

# STATISZTIKA

## Grafikus ábrázolási módszerek

Köszönetnyilvánítás: Balogh Péter,  
Jeney László, Michael Friendly

- **A statisztikai analízis célja:** összegzés;
- **A grafikus analízis céljai:** interpretálás, bemutatás;
  - Gyakran hasznosabb, ha vizuális összegzést is tartalmaz (pl. illesztett görbe, adat ellipszis);
- **Különböző célokra különböző grafikonok:**
  - □ megismerés (áttekintés);
  - □ feltárás (áttekintő sablonok, trendek);
  - □ modell diagnózis (feltételezések, kiugró értékek);

- A többváltozós adatok újszerű grafikonokat igényelnek a változók közötti kapcsolatok prezentálására;
  - szórásgörbe mátrixok;
  - vizuális elvékonyítás: a kevesebb gyakran több;
  - alacsony-D nézet (kettős rajzok);
  - HE rajzok a többváltozós próbák vizualizálására;
  - mozaik rajzok az n-szeres gyakorisági táblázatok vizualizálására;

# Grafikus ábrázolási módszerek

A statisztikai elemzés egyik alapvető módszere az összehasonlítás. Ez történhet különböző számítások révén, illetve grafikus ábrázolás segítségével.

## Előnyei:

- Szemléletes;
- Arányokat is érzékeltet;

# Grafikus ábrázolási módszerek

- Grafikus ábrázolási módszerek típusai;
  - Általános statisztikai grafikus módszerek;
  - Térképészeti eljárások;
- Funkciói: eszköz és cél;
  - Kutatási munkában elemzési eszköz;
  - Dolgozatban, prezentációban illusztrációs cél;
- Jó, ha szöveg nélkül is megállja a helyét (főleg PowerPoint-ban);
  - De: nem helyettesítheti az elemzést: (minden ábrához legyen szöveg);

# Minden lényeges információ rajta legyen (ismétlődés nélkül)

- Inkább a címben
  - Vizsgált terület: pl. Magyarország (területi szint: pl. régiók);
  - Vizsgált jelenség: pl. regionális gazdasági fejlettségi különbségek;
  - Mutató: pl. egy főre jutó GDP;
  - Vizsgált idő (vagy időszak): pl. 2004 (vagy 2004–2012);
- Inkább a kategória tengely feliratainál;
  - Mértékegység, pl. USA dollár/fő;
- Egy információ se szerepeljen egyszerre két helyen (vagy a címben, vagy a kategóriatengelyen);
- Ritkán szerepel a cím magán az Excel ábrán (nem hiba);
  - Word: ábra alá külön sorba (utólag is könnyen módosítható), PowerPoint: előfordulhat, hogy már az Excelben felkerül);
  - Mindig legyen forrásmegjelölés (PowerPoint-ban is);

# Mindig a jelenséghez tartozó ábratípust válasszunk

## ÁBRÁZOLÁSI TÍPUSOK (1);

- **mértani alakzatok felhasználásával:** többnyire koordináta rendszerben, diagramok segítségével;
- **koordináta-rendszer nélkül:** a leggyakoribb fajtája a kördiagram;

# Mindig a jelenséghez tartozó ábratípust válasszunk

## ÁBRÁZOLÁSI TÍPUSOK (2);

- **térképes ábrázolás:** a vetületi ábrázolás leggyakoribb módja; történhet szín szerinti megkülönböztetéssel, illetve lokális diagramok alkalmazásával;
- **figurális ábrázolás:** népszerűsítési célokat szolgál, a statisztikához komolytalan, csak hozzávetőlegesen tájékoztat;
- **Egyszerűbb grafikus ábrázolási módszerek;**
  - pontdiagram;
  - vonaldiagram (grafikon);
  - oszlopdiagram;
  - szalagdiagram;
  - kördiagram: ritkábban ajánlott (csak kevés körcikkel);
  - buborékdiagram;
  - radar- (sugár-)diagram;
  - háromszögdiagram;
  - hisztogram;



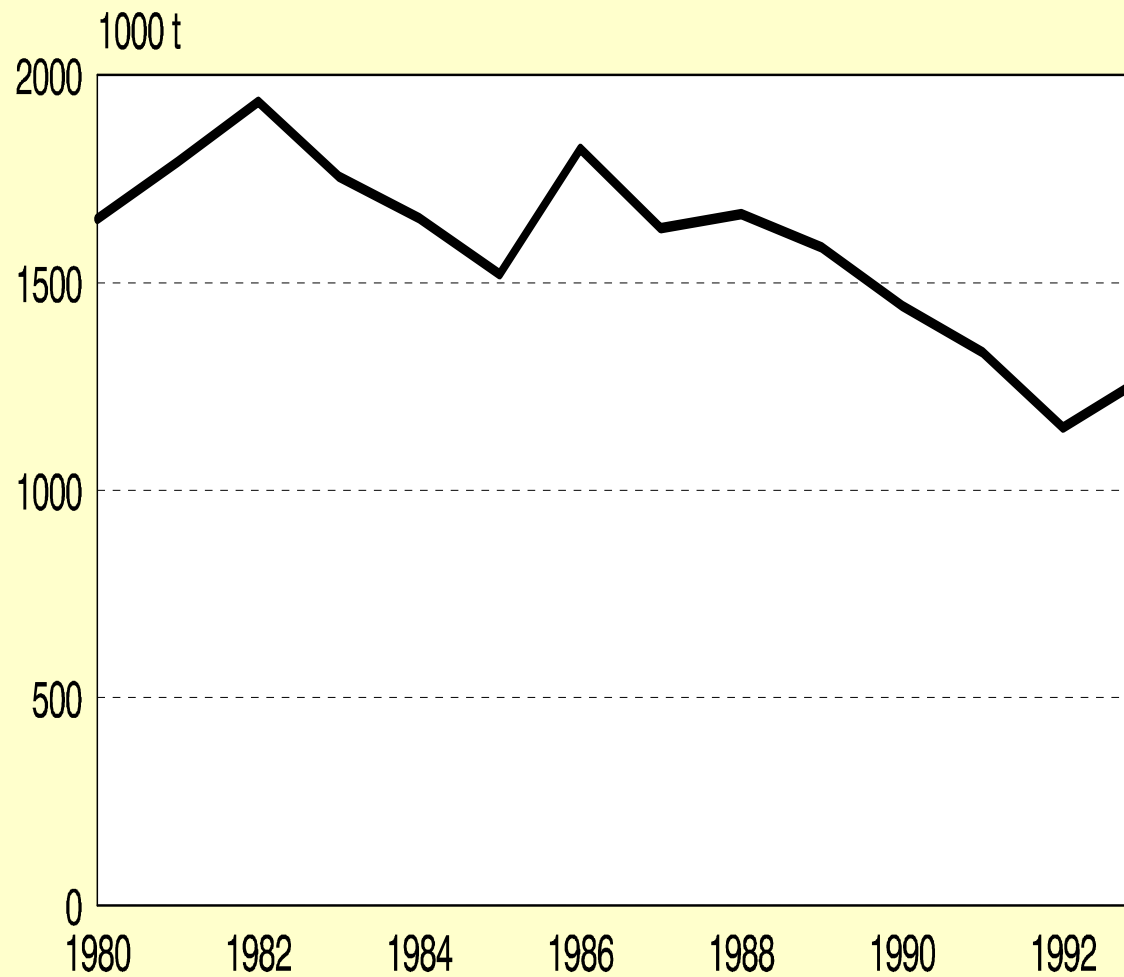
# Az ábrázolás alapelvei

- Az ábra legyen egyszerű, világos.
- A statisztikai táblához hasonlóan legyen címe, lábjegyzetben jelöljük meg az adatok forrását!
- Legyen jelmagyarázat! A táblában alkalmazott minden jelölés legyen benne! Tüntessük fel az alkalmazott mértékegységeket!
- Az ábra az összes tartozékaival (cím, lábjegyzet, jelmagyarázat, táblába írt megnevezések) a szöveggörnyezettől függetlenül is legyen világos, minden részletében érthető.
- Az ábra biztosítsa az összehasonlíthatóságot! Ha koordináta-rendszert használunk, feltétlenül tüntessük fel a tengelybeosztást, az esetleges tengelymegszakítást! Rácsozással (grid) segíthetjük a leolvasás pontosságát.

# Koordináta rendszerek

- Az ábrákat legtöbbször koordináta rendszerben helyezzük el. Ennek segítségével tudjuk leolvasni az ábrázolni kívánt méreteket.
- A tengelyeken tüntessük fel a skálabeosztást. Ha a beosztás a tengelyek metszéspontjától, 0-tól nem folyamatos, úgy a megszakítást két hullámvonallal jelöljük.
- A legtöbbször derékszögű koordináta rendszert használunk.

## Gyümölcsstermelés alakulása Magyarországon

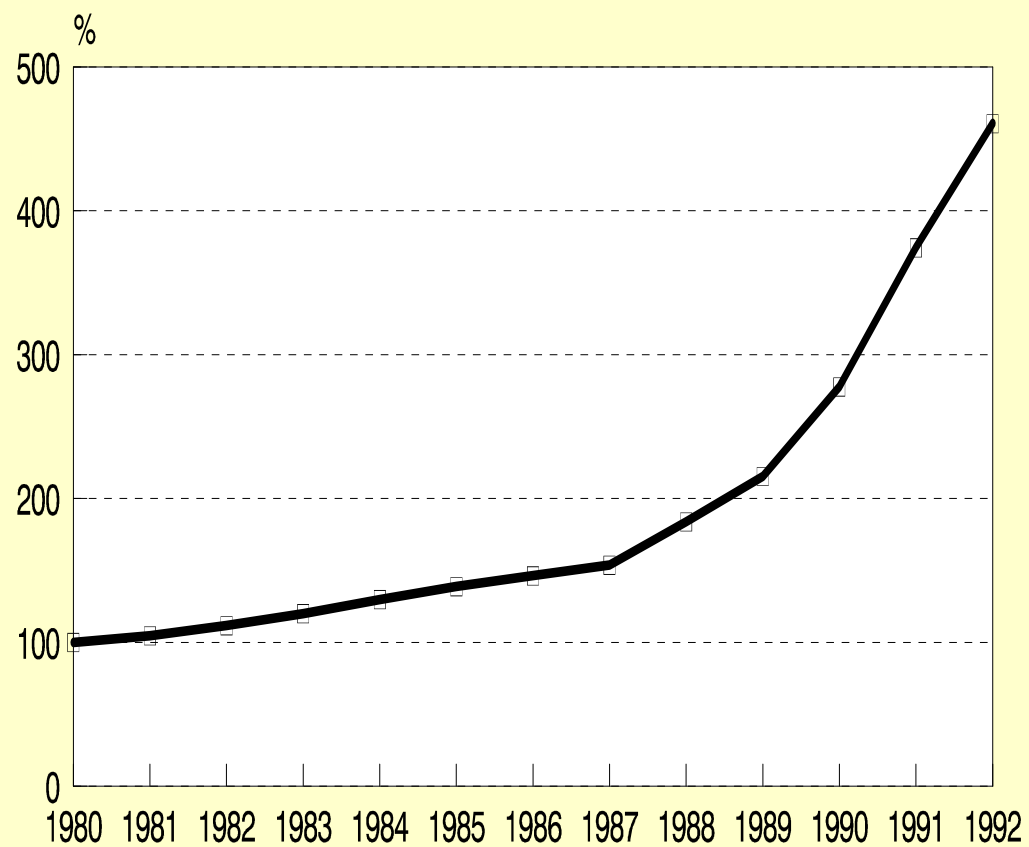


Magyar statisztikai Évkönyv 1993

# Fogyasztói árindex

(1990-ig munkás-alkalmazotti)

1980=100



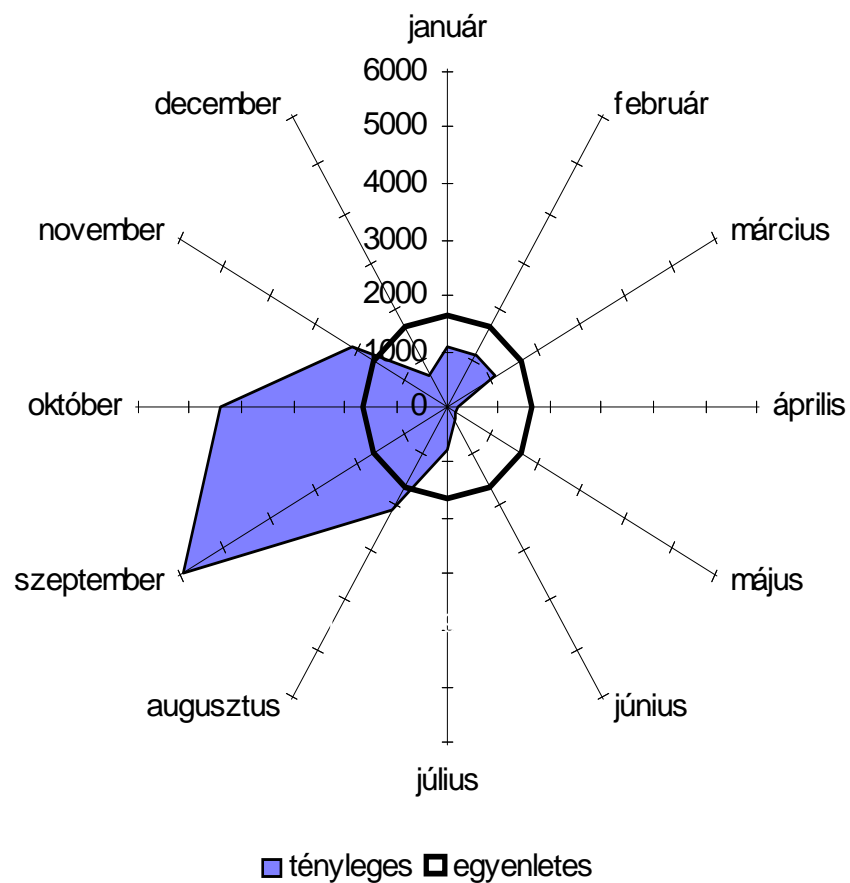
Forrás: Magyar statisztikai évkönyv

# Polár koordináta rendszer

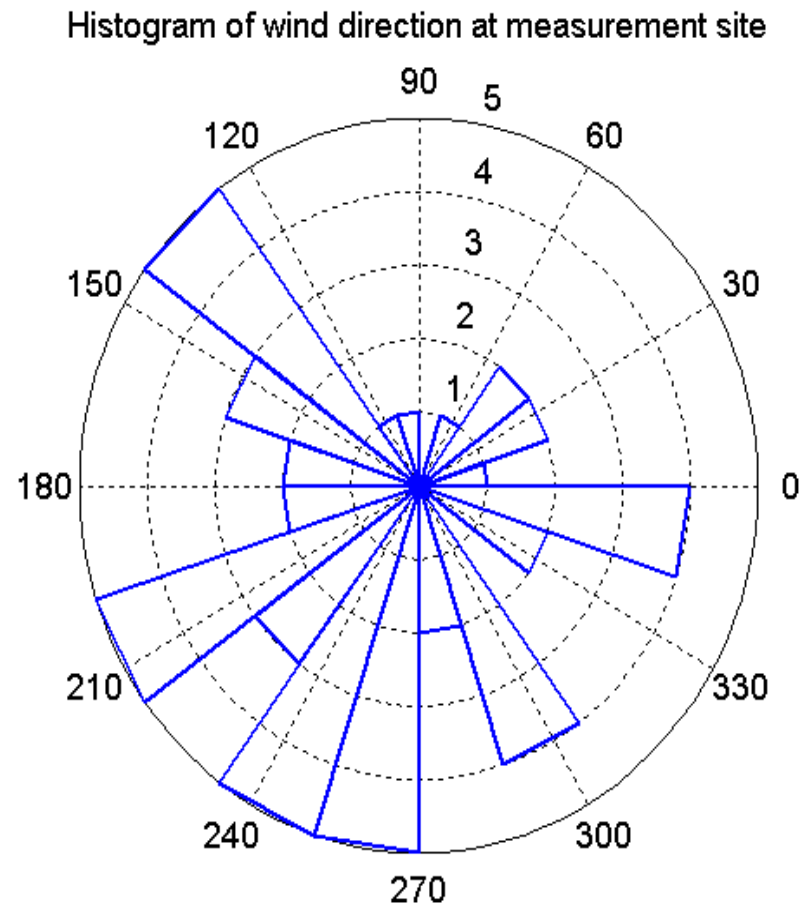
- Polár koordináta rendszerben a tengelyek más szöget zárnak be egymással, és csak a pozitív féltengelyeket szoktuk ábrázolni.
- Használatos az éven belüli szezonális ábrázolására, ahol a 12 hónapot egy-egy tengely képviseli. A szezonális hiányát jelzi, ha az adatok egy origo középpontú körön helyezkednek el. Minél nagyobb az eltérés ettől, annál erősebb a szezonális.

# Polár koordináta rendszer

A vöröshagyma felvásárlás (t) alakulása, 1993



# Polár koordináta rendszer - rózsa diagram



# Polár koordináta rendszer

- Használatos még pl. az élelmiszerek érzékszervi vizsgálata eredményeinek közlésénél.
- Ha egy terméket több tulajdonságával jellemzünk, mindegyiket egy-egy tengelyen feltüntetve a termék minőségét jellemző poláris görbét (más szóval pókháló diagramot) kapunk.
- Ezek összehasonlítása megkönnyíti a tulajdonságok együttes figyelembe vételét.



# Grafikus ábrák típusai

- vonaldiagram;
- hisztogram;
- oszlop- és szalagdiagram;
- kör- és egyéb diagram;
- kartogram (térkép);
- piktogram (jelenség "képe");

# Hisztogram

A **hisztogram** olyan oszlopdiagram, amelyen az oszlopok **területe** arányos a gyakorisággal.

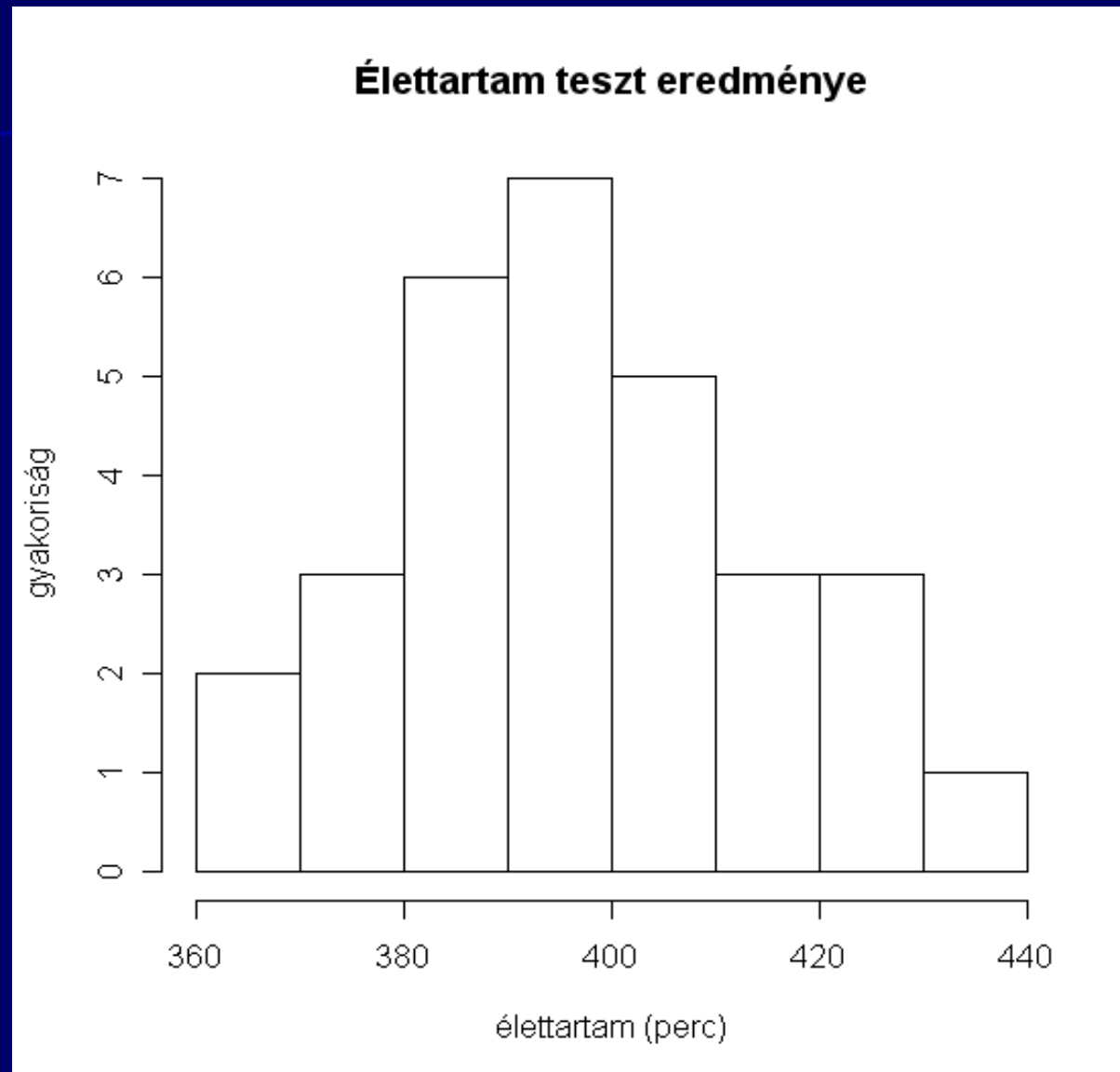
A vízszintes tengelyen: osztályközök;

A függőleges tengelyen: **sűrűség**: *egységnyi osztályközre* jutó (relatív) gyakoriság.

## FIGYELEM!

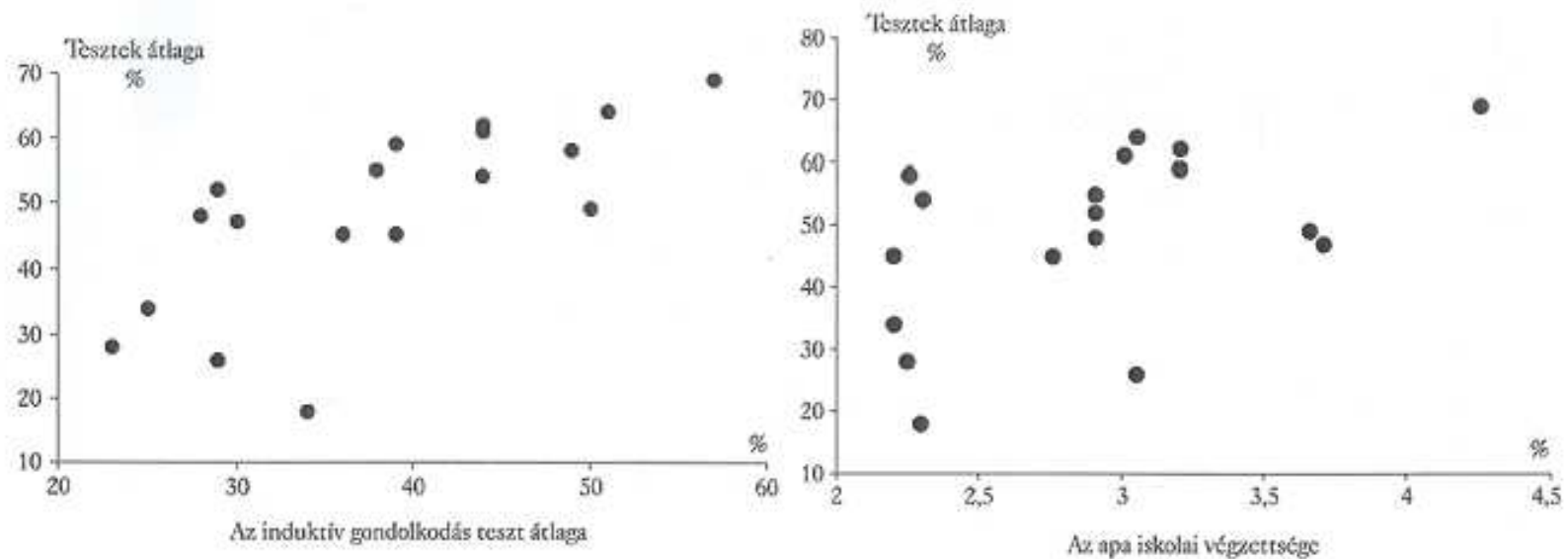
- **Nem egyenlő** osztályközök esetén csak a **sűrűség** lehet a függőleges tengelyen!
- **Egyenlő** osztályközök esetén a (relatív) **gyakoriság** is jó, mert **az** arányos a **sűrűséggel**.

# Hisztogram

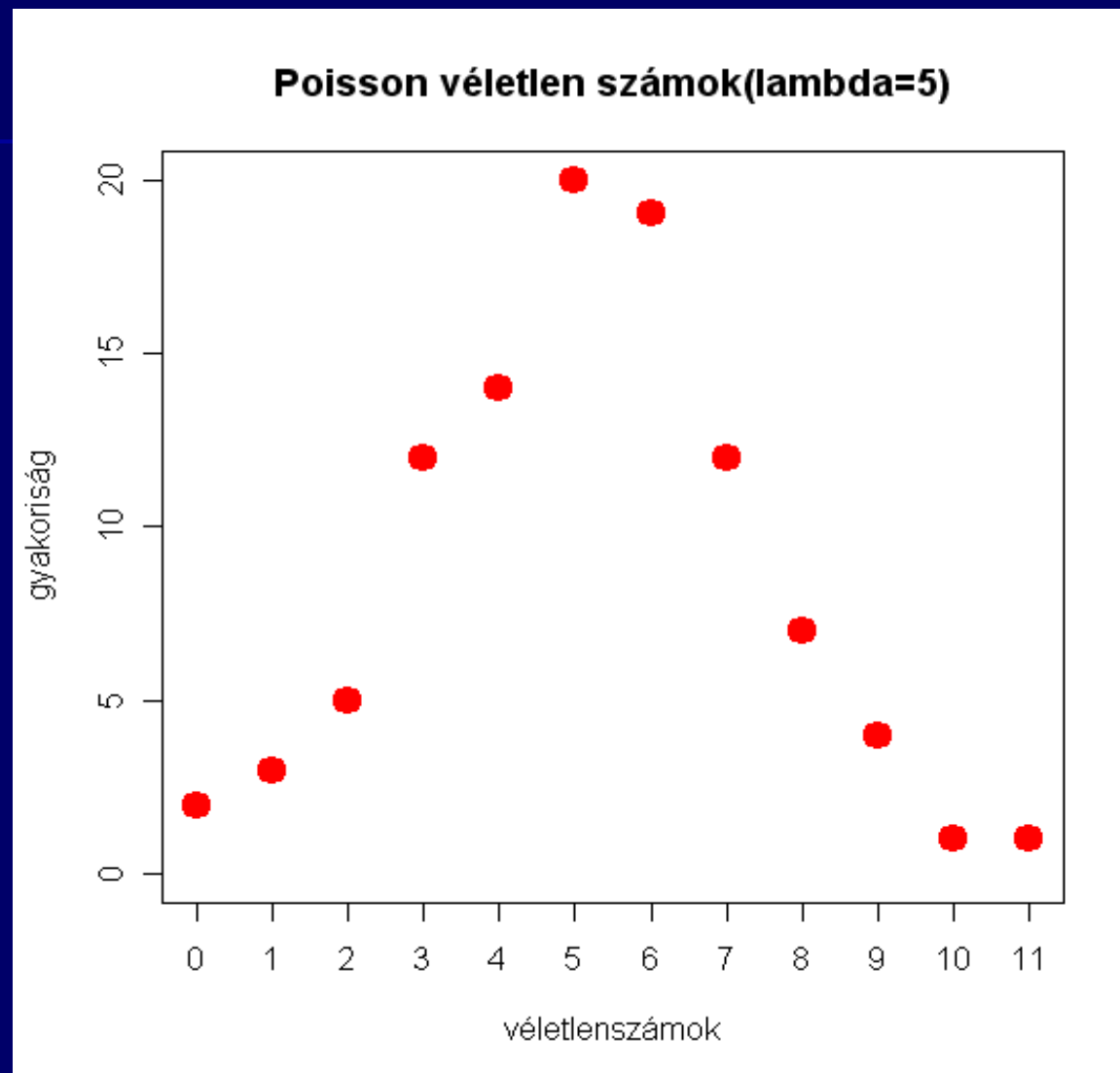


# Pontdiagram

- Pontdiagram: két egymással összefüggésben lévő mennyiségi ismerv értékeinek ábrázolása koordinátarendszerben. (idő és menny. sorok)

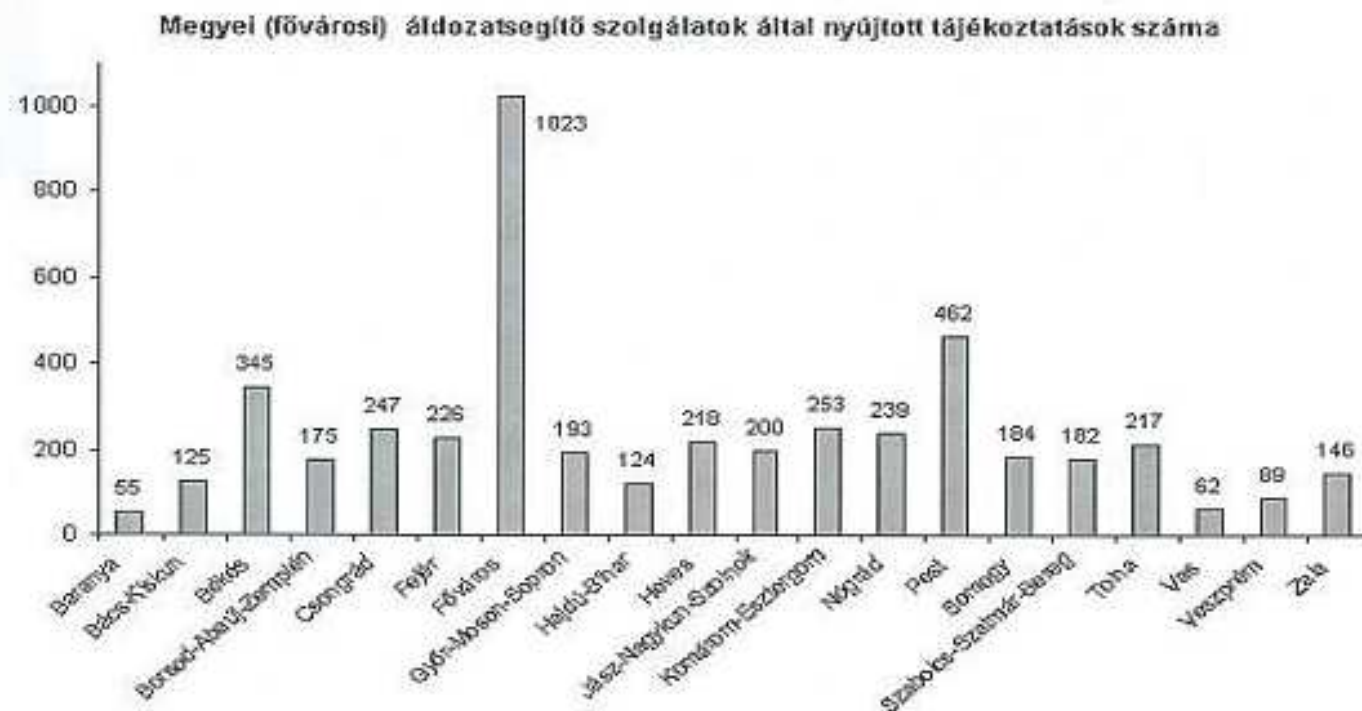


# Pontdiagram

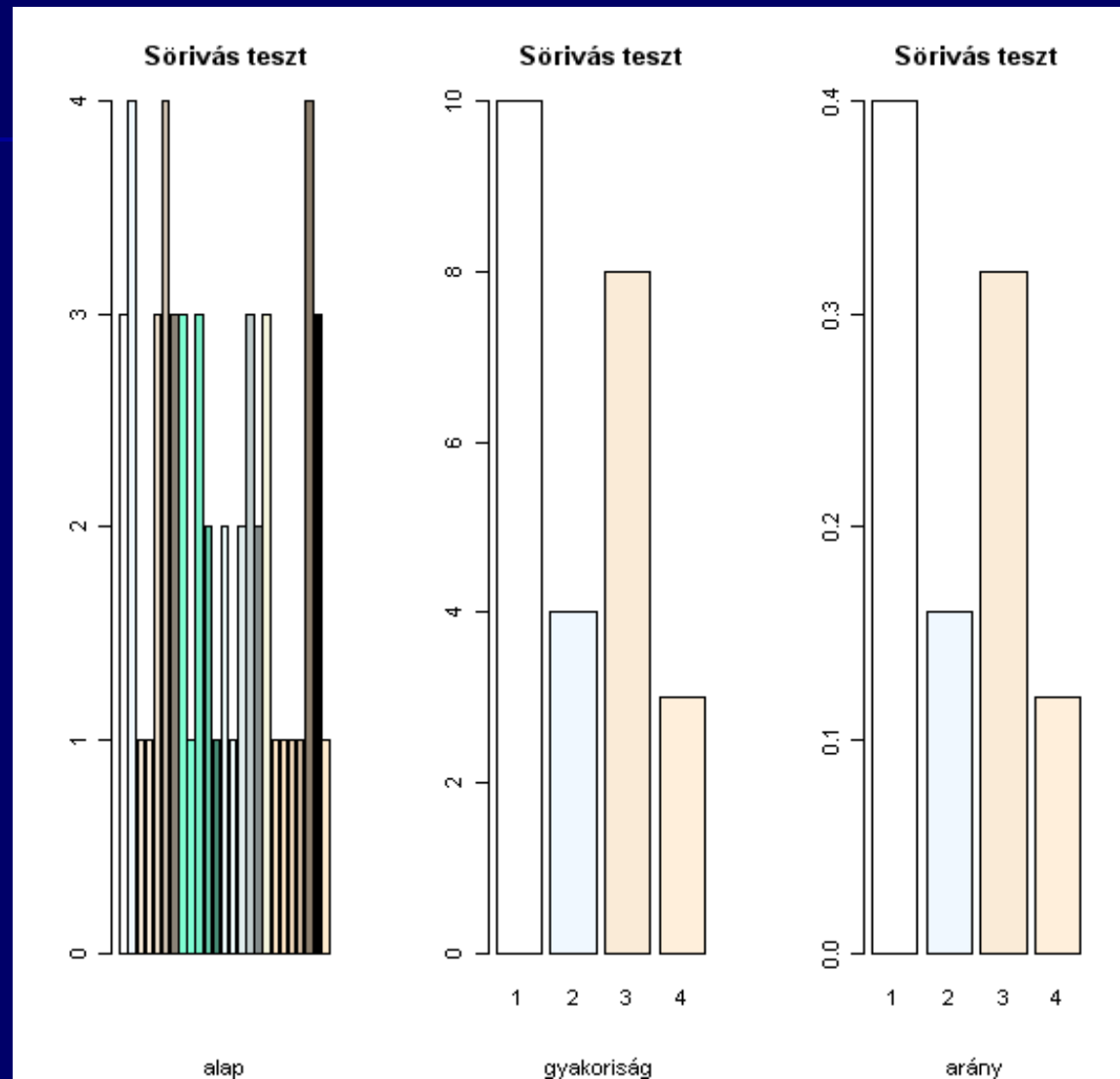


# Oszlopdiagram

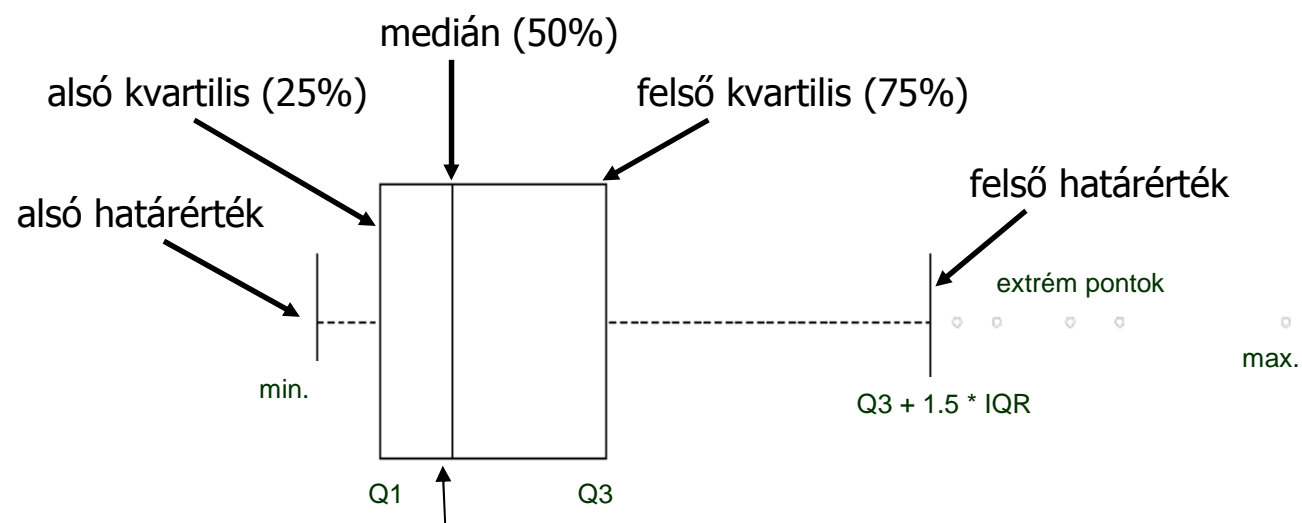
- Oszlopdiagram: összehasonlítás az oszlopok magasságával. (összehasonlítás)



# Oszlopdiagram

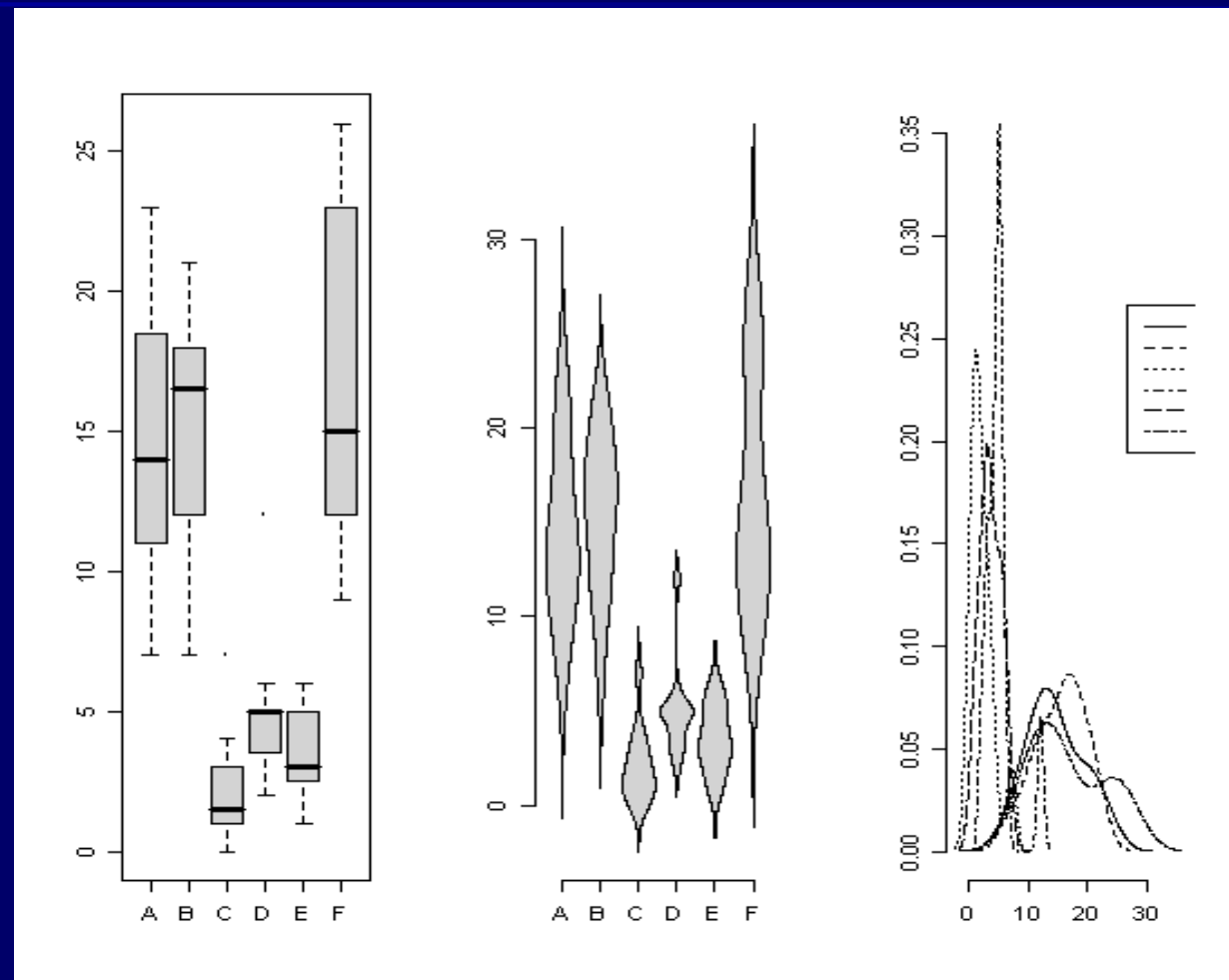


# Box-plot ábrázolás





# A violindiagram ábrázolása a boxplot és a sűrűségdiagram társaságában



# Egy grafikus ábra szerkesztésének a szabályai

- ❑ Ki kell választani a legmegfelelőbb ábrázolási módot;
- ❑ A kiválasztott grafikus ábra legyen egyszerű, világos, áttekinthető, kifejező;
- ❑ A kiválasztott ábra segítse elő az általa kifejezett jelenségek arányainak, összefüggéseinek helyes megértését;
- ❑ Minden grafikon két részből álljon:
  - ábra;
  - magyarázó jelölések (cím, skála, jelmagyarázat);
  - esetleges lábjegyzet;

# Vonaldiagram

- Hézagmentes, egyenlő közű idősorok és azonos csoportközű mennyiségi sorok ábrázolására szolgál.
- Lényeges a helyes léptékarány megválasztása.
- Gyakran előfordul, hogy egy grafikus ábrán vizsgáljuk több egymással összefüggő jelenség alakulását.
- Abszolút adatok helyett célszerűbb közös bázisból kiindulni.

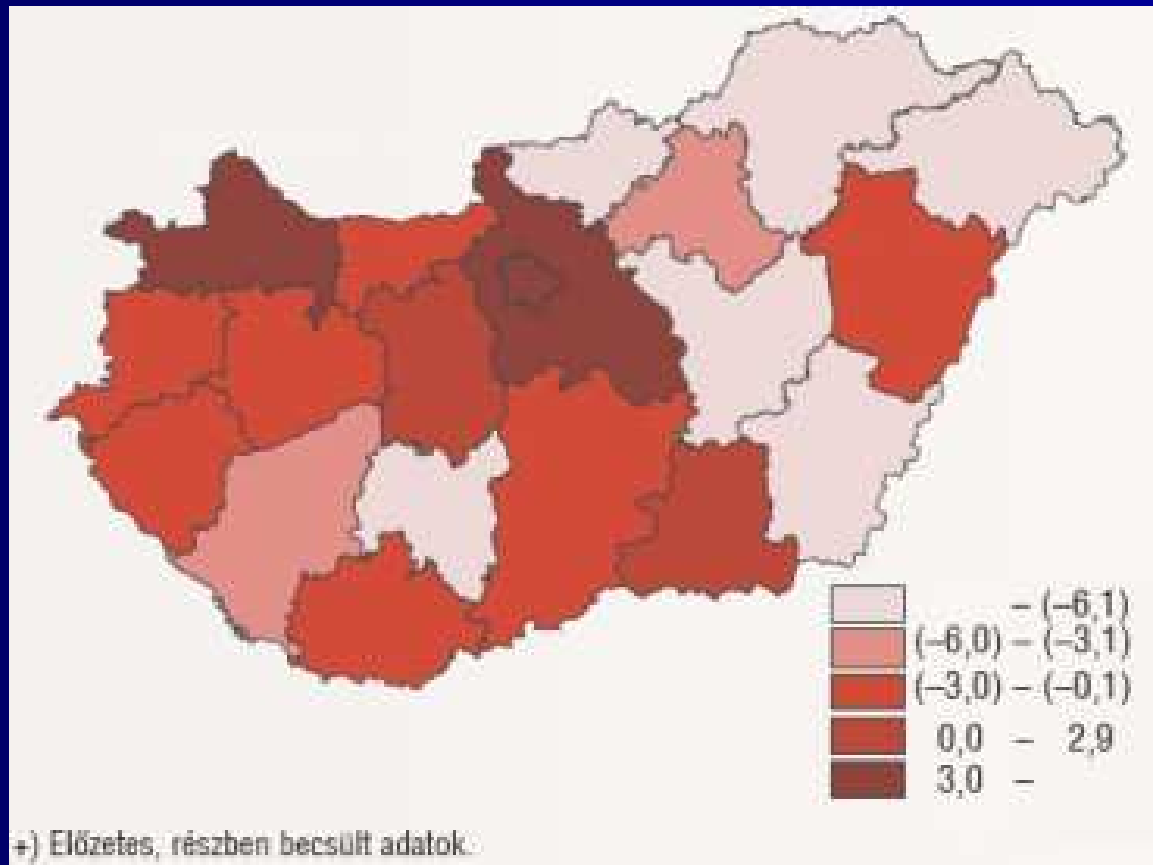
# Oszlop- és szalagdiagram

- A leíró sor kivételével bármely statisztikai sort ábrázolhatjuk ilyen módon.
- Célszerű egyenlő alapú oszlopokkal végezni az elemzést, mert így az adatok közötti arányt az oszlopok magassága érzékelteti.
- Idősorok ábrázolásánál vonaldiagram helyett ezt a formát akkor használjuk, ha
  - ☐ csak néhány időszakra vonatkozó adat áll rendelkezésre;
  - ☐ nem hézagmentes idősor alapján végezzük az ábrázolást;
  - ☐ nem egyenlő időközökre vonatkozó adatokat hasonlítunk össze;

# Kartogram, illetve kartodiagram

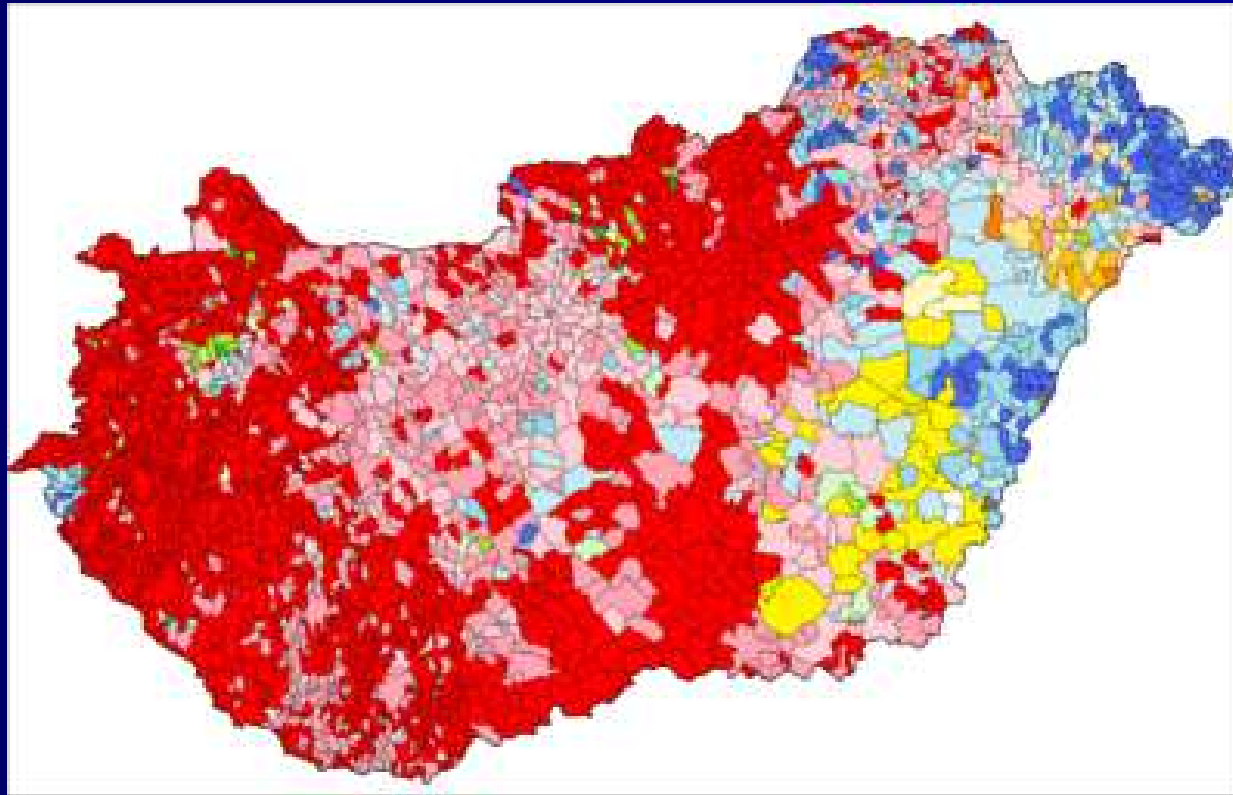
- Ezt a fajta ábrázolási módot területi sorok elemzésénél alkalmazzák, ha a területi adatok teljesek.
- Az ábrázolás lényege, hogy **a térképen a területi egységek valamilyen jellemzőjét ábrázoljuk különböző színezéssel, vonalazással.**
- **Kartodiagram** esetén a mennyiségi adat egy felületre, a térkép meghatározott területére vonatkozik. A vonatkoztatási terület bármilyen nagyságú lehet: általában statisztikai, közigazgatási területegység. A diagramot úgy kell elhelyezni a területen belül, hogy a hovatartozása egyértelműen, könnyen felismerhető legyen. Kis felület esetén annak közelébe, nagyobb felületnél annak súlypontjába helyezzük.

- **Kartogram** – a színezés jelzi az adott mutató nagyságát;



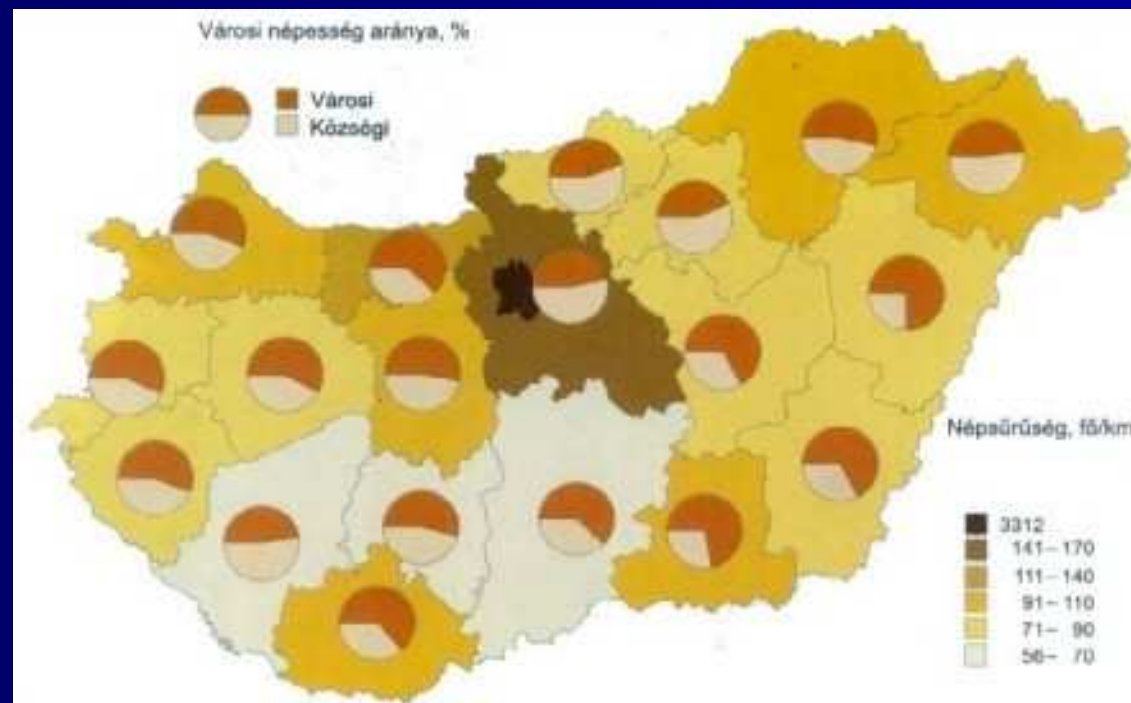
Ezer lakosra jutó belföldi vándorlási különbözet 2009 (Forrás: Statisztikai tükör 2010/42)

- **Kartogram** – a színezés jelzi az adott mutató minőségét



A legnagyobb vallási csoportok Magyarországon. A **piros** és **rózsaszín**: római katolikus; **kék**: református; **zöld**: evangélikus; **barna**, vagy **narancssárga**: görög katolikus; **citromsárga**: a többség vallási közösséghez nem tartozik (2001).

- **Kartodiagram** – A térkép különböző pontjain diagramokat helyeznek el, ezzel szemléltetve a mutató nagyságát

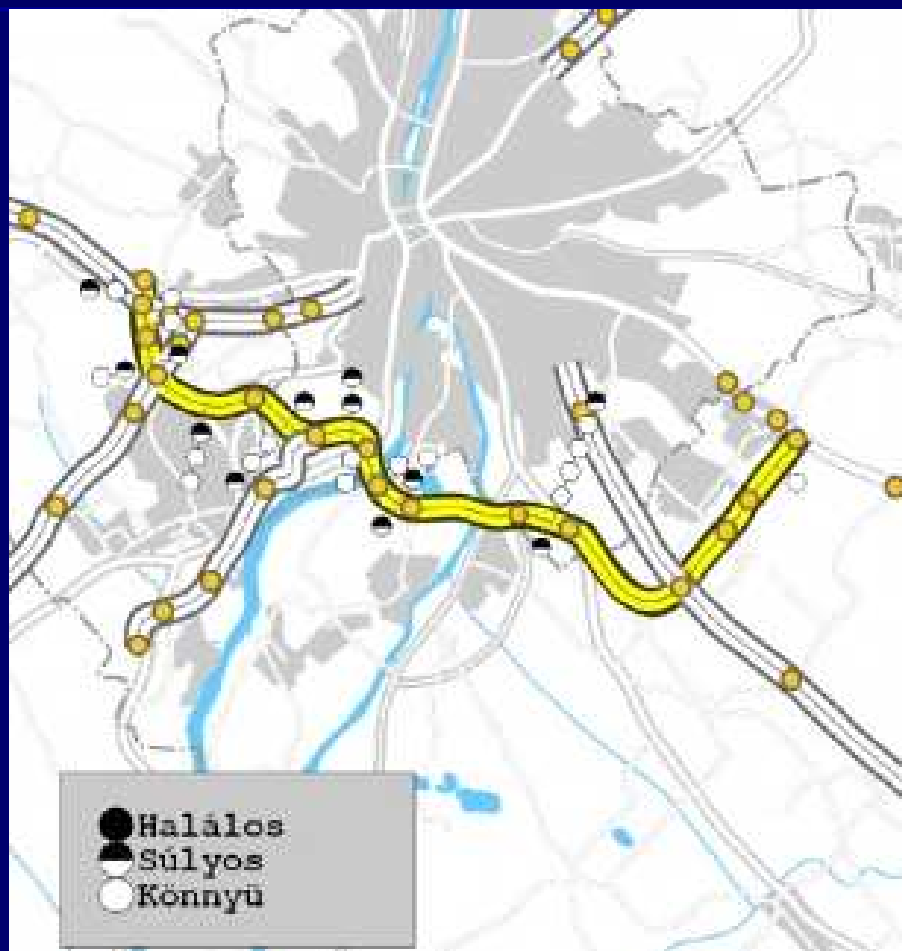


A népsűrűség és a városi népesség aránya 2002 január 1 (Forrás: sdt.sulinet.hu)



# Ponttérkép

- **Ponttérkép** – egyfajta sűrűségi mutató, a pontok mennyisége jelzi a mutató nagyságát



M0 autóbaleseti ponttérkép 2007. I-VIII. hó (Forrás: [www.police.hu](http://www.police.hu))

# Piktogram

- Ez az ábrázolási fajta (egy fogalmat megjelenítő jel) az adatok közötti arányt, a jelenséget ábrázoló "képek" területének egymás közötti arányával fejezik ki.  
Pl. közlekedési útjelző táblák; tájékoztató ábra, melyet felirat helyett alkalmaznak); képírásban használt jel);

# Piktogram

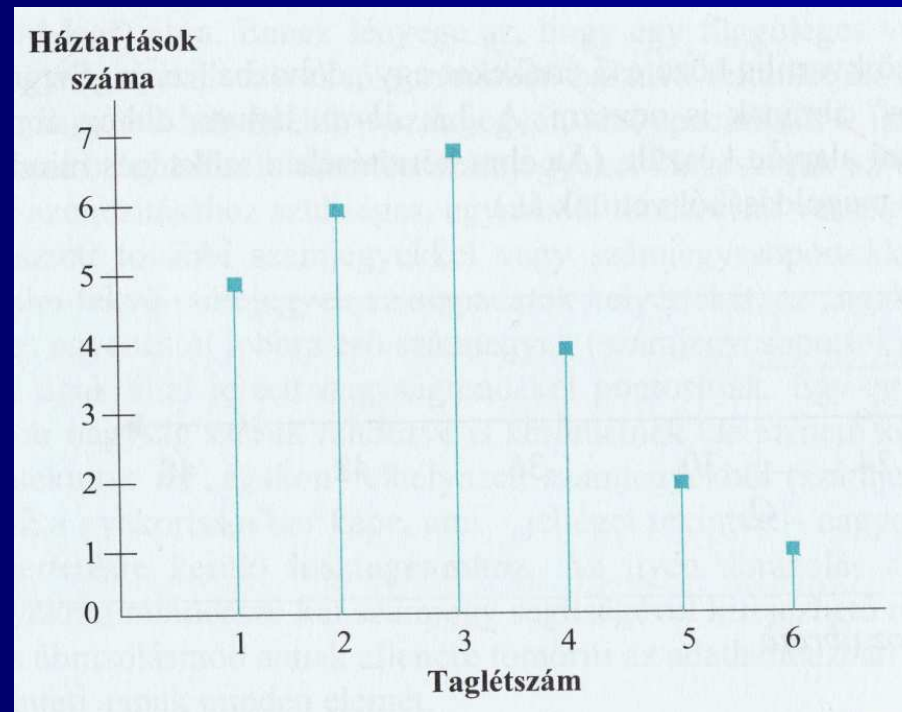
- Piktogram:  
figurális  
ábrázolás, mely  
a jelenséget  
megtestesítő  
különböző  
nagyságú  
figurák alapján  
fejezi ki a  
nagyságrendi  
relációt.



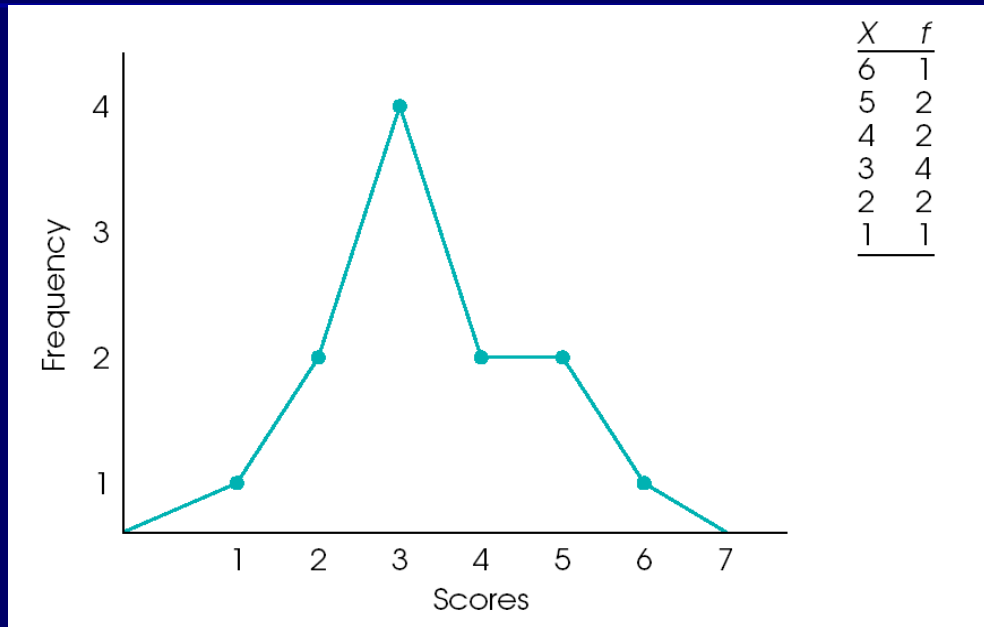
# Gyakorisági eloszlások ábrázolása

Pálcika diagram;

- Néhány értéket felvevő diszkrét mennyiségi ismérvek esetében;



# Gyakorisági poligon (1)



A smoothed out histogram

Make a point representing  $f$  of each value

Connect dots

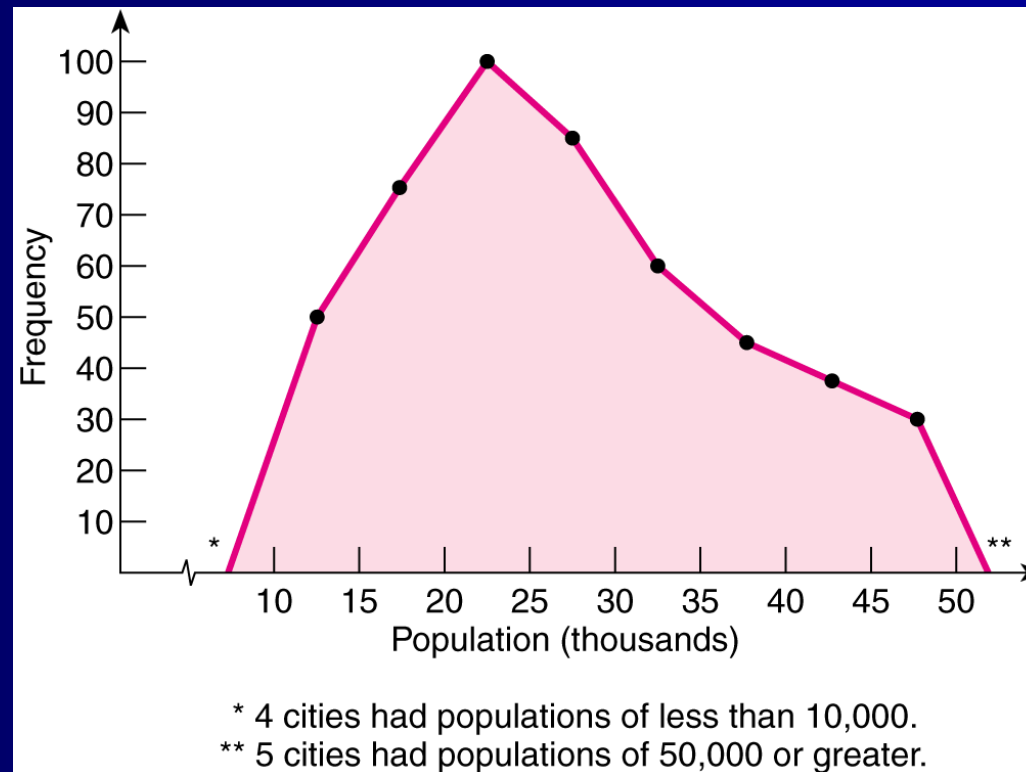
*Anchor* line on x axis

Useful for comparing distributions in two samples (in this case, plot  $p$  rather than  $f$ )

# Gyakorisági poligon (2)

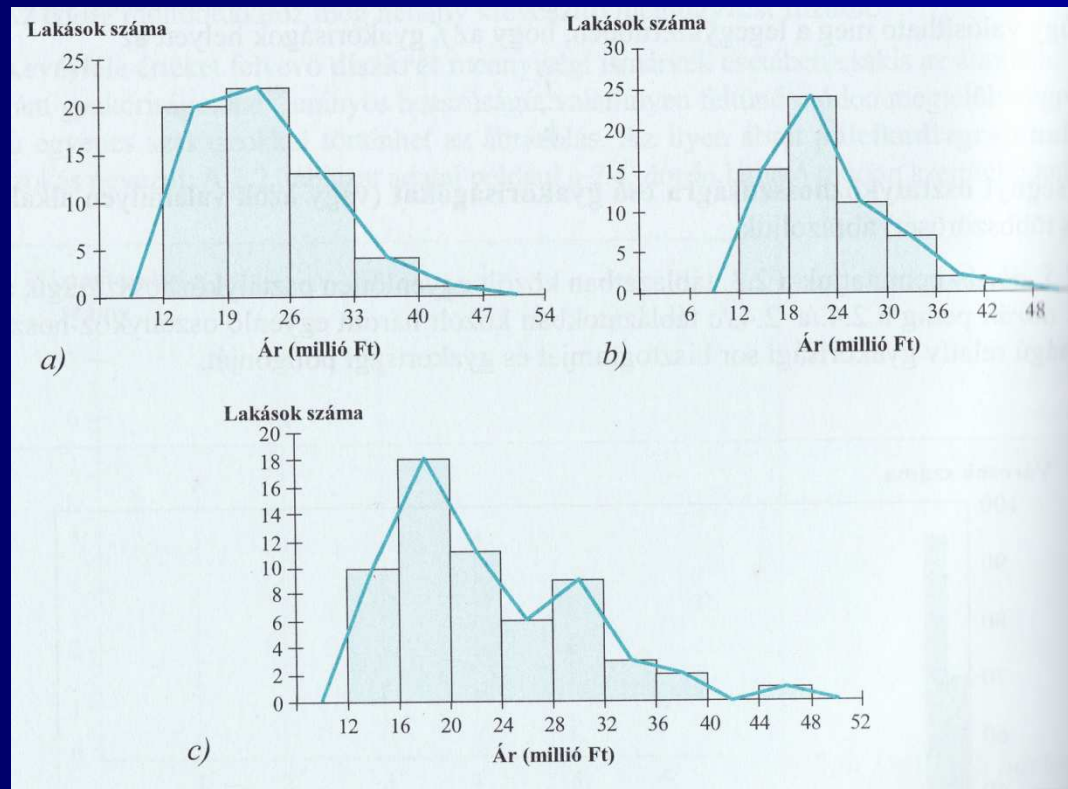
## ■ Frequency Polygons

- Depicts information from a frequency table or a grouped frequency table as a **line graph**

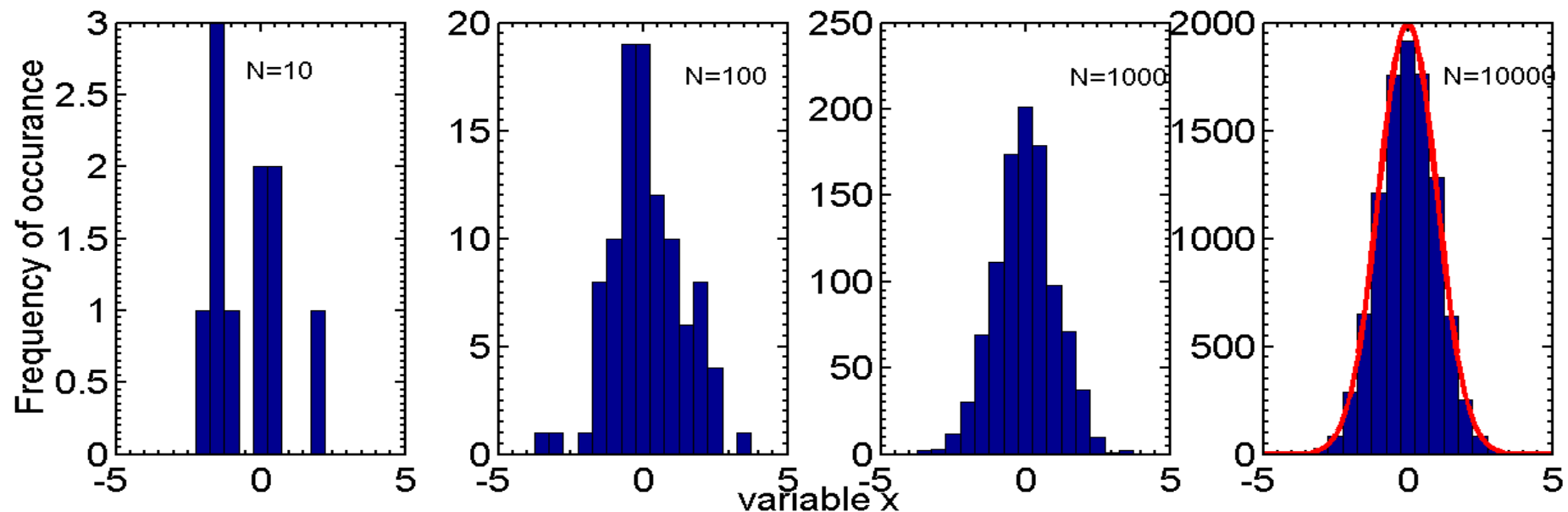


# Hisztogram, poligon

- Az oszlopok területe arányos kell, hogy legyen az ábrázolt gyakorisággal.
- Eltérő osztályköz hosszúság esetén az  $f_i$  gyakoriságok helyett **az  $f_i/h_i$  egységnyi osztályköz hosszúságára eső gyakoriságokat ábrázoljuk.**



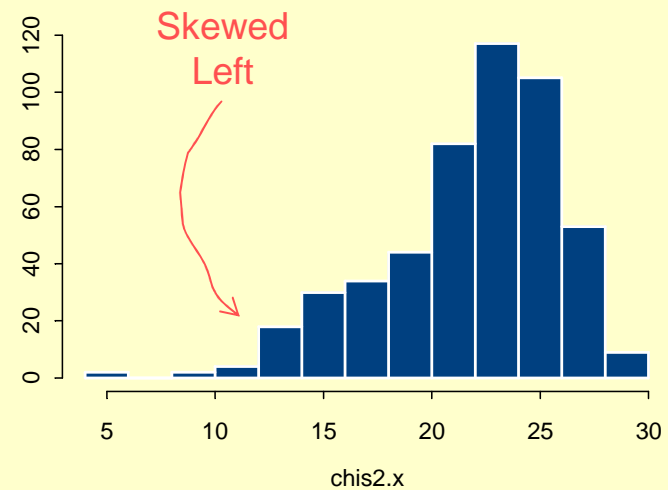
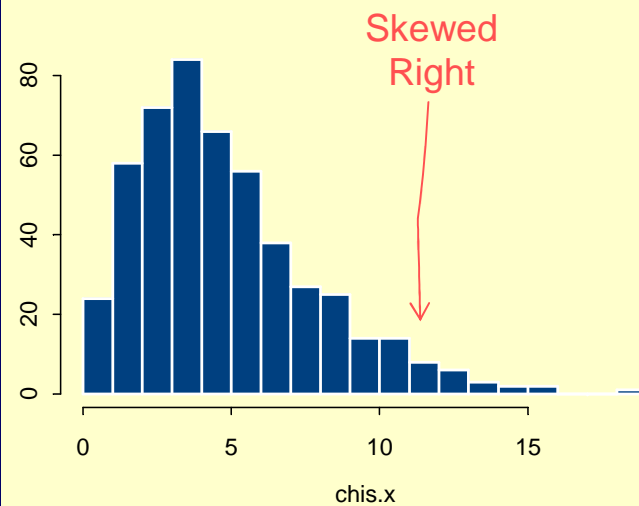
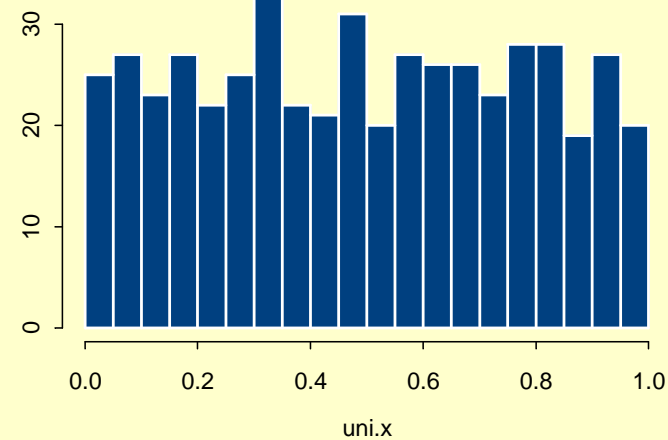
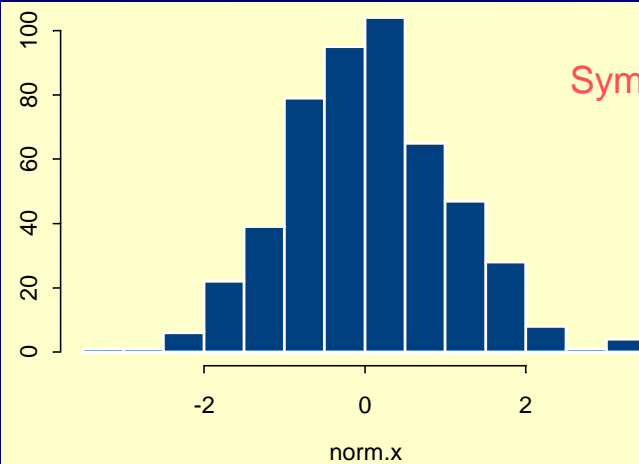
# Véletlen hibák gyakorisági eloszlása - oszlopdiagram, hisztogram



- ❖ As number of measurements increases the distribution becomes more stable;
  - The larger the effect the fewer the data you need to identify it;
- ❖ Many measurements of continuous variables show a bell-shaped curve of values, known as a Gaussian distribution.

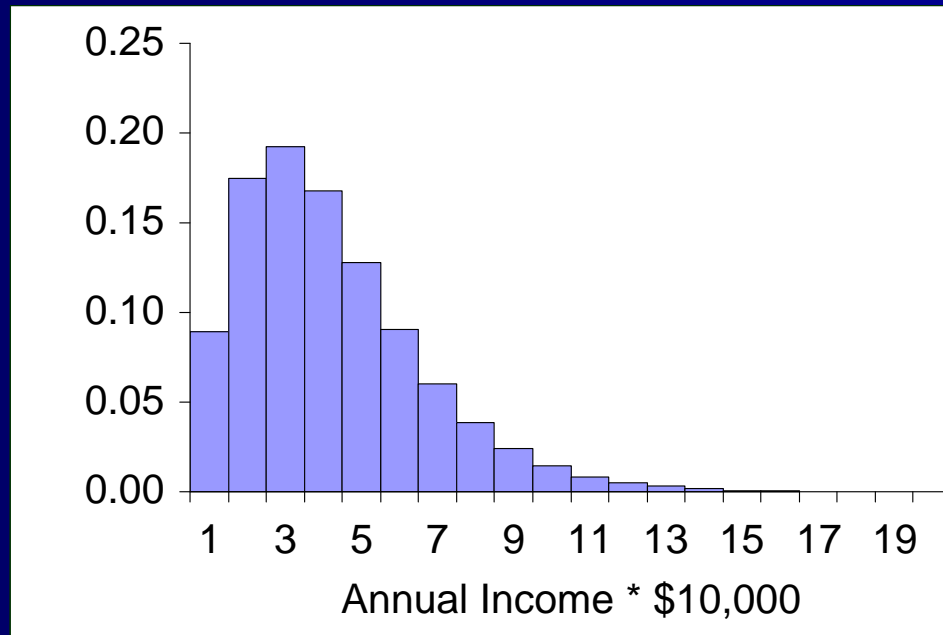


# Szimmetrikus, vs. ferde



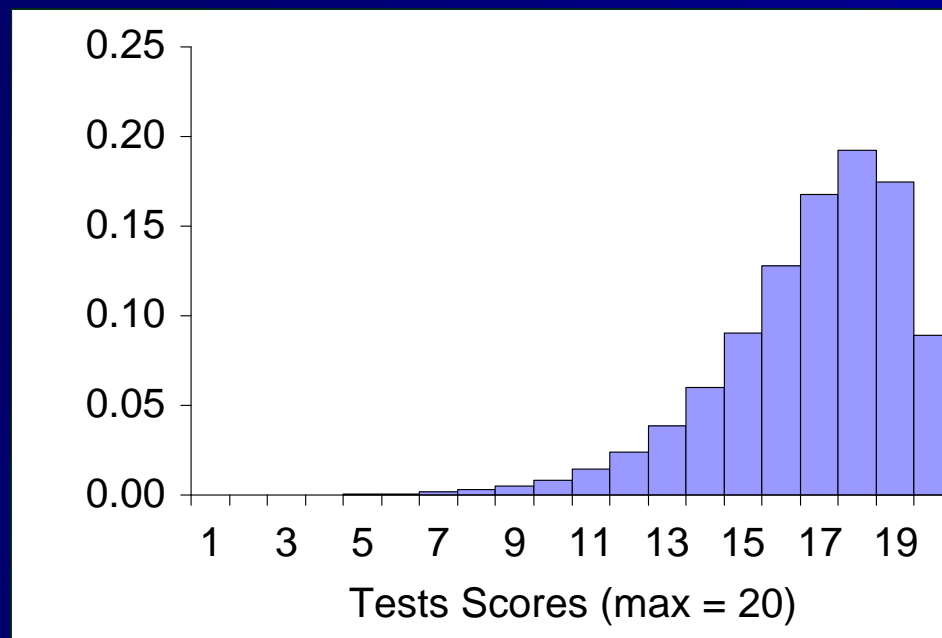
# Ferde gyakorisági eloszlás

- Positively skewed;
  - Skewed right;
  - Tail trails to the right;
  - \*\*\* *The skew describes the skinny end* \*\*\*

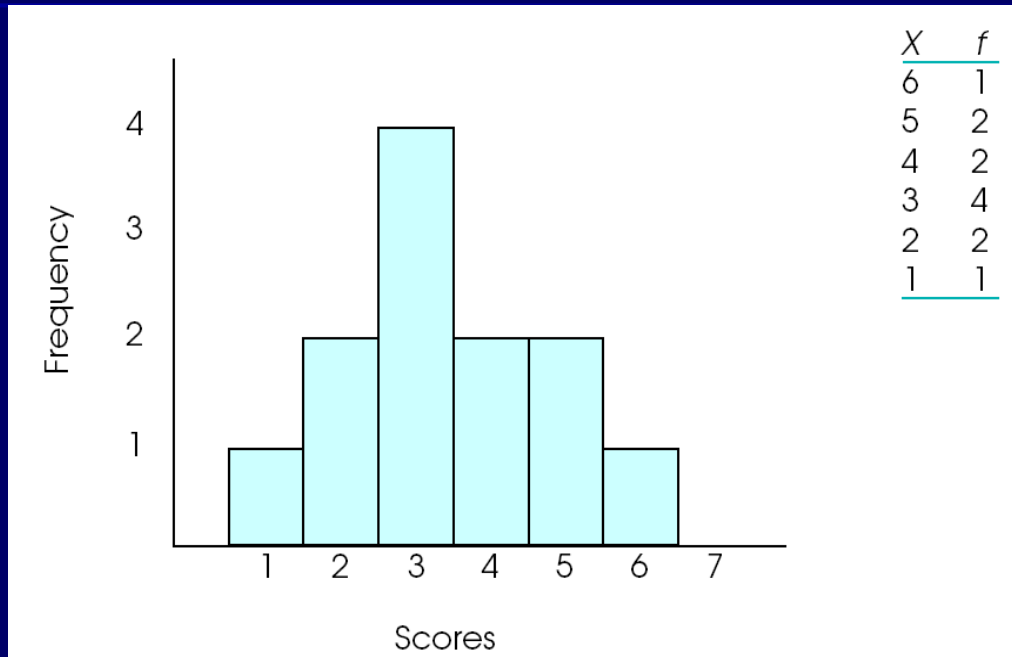


# Ferde gyakorisági eloszlás

- Negatively skewed;
  - Skewed left;
  - Tail trails to the left;

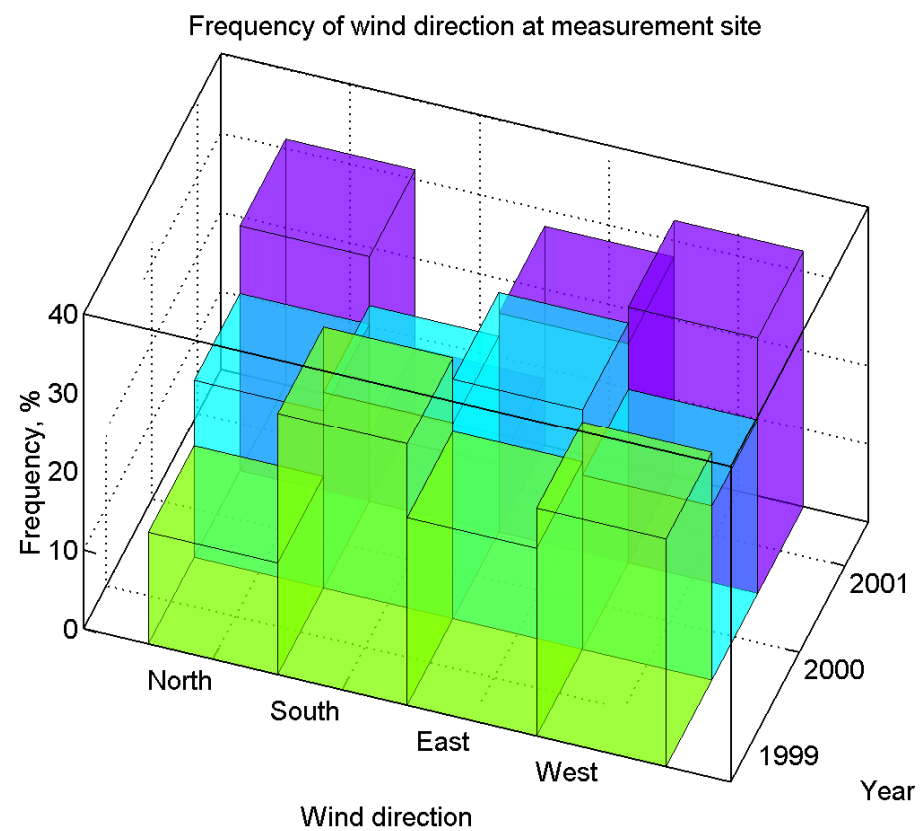


# Egyváltozós hisztogram



- $f$  on  $y$  axis (could also plot  $p$  or % )
- $X$  values (or midpoints of class intervals) on  $x$  axis
- Plot each  $f$  with a bar, equal size, touching
- ***No gaps between bars***

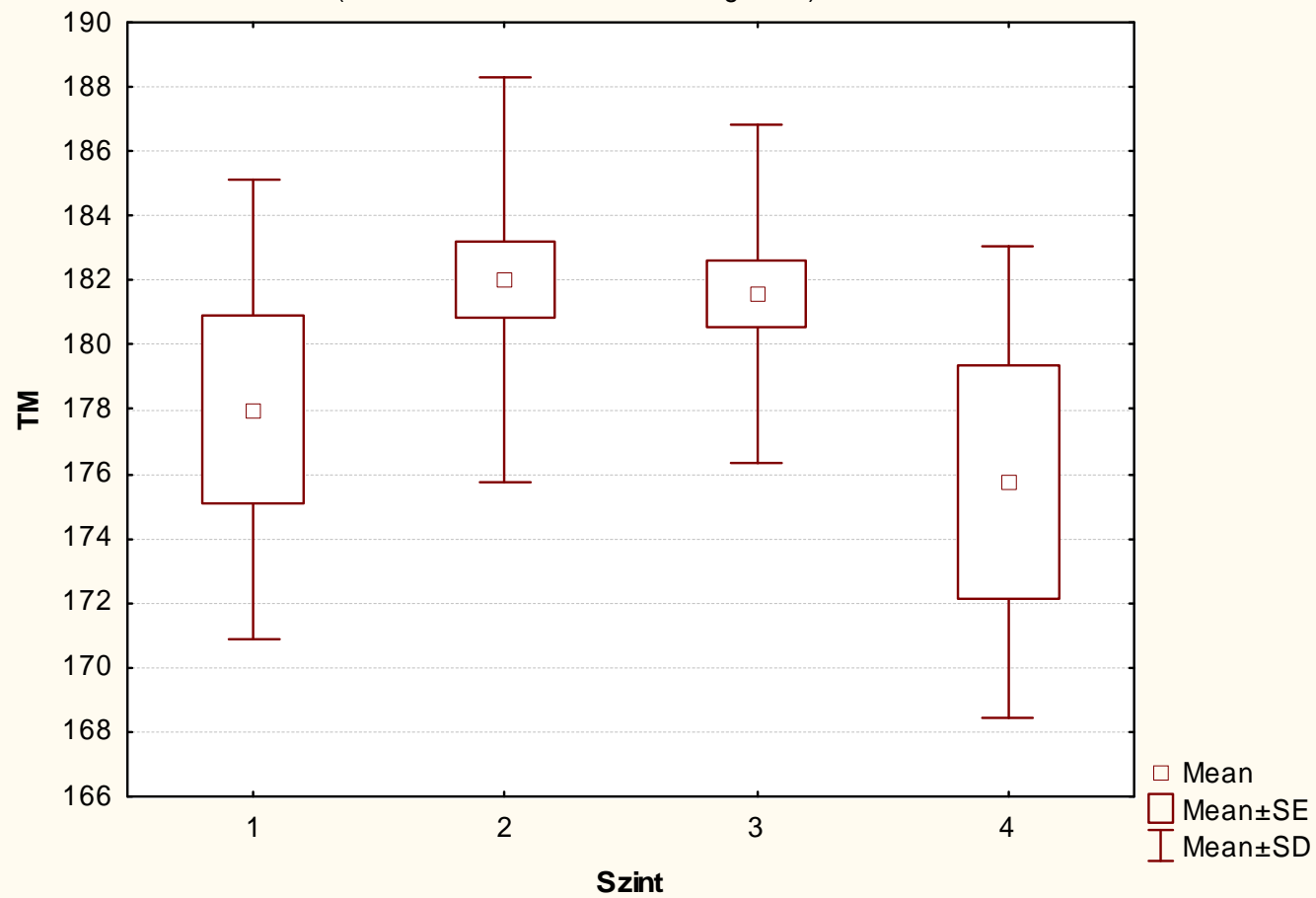
# Kétváltozós hisztogram



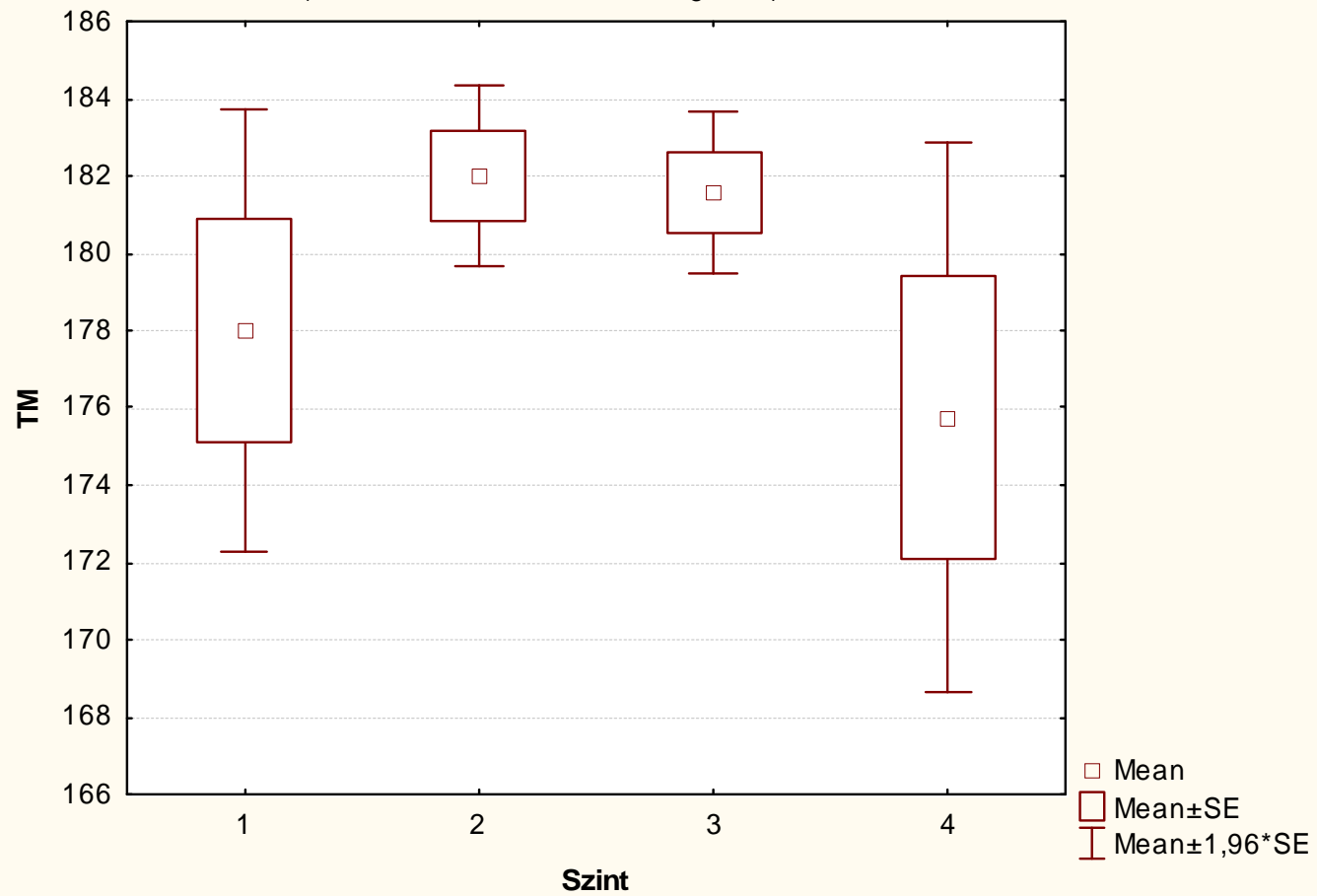
# Box plot (doboz ábra)

- Az eredmények ábrázolásakor a diagramokon az átlagot és az átlag hibáját, vagy a szórást szokták feltüntetni.
- A statisztikai programokban ezt általában külön be lehet állítani, egyes grafikon típusok pedig mindkét értéket képesek megjeleníteni.
- Az értékeket  $\pm$  értelemben értelmezzük és általában így is ábrázoljuk.

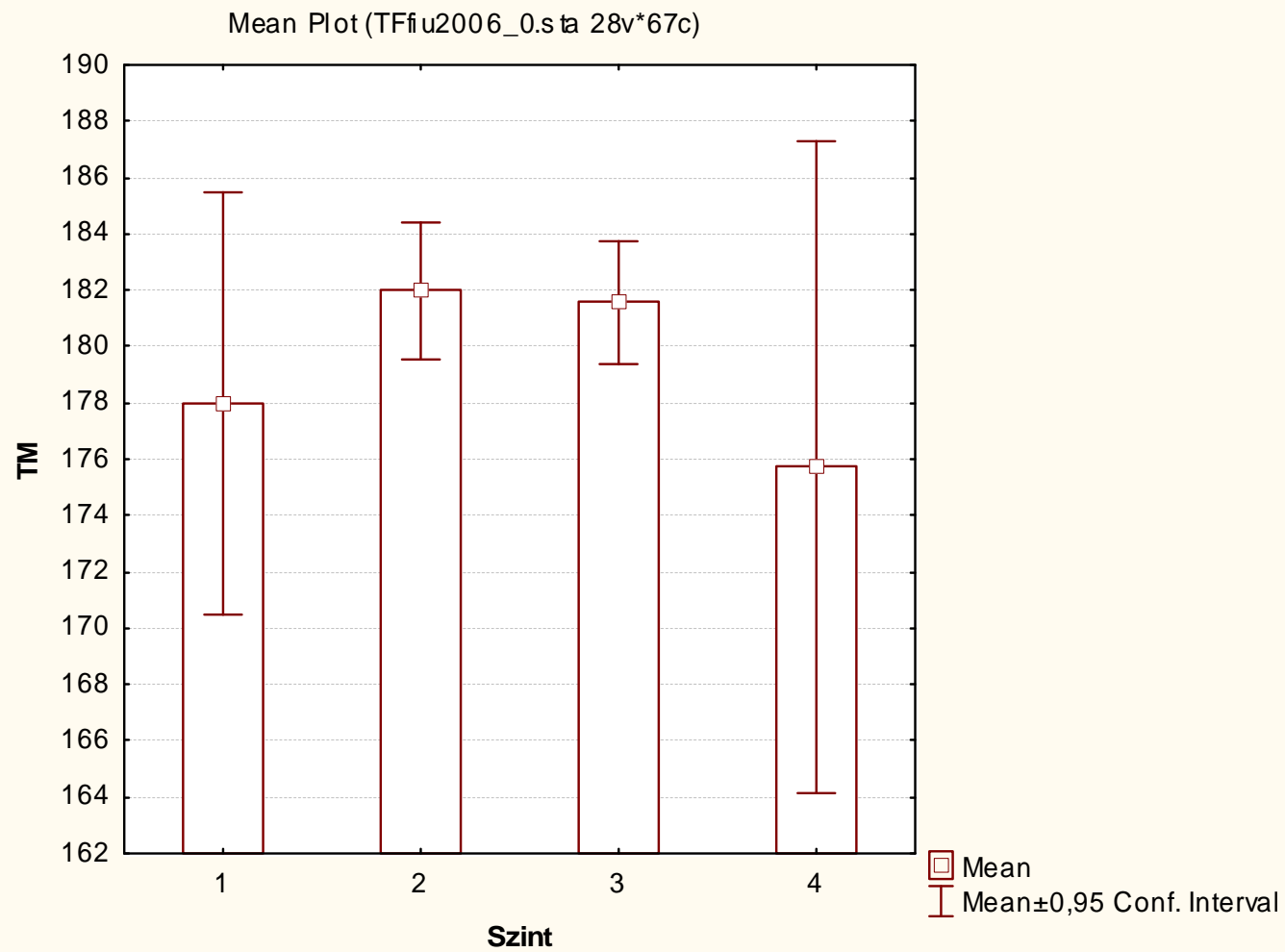
Categ. Box & Whisker Plot: **TM**  
(Casewise deletion of missing data)



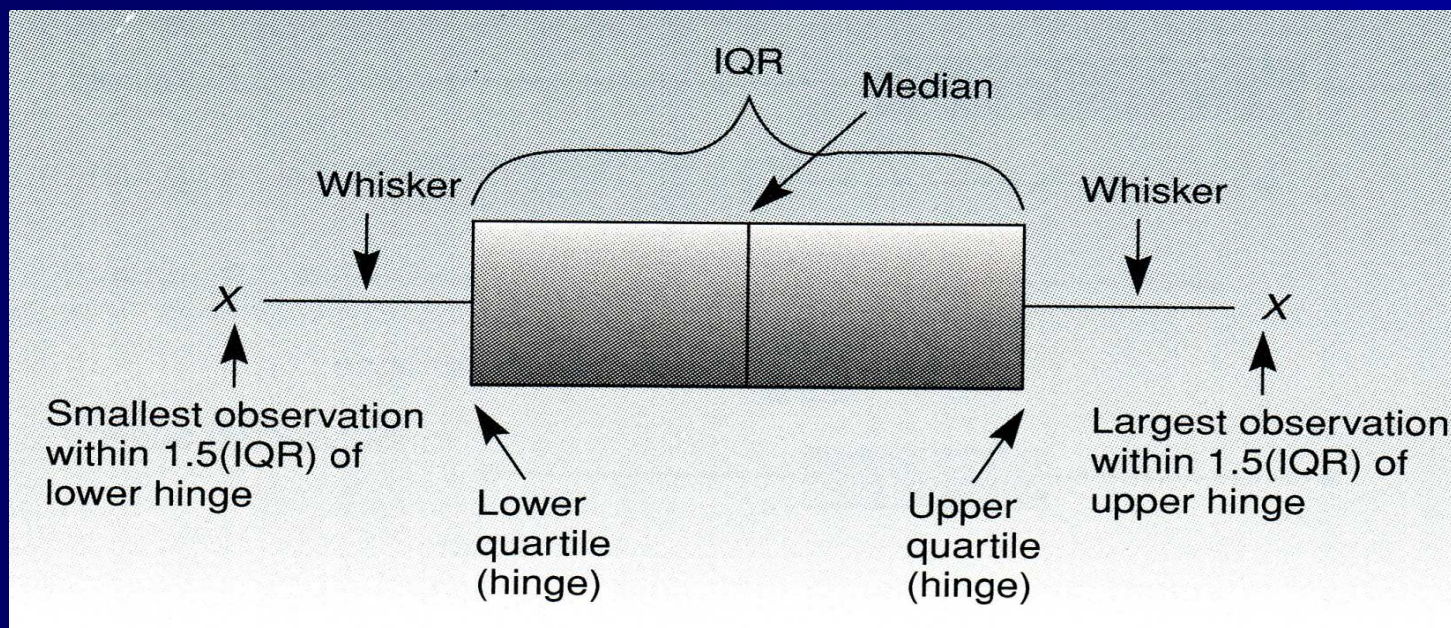
Categ. Box & Whisker Plot: **TM**  
(Casewise deletion of missing data)





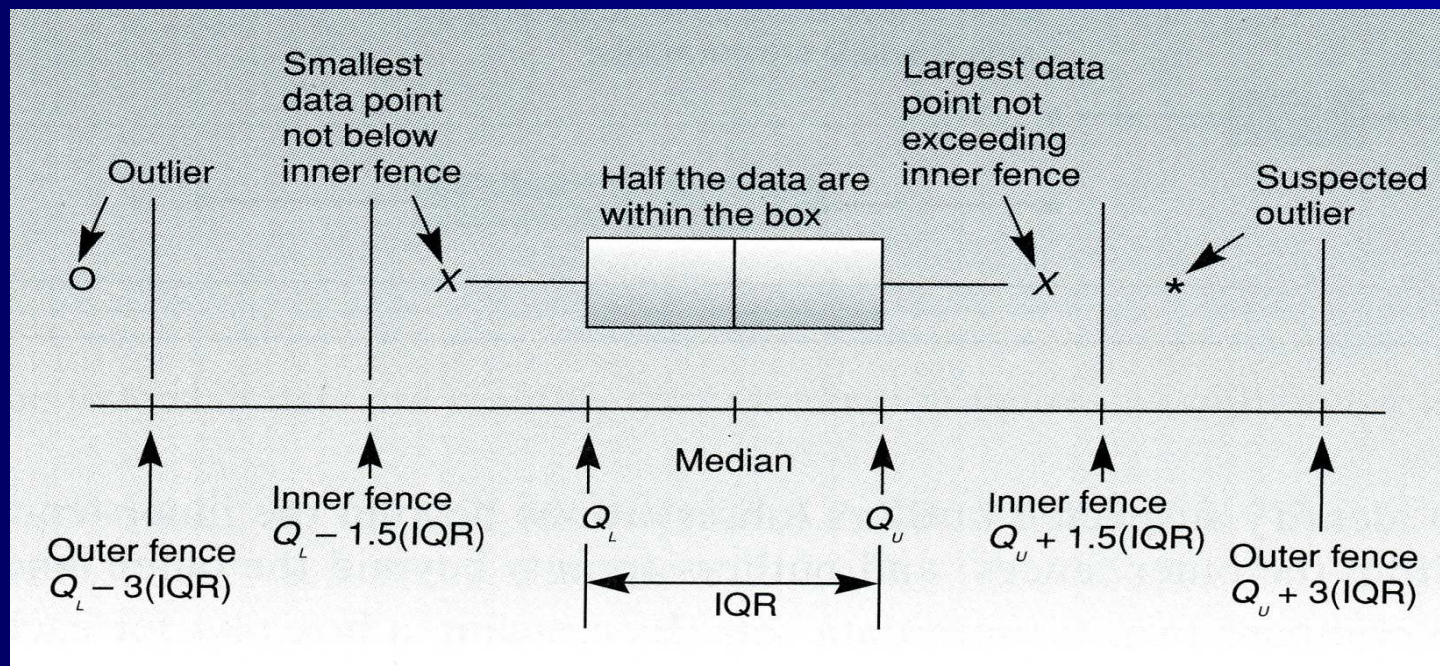


# Box plot (doboz ábra)

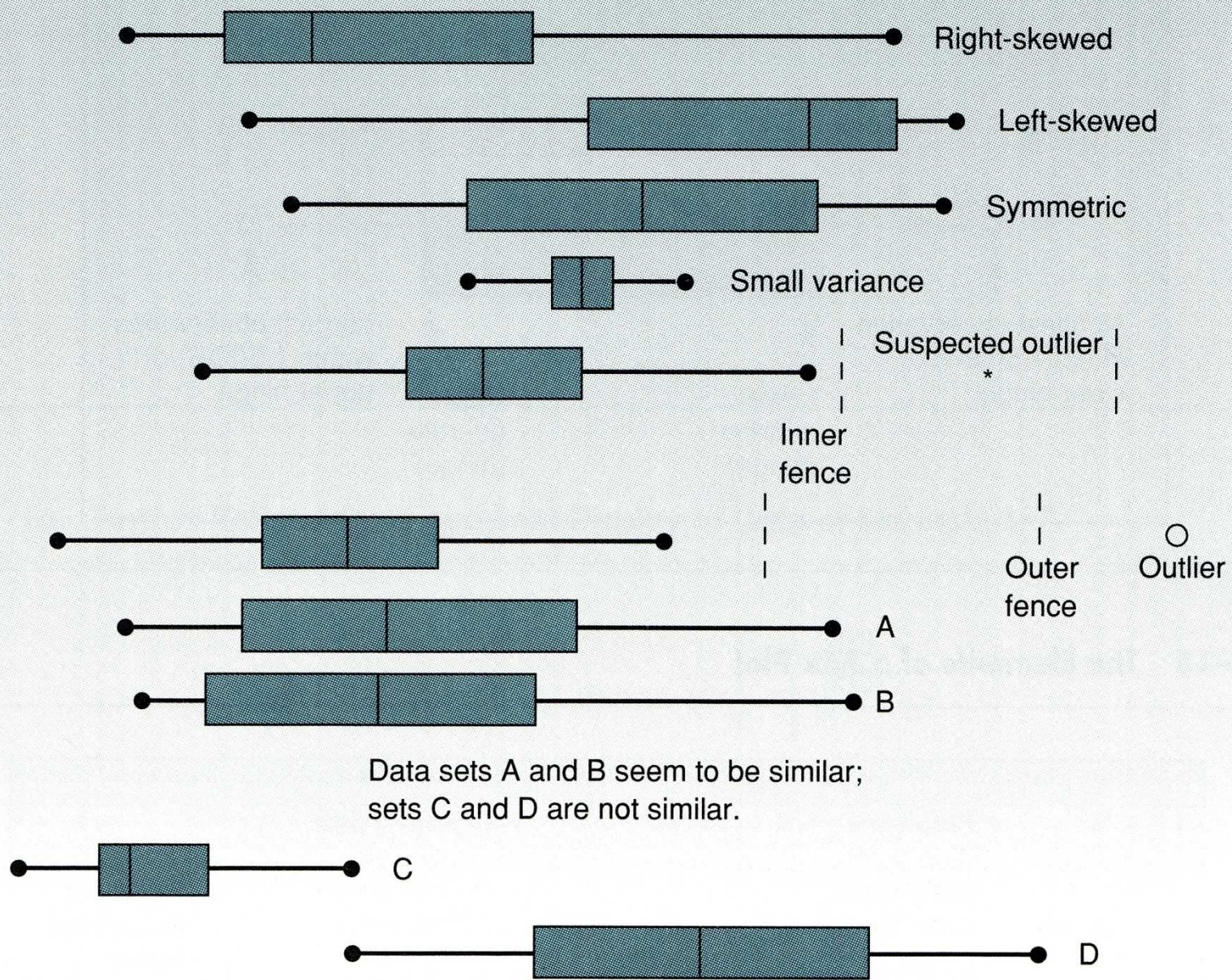




# A box plot ábra elemei

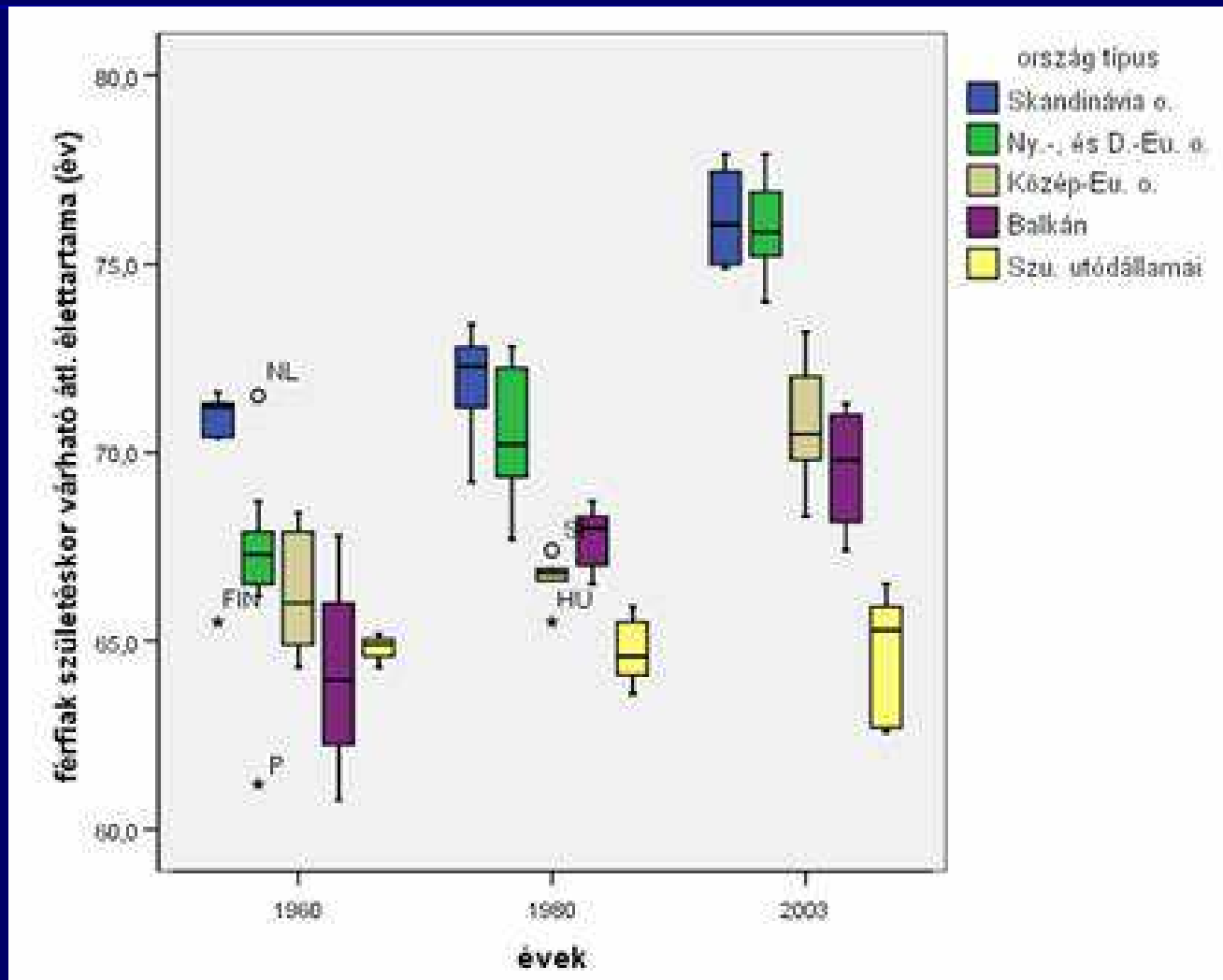






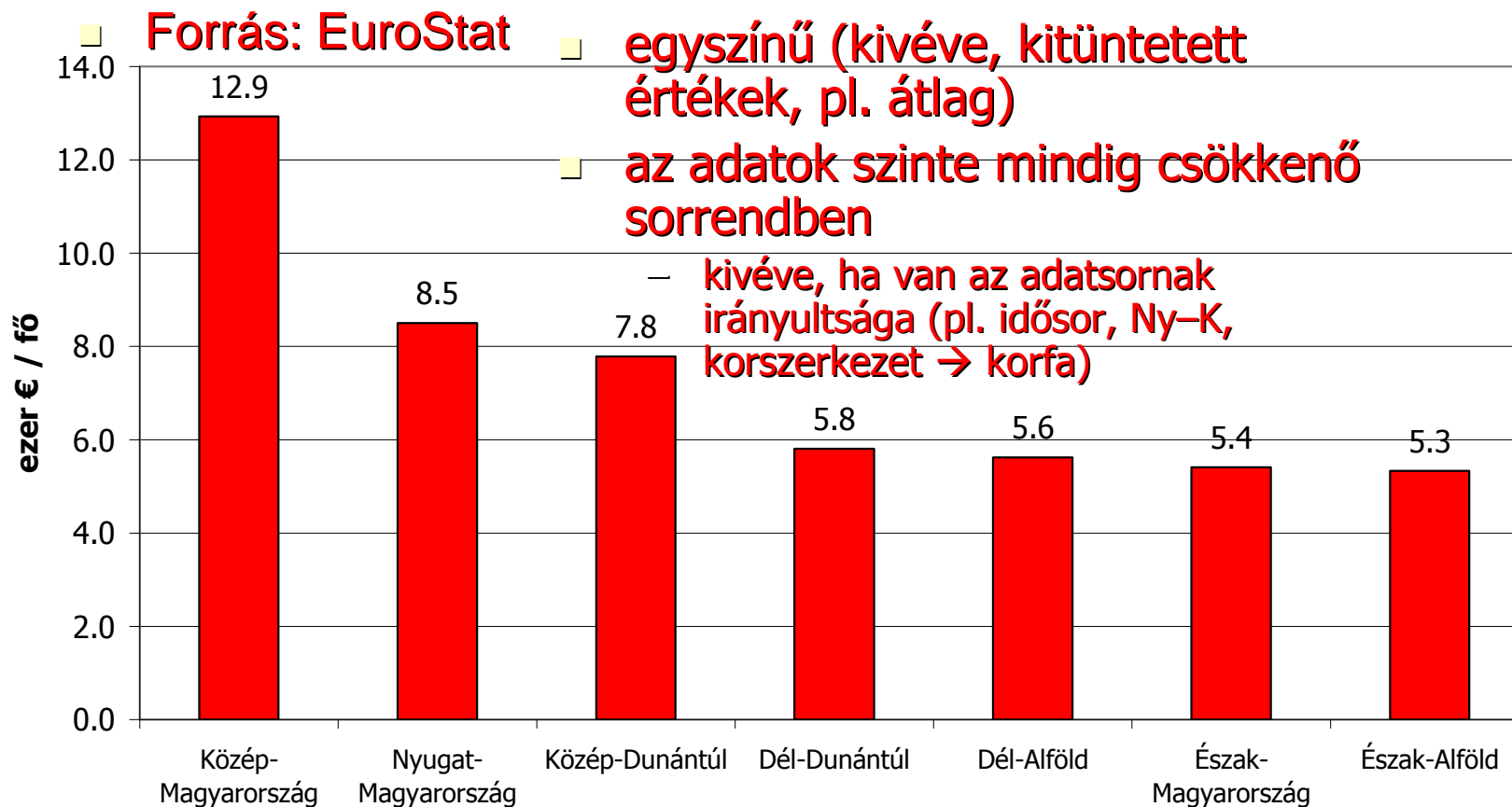


# Férfiak születéskor várható átlagos élettartamának box-plot ábrái



# Egyszerű oszlopdiagram

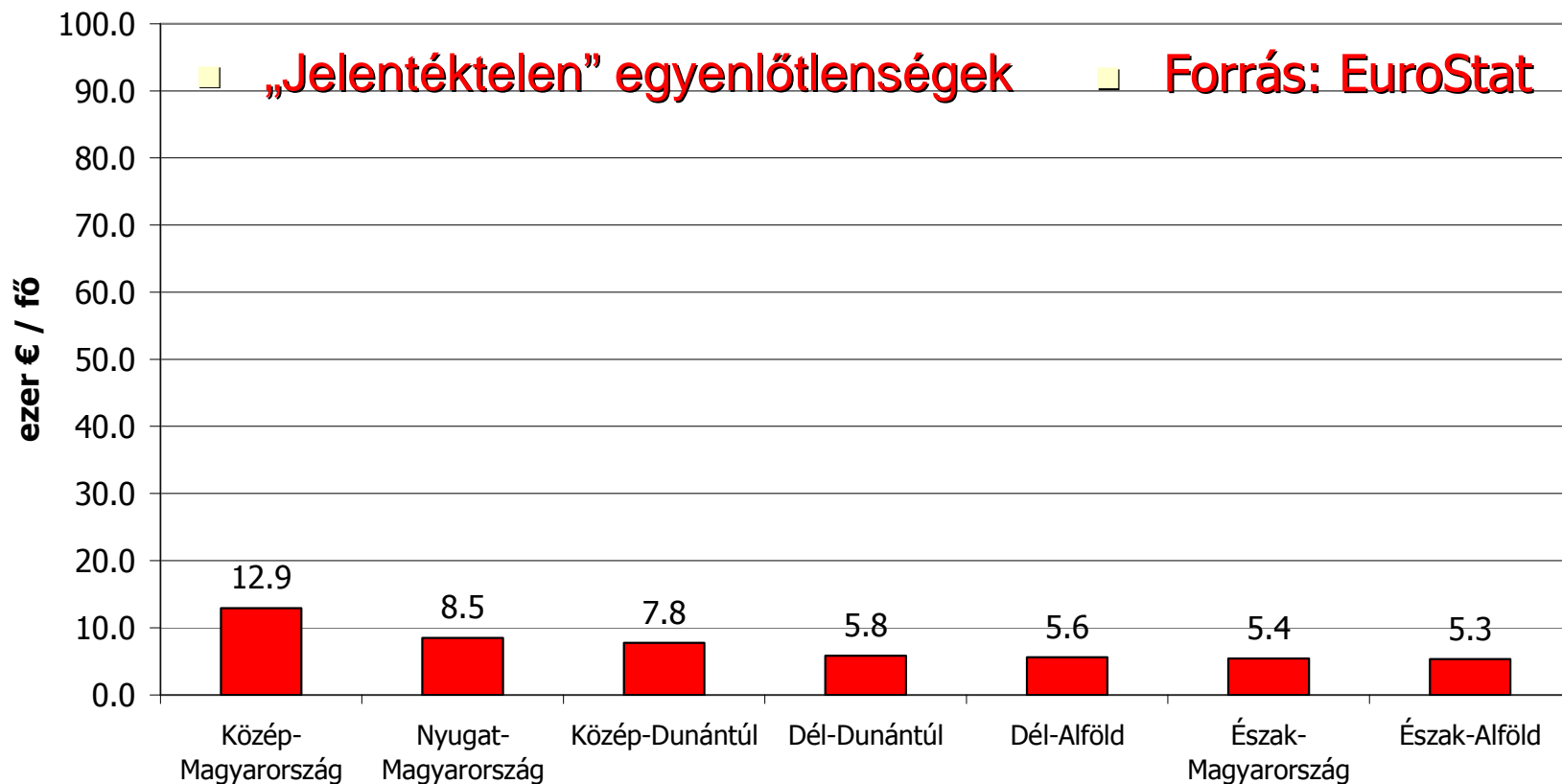
Regionális gazdasági fejlettségi különbségek Magyarországon  
2004-ben (NUTS2-es szinten, egy főre jutó GDP alapján)



# Semleges értékhatárok a kategóriatengelyen

Hibás ábra

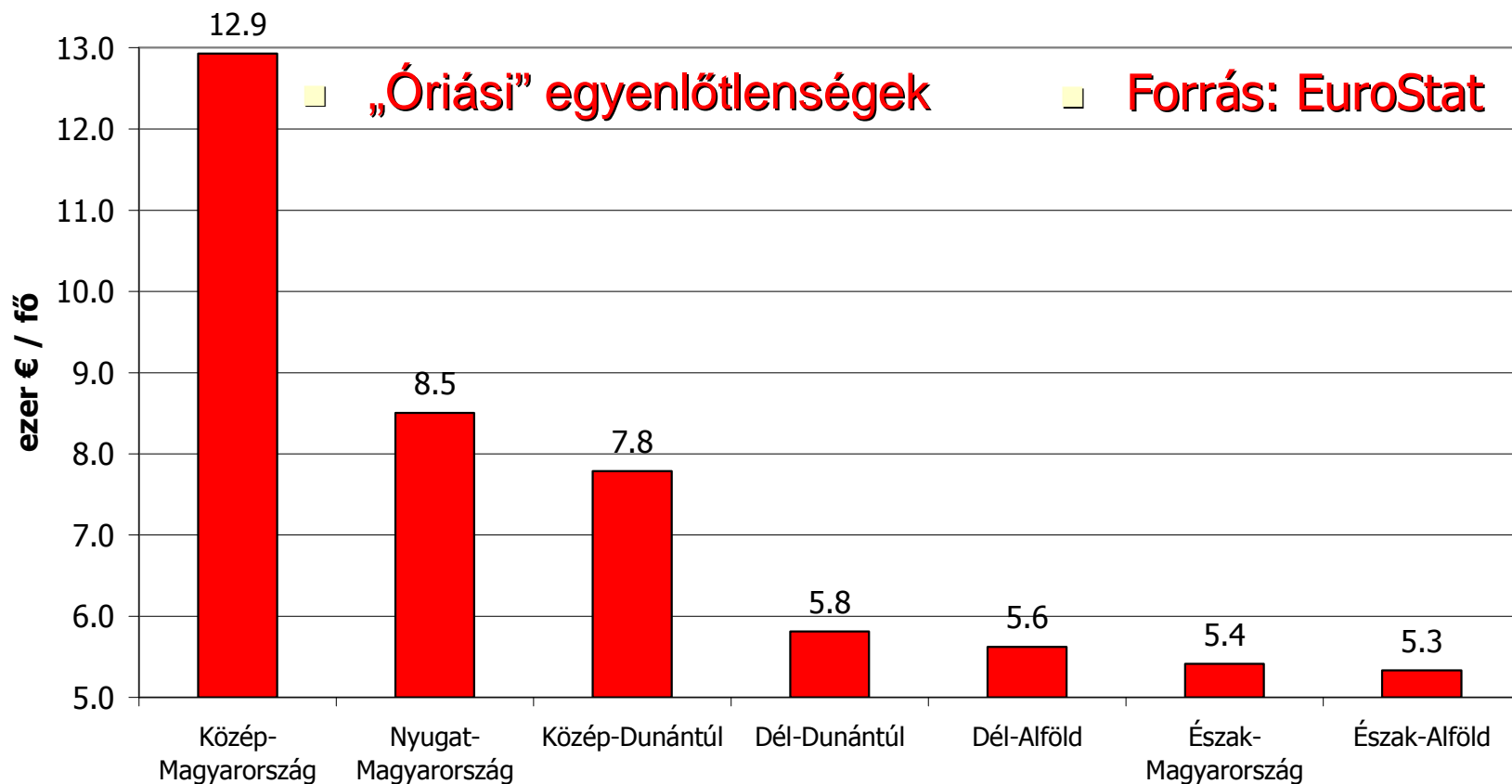
Regionális gazdasági fejlettségi különbségek Magyarországon  
2004-ben (NUTS2-es szinten, egy főre jutó GDP alapján)



# Semleges értékhatárok a kategóriatengelyen

Hibás ábra

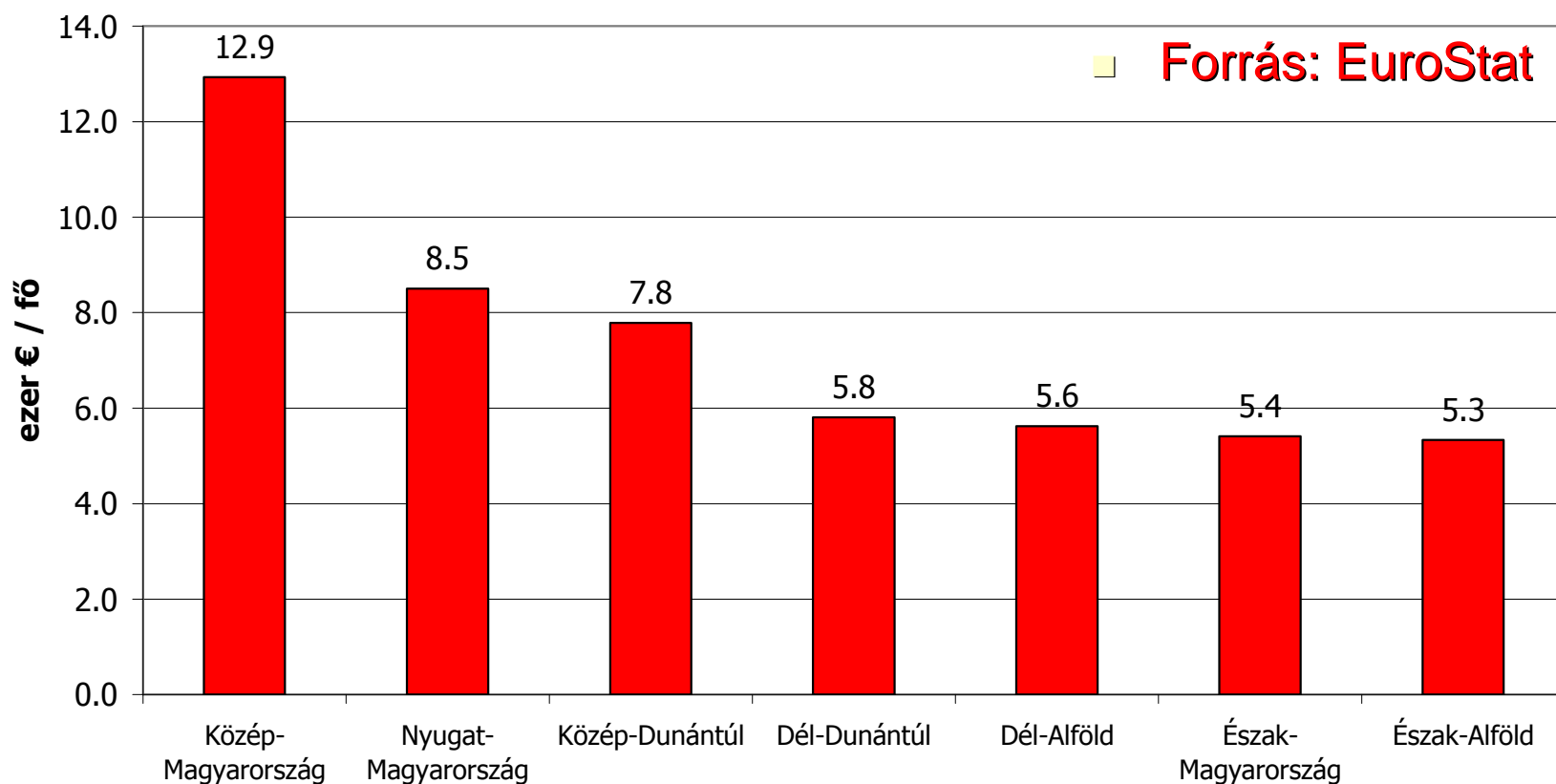
Regionális gazdasági fejlettségi különbségek Magyarországon  
2004-ben (NUTS2-es szinten, egy főre jutó GDP alapján)





# A jól elkészített diagram ismérvei

**Regionális gazdasági fejlettségi különbségek Magyarországon  
2004-ben (NUTS2-es szinten, egy főre jutó GDP alapján)**



# Hibás csoportosított oszlopdiagram

Hibás ábra

A regionális gazdasági fejlettségi különbségek alakulása  
Magyarországon 1995–2004 között (megyei szinten, egy főre  
jutó GDP alapján)

Többszínű

Értékeket hasonlít össze kategóriák mentén

Alkalmas pl. területi egyenlőtlenségek időbeli változásának  
a vizsgálatára

Fontos: az összehasonlíthatóság érdekében mindkét időpontban %-os  
értékek szerepeljenek

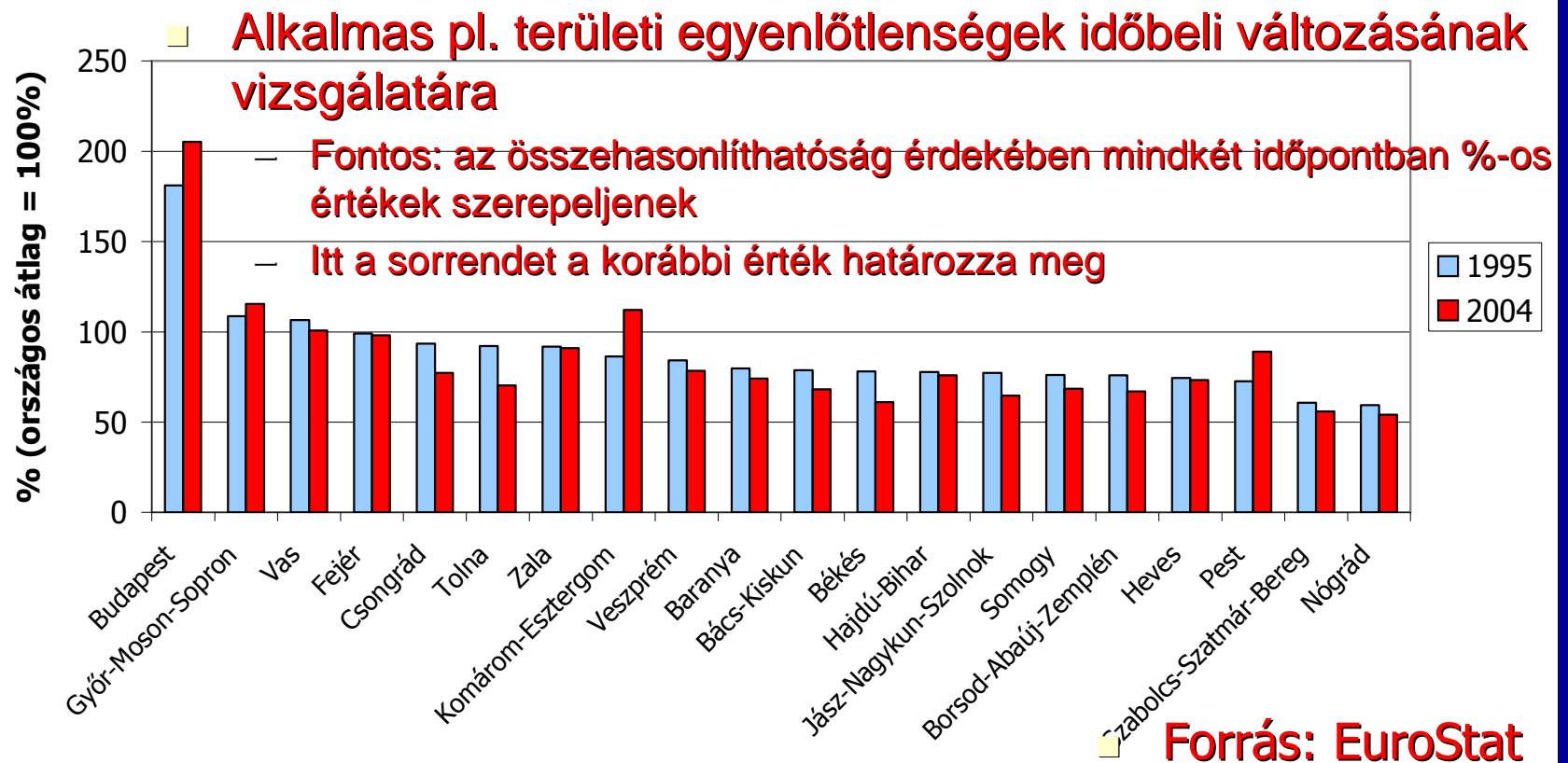
A sorrendet a korábbi érték határozza meg



Forrás: EuroStat

# Jó csoportosított oszlopdiagram (divergencia Magyarországon)

A regionális gazdasági fejlettségi különbségek alakulása  
Magyarországon 1995–2004 között (megyei szinten, egy főre  
jutó GDP alapján)



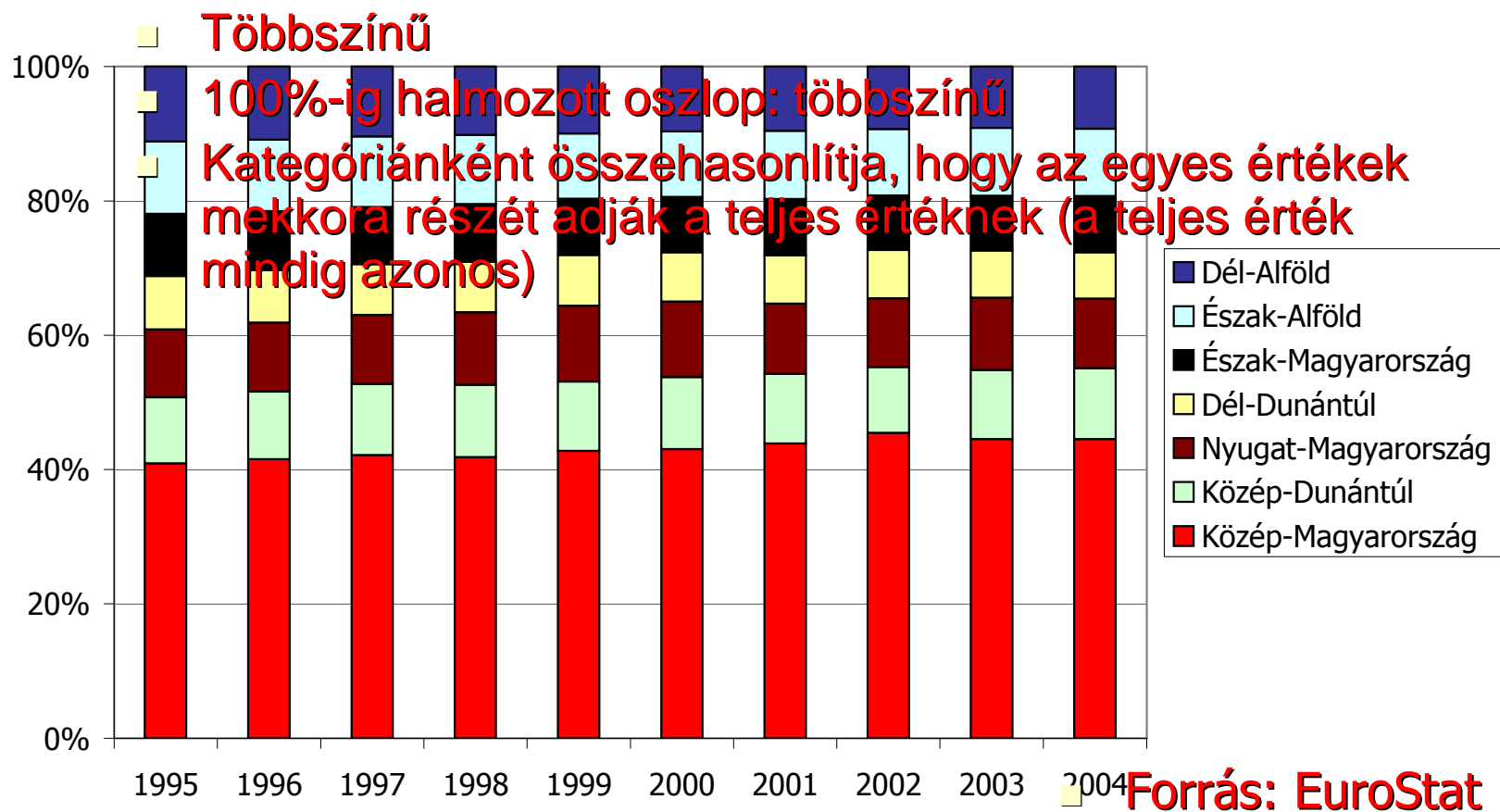
# Halmazozott oszlopdiagram: csak abszolút mutatóknál!

A GDP régiók közötti megoszlásának alakulása  
Magyarországon 1995–2004 között (NUTS2-es szinten)



# 100%-ig halmozott oszlopdiagram: csak abszolút mutatóknál!

A GDP régiók közötti megoszlásának alakulása  
Magyarországon 1995–2004 között (NUTS2-es szinten)



# Jó színezés legyen (hasonló színű szomszédos egységek fekete-fehérben ne mosódjanak egybe)

Hibás ábra

A GDP régiók közötti megoszlásának alakulása Magyarországon  
1995–2004 között (NUTS2-es szinten)

Forrás: EuroStat

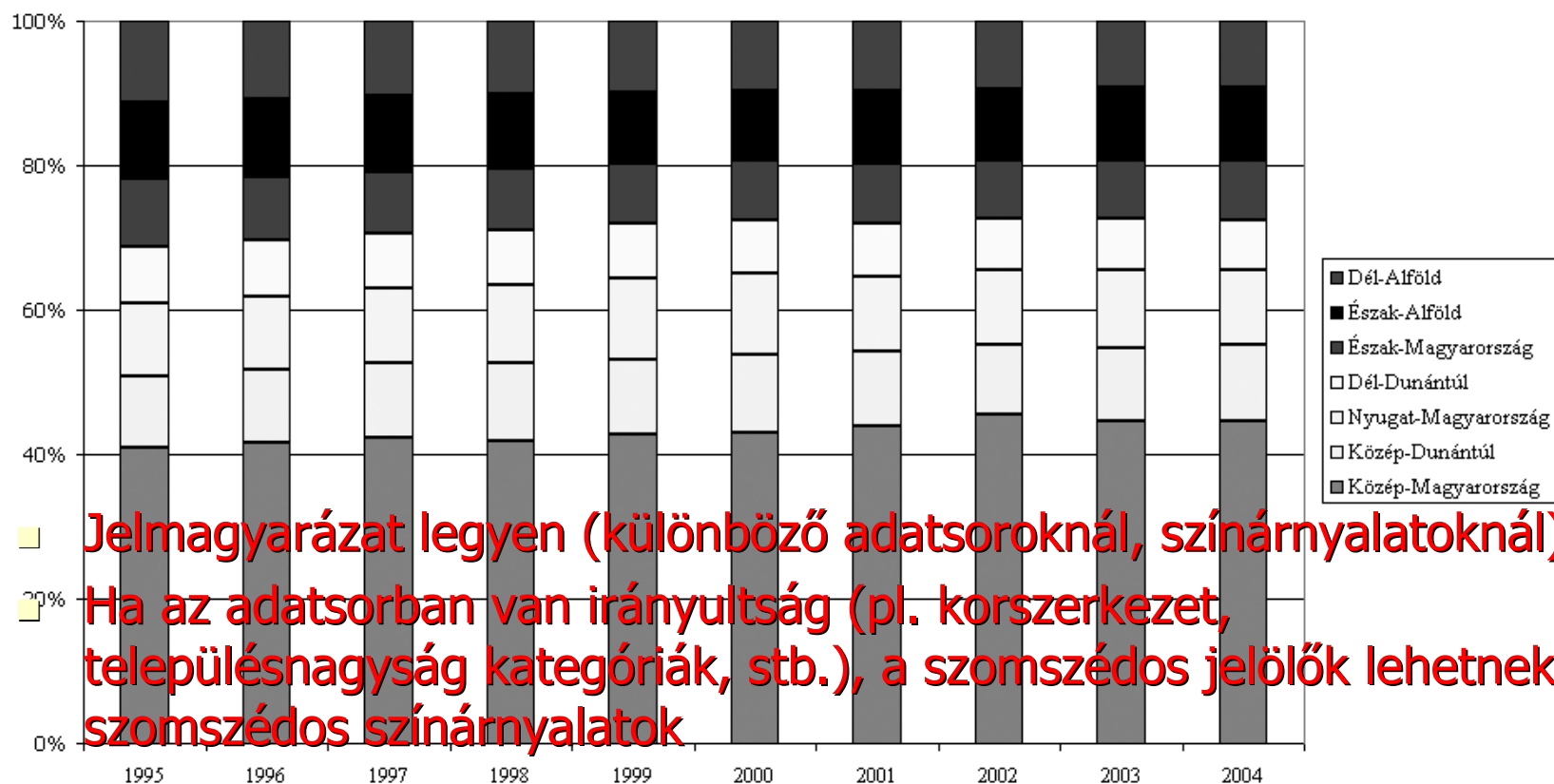


# Jó színezés legyen (hasonló színű szomszédos egységek fekete-fehérben ne mosódjanak egybe)

Hibás ábra

A GDP régiók közötti megoszlásának alakulása Magyarországon  
1995–2004 között (NUTS2-es szinten)

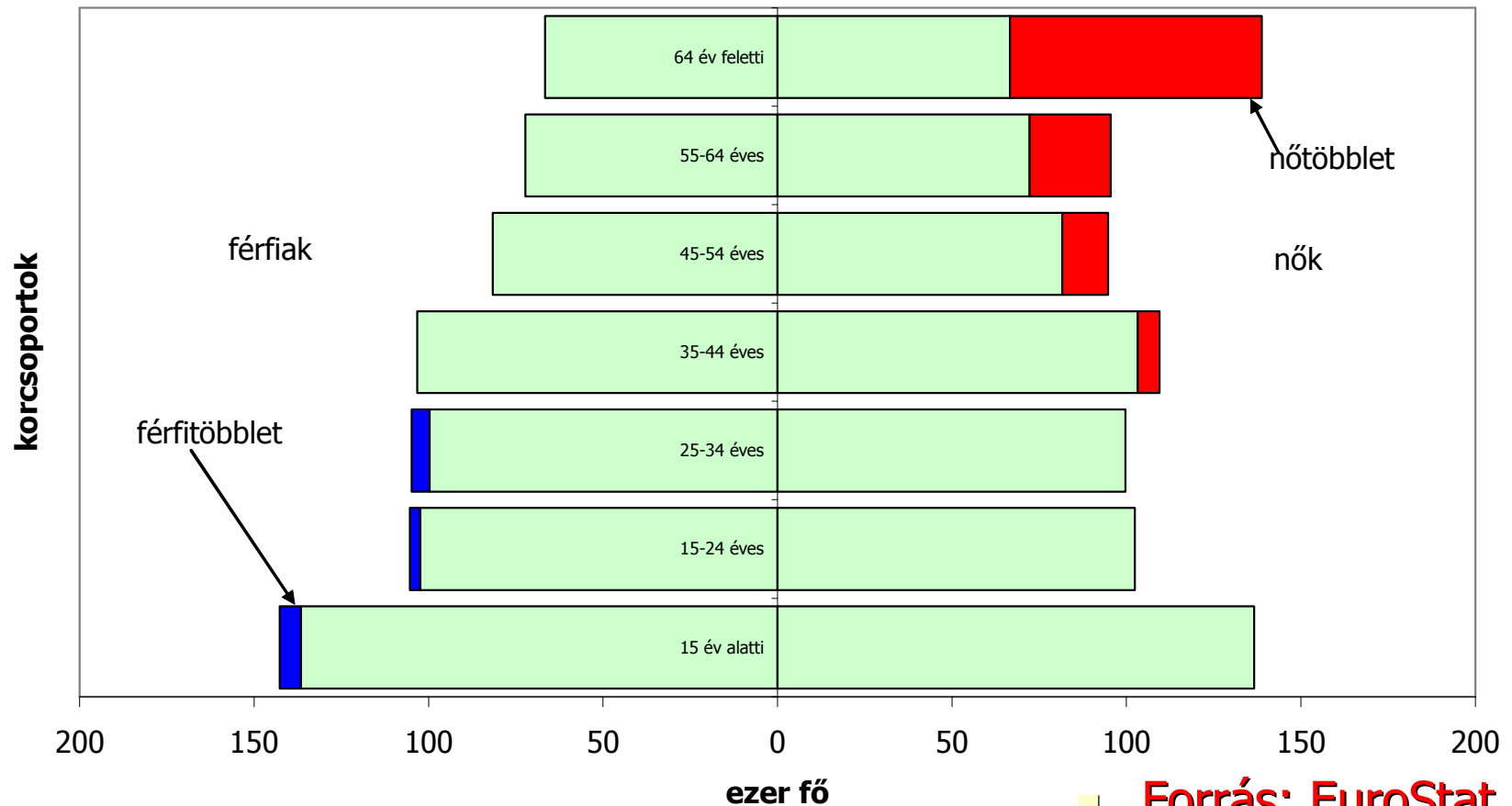
Forrás: EuroStat



Jelmagyarázat legyen (különböző adatsoroknál, színárnyalatoknál)  
Ha az adatsorban van irányultság (pl. korszerkezet, településnagyság kategóriák, stb.), a szomszédos jelölők lehetnek szomszédos színárnyalatok

# Halmazott sávdigram; speciális fajtája: a korfa

Magyarország korfája, 1998

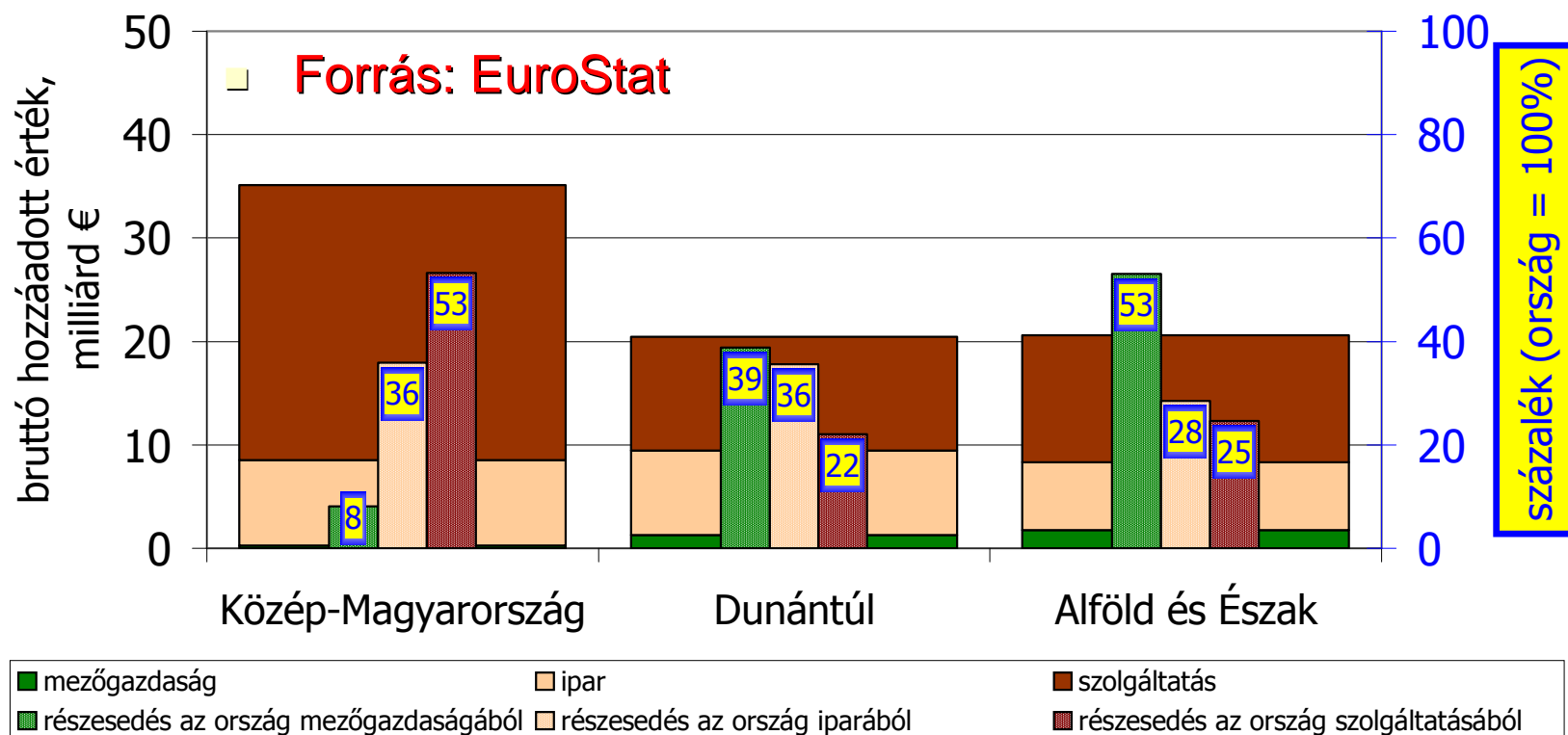




# Célorientált, áttekinthető legyen, ne túlságosan összetett (inkább külön diagramokon)!

**Hibás ábra**

A magyar régiók gazdasága ágazatonként, illetve a régiók aránya az ágazatokból (NUTS1-es szinten, bruttó hozzáadott érték alapján), 2004-ben



# Egyszerű legyen, ne túldizájnjolt!

## Hibás ábra

Regionális gazdasági fejlettségi különbségek Magyarországon  
2004-ben (NUTS2-es szinten, egy főre jutó GDP alapján)

- **Kerüljük a színes vagy mintás háttér**

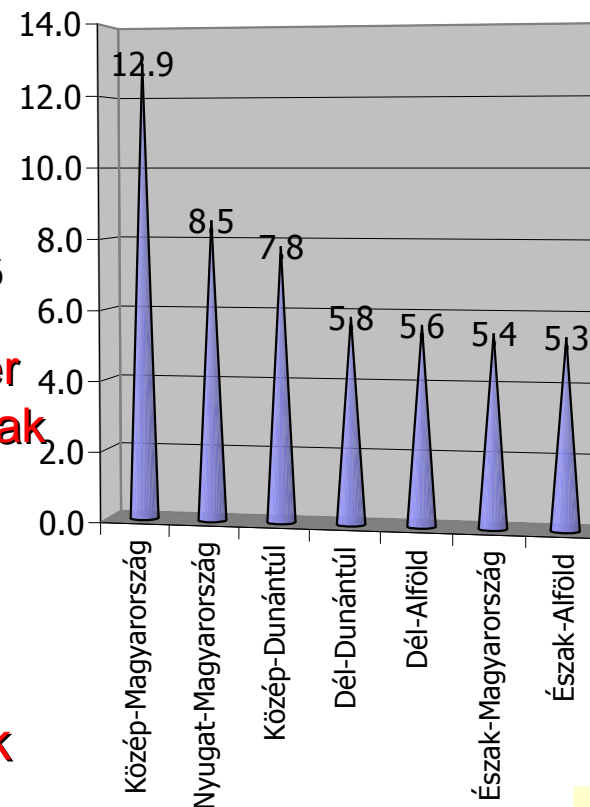
- Elegánsabb a fehér háttér

- Régi Excel: ezer € / fő  
szereplő szürke háttér előtt kevésbé látszanak a szürkés-kék jelölők

- Nyomtatásnál felesleges „festékpazarlás”

- **Kúpok helyett oszlopok**

- **3D helyett 2D**



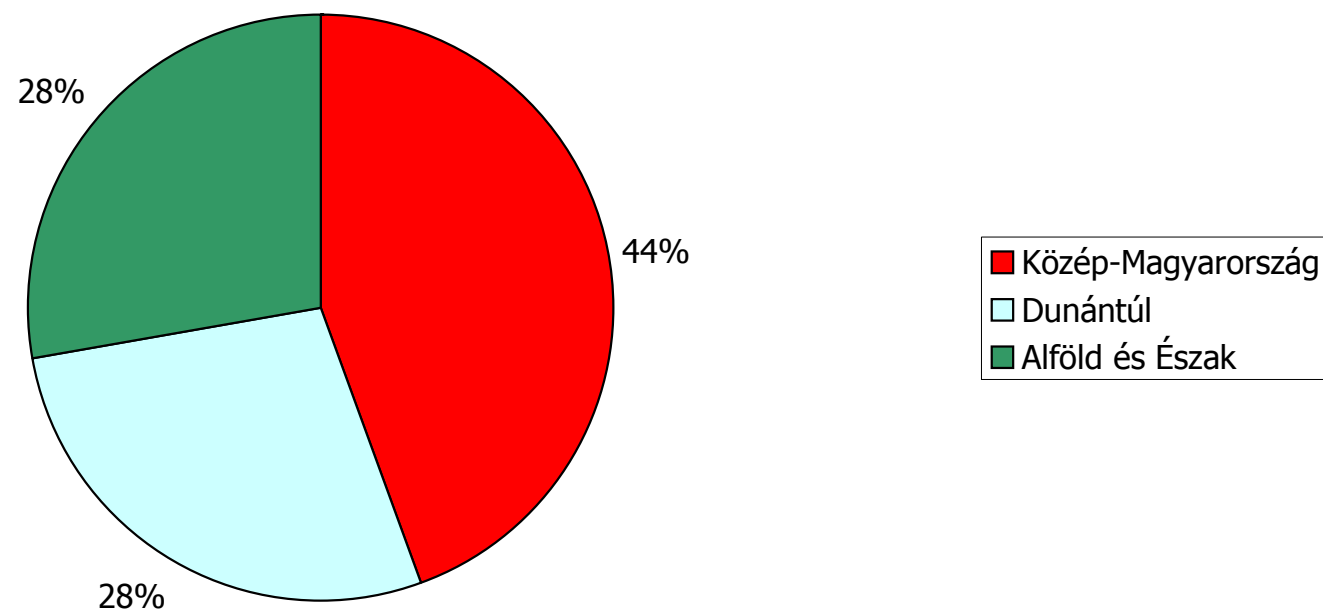
■ **Forrás: EuroStat**

# Kördiagram: csak abszolút mutatóknál!

- Nem ajánlott (helyette inkább oszlopdiagram);
- Nehezen mérhető az összetétel változása (perek diagram);
- 3-4 körcikknél nem lehet több (össze kell vonni a kisebb értékeket → egyéb kategória);
- Végképp rossz:
  - 3D, robbantott kör;

# Jobb kördiagram

**A GDP megoszlása Magyarország régiói között 2004-ben  
(NUTS1-es szinten)**

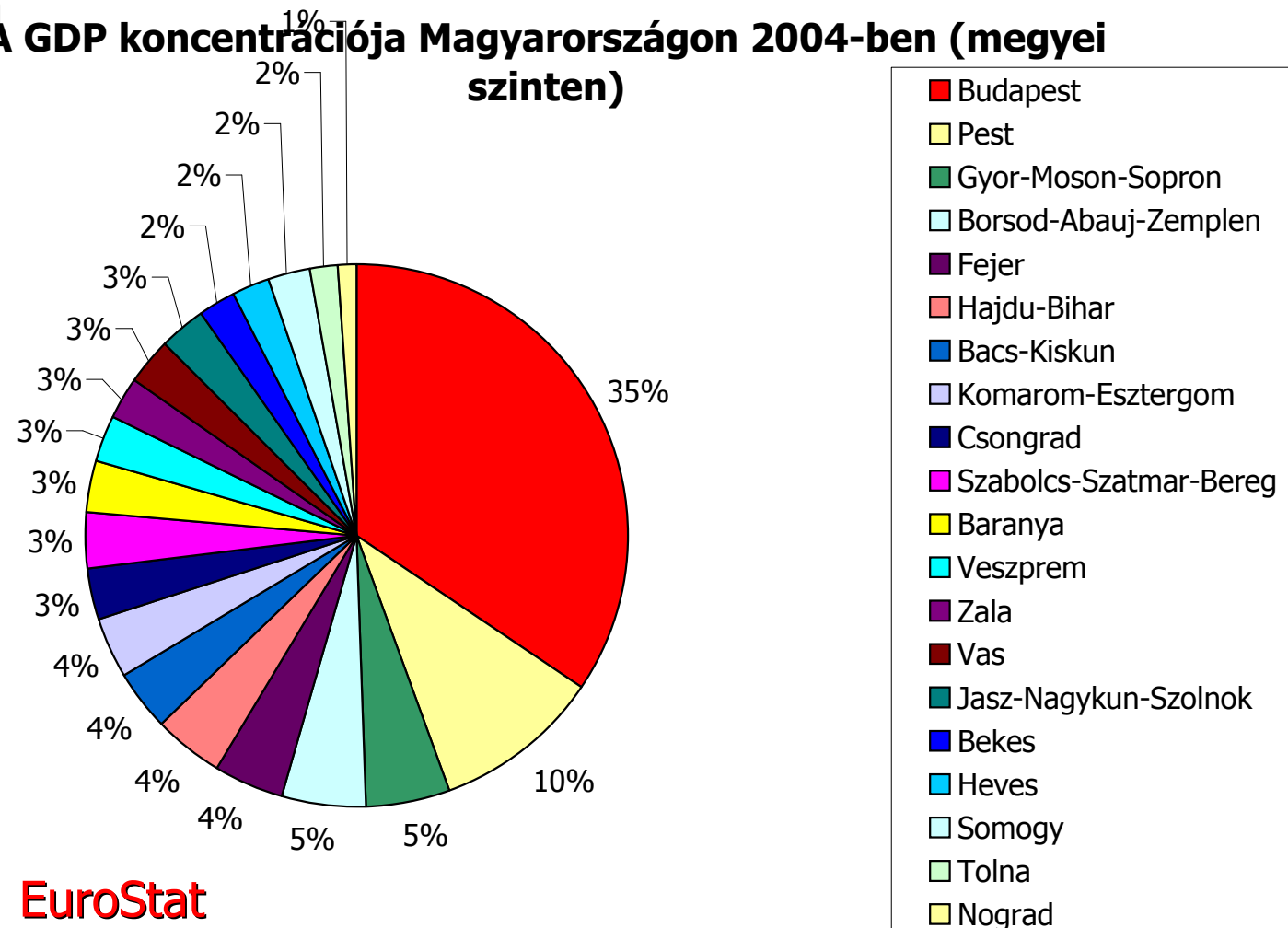


■ **Forrás: EuroStat**

# Rossz kördiagram (ne legyen 3-4-nél több körcikk)

Hibás ábra

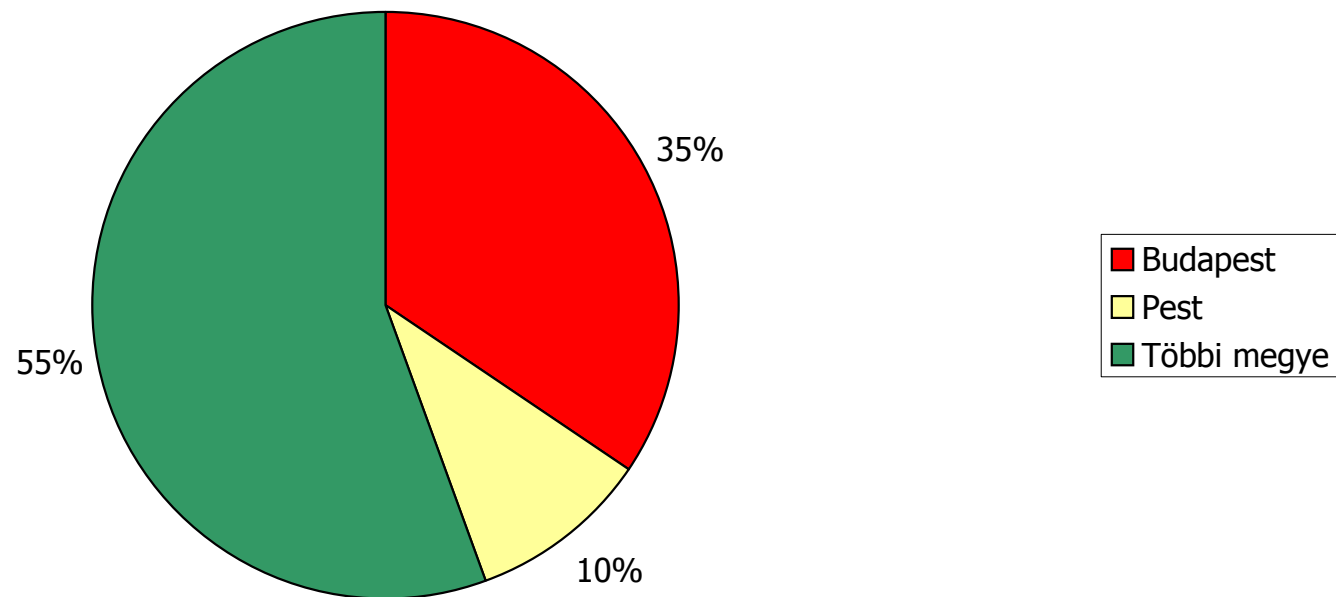
A GDP koncentrációja Magyarországon 2004-ben (megyei szinten)



Forrás: EuroStat

# Jobb kördiagram: nagyobb elemszámnál aggregálni kell

A GDP koncentrációja Magyarországon 2004-ben (megyei szinten)



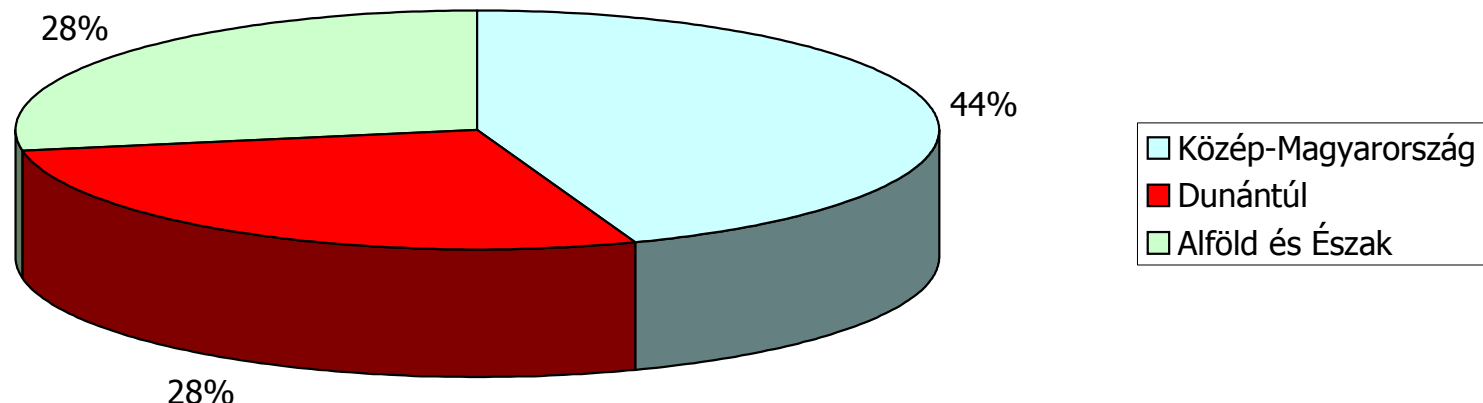
■ Forrás: EuroStat

# Optikailag semleges legyen, kerüljük a térhatást (3D-t)!

Hibás ábra

A GDP megoszlása Magyarország régiói között 2004-ben  
(NUTS1-es szinten)

■ Forrás: EuroStat

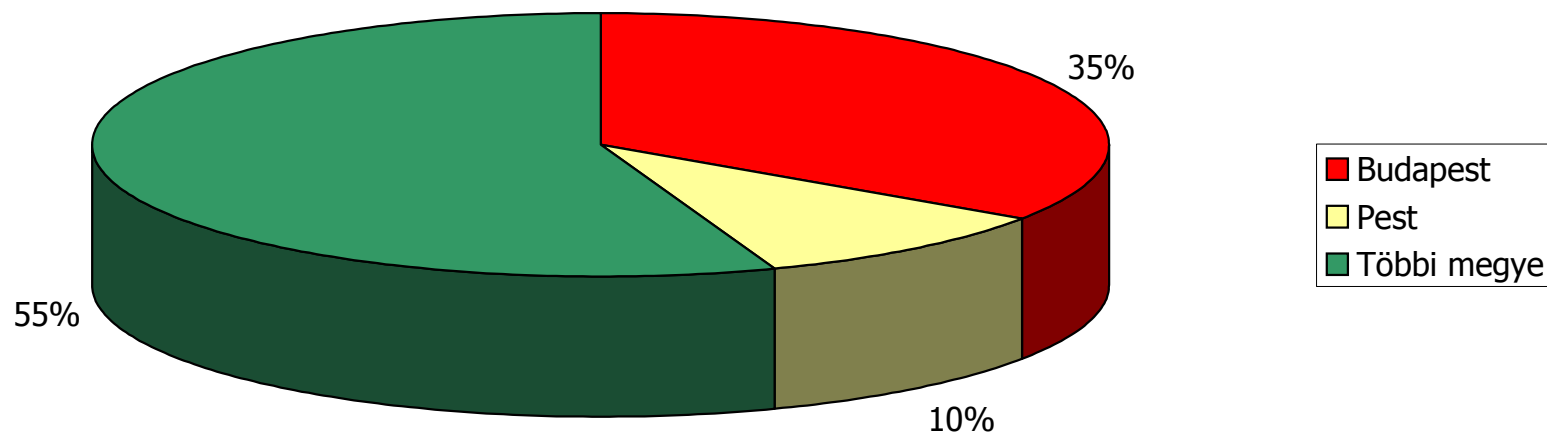


- A térhatású kördiagram nem jó;
  - Az előtérben lévő körcikkek nagyobbak látszanak;
  - Térhatás, komolytalan dizájnolás (oszlopdiagramnál is);

# Rossz kördiagram (ne legyen térhatású)

Hibás ábra

A GDP koncentrációja Magyarországon 2004-ben (megyei szinten)



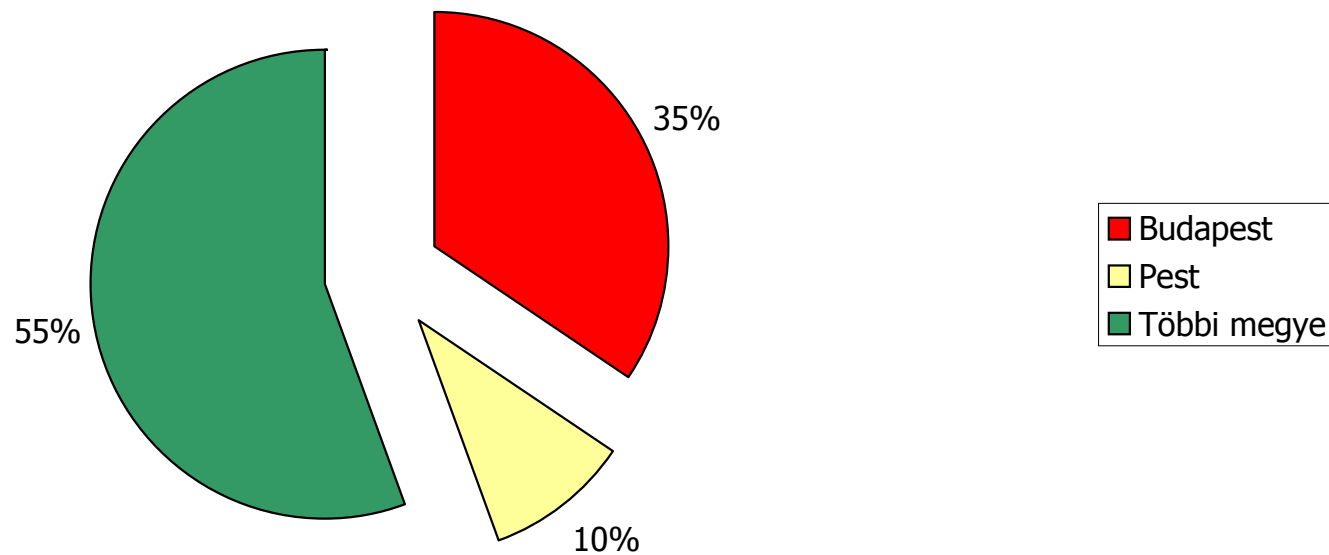
Forrás: EuroStat



# Rossz kördiagram (ne legyen robbantott)

Hibás ábra

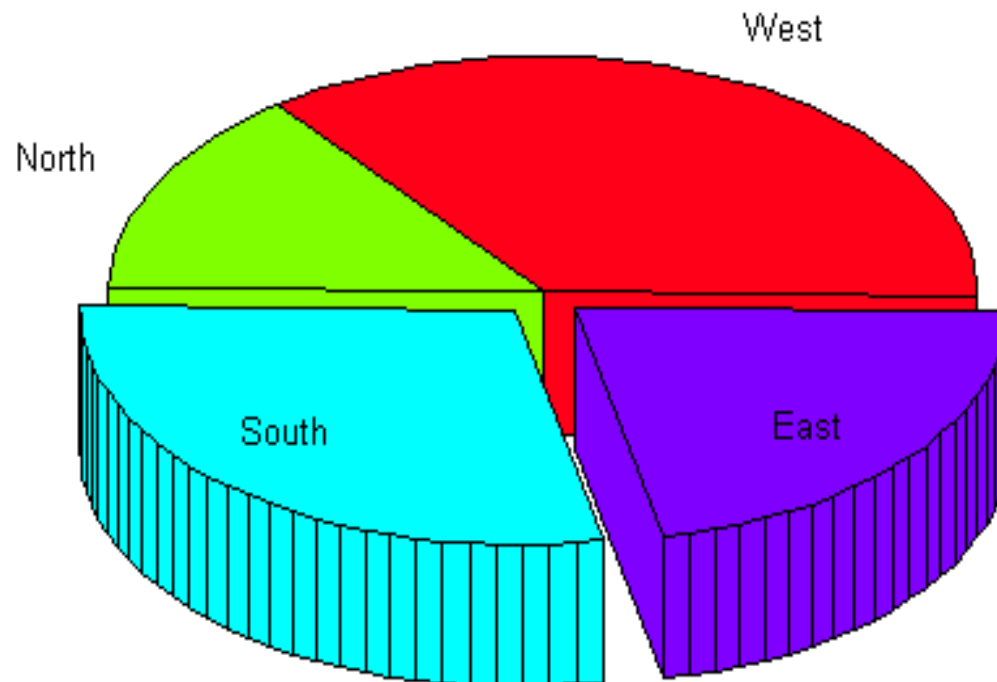
A GDP koncentrációja Magyarországon 2004-ben (megyei szinten)



Forrás: EuroStat

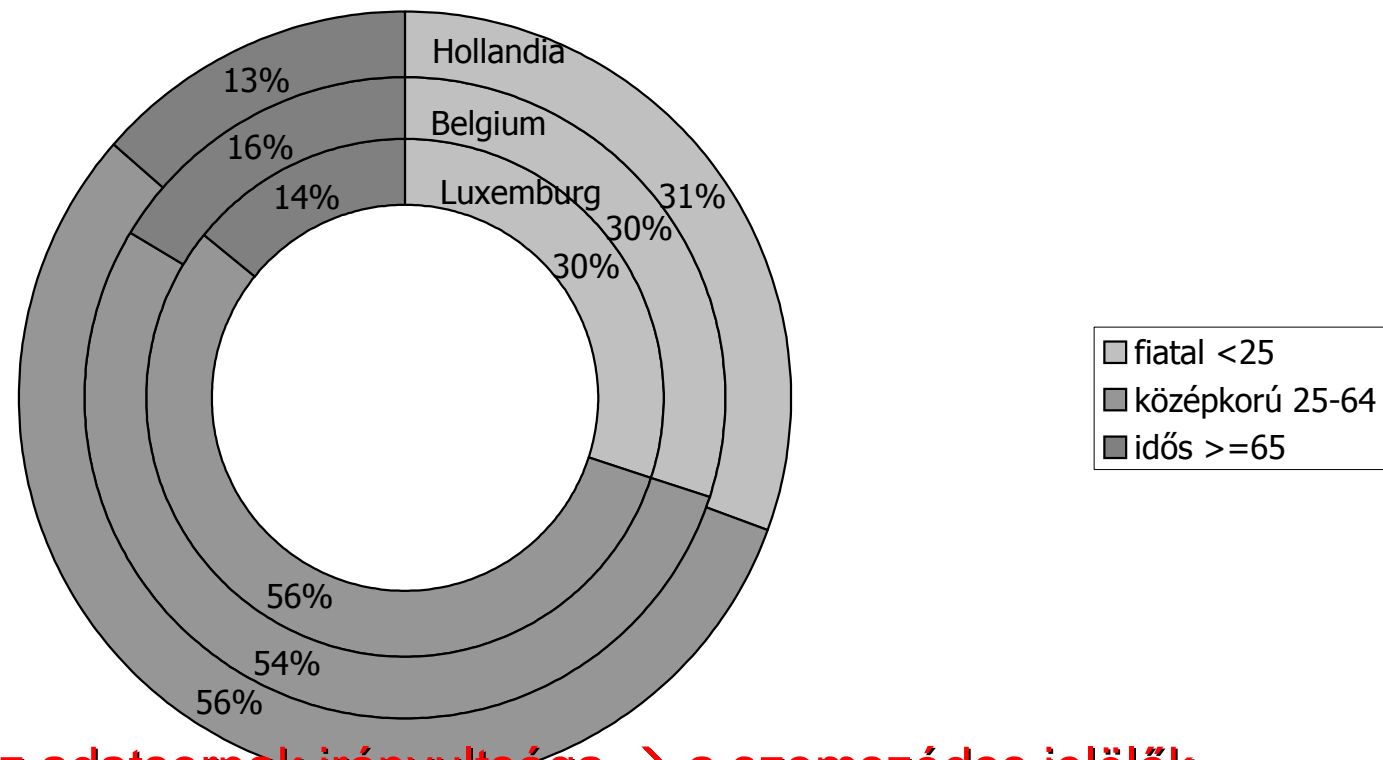
# Rossz kördiagram (az adott ismértvhez nem a megfelelő ábrázolási módot választottuk)

Frequency of wind direction at measurement site



# Jobb kördiagram: színezés

A korszerkezet különbségei a Benelux államokban (1998)

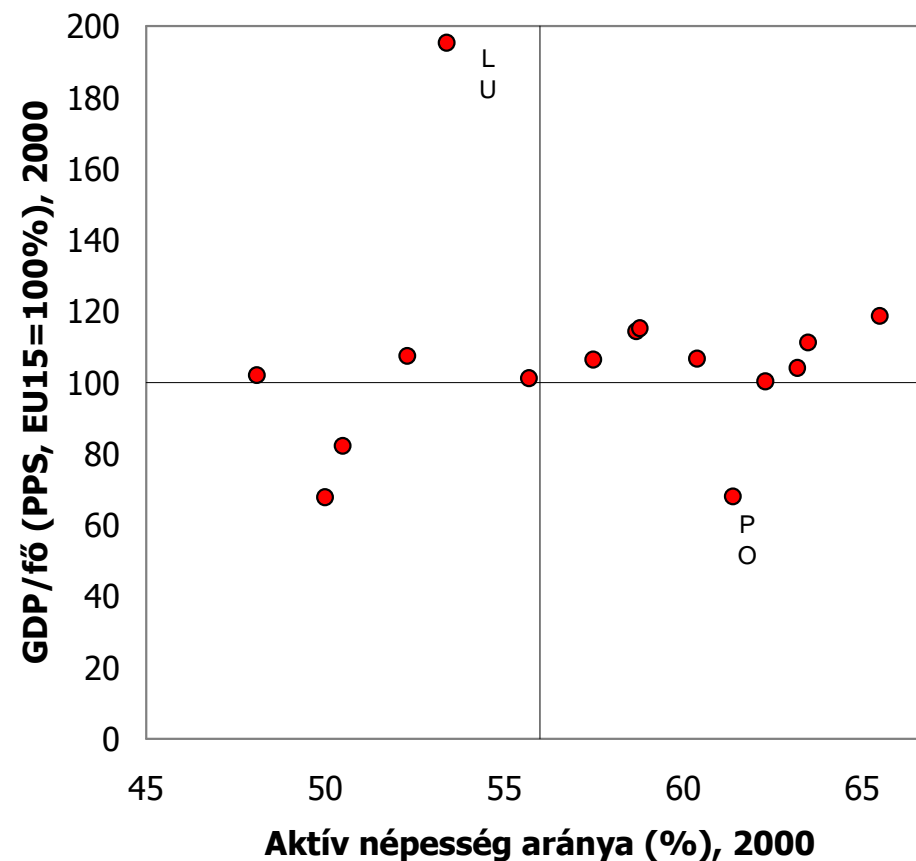


■ Ha van az adatsornak irányultsága → a szomszédos jelölők szomszédos színárnyalatok legyenek

■ Forrás: EuroStat

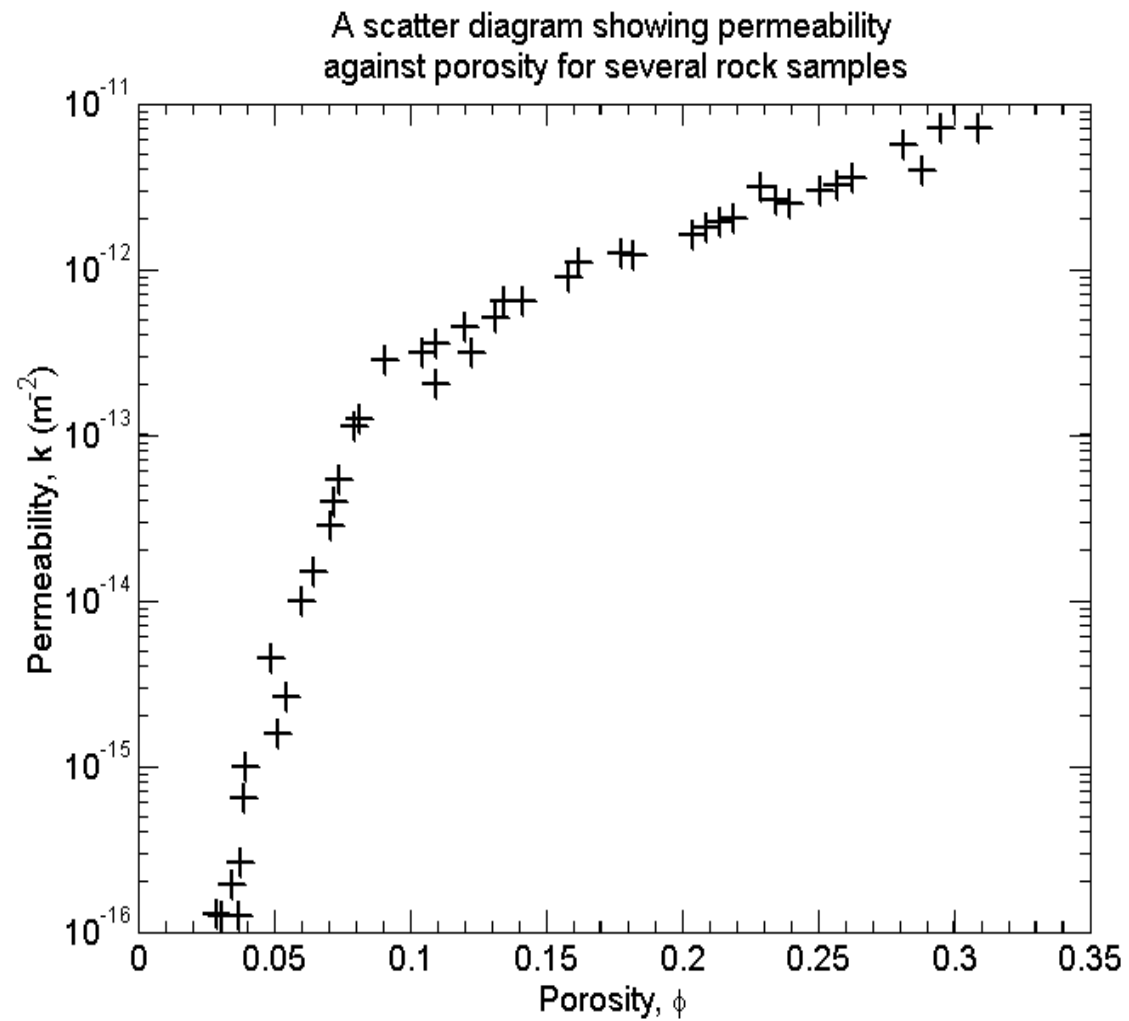
# Pontdiagram: kétdimenziós összehasonlítás

Az EU tagállamainak csoportosítása két jelzőszám szerint

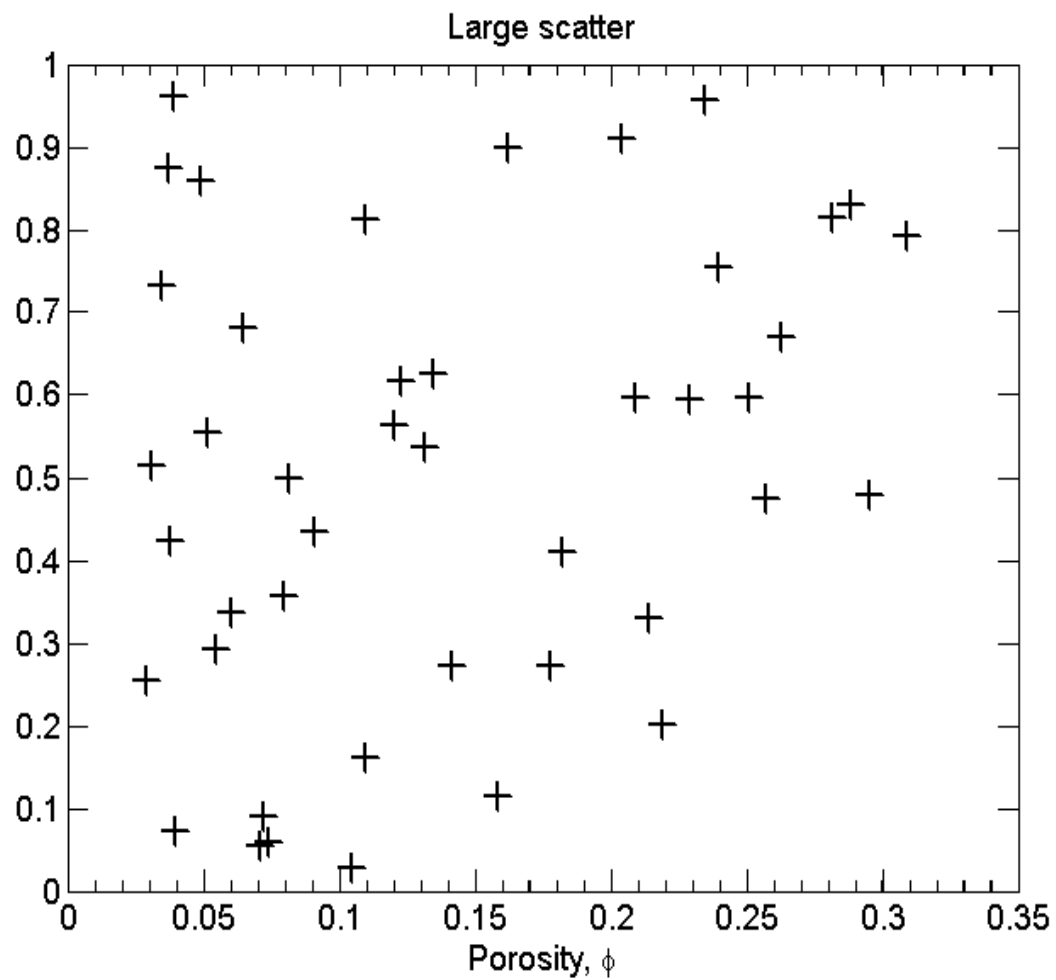


Forrás: EuroStat

# Scatter diagram

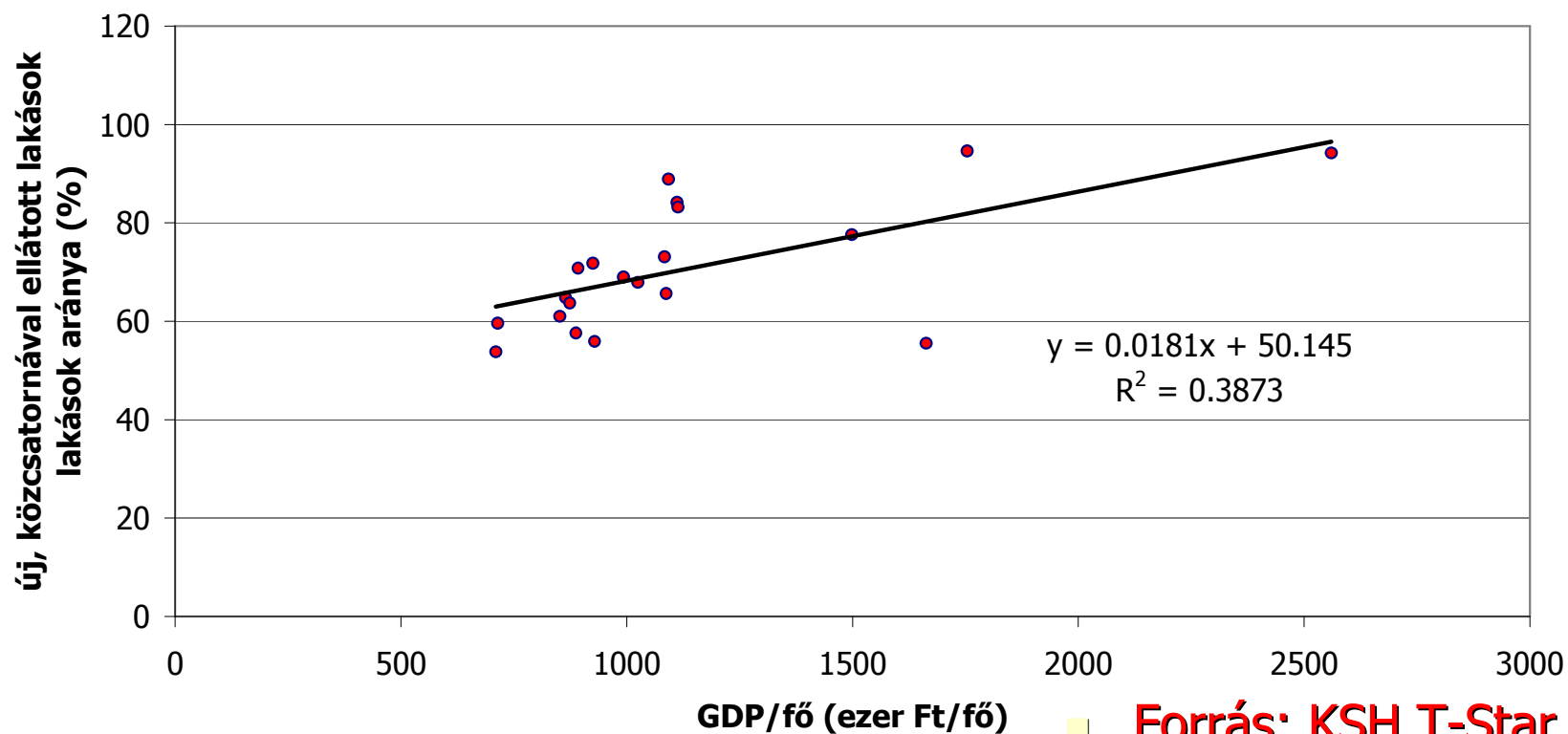


# Scatter diagram



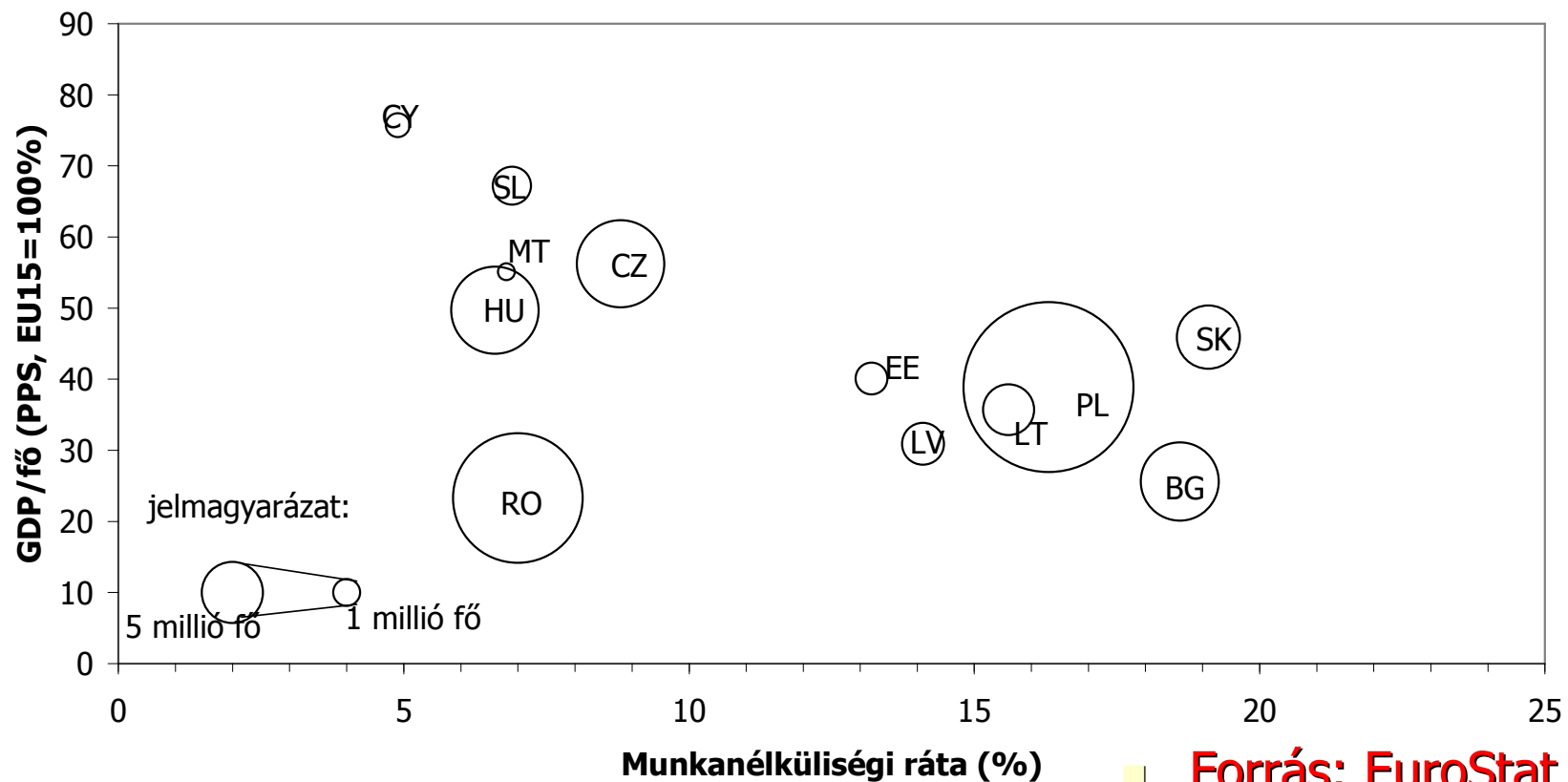
# Pontdiagram speciális típusa a regressziós diagram

**Értéktermelő-képesség és az újonnan épített lakások  
csatornával való ellátottságának összefüggése a magyar  
megyékben (2000)**



# Buborékdiagram: háromdimenziós összehasonlítás

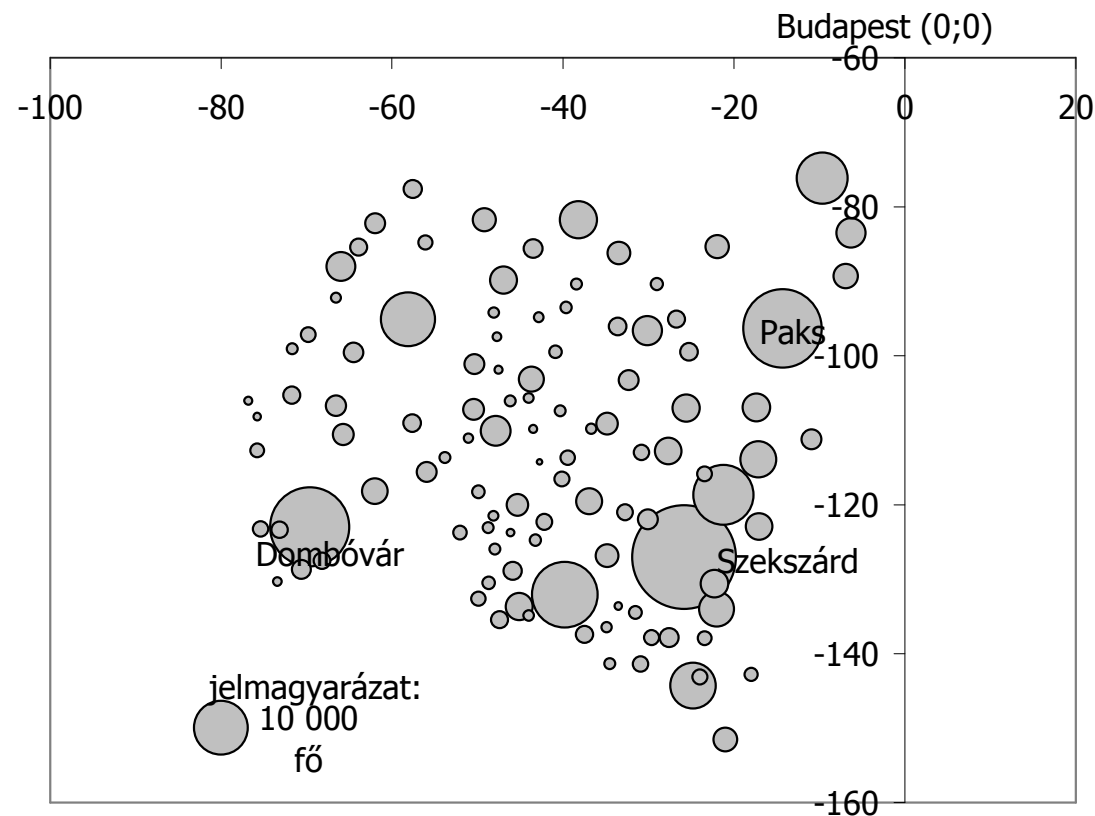
Az EU csatlakozásra váró országok három jelzőszám tükrében  
(2000)





# Buborékdiagram: sajátos esete a piktogrammos térkép

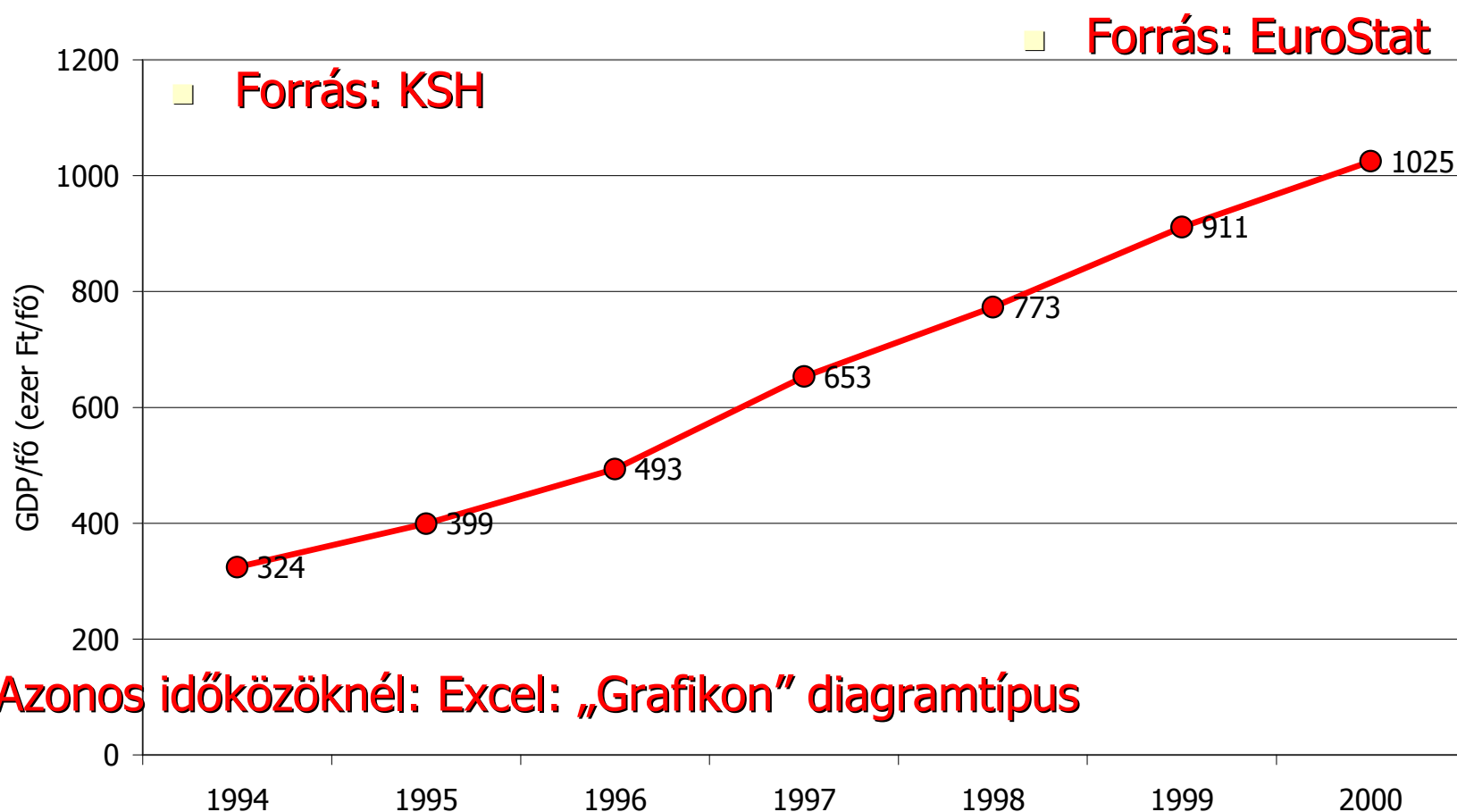
**Tolna megye népességének települési megoszlása (2000)**



■ **Forrás: KSH T-Star**

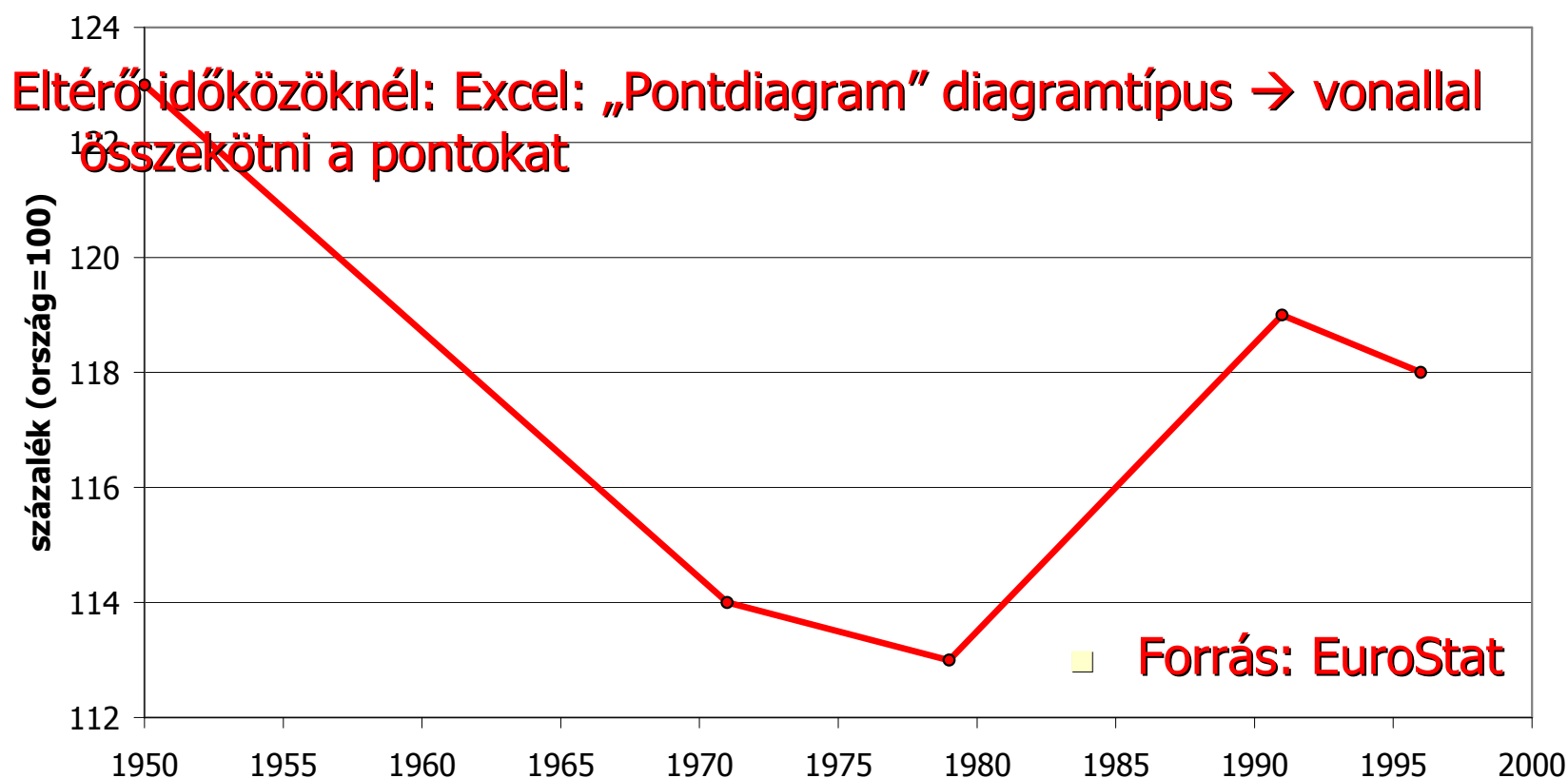
# Vonaldiagram (grafikon): két-dimenziós összehasonlítás (az egyik dimenzió az idő)

Pest megye gazdasági fejlődése (1994-2000)



# Vonaldiagram (grafikon): két-dimenziós összehasonlítás (az egyik dimenzió az idő)

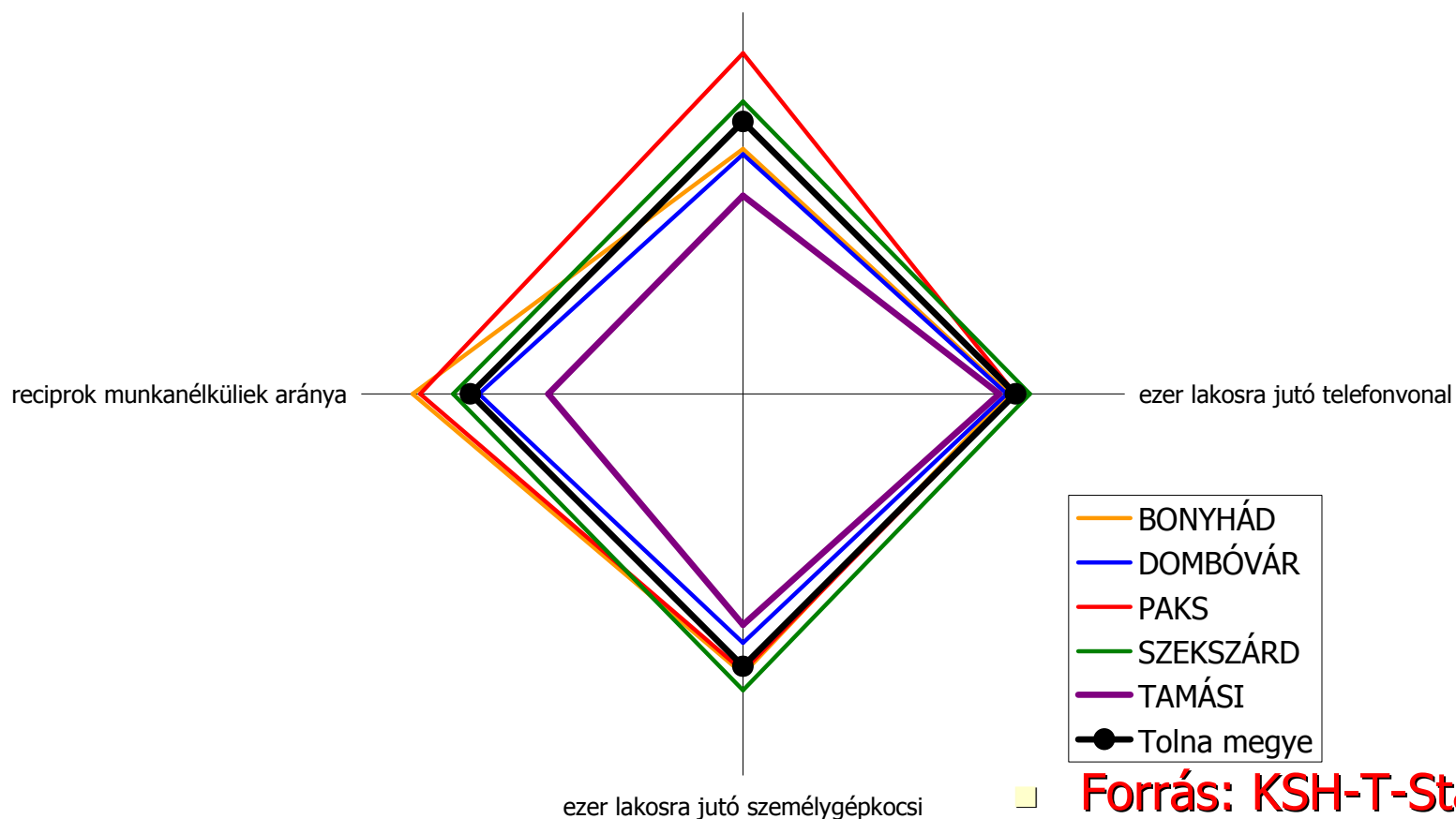
Délkelet-Anglia gazdasági fejlődése az egy főre jutó GDP alapján (1950-1996)



# Radar- (sugár-)diagram: sok-dimenziós összehasonlítás

**Tolna megye kistérségeinek fejlettsége (2000) (megyei átlag = 100%)**

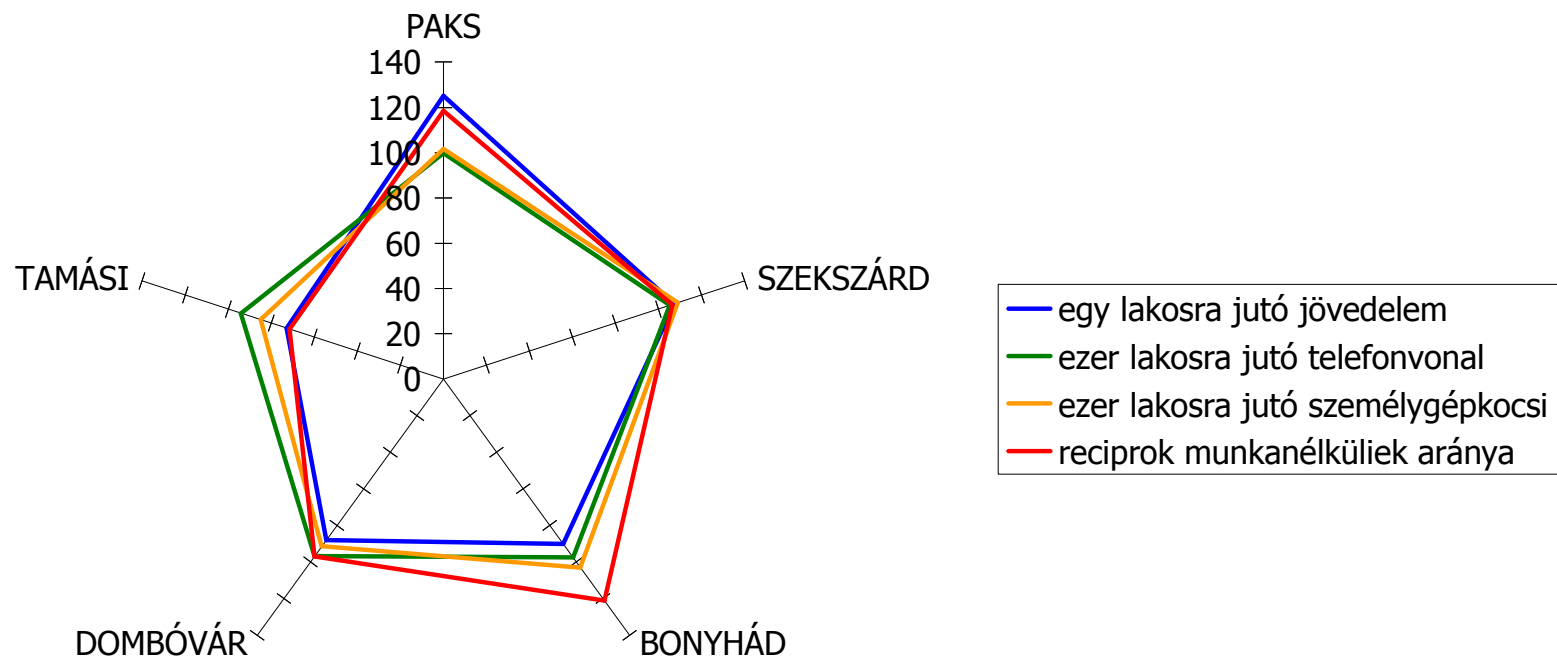
egy lakosra jutó jövedelem



Forrás: KSH-T-Star

# Radar- (sugár-)diagram: sok dimenziós összehasonlítás

**A fejlettség területi egyenlőtlenségei Tolna megyében (2000)**  
(megyei átlag=100%)





Nézzük mindig a dolgok napos  
oldalát!

**Mára befejeztük, viszontlátásra!**