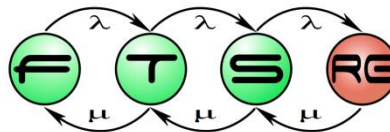


Vizuális adatelemzés

Salánki Ágnes, Guta Gábor, PhD

Dr. Pataricza András

**Budapest University of Technology and Economics
Fault Tolerant Systems Research Group**



Mi is lesz?

Miért vizualizálunk?



Mit vizualizálunk?



Hogyan vizualizálunk?



Mire következtetünk?

Mi is lesz?

Miért vizualizálunk?



Mit vizualizálunk?



Hogyan vizualizálunk?



Mire következtetünk?

A vizualizáció hétköznapijai

Analóg megjelenítés



Digitális megjelenítés



Analóg + koord. rei

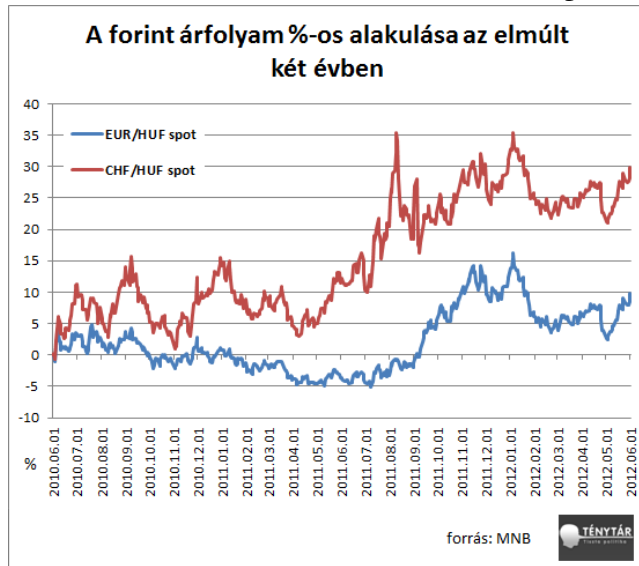


id megjelenítés



A vizualizáció hétköznapijai

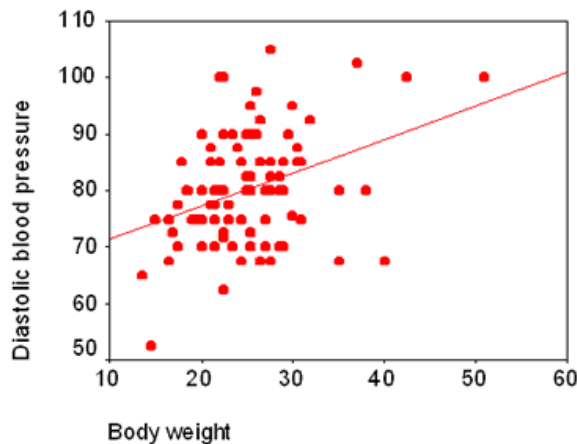
Trend analízis és előrejelzés



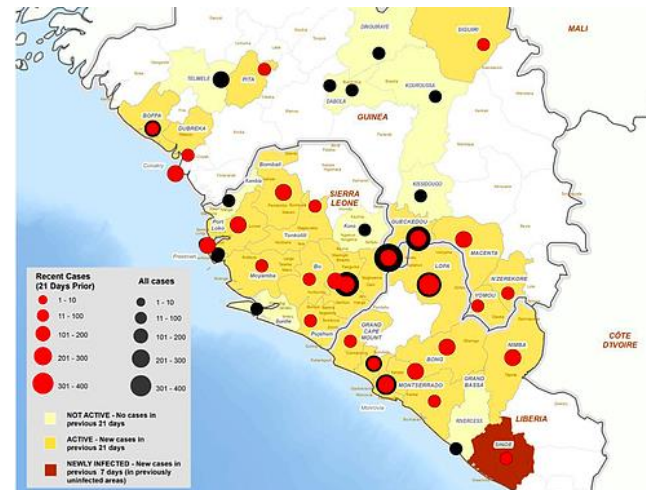
Idősor analízis



Korrelációanalízis

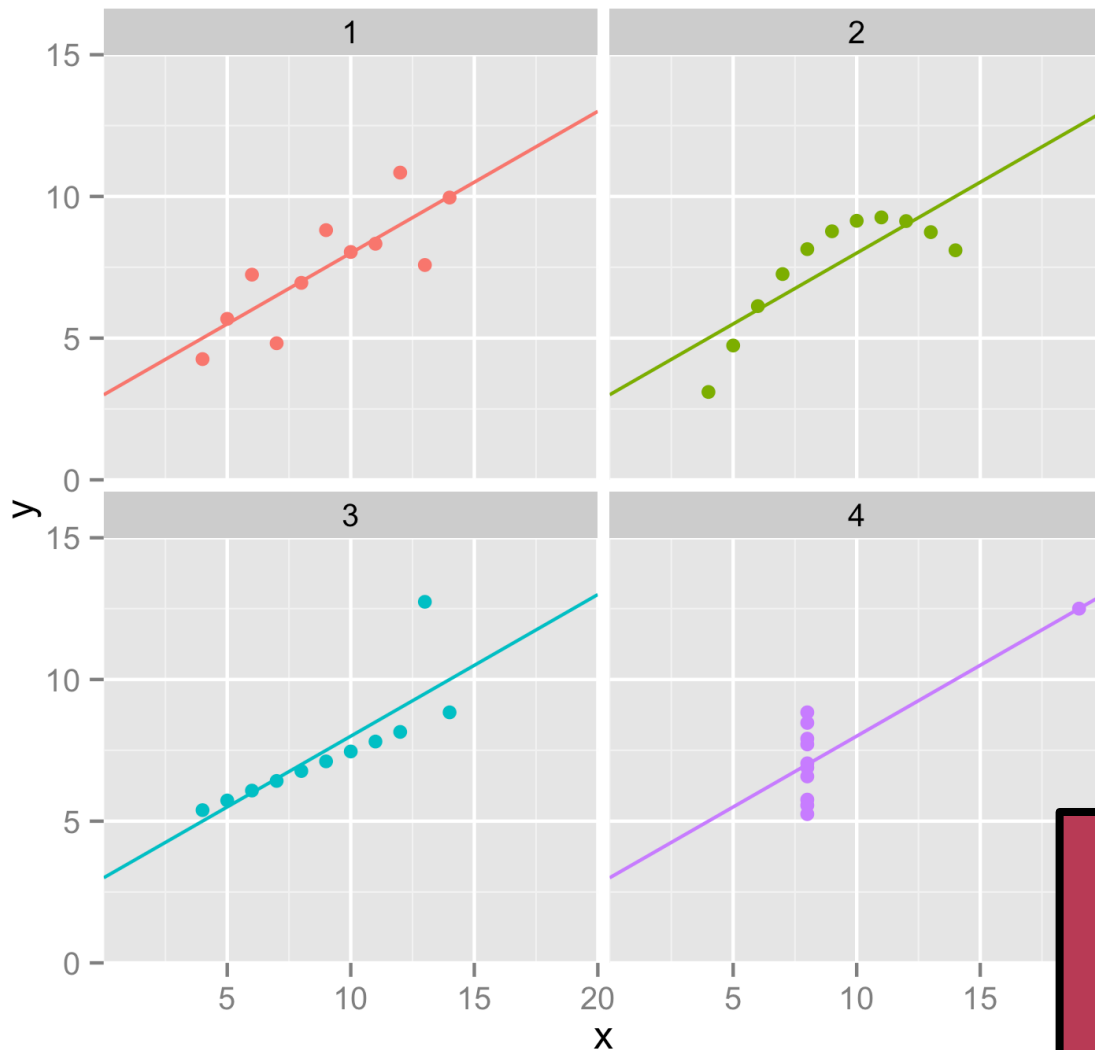


Térbeli analízis



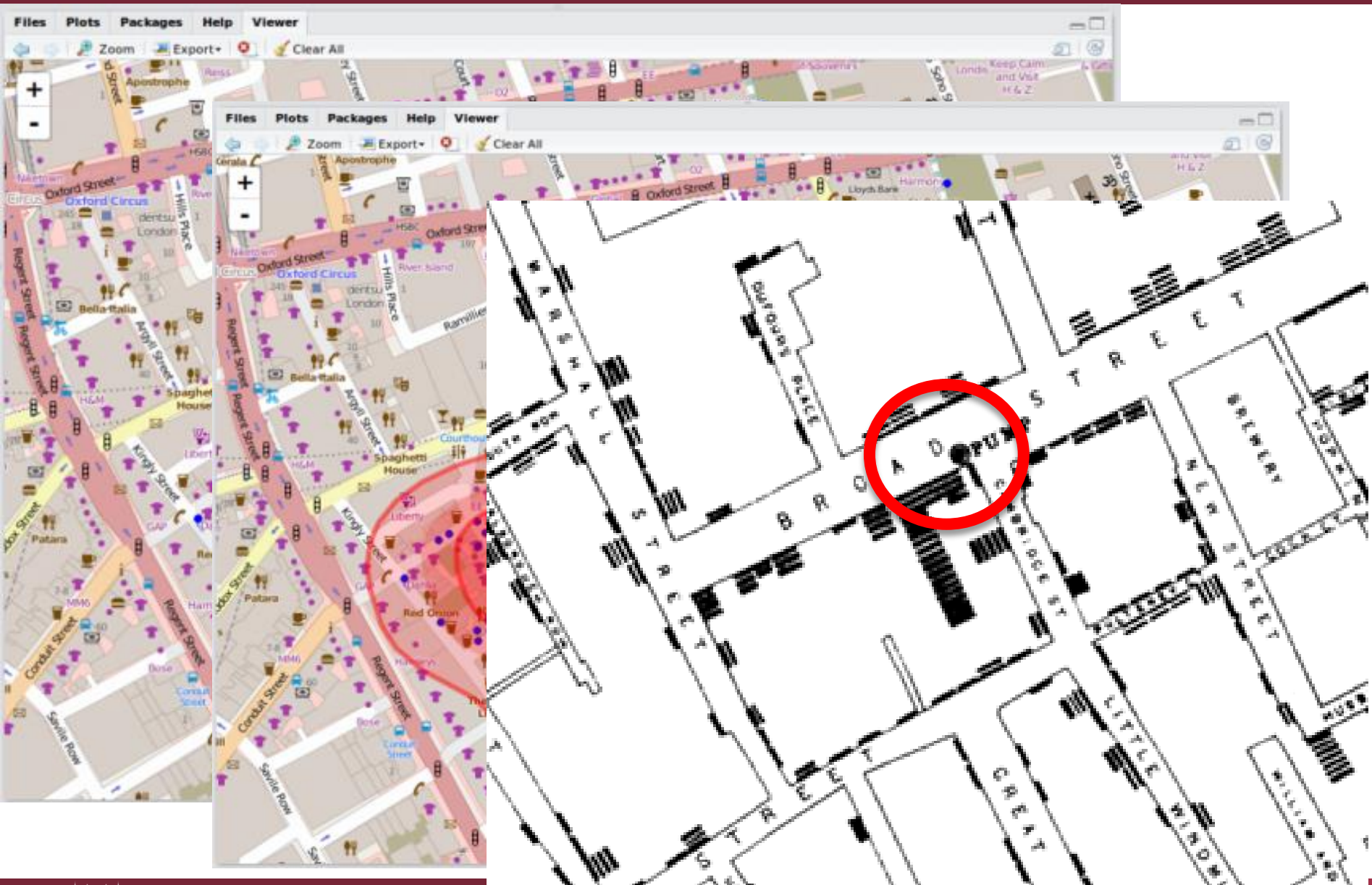
Számítások ellenőrzése

Anscombe's Quartet



Hibás feltételezések
elkerülése... és intuíció

Összefüggések feltárása



Mindent a szemnek!

„Masszív” erőforrások

- 120.000.000 szenzor



3. Vizuális kiválasztás és manipuláció

4. Interpretáció, korreláció más modellekkel, kiértékelés

Mi is lesz?

Miért vizualizálunk?



Mit vizualizálunk?



Hogyan vizualizálunk?



Mire következtetünk?

Emlékeztető: táblázatok

- **Sor** = elem, megfigyelés
- **Oszlop** = jellemző

Típus

Numerikus
érték

név	fénykard színe	nem	holdak száma	keringési idő
Alderaan			1	364
Coruscant			4	368
Darth Vader	piros	férfi		
Jabba		híműs		
Leia Organa		nő		
Luke Skywalker	zöld	férfi		
Mace Windu	lila	férfi		
Tatooine			3	304
Yoda	zöld	férfi		

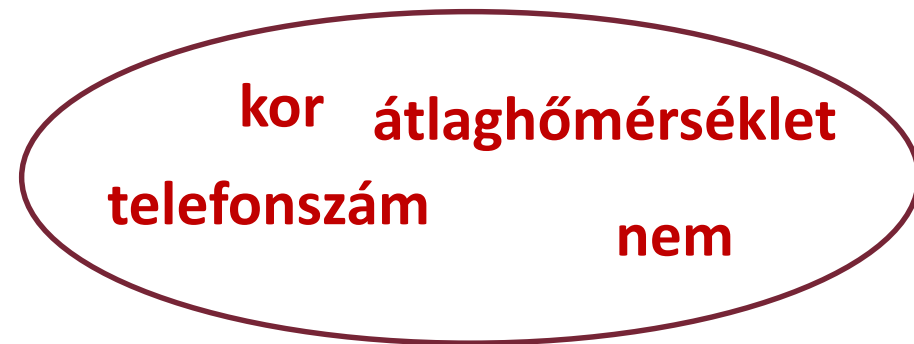
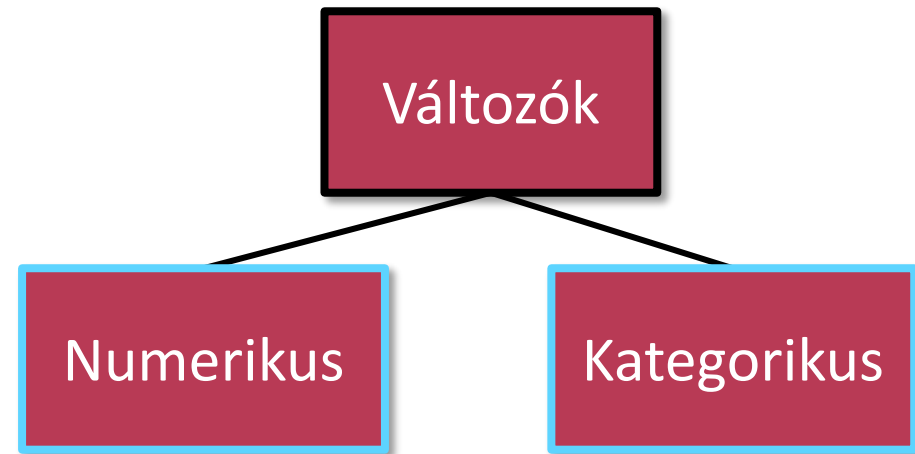
Numerikus és kategorikus változók

- Numerikus (numerical)

- az alapvető aritmetikai műveletek értelmesek

- Kategorikus (categorical)

- Csak megkülönböztetés miatt, már reláció sincs



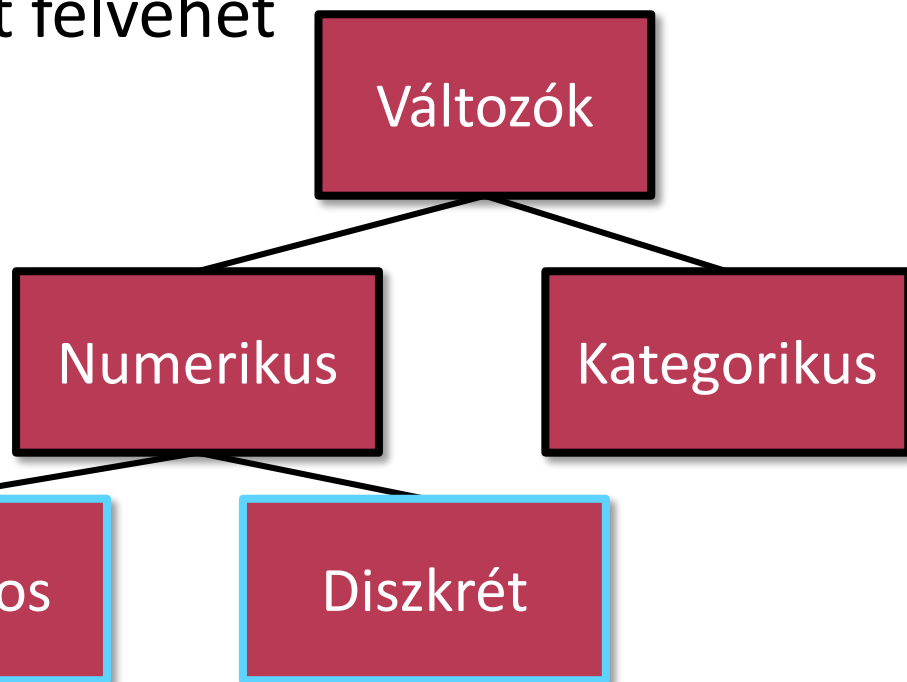
Numerikus változók

■ Folytonos

- Mért – tetszőleges értéket felvehet

- adott tartományon belül
- adott pontosság mellett

- Pl. a teremben ülők
ZH pontszámának átlaga

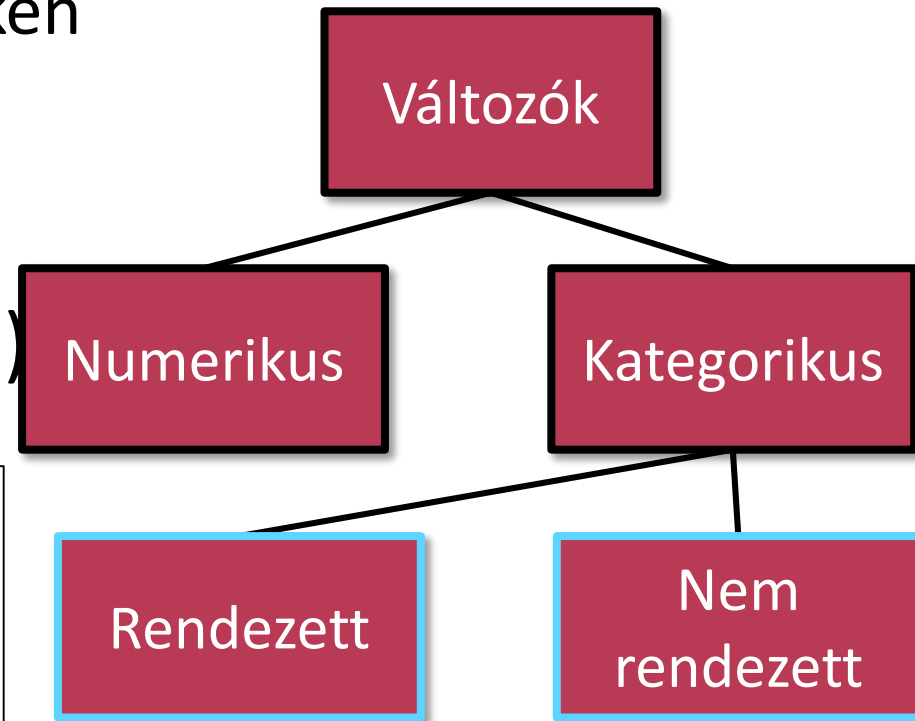


■ Diszkrét

- Számolt – véges sok értéket vehet fel adott tartományban
- Pl. az előadáson ülők száma

Kategorikus változók

- Rendezett (ordinális)
 - Teljes rendezés az értékeken
 - Pl. szállodai csillagok
- Nem rendezett (reguláris)



9. Ajánlanád-e a tárgyat másoknak?

- ☐ Mindenkit rábeszelnék
- ☐ Nyugodtan ajánlanám
- ☐ Esetleg ajánlanám
- ☐ Inkább lebeszelném róla
- ☐ Feltétlenül lebeszelném
- ☐ Nem kívánok válaszolni

Mi is lesz?

Miért vizualizálunk?



Mit vizualizálunk?



Hogyan vizualizálunk?



Mire következtetünk?

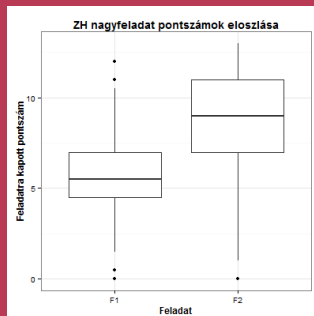
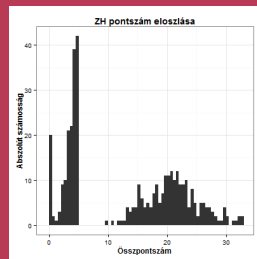
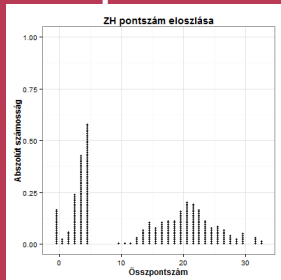
1 változó – eloszlásokra

Változók

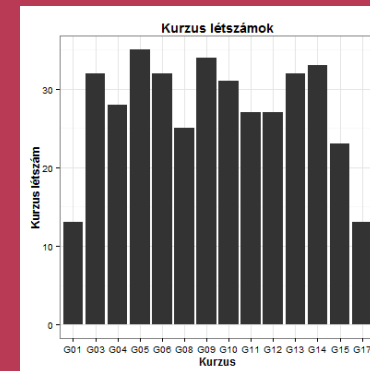
Numerikus

Kategorikus

ZH pontszám: [13, 15, 2, ...]



Kurzus: [G01, G03, G15, G17, ...]

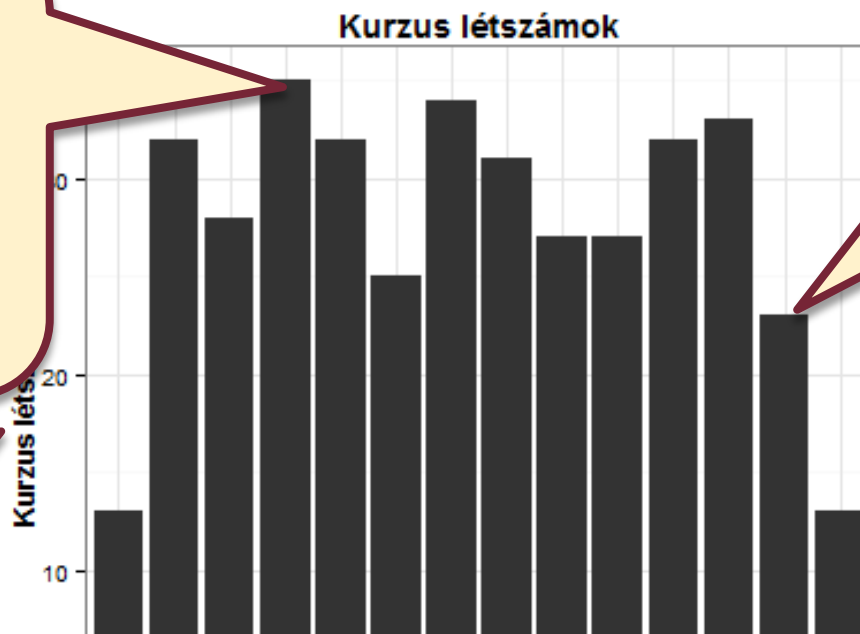


Oszlopdiagram

- Bemenő változó: kurzus kód
- Kérdés: az egyes kurzusokra hányan járnak?

Vannak nagyon népszerű időpontok/gyakvezek?

abszolút gyakoriság!



Oszlop-magasság: adott érték gyakorisága

Tervezői döntés: értékkészlet darabolása
Pl.: kedd-csütörtök-péntek

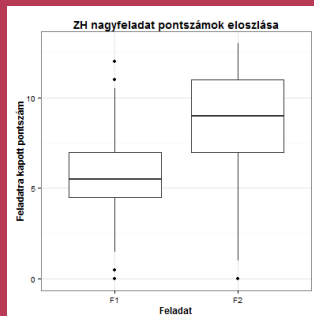
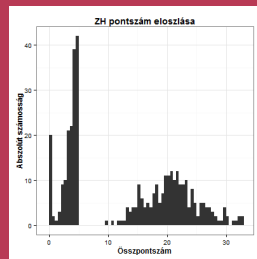
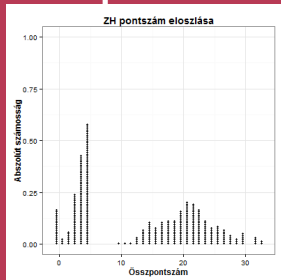
1 változó – eloszlásokra

Változók

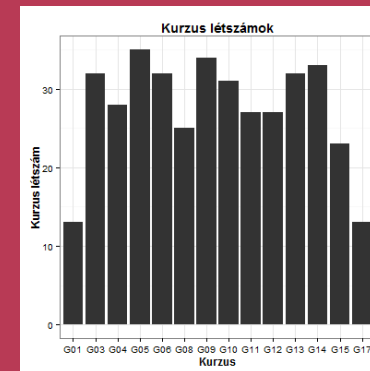
Numerikus

Kategorikus

ZH pontszám: [13, 15, 2, ...]

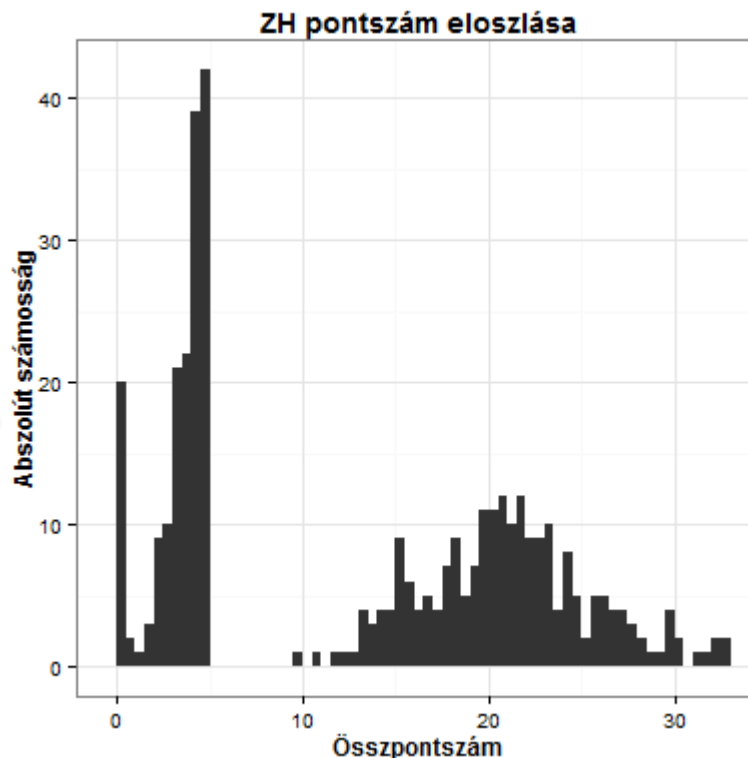


Kurzus: [G01, G03, G15, G17, ...]



Hisztogram

- Bemenő változó: ZH összpontszám
- Kérdés: hogyan alakultak a ZH pontszámok?



abszolút
gyakoriság!

Oszlop-
magasság:
adott
intervallum
számossága

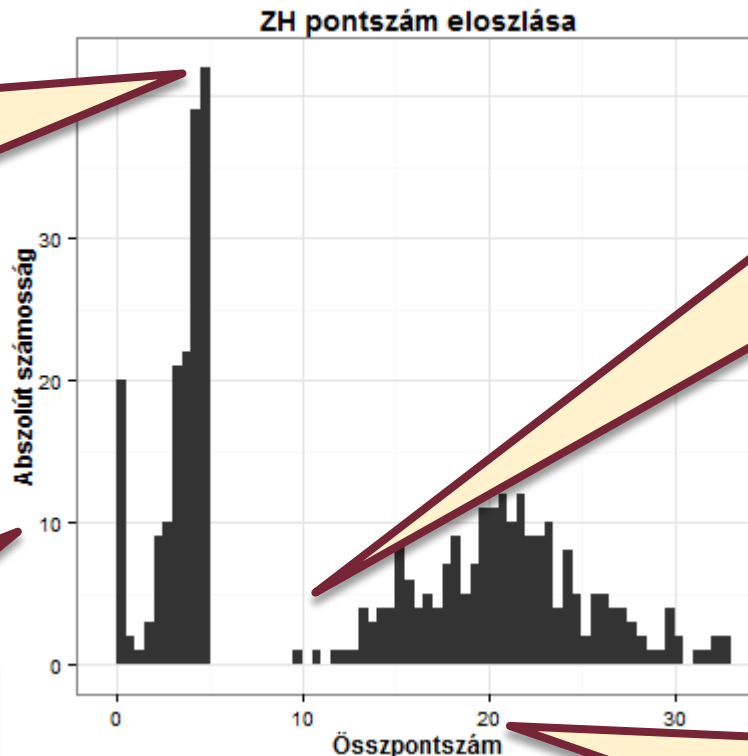
Tervezői döntés: mekkora legyen az intervallum hossza?
Pl.: elég 1 pontos felbontással, vagy menjünk fél pontokig?

Hisztogram

- Bemenő változó: ZH összpontszám
- Kérdés: hogyan alakultak a ZH pontszámok?

Sokan voltak a határon

Akik el sem jöttek

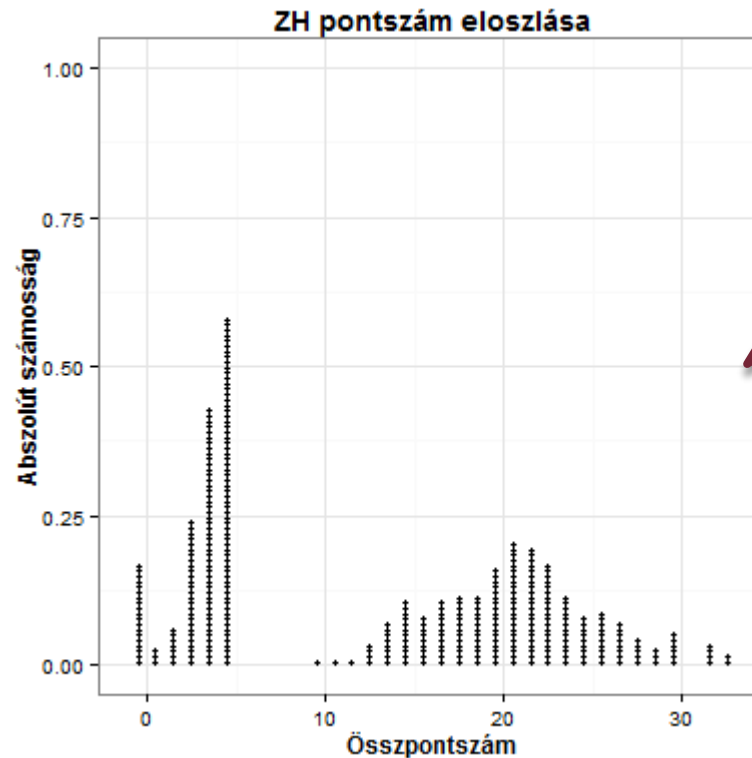


Akik átmentek a beugrón, valószínűleg át is mentek

20 pont körül volt az átlag/medián

Relatív gyakoriságok

- Bemenő változó: ZH összpontszám
- Kérdés: hogyan alakultak a ZH pontszámok?

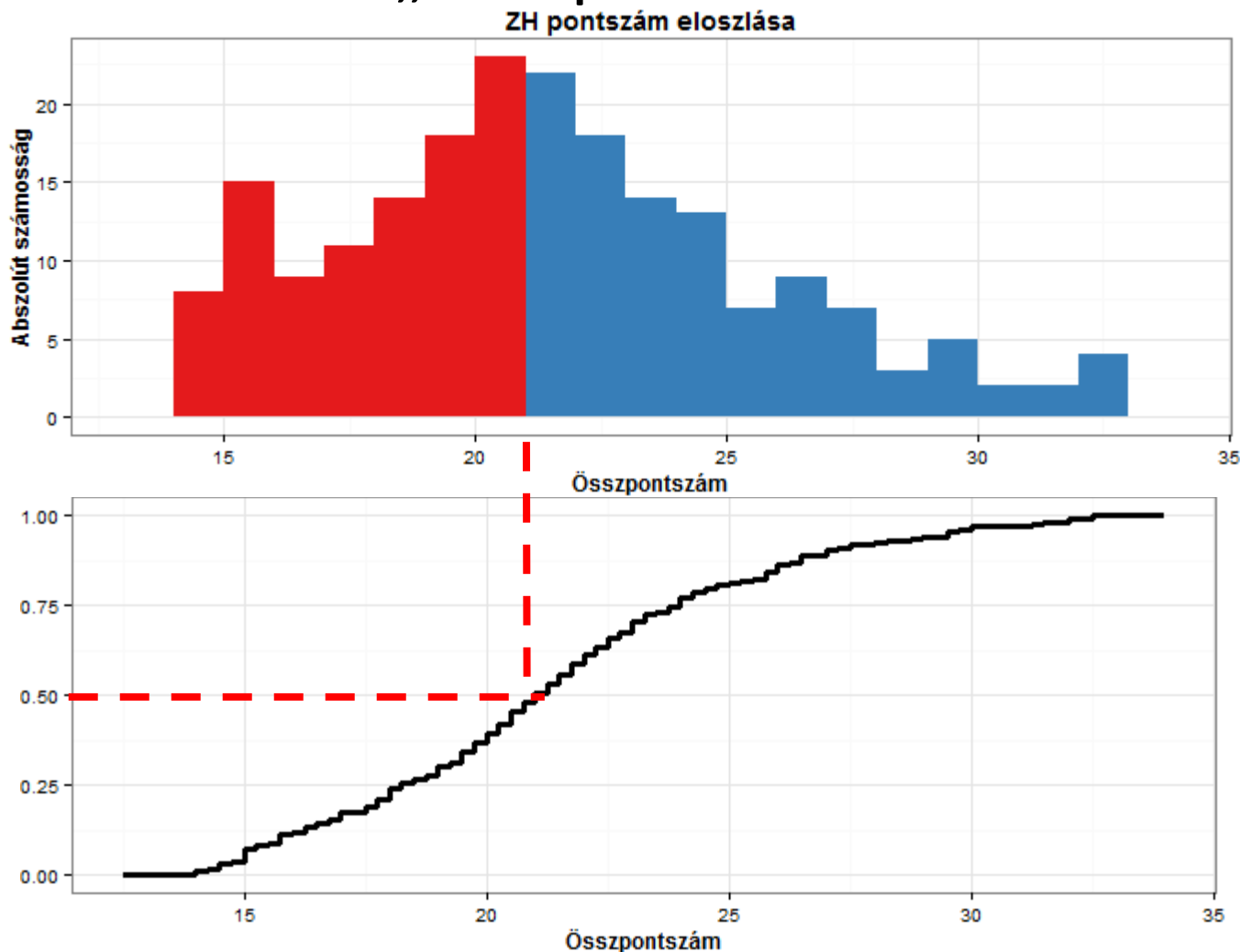


De relatív
gyakoriság!

Ugyanaz mint
hisztogramnál

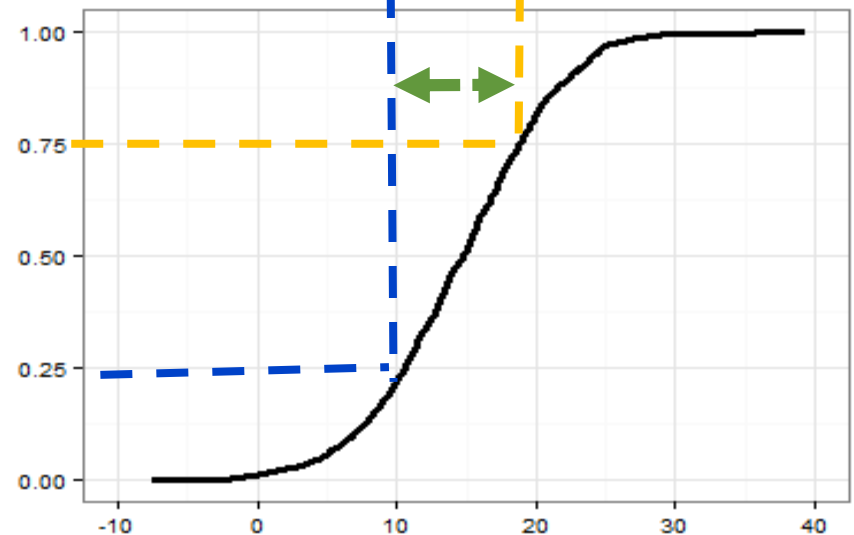
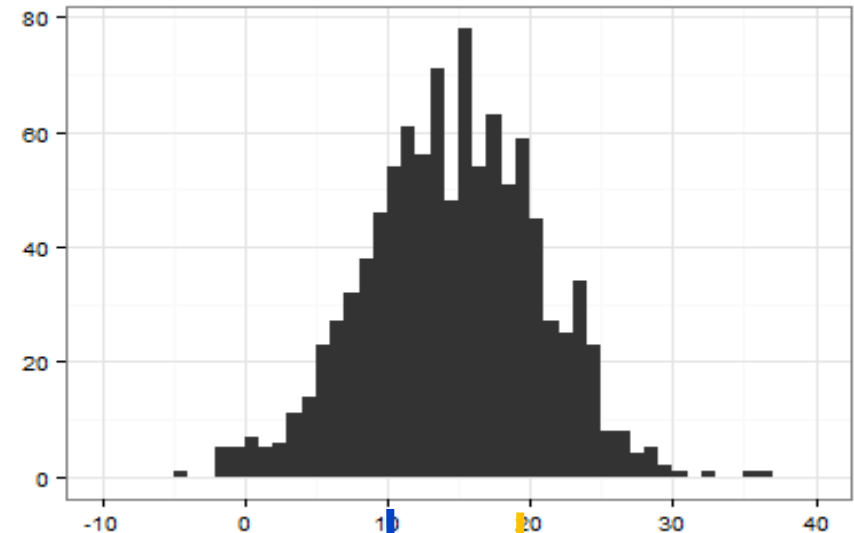
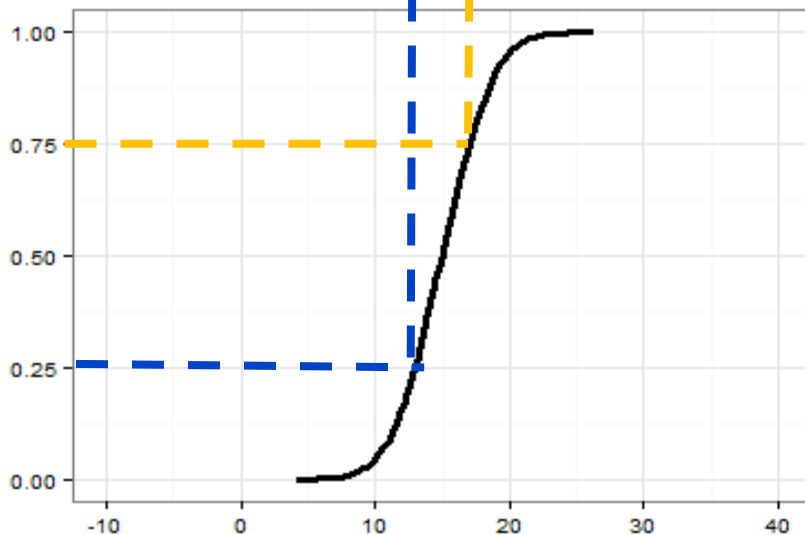
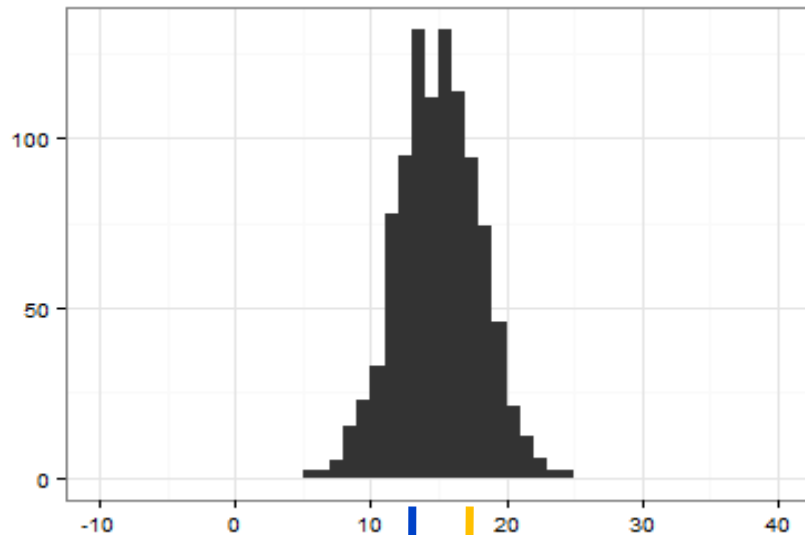
Egyszerű statisztikai jellemzés

- Hol van az adatok „közepe”?



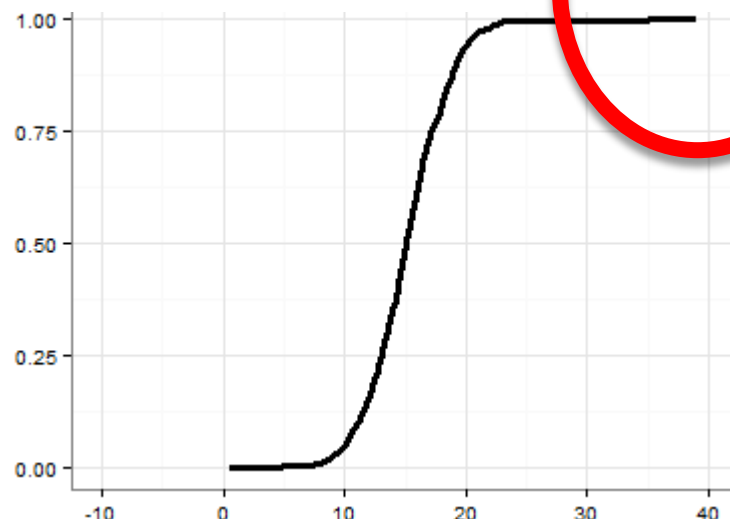
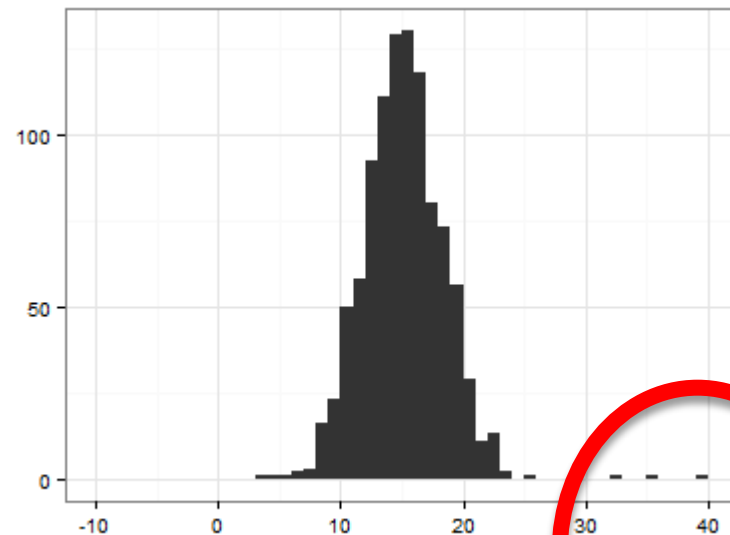
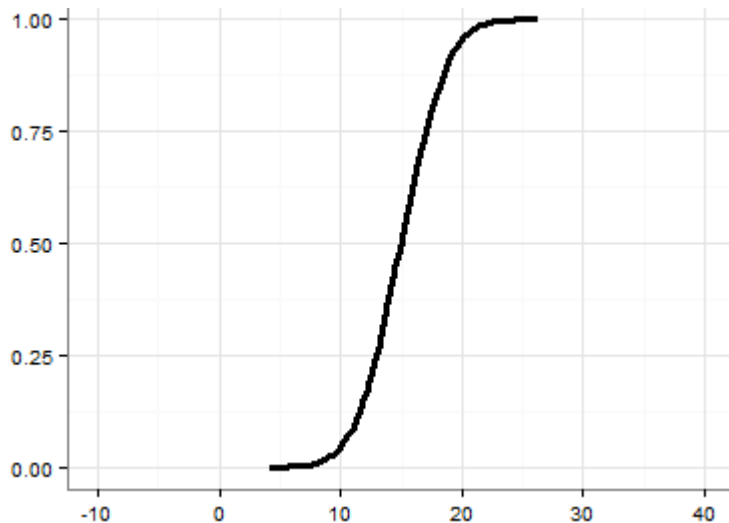
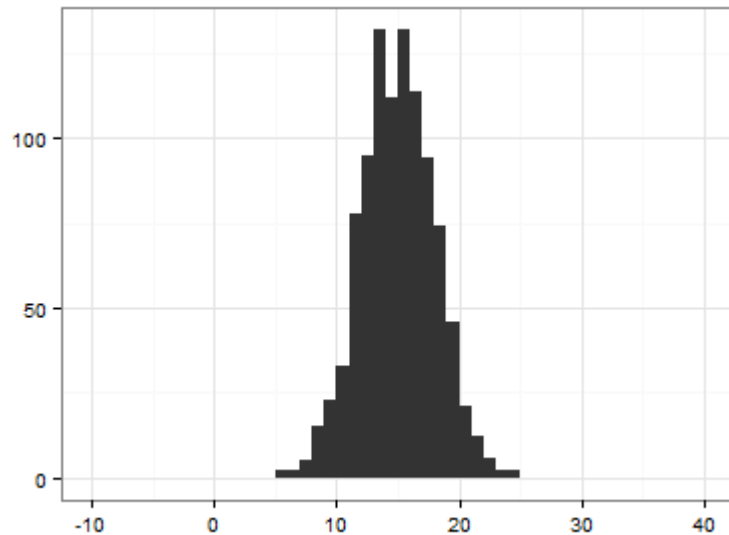
Egyszerű statisztikai jellemzés

- Mennyire „szórtak” az adatok?



Egyszerű statisztikai jellemzés

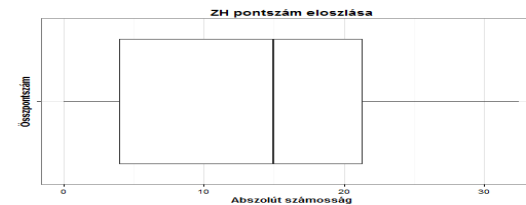
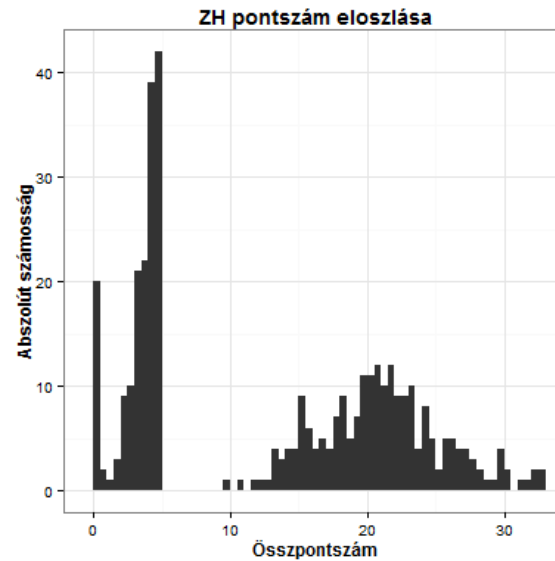
■ Vannak-e kilógók?



Boxplot

- Bemenő változó: ZH összpontszám
- Kérdés: hogyan alakultak a ZH pontszámok úgy nagyjából?

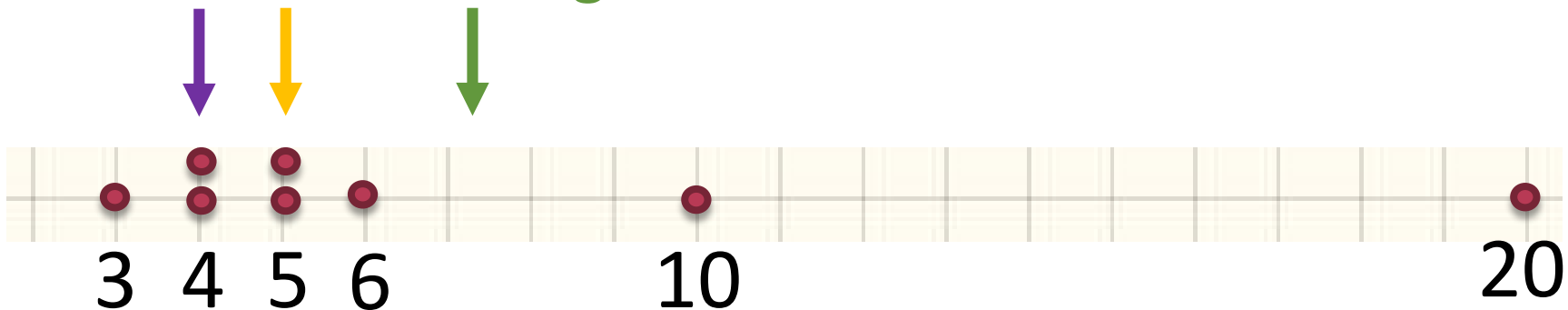
Egyfajta absztrakció itt is:
legyenek intervallumok,
felesleges minden pontot
kirajzolni



(Folytonos) megfigyelések jellemzése

- A „központ” jellemzése
 - Átlag, **medián**, módusz
 - {3, 4, 4, 5, 5, 6, 10, 20}
 - Átlag: ~ 7.125
 - Medián: 5
 - Módusz: 4 és 5
- A „terjedelem” jellemzése?

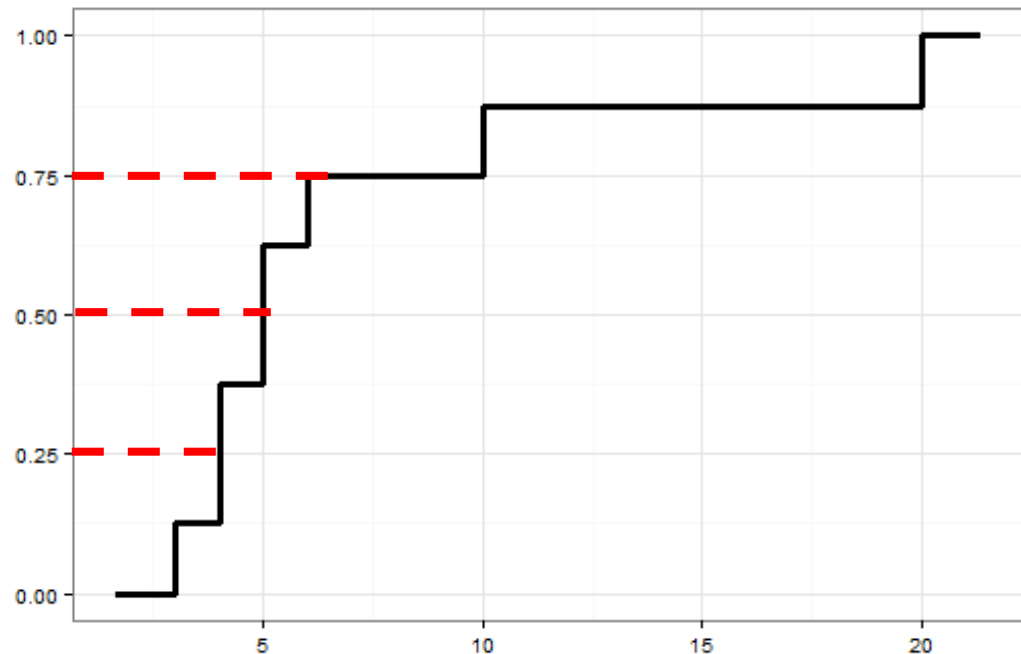
módusz medián átlag



Percentilisek

■ Percentilis

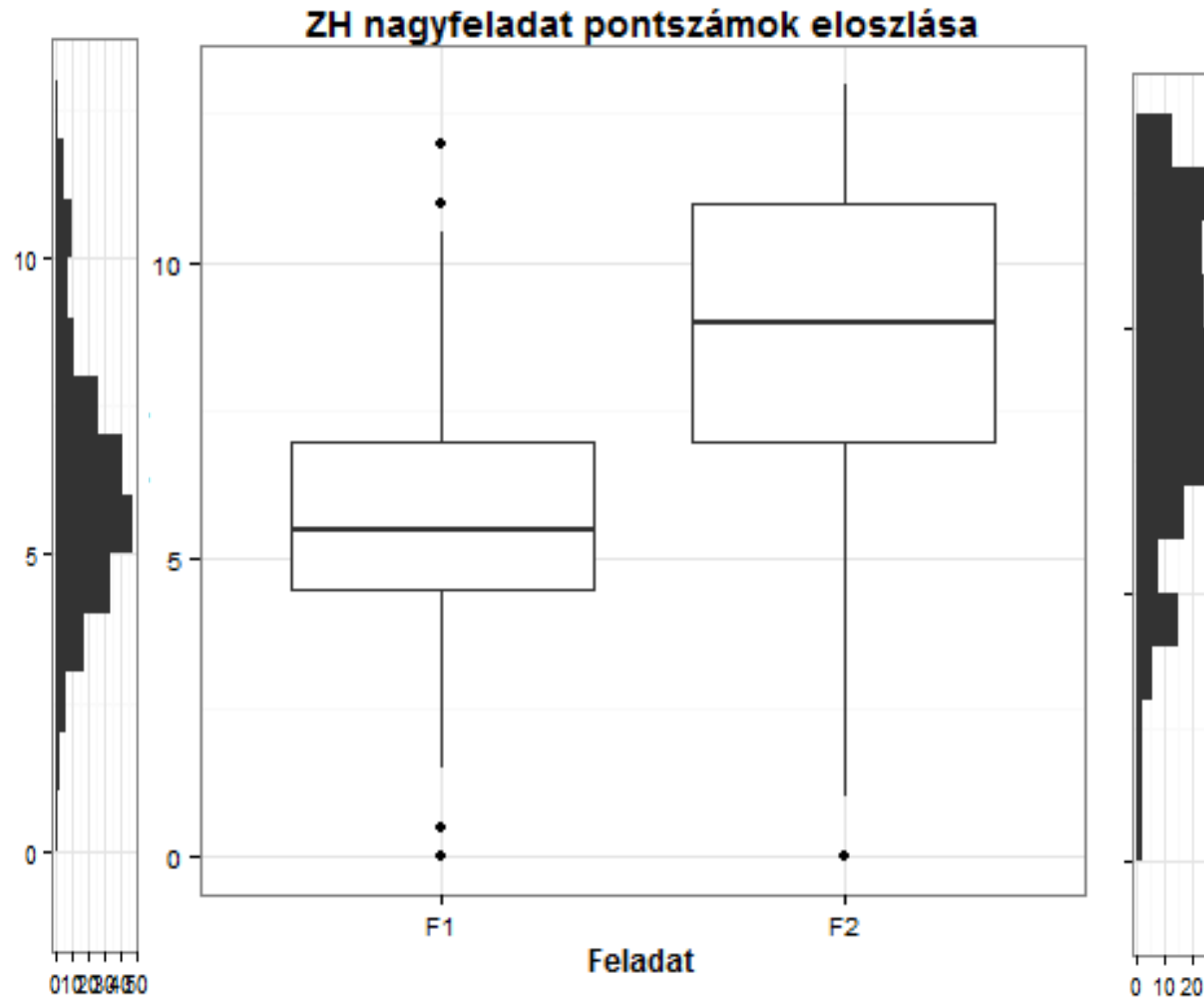
- Az n -edik percentilnél az adatok $n\%$ -a kisebb
- $\{3, 4, 4, 5, 5, 6, 10, 20\}$
 - 50. percentilis: 5
 - 25. percentilis: 4
 - 75. percentilis: 6



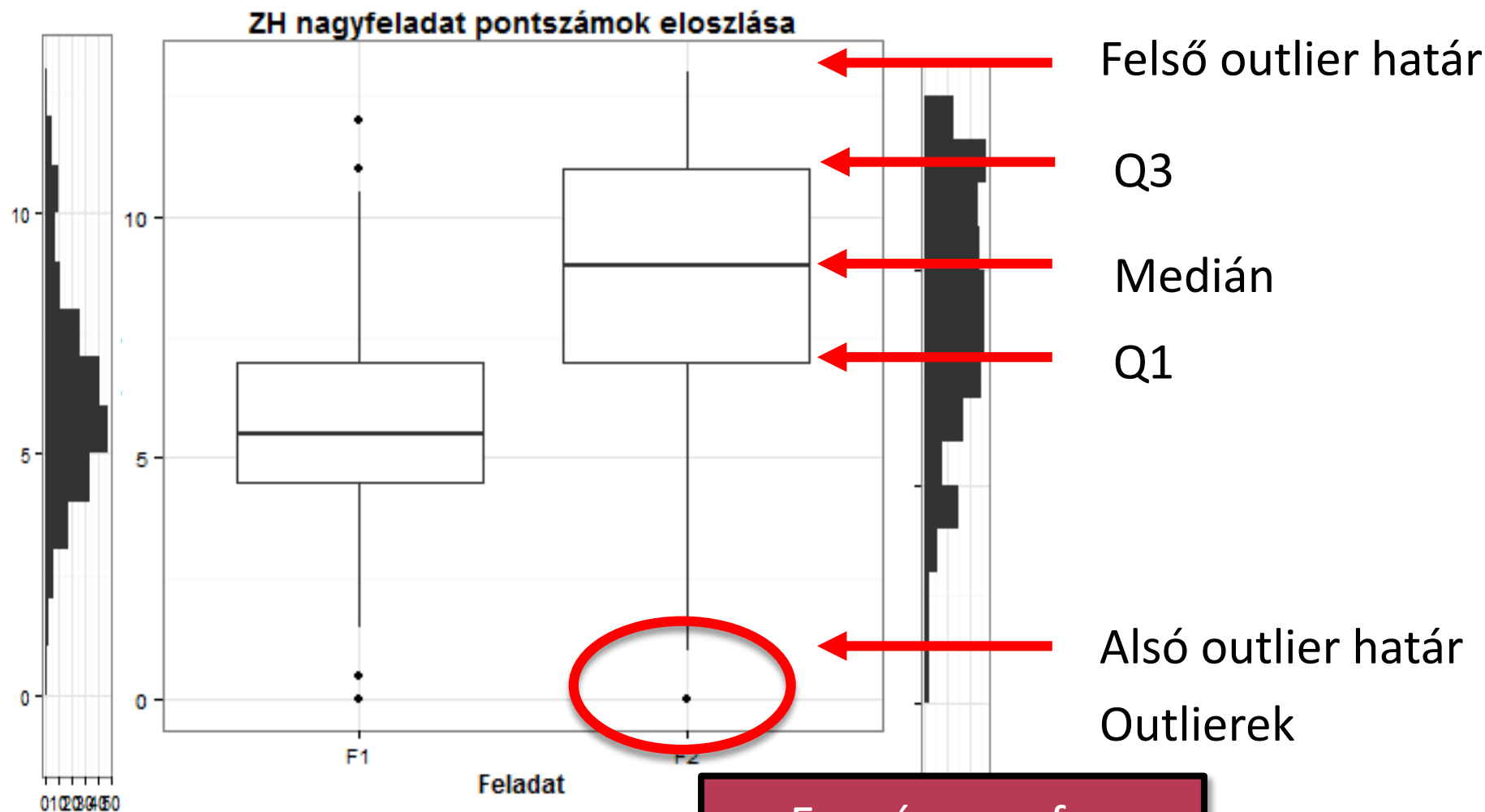
■ Kvartilis

- Q1: 25. percentilis
- Q3: 75. percentilis
- **Q2: medián**

Boxplot (Box and whisker plot)

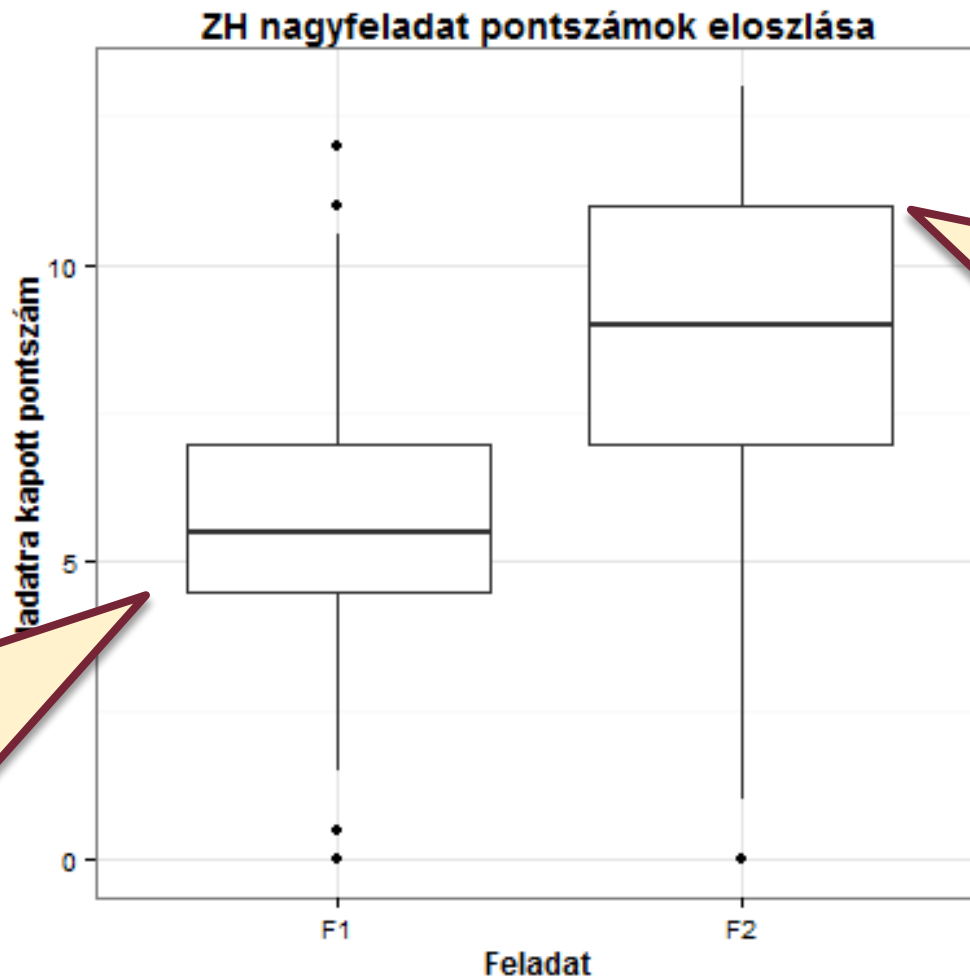


Boxplot (Box and whisker plot)



Ez már nem fog
menni Excelben. (?)

Boxplot (Box and whisker plot)

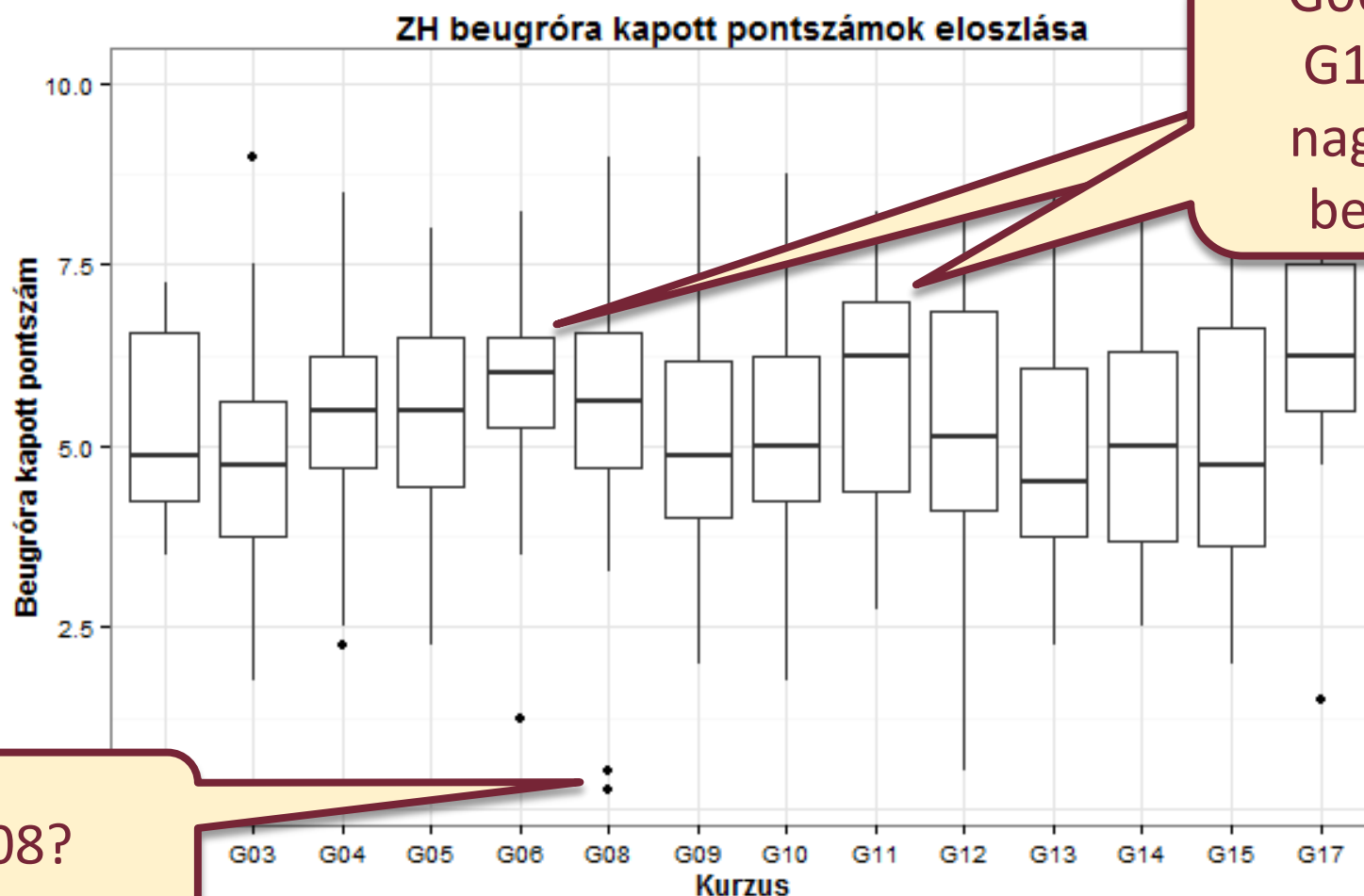


Az F1
pontszámok
50%-a
4.5 és 7.5
között volt

F2-re
általában több
pontot kaptak,
mint F1-re

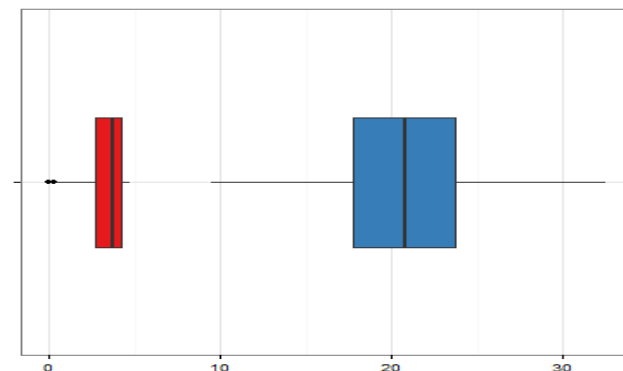
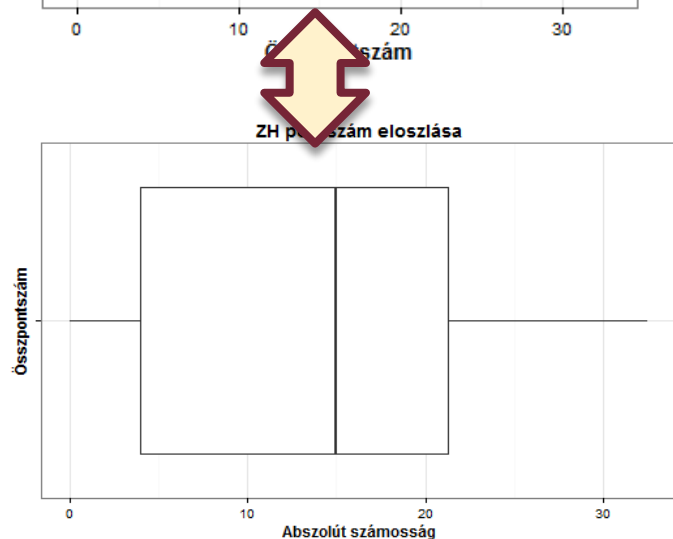
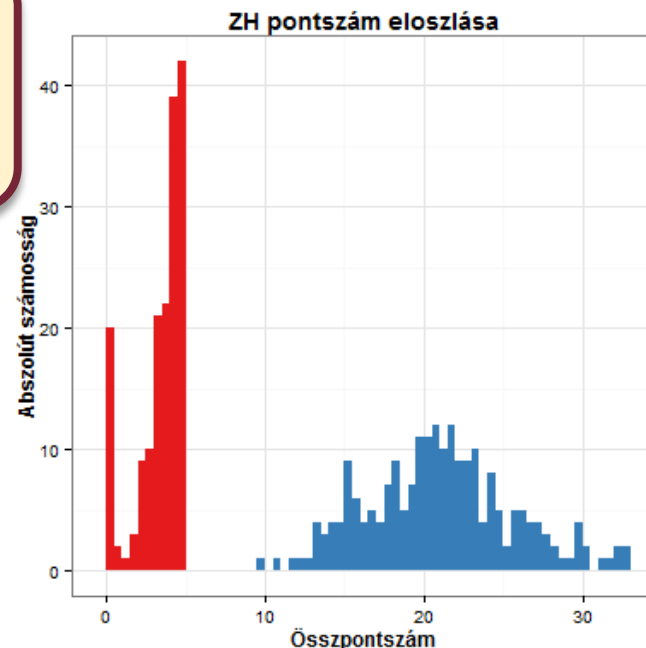
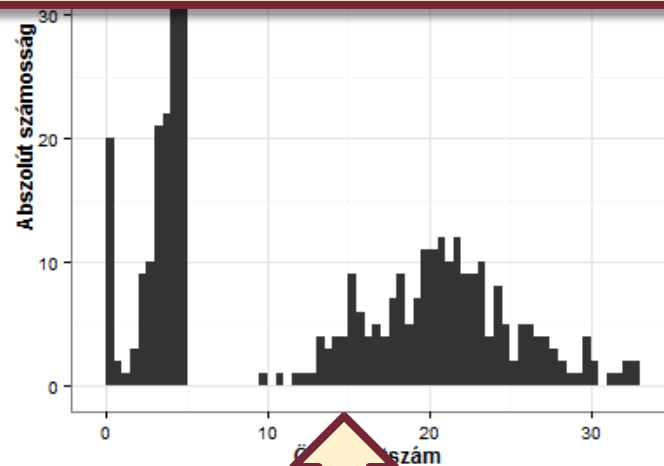
Boxplot (Box and whisker plot)

- Melyik csoportban hogyan sikerültek a beugrók?



Boxplot (Box and whisker plot)

Absztrakció: a boxplottal fontos információt is veszíthetünk!



Miért medián, miért nem átlag?

■ Alaphalmaz

○ 1000 pont $\sim U(1, 5)$ egyenletes eloszlás

- *átlag = medián = 3 ms*



3ms $\pm$ 2 ms



Válaszidő

Új medián: `sort(resp. times)[501] = 3.02 ms`

Vál. medián

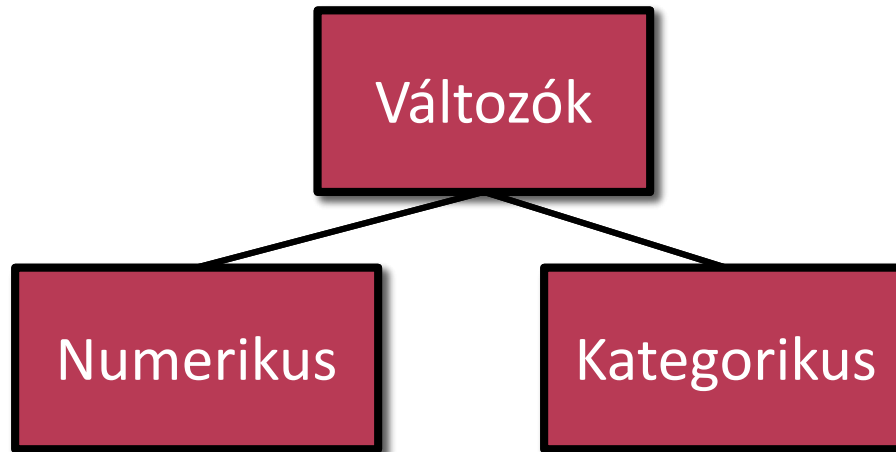
→ ✓ Robusztus

Vál. átlag

→ ✗ Nem rob.

Új átlag: $(2 * 10^4 + 3 * 10^3) / 1001 = 25 \text{ ms!}$

2 változó kapcsolata

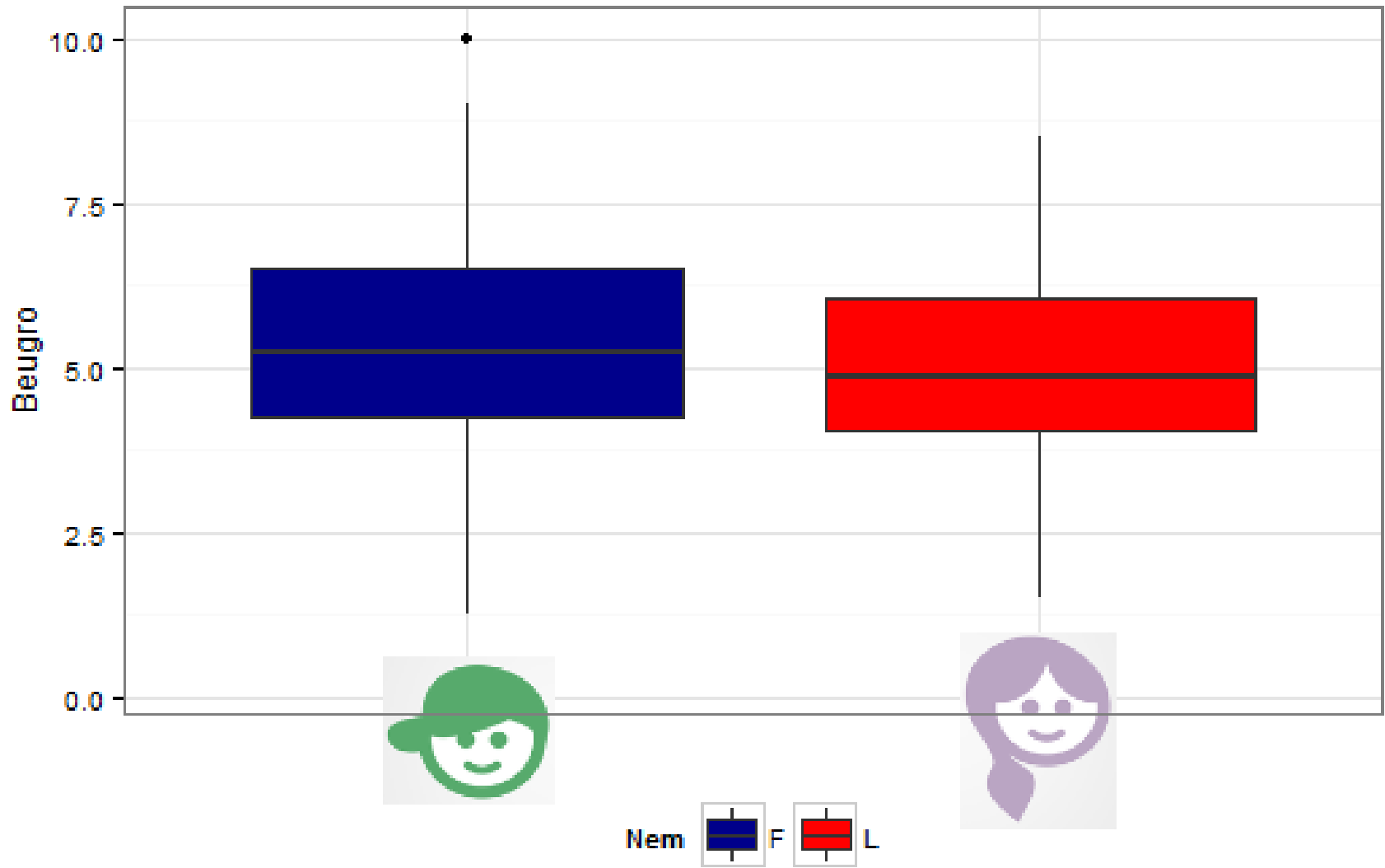


2 numerikus

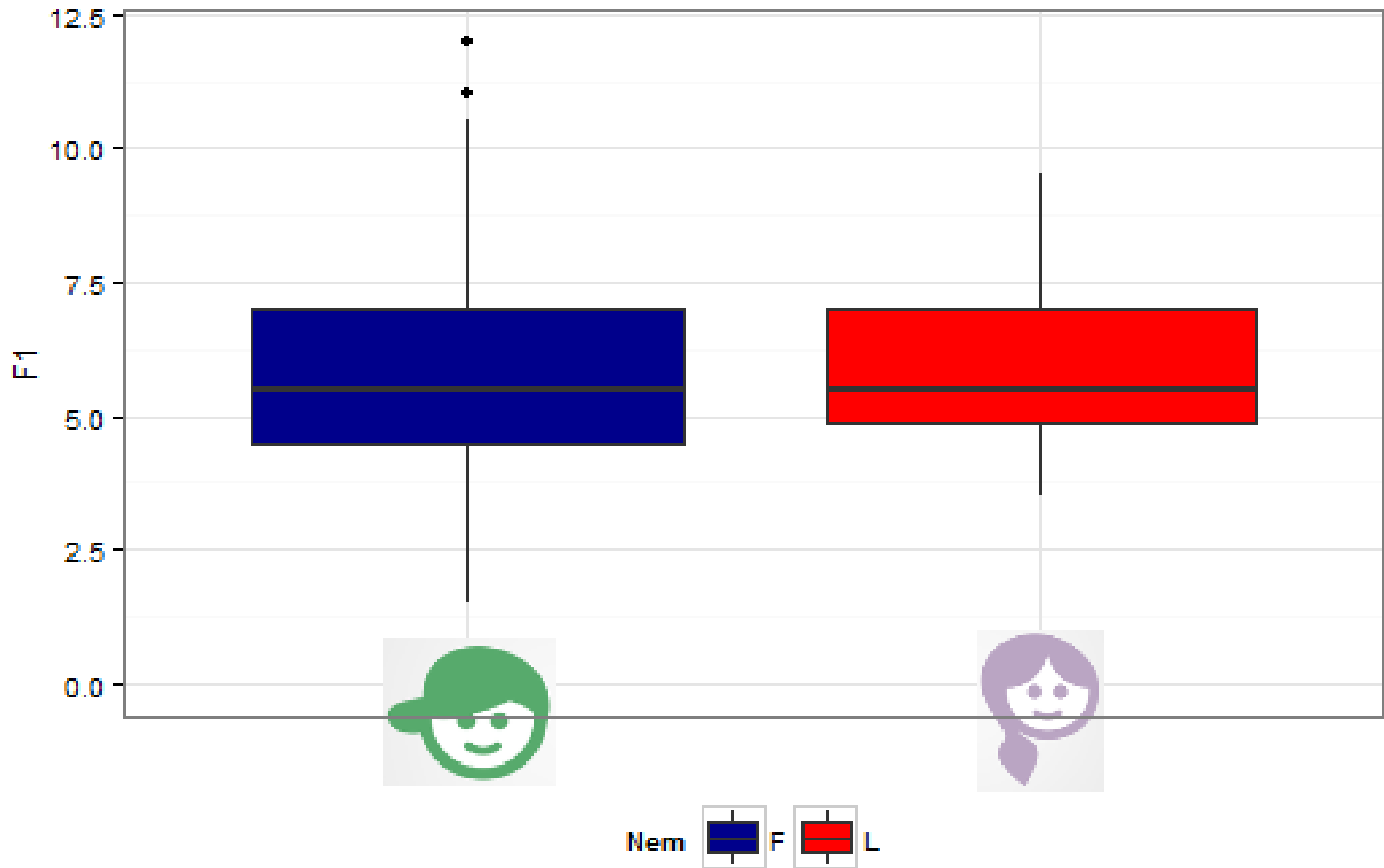
1 numerikus,
1 kategorikus

2 kategorikus

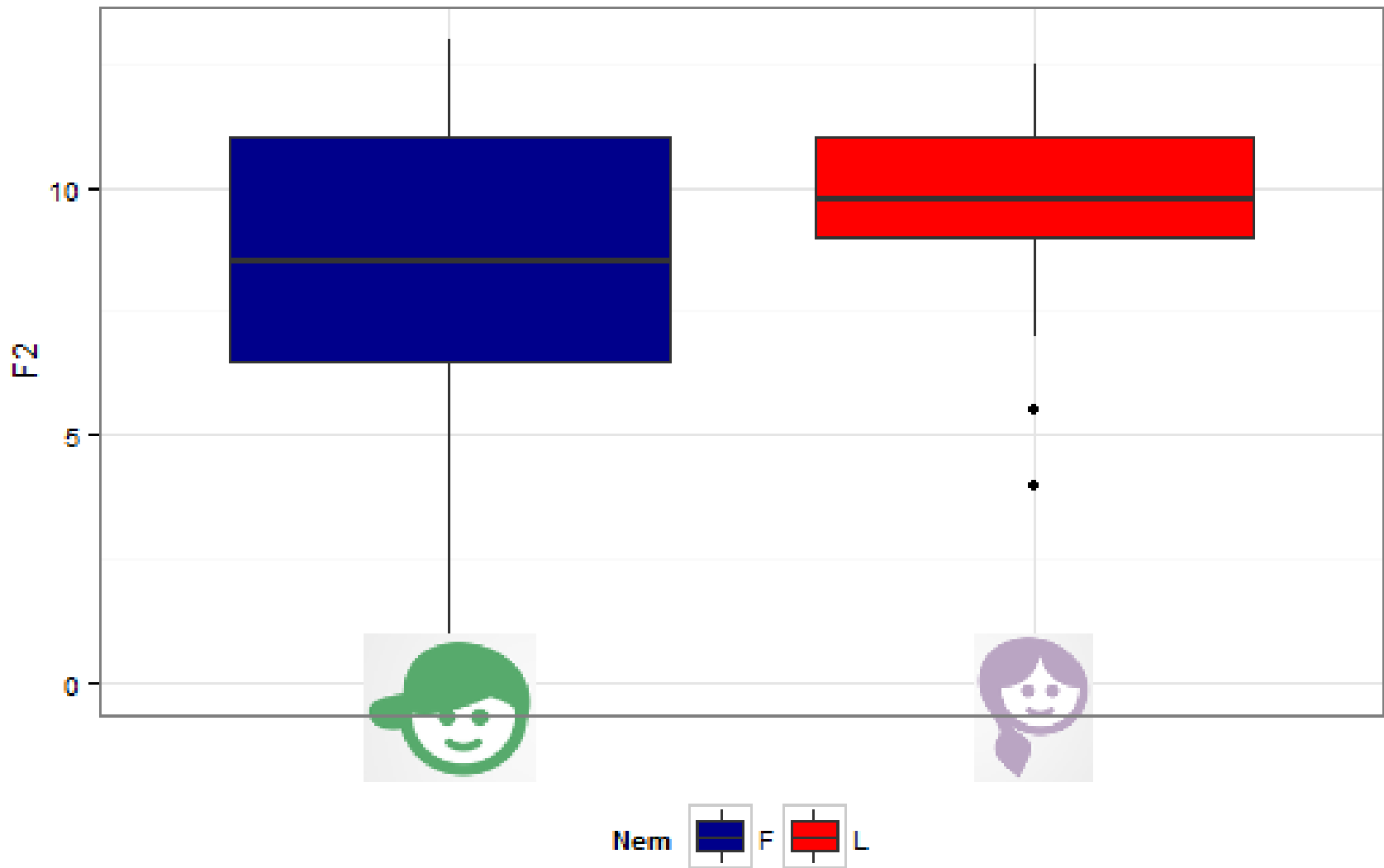
Numerikus kategóriánként



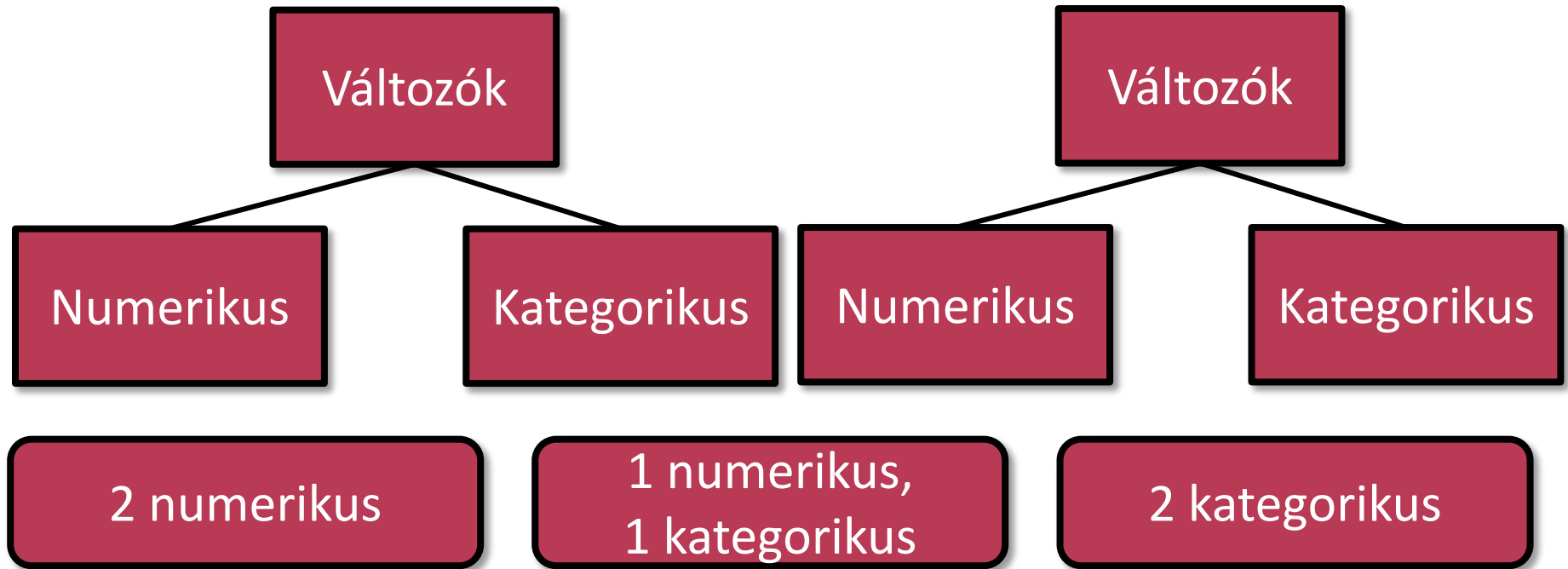
Numerikus kategóriánként



Numerikus kategóriánként



2 változó kapcsolata

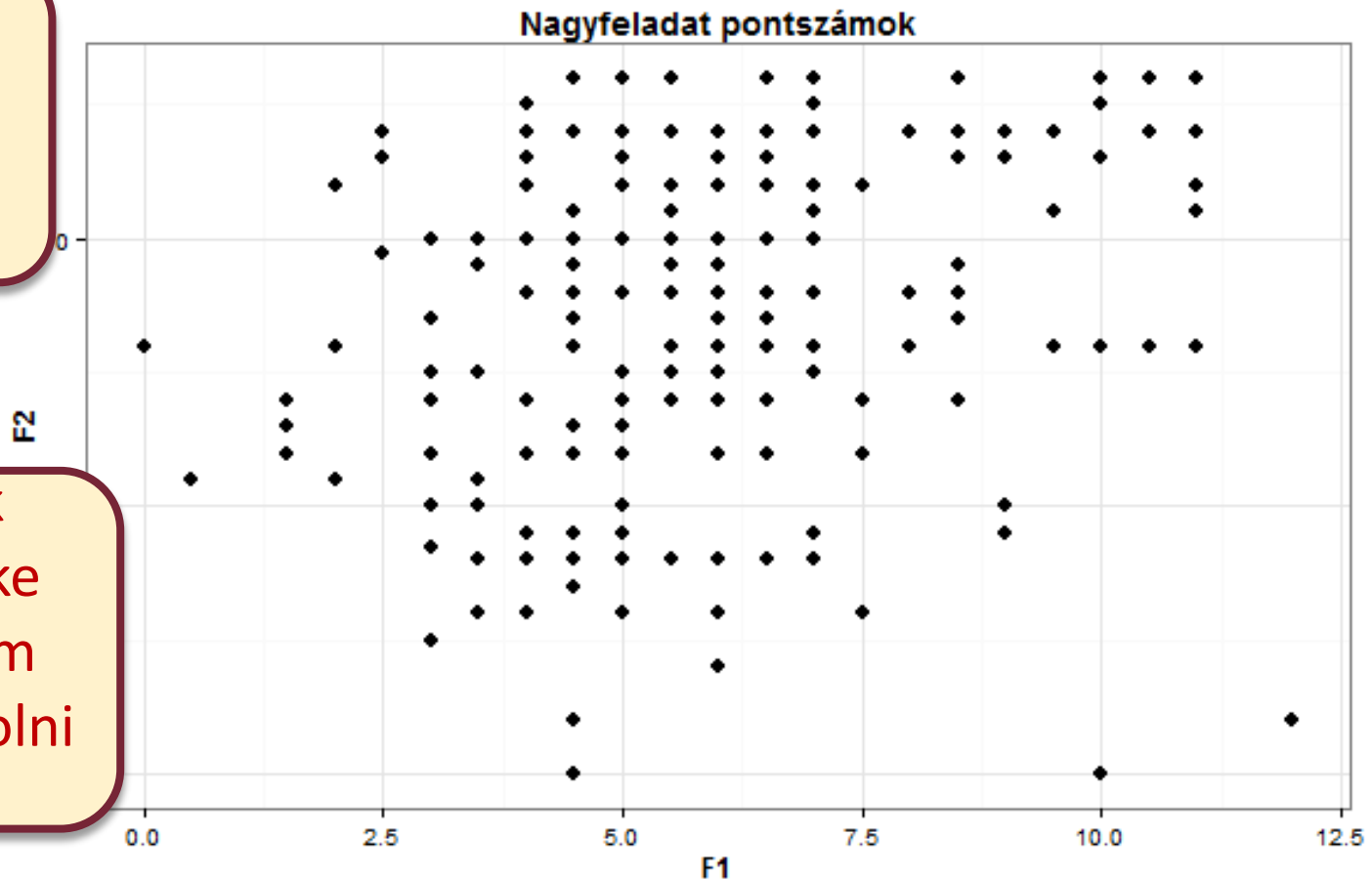


Pont – pont diagram (scatterplot)

- Bemenő változó: nagyfeladatokra kapott pontok
- Kérdés: hogyan viszonyulnak egymáshoz?

Együttesen
előforduló
pontpárokat
vizualizálunk

Ha az egyik
változó értéke
hiányzik, nem
tudjuk felrajzolni



Pont – pont diagram (scatterplot)

- Bemenő változó: nagyfeladatokra kapott pontok
- Kérdés: hogyan viszonyulnak egymáshoz?

Nem biztos,
hogy akinek
megy az F1,
megy az F2 is

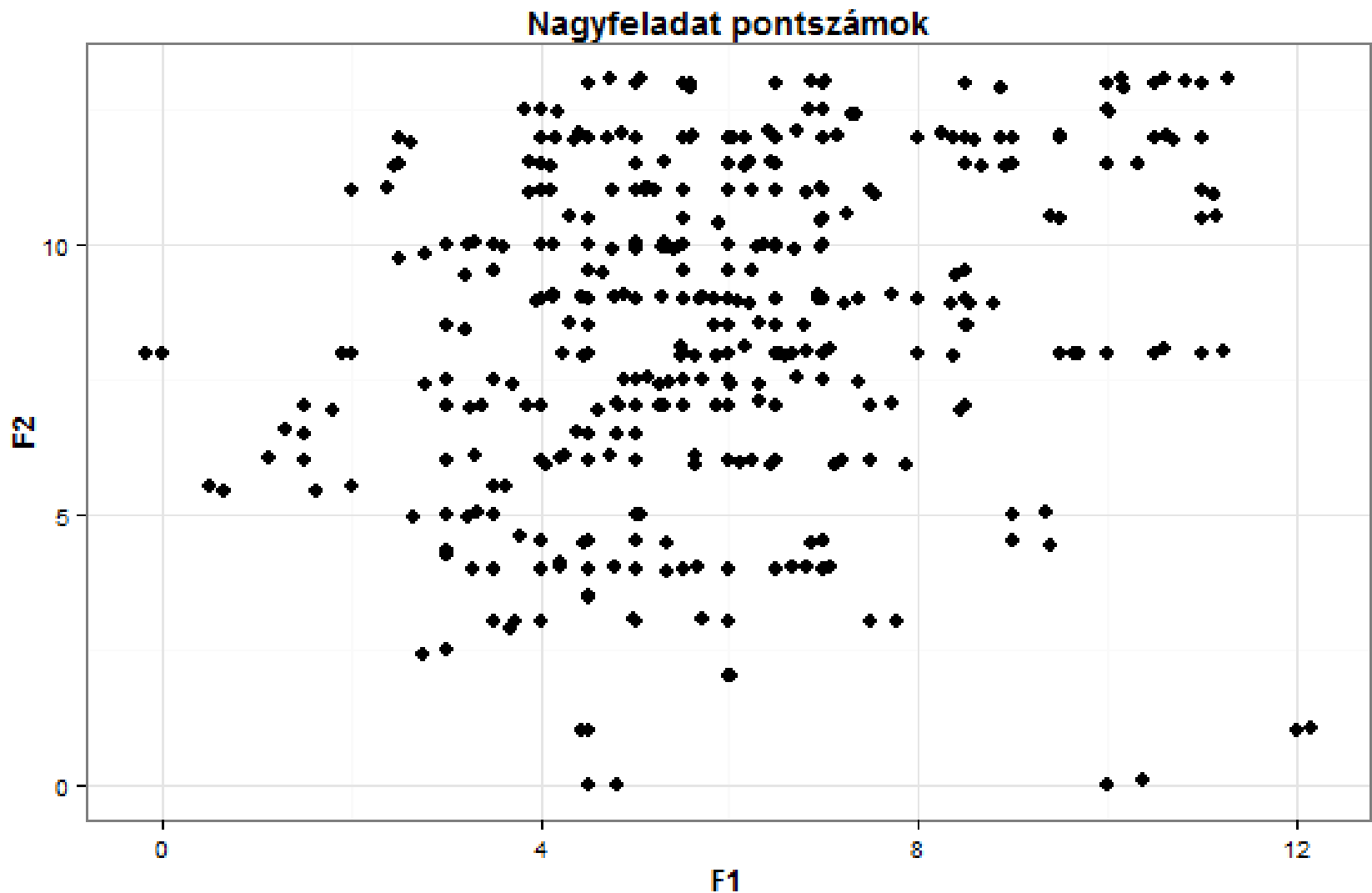


Hogyan kezeljük a takarásokat?

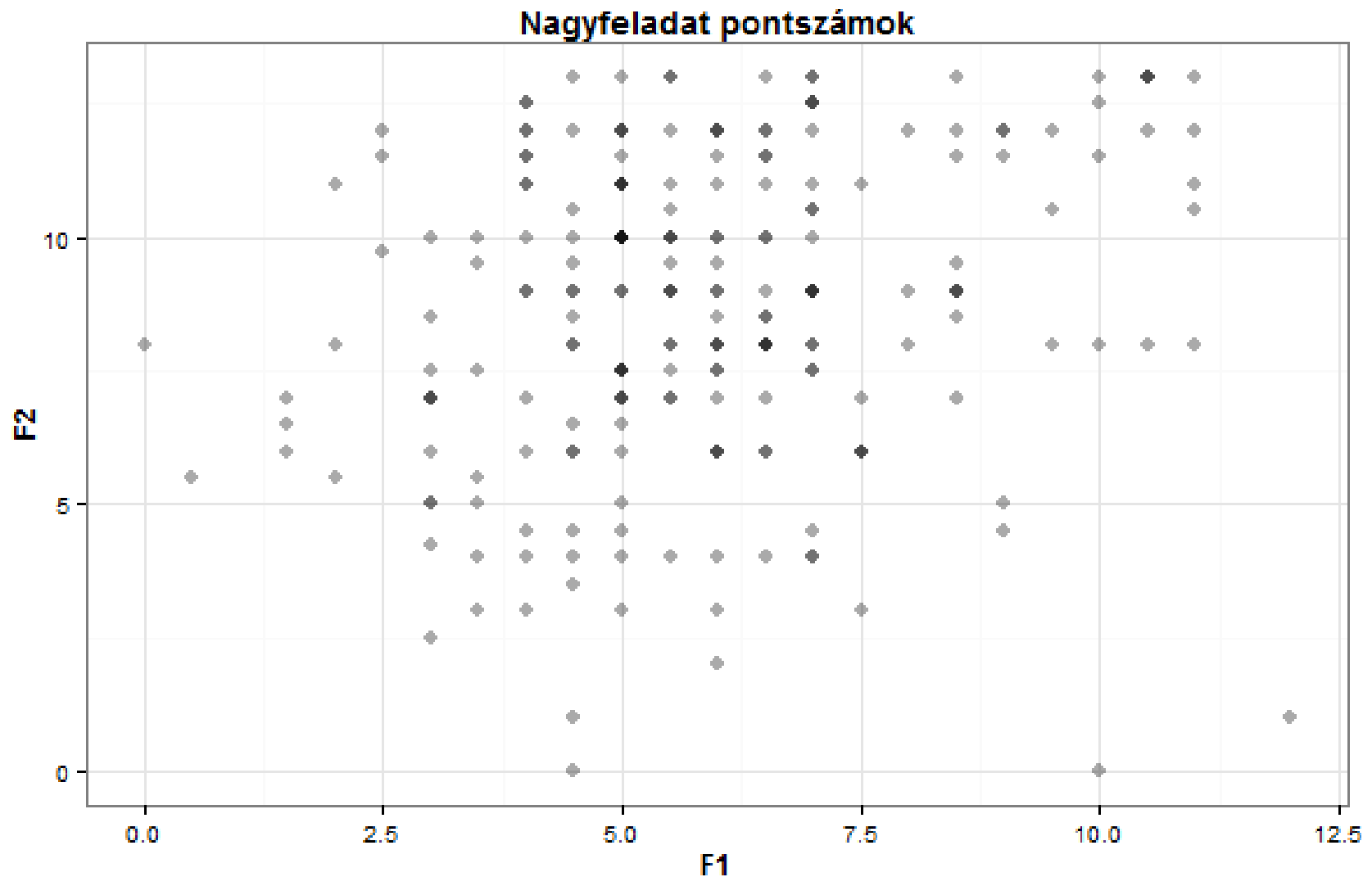
Overplotting



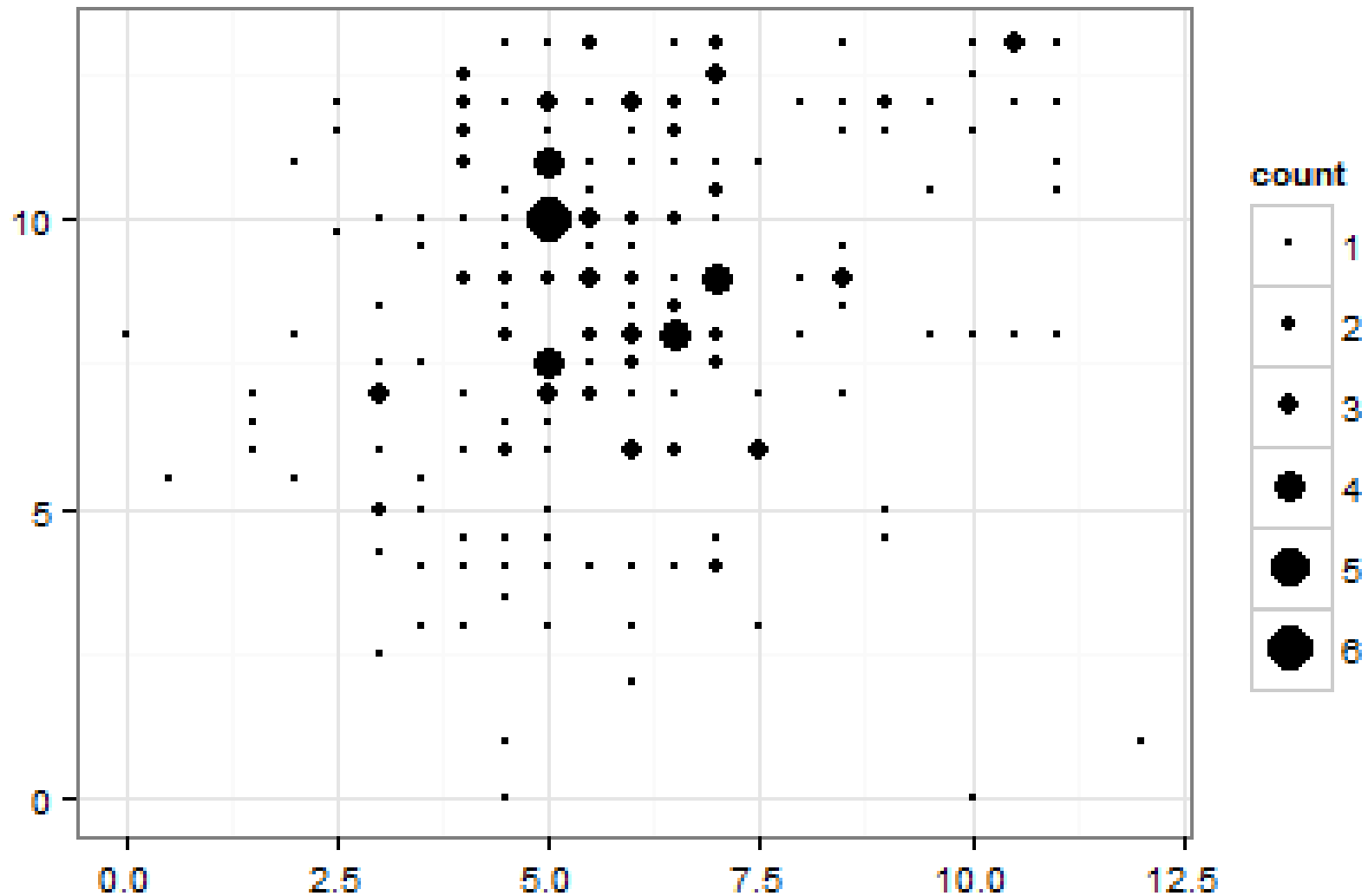
Overplotting megoldások 1: jitter



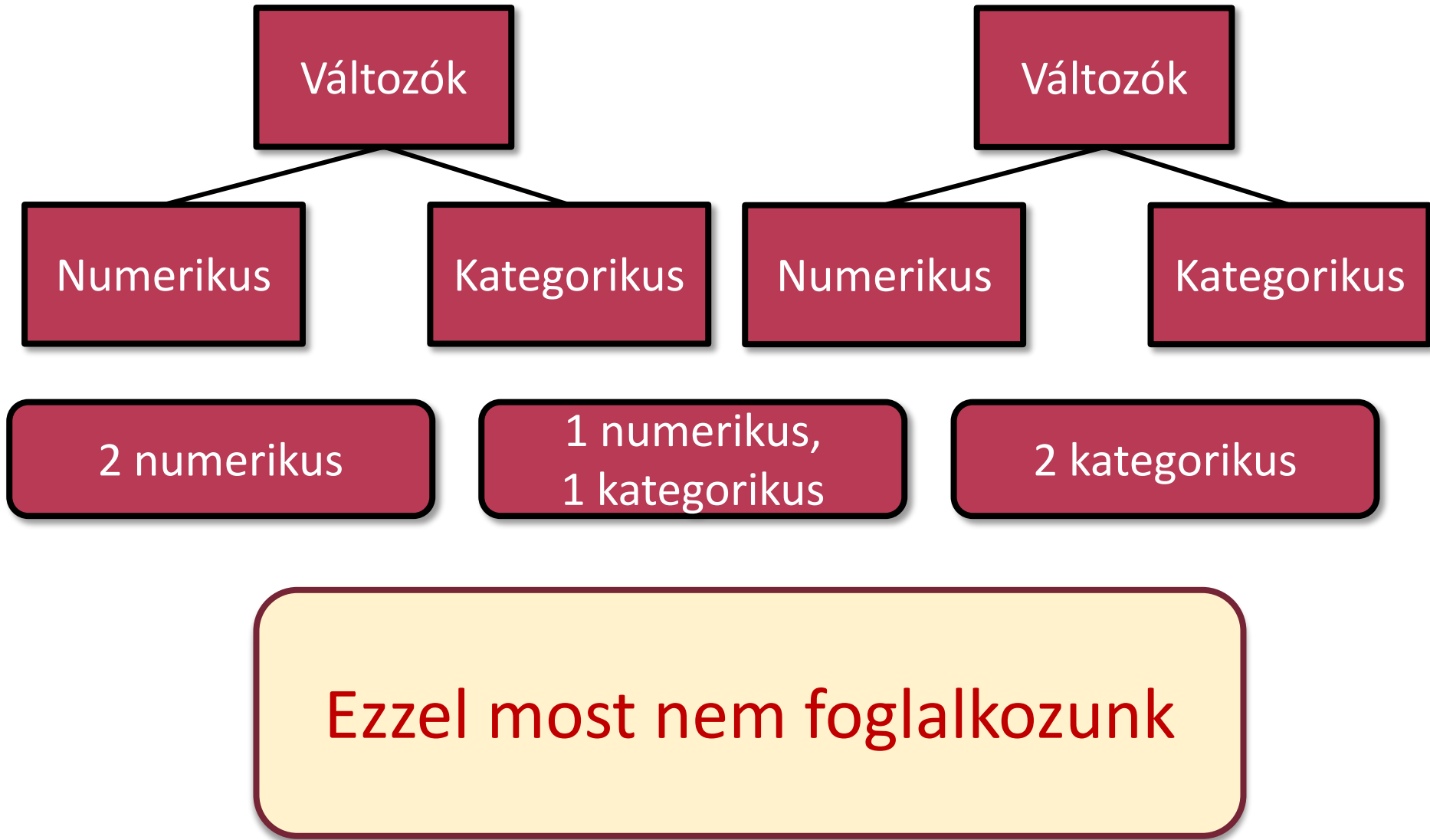
Overplotting megoldások 2: átlátszóság



Overplotting megoldások 3: méret



2 változó kapcsolata

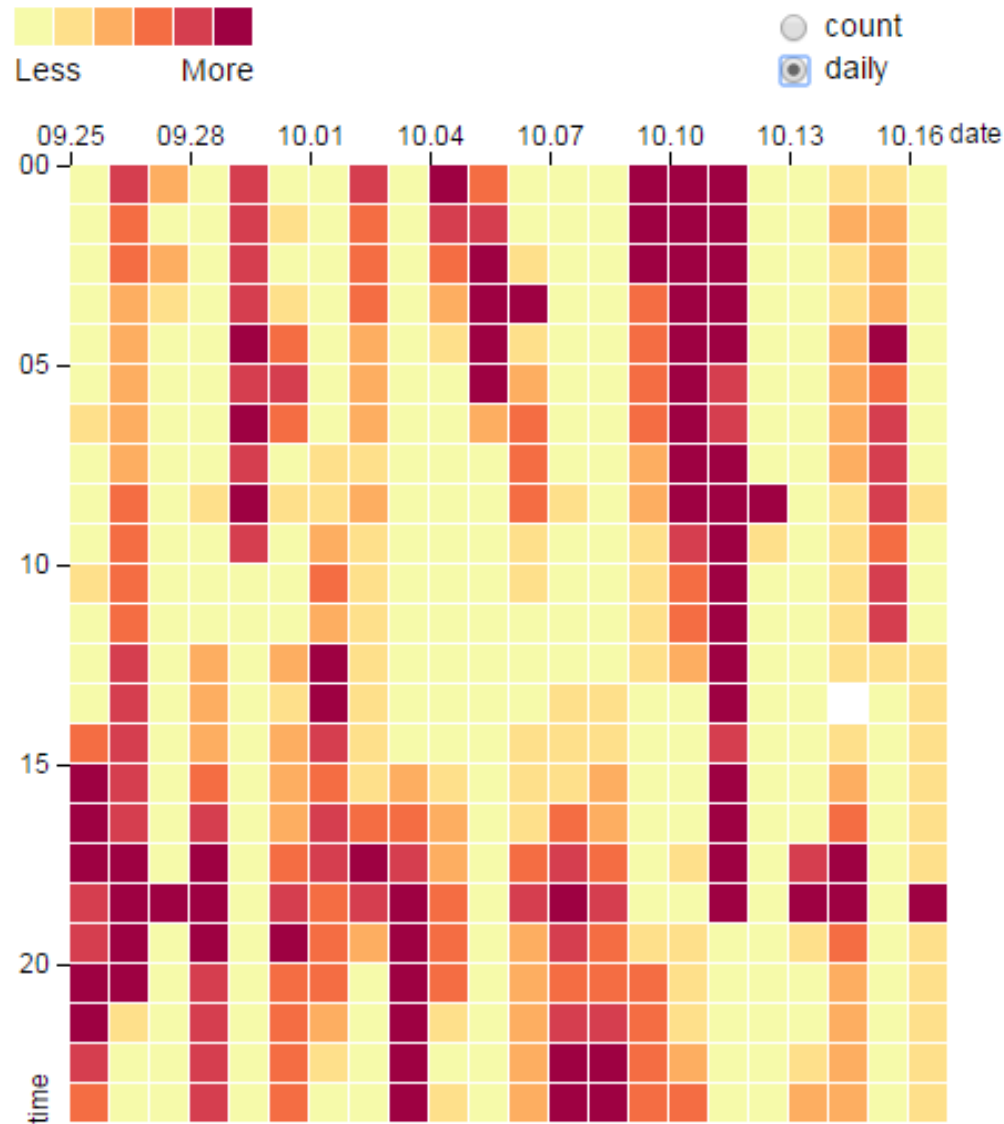


SOK VÁLTOZÓ

≥ 3 változó

- A grafikai objektumok attribútumait változtatom
 - Szín
 - Méret
 - Textúra
 - Hely – ez triviálisnak tűnik, de a treemapnél van jelentősége
- Pl. heatmap, bubble chart, treemap

Heat Map: Napi hőmérséklet



Bubble chart: átlagéletkor változása régiónként

GAPMINDER WORLD

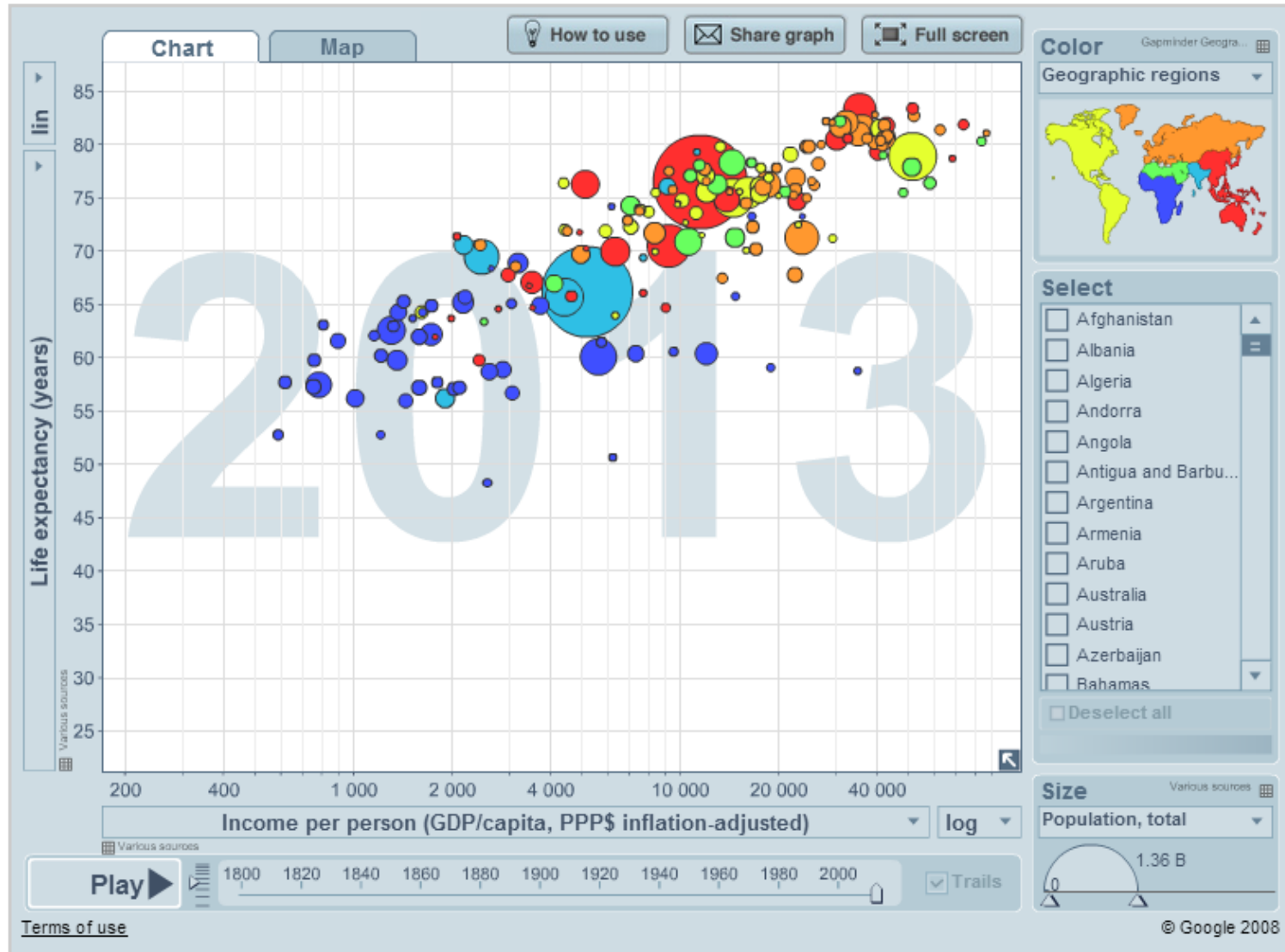
VIDEOS

DOWNLOADS

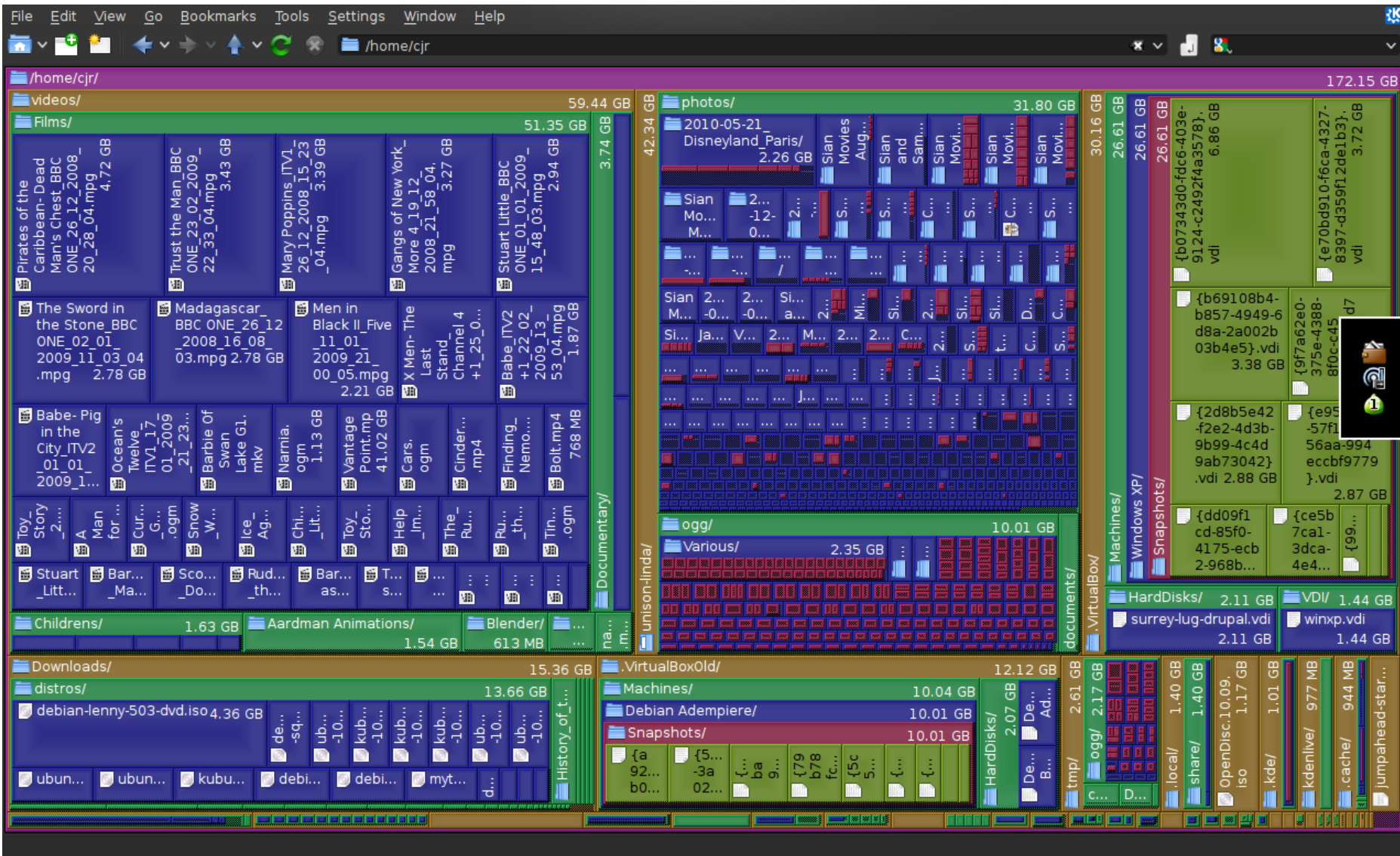
TEACH

IGNORANCE

DATA

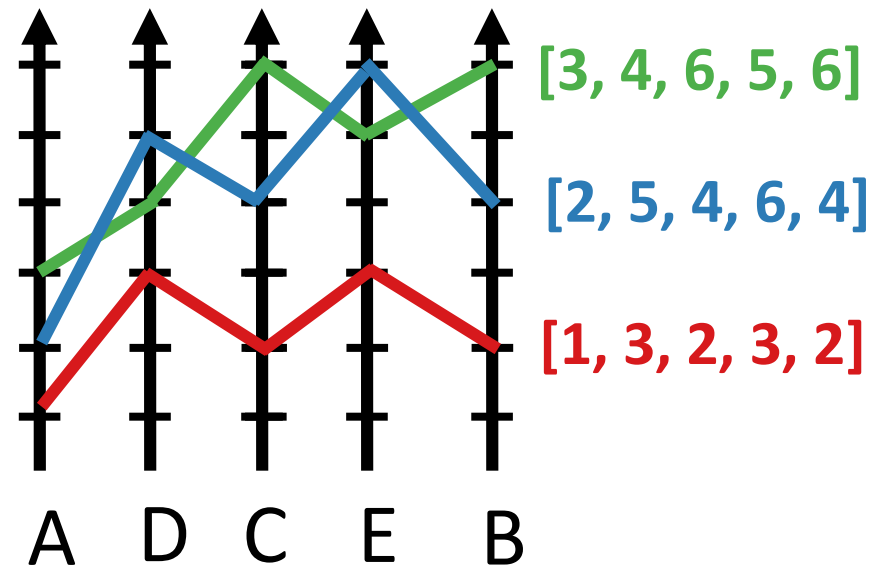
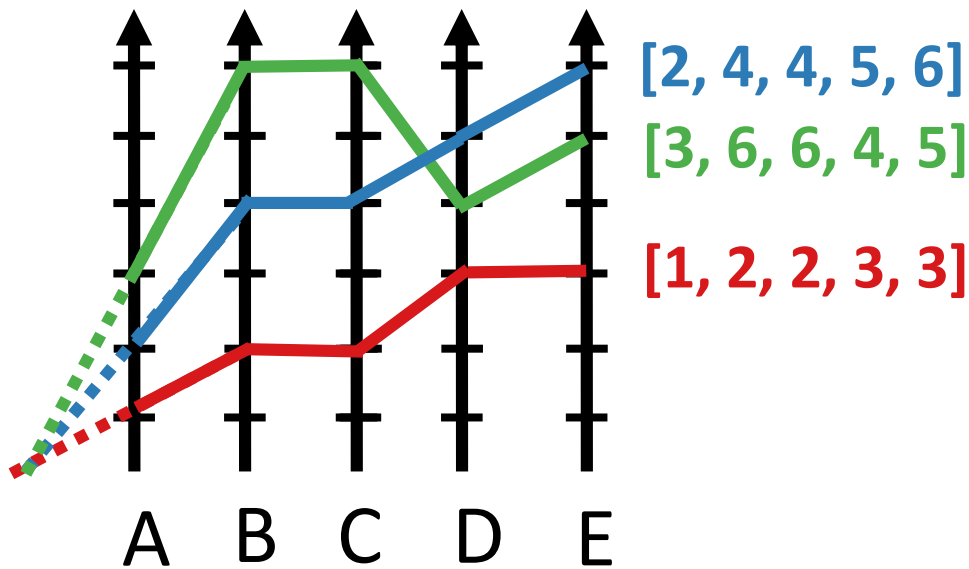


Treemap: állományrendszer



Párhuzamos koordináták

- Sokdimenziós
- Kompakt és skálázható 😊
- Koordináta sorrend?



Radar chart: egy párhuzamos koord. kiterjesztés

