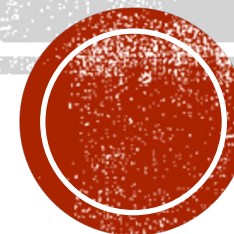


CURS 3 — ANALIZA DE FRECVENTE

Lect. univ. dr. Adrian Gorbănescu



INTRODUCERE

- **Profesorul de statistică aplică un chestionar de evaluare a conștiinciozității studenților săi. Tabelul 3.1 prezintă scorurile obținute de cei 50 de studenți care au completat chestionarul.**

Tabel 3.1 – Scorurile pentru variabila conștiinciozitate obținute de cei 50 de studenți

57	59	51	64	66
55	68	67	71	61
61	54	52	57	61
59	67	55	74	72
64	65	51	72	76
52	73	53	67	57
51	45	52	59	70
66	48	48	71	71
51	75	74	45	71
66	75	54	63	63



INTRODUCERE

- *Poți descrie pe scurt scorurile obținute de studenți?*
- *Care este cel mai mic scor, dar cel mai mare?*
- *Care este scorul care apare de cele mai multe ori, dar cel cu cea mai mică frecvență?*



FRECVENTA ABSOLUTA

- ***Frecvența absolută (fa) este numărul de apariții al fiecărei valori în distribuție.***
- ***frecvența absolută a valorii 45 este 2 (doi studenți au obținut scorul 45)***
- ***pentru valoarea 67 avem frecvența absolută egală cu 3.***
- ***Valorile cu frecvența 0 nu apar în tabel tocmai pentru a face mai ușoară interpretarea datelor.***
- ***Suma tuturor frecvențelor absolute este egală cu numărul total de scoruri din distribuție (în cazul nostru, 50).***



FRECVENTA ABSOLUTA

	Frecvența absolută (fa)
45	2
48	2
51	4
52	3
53	1
54	2
55	2
57	3
59	3
61	3
63	2
64	2
65	1
66	3
67	3
68	1
70	1
71	4
72	2
73	1
74	2
75	2
76	1
Total	50



FRECVENTA CUMULATA

- ***Frecvența cumulată (fc) reprezintă numărul total de valori începând de la cel mai mic scor din distribuție până la o anumită valoare, inclusiv.***
- ***De exemplu, în Tabelul 3.2 avem opt cazuri/participanți până la scorul 51.***
- ***30 dintre studenții care au răspuns la chestionar au obținut un scor mai mic sau egal cu 65.***
- ***Întotdeauna frecvența cumulată a ultimului scor din distribuție coincide cu suma frecvențelor absolute. În acest exemplu, frecvența cumulată pentru scorul 76 este 50.***



FRECVENTA CUMULATA

	Frecvența absolută (fa)	Frecvența cumulată (fc)
45	2	2
48	2	4
51	4	8
52	3	11
53	1	12
54	2	14
55	2	16
57	3	19
59	3	22
61	3	25
63	2	27
64	2	29
65	1	30
66	3	33
67	3	36
68	1	37
70	1	38
71	4	42
72	2	44
73	1	45
74	2	47
75	2	49
76	1	50
Total	50	



FRECVENTA RELATIVA PROCENTUALA

- ***Frecvența relativă procentuală ($fr\%$) exprimă procentul care corespunde unei valori din cadrul distribuției.***
- ***Suma tuturor frecvențelor procentuale este egală cu 100%.***

$$fr\% = \frac{fa}{\sum fa} * 100$$

(formula 3.1)

- ***4% din studenți au obținut scorurile 45 și 48.***
- ***8% au obținut scorul 51 la chestionarul de conștiinciozitate.***



FRECVENTA RELATIVA PROCENTUALA

		Frecvența absolută (fa)	Frecvența cumulată (fc)	Frecvența procentuală (fr%)
Valid	45	2	2	4,0
	48	2	4	4,0
	51	4	8	8,0
	52	3	11	6,0
	53	1	12	2,0
	54	2	14	4,0
	55	2	16	4,0
	57	3	19	6,0
	59	3	22	6,0
	61	3	25	6,0
	63	2	27	4,0
	64	2	29	4,0
	65	1	30	2,0
	66	3	33	6,0
	67	3	36	6,0
	68	1	37	2,0
	70	1	38	2,0
	71	4	42	8,0
	72	2	44	4,0
	73	1	45	2,0
	74	2	47	4,0
	75	2	49	4,0
	76	1	50	2,0
	Total	50		100,0



FRECVENTA RELATIVA CUMULATA PROCENTUAL

- ***Frecvența relativă cumulată procentuală (frc%) ne indică procentul cumulat al scorurilor din distribuție până la o anumită valoare, inclusiv.***
- ***Pentru cel mai mare scor din distribuție întotdeauna frecvența cumulată procentuală este 100%.***
- ***Pentru scorul 55 avem o frecvență cumulată procentuală de 32%.***
- ***Acest rezultat se traduce prin faptul că 32% dintre studenți au un scor la chestionarul de conștiinciozitate mai mic sau egal cu 55.***



FRECVENTA RELATIVA CUMULATA PROCENTUAL

	Frecvența absolută (fa)	Frecvența cumulată (fc)	Frecvența procentuală (fi%)	Frecvența cumulată procentual (frc%)
45	2	2	4,0	4,0
48	2	4	4,0	8,0
51	4	8	8,0	16,0
52	3	11	6,0	22,0
53	1	12	2,0	24,0
54	2	14	4,0	28,0
55	2	16	4,0	32,0
57	3	19	6,0	38,0
59	3	22	6,0	44,0
61	3	25	6,0	50,0
63	2	27	4,0	54,0
64	2	29	4,0	58,0
65	1	30	2,0	60,0
66	3	33	6,0	66,0
67	3	36	6,0	72,0
68	1	37	2,0	74,0
70	1	38	2,0	76,0
71	4	42	8,0	84,0
72	2	44	4,0	88,0
73	1	45	2,0	90,0
74	2	47	4,0	94,0
75	2	49	4,0	98,0
76	1	50	2,0	100,0
Total	50		100,0	



FRECVENTA RELATIVA CUMULATA PROCENTUAL

- *Frecvența relativă cumulată procentual se numește rang percentil.*
- *Scorul 64 are frecvența cumulată procentuală 58,0 și putem spune că îi corespunde rangul percentil 58.*
- *Valoarea corespunzătoare unui rang percentil poartă denumirea de percentilă.*
- *scorul 64 este percentila 58.*



FRECVENTA RELATIVA CUMULATA PROCENTUAL

- *percentila 25 (quartila 1; corespunde rangului percentil 25).*
- *percentila 50 (quartila 2 sau mediană; corespunde rangului percentil 50).*
- *percentila 75 (quartila 3; corespunde rangului percentil 75).*



ANALIZA DE FRECVENTE ÎN R

- Analiza de frecvențe solicită pachetul **psych**
- `library(psych)`
- Vom analiza frecvențele variabilei constiințiozitate.



ANALIZA DE FRECVENTE ÎN R

1. Frecvența absolută

- vom crea un obiect pe care îl vom numi sugestiv faconst.
 - Această denumire ne permite să înțelegem că este vorba despre frecvența absolută a variabilei conștiinciozitate
-
- `faconst <- table(bdcap3$conștiinciozitate)`
 - `cbind(faconst)`
-
- `table` — este funcția din R care calculează frecvențele absolute.
 - `bdcap3` — reprezintă denumirea bazei de date.
 - `conștiinciozitate` — este variabila ale cărei frecvențe absolute dorim să le calculăm.
 - `cbind` — reprezintă funcția care va genera un tabel cu frecvența absolută a fiecărei valori.



ANALIZA DE FRECVENTE ÎN R

```
# Frecvența absolută  
faconst<-table(bdcap3$constiinciozitate)  
cbind(faconst)
```

	faconst
45	2
48	2
51	4
52	3
53	1
54	2
55	2
57	3
59	3
61	3
63	2
64	2
65	1
66	3
67	3
68	1
70	1
71	4
72	2
73	1
74	2
75	2
76	1



ANALIZA DE FRECVENTE ÎN R

2. Frecvența cumulată

- vom genera un obiect (fcumconst) care să primească rezultatele frecvențelor cumulate.
- `fcumconst <- cumsum(faconst)`
- `cbind(faconst, fcumconst)`
- `cumsum` — este funcția necesară pentru a calcula frecvențele cumulate.
- `faconst` — reprezintă frecvențele absolute ale variabilei conștiinciozitate, care vor fi adunate pentru a calcula frecvențele cumulate.
- `cbind` — va genera un tabel care va conține atât frecvențele absolute, cât și frecvențele cumulate ale variabilei conștiinciozitate.



ANALIZA DE FRECVENTE ÎN R

```
# Frecența cumulată
```

```
fcumconst <- cumsum(faconst)  
cbind(faconst, fcumconst)
```

	faconst	fcumconst
45	2	2
48	2	4
51	4	8
52	3	11
53	1	12
54	2	14
55	2	16
57	3	19
59	3	22
61	3	25
63	2	27
64	2	29
65	1	30
66	3	33
67	3	36
68	1	37
70	1	38
71	4	42
72	2	44
73	1	45
74	2	47
75	2	49
76	1	50



ANALIZA DE FRECVENTE ÎN R

3. Frecvența relativă

- trebuie creat un obiect care să primească rezultatele acestui tip de frecvență. Vom denumi acest obiect `frconst`.
 - `frconst <- prop.table(faconst)`
 - `cbind(faconst, fcumconst, frconst)`
-
- `frconst` — este obiectul care conține frecvențele relative ale variabilei conștiinciozitate.
 - `prop.table` — este care calculează frecvența relativă; `faconst` este frecvența absolută (în funcție de ea se calculează frecvența relativă)
 - `cbind` — generează tabelul care cuprinde frecvențele absolute, frecvențele cumulate și frecvențele relative.



ANALIZA DE FRECVENTE ÎN R

```
# Frecvența relativă  
frconst<-prop.table(faconst)  
cbind(faconst, fcumconst, frconst)
```

	faconst	fcumconst	frconst
45	2	2	0.04
48	2	4	0.04
51	4	8	0.08
52	3	11	0.06
53	1	12	0.02
54	2	14	0.04
55	2	16	0.04
57	3	19	0.06
59	3	22	0.06
61	3	25	0.06
63	2	27	0.04
64	2	29	0.04
65	1	30	0.02
66	3	33	0.06
67	3	36	0.06
68	1	37	0.02
70	1	38	0.02
71	4	42	0.08
72	2	44	0.04
73	1	45	0.02
74	2	47	0.04
75	2	49	0.04
76	1	50	0.02



ANALIZA DE FRECVENTE ÎN R

4. Frecvența relativă procentuală

- `Frpconst` este obiectul pe care îl generăm pentru a primi frecvențele procentuale ale variabilei conștiințiozitate.
 -
 - `frpconst <- frconst*100`
 - `cbind(faconst, fcumconst, frconst, frpconst)`
-
- `frpconst` — este obiectul care conține frecvențele procentuale ale variabilei analizate.
 - `frconst*100` - este formula în funcție de care se calculează frecvența procentuală (vezi formula 3.1).
 - `cbind` — generează tabelul care cuprinde frecvențele absolute, frecvențele cumulate, frecvențele relative și frecvențele relative procentuale.



ANALIZA DE FRECVENTE ÎN R

```
#Frecvența relativă procentuală  
frpconst <- frconst*100  
cbind(faconst, fcumconst, frconst, frpconst)
```

	faconst	fcumconst	frconst	frpconst
45	2	2	0.04	4
48	2	4	0.04	4
51	4	8	0.08	8
52	3	11	0.06	6
53	1	12	0.02	2
54	2	14	0.04	4
55	2	16	0.04	4
57	3	19	0.06	6
59	3	22	0.06	6
61	3	25	0.06	6
63	2	27	0.04	4
64	2	29	0.04	4
65	1	30	0.02	2
66	3	33	0.06	6
67	3	36	0.06	6
68	1	37	0.02	2
70	1	38	0.02	2
71	4	42	0.08	8
72	2	44	0.04	4
73	1	45	0.02	2
74	2	47	0.04	4
75	2	49	0.04	4
76	1	50	0.02	2



ANALIZA DE FRECVENTE ÎN R

5. *Frecvența relativă procentuală*

- `Frcpconst` este obiectul generat pentru a primi frecvențele relative cumulate procentuale ale variabilei conștiinciozitate.
-
- `frpconst<-cumsum(frpconst)`
- `cbind(faconst, fcumconst, frconst,frpconst,frcpconst)`
- `cumsum` — este funcția necesară pentru a calcula frecvențele cumulate.
- `frpconst` — reprezintă frecvențele relative procentuale ale variabilei conștiinciozitate, care vor fi adunate pentru a calcula frecvențele cumulate procentuale.
- `cbind` — generează tabelul care cuprinde frecvențele absolute, frecvențele cumulate, frecvențele relative, frecvențele relative procentuale și frecvențele relative cumulate procentuale.



ANALIZA DE FRECVENTE ÎN R

```
#Frecvență relativă cumulată procentuală
```

```
frcpconst<-cumsum(frpconst)
```

```
cbind(faconst, fcumconst, frconst, frpconst, frcpconst)
```

	faconst	fcumconst	frconst	frpconst	frcpconst
45	2	2	0.04	4	4
48	2	4	0.04	4	8
51	4	8	0.08	8	16
52	3	11	0.06	6	22
53	1	12	0.02	2	24
54	2	14	0.04	4	28
55	2	16	0.04	4	32
57	3	19	0.06	6	38
59	3	22	0.06	6	44
61	3	25	0.06	6	50
63	2	27	0.04	4	54
64	2	29	0.04	4	58
65	1	30	0.02	2	60
66	3	33	0.06	6	66
67	3	36	0.06	6	72
68	1	37	0.02	2	74
70	1	38	0.02	2	76
71	4	42	0.08	8	84
72	2	44	0.04	4	88
73	1	45	0.02	2	90
74	2	47	0.04	4	94
75	2	49	0.04	4	98
76	1	50	0.02	2	100



FRECVENTA RELATIVA CUMULATA PROCENTUAL

- *Frecvența relativă cumulată procentual se numește rang percentil.*
- *Scorul 64 are frecvența cumulată procentuală 58,0 și putem spune că îi corespunde rangul percentil 58.*
- *Valoarea corespunzătoare unui rang percentil poartă denumirea de percentilă.*
- *scorul 64 este percentila 58.*



FRECVENTA RELATIVA CUMULATA PROCENTUAL

- *percentila 25 (quartila 1; corespunde rangului percentil 25).*
- *percentila 50 (quartila 2 sau mediană; corespunde rangului percentil 50).*
- *percentila 75 (quartila 3; corespunde rangului percentil 75).*



FRECVENTA RELATIVA CUMULATA PROCENTUAL

- *percentila 25 (quartila 1; corespunde rangului percentil 25).*
- *percentila 50 (quartila 2 sau mediană; corespunde rangului percentil 50).*
- *percentila 75 (quartila 3; corespunde rangului percentil 75).*



FRECVENTA RELATIVA CUMULATA PROCENTUAL

- În R obținem percentilele 25, 50 și 75 (Q1, Q2 și Q3) folosind funcția **quantile()**.
 - `quartconst <- quantile (bdcap3$constiinciozitate, probs = c(0.25, 0.50, 0.75))`
 - `quatconst`
-
- `quartconst` — este obiectul creat pentru a primi valorile corespunzătoare celor trei quartile.
 - `bdcap3$constiinciozitate` — reprezintă numele bazei de date, respectiv variabila analizată.
 - `probs` — indică cele trei valori corespunzătoare rangurilor percentile.
 - `c` — permite cumulara mai multor valori în aceeași funcție.



FRECVENTA RELATIVA CUMULATA PROCENTUAL

```
quartconst<-quantile(bdcap3$constiinciozitate, probs = c(0.25,0.50,0.75))  
quartconst
```

25%	50%	75%
54.0	62.0	69.5



FRECVENTA RELATIVA CUMULATA PROCENTUAL

- dacă dorim să calculăm percentilele 22, 60 72 și 84 linia de cod devine:
- `quartconst <- quantile (bdcap3$constiinciozitate, probs = c(0.22, 0.60, 0.72, 0.84))`
- `quartconst`

```
quartconst<-quantile(bdcap3$constiinciozitate, probs = c(0.22,0.60,0.72, 0.84))  
quartconst
```

22%	60%	72%	84%
52.78	65.40	67.28	71.16



REPREZENTĂRI DE TIP GRAFIC

1. Graficul de tip plăcintă

- Să ne imaginăm că cei 50 de studenți au răspuns și la întrebarea „În ce domeniu doriți să lucrați după absolvirea facultății?”, opțiunile fiind:
 - a) Psihoterapie
 - b) Psihologie organizațională
 - c) Securitate națională
 - d) Psihologie educațională
- 17 studenți au optat pentru psihoterapie, 11 pentru psihologie organizațională, 13 pentru securitate națională, iar 9 pentru psihologie educațională.



REPREZENTĂRI DE TIP GRAFIC

1. Graficul de tip plăcintă

- Generăm un obiect (**optiuni**) care să cuprindă opțiunile/răspunsurile studenților.
- Deoarece în baza de date opțiunile sunt codate cu numere 1 = Psihoterapie, 2 = Psihologie organizațională, 3 = Securitate națională și 4 = Psihologie educațională) vom crea un obiect (etichete) care să facă corespondența dintre fiecare opțiune și valoarea numerică.
- `optiuni <- c(17, 11, 13, 9)`
- `etichete<-c("Psihoterapie", "Psihologie organizațională", "Securitate națională", "Psihologie educațională")`

```
optiuni<-c(17, 11, 13, 9)
etichete<-c("Psihoterapie", "Psihologie organizațională", "Securitate națională", "Psihologie educațională")
```



REPREZENTĂRI DE TIP GRAFIC

1. Graficul de tip plăcintă

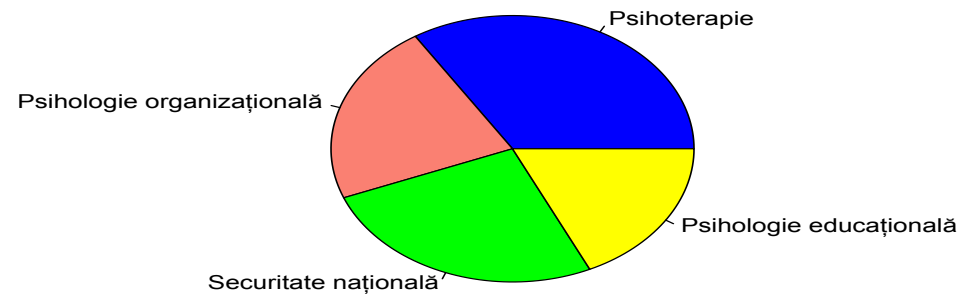
- `pie(optiuni, etichete, col = c("blue", "salmon", "green", "yellow"))`
- `pie` — reprezintă funcția destinată graficului de tip plăcintă
- `optiuni` — este obiectul în care sunt cuprinse răspunsurile studenților
- `etichete` — este obiectul care face corespondența dintre valorile numerice și eticheta lor
- `col` — este funcția din R care permite colorarea fiecărei „felii” din grafic.

```
pie(optiuni, etichete, col = c("blue", "salmon", "green", "yellow"))
```



REPREZENTĂRI DE TIP GRAFIC

1. Graficul de tip plăcintă



REPREZENTĂRI DE TIP GRAFIC

2. Graficul de tip bară

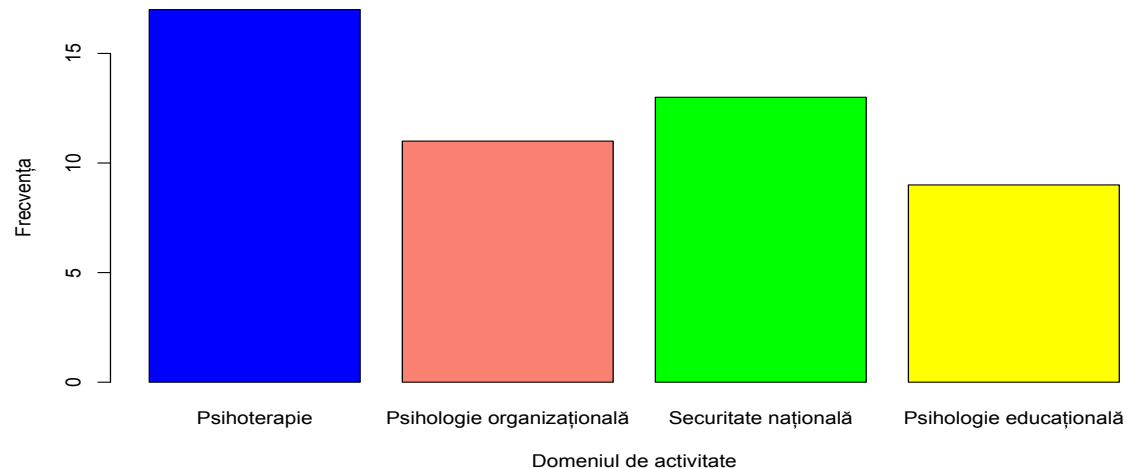
- `barplot(optiuni, names.arg=etichete, xlab="Domeniul de activitate", ylab="Frecvența", col = c("blue", "salmon", "green", "yellow"))`
- `barplot` — reprezintă funcția din R care generează graficul de tip bară
- `optiuni` — face trimitere la obiectul în care sunt cuprinse răspunsurile studenților
- `names.arg` — este un vector care va permite etichetarea fiecărei bare din grafic
- `xlab` — permite denumirea axei X
- `ylab` — permite denumirea axei Y
- `col` - permite colorarea fiecărei bare din grafic



REPREZENTĂRI DE TIP GRAFIC

2. Graficul de tip bară

```
barplot(optiuni, names.arg=etichete, xlab="Domeniul de activitate",  
        ylab="Frecvența", col = c("blue", "salmon", "green", "yellow"))
```



REPREZENTĂRI DE TIP GRAFIC

3. Graficul de tip histogramă

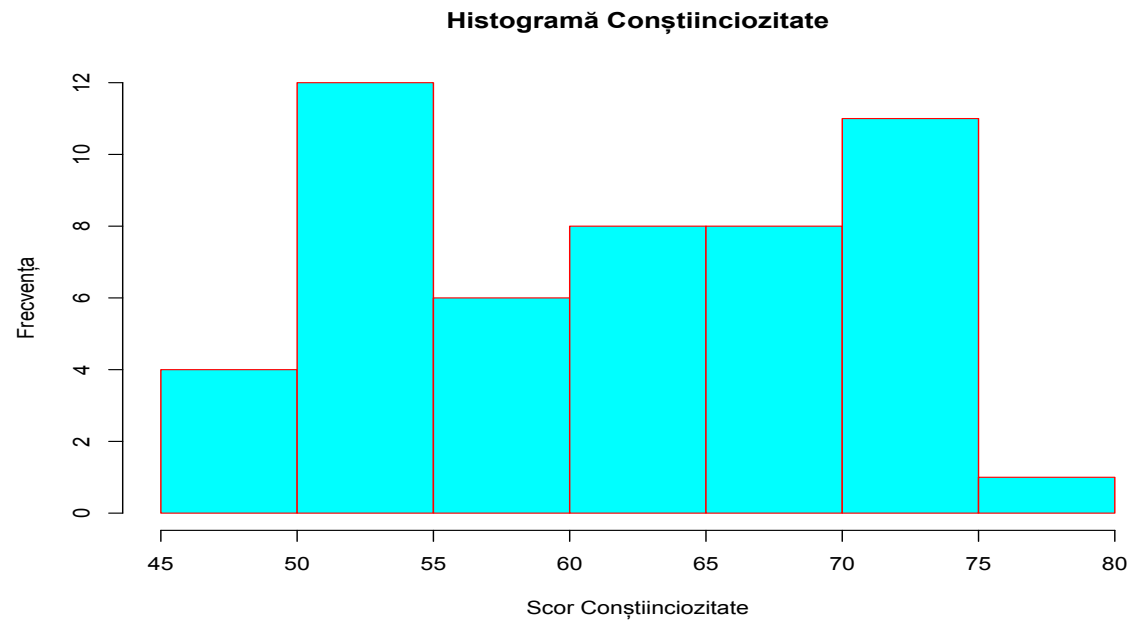
- `hist(bdcap3$conștiinciozitate, xlab="Scor Conștiinciozitate", ylab="Frecvența", col="cyan", border="red", main="Histogramă Conștiinciozitate")`
- `hist` – funcția necesară pentru a genera graficul de tip histogramă
- `bdcap3$conștiinciozitate` – reprezintă baza de date și variabila analizată
- `xlab` – permite denumirea axei X
- `ylab` – permite denumirea axei Y
- `col` – funcția necesară pentru a colora graficul
- `border` – permite colorarea conturului histogramelor
- `main` – este utilizat pentru a da un titlu general graficului



REPREZENTĂRI DE TIP GRAFIC

3. Graficul de tip histogramă

```
hist(bdcap3$constiinciozitate, xlab="Scor Conștiinciozitate", ylab="Frecvența",  
     col="cyan", border="red", main="Histogramă Conștiinciozitate")
```



VA MULTUMESC PENTRU PREZENTA

- https://www.youtube.com/watch?v=-Qo_o7sYeDI

Multumesc

