

# WIZUALIZACJA INFORMACJI

## Grafika informacyjna

### Wprowadzenie



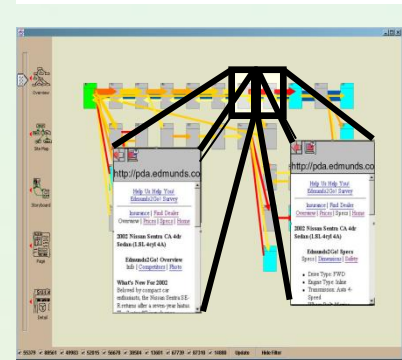
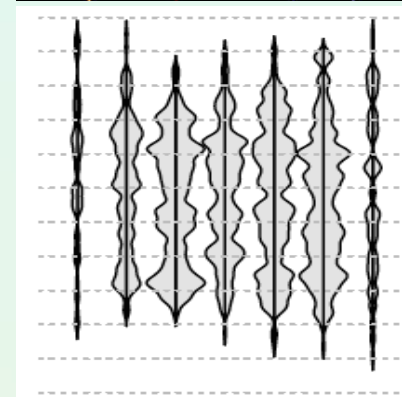
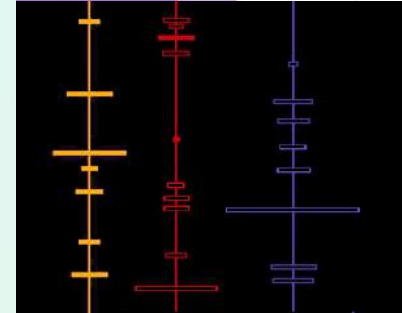
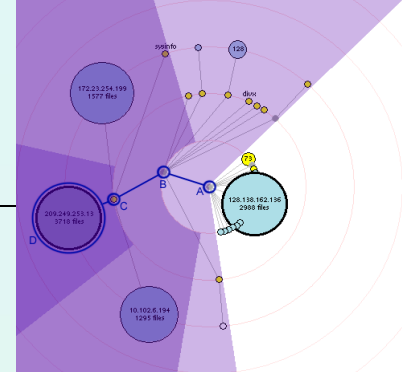
Jerzy Stefanowski

Institute of Computing Sciences,  
Poznań University of Technology

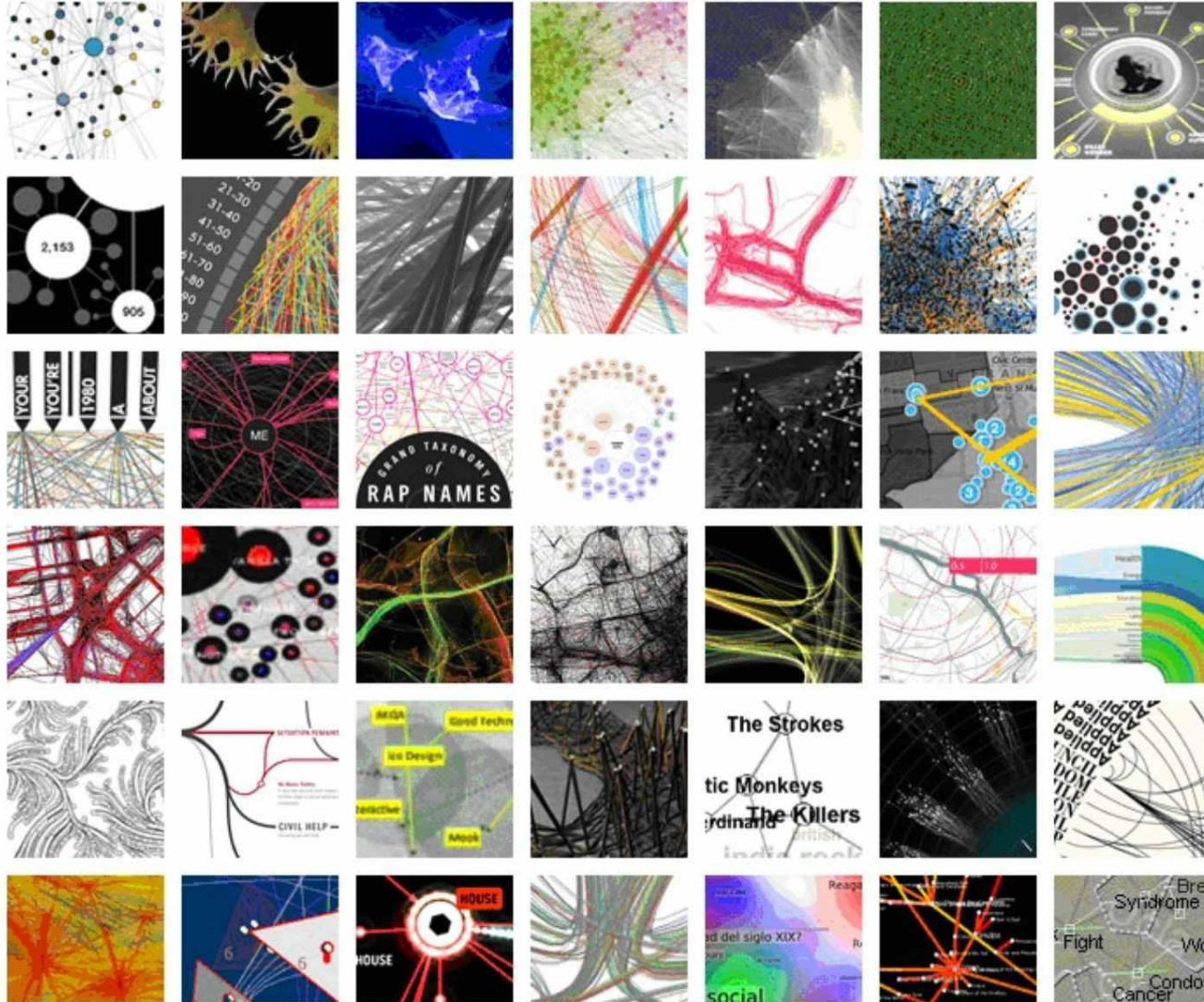
Październik, 2015, Poznań

# Plan Wykładu

1. Przykłady grafiki informacyjnej
2. Wizualizacja informacji → definicje
3. Cele wizualizacji  
→ “Record, Explore, Convince”
4. Perspektywa historycznych wizualizacji
5. Percepcja człowieka i błędy prezentacji
6. Złożone prezentacje graficzne
7. Architektura wyszukiwania informacji
8. Infografiki
9. Organizacja przedmiotu

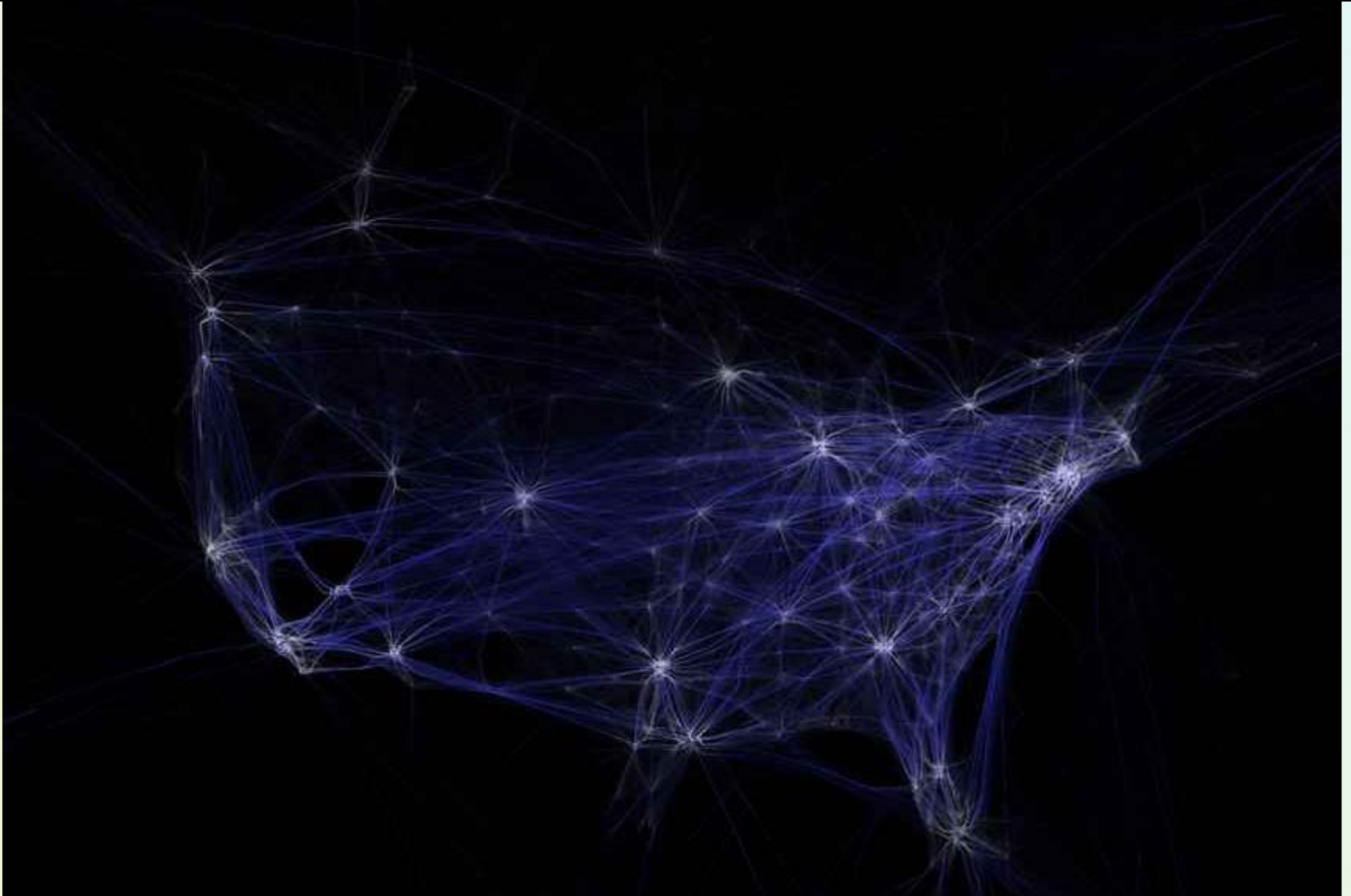


# Graphics are everywhere



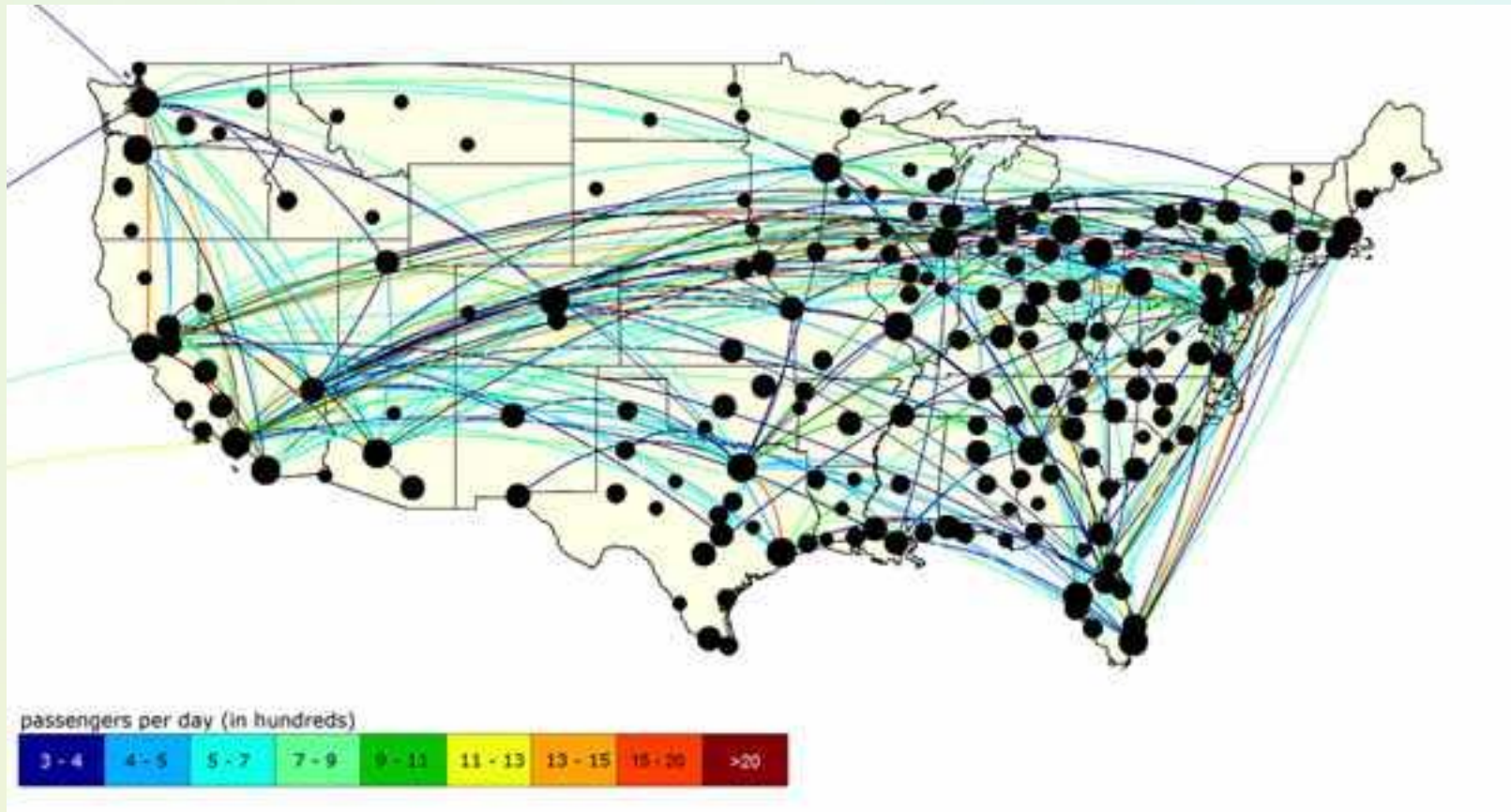
# Przykłady wizualizacji

---

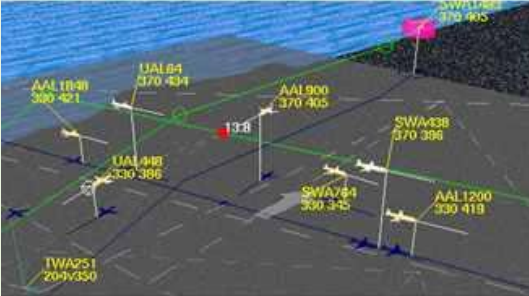
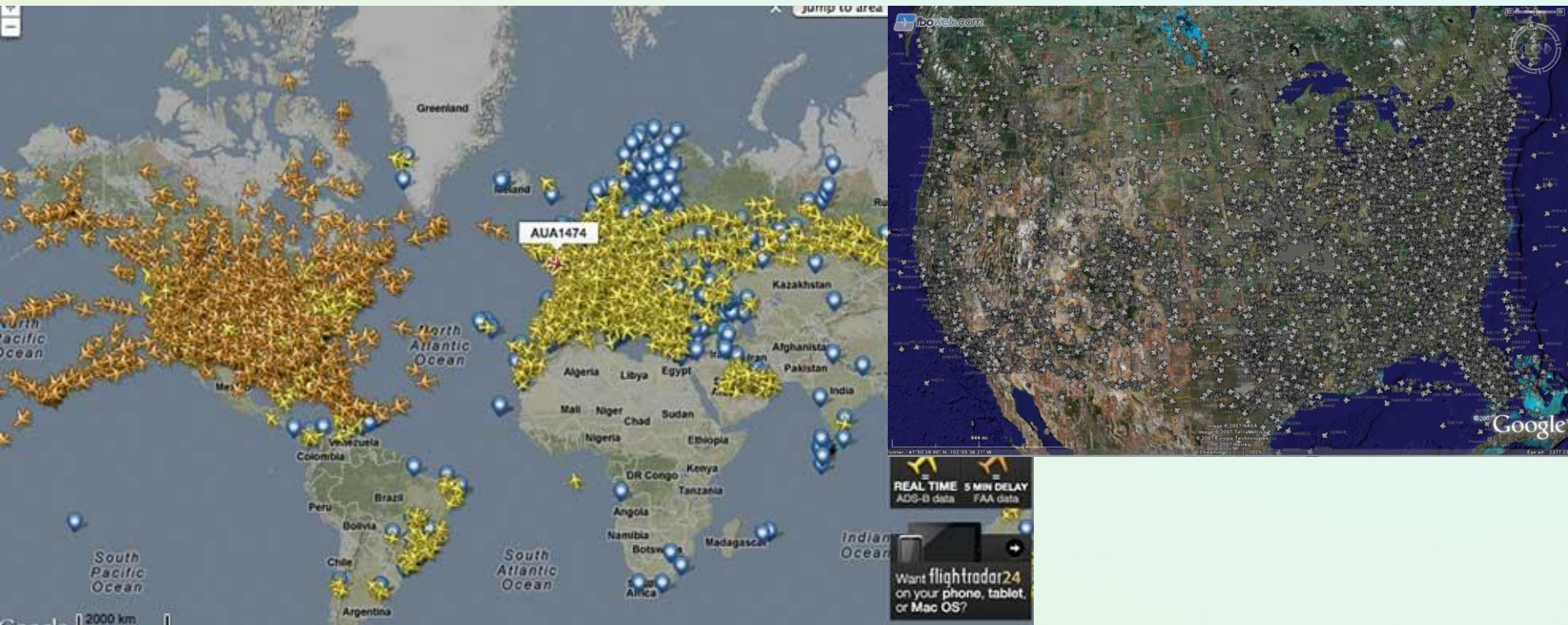


# US flight tracks

Połączenie lotnicze



# Inne wizualizacje → Flight Radar

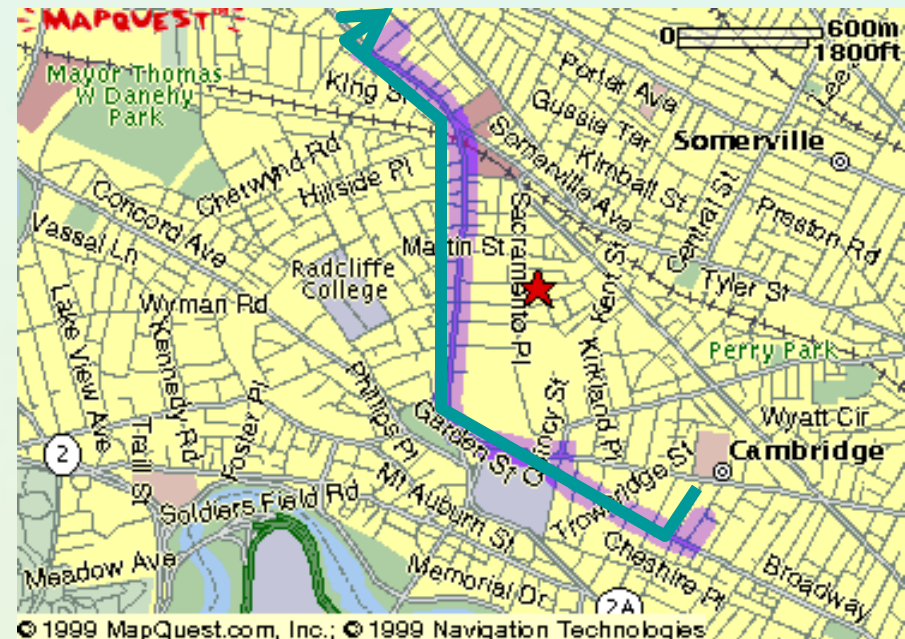


# Wspomaganie poszukiwania drogi

1. Start out going Southwest on ELLSWORTH AVE  
Towards BROADWAY by turning right.
2. Turn RIGHT onto BROADWAY.
3. Turn RIGHT onto QUINCY ST.
4. Turn LEFT onto CAMBRIDGE ST.
5. Turn SLIGHT RIGHT onto MASSACHUSETTS AVE.
6. Turn RIGHT onto RUSSELL ST.

Obraz z mapquest.com

Przykład z wykładu Marti Hearst  
- SIMS 247: Information Visualization and  
Presentation, Univeristy of Berkley



# Electronic Health (Patient) Record

John Doe, 47 year old male - DOB: 07/08/1962 - Follow-up Visit - 10/07/2005 07:12 pm - Next App...

Encounter - Follow-up Visit - Friday, 10/07/2005 07:12 pm - Open

**Summary II**

Name: Doe, John  
DOB: 07/08/1962  
Gender: Male  
Race: Cash-paying patient

**Problems**

| Onset      | ICD   | Description   |
|------------|-------|---|
| 01/05/2010 | 796.2 | Elevated blood pressure reading without diagnosis of hypertension |
| 01/05/2010 | 401.9 | Unspecified essential hypertension                                |
| 01/05/2010 | 401.1 | Essential hypertension, benign course                             |

**Medication List**

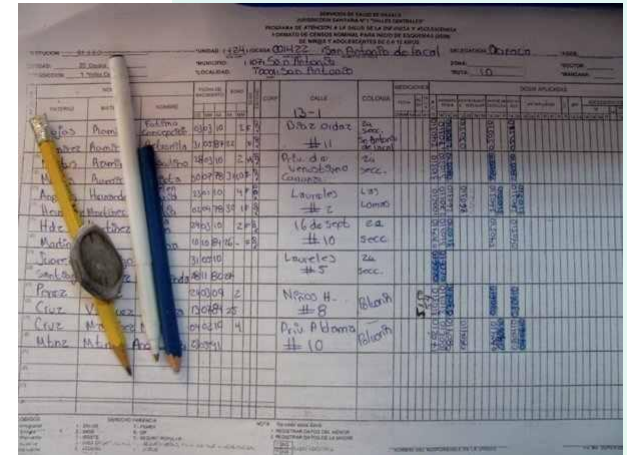
| Date       | Drug Name             | Qty | Issue Date |
|------------|-----------------------|-----|------------|
| 02/07/2005 | metaxin 40-mg capsule | 30  | 02/07/2005 |

**Procedures**

| Date       | Description                            |
|------------|--|
| 01/05/2010 | 663 Rhythmic Strip v. Integ. & results |
| 05/07/2005 | 663 Rhythmic Strip v. Integ. & results |
| 11/07/2004 | Injection consent                      |
| 02/08/2004 | Diast. Injection                       |
| 12/04/2003 | Lesson                                 |
| 12/04/2003 | Lesson                                 |
| 12/04/2003 | Lesson                                 |

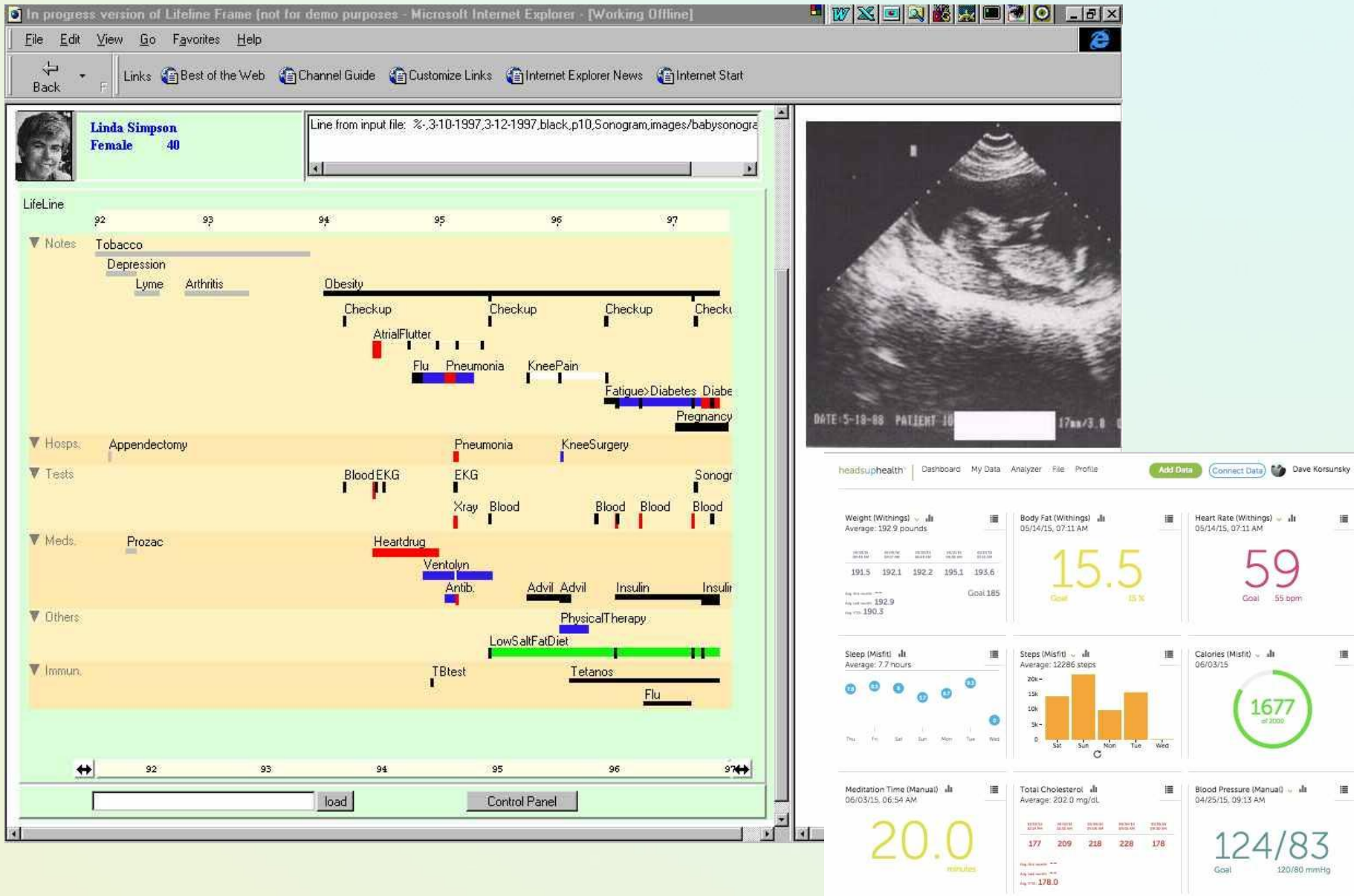
**Screened Documents**

| Date       | Document Type                                 | Description                                   |
|------------|---|---|
| 10/06/2005 | Anesthesia Record                             | Anesthesia Record                             |
| 10/06/2005 | Dr. Substather - Pre-Op Anesthesia Flow Sheet | Dr. Substather - Pre-Op Anesthesia Flow Sheet |
| 10/06/2005 | Pre-Op Evaluation - Pre-Operative Document    | Pre-Op Evaluation - Pre-Operative Document    |
| 10/06/2005 | Ultrasound Pictures                           | Ultrasound Pictures                           |
| 10/07/2005 | Abdomen                                       | Abdomen                                       |
| 10/07/2005 | Abdomen                                       | Abdomen                                       |
| 06/28/2005 | Rx  | Rx  |
| 10/07/2005 | Consult Reports                               | Consult Reports                               |





# Inne wizualizacje danych pacjenta



# Inne wizualizacje danych pacjenta

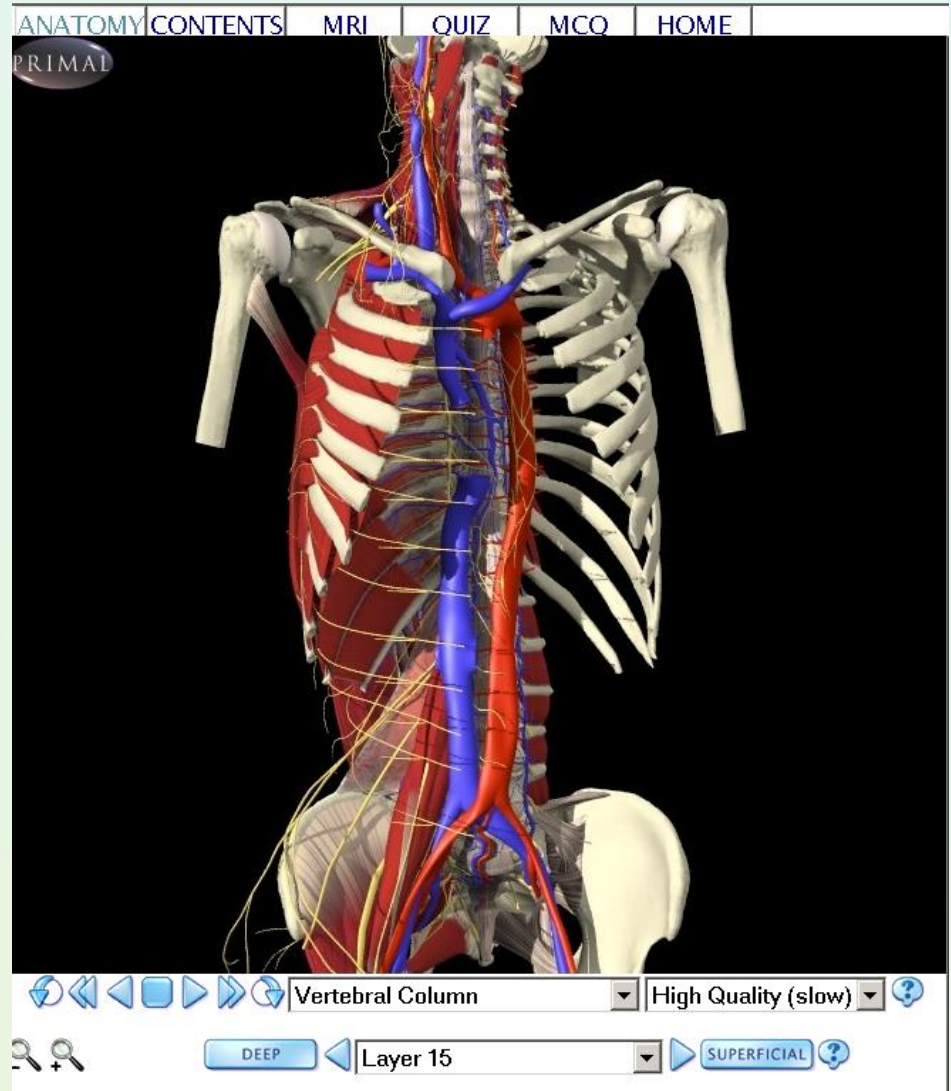


# Medycyna → 3D wizualizacje anatomiczne

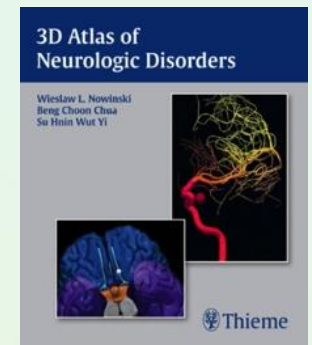
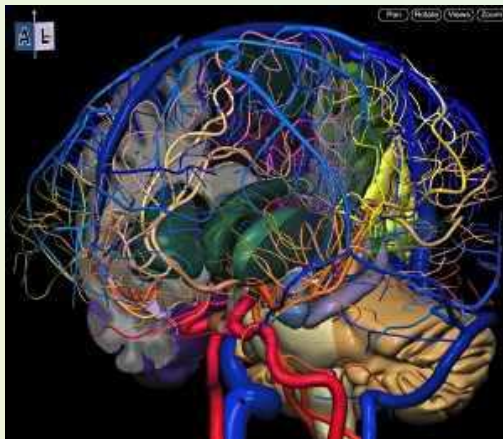
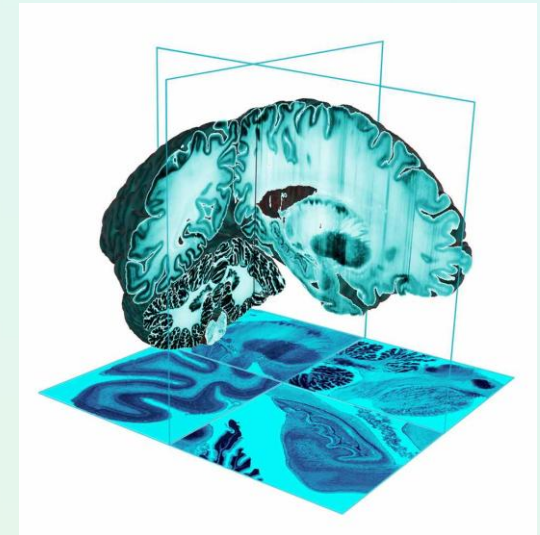
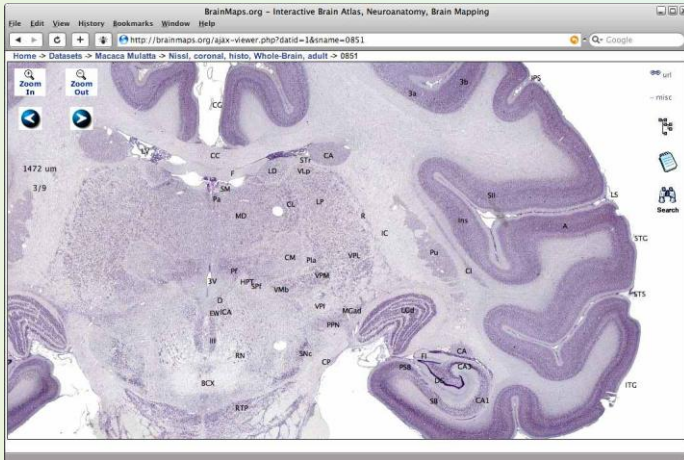
Z wielu zdjęć przekrojowych, ciała ludzkiego stwórz wizualizacje 3D oraz odpowiednio wskaż interesujące organy

Anatomy.TV  
Wykorzystywane do nauczania anatomii

**Look under Anatomy.TV in  
Library electronic resources**



# 3D Human Brain Visualization

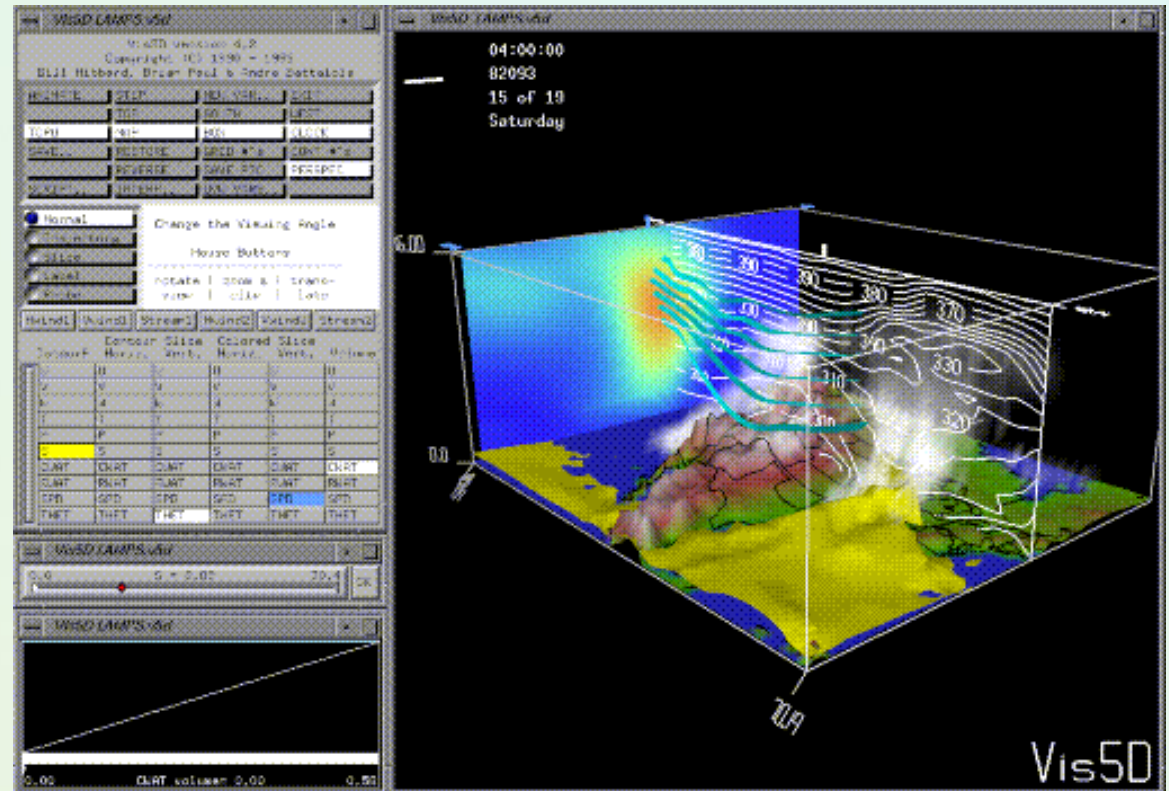


Wiesław L. Nowiński – Human Brain Atlas  
Spójrz <http://www.wieslawnowinski.com/>

# Meteorologia

Ciśnienie w różnych  
Warstwach atmosfery  
Ilustracja poziomiami  
w powierzchniach  
przecięcia

Generated by  
the Vis5D system  
from University of  
Wisconsin (now  
Vis5d+)

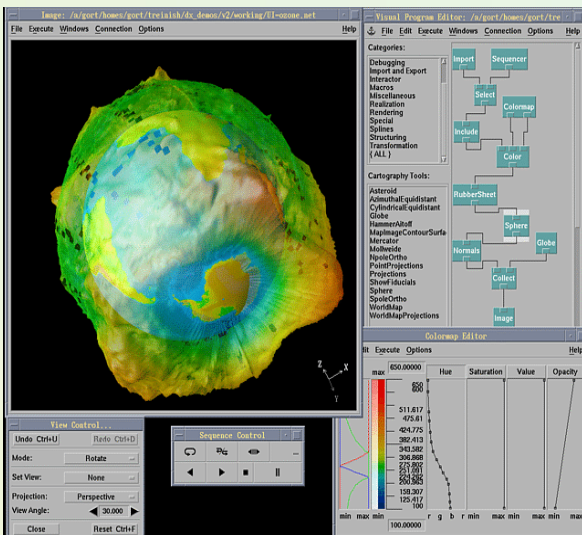


Vis5d: <http://www.ssec.wisc.edu/~billh/vis5d.html>

Vis5d+ : <http://vis5d.sourceforge.net>

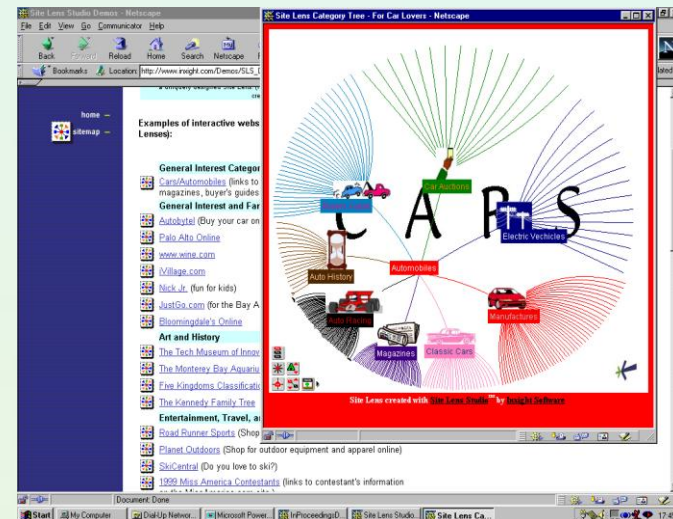
# Dwa oblicza wizualizacji

- ❑ Scientific Visualization (Wizualizacja naukowa)
  - Wizualizacja obiektów i procesów fizycznie istniejących



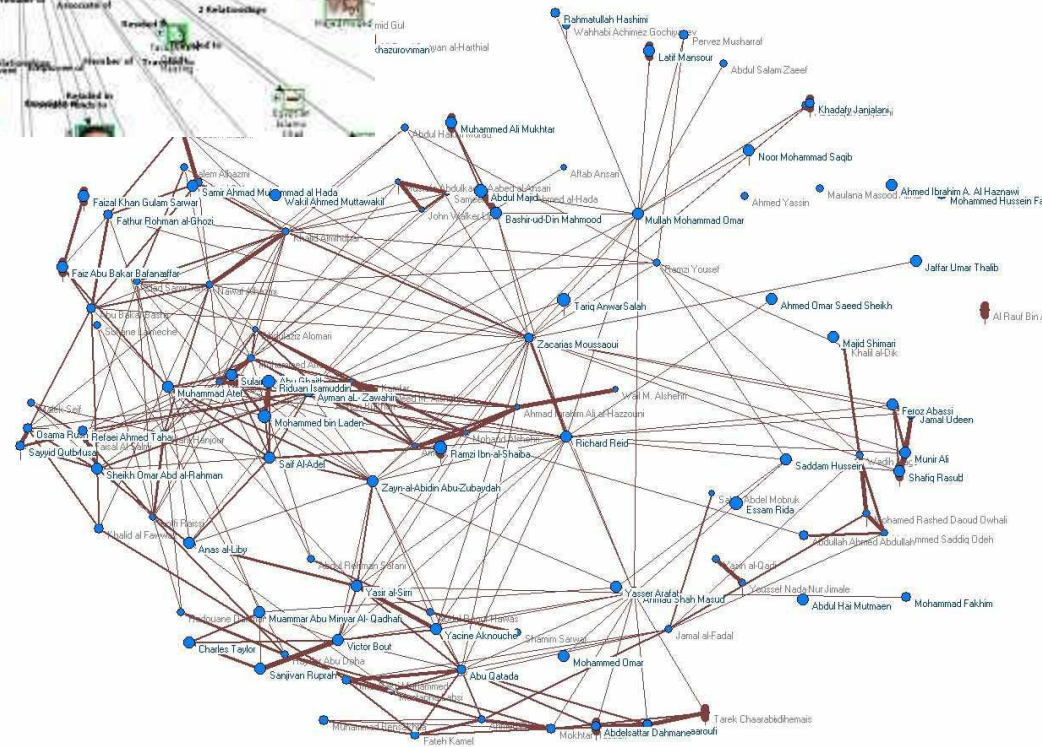
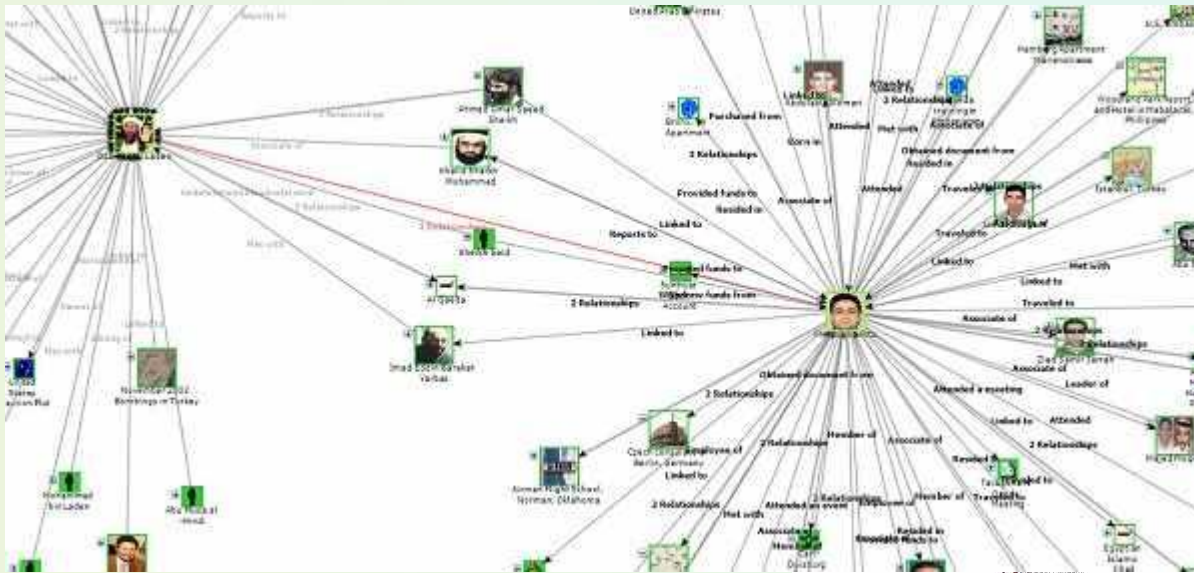
Ozone layer around earth  
IBM Data Explorer  
<http://www.research.ibm.com/dx/>

- ❑ Information Visualization
  - Wizualizacja danych abstrakcyjnych



Automobile web site  
- visualizing links

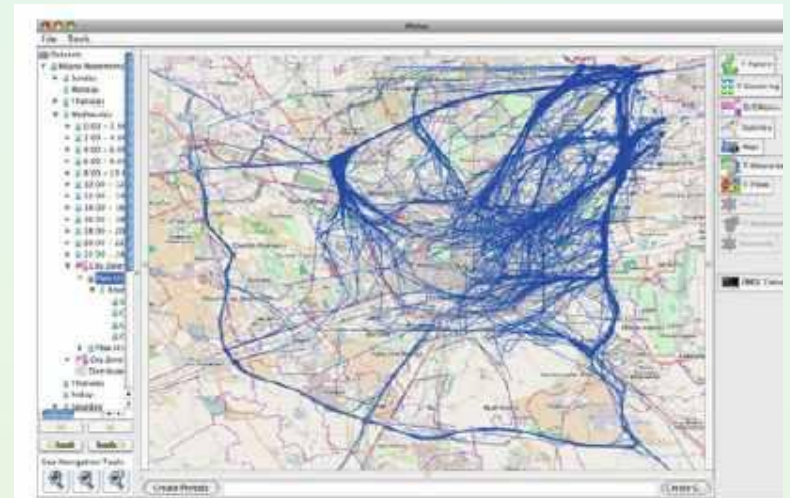
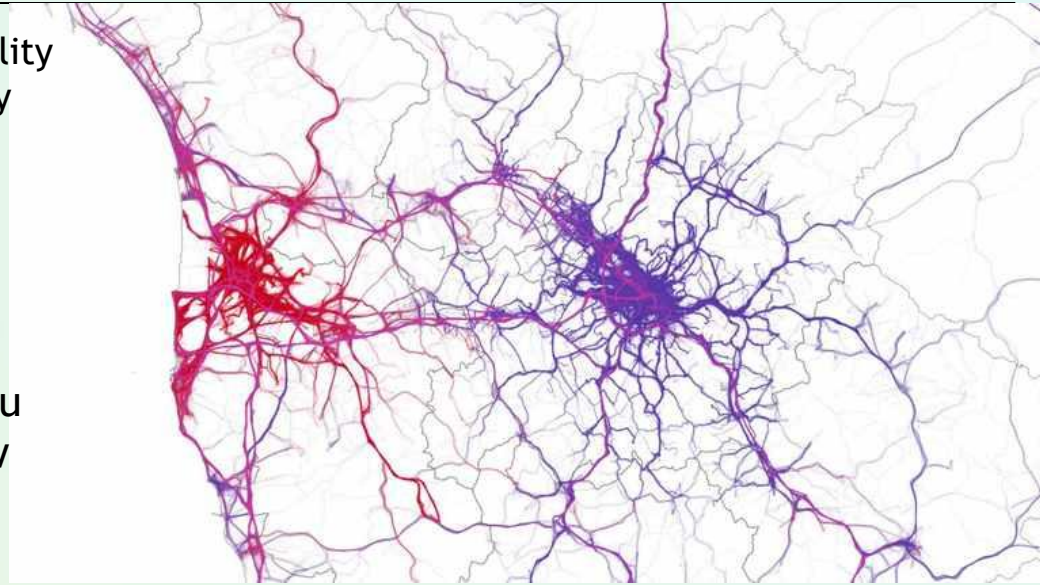
# Analiza powiązań użytkowników Internetu



□ Analiza sieci informacyjnych

# Analiza mobilności ludzi

- ❑ “Unveiling the complexity of human mobility by querying and mining massive trajectory data”. Fosca Giannotti, Mirco Nanni, Dino Pedreschi, Fabio Pinelli, Chiara Renso, Salvatore Rinzivillo, Roberto Trasarti
- ❑ KDD Lab, University of Pisa
- ❑ Złożone niejednorodne repozytoria danych → w tym eksploracja b. wielu zapisów (trajektorii) ruchu pojazdów (GPS devices)
- ❑ Przykładowe pytania i zadania:
  - Odkryj wzorce podróży
  - Gdzie są obszary intensywne ruchu samochodowego? Jak się zmieniają w czasie?
  - Co wpływa na zmiany ruchu?
  - Powiązanie danych z samochodów z danymi demograficznymi.
- ❑ M-Atlas → visual exploration of the discovered behavioral patterns and models



**Fig. 17** The result of T-Clustering from the trajectories moving from the center to the North-East area. *Left* The input data set for the clustering algorithm: the trajectories moving from the center to the North-East



# Poszukiwanie wzorców zachowań

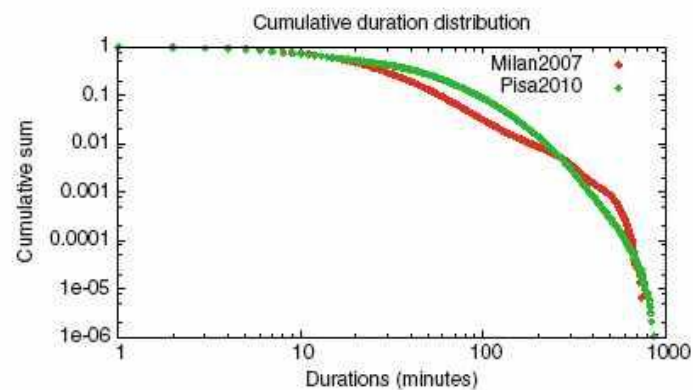
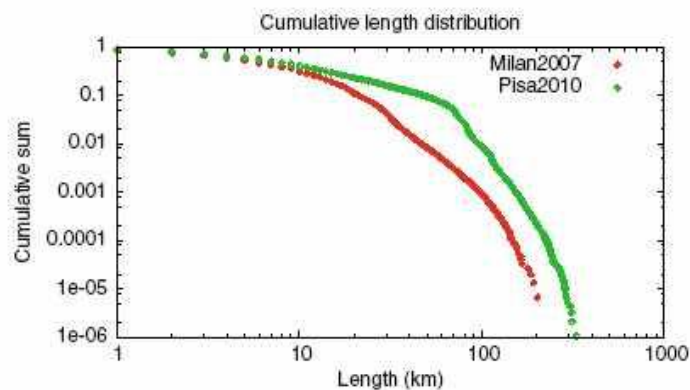
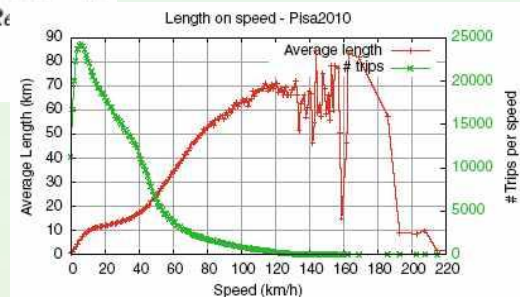


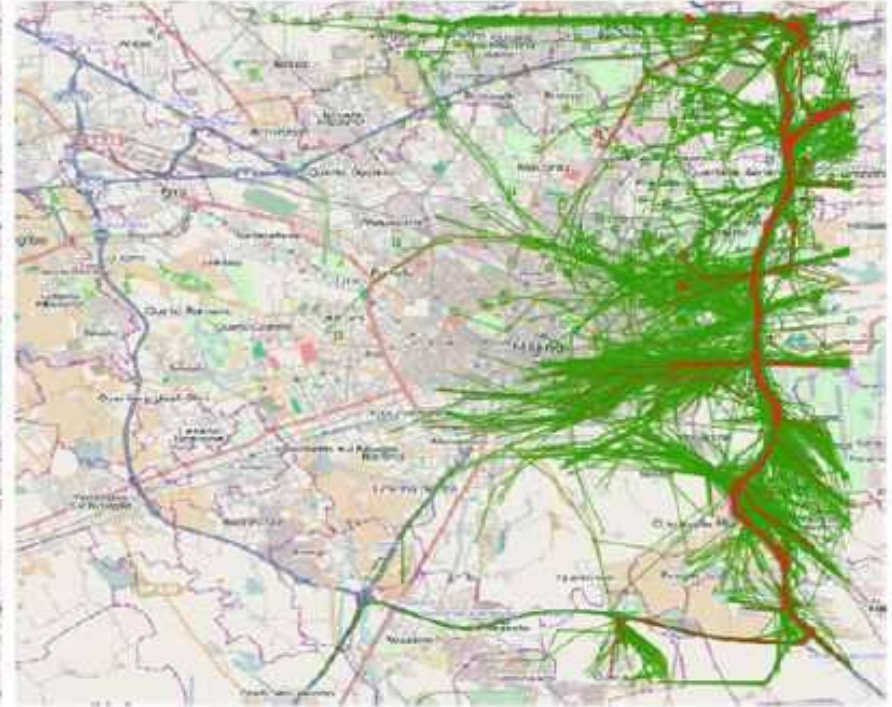
Fig. 3 Trip length cumulative distribution in log-log scale (left), trip duration cumulative distribution in log-log scale (right). *Re* Milano2007 and *green lines* for Pisa2010



# Ruch drogowy - zmiany



**Fig. 12** Classification of new trajectories using a set of specimens from `WednesdaySpecimens`. *Left*, blue lines represent the trajectories of a single cluster of Wednesday, April 4, and the *red* lines are the specimens learned for the selected cluster. *Right*, green lines are the trajec-



ories of the entire week classified by the same set of specimens. Visual inspection confirms that cluster shape is preserved, albeit the size of the second data set is 7 times larger. Quantitative measures of cluster quality, such as silhouette coefficients, can be easily computed

# Grupowanie podobnych mobilności

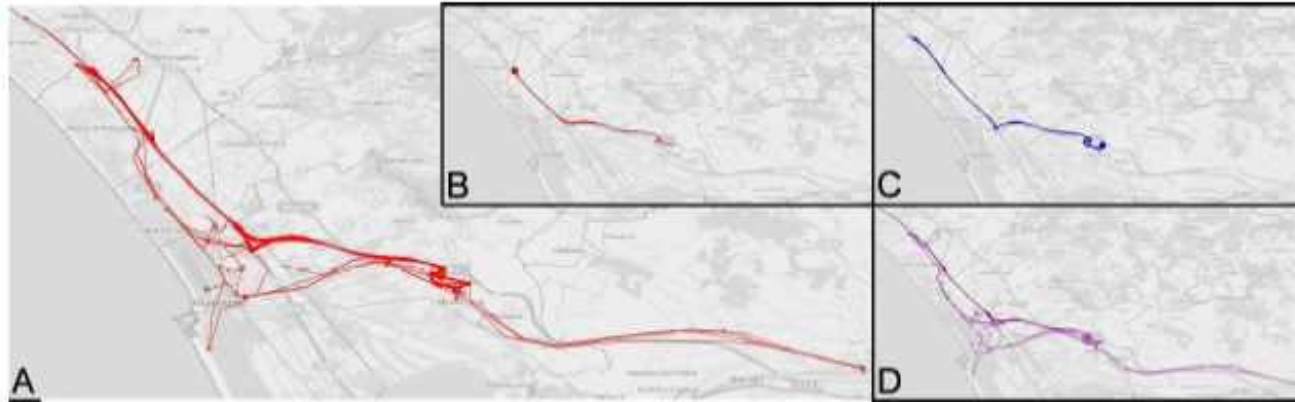


Figure 2: An example of mobility profile extraction: (a) The entire set of trips of a user, (b,c) the two clusters extracted, and (d) the remaining trips which are not periodic.

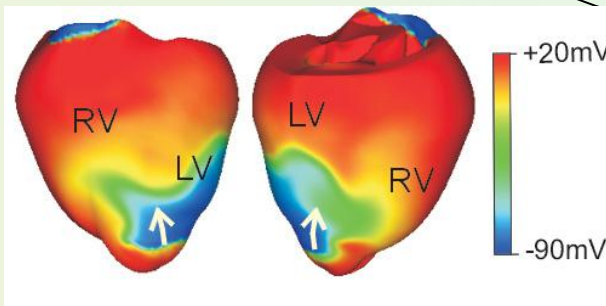


Fig. 3: (Left) The clusters obtained with grid cell size of 2000m. (Right) The clusters determined by the level 2 of the Infomap hierarchy for the same grid resolution.

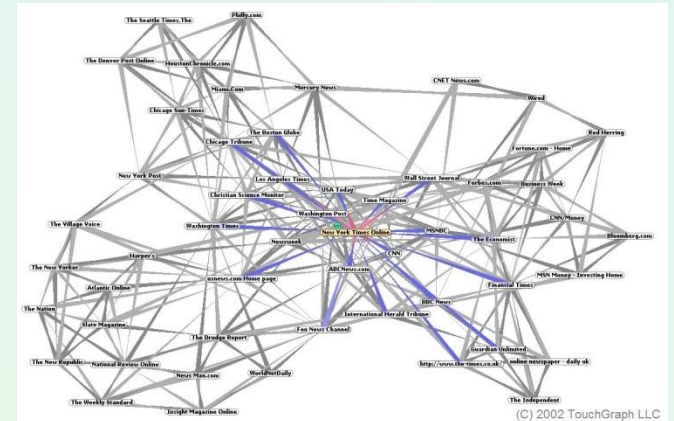
# Visualization - Bridging the Gap between the Computer and the Mind



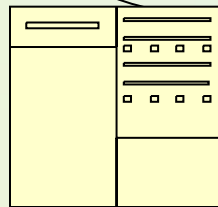
Human Information Analyst



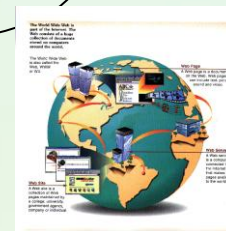
Visualization



Computation Resources



Information Resources



# Visualization? Definicje anglosaskie

---

Visualize: to form a mental image or vision of ...

Visualize: to imagine or remember as if actually seeing.

American Heritage dictionary, Concise Oxford dictionary

The use of computer-supported, interactive visual representations of data to amplify cognition

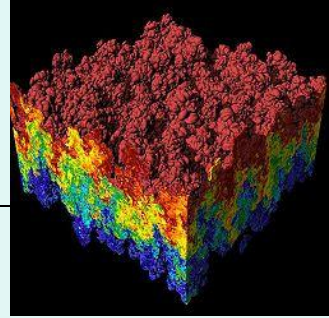
Readings in Information Visualizations: Using Vision to Think

Visualization is any technique for creating images, diagrams, or animations to communicate a message [Wiki]

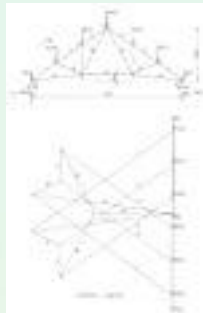
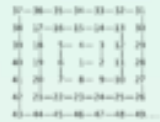
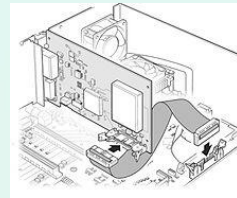
Lecz także

To convey information through visual representations

# Wykorzystanie wizualizacji

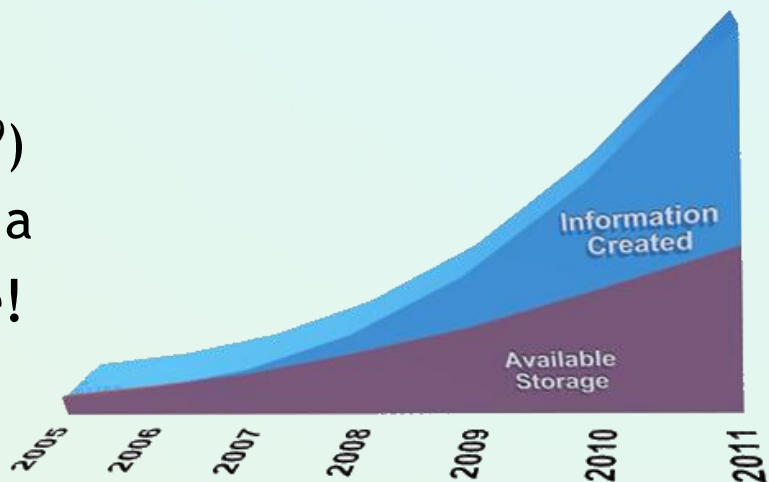


- ❑ Wizualizacja naukowa
- ❑ Wizualizacja edukacyjna
  - Często powiązania z symulacją i animacją
- ❑ **Wizualizacja informacji**
- ❑ **Wizualna eksploracja danych (Visual analytics)**
- ❑ Rysunki techniczne
  - Plany konstrukcyjne lub zasady działania różnych przedmiotów
  - Ilustracje techniczne - dla szerszego grona niewykwalifikowanych odbiorców
- ❑ Kartografia
- ❑ Specjalizowane wykresy matematyczne
- ❑ Wizualizacja przestrzenna
  - projekcje 2D dla trójwymiarowych obiektów (tomografia, MRI)
- ❑ Interaktywne wizualizacje
  - „Virtual Reality”
- ❑ **Komunikacja wizualna → infografiki**



# Rozwój technologii informatycznej & Big Data

- ❑ Gwałtowny wzrost rozmiaru świata danych elektronicznych
  - 2008 zebrano 4 exabytes ( $4.0 \times 10^{19}$ )
  - Co kilka lat ilość danych się podwaja
- ❑ Big Data to nie tylko masywne dane!
- ❑ Big Data - V definicja
  - Volume
  - Velocity (time ...)
  - Variety (heterogeneous, complex data representations)
  - Veracity (trustworthiness)
  - ...
- ❑ Inne „V” → **wizualizacja!**



J. Gantz, D. Reinsel, Extracting Value from Chaos, IDC, 2011



# Information Visualization and Big Data

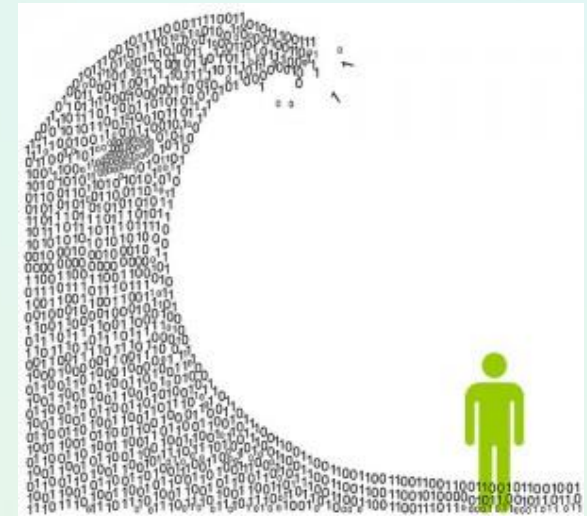
---

## ❑ Wyzwania

- “Masywne dane” lub tzw. “Big Data”
  - Zalew nadmiarem informacji
  - Jak je analizować i wyciągać użyteczne wnioski?

## ❑ Rozwiązania

- Lepiej wykorzystać zdolności ludzkiej percepcji
- Zamień dane szczegółowe w reprezentacje graficzne
- Wizualna eksploracja danych!





# Wizualizacja informacji

---

- ❑ Metodologia dotycząca **analizy danych abstrakcyjnych**, np. danych statystycznych, charakterystyk zachowania użytkowników, notowań giełdowych, pomiarów.
  - Różnice w stosunku do wizualizacji naukowej (zjawiska naturalne i procesy fizyczne)
- ❑ Preferuje się zastosowanie graficznej wizualizacji zamiast reprezentacji liczbowo-tekstowej oraz właściwe wykorzystanie ludzkiej percepcji

Za: Card, Mackinlay, Shneiderman: Readings in Information Visualization: Using Vision to Think (1999).

# Inne definicje InfoViz

---

- ❑ “Transformation of the symbolic into the geometric (..) Visualization offers a method for seeing the unseen“  
(McCormick et al., 1987)
- ❑ “... finding the artificial memory that best supports our natural means of perception.”  
(Bertin, 1983)
- ❑ The depiction of information using spatial or graphical representations, to facilitate comparison, pattern recognition, change detection, and other cognitive skills by making use of the visual system  
(Hearst 2003)
- ❑ The process of representing abstract data as images that can aid in understanding the meaning of the data  
(Peggy Parskey 2009)

# Wizualizacja informacji, cd.

---

- ❑ Nie tylko prosta transformacja danych, lecz także wspomaganie ich zrozumienia oraz interpretacji (S.Few 2009)
- ❑ Jakościowe rozumienie a nie szczegółowe liczby
- ❑ “W obrazach a nie liczbach znajdziemy najefektywniejszy sposób opisu, analizy i podsumowań dużych zbiorów danych ilościowych” (Edward Tufte)
  - Well designed graphics are usually simplest and at the same time the most powerful

# Exploratory Data Analysis

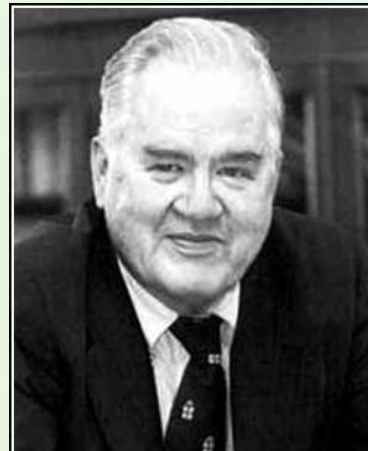
---

- “Information visualization is ideal for exploratory data analysis. Our eyes are naturally drawn to trends, patterns, and exceptions that would be difficult or impossible to find using more traditional approaches, such as tables or text, including pivot tables. When exploring data, even the best statisticians often set their calculations aside for a while and let their eyes take the lead”

S. Few: Now you see it

**Eksploracyjna Analiza Danych** → To coś więcej, zapoznaj się z pracami John’a Tukey’a

“ EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task (..) Summarize the main characteristics of data, often with visual methods ”



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

— John Tukey —

# Cele wizualizacji informacji

---

- Zwięzłe podsumowanie wielkich wolumenów danych
- Jakościowa analiza złożonych, trudnych danych
- Prezentacja informacji z różnych punktów widzenia
- Udostępnienie wielu poziomów szczegółowości prezentacji (“zoom” od ogółu do szczegółu)
- “Tell stories about the data “

Marti Hearst -- SIMS 247: Information Visualization and Presentation, Univeristy of Berkley

## Shneiderman's Mantra:

- Overview first, zoom and filter, then details-on-demand
- Overview first, zoom and filter, then details-on-demand

**Odkrywać! Ujawniać! Objasniać!**

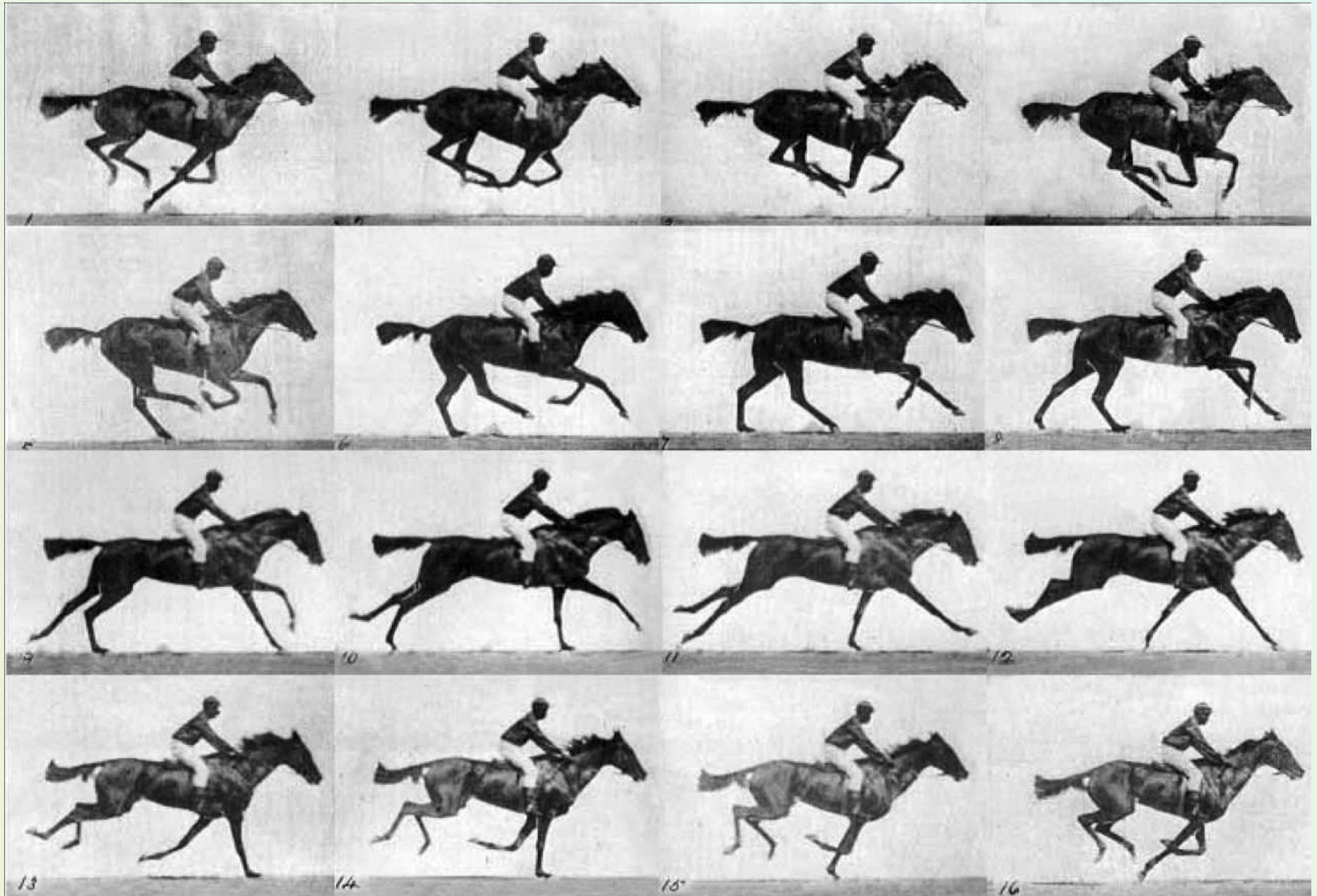
Przemysław Biecek

# The value of visualization

---

- ❑ **Rejestruj** → Record information (Register)
  - Blueprints, photographs, seismographs, ...
- ❑ **Eksploruj dane** → Explore
  - Analizuj, aby wspomagać decyzje
  - Poszukuj i formułuj nowe hipotezy
  - Znajdź ukryte wzorce w danych (KDD - hidden patterns)
- ❑ **Komunikuj i przekonuj innych** → Convince & Communicate information to others
  - Grafika prezentacyjna i infografiki
  - Przekonuj innych odbiorców
  - Współpracuj i uaktualniaj

# Rejestracja



Gallop, Bay Horse "Daisy" [Muybridge 1884-86]

# Eksploruj i odkrywaj ukryte wzorce

Kwartet Anscombe'a → Rola graficznej analizy w statystyce

| Set A |       | Set B |      | Set C |       | Set D |      |
|-------|-------|-------|------|-------|-------|-------|------|
| X     | Y     | X     | Y    | X     | Y     | X     | Y    |
| 10    | 8.04  | 10    | 9.14 | 10    | 7.46  | 8     | 6.58 |
| 8     | 6.95  | 8     | 8.14 | 8     | 6.77  | 8     | 5.76 |
| 13    | 7.58  | 13    | 8.74 | 13    | 12.74 | 8     | 7.71 |
| 9     | 8.81  | 9     | 8.77 | 9     | 7.11  | 8     | 8.84 |
| 11    | 8.33  | 11    | 9.26 | 11    | 7.81  | 8     | 8.47 |
| 14    | 9.96  | 14    | 8.1  | 14    | 8.84  | 8     | 7.04 |
| 6     | 7.24  | 6     | 6.13 | 6     | 6.08  | 8     | 5.25 |
| 4     | 4.26  | 4     | 3.1  | 4     | 5.39  | 19    | 12.5 |
| 12    | 10.84 | 12    | 9.11 | 12    | 8.15  | 8     | 5.56 |
| 7     | 4.82  | 7     | 7.26 | 7     | 6.42  | 8     | 7.91 |
| 5     | 5.68  | 5     | 4.74 | 5     | 5.73  | 8     | 6.89 |

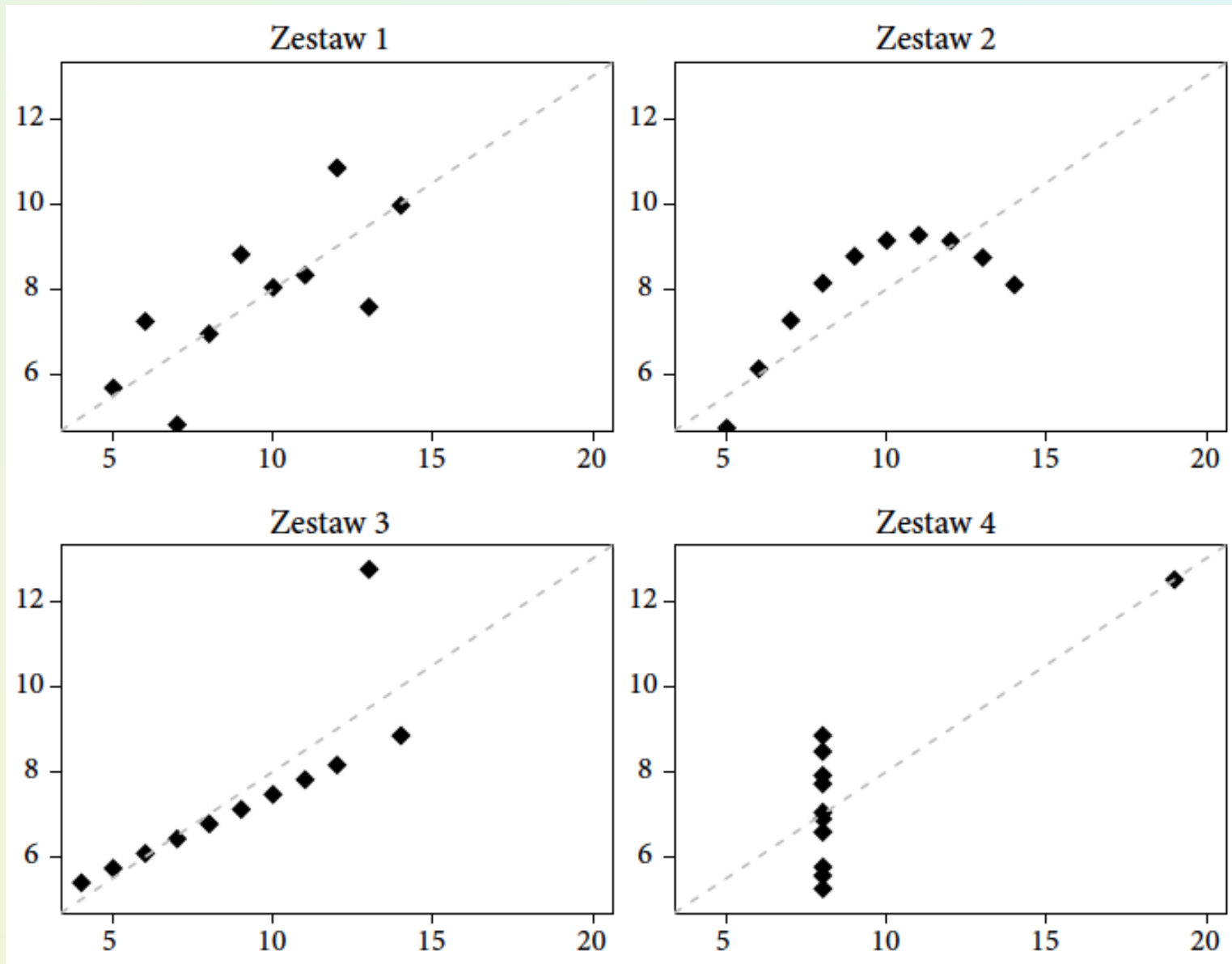
  

|                                  |                          |
|----------------------------------|--------------------------|
| <b>Summary Statistics</b>        | <b>Linear Regression</b> |
| $\mu_X = 9.0$ $\sigma_X = 3.317$ | $Y = 3 + 0.5 X$          |
| $\mu_Y = 7.5$ $\sigma_Y = 2.03$  | $R^2 = 0.67$             |

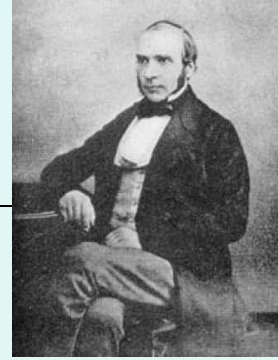
**[Anscombe 73]**



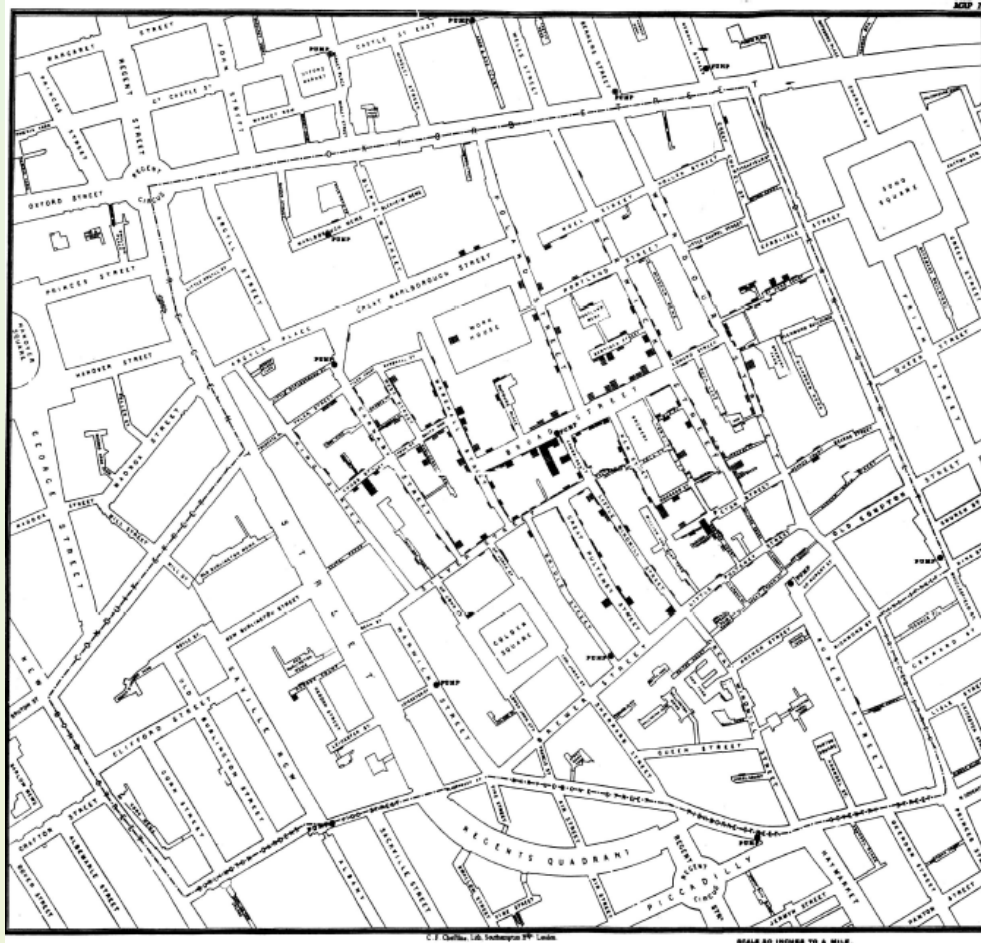
# Wykresy kwartetu Anscombe'a



# Epidemia cholery w Londynie (1854)



- ❑ Soho - rosnące zachorowania i zgony
- ❑ Poszukiwanie przyczyny - John Snow wywiady

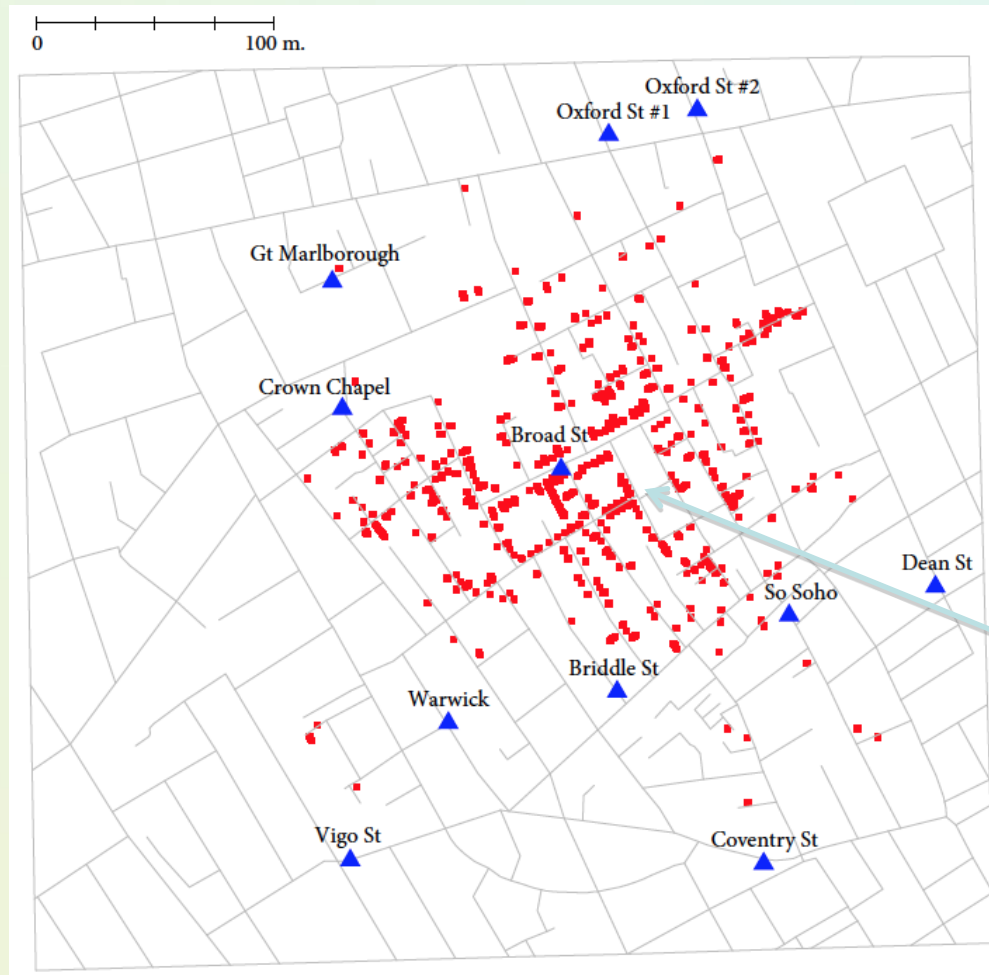


naniesienie na mapę dzielnicy punktów symbolizujących zachorowanie / zgon danej osoby

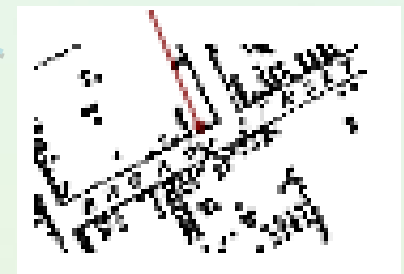
za Visual Explanations by Edward Tufte, Graphics Press, 1997

# Epidemia cholery badanie J.Snow'a

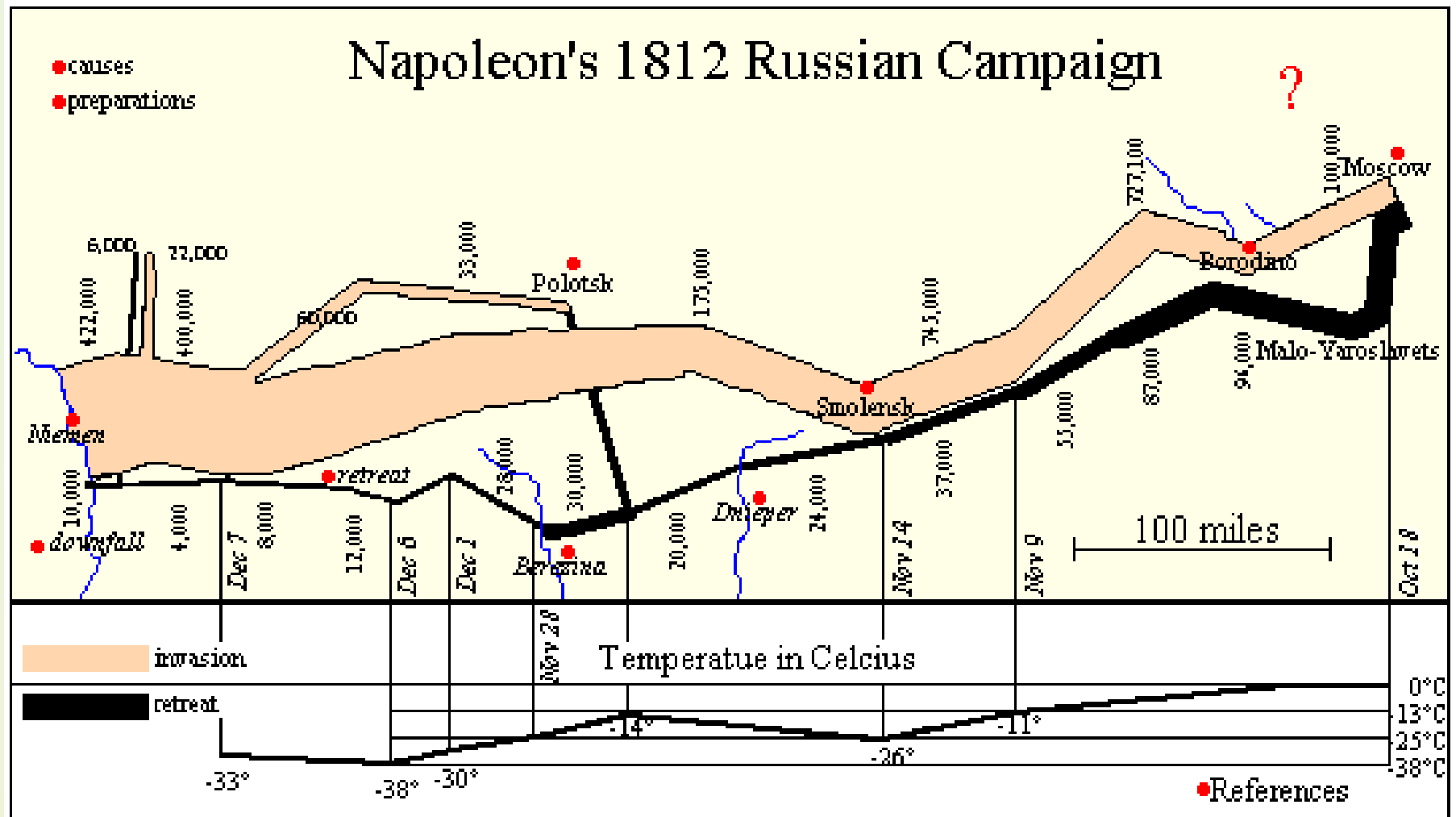
- Hipoteza - przypadki zachorowań i śmierci są związane z pompą wody na Broad Street



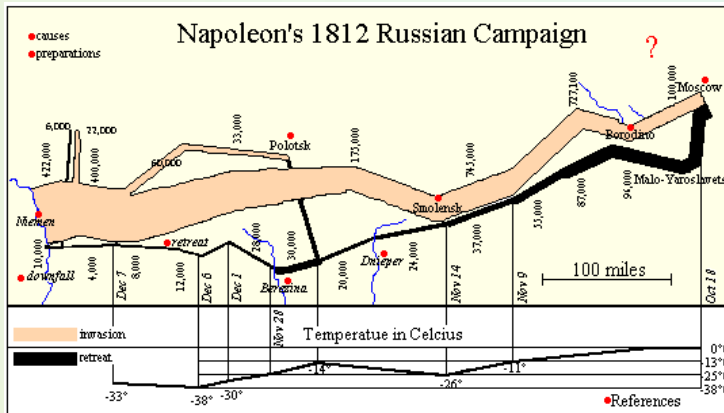
Skupienie  
największej  
liczby zgonów



# Tufte - Minard's *Napoleon's March to Moscow*



# E. Tufte – opowieść o kampanii rosyjskiej Napoleona



## Enforce Visual Comparisons

Width of tan and black lines gives you an immediate comparison of the size of Napoleon's army at different times during march.

## Show Causality

Map shows temperature records and some geographic locations that shows that weather and terrain defeated Napoleon as much as his opponents.

## Show Multivariate data

*Napoleon's March* shows six: army size, location (in 2 dimensions), direction, time, and temperature.

## Use Direct Labeling

Integrate words, numbers & images

Don't make user work to learn your "system."

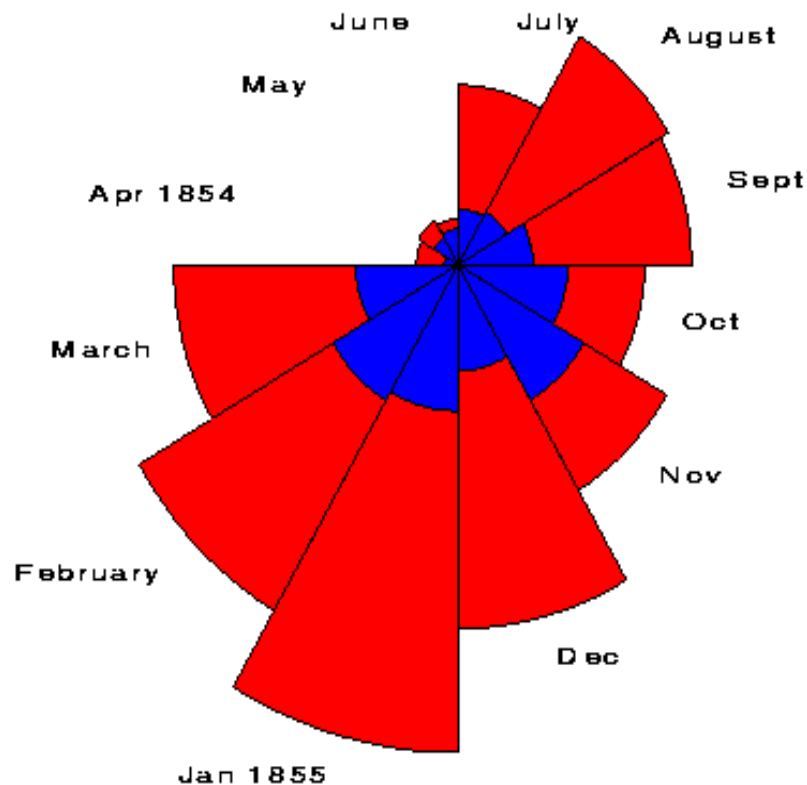
Legends or keys usually force the reader to learn a system instead of studying the information they need.

## Design Content-Driven

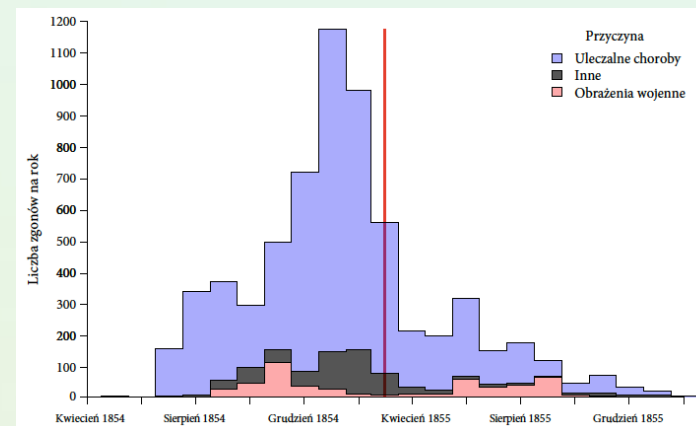
# F. Nightingale (1856) - perswazja

Causes of Mortality in the Army in the East  
April, 1854 to March 1855

■ Non-Battle  
■ Battle



From: F. Nightingale, "Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army", 1858



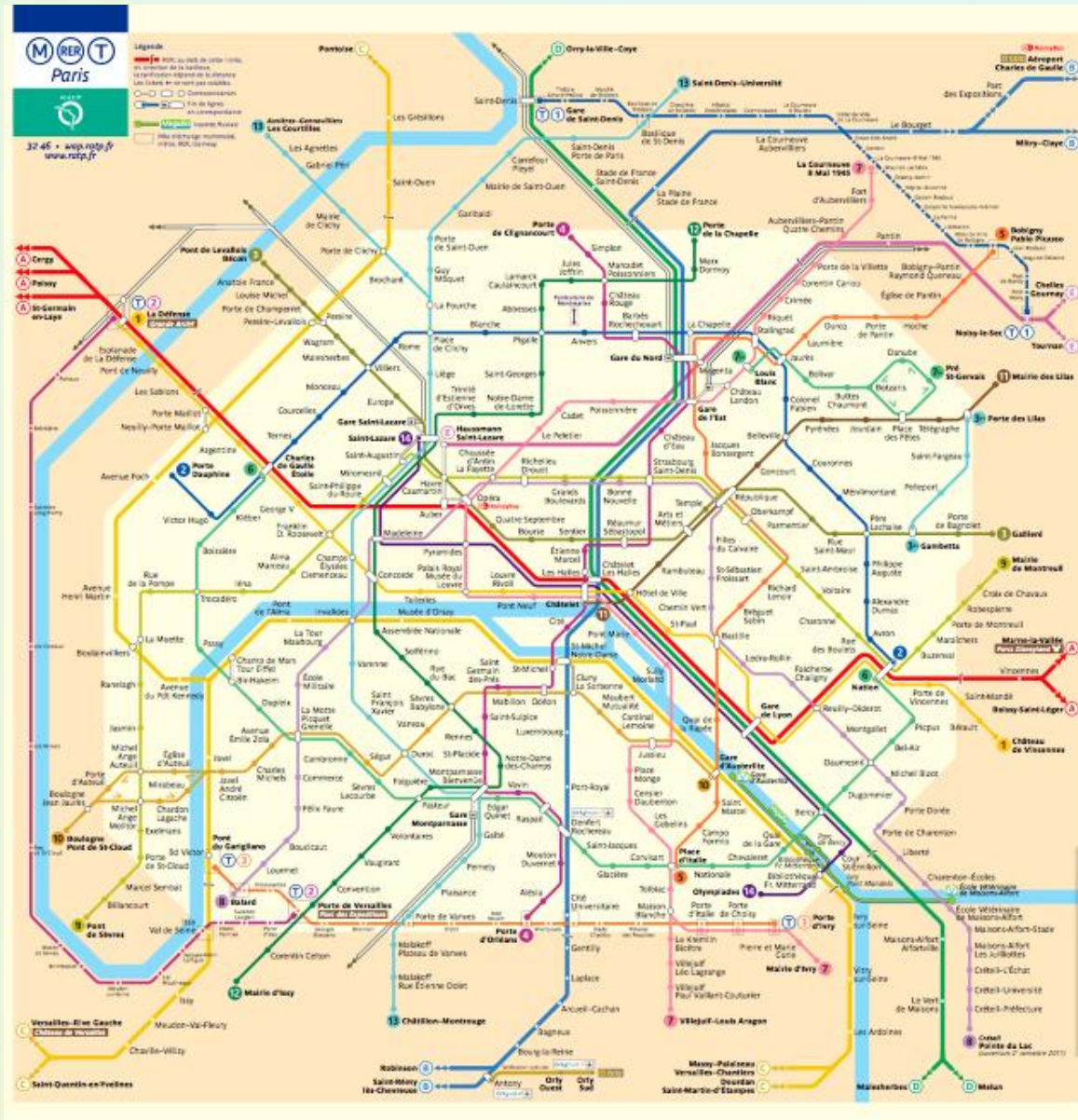
# Clarify - prosto wyjaśnij

## Mapy metra

- Tradycyjne - linie naniesione na tle mapy Paryża (1930)



# Schematy informacyjne dla pasażerów





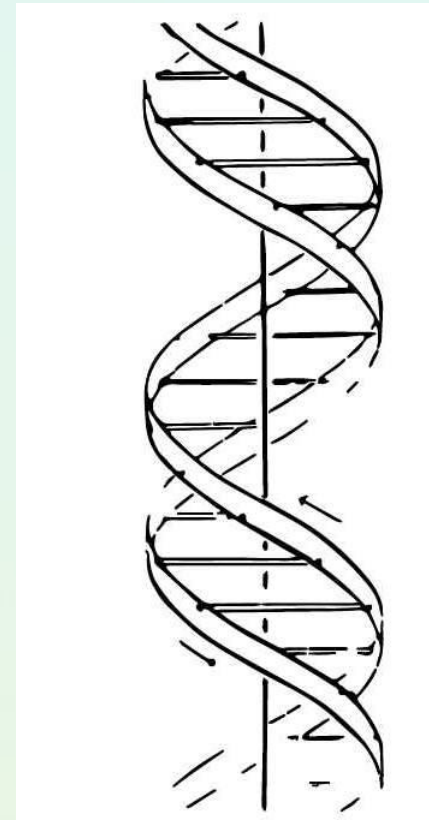
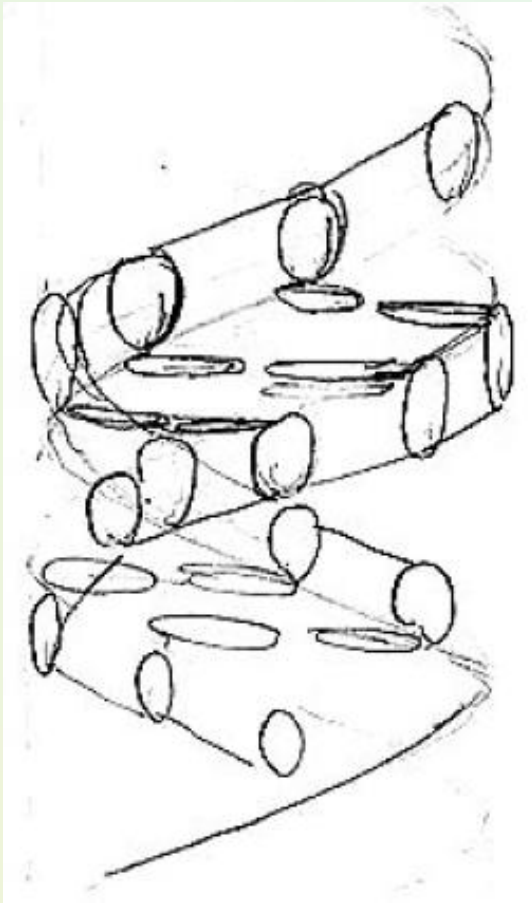
# Harry Beck – mapa dla London Tube (1931)



# Inspiruj

---

- ❑ Podwójna helisa [Watson & Creek 1953]



# Pytania badawcze

---

- ❑ Jak dokonać transformacji danych w postać graficzną zachowując ich oryginalne znaczenie i dostarczyć nowego spojrzenia?
- ❑ Jak efektywnie przedstawić złożone, wielowymiarowe dane / zjawisko w prosty i czytelny sposób?
  - Jakie techniki transformacji zastosować?
  - Które reprezentacje graficzne i w jakich sytuacjach mogą być skuteczniejsze?
  - Nowe środki technologiczne (IT)
- ❑ Jak najlepsze uwzględnienie aspektów poznawczych
  - “The purpose of visualization is insight, not pictures”

# Wizualizacja BC → Historyczne przykłady

---

- ❑ Wizualizacja i wyobrażenia, oraz wykorzystanie wykresów zawsze było ważnym składnikiem badań naukowych.
  - Kto to powiedział? I kiedy?
- ❑ Wiele przykładów historycznych *Before Computers* (BC)
  - Wykresy (X wiek)
  - Schematy prezentacyjne (Leonardo da Vinci)
  - Grafiki ekonomiczne XVIII (William Playfair)
  - Specjalne mapy XVII (Halley)
  - i wiele innych

Odpowiedź: Rene Descartes, 1637

# Krótką historia grafik informacyjnych

**3,000 BC**

Good examples of early infographics are **Egyptian hieroglyphics** which formed language through the use of graphic symbols and icons.



**30,000 BC**

The first examples of infographics date back to the **Late Stone Age** when our ancestors began painting animal portraits on cave walls in the south of France.



**1350**

Medieval French philosopher **Nicole d'Orseme** created one of the first graphs in order to help explain how to measure a moving object.

**1510**

**Leonardo da Vinci** blended written instruction with illustrations to create a comprehensive guide on human anatomy.



**1786**

Scottish engineer **William Playfair** pioneered data visualization. His book "The Commercial and Political Atlas and Statistical Breviary" was the first to explain numeric data through the use of linear graphs, pie charts and bar graphs.

**1857**

English nurse **Florence Nightingale** combined stacked bar/ pie charts (Coxcomb chart) to illustrate the monthly number of casualties and causes of death explain during the Crimean War. She used these infographics to help convince Queen Victoria to improve conditions in military hospitals.



**1850-1870**

**Charles Joseph Minard**, a Civil Engineer from France, began combining maps with flow charts in order to explain geographical statistics. One of his most famous data visualizations illustrated the causes of Napoleon's failed attempt to invade Russia. He captured a complex data set for the period (map location, direction travelled, decline in troops and temperature) in a single infographic.

**1970-1990**

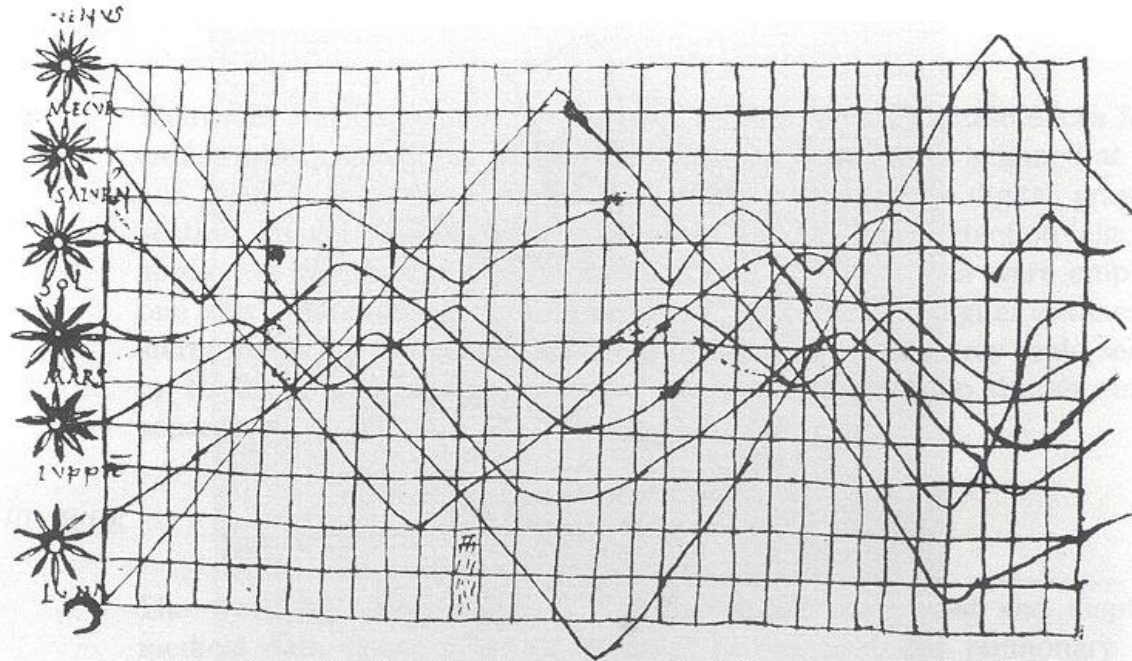
Infographics became more popular as **mainstream news publications** like The Sunday Times (UK), Time Magazine and USA Today began using them to simplify information and enhance comprehension of complicated issues and news stories.



**1930-1940**

The modern era ushered in Isotype, a visual communication model developed by **Otto Neurath** to teach ideas and concepts through the use of icons and pictures.

# Historyczne wykresy astronomia

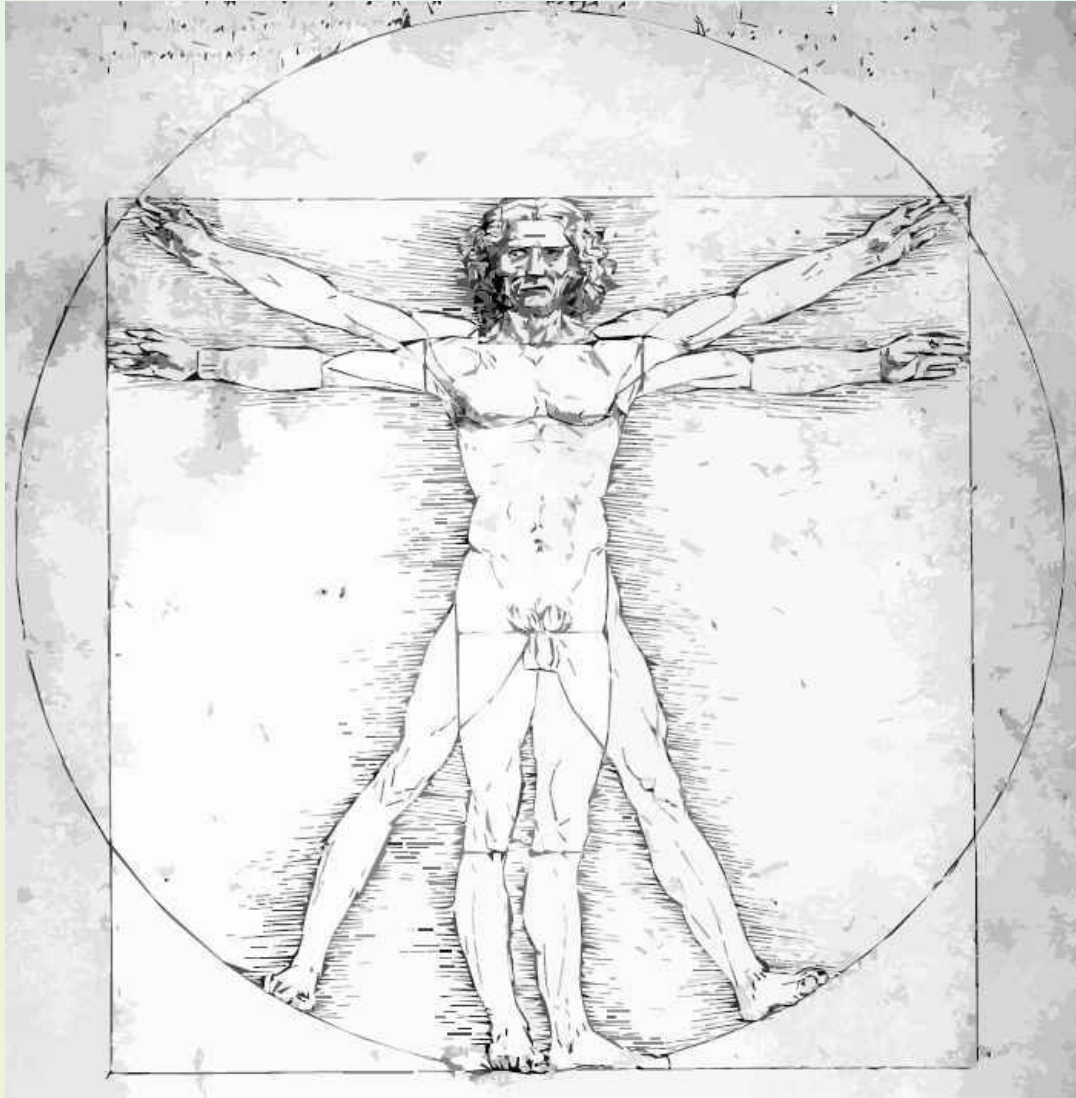


**Figure 1.2.** Line charts(S): illustration of the inclination of the planetary orbits as a function of time (10th Century).

This and following two pictures are taken from Brian Collins 'Data Visualization - Has it all been seen before?' in 'Animation and Scientific Visualization', Academic Press

# Leonardo da Vinci (1452-1519)

---



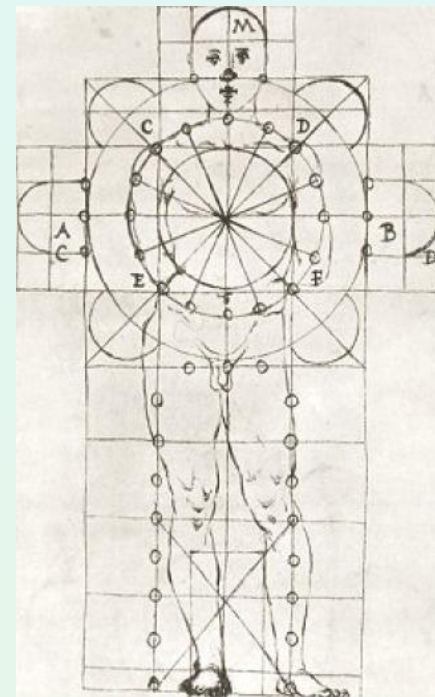
# Człowiek witruwiański

- Wyzwanie dla malarzy, rysowników
  - zrozumienie proporcji ciała idealnie zbudowanego człowieka

- Leonardo da Vinci

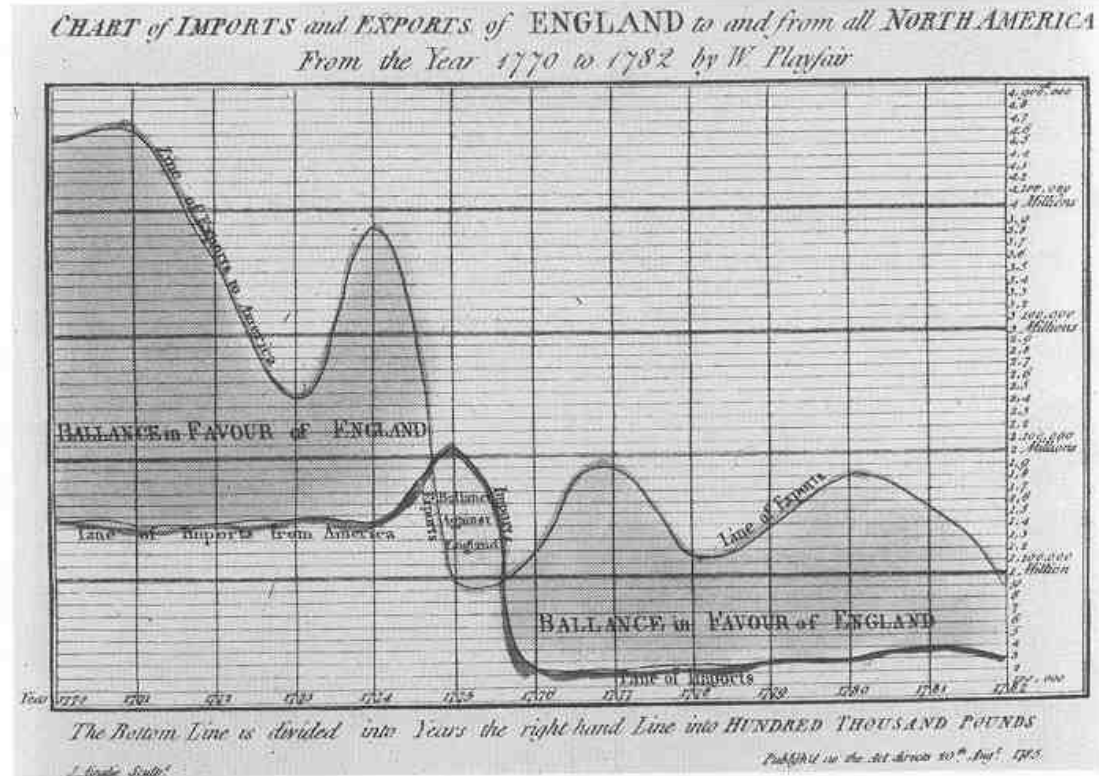
Wg. rzymskiego architekta Witruwiusza:

- Wysokość idealnie zbudowanego człowieka odpowiada szerokości rozłożonych ramion
- Sylwetkę człowieka można podzielić w pionie na cztery równe części, których granice przebiegają na wysokości kolan, podbrzusza i piersi
- Wysokość człowieka i długości jego kończyn można również wyrazić jako wielokrotności długości jego ramion, nadgarstka, łokcia czy barku





# The First Business Graphics

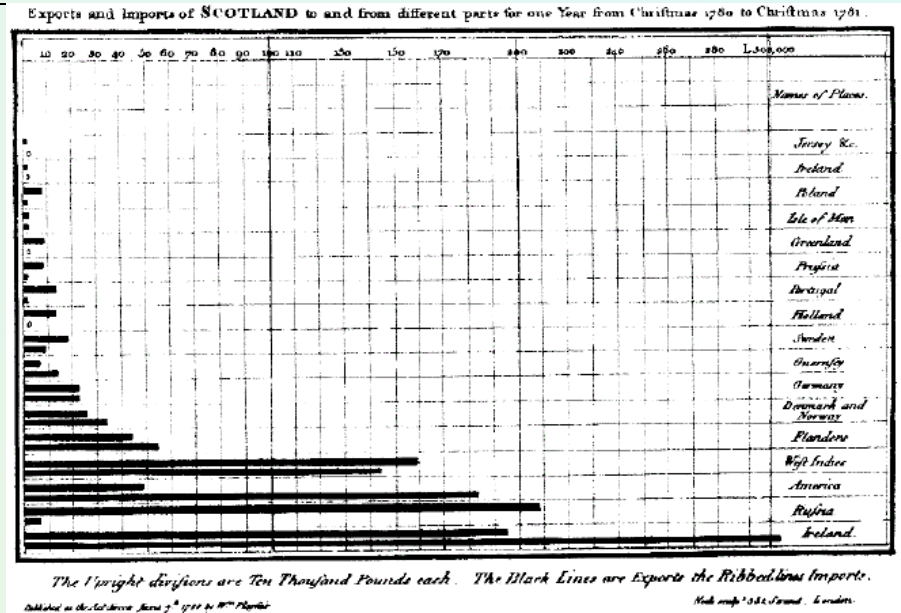


**Figure 1.3.** Line charts(S): chart of imports and exports from the year 1770 to 1782 (1785, William Playfair).

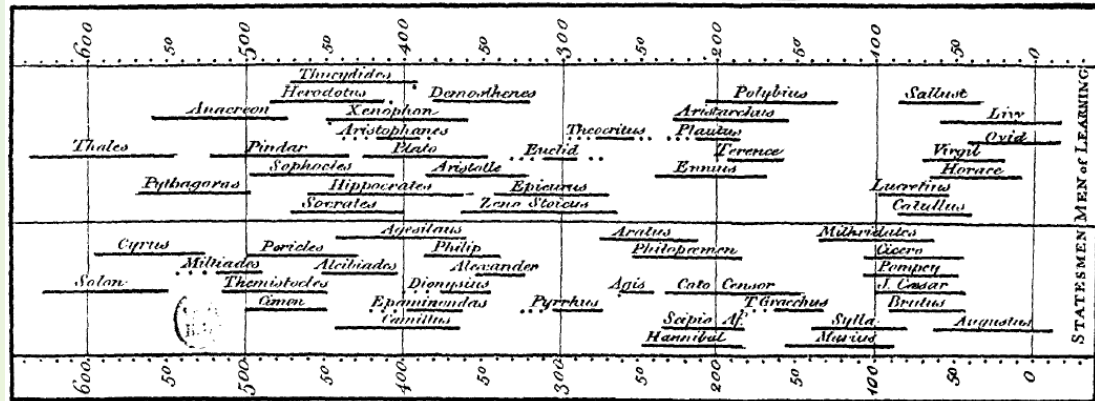
# Trochę historii wizualizacji danych

□ William Playfair

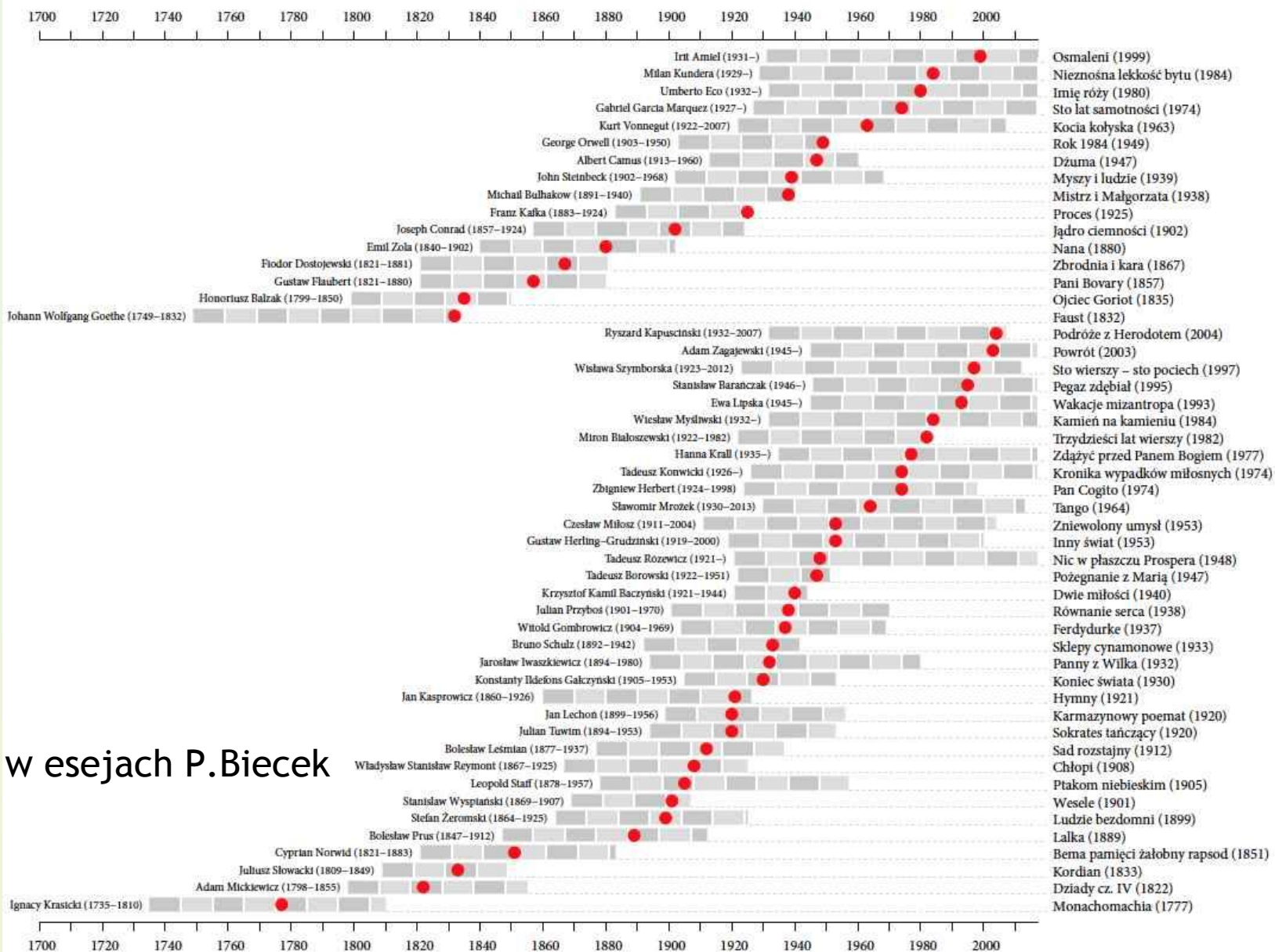
□ Joseph Priestley



## A Specimen of a Chart of Biography.



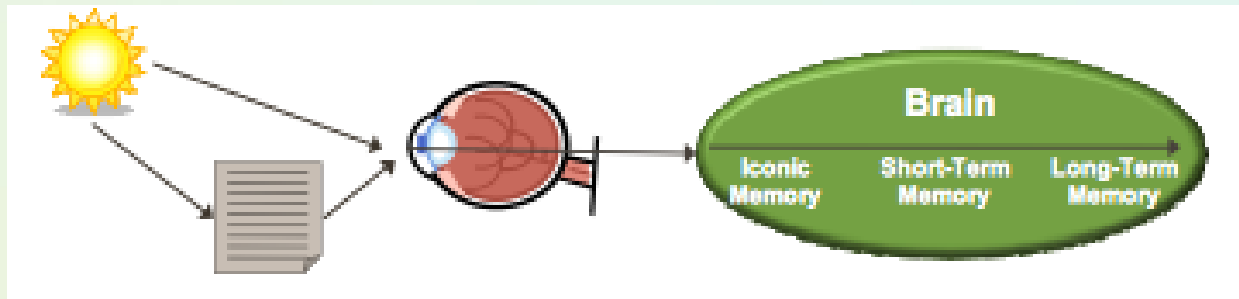
# Biecek prezentacja twórczości literackiej



- Więcej w esejach P.Biecek

# Lepsze zrozumienie percepcji człowieka

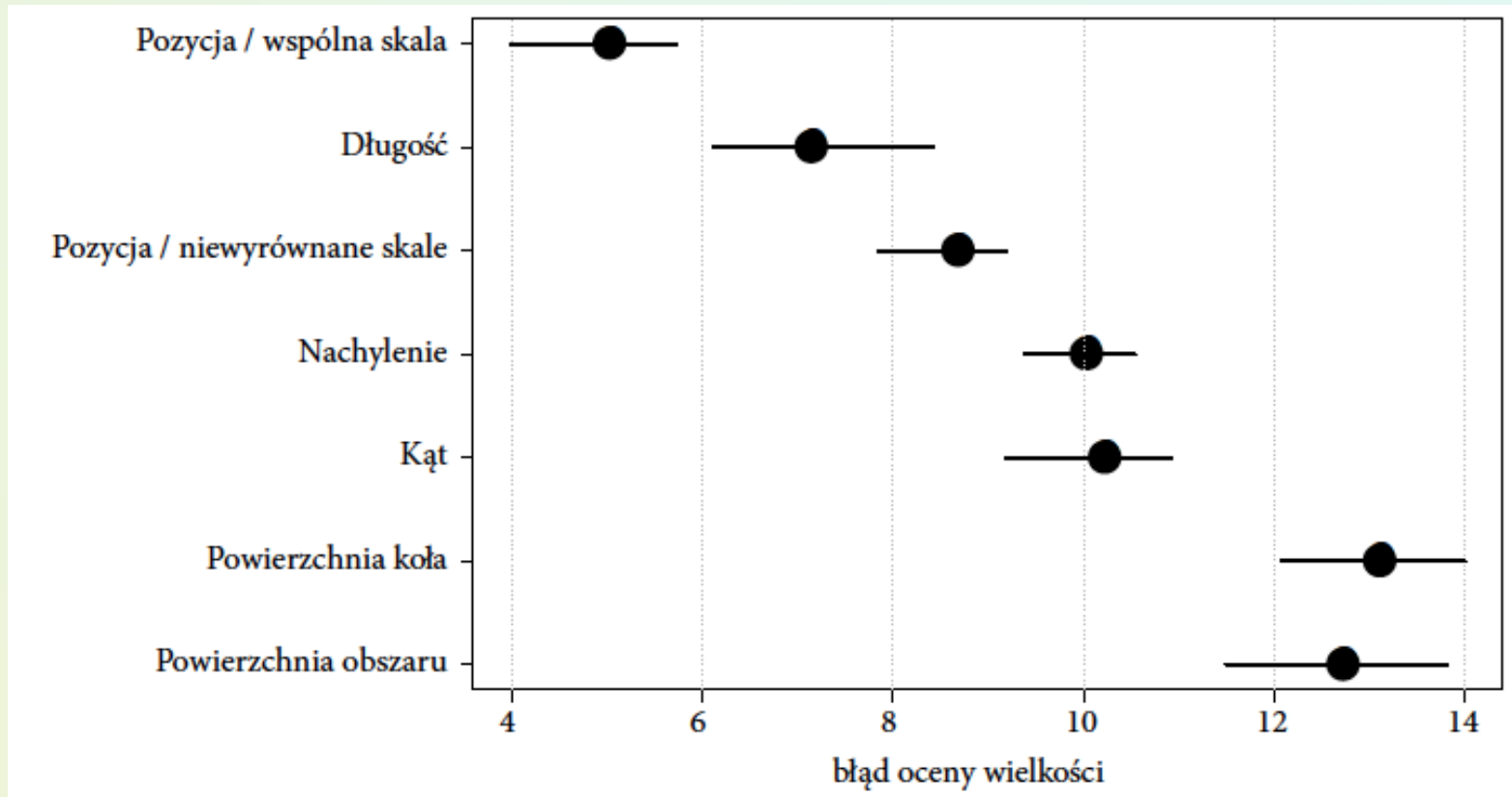
- “The way in which problems are represented has much to do with the quality of the solutions that are found”
  - Herbert Simon



- Człowiek przetwarza szybciej obrazy w porównaniu z tekstem
- Lepiej zapamiętuje informacje powiązaną z obrazami (zwłaszcza w tzw. pamięci długoterminowej)
  - Po 3 dniach do 65% informacji wizualnej vs. 20% czysto tekstowej
- Zmiany zdolności skupienia słuchaczy i czytelników

# Badanie nad percepcją podstawowych elementów

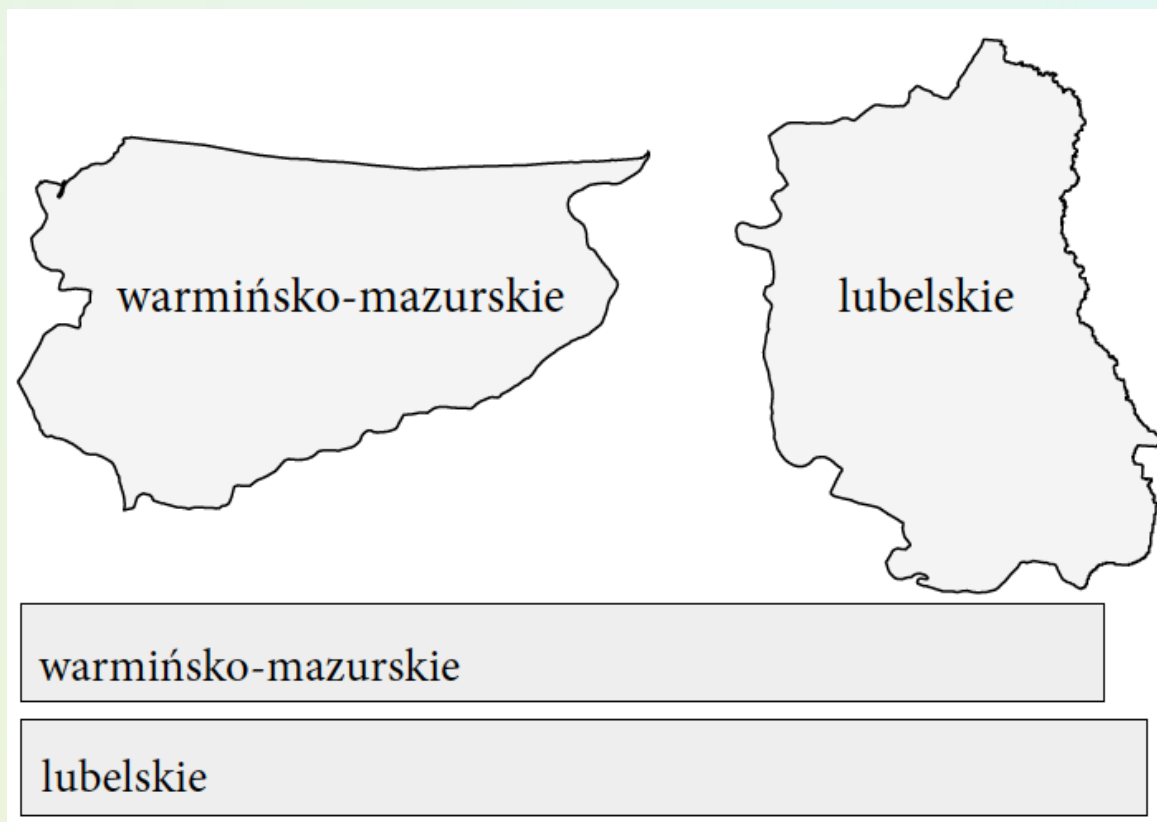
- ❑ Jacques Bertin oraz William Cleveland
- ❑ Średni względny błąd percepcji wielkości prezentowanych za pomocą różnych charakterystyk



# Przykłady błędów percepcji

---

- ❑ Ilustracje Przemysław Biecek - eseje o sztuce prezentowania danych
- ❑ Które województwo jest większe i o ile?

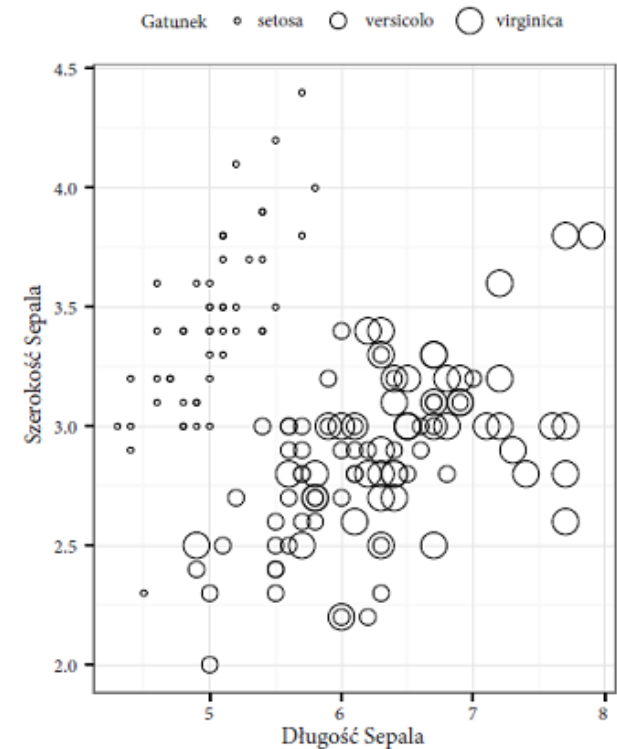
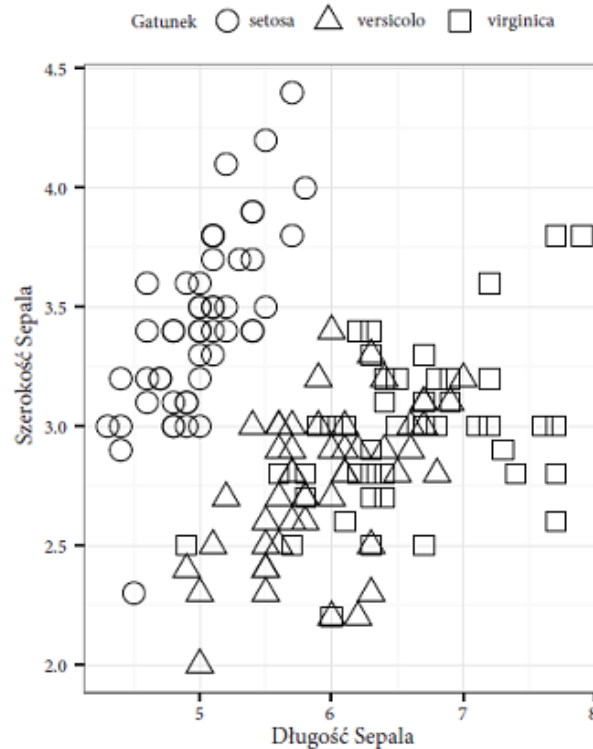
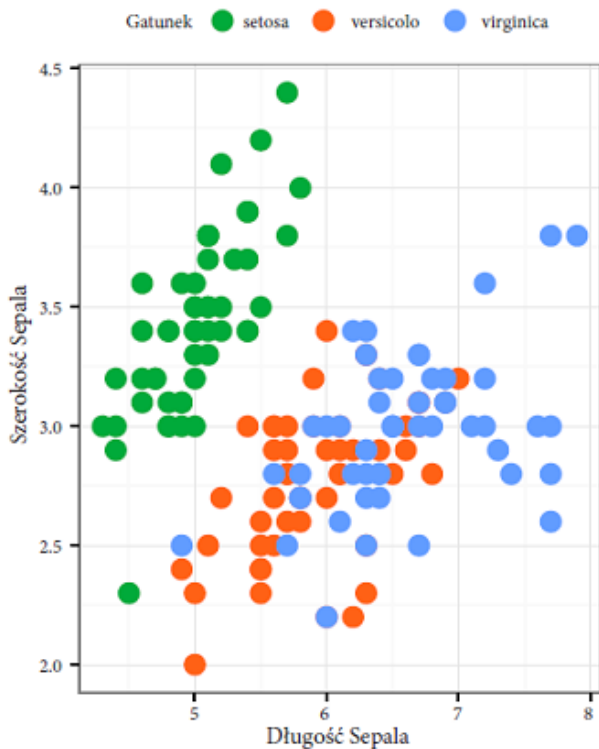


- ❑ Warmińsko-mazurskie 24 173 km<sup>2</sup>
- ❑ Lubelskie 25 133 km<sup>2</sup>

# Percepcja → wykorzystanie koloru



## Fisher's Iris data



# “Infoviz – Toolbox” podstawowe elementy

---

## Perceptual Coding

|               |  |
|---------------|--|
| Position      |  |
| Size          |  |
| Orientation   |  |
| Texture       |  |
| Shape         |  |
| Color         |  |
| Shading       |  |
| Depth Cues    |  |
| Surface       |  |
| Motion        |  |
| Stereo        |  |
| Proximity     |  |
| Similarity    |  |
| Continuity    |  |
| Connectedness |  |
| Closure       |  |
| Containment   |  |

## Interaction

|                        |  |
|------------------------|--|
| Direct Manipulation    |  |
| Immediate Feedback     |  |
| Linked Displays        |  |
| Animate Shift of Focus |  |
| Dynamic Sliders        |  |
| Semantic Zoom          |  |
| Focus+Context          |  |
| Details-on-Demand      |  |
| Output → Input         |  |

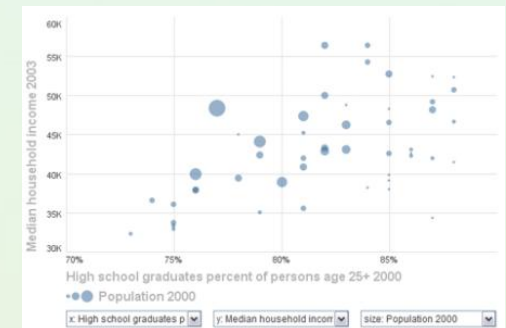
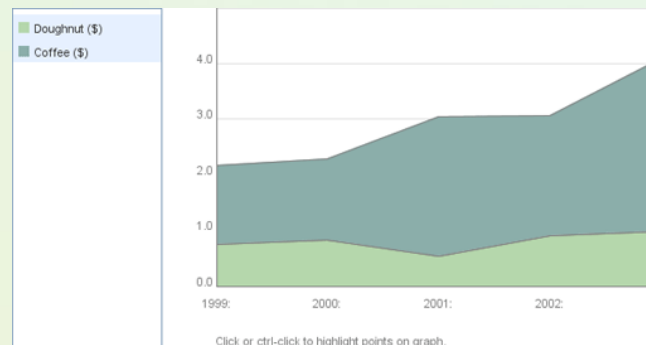
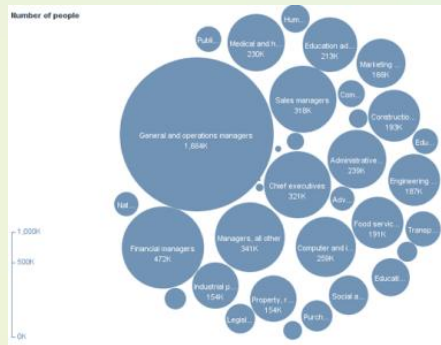
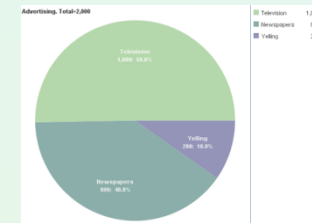
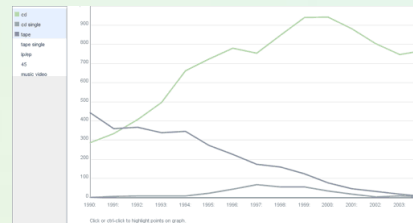
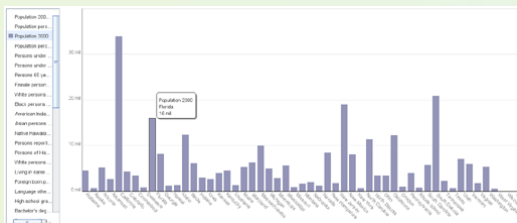
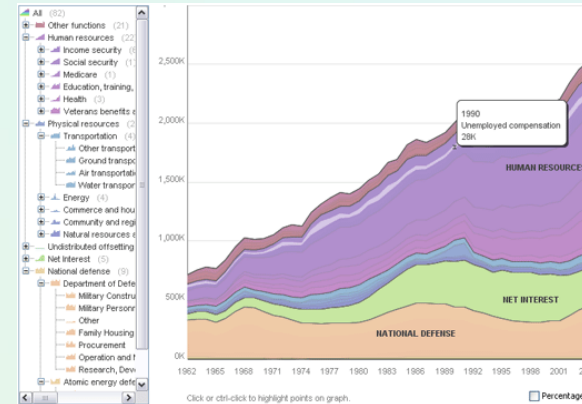
## Information Density

|                         |  |
|-------------------------|--|
| Maximize Data-Ink Ratio |  |
| Maximize Data Density   |  |
| Minimize Lie factor     |  |



# Standardowe grafiki statystyczne

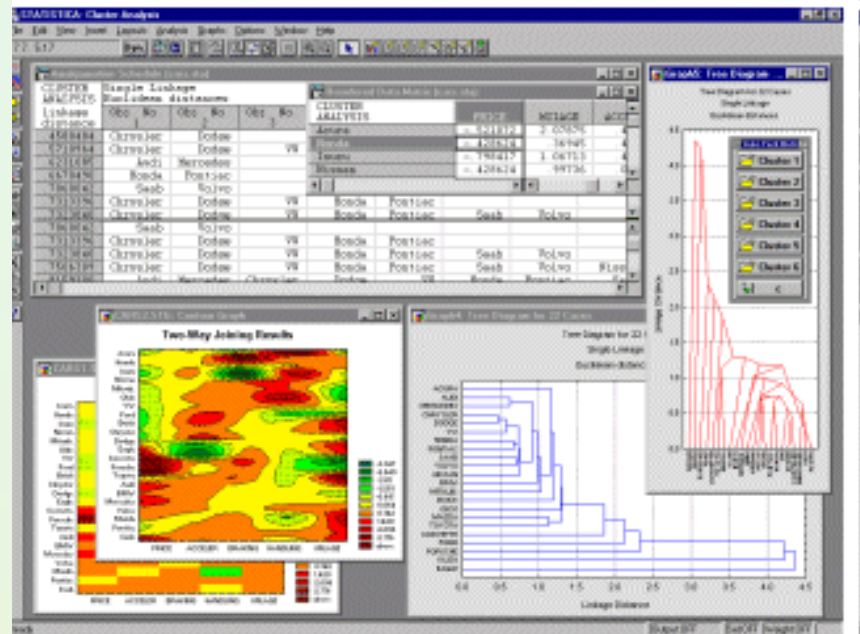
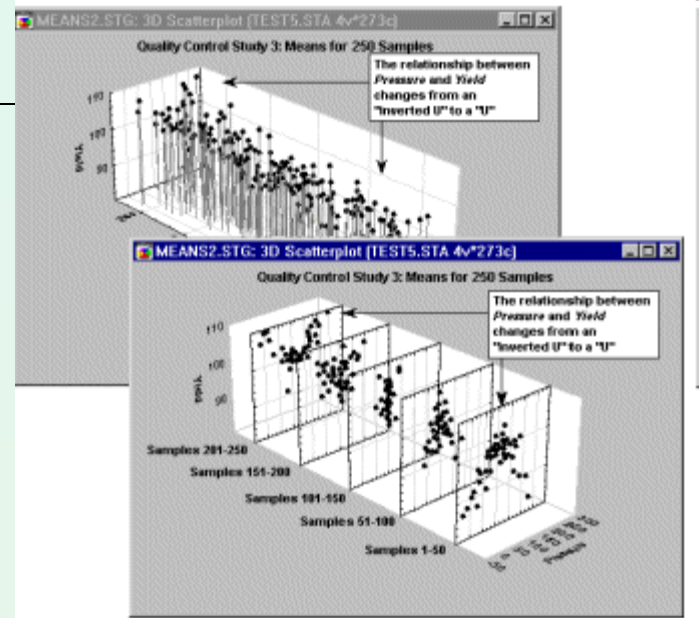
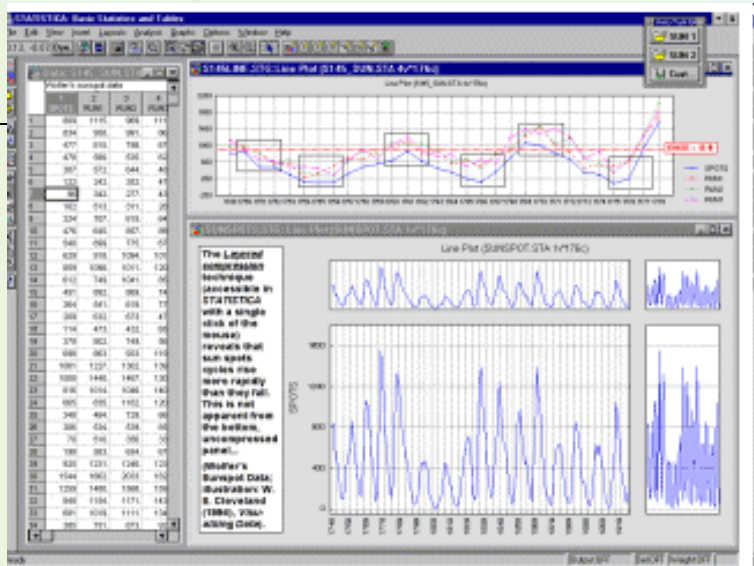
- Bar Charts 
- Line Charts 
- Pie Charts 
- Bubble Charts 
- Stacked Charts 
- Scatterplots 



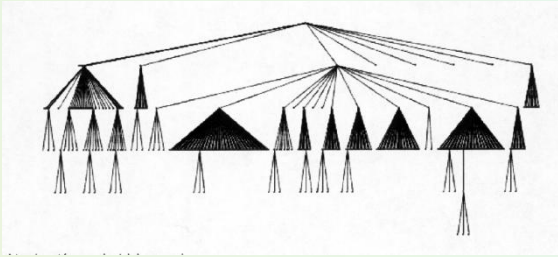
# Różne rodzaje danych i wykresów

---

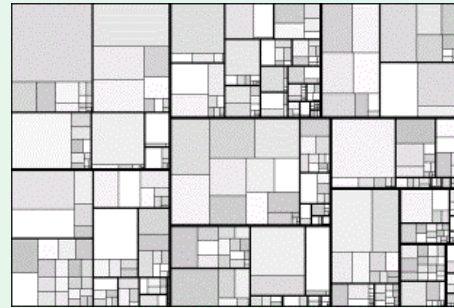
- ❑ Typowe skale, klasyczne statystyczne wykresy
  - Oparte na typowej 1-, 2- lub 3-D geometrii
  - 3D (shaps) dziedziczące z wizualizacji naukowej
- ❑ Abstract, N-wymiarowe specjalizowane
  - Challenge of creating intuitive mapping
  - **Scatterplot and Dimensional Stacking**
  - **Glyphs**
    - Chernoff Faces
  - **Parallel Coordinates** and **Table Lens**
    - **Brushing**
  - **Hierarchies:** Treemaps, Brain, Hyperbolic Tree
  - **Graphs** (Networks)



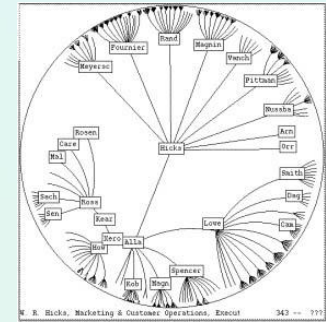
# Abstract → Hierarchical Information – Przykłady



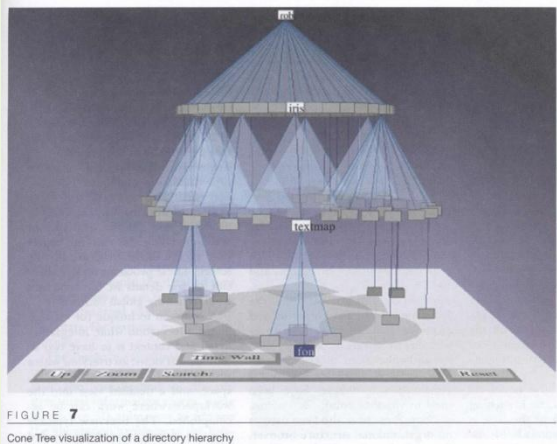
Traditional



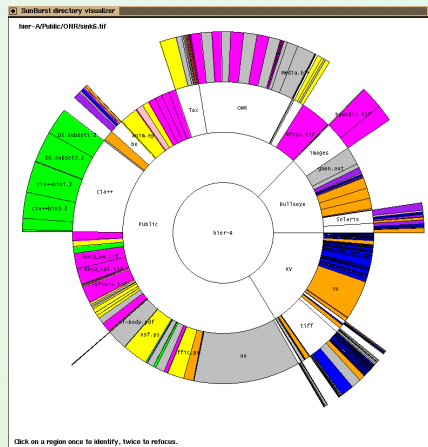
Treemap



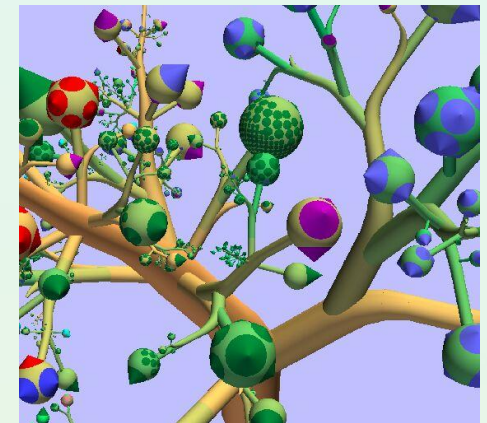
Hyperbolic Tree



ConeTree



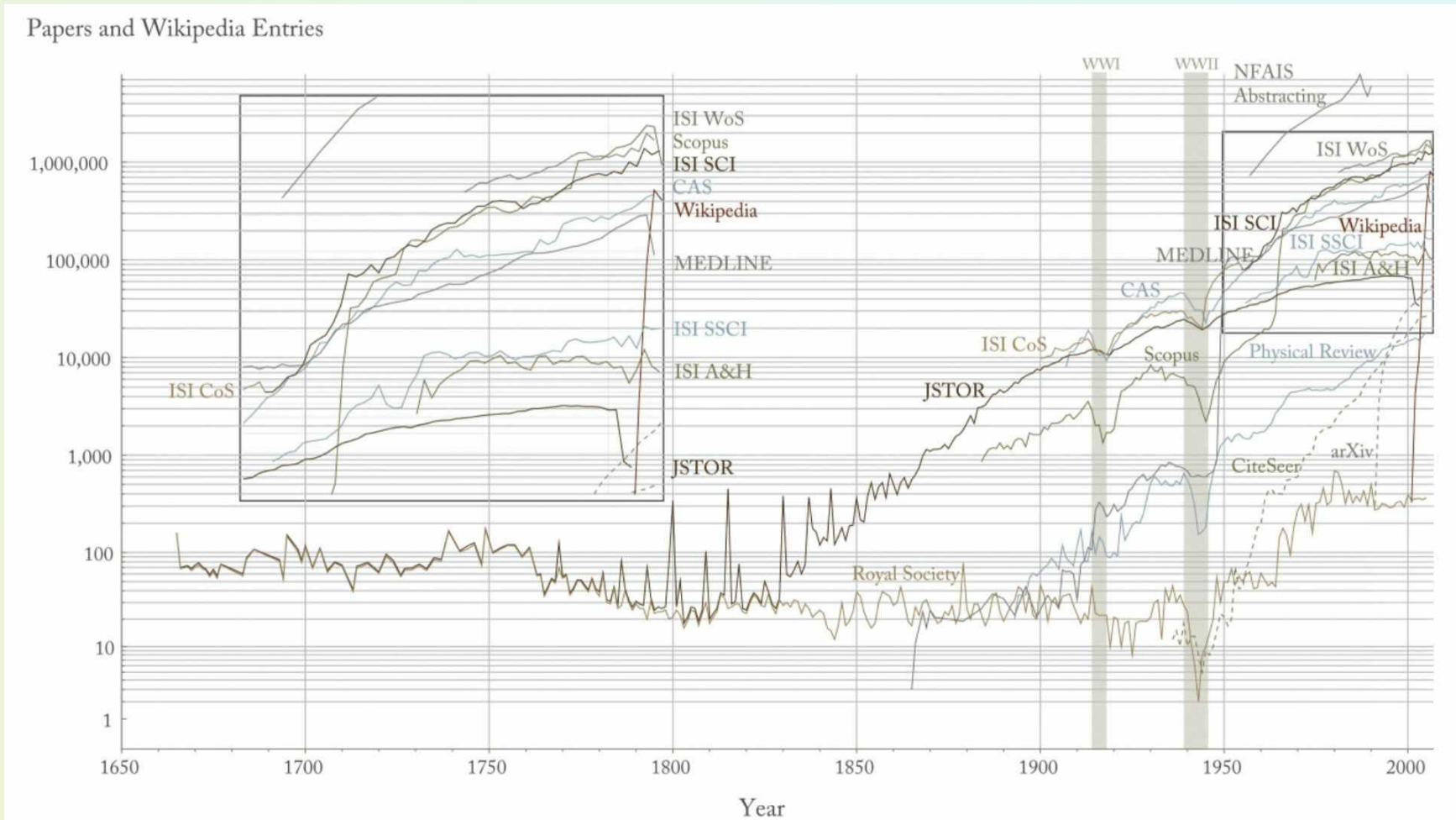
SunTree



Botanical

# Grafika a wyszukiwanie informacji

## ❑ Analiza bibliometryczna literatury naukowej



# Kto Nas czyta?



**Jerzy Stefanowski**

Poznan University of Technology  
 machine learning, data mining, data streams, rule induction, ensembles  
 Verified email at cs.put.poznan.pl - Homepage



Title 1-20 Cited by Year

Incomplete information tables and rough classification  
 J Stefanowski, A Tsoukiàs 382 2001  
 Computational Intelligence 17 (3), 545-566

On the extension of rough sets under incomplete information  
 J Stefanowski, A Tsoukiàs 348 1999  
 New Directions in Rough Sets, Data Mining, and Granular-Soft Computing, 73-81

On rough set based approaches to induction of decision rules  
 J Stefanowski 264 1998  
 Rough sets in knowledge discovery 1 (1), 500-529

Lingo: Search results clustering algorithm based on singular value decomposition  
 S Osiński, J Stefanowski, D Weiss 234 2004  
 Intelligent information processing and web mining, 359-368

ROSE-software implementation of the rough set theory  
 B Predki, R Słowiński, J Stefanowski, R Susmaga, S Wilk 154 1998  
 Rough Sets and Current Trends in Computing, 605-608

Variable consistency model of dominance-based rough sets approach  
 S Greco, B Matarazzo, R Słowiński, J Stefanowski 142 2001  
 Rough Sets and Current Trends in Computing, 170-181

An algorithm for induction of decision rules consistent with the dominance principle  
 S Greco, B Matarazzo, R Słowiński, J Stefanowski 139 2001  
 Rough sets and current trends in computing, 304-313

Rough classification in incomplete information systems  
 R Słowiński, J Stefanowski 134 1989  
 Mathematical and Computer Modelling 12 (10), 1347-1357

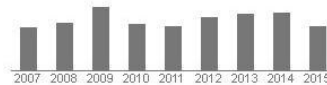
Rough classification with valued closeness relation  
 R Słowiński, J Stefanowski 113 1994  
 New approaches in classification and data analysis, 482-489

Evaluation of vibroacoustic diagnostic symptoms by means of the rough sets

Google Scholar

Get my own profile

| Citation indices | All  | Since 2010 |
|------------------|------|------------|
| Citations        | 4416 | 1952       |
| h-index          | 34   | 23         |
| i10-index        | 76   | 52         |



Scopus

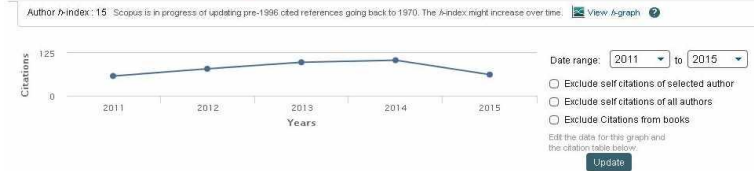
Co-authors View all...

- Roman Słowiński
- Szymon Wilk
- Dawid Weiss
- Salvatore Greco
- Dariusz Brzezinski
- Jerzy Błaszczyński
- Robert Susmaga
- Magdalena Deckert
- Krzysztof Krawiec
- Stanisław Osinski
- Radostaw Z. Ziemiński
- Sławomir Nowaczyk
- Tomasz Maciejewski
- Francisco Herrera
- Julián Luengo Martín
- Mark Last
- Georg Krempel
- Eyke Hüllermeier
- Vincent Lemaire
- Masahiro Inuiguchi

Search Alerts My list

Citation overview This is an overview of citations for this author

67 Cited Documents from "Stefanowski, Jerzy"  
 Author ID:6701756683 Back to author details Save these documents to My list



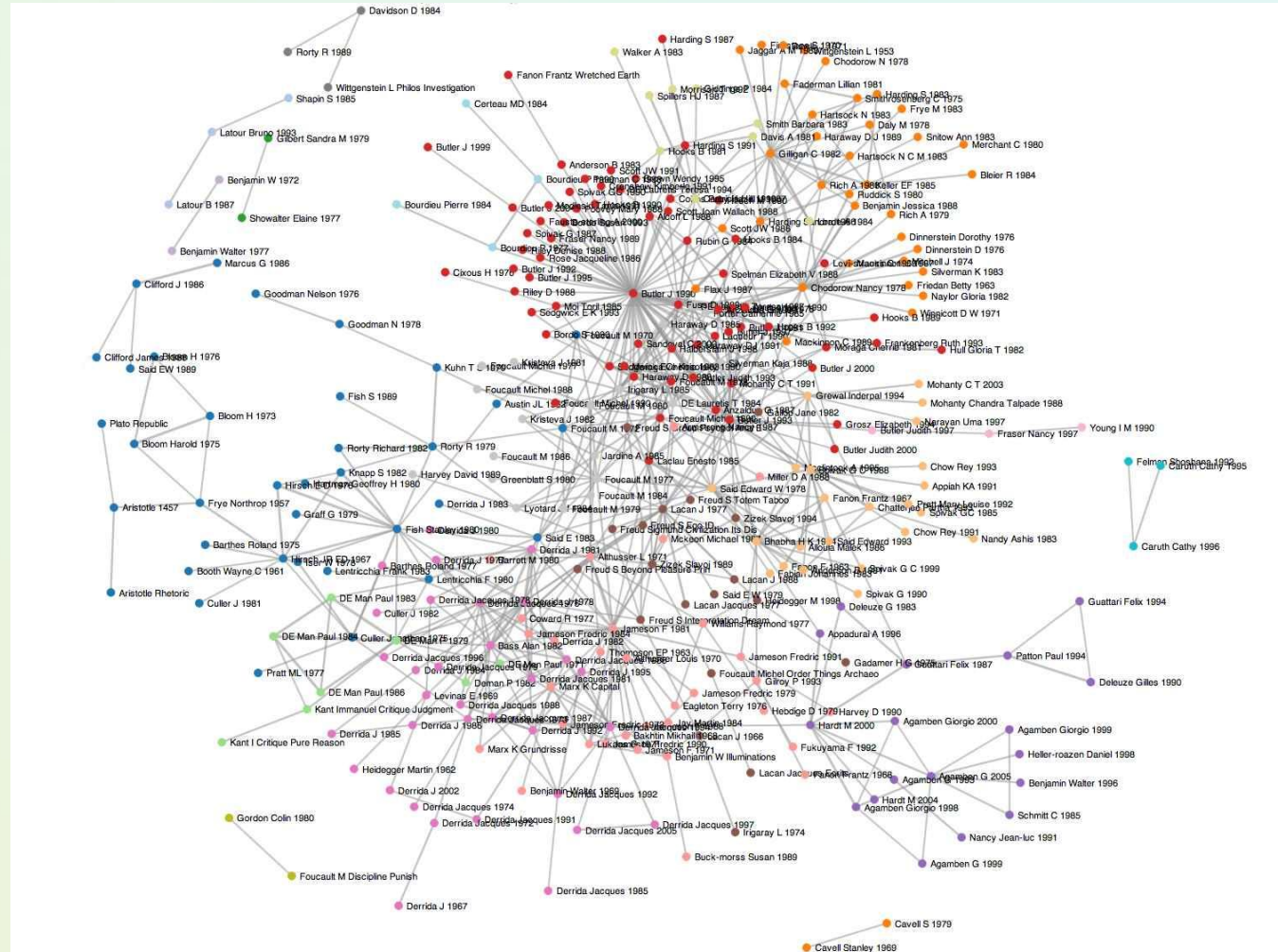
Documents

Citations

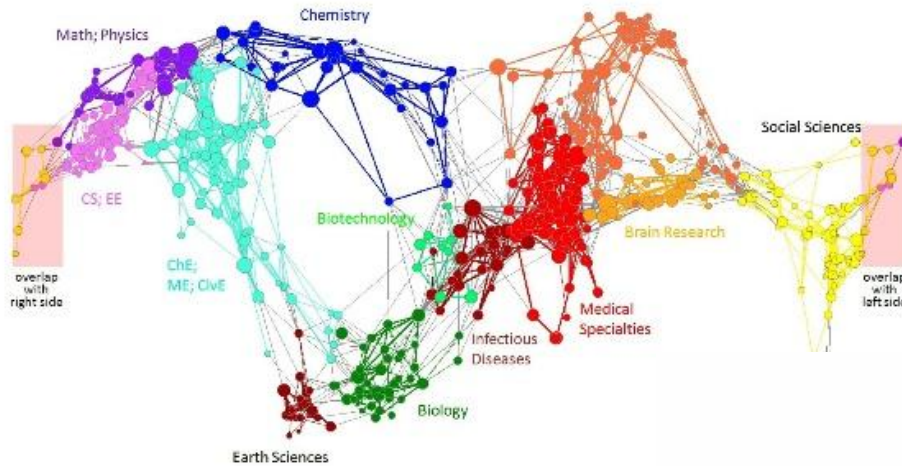
| Sort on: Date (newest) Citation count (descending) ...             | Total | Citations |      |      |      |      |      |    | Subtotal | >2015 | Total |
|--|-------|-----------|------|------|------|------|------|----|----------|-------|-------|
|  |       | >2011     | 2011 | 2012 | 2013 | 2014 | 2015 |    |          |       |       |
| 1 Incomplete information tables and rough classification           | 2001  | 124       | 17   | 23   | 26   | 18   | 2    | 86 | 210      |       |       |
| 2 Rough classification in incomplete information systems           | 1989  | 60        | 4    | 4    | 3    | 7    |      | 18 | 78       |       |       |
| 3 Evaluation of vibroacoustic diagnostic symptoms by means of ...  | 1992  | 44        | 2    | 2    | 3    |      | 1    | 8  | 52       |       |       |
| 4 Three discretization methods for rule induction                  | 2001  | 33        | 2    | 1    | 9    | 3    | 3    | 18 | 51       |       |       |
| 5 Rough-set reasoning about uncertain data                         | 1996  | 37        | 2    | 2    | 1    | 3    | 1    | 9  | 46       |       |       |
| 6 Induction of decision rules in classification and discovery...   | 2001  | 23        | 4    | 2    | 4    | 4    |      | 14 | 37       |       |       |
| 7 Rough sets approach to analysis of data from peritoneal lava...  | 1988  | 28        | 3    |      | 1    | 3    | 1    | 8  | 36       |       |       |
| 8 A comparison of two approaches to data mining from imbalance...  | 2005  | 12        | 2    | 7    | 6    | 3    |      | 18 | 31       | 1     |       |
| 9 Rough set based processing of inconsistent information in de...  | 2000  | 17        | 4    | 3    |      | 1    | 1    | 9  | 26       |       |       |
| 10 On combined classifiers, rule induction and rough sets          | 2007  | 8         | 2    | 4    | 5    | 2    | 1    | 14 | 22       |       |       |
| 11 Reacting to different types of concept drift. The accuracy u... | 2014  |           |      |      |      | 12   | 9    | 21 | 21       |       |       |
| 12 Feature subset selection for classification of histological ... | 1997  | 15        | 2    | 2    | 1    | 1    |      | 6  | 21       |       |       |
| 13 Local neighbourhood extension of SMOTE for mining imbalanced... | 2011  |           |      | 1    | 4    | 5    | 5    | 15 | 16       | 1     |       |
| 14 Rule induction via clustering decision classes                  | 2008  | 9         | 2    | 2    | 2    | 1    |      | 7  | 16       |       |       |
| 15 Incremental versus non-incremental rule induction for multic... | 2004  | 11        | 1    |      | 3    |      | 1    | 5  | 16       |       |       |
| 16 Interactive relation processing of imbalanced data with         | 2010  |           | 1    | 3    | 7    | 3    | 3    | 14 | 14       | 44    |       |

# Citations? Grafiki w prezentacji dokumentów

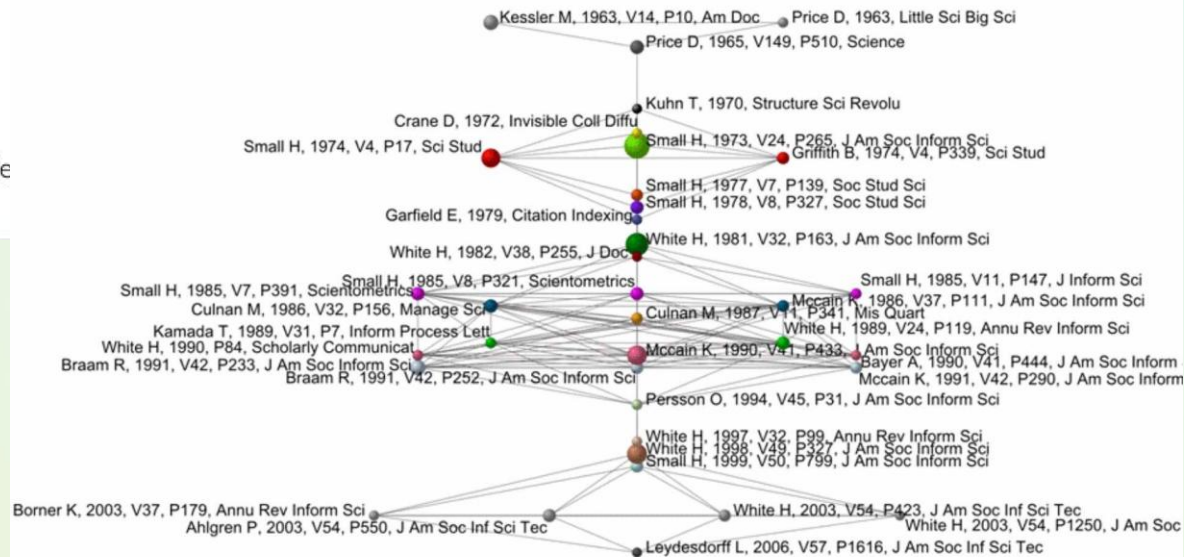
- Wykorzystaj dane (Internet, wyszukiwarki, wydawnictwa)
- Citation graphs



# Mapy rozwoju nauki i inspiracji ....



Citation networks and Maps of science  
[Börner et al., 2012]



Za wykładem Veslava Osinska



# Wizualizacja słów kluczowych

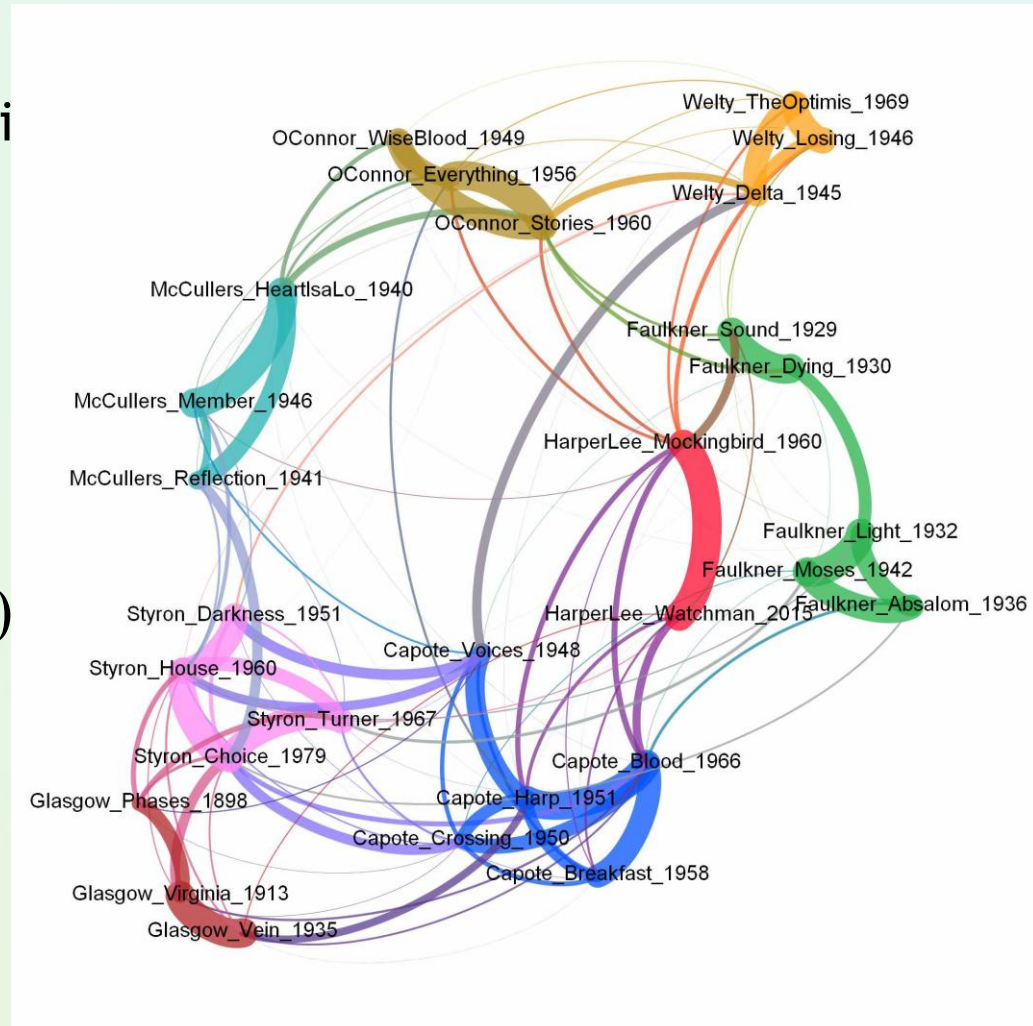
## □ “Clouds of words”



Przykładowe oprogramowanie → Wordle; TagCrowd; Word It Out ; ..

# Analiza podobieństwa tekstów amerykańskich pisarzy

- ❑ NLP, powiązania i inspiracje
- ❑ “Śledztwo literackie” - M.Eder i J.Rybicki 2015 - poszukiwanie podobieństwa stylu pisania słynnych książek
- ❑ Opis oceny autorstwa książki Harper Lee
  - Zabić drozda (1960)
  - “Go Set a Watchman” (2015)
  - Czy jest autorką, czy widać “odciski” innych osób w stylu tekstu?
- ❑ Więcej - artykuł Gazeta Wyborcza (31 07 2015)



Spójrz na blog BIQDATA.pl

# Infografika

---

- ❑ Obraz łączący w sobie dane z wzornictwem graficznym, który służy do zwięzłego przekazywanie odbiorcą komunikatów

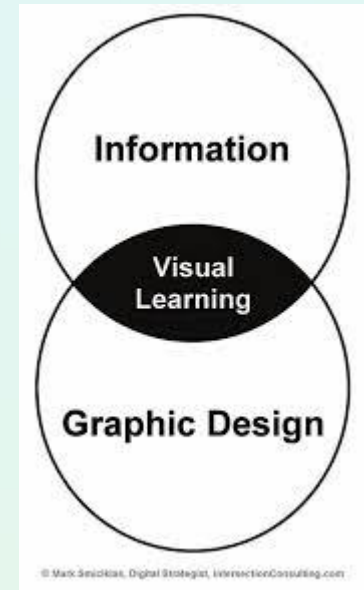
Mark Smiciklas

- ❑ Grafiki objaśniające [Nigel Holmes]

- ❑ Wizualne przyswajanie informacji - korzyści

- Większa zrozumiałość informacji, pojęć, pomysłów
- Rozwinięcie myślenia krytycznego oraz zdolności do opracowywania i organizowania pomysłów
- Sprawniejsze zapamiętywanie i przywoływanie wiedzy

Spójrz do: Mark Smiciklas: Inforgrafiki - praktyczne zastosowanie w biznesie 2012 lub jego blog



# Iconographics



- In 2000, human attention span equaled 13 seconds
- In 2013, attention span decreased to 8 seconds



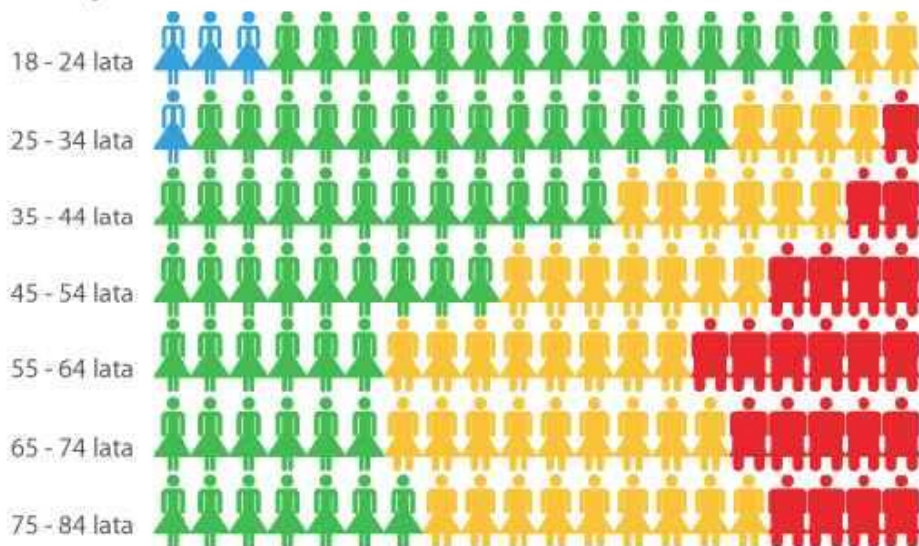
44% of people find it hard to focus at school or at work



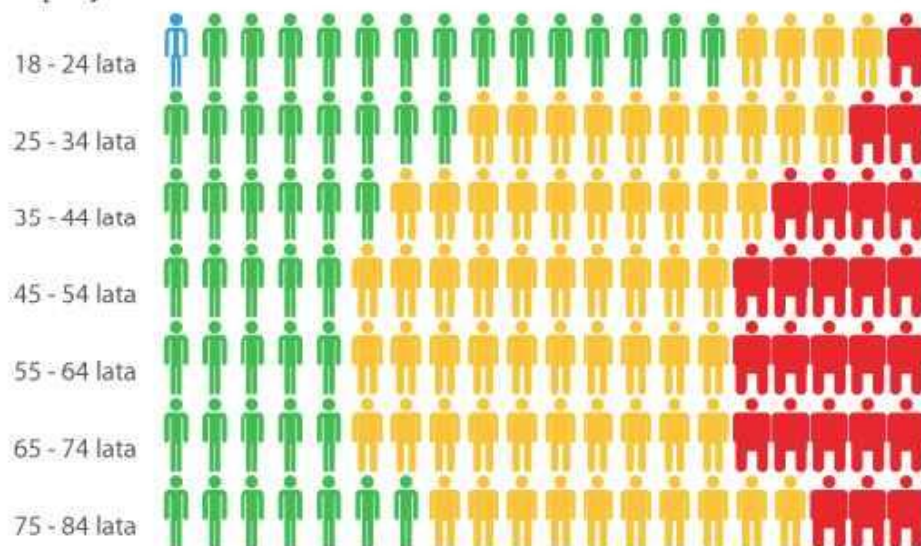
# Wizualne symbole zgodnie z regułami języka ISOTYPE

## BMI a wiek - Polacy w 2008 r.

### kobiety



### mężczyźni



5%  
niedowaga  
waga prawidłowa  
nadwaga  
otyłość

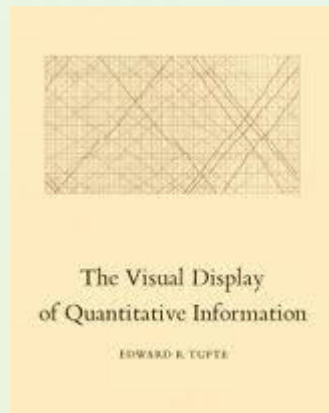
*Udział osób z niedowagą lub nadwagą w różnych grupach wiekowych: autorstwa Magdy Małczyńskiej-Umedy - Młodzi XXL, Eurostat a problem nadwagi i otyłości*

Język ISOTYPE - zasady edukacyjnej prezentacji, w tym listy precyzyjnie zaprojektowanych znaków/piktogramów o uniwersalnych znaczeniach oraz reguły czytelnego i dokładnego używania tych piktogramów w celu prezentowania informacji

# W stronę najlepszego projektu

---

- ❑ Wizualizacja danych to przecięcie sztuki, statystyki i twórczego projektu
- ❑ Wyznaczniki dobrej wizualizacji  
→ zasady Edwarda Tuftego
- ❑ Jesteśmy inspirowani ....



- ❑ Spójrz także [www.edwardtufte.com](http://www.edwardtufte.com)

# Książki i materiały

---

## ❑ Polskojęzyczne:

- **Przemysław Biecek: Odkrywać! Ujawniać! Objasniać! 2014**
- Także blogi <http://biecek.pl/Eseje/>  
oraz V. Osińska <http://www.wizualizacjainformacji.pl/>

## ❑ Klasyczne:

- **Edward Tufte: Visual Display of Quantitative Information**
- **Stephen Few: Show Me the Numbers; Now You See It**
- Naomi Robbins: Creating More Effective Graphs
- Howard Wainer: Graphic Discovery
- William Cleveland: Visualizing Data

# Cele przedmiotu

---

Zainteresować tematyką wizualizacji informacji

Uświadomić wartość wizualizacji danych ilościowych (np. inżynierskich) ...

... lecz także nie liczbowych (np. z sieci, tekstów)

Przedstawić podstawowe elementy grafiki informacyjnej i zasad ich stosowania

Zainspirować zasadami projektowania graficznego przekazu informacyjnego (Edward Tufte)

Stworzyć praktyczny mini-projekt wizualizacji danych



# Trzeba coś zrobić,....

---

- ❑ Uczestniczyć w zajęciach
- ❑ Projekty - nauczymy się czegoś praktycznego
  - Wykonać zespołowy mini-projekt wizualizacji danych
  - Studium literaturowe (poczytajmy, oglądajmy) - przygotować podsumowanie (esej) i “porywającą” prezentację
  - Dokonać analizy / porównania / oceny oprogramowanie i / lub bibliotek programistycznych do wizualizacji informacji
- ❑ Być aktywnym, zainteresować się, ...
- ❑ To have some fun ....

# Plan dalszych wykładów

---



- ❑ Percepcja (kolor, dane)
  - Możliwe błędy i przekłamania wizualizacji
  - Dobór środków prezentacji do możliwości poznawczych
- ❑ “Praktyczna kuchnia”
  - Podstawowe środki grafiki informacyjnej
  - Jak poprawnie konstruować wykresy
- ❑ Metody wizualizacji złożonych danych wielowymiarowych
- ❑ Zasady projektowanie wizualizacji wg. E.Tufte’go
- ❑ Wizualizacyjna eksploracja Big Data
  - Sieci społeczne, grafy, data mining (eksploracja danych)
- ❑ Wyszukiwanie i analiza dokumentów
  - Tekst, tekst,..., powiązania
- ❑ Dobre środki komunikowania informacji

*Koniec tej części*



Poszukuj samodzielnie dodatkowych materiałów!

Wypróbuj różne oprogramowanie