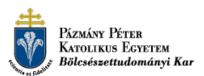
### Statisztika

### Politológus képzés

Daróczi Gergely

Politológia Tanszék

2012. február 28.



### **Outline**

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

### Változók és mérési szintek

Egy konkrét példa

### Hány éves Ön?

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

Mutassa be a fenti példában a változót és annak attribútumait!

### Változók és mérési szintek

Tipológia

#### 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!



### Mérési szintek

Kvalitatív és kvantitatív változók

#### Kvalitatív változók:

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

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

#### Kvantitatív változók:

- Intervallum: u.a., mint fentebb + a kategóriák egyenlő távolságra találhatóak
- 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

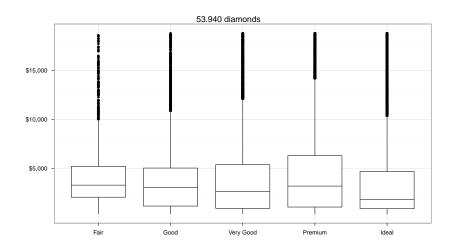
# Milyen változók, milyen mérési szintek feleltethetőek meg az alábbi kérdéseknek/kijelentéseknek?

- 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?



#### Egy grafikus ábrázolás



```
ggplot(diamonds, aes(cut, price)) + geom_boxplot() + xlab('') + ylab('') + yscale_y_continuous(formatter="dollar") + theme_bw() + opts(title="53.940 diamonds") + D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4 D + 4
```

### Az adatbázis szerkezete

#### ggplot2/diamonds

Prices of 50,000 round cut diamonds

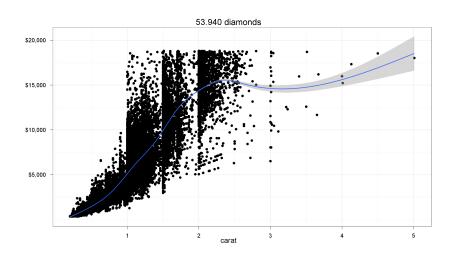
#### Description:

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 / mean(x, y) = 2 \* z / (x + y) (43--79)
- table. width of top of diamond relative to widest point (43--95)

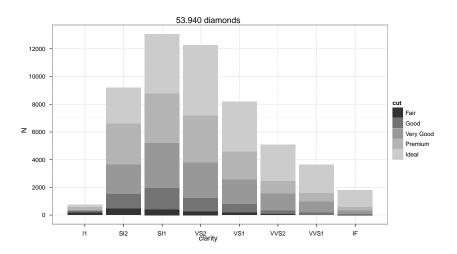


#### Egy grafikus ábrázolás



ggplot(diamonds, aes(carat, price)) + geom\_point() + geom\_smooth() + ylab('') +
scale\_y\_continuous(formatter="dollar") + theme\_bw() + opts(title="53.940 diamonds")

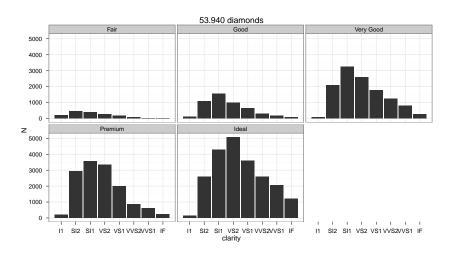
#### Egy grafikus ábrázolás



 $\label{eq:general_general} \mbox{ggplot(diamonds, aes(clarity, fill=cut)) + geom\_bar() + ylab("N") + theme\_bw() + opts(title="53.940 diamonds")}$ 

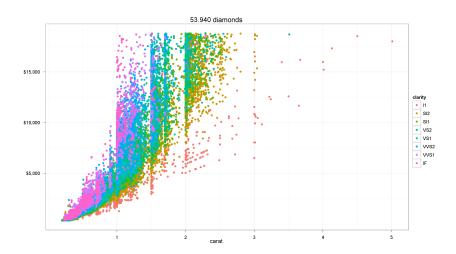
11 / 25

#### Egy grafikus ábrázolás



 $\label{eq:ggplot} $$\gcd(diamonds, aes(clarity)) + \gcd_bar() + ylab("N") + facet_wrap(~ cut) + theme_bw() + opts(title="53.940 diamonds")$ 

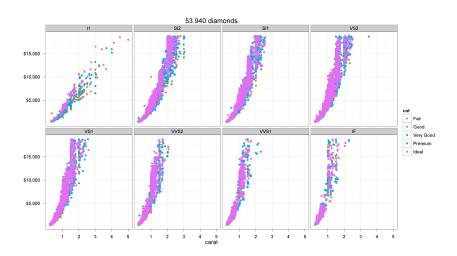
#### Egy grafikus ábrázolás



 $\label{eq:ggplot} $$\gcd(diamonds, aes(carat, price, color=clarity)) + geom\_point() + ylab('') + scale\_y\_continuous(formatter="dollar") + theme\_bw() + opts(title="53.940 diamonds")$ 

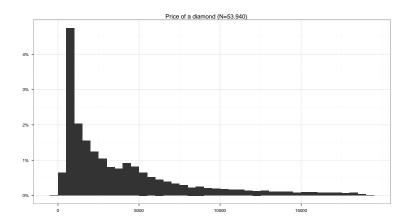
13 / 25

#### Egy grafikus ábrázolás



 $\label{eq:ggplot} $$\gcd(diamonds, aes(carat, price, color=cut)) + geom\_point() + ylab('') + facet\_wrap(~ clarity,nrscale\_y\_continuous(formatter="dollar") + theme\_bw() + opts(title="53.940 diamonds")$ 

Egy grafikus ábrázolás



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

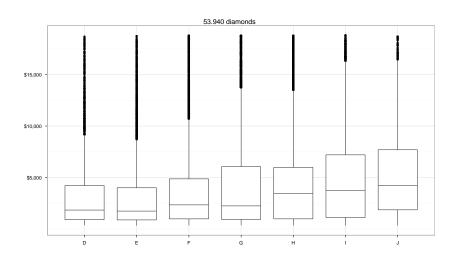
### A változók közötti kapcsolat lehetséges típusai

### Lehetséges kapcsolat két változó között:

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

szignifikáns...

#### A hatás iránya



17 / 25

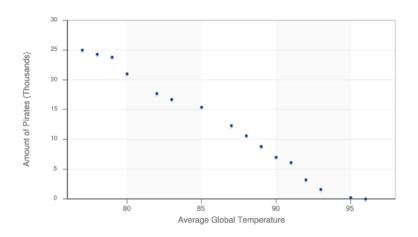
### Puska

#### A gyémánt színei

зіа	Status: current	AGS	Status: coment	AGS	Status: historical: pre 1995	CIBJO Status: current	IDC	Status: current	Sean. D.N.	Status: current	Old World Terms	Status: historica
yade and description [6]		grade and electronic colorimeter scale[7]		grade and electronic colorimeter scale [1]		grade <sup>[8]</sup>		grade and description [8]	grade for .50ct and over[F]	grade for under .50ct	series 1 scale[9]	zeries 2 scale
D		0	0-0.49	0	0-0.75	Exceptional white +	Exceptional white +	Colorless	River	White	Finest White	Jager
Е	Colorless	0.5	0.5-0.99	1	0.76-1.35	Exceptional white	Exceptional white	Coloness				River
F		1.0	1.0-1.49	2	1.38-2.00	Rare white +	Rare white +	Colorless when viewed through the crown	Top Wesselton Wesselton		Fine White	Kirel
G		1.5	1.5-1.99			Rare white	Rare white					Top Wesselton
н	Near Colonless	2.0	2.0-2.49	3	2.01-2.50	White	White				White	Wesselton
1	Near Coroness	2.5	2.5-2.99	4	2.51-3.0	Slightly tinted white		Slightly colored	Top Crystal	Slightly tinted white	Commercial White	Top Crystal
J		3.0	3.0-3.49	5	3.01-3.75	Slightly tinted white	Slightly tinted white		Crystal		Top silver cape	Crystal
K		3.5	3.5-3.99		3.76-4.5	Tinted white	Tinted white		Top cape	Tinted white	Silver cape	Тор саре
L	Faint Yellow	4.0	4.0-4.49	l °								
м		4.5	5 4.5-4.99	7	4.51-6.50 5.51-7.0	Tinted color 1		Slightly aclosed to colored	Cape	Tinted color	Light cape	Cape
N		5.0	5.0-5.49	17		Tinted color 2	Tinted color					Low Cape
0		5.5	5.5-5.99	8					Light yellow		Cape	Very light yellow
Р	Very Light Yellow	6.0	6.0-6.49									Light yellow
Q		6.5	6.5-6.99									
R		7.0	7.0-7.49		7.01-8.5						Dark cape	
s		7.5	7.5-7.99			Tinted color 3			Yellow			
т		8.0	8.0-8.49									
U		8.5	8.5-8.99	10	8.51-10							
v	Light Yellow	9.0	9.0-9.49									
w		9.5	9.5-9.99									
х				, and the second								
Υ		10 10+		10+								
z			i i									

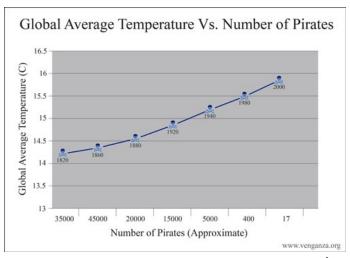
Forrás: http://en.wikipedia.org/wiki/Diamond\_color

Magas korreláció



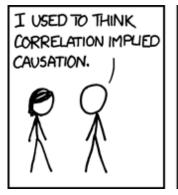
A korrelációs együttható magas értéke nem cáfolható. Vajon miért?

Magas korreláció



A korrelációs együttható magas értéke nem cáfolható. És?

Correlation does not imply causality!

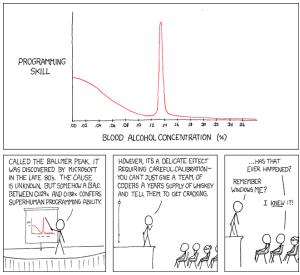




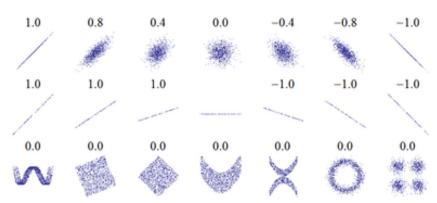


Forrás:http://xkcd.com/552/

Alacsony korrelációs együttható. Nincs kapcsolat?



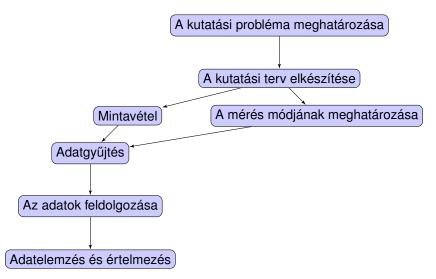
Korreláció



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