

# Statisztika

## Politológus képzés

Daróczi Gergely

Politológia Tanszék

2012. február 28.



PÁZMÁNY PÉTER  
KATOLIKUS EGYETEM  
*Bölcsészettudományi Kar*

- 1 Változók és mérési szintek
  - Mérési szintek
  - Példák
- 2 A változók közötti kapcsolatról
  - Grafikus példák
  - A relációk típusai
  - További példák
  - A korrelációs együttható
- 3 A kutatás főbb fázisai

## Hány éves Ön?

- >18
- 19-24
- 25-30
- 31-40
- 41-100
- 100<

**Mutassa be a fenti példában a változót és annak attribútumait!**

## A változók típusai

- Belső változók
  - Függő változó
    - kvalitatív változók
    - kvantitatív változók
  - Független változó
    - kvalitatív változók
    - kvantitatív változók

## Külső változók

- Kontroll-változó
- Egyéb változók

**Hozzon 1-1 példát a fentebbi típusokra!**

### Kvalitatív változók:

- Nominális: egymást kizáró, a teljes válasz-univerzumot lefedő kategóriák
- Ordinális: u.a., mint fentebb + a kategóriák sorbarendeázhetők

	Nominális	Ordinális	Intervallum	Arány
Klasszifikáció	X	X	X	X
Sorrend		X	X	X
Egyenlő távolságok			X	X
Nullpont				X

### Kvantitatív változók:

- Intervallum: u.a., mint fentebb + a kategóriák egyenlő távolságra találhatók
- Arány: u.a., mint fentebb + nullpont

### Határozza meg a következő változók mérési szintjét!

- nem
- iskolai végzettség
- fizetés
- IQ
- tanulmányi átlag
- kor
- kedvenc szín
- születési hely

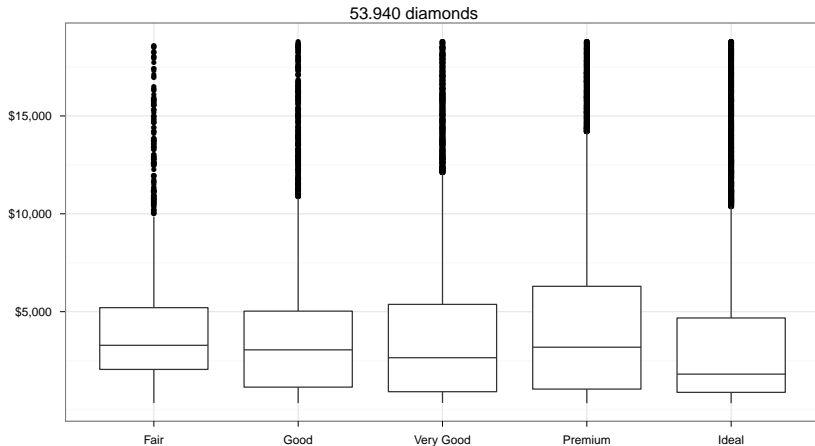
## Milyen változók, milyen mérési szintek feleltethetők meg az alábbi kérdéseknek/kijelentéseknek?

- Az idősebb emberek jobban félnek a bűnözőktől, mint a fiatalabbak.
- A gazdasági fejlődés valóban a várható életkor csökkenésével jár együtt?
- Melyik a legnagyobb magyar focidrukker csapat?
- Minél hosszabb az eljegyzés, annál hosszab a házasság.
- Mely foglalkozás bír a legnagyobb presztízzsel?

## Attribútumok?

# A változók közötti kapcsolatról

## Egy grafikus ábrázolás



```
ggplot(diamonds, aes(cut, price)) + geom_boxplot() + xlab("") + ylab("") +  
scale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds")
```



Prices of 50,000 round cut diamonds

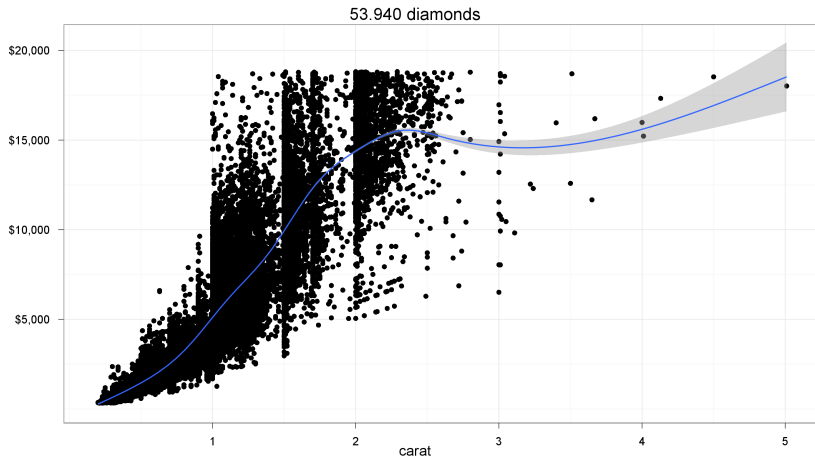
### Description:

A dataset containing the prices and other attributes of almost 54,000 diamonds. The variables are as follows:

- price. price in US dollars (\\$326--\\$18,823)
- carat. weight of the diamond (0.2--5.01)
- cut. quality of the cut (Fair, Good, Very Good, Premium, Ideal)
- colour. diamond colour, from J (worst) to D (best)
- clarity. a measurement of how clear the diamond is (I1 (worst), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (best))
- x. length in mm (0--10.74)
- y. width in mm (0--58.9)
- z. depth in mm (0--31.8)
- depth. total depth percentage =  $z / \text{mean}(x, y) = 2 * z / (x + y)$  (43--79)
- table. width of top of diamond relative to widest point (43--95)

# A változók közötti kapcsolatról

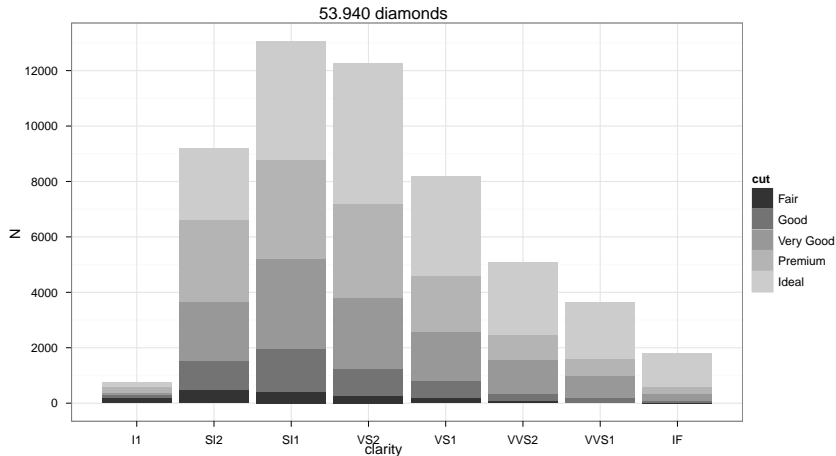
## Egy grafikus ábrázolás



```
ggplot(diamonds, aes(carat, price)) + geom_point() + geom_smooth() + ylab('') +  
scale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds")
```

# A változók közötti kapcsolatról

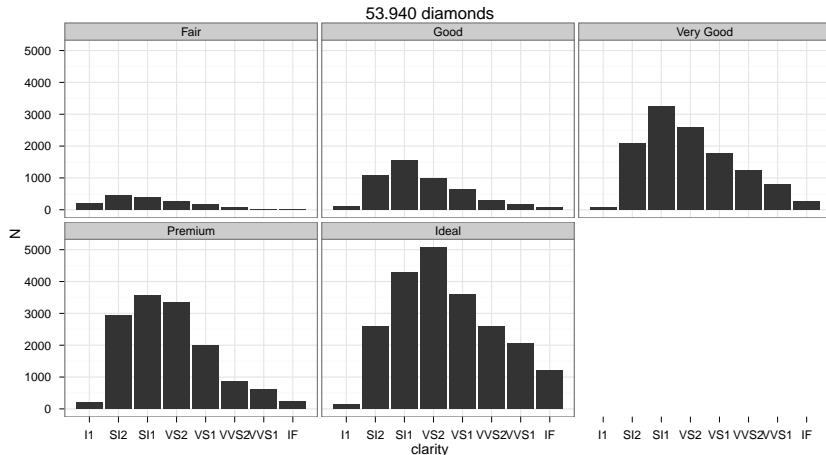
## Egy grafikus ábrázolás



```
ggplot(diamonds, aes(clarity, fill=cut)) + geom_bar() + ylab("N") +  
theme_bw() + opts(title="53.940 diamonds")
```

# A változók közötti kapcsolatról

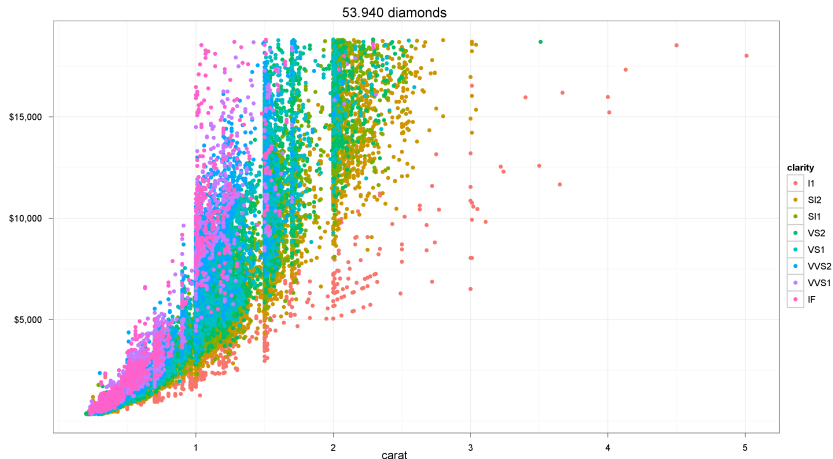
## Egy grafikus ábrázolás



```
ggplot(diamonds, aes(clarity)) + geom_bar() + ylab("N") + facet_wrap(~ cut) +  
theme_bw() + opts(title="53.940 diamonds")
```

# A változók közötti kapcsolatról

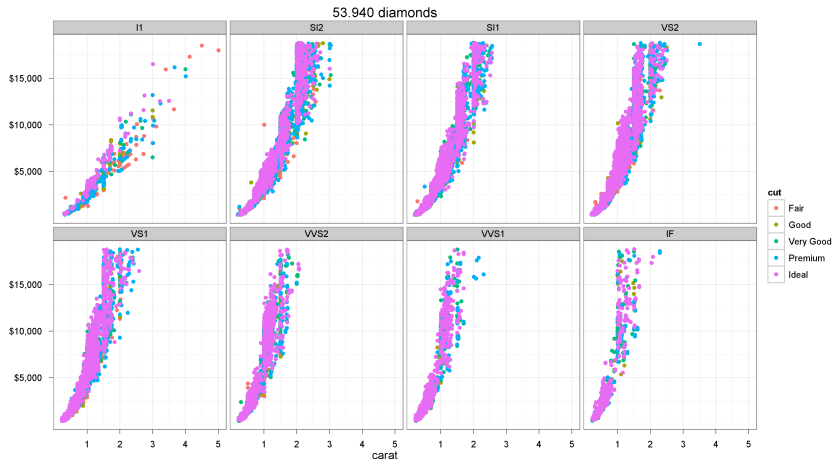
## Egy grafikus ábrázolás



```
ggplot(diamonds, aes(carat, price, color=clarity)) + geom_point() + ylab("") +  
scale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds")
```

# A változók közötti kapcsolatról

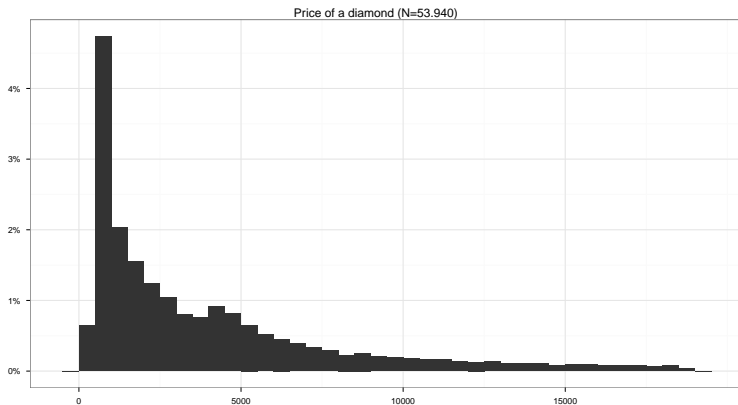
## Egy grafikus ábrázolás



```
ggplot(diamonds, aes(carat, price, color=cut)) + geom_point() + ylab("") + facet_wrap(~ clarity, nr=2) +  
scale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds")
```

# A változók közötti kapcsolatról

## Egy grafikus ábrázolás



Online kutatást végeztünk az Interneten gyémántot árusító emberek között.  
**Mit gondolhatunk a kutatás érvényességéről és megbízhatóságáról?**

## Lehetséges kapcsolat két változó között:

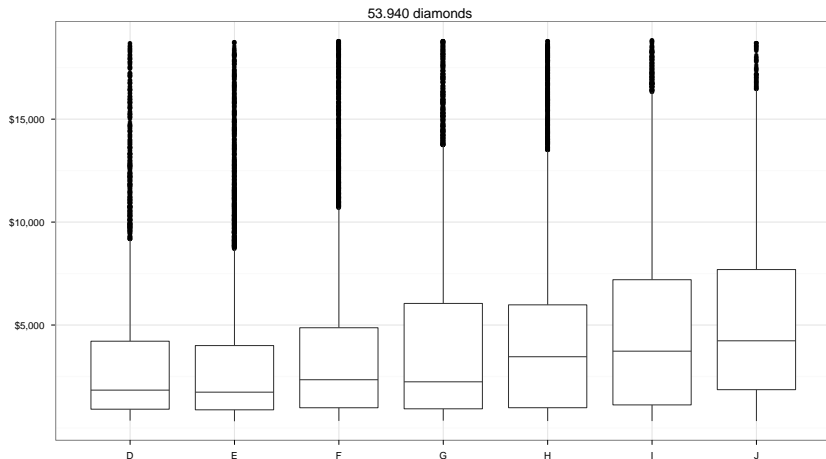
- együttjárás, asszociáció,
- korreláció ,
- hamis/látszólagos összefüggés,
- hatás,
- a hatás iránya,
- **okság.**

**szignifikáns...**



# A változók közötti kapcsolatról

## A hatás iránya



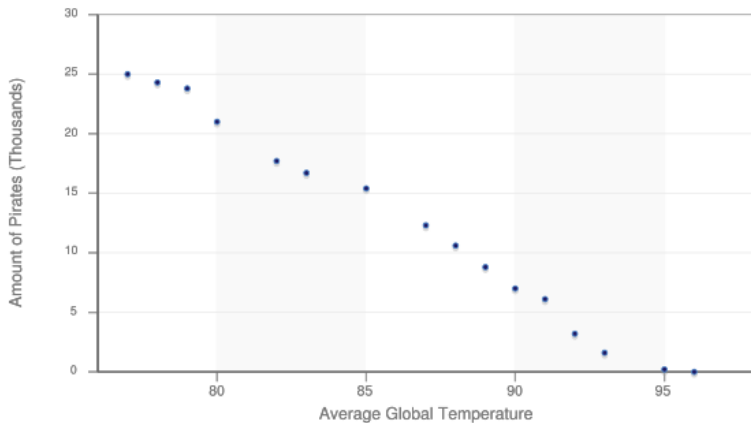
```
ggplot(diamonds, aes(color, price)) + geom_boxplot() + xlab("") + ylab("") +  
scale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds")
```

[illegible]

Forrás: [http://en.wikipedia.org/wiki/Diamond\\_color](http://en.wikipedia.org/wiki/Diamond_color)

# A változók közötti kapcsolatról

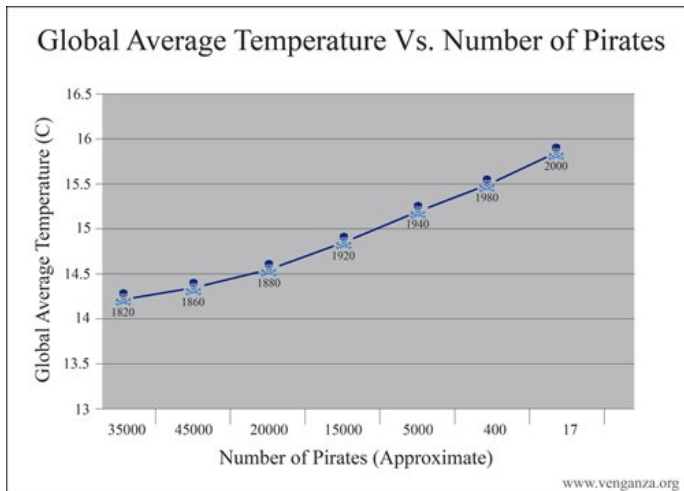
Magas korreláció



A korrelációs együttható magas értéke nem cáfolható. Vajon miért?

# A változók közötti kapcsolatról

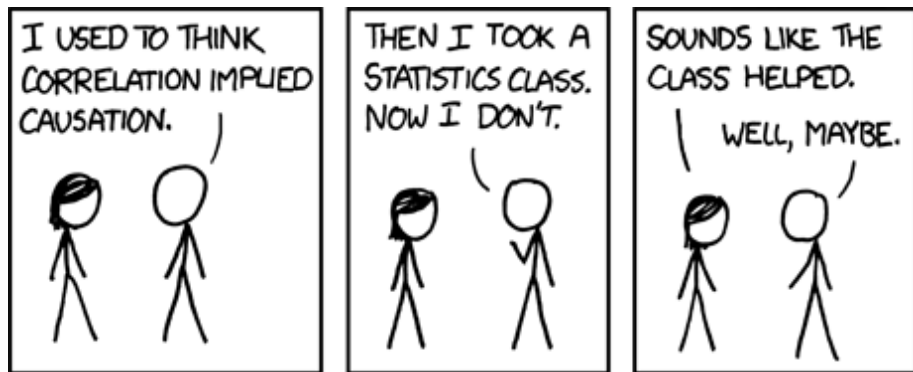
Magas korreláció



A korrelációs együttható magas értéke nem cáfolható. És?

# A változók közötti kapcsolatról

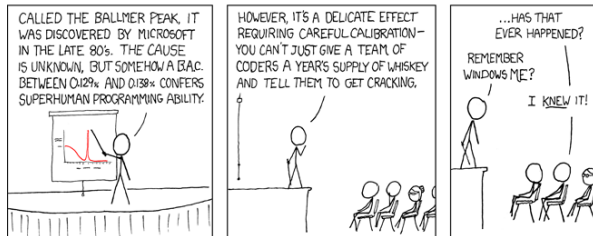
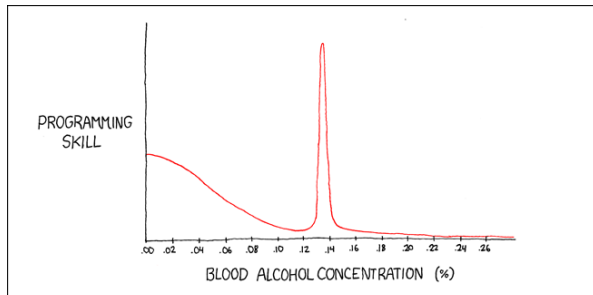
Correlation does not imply causality!



Forrás: <http://xkcd.com/552/>

# A változók közötti kapcsolatról

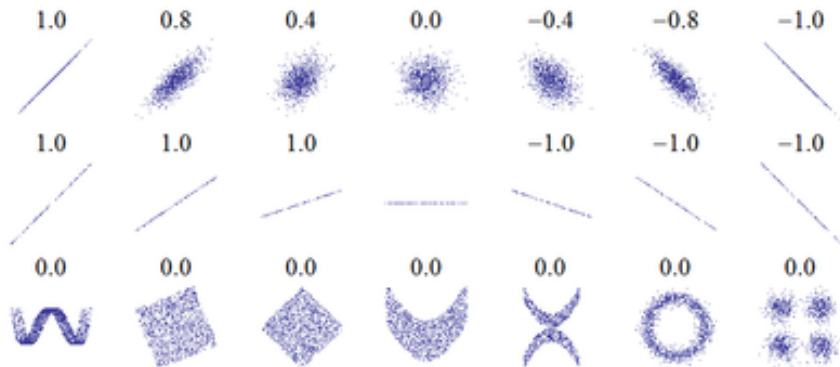
Alacsony korrelációs együtttható. Nincs kapcsolat?



Forrás: <http://xkcd.com/323/>

# A változók közötti kapcsolatról

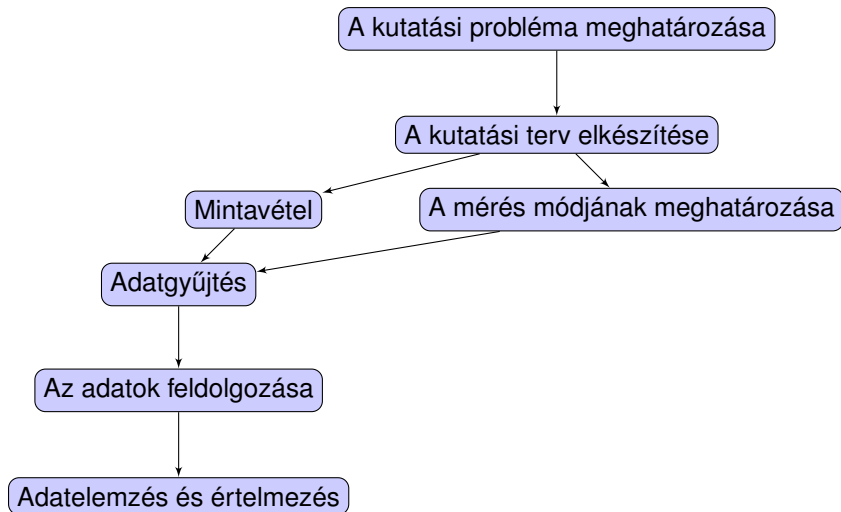
## Korreláció



Pozitív („egyenes arányosság”:  $R = 1$ ), negatív („fordított arányosság”:  $R = -1$ ), lineáris, négyzetes ... és korrelálatlan ( $R = 0$ ) kapcsolat  
R: korrelációs együttható

# A kutatás főbb fázisai

Egy folyamatábra





# Köszönöm a figyelmet!

Daróczi Gergely  
*[daroczi.gergely@btk.ppke.hu](mailto:daroczi.gergely@btk.ppke.hu)*