors

---

Różne pytanie do analizy

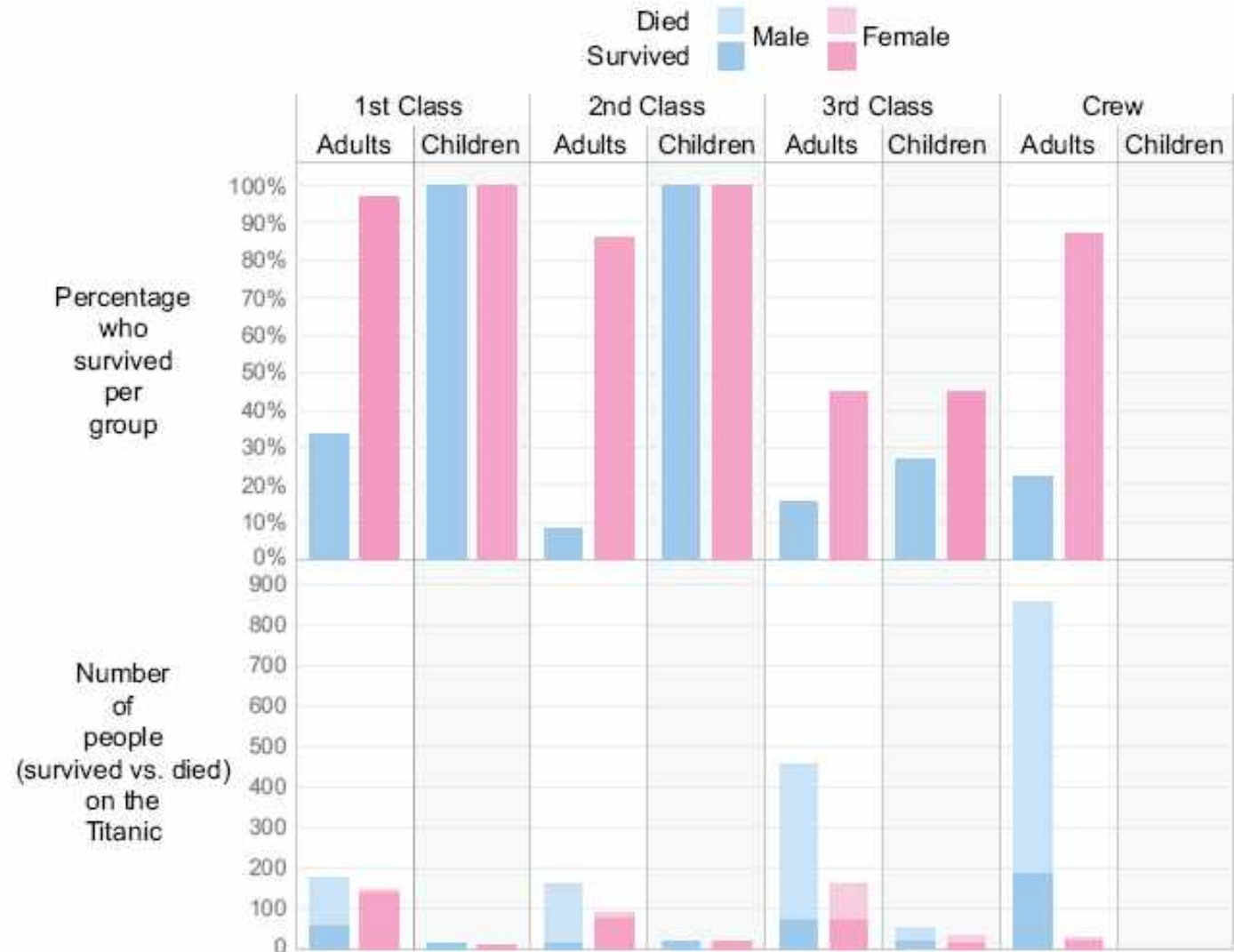
- Jaki jest związek między klasą podróży a śmiercią pasażera?
- Czy przeżyło więcej mężczyzn czy kobiet?
- Wśród załogi - czy przeżyło więcej mężczyzn czy kobiet?
- Jaki był związek między klasą a liczbą dzieci?

Pamiętaj, że

- Nie widzimy bezpośrednio wartości liczbowych

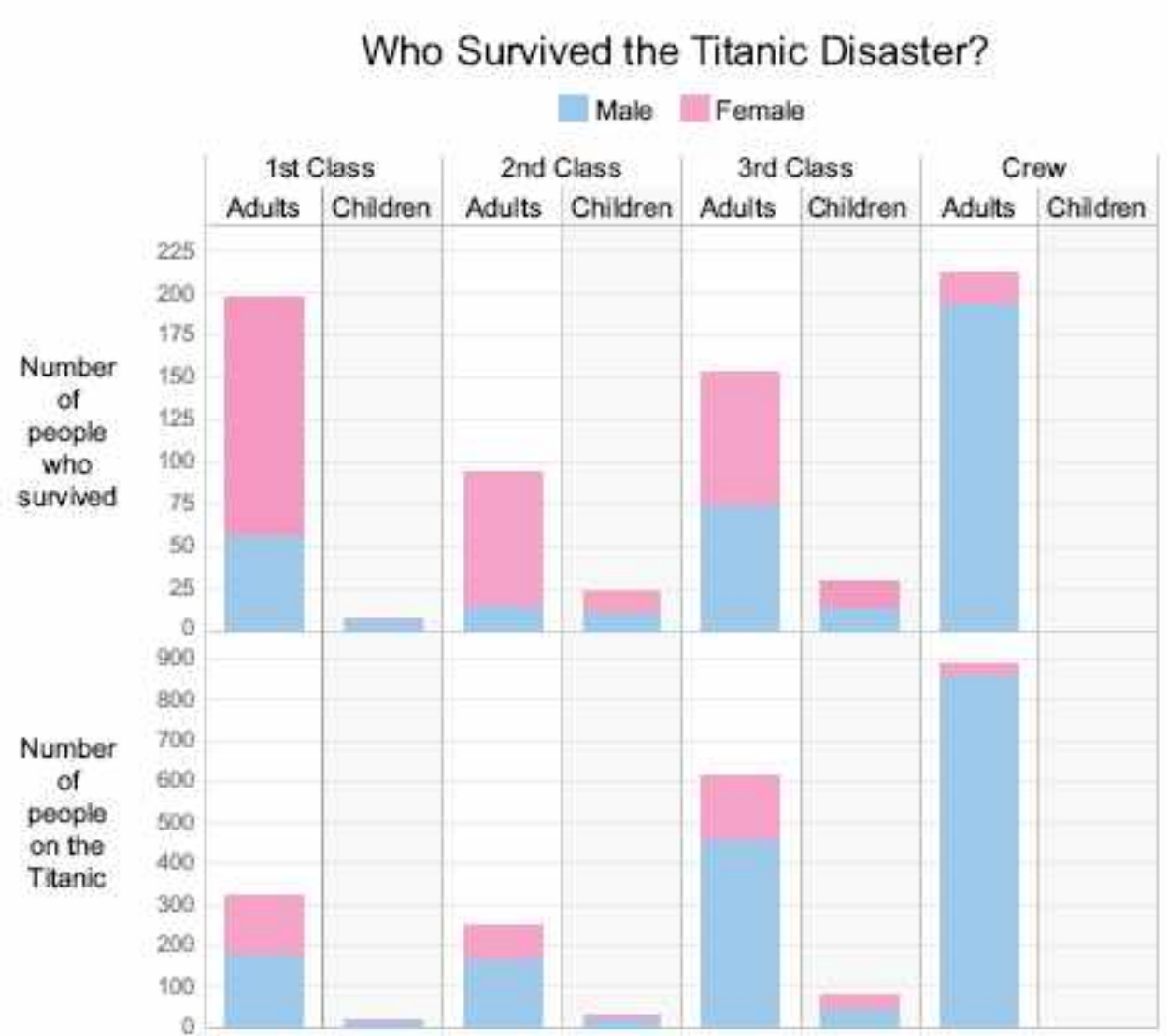
# Inna wizualizacja ocalonych

## Who Survived the Titanic Disaster?

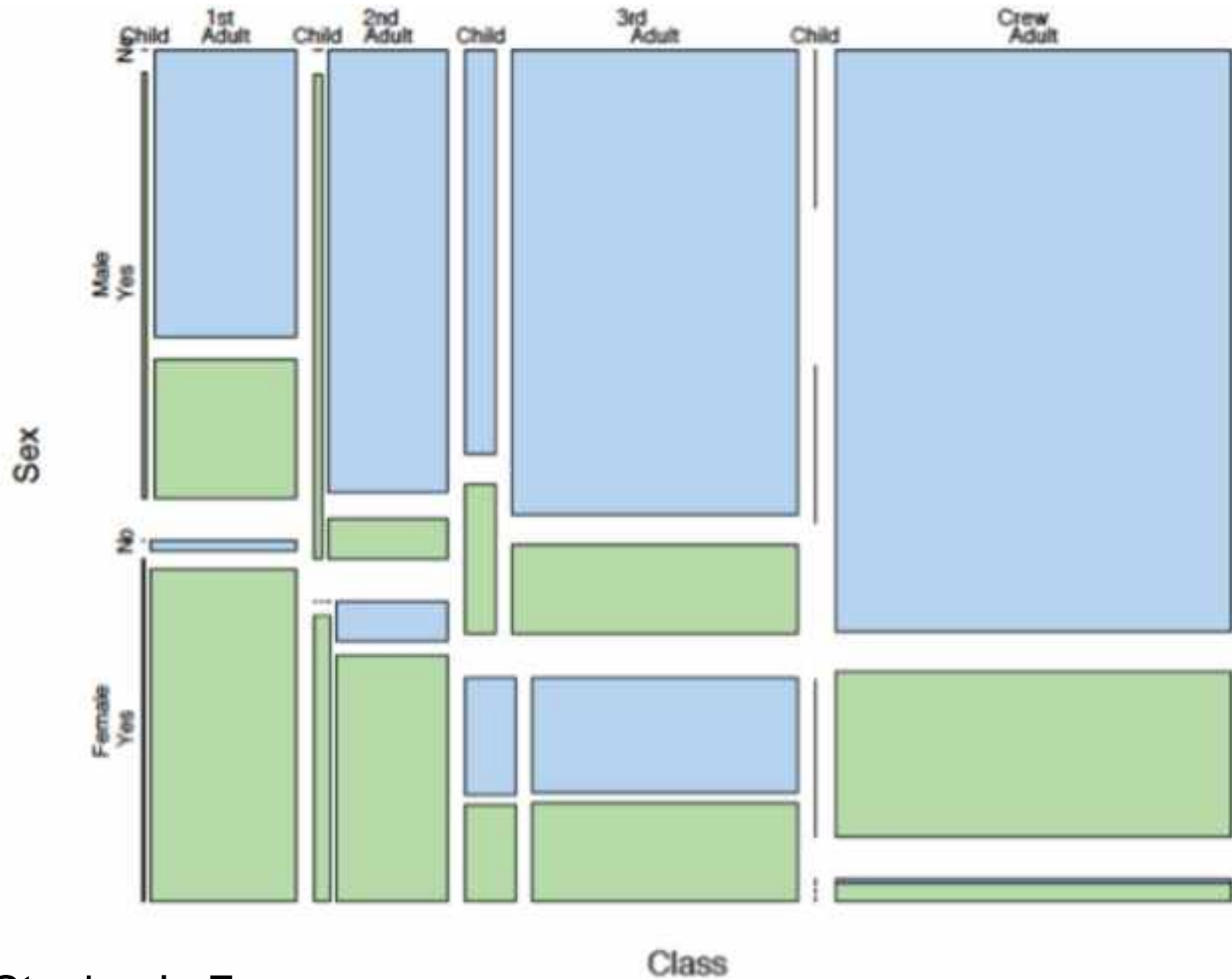


Przykład Stephen'a Few

# Dokładniejsza analiza liczb ocalnych



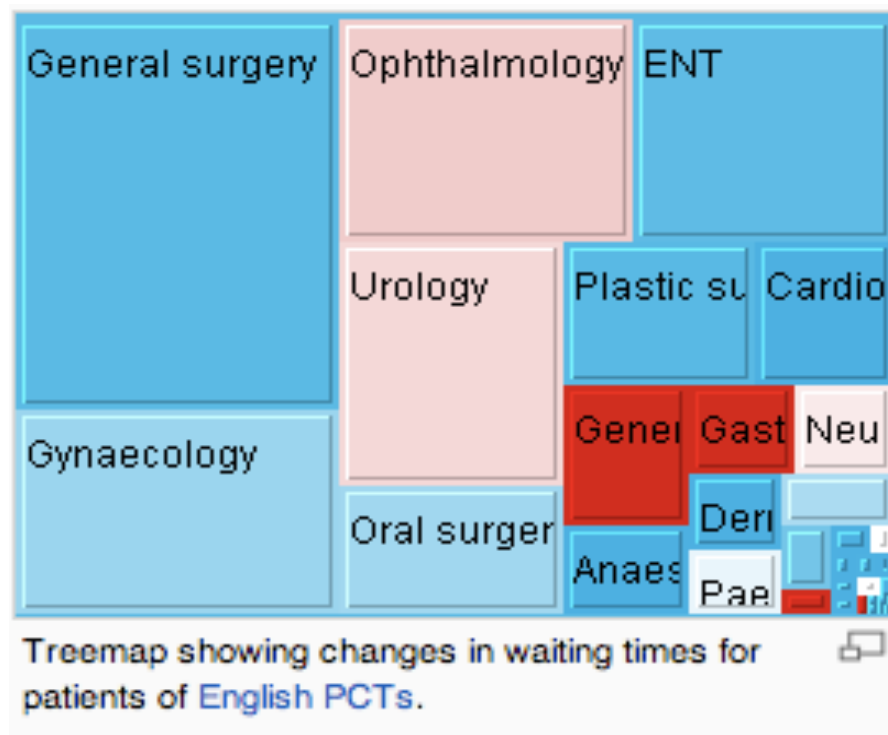
# Złe skategoryzowanie - wizualizacja ocalonych



Przykład Stephen'a Few

# Hierarchiczne wizualizacje – Treemaps (Johnson and Shneiderman 1991)

---



- Treemaps display hierarchical data using rectangles. Each branch of the tree is assigned a rectangle. Then each sub-branch gets assigned to a rectangle and this continues recursively until a leaf node is found.
- A leaf node's rectangle has an area proportional to a specified dimension (cardinality) on the data. (Illustration -- proportional to a waiting time)
- Depending on choice the rectangle representing the leaf node is colored, sized or both according to chosen attributes.



# Treemaps – zasady budowy wizualizacji

---

Rekurencyjny podział przestrzeni w odniesieniu do binarnego drzewa:

- Najpierw podziel obszar horyzontalnie zgodnie z proporcją przypadków w gałęziach drzewa.
- W następnym kroku dziel “wertykalnie” wg. proporcji wśród węzłów potomnych
- Proces powtarzaj, aż dojdiesz do węzłów końcowych drzewa

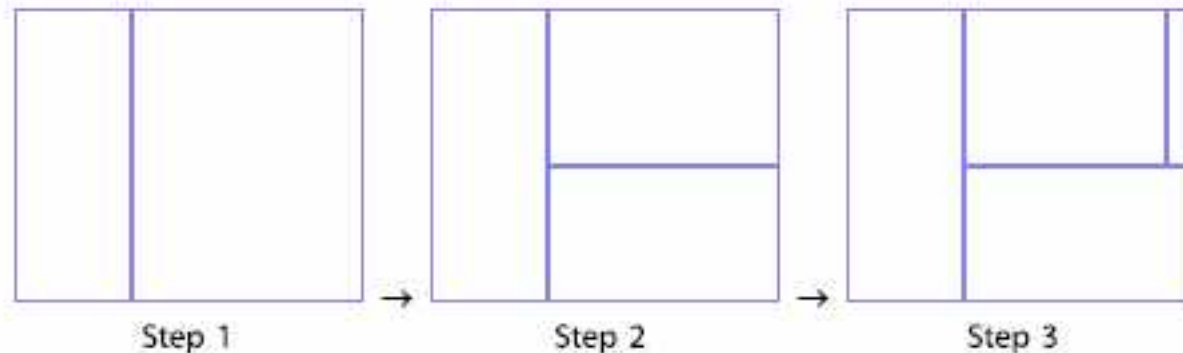


Figure 10.5. Construction of a treemap consisting of three subsequent binary splits

# Treemaps – przykład hierarchicznego drzewa

---

Rekurencyjny podział przestrzeni w odniesieniu do binarnego drzewa:

- Najpierw podziel obszar horyzontalnie zgodnie z proporcją przypadków w gałęziach drzewa
- W następnym kroku dziel “wertykalnie” wg. proporcji wśród węzłów potomnych
- Proces powtarzaj, aż dojdiesz do węzłów końcowych drzewa

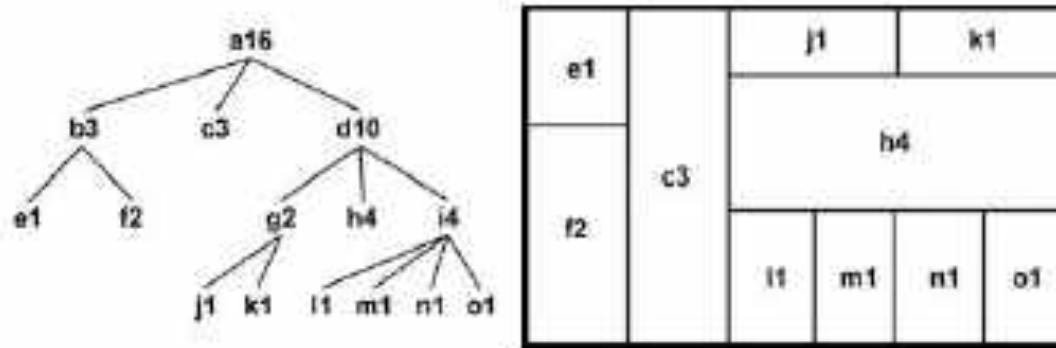


Figure 2: Tree diagram and corresponding Treemap

# Treemaps – złożona hierarchia kształtów

---

Rekurencyjny podział przestrzeni w odniesieniu do binarnego drzewa  
Lecz mogą powstawać złożone zagnieżdżenia prostokątów

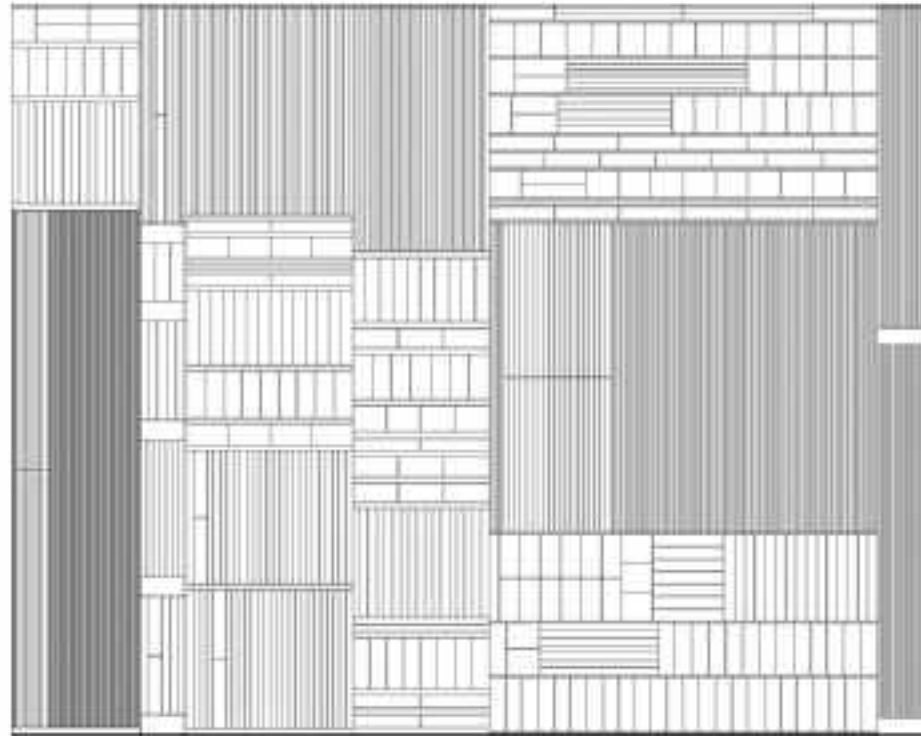


Figure 3: Aspect ratio problem of the original Slice-and-Dice Treemaps

# Hierarchiczne wizualizacje – Schneiderman 92

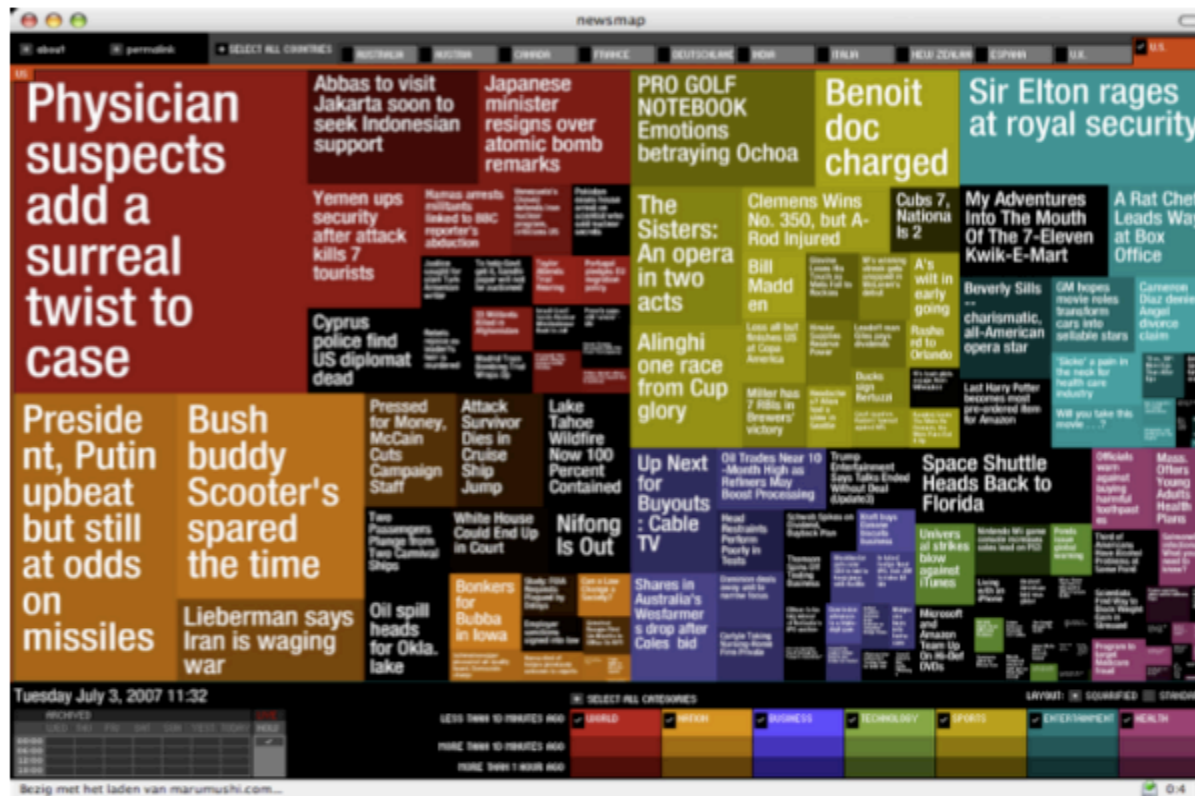


Figure 4: Treemap used for displaying news <http://newsmap.jp/>

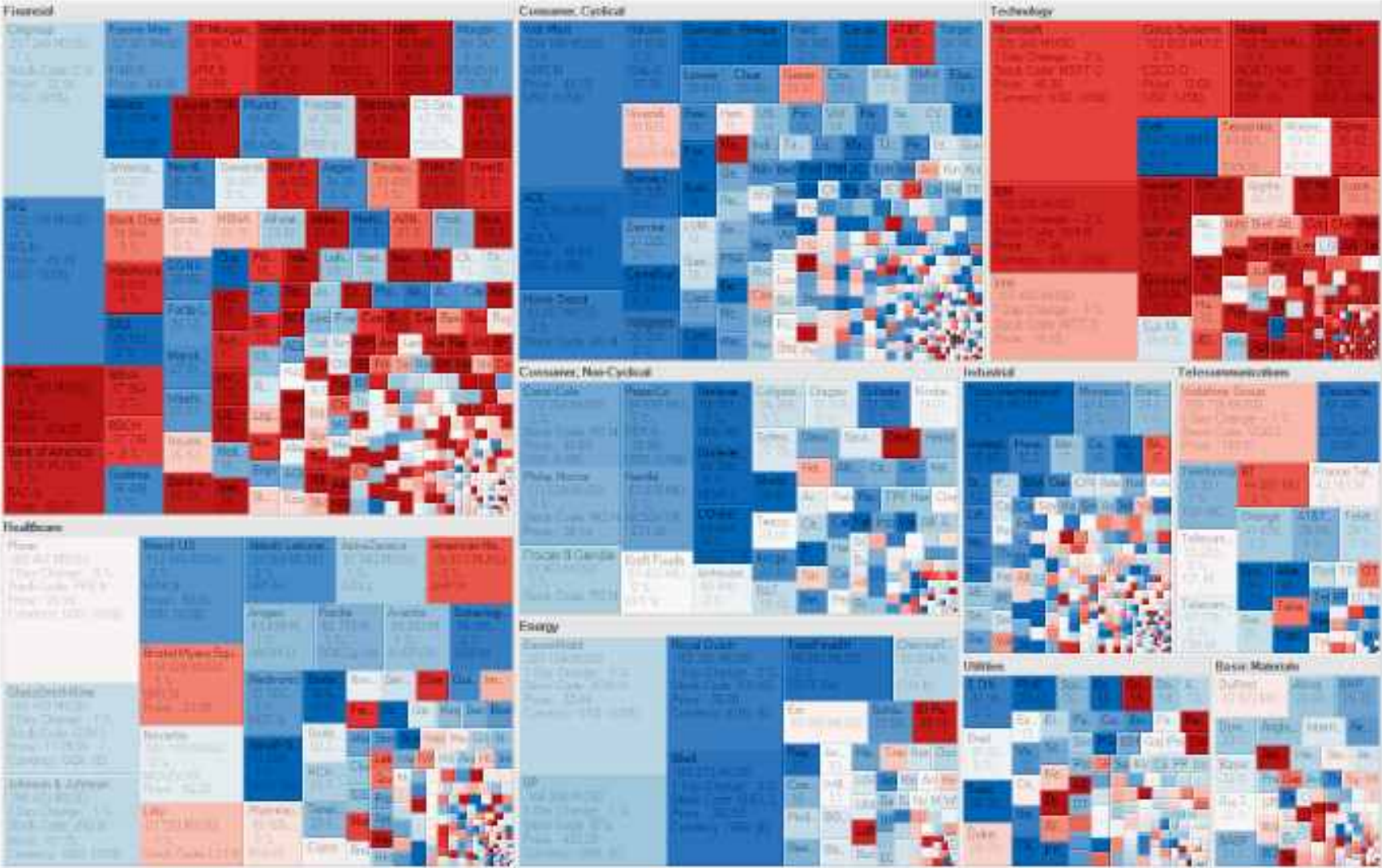
- Treemaps wykorzystywane w Google Tools
- Ilustracja wiadomości dostępnych w sieci

# Zagnieżdżanie kształtów w Treemaps



**Figure 1:** Squarified treemaps summarise the dataset by category through alternative hierarchies. (a) year»month; (b) type»year»month; (c) postcode»type. Local linear colour schemes are used (ColorBrewer "Oranges"; <http://www.colorbrewer.org/>) and ordering is by size from top left to bottom right. The house price data are Crown copyright (Land Registry) and were retrieved from <http://www.houseprices.co.uk/>.

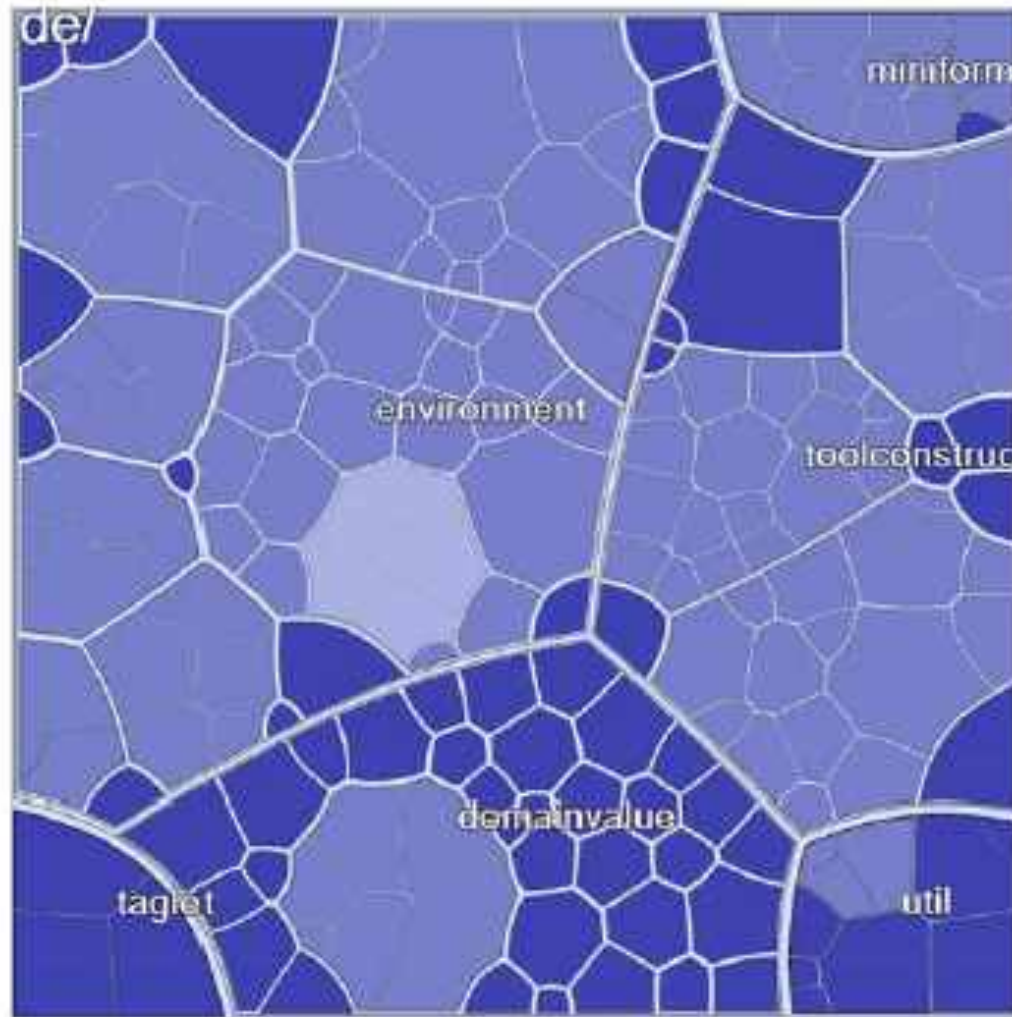
# Bardziej złożone Treemaps



# Inne rodzaje wizualizacji

---

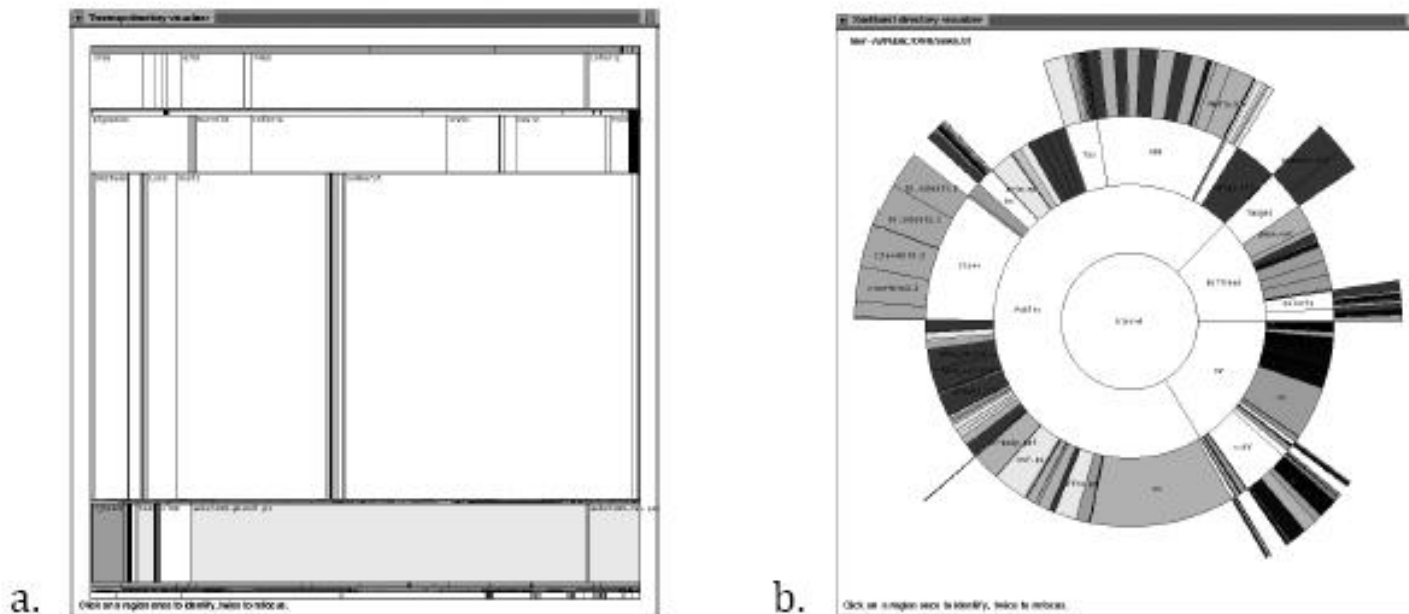
Od Treemaps do Voronoi maps



# Prostokąty vs. hierarchia pierścieniowa

---

Cel – lepsza wizualizacja zależności hierarchicznej



Rysunek 1. Strategia prostokątna (a) i pierścieniowa (b) wizualizacji zasobów katalogowych

Źródło: J. Stasko, R. Catrambone, M. Guzdial, K. McDonald, *An evaluation of space-filling information visualizations for depicting hierarchical structures*, „International Journal of Human-Computer Studies” 2000, vol. 53, iss. 5, s. 667–668.



# Literatura

---

- Chen, C.H., Härdle, W. and Unwin, A. (eds). *Handbook of Data Visualization*, Springer, Heidelberg (2008).
- Unwin, Theus, Hofmann: *Graphics of Large Datasets. Visualizing a Million*, Springer (2006)
- Inselberg, A.: The Plane with Parallel Coordinates, *The Visual Computer* (1985)
- Tufte E.: *The Visual Display of Quantitative Information*
- Tufte E.: *Visual Explanations*

Także materiały wykładowe:

- Chris North: cs5984 Information Visualization (Web site: [www.cs.vt.edu/~north/infoviz/](http://www.cs.vt.edu/~north/infoviz/))
- Ken Brodrie: Envisioning Information Lectures, University of Leeds
- ...





Rodzinnych i radosnych Świąt Bożego Narodzenia  
oraz szczęśliwego Nowego Roku!