



Бизнис статистика

Предавање 2:

Дескриптивни статистики

Претставување на податоци со
табели и графици



Организација на податоците

- Податоците во првобитен облик (добиеени од некој извор на информации) најчесто не можат лесно да се разберат и да се користат за некакво донесување на одлуки.
- Првичната анализа на податоците се состои во нивна организација и систематизација како и определување на одредени бројни карактеристики.
- Организацијата на податоците се состои од
 - Табеларен приказ на распределбата на честоти (фреквенции)
 - Графички прикази на податоците и честотите
 - Определување бројни карактеристики на податоците



Организирање на податоците

- Техники за организирање и прикажување на податоци:

Категориски променливи

- Табели на честоти
- Столбести дијаграми
- Пити

Квантитативни нумерички променливи

- Табели на распределба на честоти
- Линиски графици
- Хистограми
- Стебло-лист дијаграми
- Графици на расејување

Организирање и прикажување на квантитативни податоци



Квантитативни податоци

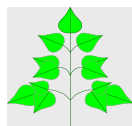
41, 24, 32, 26, 27, 27, 30, 24, 38, 21

Подреден
примерок

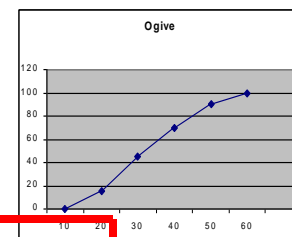
21, 24, 24, 26, 27,
27, 30, 32, 38, 41

Стебло - лист
приказ

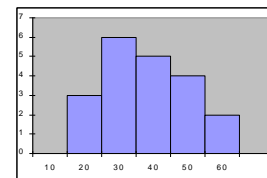
2 | 144677
3 | 028
4 | 1



Распределба на честоти
и
Кумулативна распределба
на честоти



Хистограм



Табела

Насочен
лак

Полигони



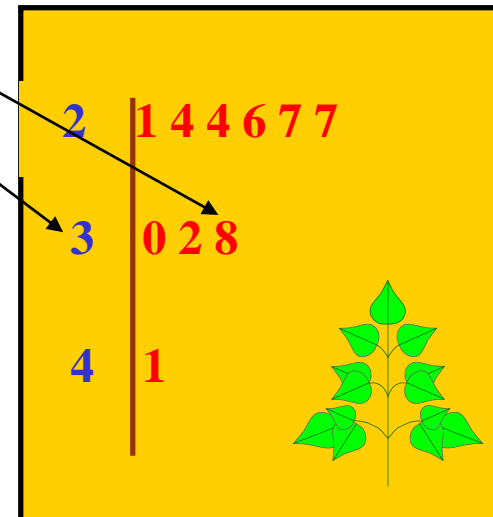
Стебло-лист дијаграм

- За да го направиме стебло-лист дијаграмот податоците
24, 26, 24, 21, 27, 27, 30, 41, 32, 38

ги подредуваме од најмал до најголем:

21, 24, 24, 26, 27, 27, 30, 32, 38, 41

- За да се направи дијаграмот секоја вредност се дели на два дела.
- Првиот дел од бројот се нарекува стебло, а вториот дел лист.
- Во примерот, стебло е првата цифра, а лист втората цифра. Кај децималните броеви, може целиот дел на бројот да биде стебло, а децималите лист.





Пример 1: Стебло-лист дијаграм

Дадени се резултатите на 30 студенти на колоквиумот по статистика

75	52	80	96	65	79	71	87	93	95
69	72	81	61	76	86	79	68	50	92
83	84	77	64	71	87	72	92	57	98

За да направиме стебло-лист дијаграм земаме првата цифра да биде стебло, а втората лист. Гледаме дека стеблото за секој податок е 5, 6, 7, 8, или 9 затоа што сите податоци се од 50 до 98.

Откако ќе ги излистаеме стеблата, ги читаме лисјата за секој податок и ги запишуваме до соодветното стебло десно од вертикалната линија.



Пример 1: Стебло лист приказ

50	52	57	61	64	65	68	69	71	71
72	72	75	76	77	79	79	80	81	83
84	86	87	87	92	92	93	95	96	98

Стебла



5

0



Лист за 50

6

1



Лист за 61

7

8

9



Пример 1: Стебло-лист дијаграмот

50	52	57	61	64	65	68	69	71	71
72	72	75	76	77	79	79	80	81	83
84	86	87	87	92	92	93	95	96	98

5	0	2	7						
6	1	4	5	8	9				
7	1	1	2	2	5	6	7	9	9
8	0	1	3	4	6	7	7		
9	2	2	3	5	6	8			

- Со користење на цифрите од податоците, може да идентификуваме 5 различни категории: 50-ки, 60-ки, 70-ки, 80-ки и 90-ки.
- Притоа, може да видиме дека 3 од податоците се 50-ки, 5 се 60-ки, 9 се 70-ки, итн.
- Овој дијаграм овозможува брзо да видиме како се распоредени податоците.



Табела на распределба на честоти

- Распределба на честоти се претставува со листа или табела.
- Табелата на распределба на честоти содржи:
 - класи (категории или интервали) во кои спаѓаат податоците
 - соодветните честоти со кои се застапени податоците во секоја класа.
- Распределба на честоти е начин да се организираат податоците во покорисен облик од првобитниот.
- Овозможува брза визуелна интерпретација на податоците.



Распределба на честоти за интервални податоци

- За да се направи табела на честоти за непрекинати квантитавни податоци потребно е податоците да се поделат во интервали.
- Затоа е потребно да определиме:

1. Број на интервали (k).

Премалку интервали можат да доведат до губење на информација.

Премногу интервали не се корисни за сумирање на податоците.

Најчесто се користат $5 \leq k \leq 15$ интервали или се користи некоја од следниве формули:

$$k = 1 + 3.322 (\log_{10} n) \quad \text{или} \quad k \approx \sqrt{n}.$$



Распределба на честоти за интервални променливи

2. Распон или Ранг на податоците (R).

Разликата меѓу најголемата и најмалата вредност во примерокот.

3. Ширина на интервалите (w).

Интервалите треба да не се преклопуваат и вообичаено е интервалите да бидат со иста ширина. Значи ако сакаме k интервали ширината на интервалите, w , треба да биде

$$w \geq R / k.$$

За да се добијат крајни точки на интервалите и ширина кои ќе биде полесни за определување на интервалите и групирање на податоците, може долната граница на првиот интервал да биде помала од најмалиот податок, а горната граница на последниот интервал да биде поголема од најголемиот податок во примерокот.

На пример, ако најмалиот податок е 12, а најголемиот 57, првиот интервал може да биде со долна граница 10, а последниот со горна граница 60.



Распределба на честоти

Пример 2: Дадени се максимални дневни температури во 20 случајно избрани зимски денови:

-4, 2, -8, -5, -4, 3, -3, 8, 14, 0, 0, -11, -11, 3, 5, 6, 7, -3, 12, -3

- Прво ги сортираме податоците во растечки редослед:
-11, -11, -8, -5, -4, -4, -3, -3, -3, 0, 0, 2, 3, 3, 5, 6, 7, 8, 12, 14
- Го определуваме распонот: $14 - (-11) = 25$.
- Избираме поделба во 5 интервали, бидејќи $\sqrt{20} = 4.47$.
- Ја пресметуваме ширината: $25/5 = 5$ (по потреба се заокружува)
- Ги определуваме границите на интервалите $[-11, -6)$, $[-6, -1)$, $[-1, 4)$, $[4, 9)$, $[9, 14]$.
- Ги броиме набљудувањата од примерокот кои се во секој од интервалите



Распределба на честоти

Подредени податоци:

-11, -11, -8, -5, -4, -4, -3, -3, -3, 0, 0, 2, 3, 3, 5, 6, 7, 8, 12, 14

Интервал	Честота (f)	Релативна честота (f/n)	Релативна честота (во %)
$[-11, -6)$	3	0.15	15
$[-6, -1)$	6	0.30	30
$[-1, 4)$	5	0.25	25
$[4, 9)$	4	0.20	20
$[9, 14]$	2	0.10	10
	20	1	100



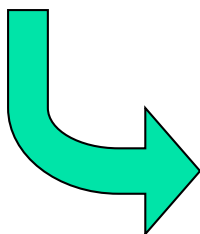
Хистограм

- Најчесто, распределбата на честотите графички се претставува со **хистограм**.
- **Крајните точки (или средината) на интервалите** се прикажани на **хоризонталната оска**.
- На **вертикалната оска** се **честотите, релативните честоти** или **процентите**.
- Над секој интервал се црта столб со висина која одговара на вредноста на честотите, релативните честоти или процентите на податоците во соодветниот интервал.



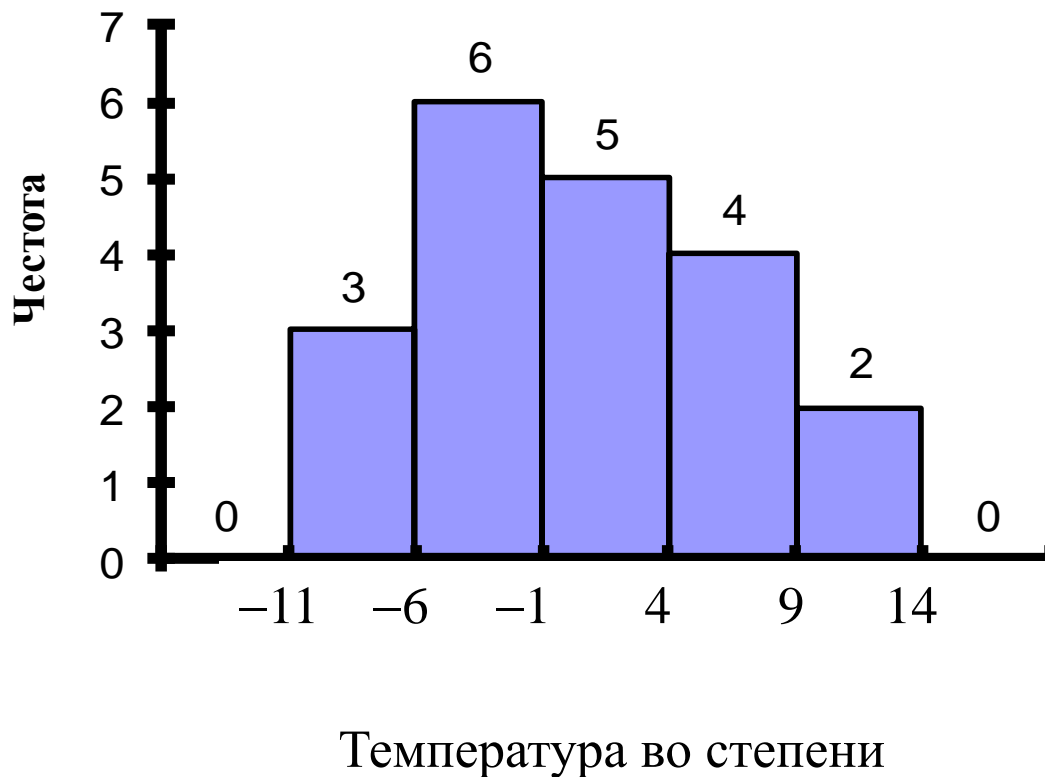
Хистограм за Пример 2

Интервал	Честота (f)
$[-11, -6)$	3
$[-6, -1)$	6
$[-1, 4)$	5
$[4, 9)$	4
$[9, 14]$	2
	20



(Нема растојание меѓу
столбовите)

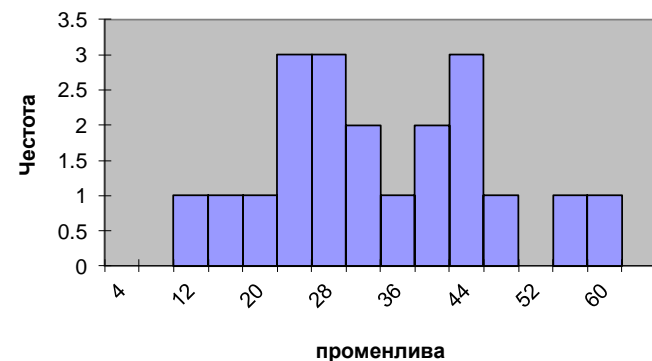
Хистограм: Дневни температури



Колку интервали да се изберат?

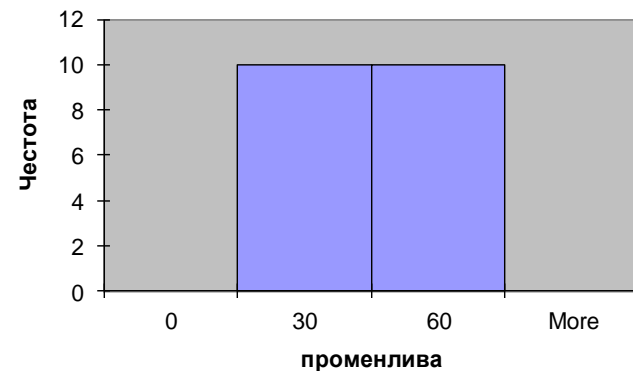
■ Многу (тесни интервали)

- Несоодветен приказ кој може да има и празни интервали.
- Слаба индикација за распределата на податоците.



■ Малку (широки интервали)

- Многу варирање групирано заедно.
- Може да се изгубат значајни однесувања на распределбата.





Пример 3

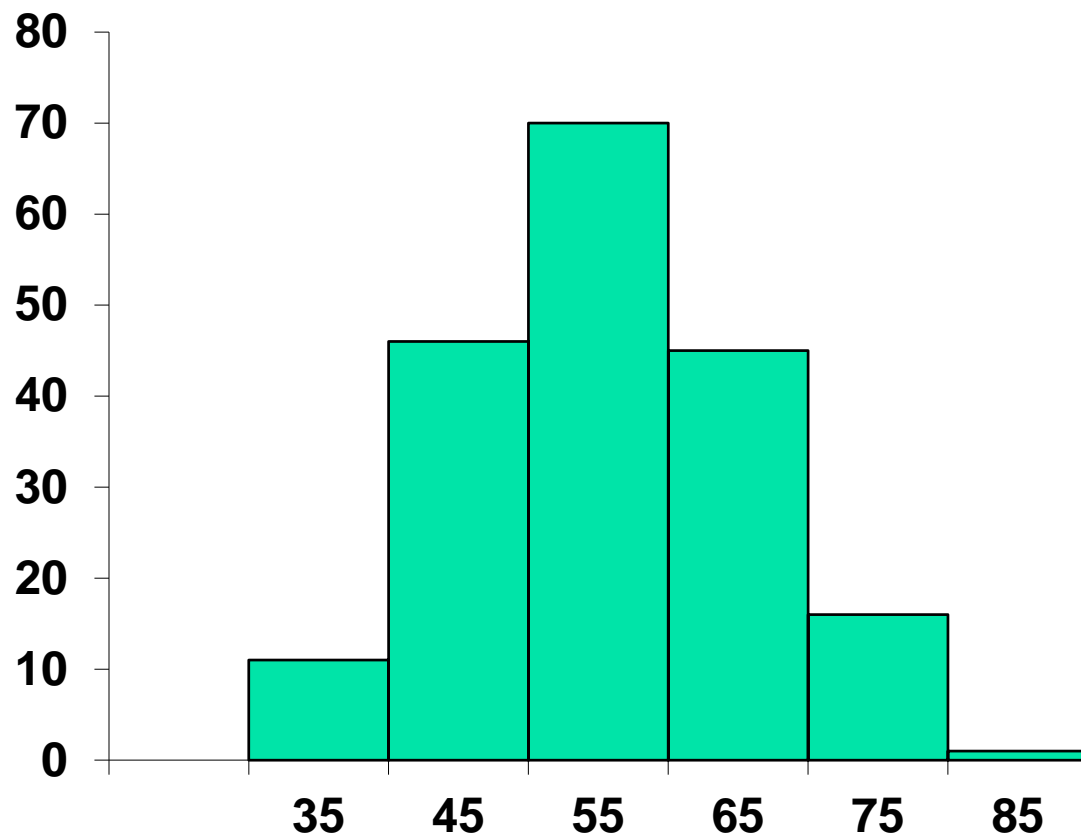
Дадени се податоци за возраста на 189 учесници на едно истражување.

Интервали	Честота
[30, 40)	11
[40, 50)	46
[50, 60)	70
[60, 70)	45
[70, 80)	16
[80, 90]	1
Вкупно	189



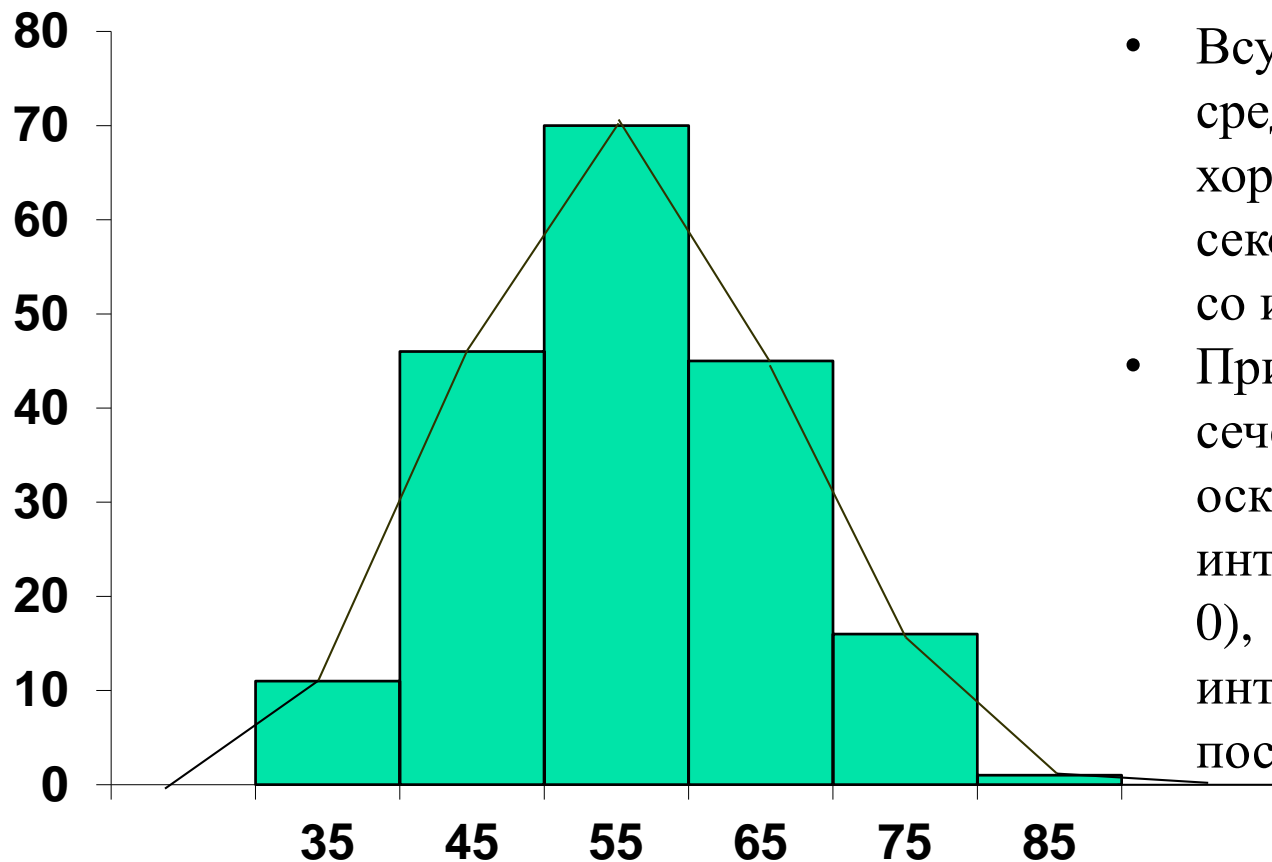
Претставување на податоците со хистограм

Интервали	Честота
[30, 40)	11
[40, 50)	46
[50, 60)	70
[60, 70)	45
[70, 80)	16
[80, 90]	1
Вкупно	189



Претставување на честотите со полигон на фреквенции

- Полигонот на фреквенции се добива со поврзување на точките со координати (средна точка на интервалот, фреквенција на интервалот) во секој од интервалите.



- Всушност, се поврзуваат средните точки на горните хоризонтални страни на секој од правоаголниците со искршена линија.
- Притоа, за да линијата ја сече хоризонталната оската, се земаат и празни интервали (со фреквенција 0), еден лево од првиот интервал и еден десно од последниот интервал.



Кумулативна честота

Кумулативна честота:

Се пресметува со собирање на последователните честоти (број на податоци во примерокот кои се наоѓаат во соодветниот интервал и интервалите пред него).

Кумулативна релативна честота:

Се пресметува со собирање на последователните релативни честоти.

Средината на интервалот:

Се собираат долната и горната граница на интервалот и се делат со 2.



За податоците од Пример 3

Интервал	Средина на интервал	Честота (f)	Кумулатив- на честота	Релативна честота $R.f = f/n$	Кумулативна релативна честота
[30, 40)	35	11	11	0.0582	0.0582
[40, 50)	45	46	57	0.2434	0.3016
[50, 60)	55	70	127	0.3704	0.6720
[60, 70)	65	45	172	0.2381	0.9101
[70, 80)	75	16	188	0.0847	0.9948
[80, 90]	85	1	189	0.0053	1
Вкупно		189		1	



Пример 4: Распределба на честоти

- Се подредуваат необработените податоци во растечки редослед: **12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58**
- Се определува распонот: **$58 - 12 = 46$**
- Се бира број на интервали: **$5 (\sqrt{20} \approx 5)$**
- Се определува ширината на интервалите: **$10 (46/5 = 9.2, \text{ се заокружува на } 10)$**
- Се определуваат границите на класите: **$(10 \cdot 5 = 50$ е вкупна ширина на 5те интервали, а распонот е 46 и затоа целиот интервал го прошируваме за 2 од лево и за 2 од десно)**
границите на интервалите се: **10, 20, 30, 40, 50, 60.**
- Се брои колку податоците од примерокот се во секој интервал.



Пример 4: Честоти, релативни честоти и проценти

Подредени податоци:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

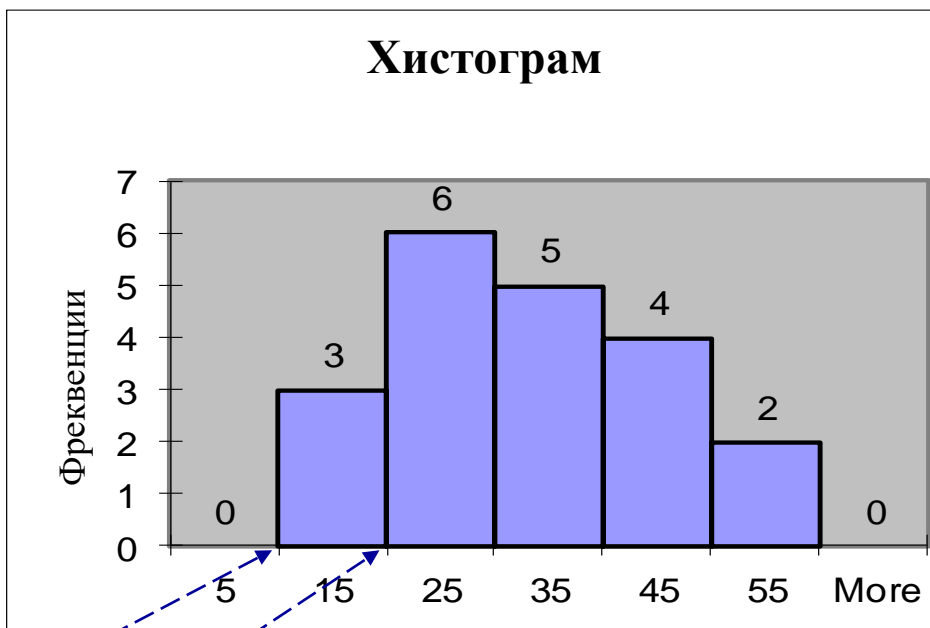
Интервал	Честота	Релативна честота	Релативна честота (во %)
[10, 20)	3	0.15	15
[20, 30)	6	0.30	30
[30, 40)	5	0.25	25
[40, 50)	4	0.20	20
[50, 60]	2	0.10	10
	20	1	100



Пример 4: Хистограм

Подредени податоци

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58



Граници на класите

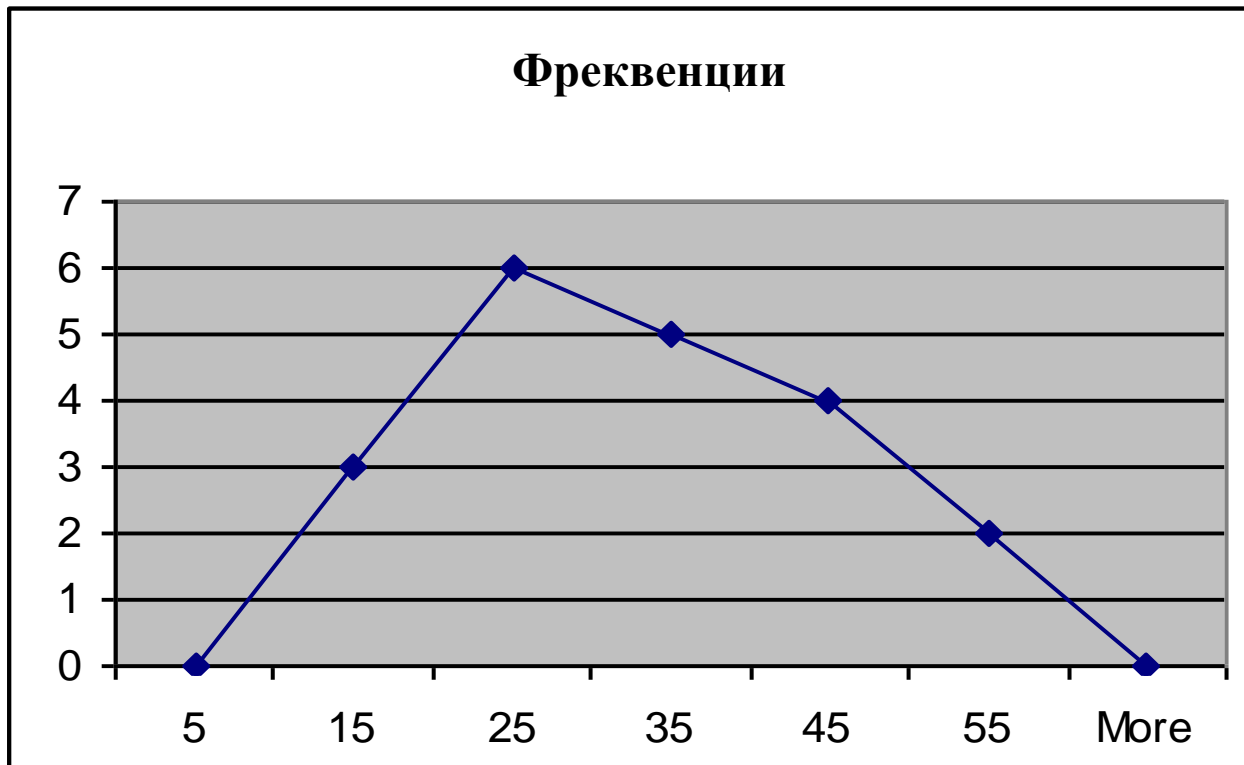
Средини на интервалите



Пример 4: Полигон на честоти

Подредени податоци:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58



Средини на интервалите



Пример 4: Кумулативни честоти

Податоците во подредена листа:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Горна граница	Кумулативна честота	Кумулативна честота (во %)
10	0	0
20	3	15
30	9	45
40	14	70
50	18	90
60	20	100

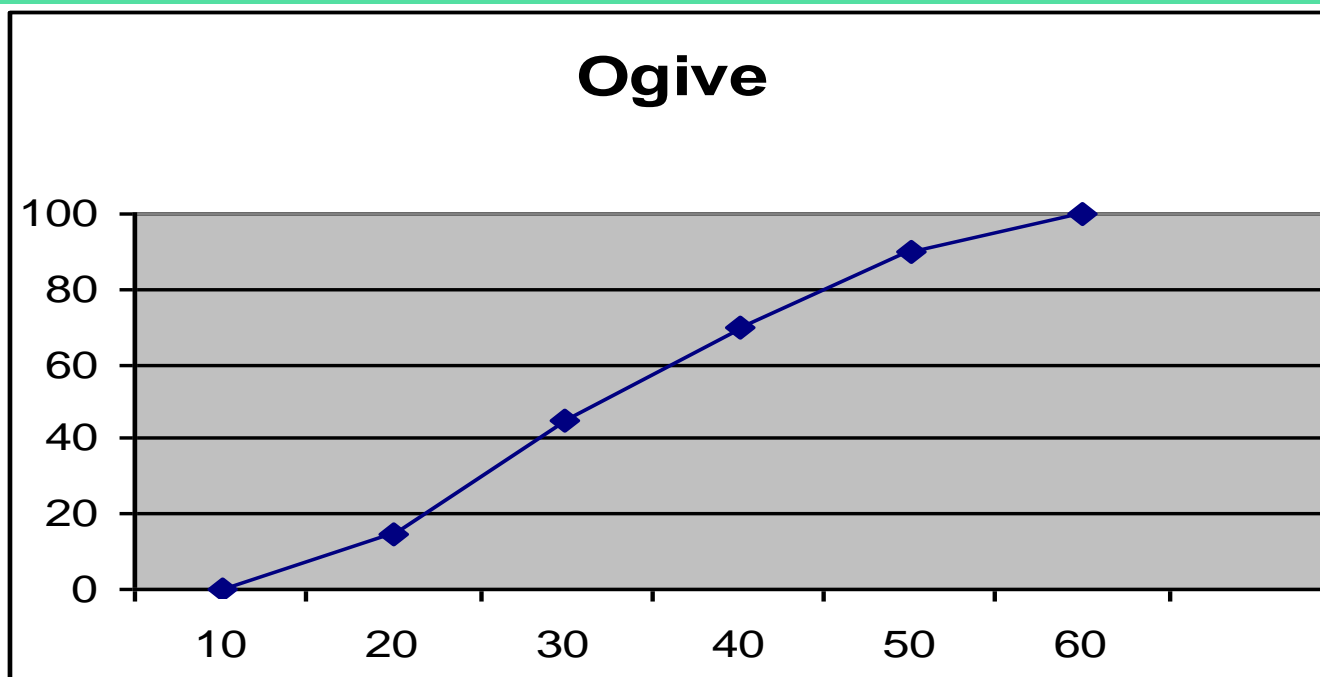


Пример 4: Ogive

(полигон на кумулативни честоти во %)

Податоци во подредена листа:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58



Граници на интервалите (*Не средини*)



Пример 5: резултати од колоквиум за 165 студенти

10`	10	7	2	17	7	17	5	8	5	26	59	20	0	8
34	0	5	15	13	7	14	35	31	15	2	17	15	19	0
35	70	33	32	20	73	29	34	43	34	64	46	37	34	8
56	8	41	25	70	44	71	35	39	38	7	29	11	39	49
26	23	0	44	35	33	9	15	2	1	22	25	47	26	27
49	2	32	12	23	35	43	40	34	67	65	71	48	40	21
30	28	44	78	46	8	17	51	2	84	40	10	40	16	5
61	55	42	35	31	44	11	0	49	34	34	2	35	16	32
7	20	24	26	40	11	13	4	46	23	44	10	13	17	8
6	44	17	34	26	35	7	66	14	32	35	4	25	35	53
24	8	11	80	12	53	49	74	19	13	35	41	1	26	1

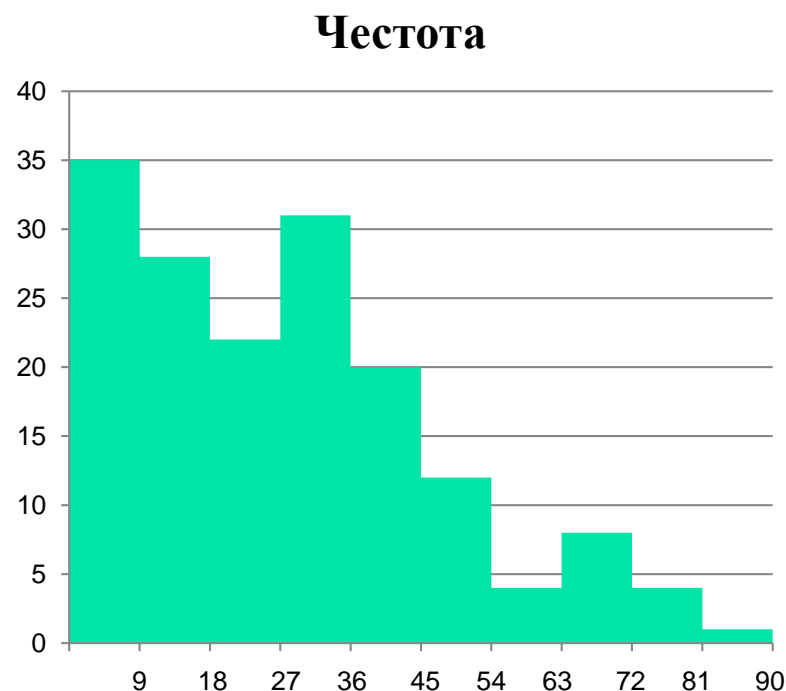


Пример 5: Определување број на интервали

- Број на податоци (обем на примеркот) $n = 165$.
- $k = 1 + 3,22 * \log_{10} 165 = 1 + 3,22 * 2,217 = 8,14$
или $k = \sqrt{165} \approx 12$. Може да земеме $k = 10$.
- Најголемиот податок, $\max = 84$.
- Најмалиот податок, $\min = 0$.
- Распонот $R = 84 - 0 = 84$.
- За да биде ширината на интервалите цел број интервалите ќе ги формираме од 0 до 90.
- Ширина на секој интервал е $90/10 = 9$.

Пример 5: Табела и хистограм на честотите

интервал	честота	рел.чес	рел. кумул. чес.
[0,9)	35	0,21	0,21
[9,18)	28	0,17	0,38
[18,27)	22	0,13	0,52
[27,36)	31	0,19	0,71
[36,45)	20	0,12	0,83
[45,54)	12	0,07	0,9
[54,63)	4	0,02	0,92
[63,72)	8	0,05	0,97
[72,81)	4	0,02	0,99
[81,90]	1	0,01	1
Вкупно	165	1	



Табели со распределба на честоти за дискретни квантитативни податоци со мало множество вредности



Ако податоците кои се разгледуваат се дискретни квантитативни и множеството податоци содржи мал број вредности, тогаш нема потреба податоците да се делат во интервали. Во овој случај, податоците се подредуваат во неопаѓачки редослед и на секоја вредност се доделува соодветната честота.

Пример 6:

- Земен е примерок од 16 студенти од ФИНКИ и добиени се податоци за бројот на положени испити од претходниот семестар:

3,5,2,4,0,1,3,5,2,3,2,3,3,2,4,1

За да се направи табела на честота:

- Прво се подредуваат вредностите од најмала до најголема:

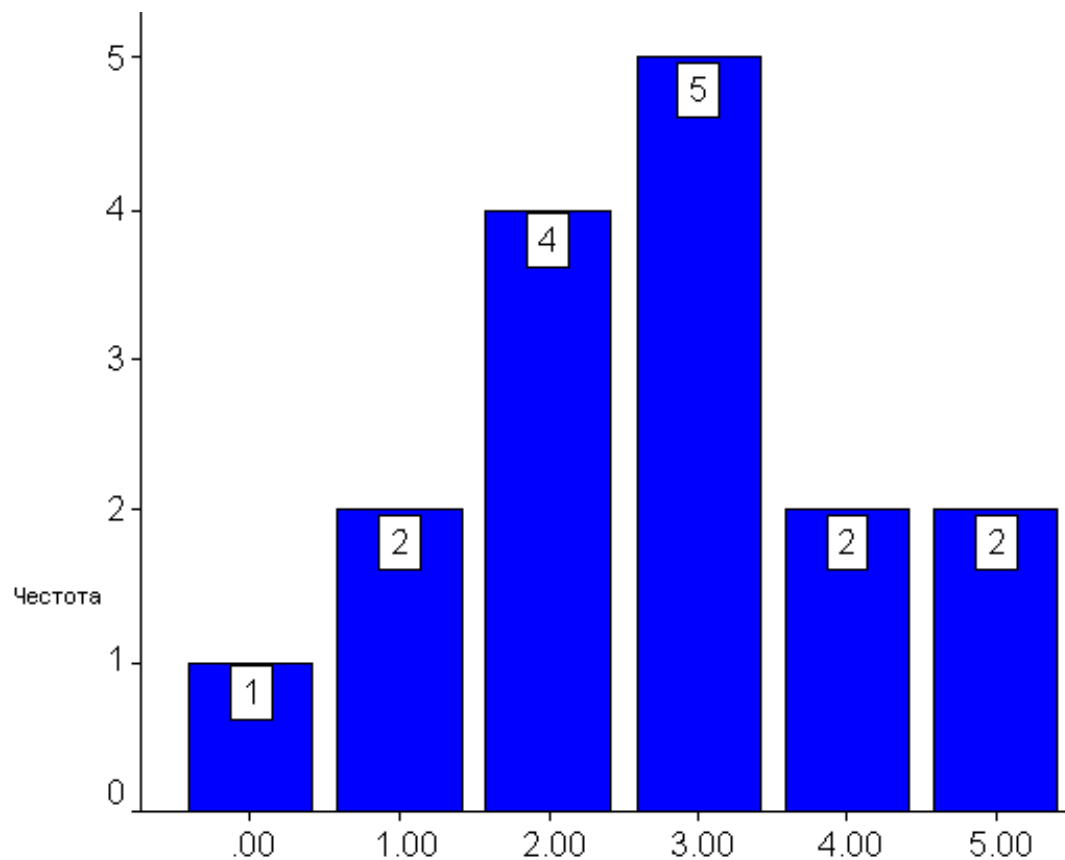
0,1,1,2,2,2,2,3,3,3,3,3,4,4,5,5

- Потоа, се определува фреквенцијата на секоја вредност.

Бр. на пол. испити	Честота	Релативна честота
0	1	0.0625
1	2	0.125
2	4	0.25
3	5	0.3125
4	2	0.125
5	2	0.125
Вкупно	16	1

Претставување на честотите со столбест дијаграм

Честотите може да ги претставиме со столбест дијаграм





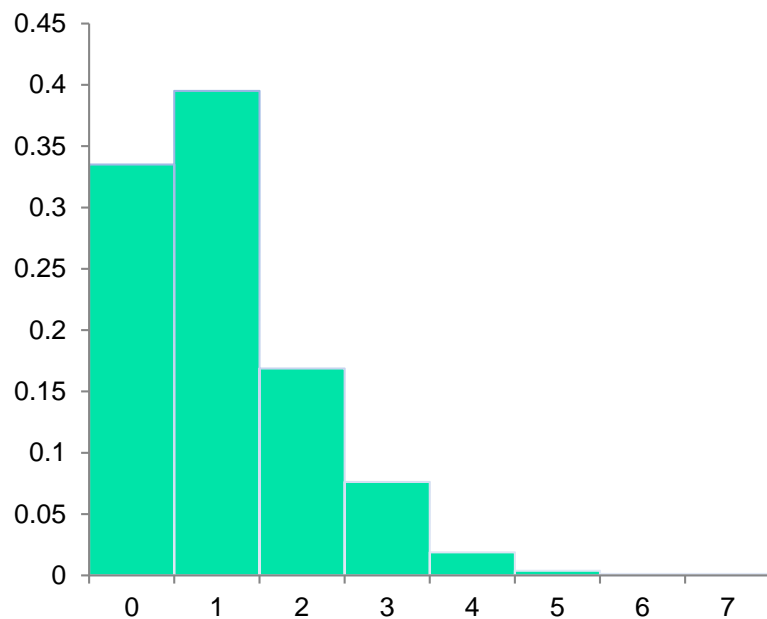
Пример 7 – број на реализирани лабораториски вежби од 800 студенти

Реализирани лаб. вежби	Честота (f_i)	Рел. честота (f_i / n)	Кумулативна рел. честота
0	268	0,335	0,335
1	316	0,395	0,730
2	135	0,169	0,899
3	61	0,076	0,975
4	15	0,019	0,994
5	3	0,004	0,998
6	1	0,001	0,999
7	1	0,001	1,000
Вкупно	800	1,000	

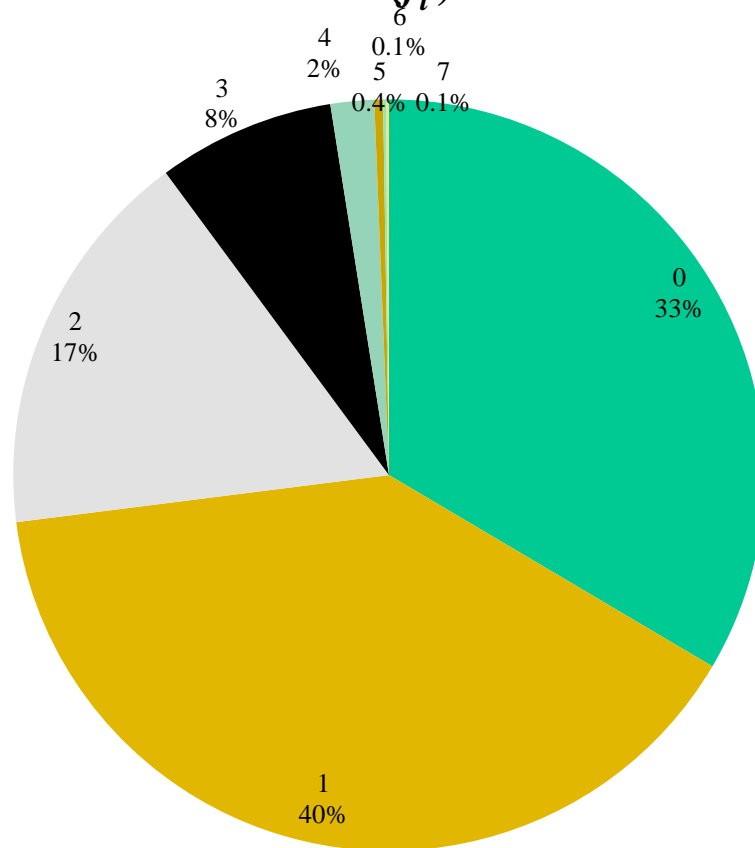


Пример 7: Графички прикази

Рел. честота (f_i/n)



Честота (f_i) во %



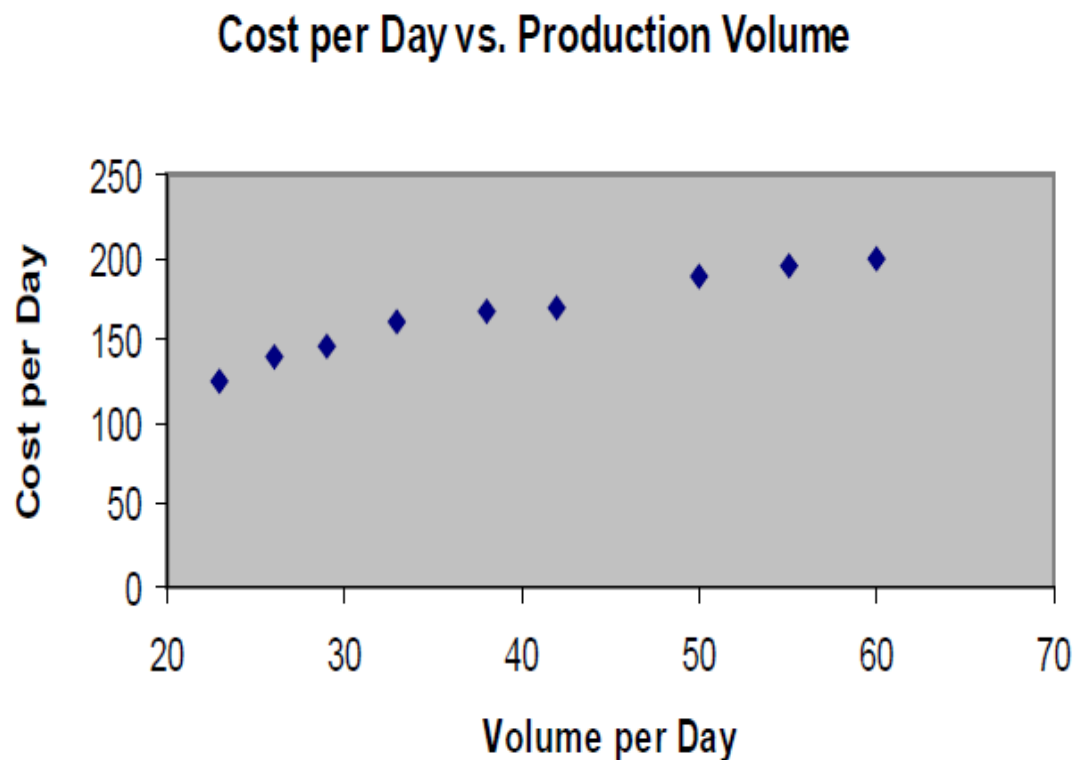


Графички прикази за дводимензионални нумерички податоци (Дијаграм на расејување)

- Некогаш за единките во примерокот (популацијата) се разгледуваат две карактеристики и се испитува дали постои некаква врска помеѓу двете карактеристики.
- Тогаш, за секоја единка во примерокот се добива пар вредности (x_i, y_i) - вредностите на двете карактеристики.
- Доколку добиените парови вредности (x_i, y_i) , $i = 1, 2, \dots, n$, се претстават како точки во координатен систем, тогаш се добива **дијаграм на расејување**.

Пример 8: Дијаграм на расејување

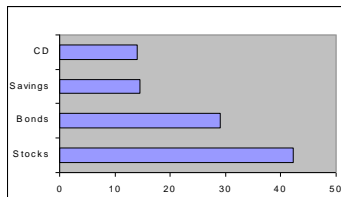
Дневен обем на производство	Цена
23	125
26	140
29	146
33	160
38	167
42	170
50	188
55	195
60	200



Табели и графици за категориски обележја

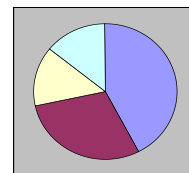
Категориски податоци

Табелирање на податоците



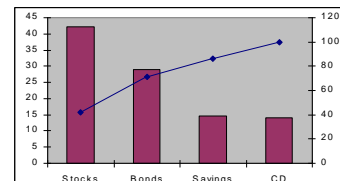
Столбести дијаграми

Графички прикази



Пити

Парето дијаграми





Пример 9:

Категориски податоци сумирани во табела

- Во табелите за секоја од категориите се запишува фреквенцијата на податоци во примерокот од таа категорија, т.е., бројот или процентот на набљудувања во секоја од категориите. На овој начин може да ги видиме разликите помеѓу категориите.

Преферирано банкарство?	Процент
АТМ	16%
Телефон	2%
Преку служби во филијалите	17%
Лично во филијалата	41%
Интернет	24%




Пример 10: Табела на честота

- Број на продадени лаптопи по градови

Град	Број на продадени лаптопи
Струмица	1052
Битола	2245
Велес	340
Охрид	552
Скопје	4630

(променливата е
категориска)





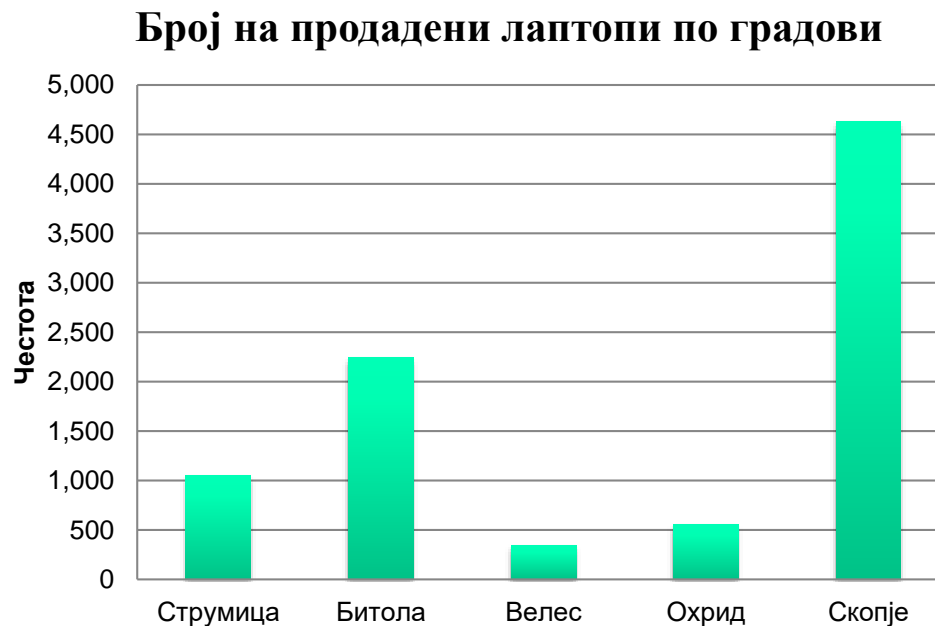
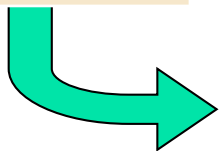
Столбести графици и пити

- Најчесто користени прикази за квалитативни (категориски) податоци се **столбести графици** и **пити**.
- Висината на столпчињата или големината на деловите од питата ја покажуваат честотата или процентот на податоци во секоја категорија.



Пример 10: Столбест дијаграм

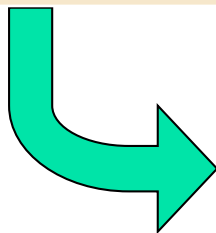
Град	Број на продадени лаптопи
Струмица	1052
Битола	2245
Велес	340
Охрид	552
Скопје	4630





Пример 10: Пита

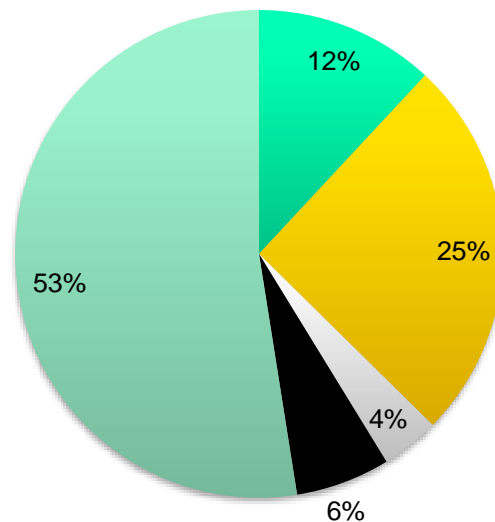
Град	Број на продадени лаптопи	% (од вкупно)
Струмица	1052	11.93
Битола	2245	25.46
Велес	340	3.85
Охрид	552	6.26
Скопје	4630	52.50



Во питата
процентите се
заокружени

Лаптопи по градови

■ Струмица ■ Битола ■ Велес ■ Охрид ■ Скопје





Пример 11: (Портфолио на инвеститор)

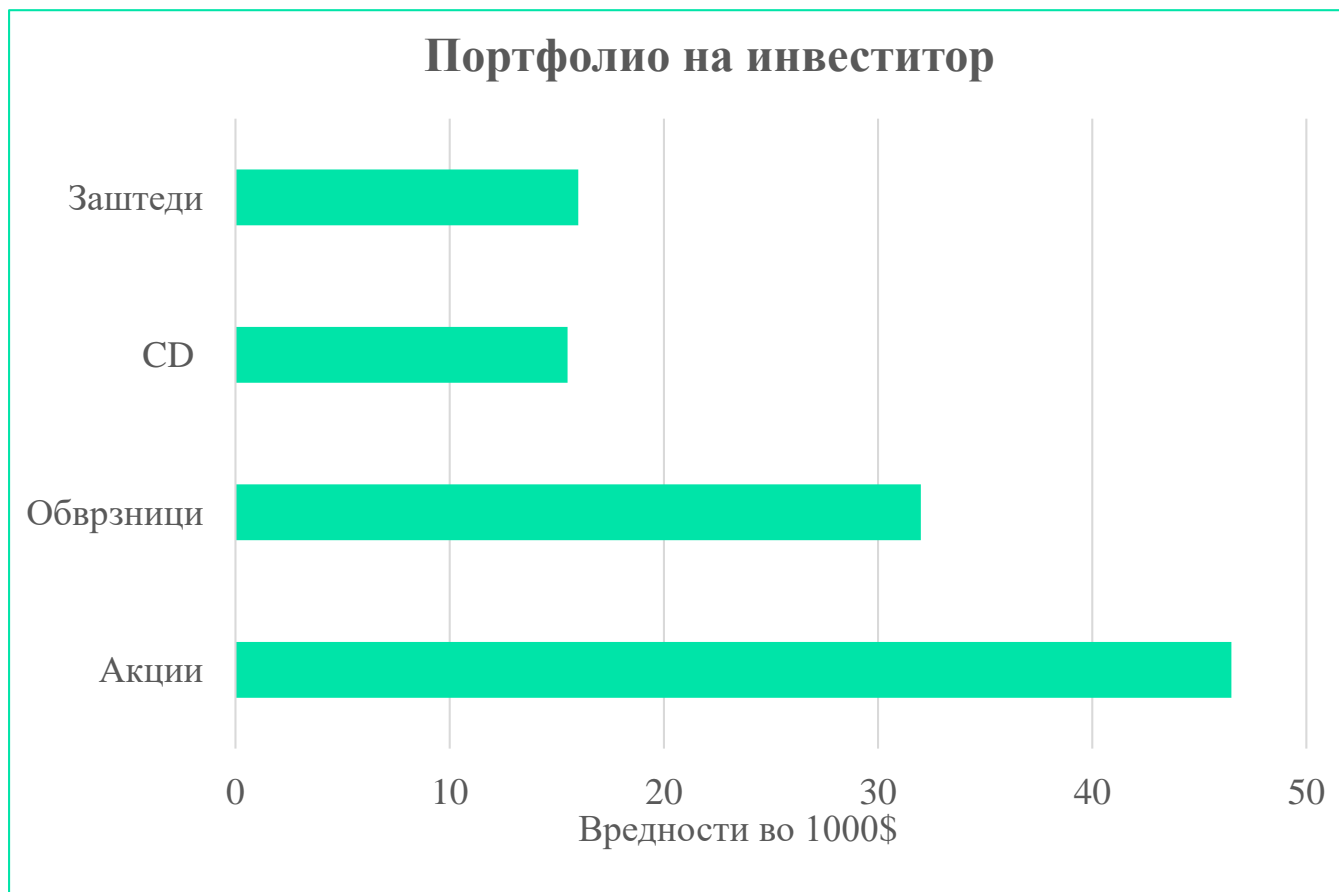
Категорија на инвестиции	Сума	% (од вкупно)
Акции	46.5	42.27
Обврзници	32.0	29.09
CD (сертификат за депозит)	15.5	14.09
Заштеди	16.0	14.55
	110	100



Податоците се категориски

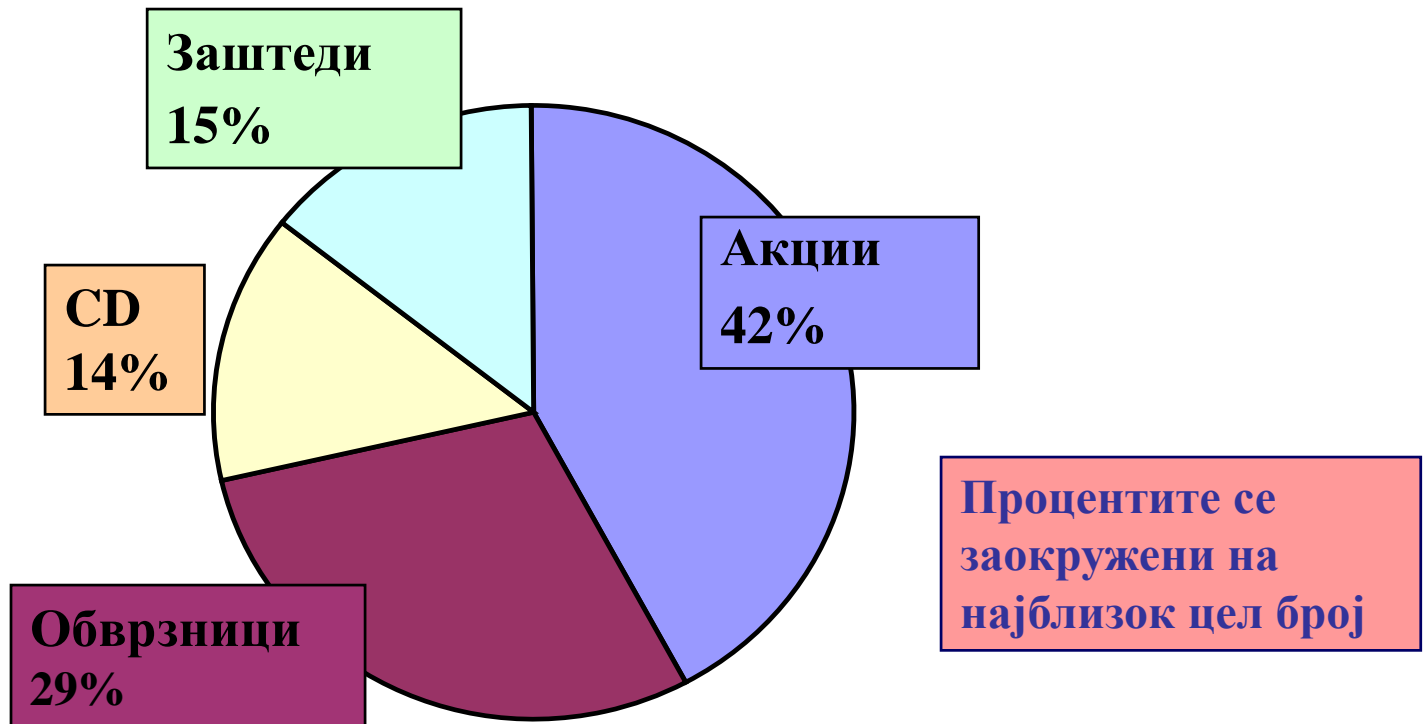


Пример 11: Столбест дијаграм





Пример 11: Претставување на процентите со пита





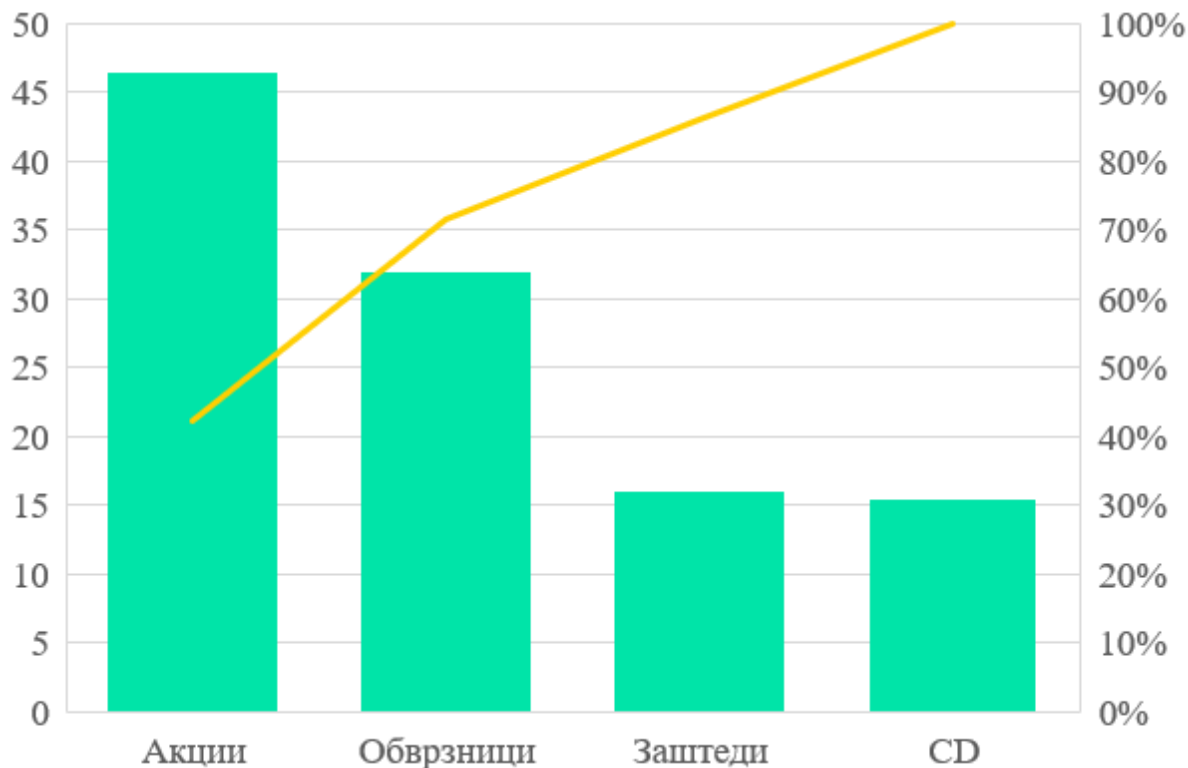
Pareto-дијаграм

- Трет вид график на квалитативни податоци е Pareto-дијаграмот, кој претставува посебна примена на столбест дијаграм.
- Pareto-дијаграмот е вид на график кој содржи и столпчиња и линиски дијаграм, каде што индивидуалните вредности се претставени во опаѓачки редослед со столпчиња, а со искршена линија е претставена кумулативната релативна фреквенција.
- На левата вертикална оска е претставена честотата на појавување, но може да претставува и цена или некоја друга единица мерка. На десната вертикална оска е претставен кумулативниот процент на вкупниот број на настани, кумулативната цена или кумулативната вредност на одредената единица мерка.



Пример 11: Парето дијаграм

Оската за столбовите е висина на инвестиции во секоја категорија



Оската за линискиот граф ги прикажува кумулативните инвестиции во %

Категорија на инвестиции	Сума	% (од вкупно)	Кумулативни %
Акции	46.5	42.27	42.27
Обврзници	32.0	29.09	71.36
Заштеди	16	14.55	85.91
CD	15.5	14.09	100
	110	100	



Табели и графици за дводимензионални категоришки обележја

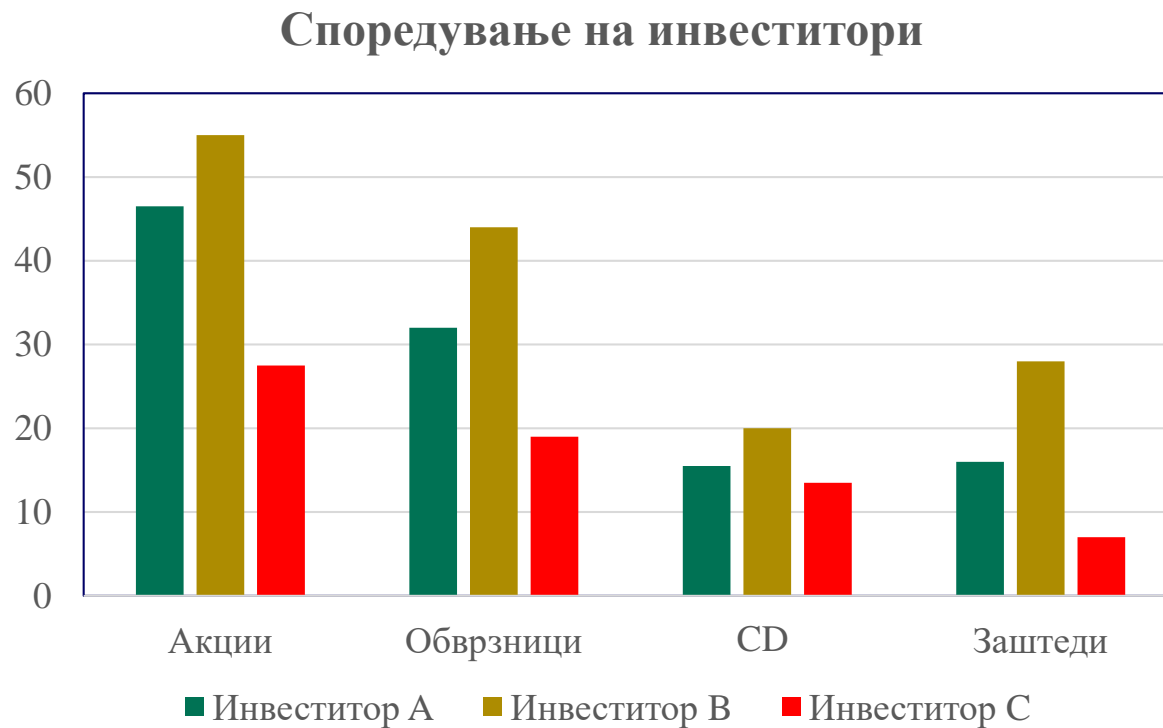
Табели на контингенција.

Пример 12. Инвестиции во илјади долари по инвеститор.

<div>Инвеститор</div> <div>Инвестиција</div>	Инвеститор А	Инвеститор В	Инвеститор С	Вкупно
Акции	46,5	55	27,5	129
Обврзници	32,0	44	19	95
CD	15,5	20	13,5	49
Заштеди	16,0	28	7	51
Вкупно	110	147	67	324

Табели и графици за дводимензионални категоришки податоци

Пример 12: Паралелни столбести дијаграми





Грешки во презентирање на податоците

- Користење на несоодветни графички прикази.
- Несоодветно споредувањето на податоците помеѓу групите.
- Компресирање на вертикалната оска.
- Нумерацијата на вертикална оска е таква што нулата не е во координатниот почеток.



Несоодветни графици



**Несоодветно
претставување**

Минимална плата



1960: \$1.00



1970: \$1.60



1980: \$3.10

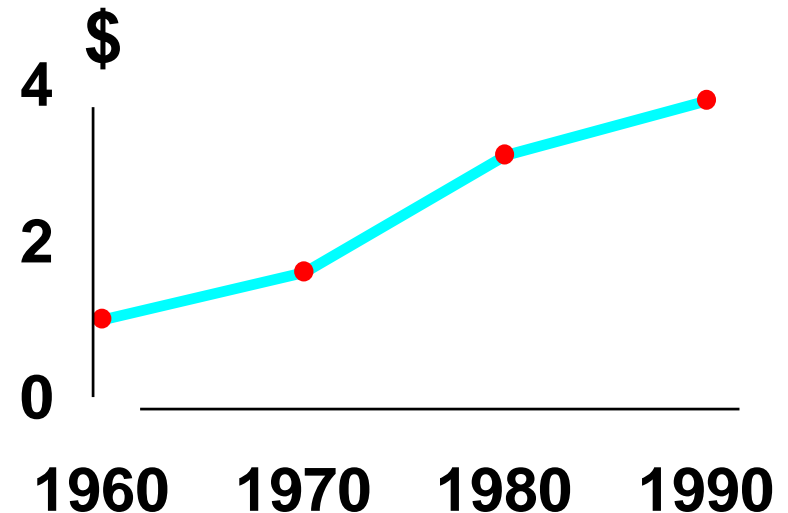


1990: \$3.80



Добро претставување

Минимална плата

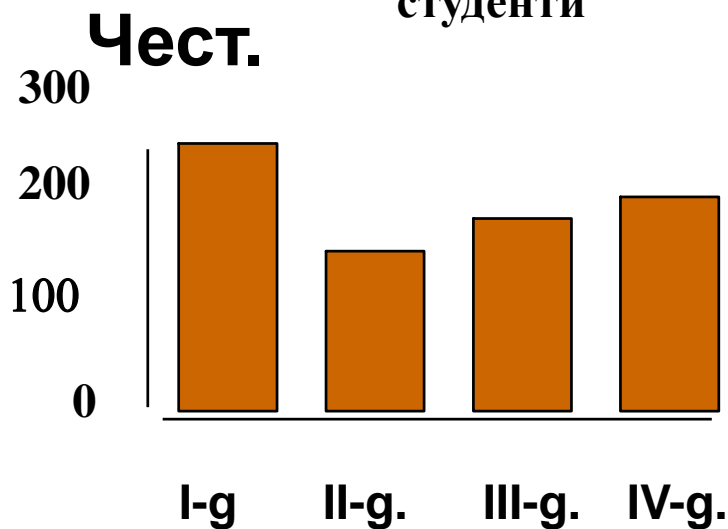


Несоодветно споредување на податоците помеѓу групите



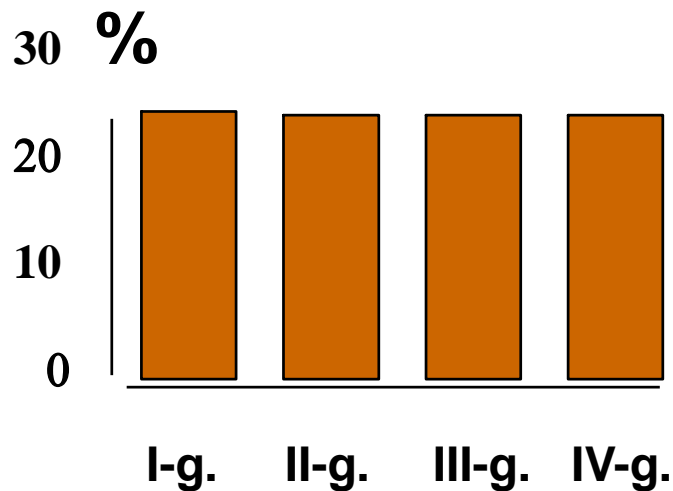
Лошо претставување

Добиени 10-ки од
студенти



Добар приказ

Добиени 10-ки од
студенти



Несоодветно е да се споредуваат бројот на 10-ки во различни години кога не се знае по колку студенти има во секоја година. Затоа е подобро да се споредуваат податоците претставени во проценти.



Компресиране на вертикална оска

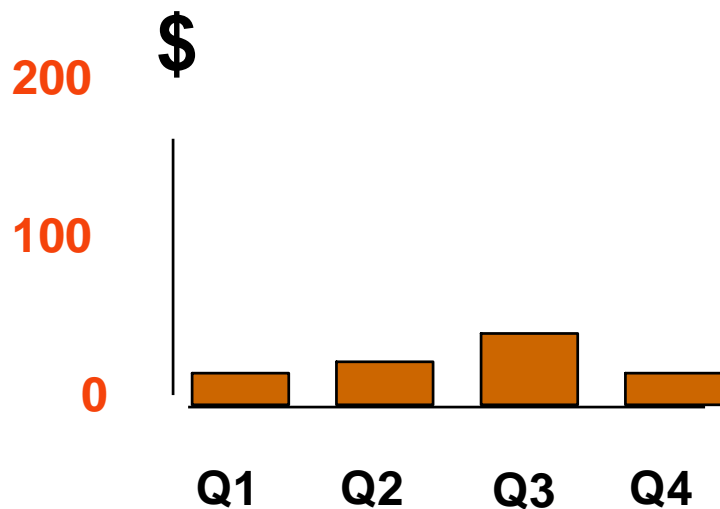


Лош приказ

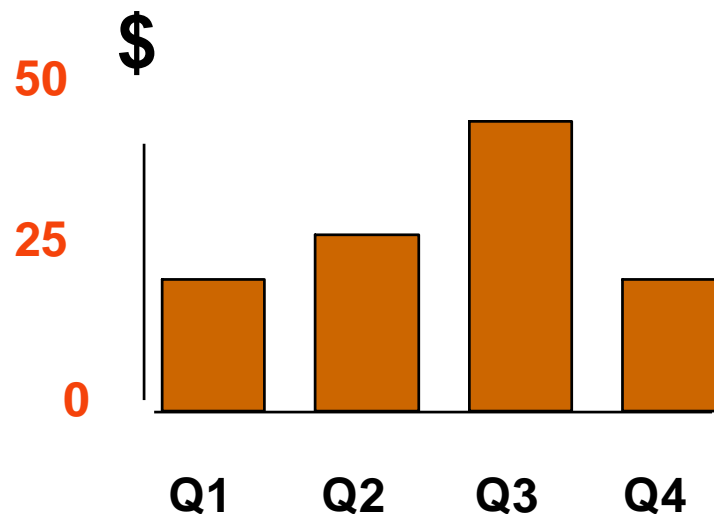


Добар приказ

Квартални плати



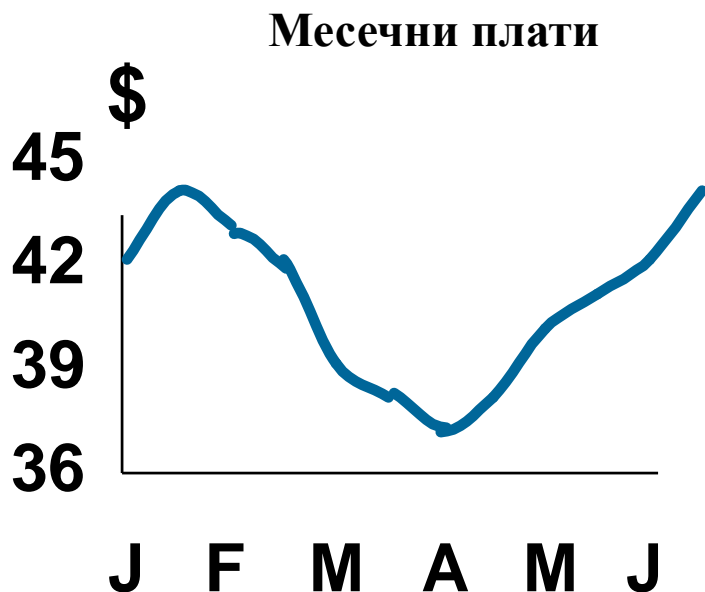
Квартални плати



Нумерацијата на вертикална оска е таква
што нулата не е во координатниот почеток

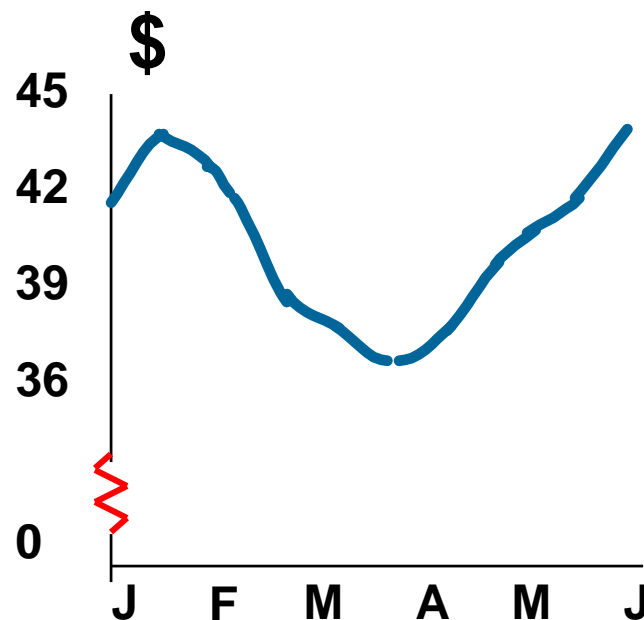


Лош приказ



Добар приказ

Месечни плати



Графички приказ за првите 6 месеци