

Információ- és tudásintegrálás: Adatvizualizáció és -integráció

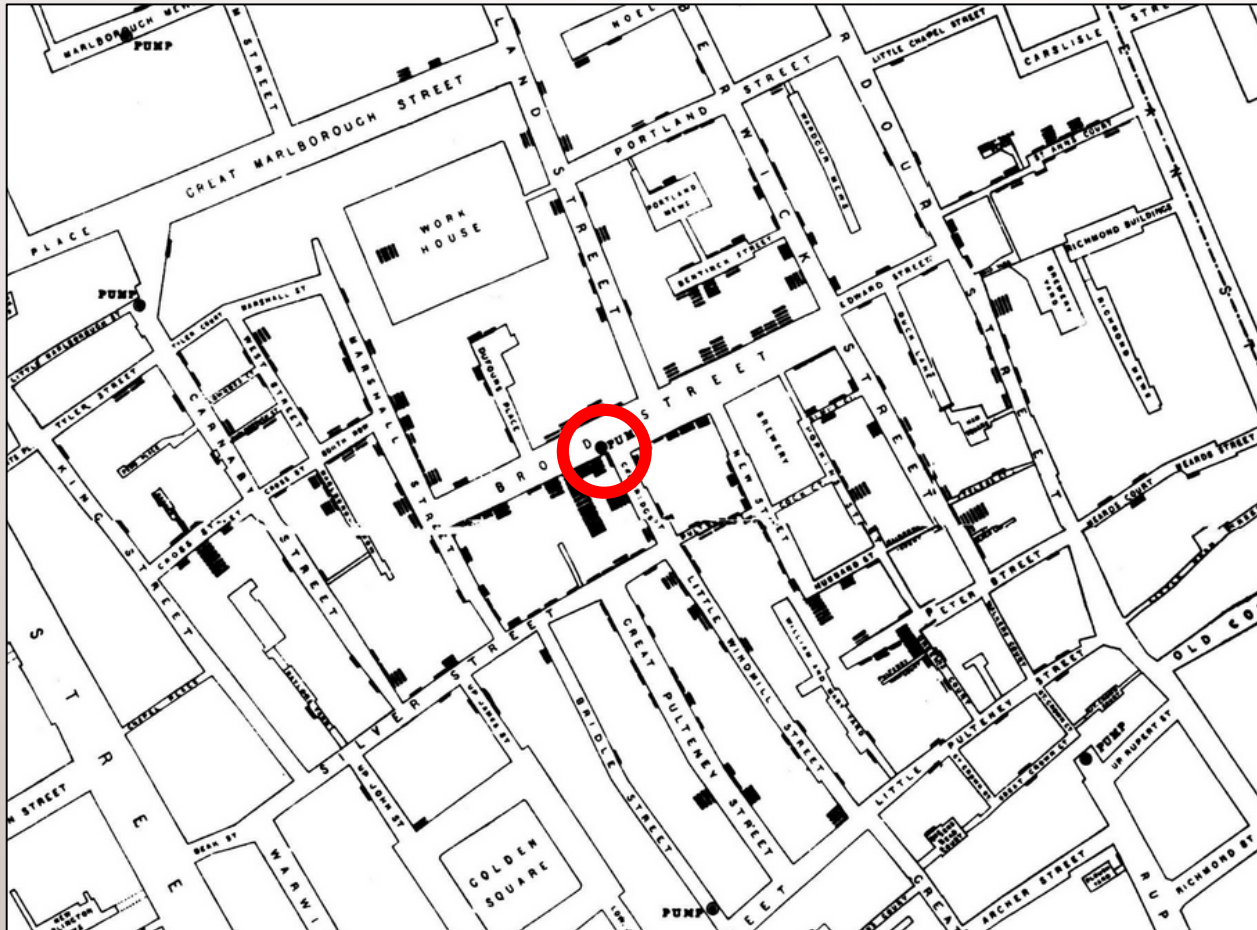


M Ű E G Y E T E M 1 7 8 2

Major Bence

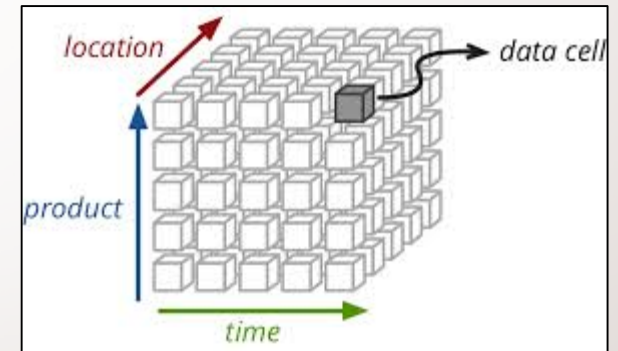
Miért fontos az adatvizualizáció?

- John Snow
 - Londoni kolerajárvány - 1854



Miért fontos az adatvizualizáció? II.

- Nagymennyiségű adatok megjelenítése
 - Üzleti intelligencia
 - Gépi tanulás
 - Bioinformatika
- Korreláció keresése
- Prezentáció (akár laikusoknak is)

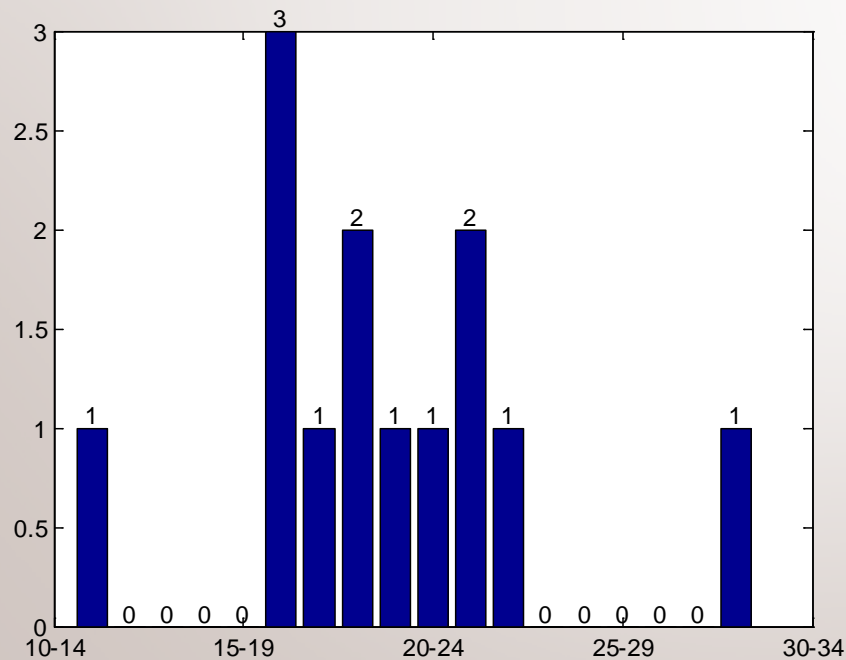




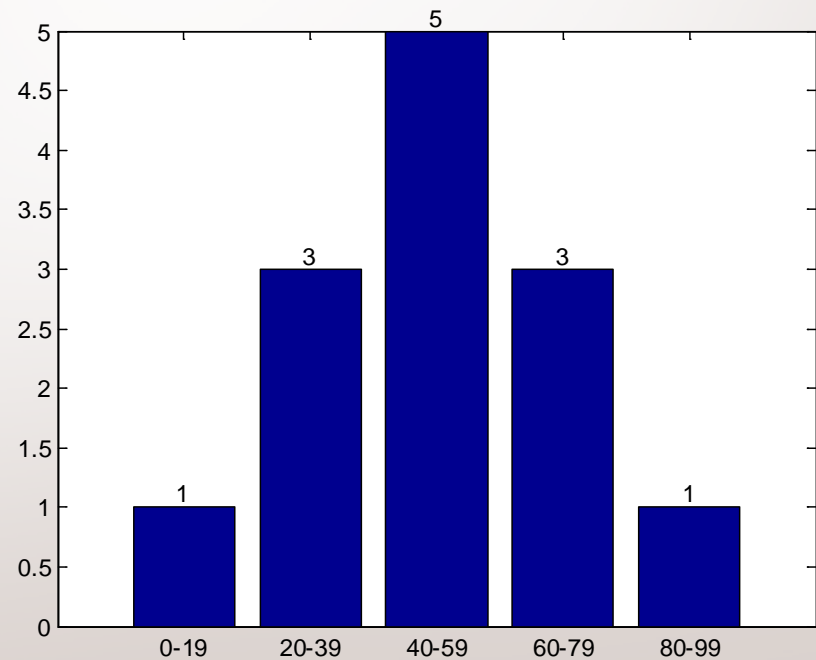
1D - Hisztogram

- Az oszlop magassága hordozza az információt
- Eldöntendő az intervallumok mérete

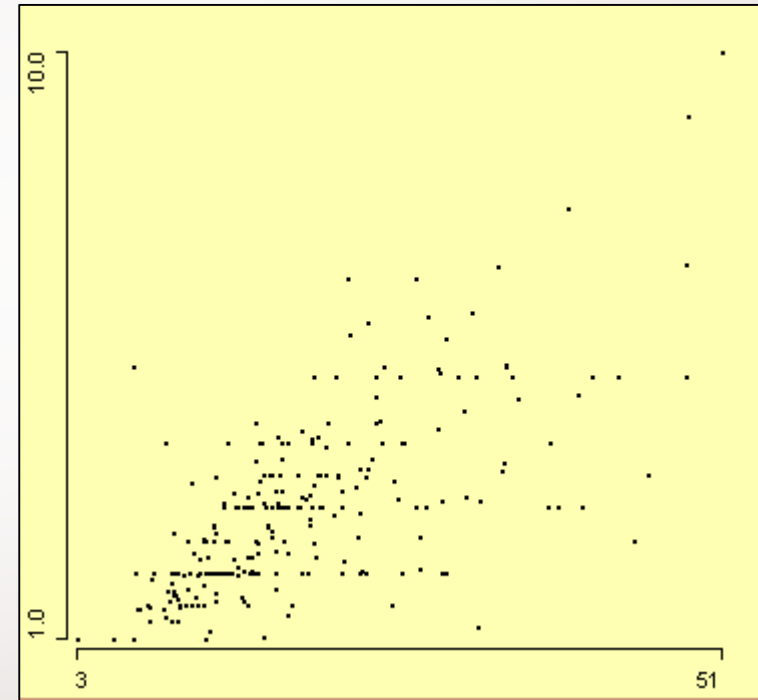
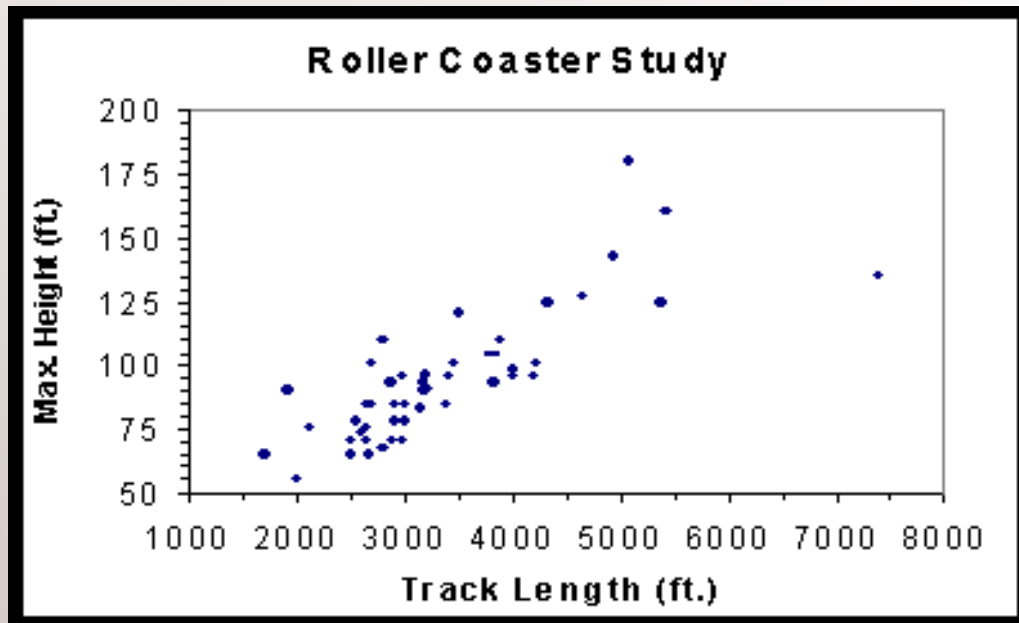
Méret = 5



Méret = 20



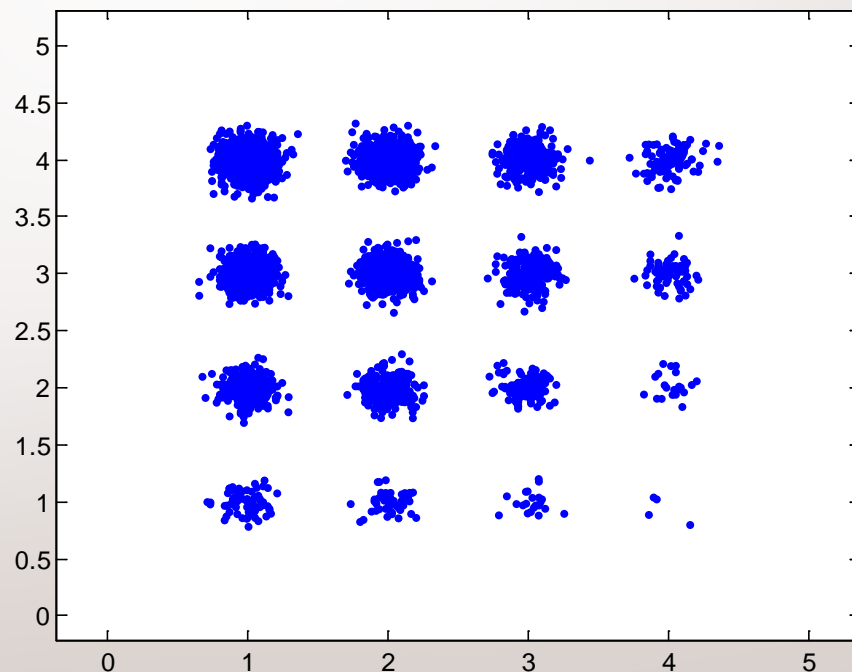
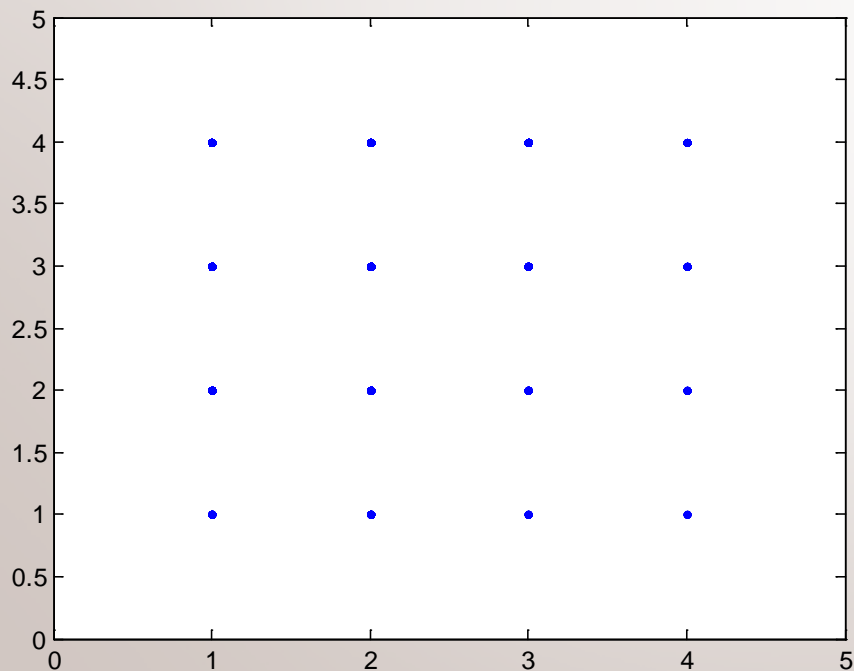
2D- Scatterplot



- Összefüggések keresése
- Problémák: diszkrét értékek kezelése, overplotting

2D – Scatterplot, problémák

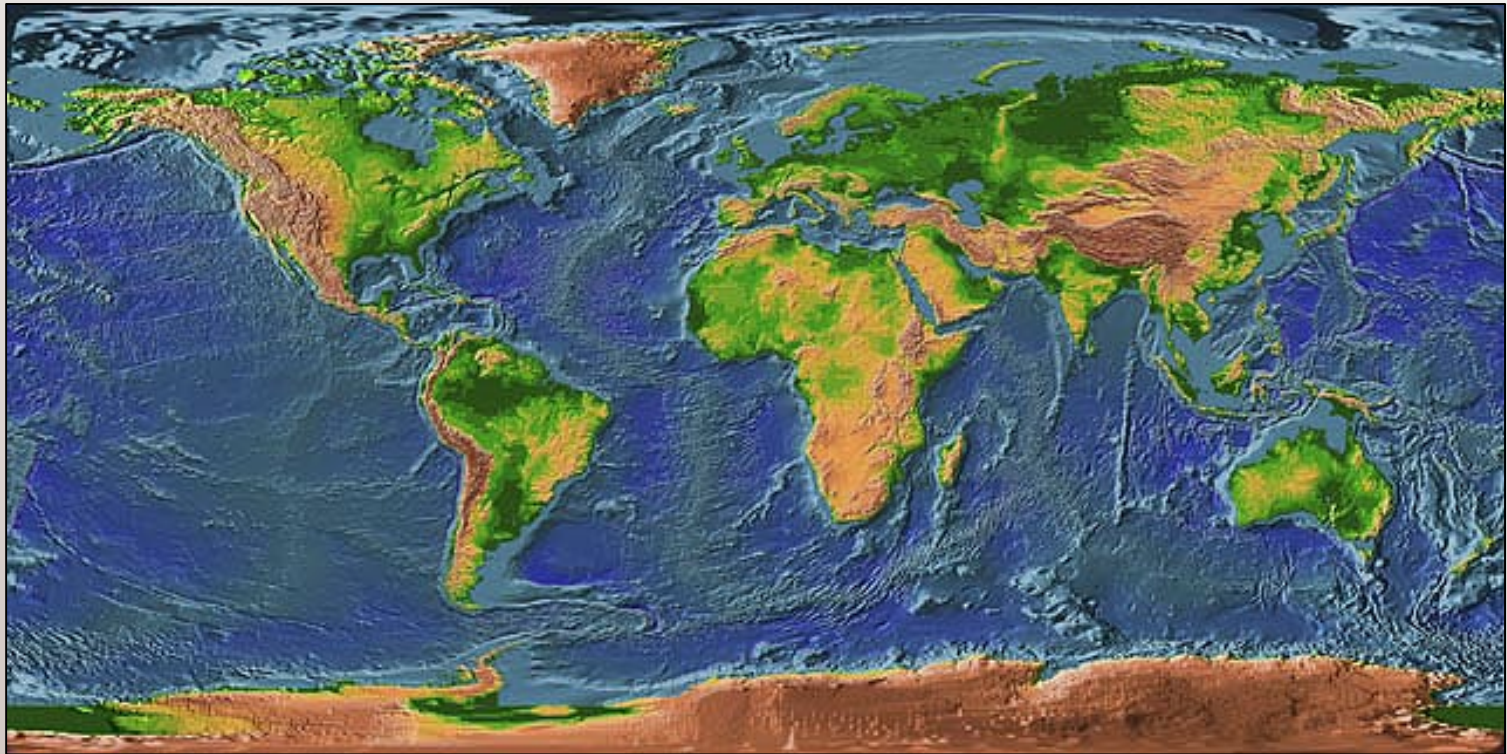
- Jittering
- Demó – Tips, Pollenek



Diszkrét értékek kezelése

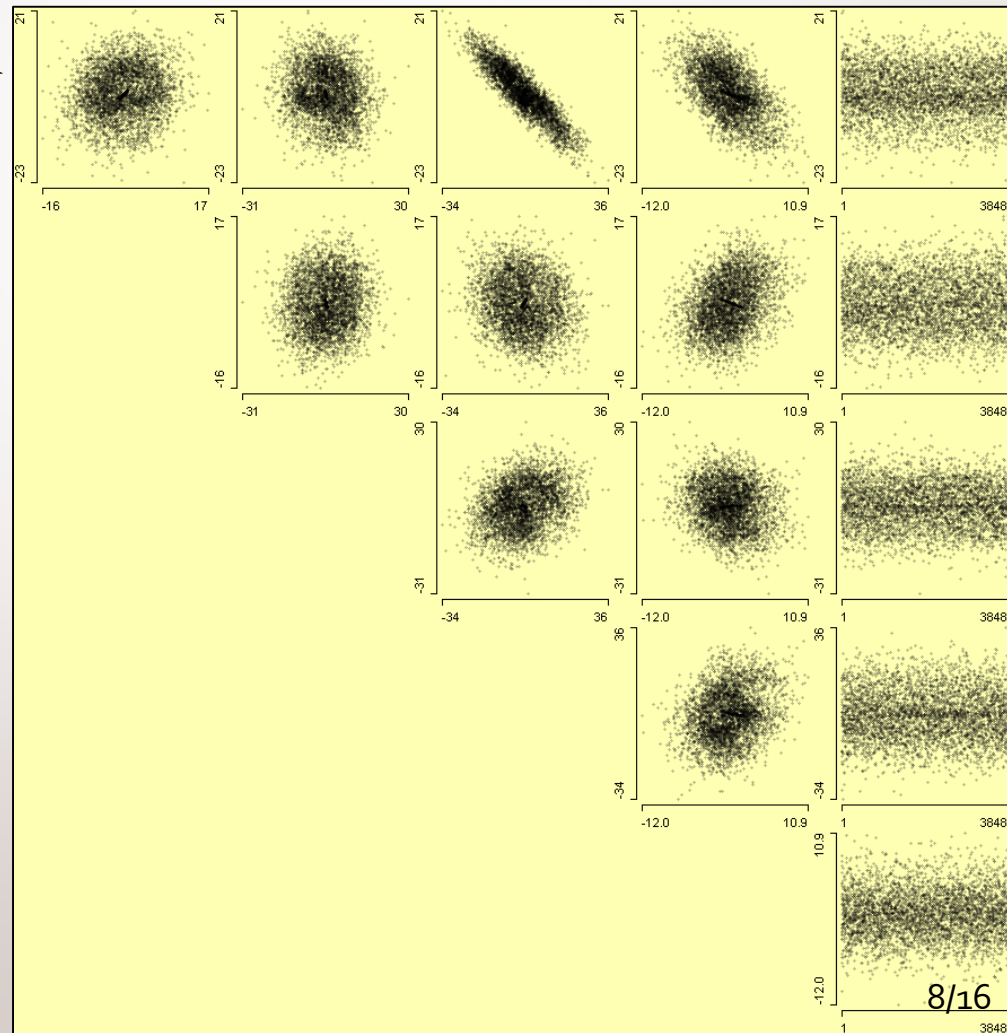
3D

- 3D Scatterplot
- Heat/Intensity map



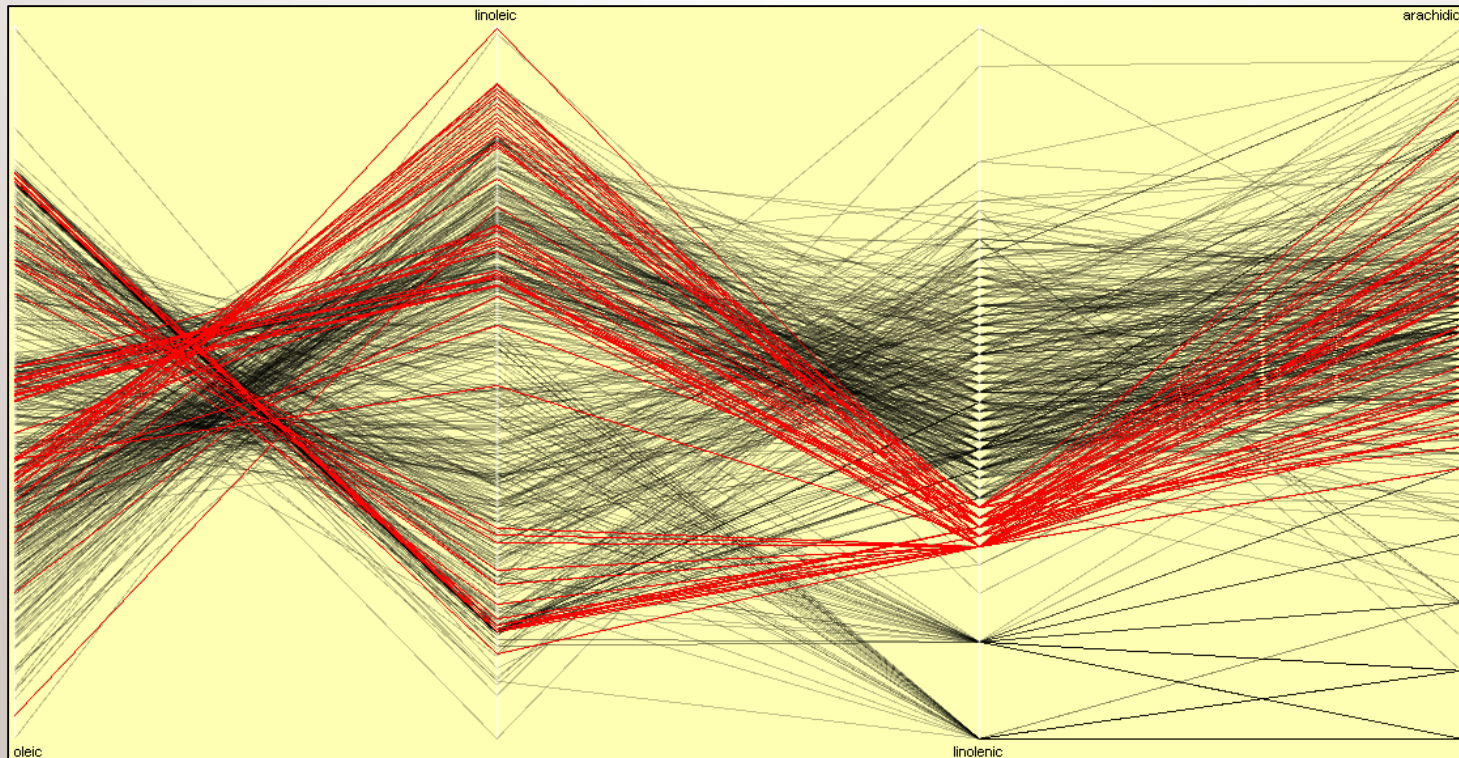
N dimenzió – Scatterplot matrix (SPLOM)

- Minden dimenzópár egy scatterplot
- Korreláció keresése gyorsan
- Demó - Pollenek



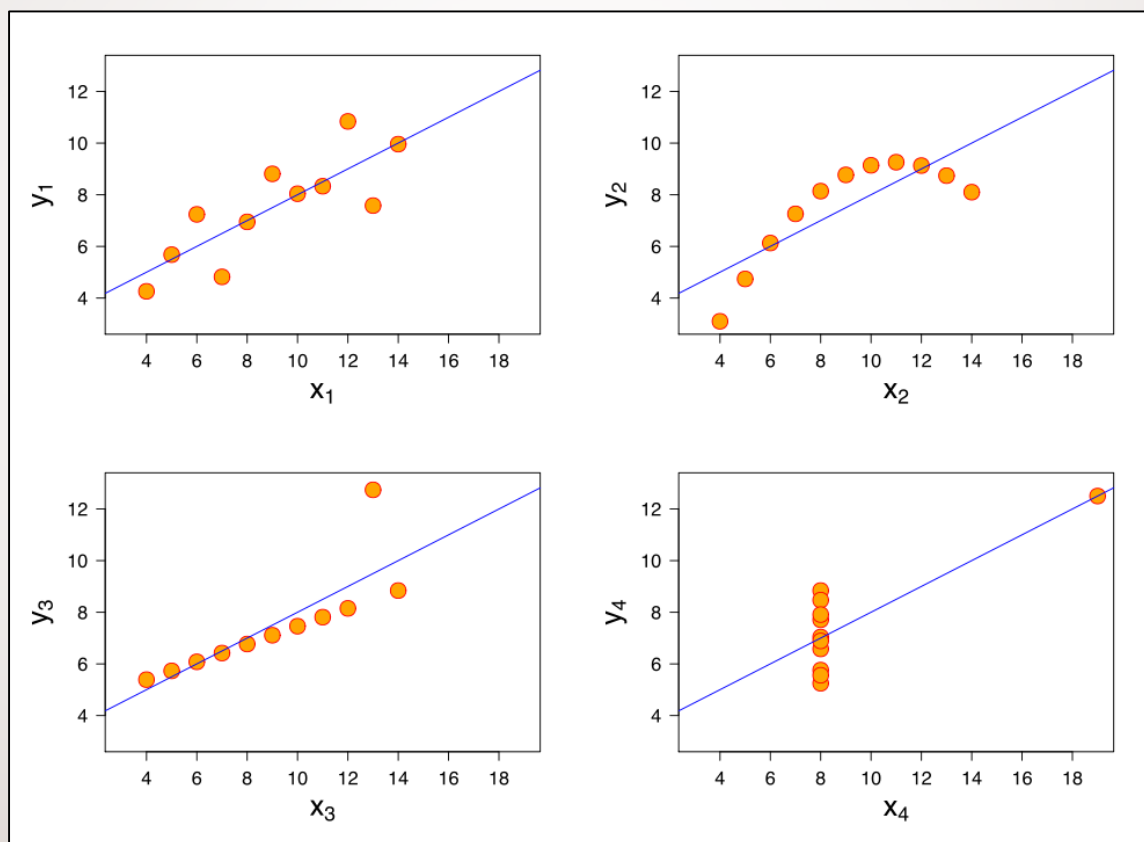
N dimenzió – Párhuzamos koordináták

- Hatékony, áttekinthető megjelenítés
- Korreláció keresése kevésbé egyszerű
- Demók – USA választások, Pollenek



Adathalmaz leírói

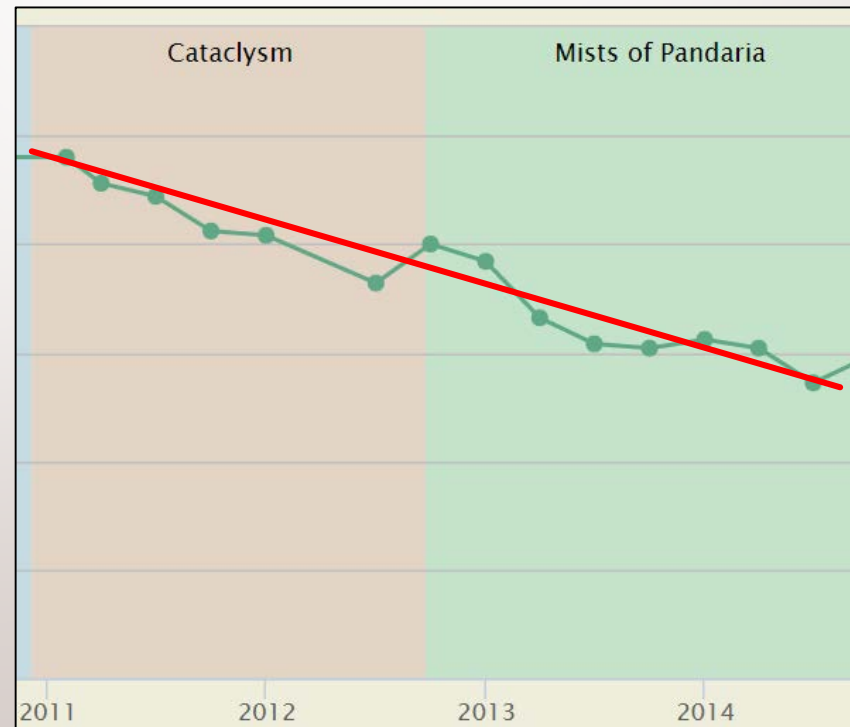
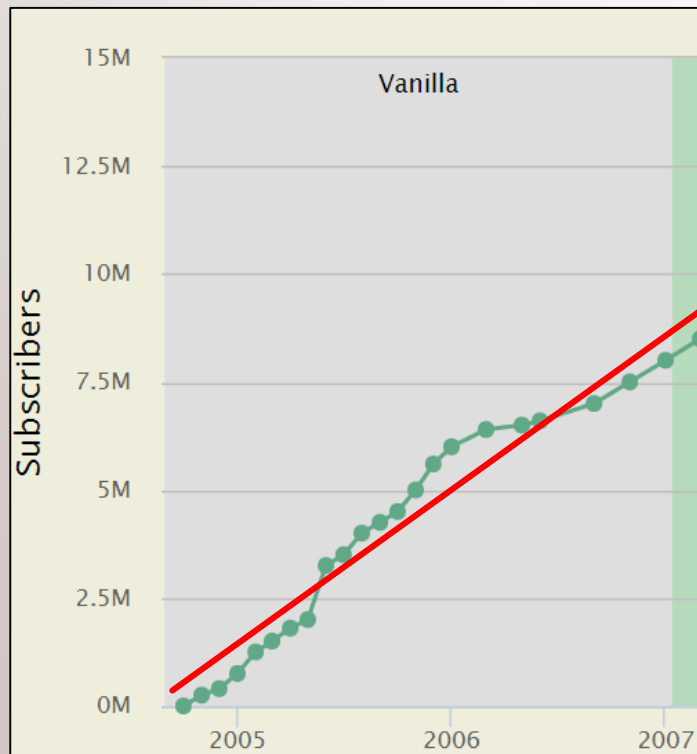
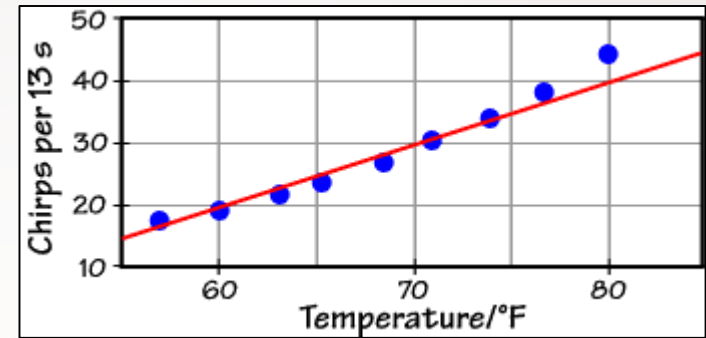
- Várható érték
- Szórás
- Módusz
- Medián



Anscombe's Quartet

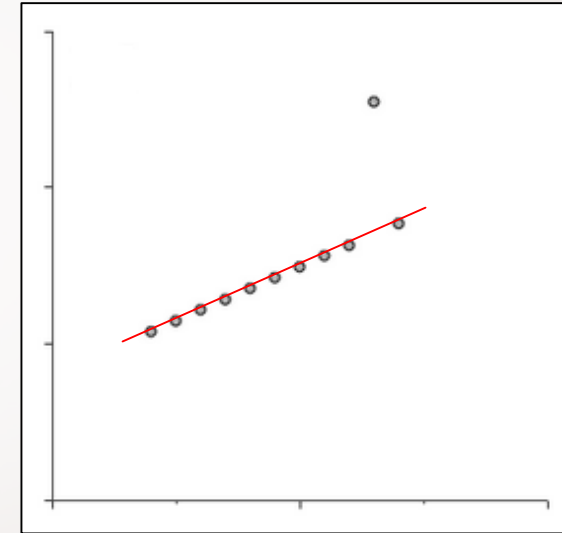
Regresszió

- Összefüggések meghatározására
- Trend meghatározására



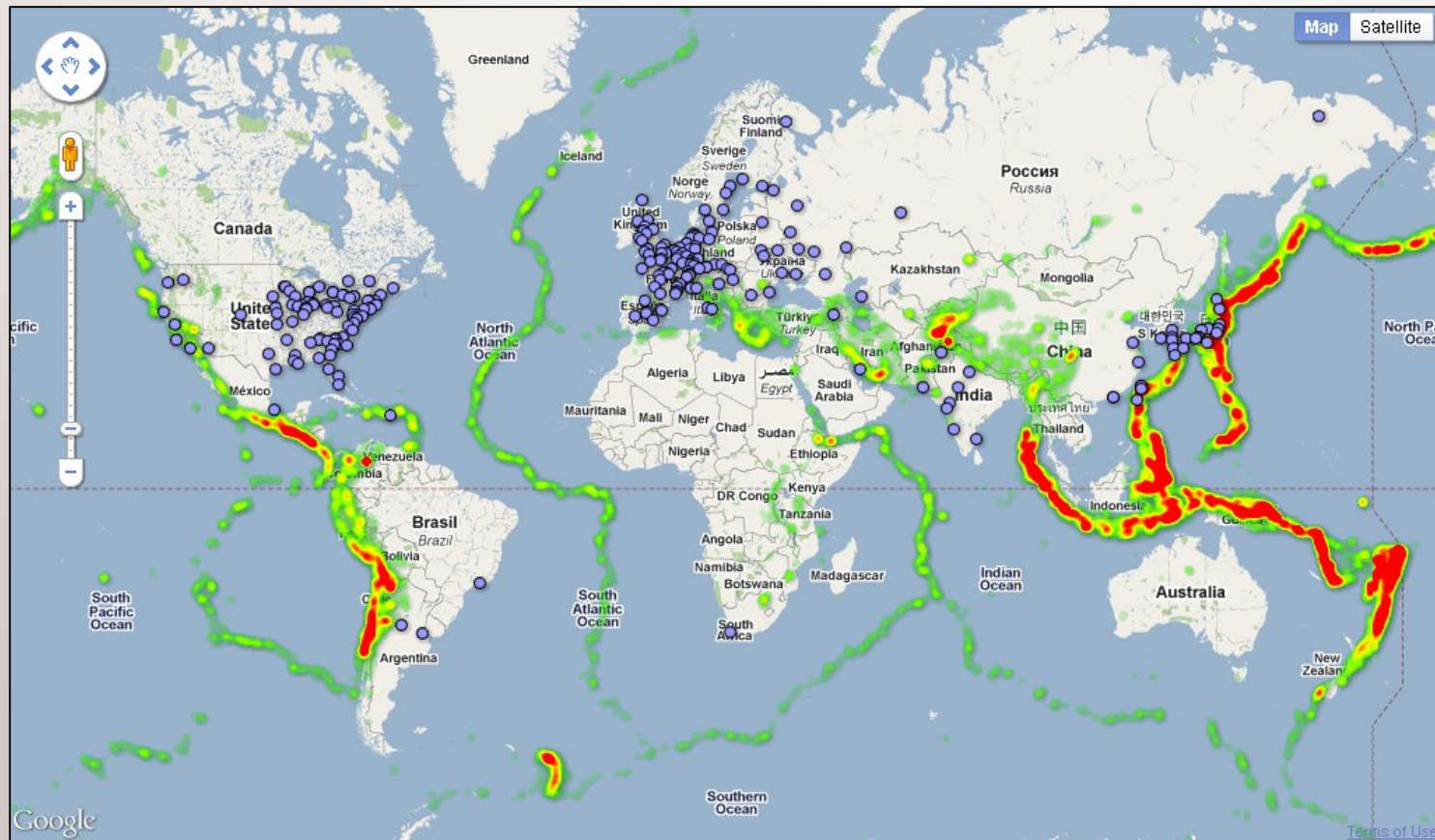
Random Sample Consensus (RANSAC)

- Modell illesztése ponthalmazra
- N : A minimális pontszám, ami egyértelműen meghatározza a modellt
- Algoritmus:
 - Kiválasztunk N véletlen mintapontot
 - Ezekre illesztjük a modellt
 - Megszámoljuk a modellel konzisztens mintapontokat
 - Ha ez nem ér el egy küszöböt, újrakezdjük








Adatintegráció

- Geografikus adatmegjelenítés
- Google Fusion Tables



Adatintegráció – Web I.

Olympiad	Year	Game	Winner	Sources
1st	776 BC	 Stadion	Coroebus of Elis	[2][3][4][5]
2nd	772 BC	 Stadion	Antimachus of Elis (or of Dyspontium)	[2][3][4][5]
3rd	768 BC	 Stadion	Androcles of Messenia (or Androclos)	[2][3][4][5]
4th	764 BC	 Stadion	Polychares of Messenia	[2][3][4][5]
5th	760 BC	 Stadion	Aeschines of Elis	[3][4][5]

Elements in our galaxy	Parts per million by mass
Hydrogen	739,000
Helium	240,000
Oxygen	10,400
Carbon	4,600
Neon	1,340
Iron	1,090
Nitrogen	960
Silicon	650
Magnesium	580
Sulfur	440
Potassium	210
Nickel	100

- Ideális eset
 - Mértékegységek
 - A táblázat tartalma kiderül a fejlécből

Adatintegráció – Web II.

- Egysoros táblázatok
- Kontextus fontossága

Életkor	Korosztály	Hivatalos versenytáv
11 évesig	Újonc	Újonc táv
12–13	Gyermek	Gyermek táv
14–15	Serdülő	Serdülő táv
16–17	Ifjúsági	Sprinttáv
18–19	Junior	Sprinttáv

Beaver
Temporal range: Late **Miocene** – Recent



North American beaver (*Castor canadensis*)

Scientific classification

Kingdom: **Animalia**
Phylum: **Chordata**
Class: **Mammalia**
Order: **Rodentia**
Family: **Castoridae**
Genus: ***Castor***

Linnaeus, 1758

Adatintegráció – Web III.

- Fontos információ a környező szövegben

This is a [list of countries](#) ordered by annual [per capita](#) consumption of [coffee](#). Data are for the year 2008 or earlier. Worldwide production was 7,358,897 metric tons,^[1] so the annual world average was 1.3 kg per person per year.^[1]

[Contents](#) [\[show\]](#)

[Countries](#) [\[edit\]](#)

Rank ↕	Country	↕	Consumption (kg/year) ↕
1	 Finland		12.0 kg
2	 Norway		9.9 kg
3	 Iceland		9.0 kg
4	 Denmark		8.7 kg
5	 Netherlands		8.4 kg
6	 Sweden		8.2 kg
7	 Switzerland		7.9 kg
8	 Belgium		6.8 kg
9	 Canada		6.5 kg
10	 Bosnia and Herzegovina		6.2 kg





Köszönöm a figyelmet!

Felhasznált források I.

- BME VIK Rendszermodellezés diásor (2014)
- Alon Halevy prezentációi (<http://homes.cs.washington.edu/~alon/>)
- Demonstrációs eszköz:
<http://www.rosuda.org/mondrian/Mondrian.html>
- http://webspaceship.edu/mtmars/mis_530/data/datarep.html
- <http://geodesy.noaa.gov/heightmod>
- <http://www.dartmouth.edu/~genchem/0102/spring/6winn/cricket.html>
- <http://www.mmo-champion.com/content/4526-WoW-Up-to-10-Million-Subscribers-Garrison-Campaign-Quests-Active-Nov-19-Hotfixes>
- http://maptd.com/map/earthquake_activity_vs_nuclear_power_plants/

Felhasznált források II.

- Wikipedia

- [http://en.wikipedia.org/wiki/John_Snow_\(physician\)](http://en.wikipedia.org/wiki/John_Snow_(physician))

- http://en.wikipedia.org/wiki/Box_plot

- http://en.wikipedia.org/wiki/Anscombe%27s_quartet

- http://en.wikipedia.org/wiki/Olympic_winners_of_the_Archaic_period

- http://en.wikipedia.org/wiki/Chemical_element

- <http://hu.wikipedia.org/wiki/Triatlon>

- <http://en.wikipedia.org/wiki/Beaver>

- http://en.wikipedia.org/wiki/List_of_countries_by_coffee_consumption_per_capita