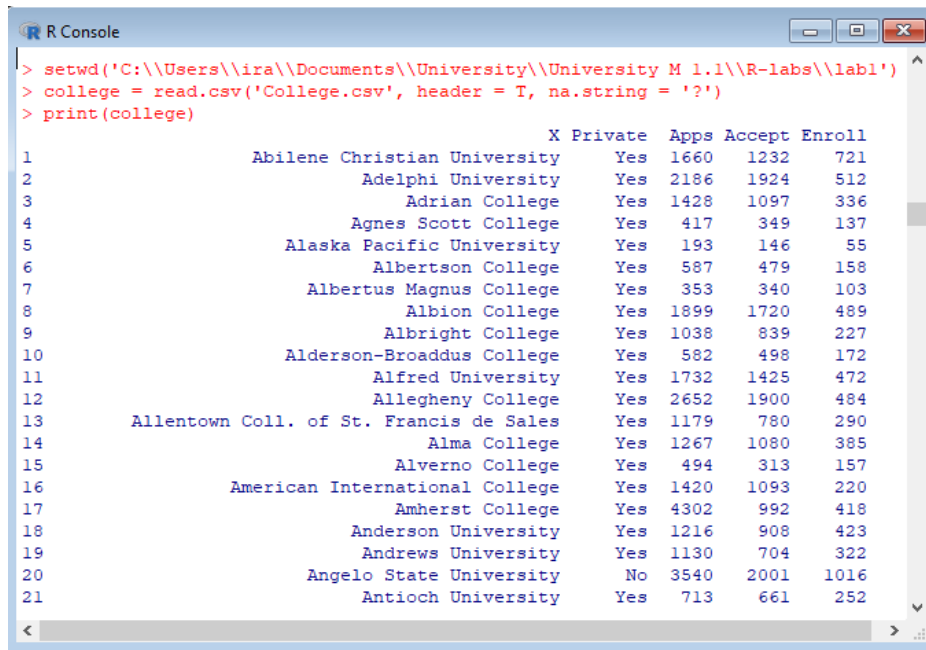


Звіт
до індивідуального завдання №1
з предмету Моделі статистичного навчання

Роботу виконала:
Мерцало Ірина Ігорівна,
студентка групи ПМІМ-11

Звдання 1

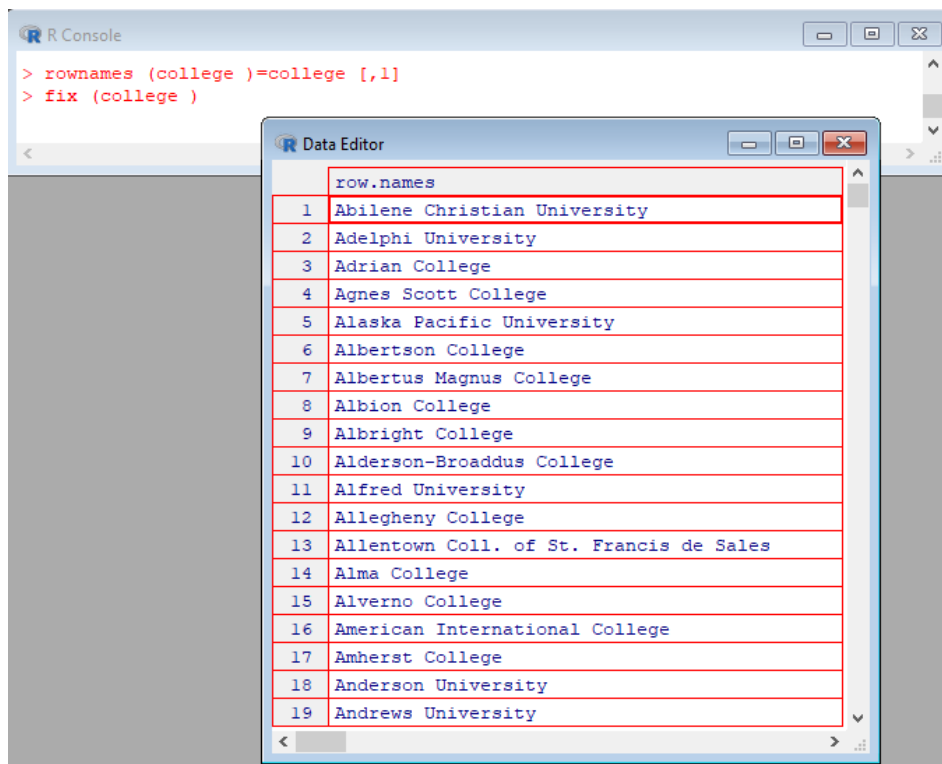
1.1 Для опрацювання даних зчитала файл 'College.csv' функцією `read.csv()` та викликала її:



```
> setwd('C:\\Users\\lira\\Documents\\University\\University M 1.1\\R-labs\\lab1')
> college = read.csv('College.csv', header = T, na.string = '?')
> print(college)
```

		X Private	Apps	Accept	Enroll
1	Abilene Christian University	Yes	1660	1232	721
2	Adelphi University	Yes	2186	1924	512
3	Adrian College	Yes	1428	1097	336
4	Agnes Scott College	Yes	417	349	137
5	Alaska Pacific University	Yes	193	146	55
6	Albertson College	Yes	587	479	158
7	Albertus Magnus College	Yes	353	340	103
8	Albion College	Yes	1899	1720	489
9	Albright College	Yes	1038	839	227
10	Alderson-Broadus College	Yes	582	498	172
11	Alfred University	Yes	1732	1425	472
12	Allegheny College	Yes	2652	1900	484
13	Allentown Coll. of St. Francis de Sales	Yes	1179	780	290
14	Alma College	Yes	1267	1080	385
15	Alverno College	Yes	494	313	157
16	American International College	Yes	1420	1093	220
17	Amherst College	Yes	4302	992	418
18	Anderson University	Yes	1216	908	423
19	Andrews University	Yes	1130	704	322
20	Angelo State University	No	3540	2001	1016
21	Antioch University	Yes	713	661	252

1.2 Переглянула дані з допомогою функції `fix()`. Ввівши наступні команди, назвам рядків присвоїлись назви університетів, які були у першому стовпці:.

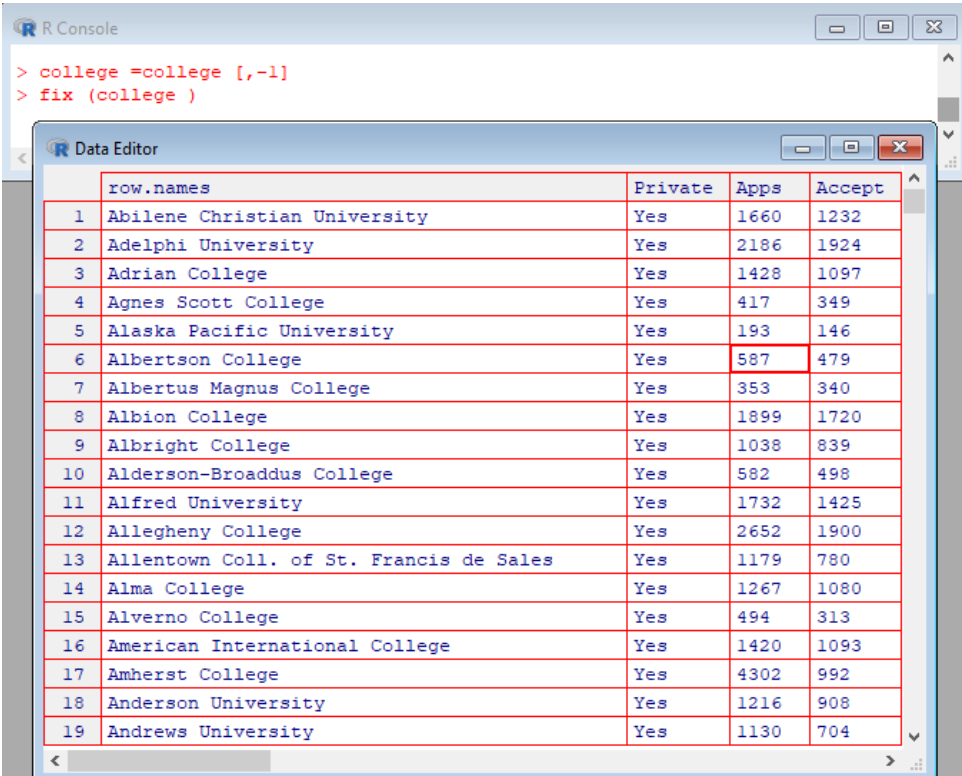


```
> rownames (college) =college [,1]
> fix (college )
```

row.names
1 Abilene Christian University
2 Adelphi University
3 Adrian College
4 Agnes Scott College
5 Alaska Pacific University
6 Albertson College
7 Albertus Magnus College
8 Albion College
9 Albright College
10 Alderson-Broadus College
11 Alfred University
12 Allegheny College
13 Allentown Coll. of St. Francis de Sales
14 Alma College
15 Alverno College
16 American International College
17 Amherst College
18 Anderson University
19 Andrews University

Наступними командами я отримала таблицю без першого стовпця, але з

доданими назвами рядків:



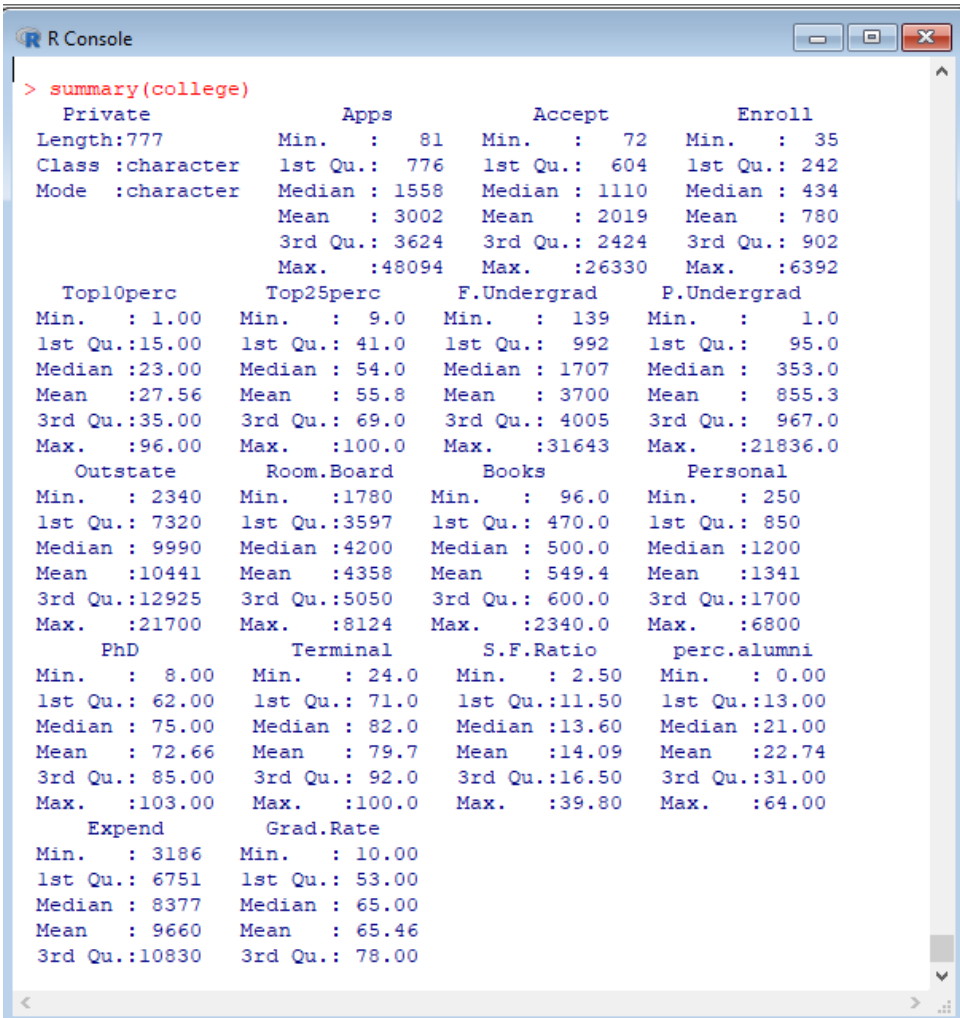
The image shows an R Console window with the following commands:

```
> college = college[, -1]
> fix(college)
```

Below the console is the 'Data Editor' window displaying the 'college' dataset. The first column, 'row.names', lists 19 college names. The other columns are 'Private', 'Apps', and 'Accept'.

row.names	Private	Apps	Accept
1 Abilene Christian University	Yes	1660	1232
2 Adelphi University	Yes	2186	1924
3 Adrian College	Yes	1428	1097
4 Agnes Scott College	Yes	417	349
5 Alaska Pacific University	Yes	193	146
6 Albertson College	Yes	587	479
7 Albertus Magnus College	Yes	353	340
8 Albion College	Yes	1899	1720
9 Albright College	Yes	1038	839
10 Alderson-Broaddus College	Yes	582	498
11 Alfred University	Yes	1732	1425
12 Allegheny College	Yes	2652	1900
13 Allentown Coll. of St. Francis de Sales	Yes	1179	780
14 Alma College	Yes	1267	1080
15 Alverno College	Yes	494	313
16 American International College	Yes	1420	1093
17 Amherst College	Yes	4302	992
18 Anderson University	Yes	1216	908
19 Andrews University	Yes	1130	704

1.3.1. Використавши функцію `summary()` отримала підсумок відносно даних:



The image shows an R Console window with the following command:

```
> summary(college)
```

The output is a summary of the 'college' dataset, showing various statistics for different variables.

Private	Apps	Accept	Enroll
Length:777	Min. : 81	Min. : 72	Min. : 35
Class :character	1st Qu.: 776	1st Qu.: 604	1st Qu.: 242
Mode :character	Median : 1558	Median : 1110	Median : 434
	Mean : 3002	Mean : 2019	Mean : 780
	3rd Qu.: 3624	3rd Qu.: 2424	3rd Qu.: 902
	Max. : 48094	Max. : 26330	Max. : 6392

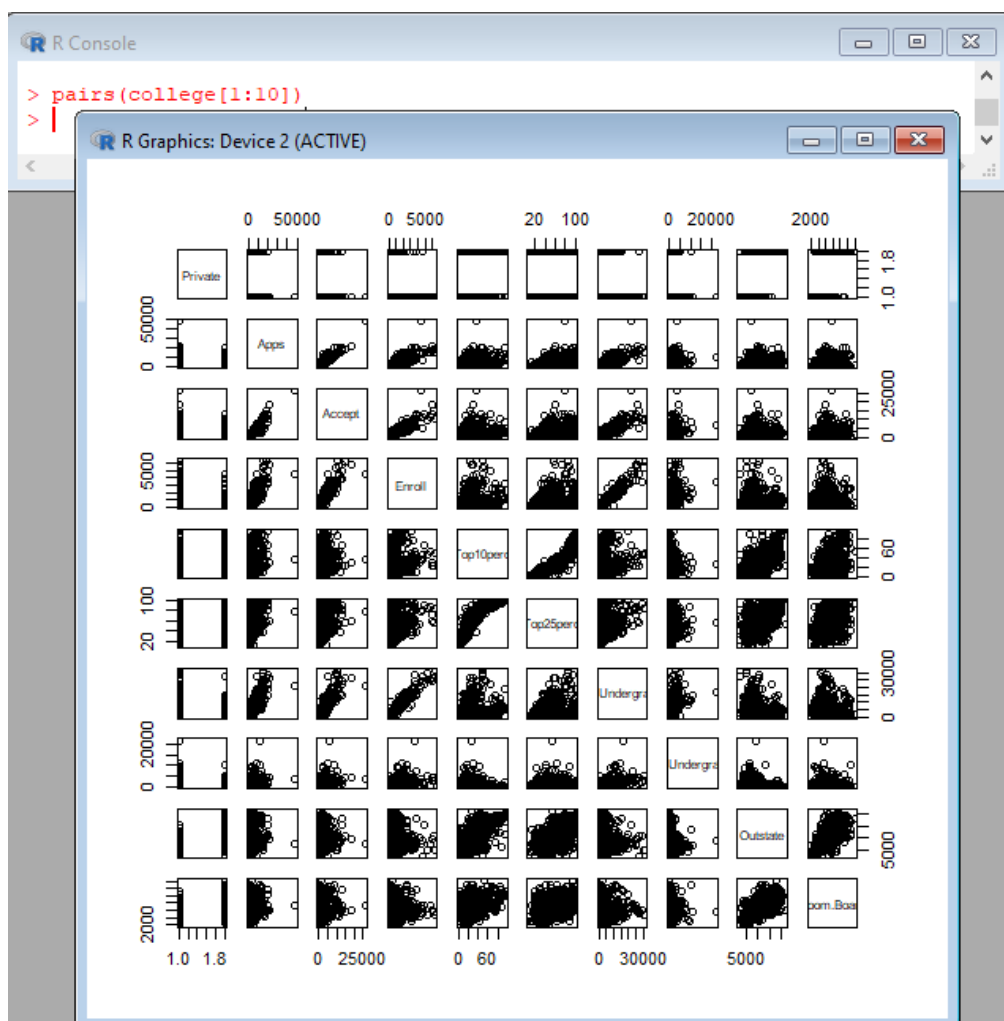
Top10perc	Top25perc	F.Undergrad	P.Undergrad
Min. : 1.00	Min. : 9.0	Min. : 139	Min. : 1.0
1st Qu.:15.00	1st Qu.: 41.0	1st Qu.: 992	1st Qu.: 95.0
Median :23.00	Median : 54.0	Median : 1707	Median : 353.0
Mean :27.56	Mean : 55.8	Mean : 3700	Mean : 855.3
3rd Qu.:35.00	3rd Qu.: 69.0	3rd Qu.: 4005	3rd Qu.: 967.0
Max. :96.00	Max. :100.0	Max. :31643	Max. :21836.0

Outstate	Room.Board	Books	Personal
Min. : 2340	Min. :1780	Min. : 96.0	Min. : 250
1st Qu.: 7320	1st Qu.:3597	1st Qu.: 470.0	1st Qu.: 850
Median : 9990	Median :4200	Median : 500.0	Median :1200
Mean :10441	Mean :4358	Mean : 549.4	Mean :1341
3rd Qu.:12925	3rd Qu.:5050	3rd Qu.: 600.0	3rd Qu.:1700
Max. :21700	Max. :8124	Max. :2340.0	Max. :6800

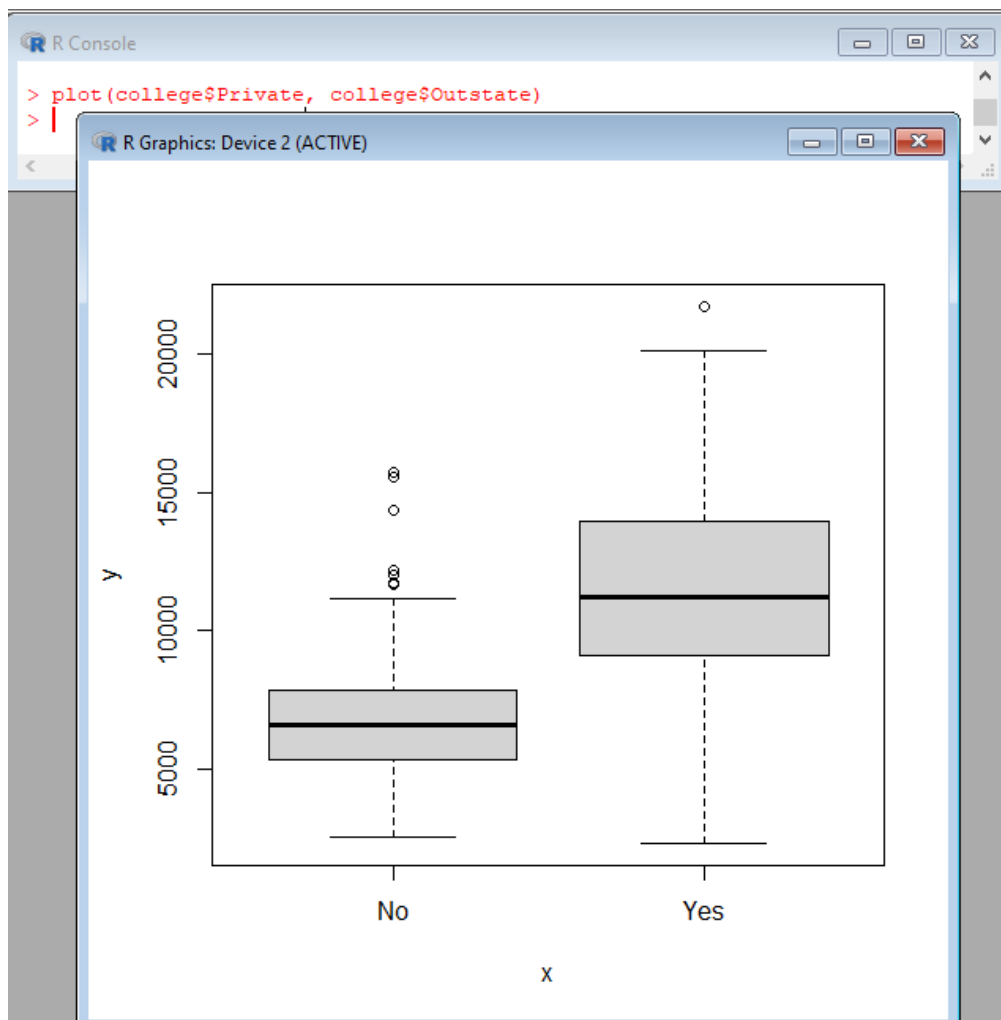
PhD	Terminal	S.F.Ratio	perc.alumni
Min. : 8.00	Min. : 24.0	Min. : 2.50	Min. : 0.00
1st Qu.: 62.00	1st Qu.: 71.0	1st Qu.:11.50	1st Qu.:13.00
Median : 75.00	Median : 82.0	Median :13.60	Median :21.00
Mean : 72.66	Mean : 79.7	Mean :14.09	Mean :22.74
3rd Qu.: 85.00	3rd Qu.: 92.0	3rd Qu.:16.50	3rd Qu.:31.00
Max. :103.00	Max. :100.0	Max. :39.80	Max. :64.00

Expend	Grad.Rate
Min. : 3186	Min. : 10.00
1st Qu.: 6751	1st Qu.: 53.00
Median : 8377	Median : 65.00
Mean : 9660	Mean : 65.46
3rd Qu.:10830	3rd Qu.: 78.00

1.3.2. Використала `pairs()` для побудови матриці графіків перших 10 стовпців даних:



1.3.3. Використала plot() для побудови діаграми Outstate vs Private

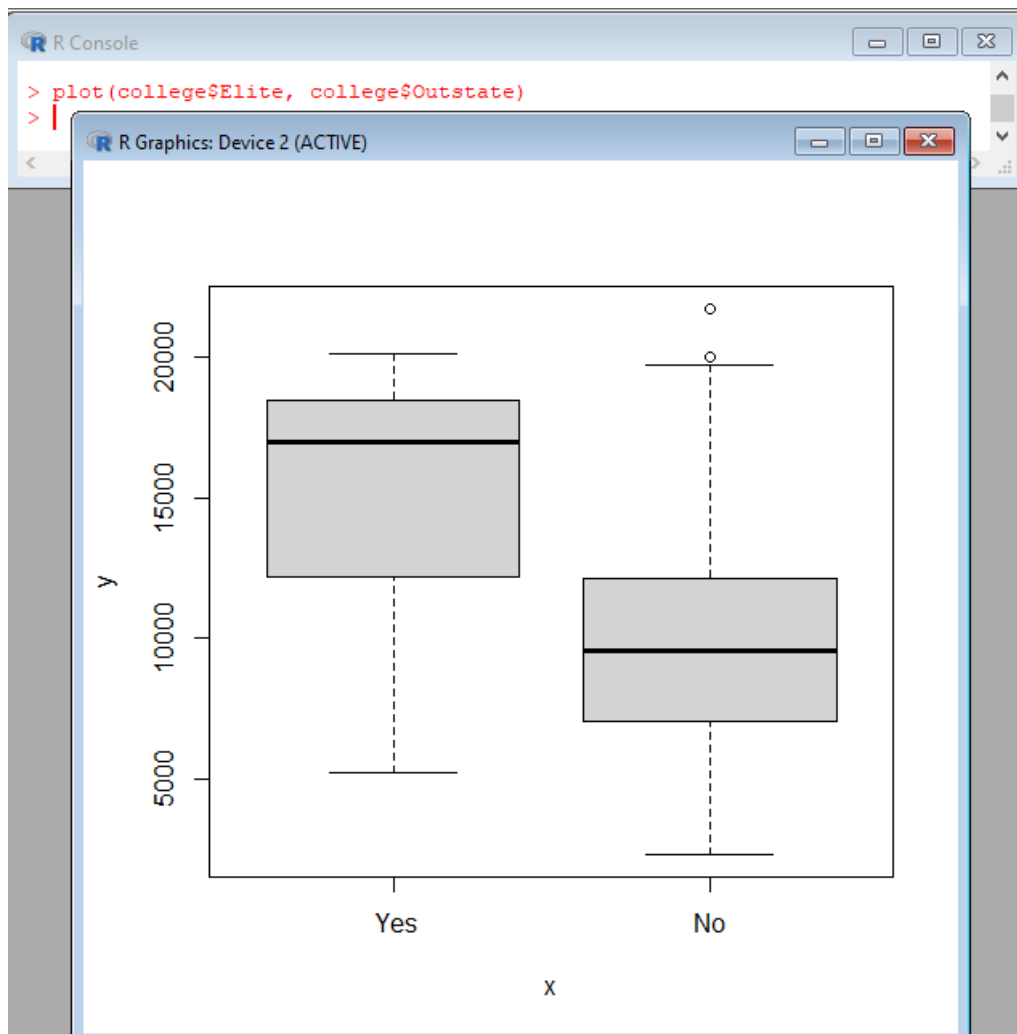


1.3.4. Використовуючи Top10perc, створила новий якісний показник Elite.

Командами нижче поділила всі університети на дві групи в залежності чи перевищує відсоток студентів з топ 10% шкіл 50% чи ні. Тоді використовуючи summary() визначила, що елітних мало, а саме 78, тоді як неелітних 699.

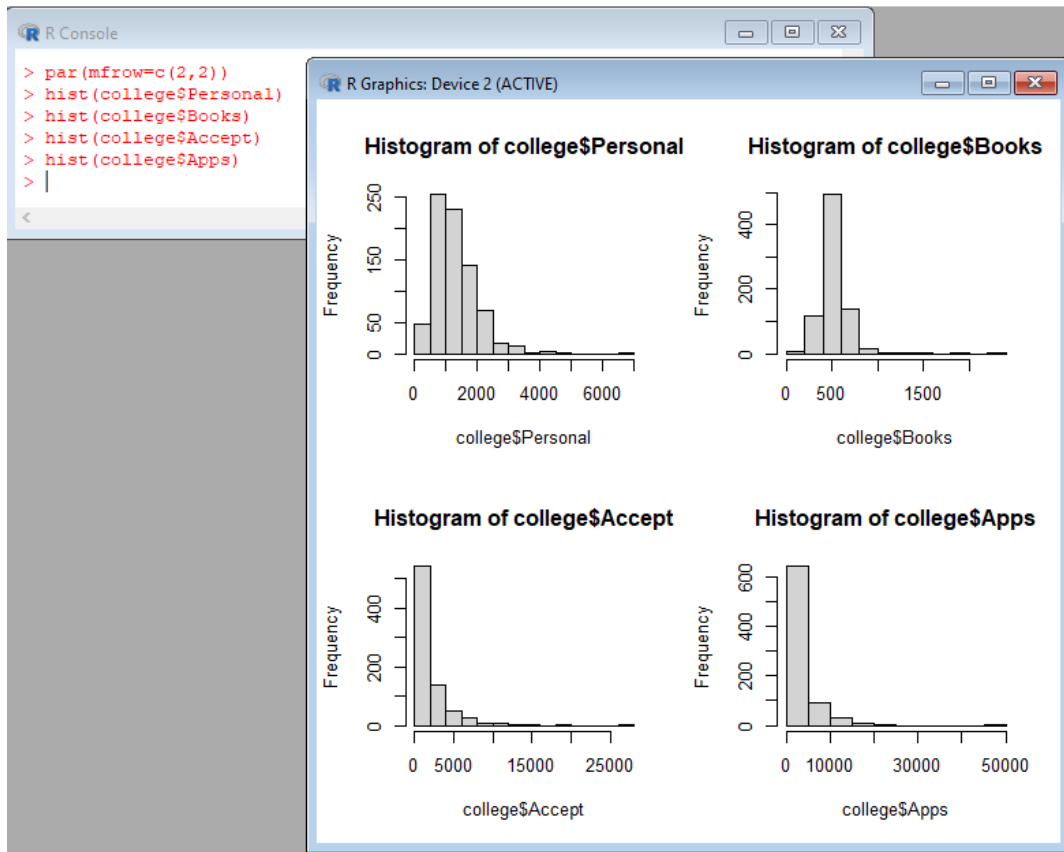
```
R Console
> Elite =rep ("No",nrow(college ))
> Elite [college$Top10perc >50]=" Yes"
> Elite =as.factor (Elite)
> college =data.frame(college ,Elite)
> summary(Elite)
Yes  No
  78 699
>
```

З допомогою plot() побудувала діаграму Outstate vs Elite.



1.3.5. Використовуючи `hist()`, побудувала декілька гістограм для кількісних показників, а саме: оціночних персональних витрат, оціночної вартості книжок, кількості отриманих заяв, кількості прийнятих заяв. Команда `par(mfrow=c(2,2))` допомогла поділити вікно графіків на 4 частини і

побудувати 4 графіки одночасно.



1.3.6. Як висновок можна сказати, що дані містять одну якісну характеристику – Privat, а решта – кількісні. Зручно підбивати підсумки по кількісних показниках, однією функцією одразу видно по кожному мінімум, максимум, середнє, медіану тощо. Наприклад так можна дізнатись, що найменша кількість поданих заяв – 81 штука, а найбільша - аж 48094. Завдяки одночасному відображенню кількох гістограмам можна простежити залежності, наприклад як на останньому скріншоті видно лінійну залежність між Apps і Асерт. Також зручно досліджувати вибірки по кількох параметрах, як наприклад тут було пораховано кількість елітних та неелітних закладів.

Завдання 2

2 Завантажила дані з файлу 'Auto.csv', видаливши пропущені значення:

```
R Console
> autos = read.csv('Auto.csv', header=T, na.strings = '?')
> autos = na.omit(autos)
> print(autos)
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	year	origin	name
1	18.0	8	307.0	130	3504	12.0	70	1	chevrolet chevelle malibu
2	15.0	8	350.0	165	3693	11.5	70	1	buick skylark 320
3	18.0	8	318.0	150	3436	11.0	70	1	plymouth satellite
4	16.0	8	304.0	150	3433	12.0	70	1	amc rebel sst
5	17.0	8	302.0	140	3449	10.5	70	1	ford torino
6	15.0	8	429.0	198	4341	10.0	70	1	ford galaxie 500
7	14.0	8	454.0	220	4354	9.0	70	1	chevrolet impala
8	14.0	8	440.0	215	4312	8.5	70	1	plymouth fury iii
9	14.0	8	455.0	225	4425	10.0	70	1	pontiac catalina
10	15.0	8	390.0	190	3850	8.5	70	1	amc ambassador dpl
11	15.0	8	383.0	170	3563	10.0	70	1	dodge challenger se
12	14.0	8	340.0	160	3609	8.0	70	1	plymouth 'cuda 340
13	15.0	8	400.0	150	3761	9.5	70	1	chevrolet monte carlo
14	14.0	8	455.0	225	3086	10.0	70	1	buick estate wagon (sw)
15	24.0	4	113.0	95	2372	15.0	70	3	toyota corona mark ii
16	22.0	6	198.0	95	2833	15.5	70	1	plymouth duster
17	18.0	6	199.0	97	2774	15.5	70	1	amc hornet
18	21.0	6	200.0	85	2587	16.0	70	1	ford maverick
19	27.0	4	97.0	88	2130	14.5	70	3	datsum pl510
20	26.0	4	97.0	46	1835	20.5	70	2	volkswagen 1131 deluxe sedan
21	25.0	4	110.0	87	2672	17.5	70	2	peugeot 504
22	24.0	4	107.0	90	2430	14.5	70	2	audi 100 ls
23	25.0	4	104.0	95	2375	17.5	70	2	saab 99e
24	26.0	4	121.0	113	2234	12.5	70	2	bmw 2002
25	21.0	6	199.0	90	2648	15.0	70	1	amc gremlin
26	10.0	8	360.0	215	4615	14.0	70	1	ford f250
27	10.0	8	307.0	200	4376	15.0	70	1	chevy c20
28	11.0	8	318.0	210	4382	13.5	70	1	dodge d200
29	9.0	8	304.0	193	4732	18.5	70	1	hi 1200d
30	27.0	4	97.0	88	2130	14.5	71	3	datsum pl510
31	28.0	4	140.0	90	2264	15.5	71	1	chevrolet vega 2300
32	25.0	4	113.0	95	2228	14.0	71	3	toyota corona

2.1 Функція `summary()` допомогла визначити якісні показники. У характеристиках `cylinders`, `origin` і `year` середнє значення не несе ніякого кількісного змісту (наприклад “5 з половиною циліндрів у двигуні”), тому вони якісні, так само як `name`. Відповідно `mpg`, `displacement`, `horsepower`, `weight`, `acceleration` – кількісні.

```
R Console
> summary(autos)
```

mpg	cylinders	displacement	horsepower	weight	acceleration	year
Min. : 9.00	Min. : 3.000	Min. : 68.0	Min. : 46.0	Min. : 1613	Min. : 8.00	Min. : 70.00
1st Qu.: 17.00	1st Qu.: 4.000	1st Qu.: 105.0	1st Qu.: 75.0	1st Qu.: 2225	1st Qu.: 13.78	1st Qu.: 73.00
Median : 22.75	Median : 4.000	Median : 151.0	Median : 93.5	Median : 2804	Median : 15.50	Median : 76.00
Mean : 23.45	Mean : 5.472	Mean : 194.4	Mean : 104.5	Mean : 2978	Mean : 15.54	Mean : 75.98
3rd Qu.: 29.00	3rd Qu.: 8.000	3rd Qu.: 275.8	3rd Qu.: 126.0	3rd Qu.: 3615	3rd Qu.: 17.02	3rd Qu.: 79.00
Max. : 46.60	Max. : 8.000	Max. : 455.0	Max. : 230.0	Max. : 5140	Max. : 24.80	Max. : 82.00

origin	name
Min. : 1.000	Length: 392
1st Qu.: 1.000	Class : character
Median : 1.000	Mode : character
Mean : 1.577	
3rd Qu.: 2.000	
Max. : 3.000	

2.2 Функцією range() визначила межі для кожного кількісного показника:

```
R Console
> range(autos$mpg)
[1] 9.0 46.6
> range(autos$displacement)
[1] 68 455
> range(autos$horsepower)
[1] 46 230
> range(autos$weight)
[1] 1613 5140
> range(autos$acceleration)
[1] 8.0 24.8
```

2.3 Обчислила середнє і стандартне відхилення кожного кількісного показника

```
R Console
> mean(autos$mpg)
[1] 23.44592
> mean(autos$displacement)
[1] 194.412
> mean(autos$horsepower)
[1] 104.4694
> mean(autos$weight)
[1] 2977.584
> mean(autos$acceleration)
[1] 15.54133
> |
```

```
R Console
> sd(autos$mpg)
[1] 7.805007
> sd(autos$displacement)
[1] 104.644
> sd(autos$horsepower)
[1] 38.49116
> sd(autos$weight)
[1] 849.4026
> sd(autos$acceleration)
[1] 2.758864
> |
```

2.4 Видалила спостереження з 10-го по 85-те

```
R Console
> autos1 = autos[-(10:84),]
> print(autos1)
```

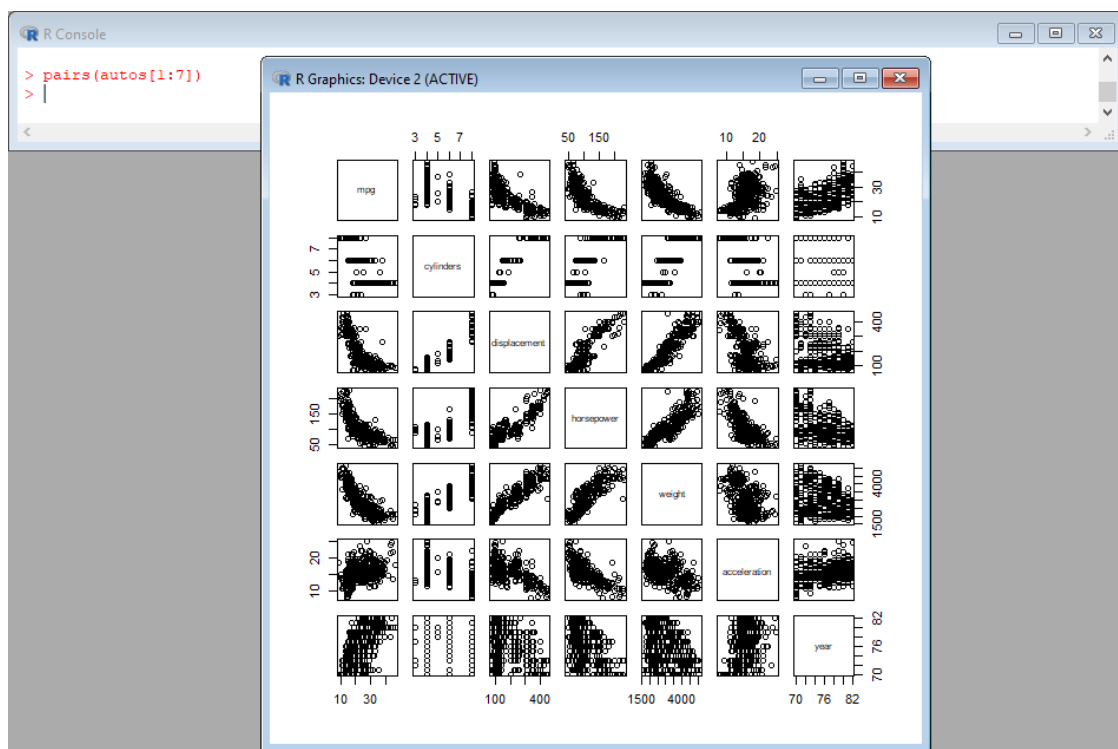
	mpg	cylinders	displacement	horsepower	weight	acceleration	year	origin	name
1	18.0	8	307	130	3504	12.0	70	1	chevrolet chevelle malibu
2	15.0	8	350	165	3693	11.5	70	1	buick skylark 320
3	18.0	8	318	150	3436	11.0	70	1	plymouth satellite
4	16.0	8	304	150	3433	12.0	70	1	amc rebel sst
5	17.0	8	302	140	3449	10.5	70	1	ford torino
6	15.0	8	429	198	4341	10.0	70	1	ford galaxie 500
7	14.0	8	454	220	4354	9.0	70	1	chevrolet impala
8	14.0	8	440	215	4312	8.5	70	1	plymouth fury iii
9	14.0	8	455	225	4425	10.0	70	1	pontiac catalina
86	13.0	8	350	175	4100	13.0	73	1	buick century 350
87	14.0	8	304	150	3672	11.5	73	1	amc matador
88	13.0	8	302	145	3988	13.0	73	1	chevrolet malibu
89	14.0	8	350	137	4042	14.5	73	1	ford gran torino
90	15.0	8	318	150	3777	12.5	73	1	dodge coronet custom
91	12.0	8	429	198	4952	11.5	73	1	mercury marquis brougham
92	13.0	8	400	150	4464	12.0	73	1	chevrolet caprice classic
93	13.0	8	351	158	4363	13.0	73	1	ford ltd
94	14.0	8	318	150	4237	14.5	73	1	plymouth fury gran sedan
95	13.0	8	440	215	4735	11.0	73	1	chrysler new yorker brougham
96	12.0	8	455	225	4951	11.0	73	1	buick electra 225 custom
97	13.0	8	360	175	3821	11.0	73	1	amc ambassador brougham
98	18.0	6	225	105	3121	16.5	73	1	plymouth valiant
99	16.0	6	250	100	3278	18.0	73	1	chevrolet nova custom
100	18.0	6	232	100	2945	16.0	73	1	amc hornet
101	18.0	6	250	88	3021	16.5	73	1	ford maverick
102	23.0	6	198	95	2904	16.0	73	1	plymouth duster
103	26.0	4	97	46	1950	21.0	73	2	volkswagen super beetle
104	11.0	8	400	150	4997	14.0	73	1	chevrolet impala
105	12.0	8	400	167	4906	12.5	73	1	ford country
106	13.0	8	360	170	4654	13.0	73	1	plymouth custom suburb
107	12.0	8	350	180	4499	12.5	73	1	oldsmobile vista cruiser
108	18.0	6	232	100	2789	15.0	73	1	amc gremlin
109	20.0	4	97	88	2279	19.0	73	3	toyota carina

Обчислила нове середнє і стандартне відхилення кожного кількісного показника:

```
R Console
> mean(autos1$mpg)
[1] 24.36845
> mean(autos1$displacement)
[1] 187.7539
> mean(autos1$horsepower)
[1] 100.9558
> mean(autos1$weight)
[1] 2939.644
> mean(autos1$acceleration)
[1] 15.7183
```

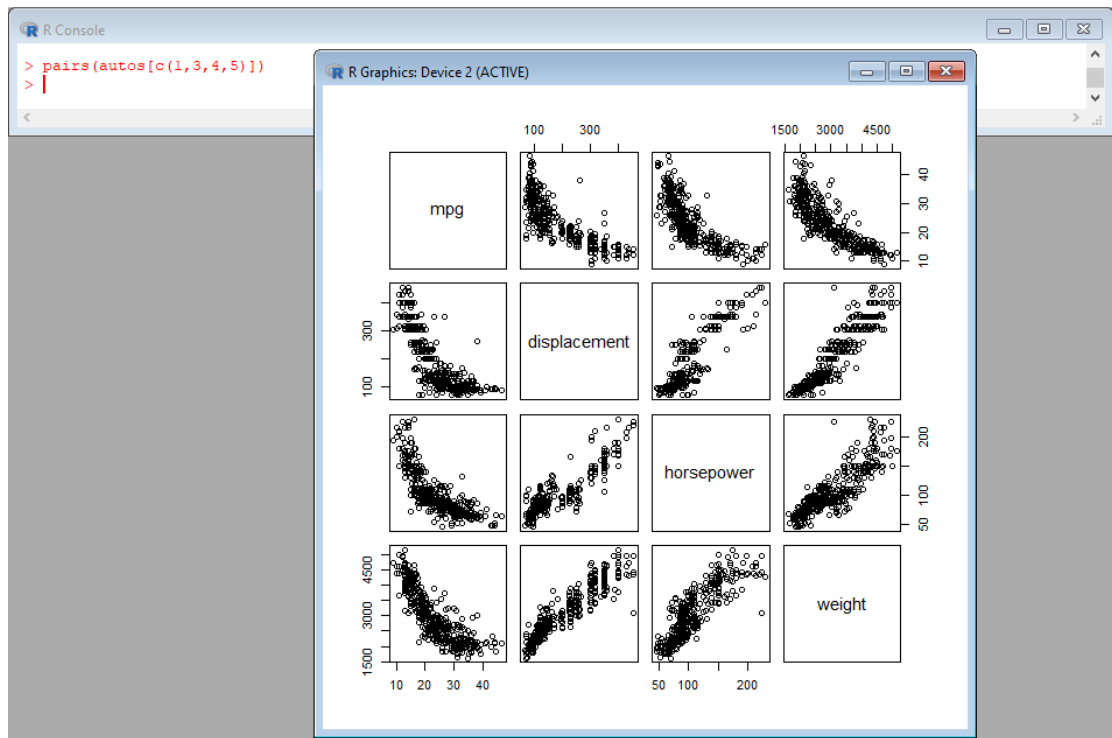
```
R Console
> sd(autos1$mpg)
[1] 7.880898
> sd(autos1$displacement)
[1] 99.93949
> sd(autos1$horsepower)
[1] 35.89557
> sd(autos1$weight)
[1] 812.6496
> sd(autos1$acceleration)
[1] 2.693813
> |
```

2.5 Використовуючи повний набір даних зобразила показники графічно функцією pairs():



Як результат, можна бачити залежності між величинами, наприклад що horsepower лінійно залежить від displacement і weight.

2.6 Щоб передбачити розхід пального (mpg) на основі інших показників, можна також скористатись графічним способом.



Як результат, видно що при зростанні таких показників як displacement, horsepower, weight – mpg спадає, тобто обернено пропорційну залежність.

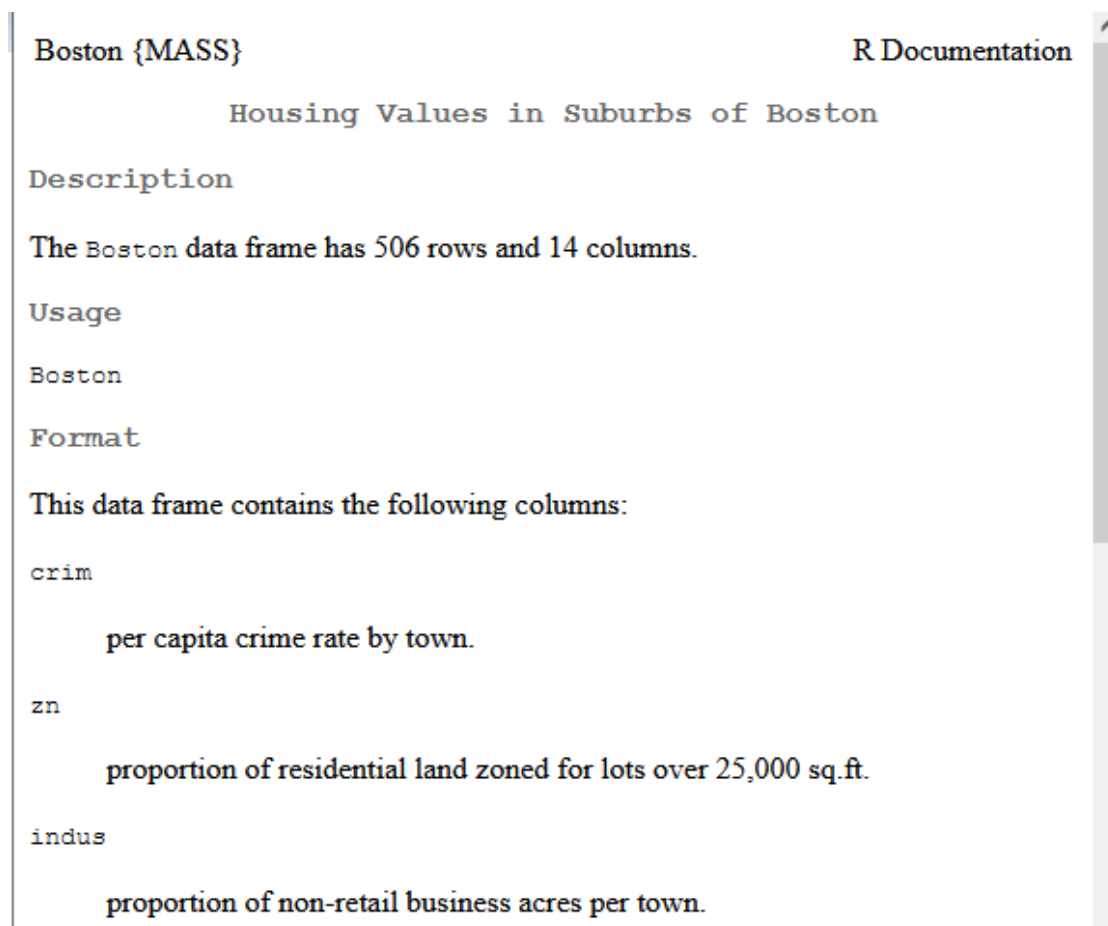
Завдання 3

3.1 Завантажила дані Boston, які є частиною бібліотеки MASS:

```
> library(MASS)
> Boston
```

	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
1	0.00632	18.0	2.31	0	0.5380	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
2	0.02731	0.0	7.07	0	0.4690	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
3	0.02729	0.0	7.07	0	0.4690	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
4	0.03237	0.0	2.18	0	0.4580	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
5	0.06905	0.0	2.18	0	0.4580	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2
6	0.02985	0.0	2.18	0	0.4580	6.430	58.7	6.0622	3	222	18.7	394.12	5.21	28.7
7	0.08829	12.5	7.87	0	0.5240	6.012	66.6	5.5605	5	311	15.2	395.60	12.43	22.9
8	0.14455	12.5	7.87	0	0.5240	6.172	96.1	5.9505	5	311	15.2	396.90	19.15	27.1
9	0.21124	12.5	7.87	0	0.5240	5.631	100.0	6.0821	5	311	15.2	386.63	29.93	16.5
10	0.17004	12.5	7.87	0	0.5240	6.004	85.9	6.5921	5	311	15.2	386.71	17.10	18.9
11	0.22489	12.5	7.87	0	0.5240	6.377	94.3	6.3467	5	311	15.2	392.52	20.45	15.0
12	0.11747	12.5	7.87	0	0.5240	6.009	82.9	6.2267	5	311	15.2	396.90	13.27	18.9
13	0.09378	12.5	7.87	0	0.5240	5.889	39.0	5.4509	5	311	15.2	390.50	15.71	21.7
14	0.62976	0.0	8.14	0	0.5380	5.949	61.8	4.7075	4	307	21.0	396.90	8.26	20.4
15	0.63796	0.0	8.14	0	0.5380	6.096	84.5	4.4619	4	307	21.0	380.02	10.26	18.2
16	0.62739	0.0	8.14	0	0.5380	5.834	56.5	4.4986	4	307	21.0	395.62	8.47	19.9
17	1.05393	0.0	8.14	0	0.5380	5.935	29.3	4.4986	4	307	21.0	386.85	6.58	23.1
18	0.78420	0.0	8.14	0	0.5380	5.990	81.7	4.2579	4	307	21.0	386.75	14.67	17.5
19	0.80271	0.0	8.14	0	0.5380	5.456	36.6	3.7965	4	307	21.0	288.99	11.69	20.2
20	0.72580	0.0	8.14	0	0.5380	5.727	69.5	3.7965	4	307	21.0	390.95	11.28	18.2
21	1.25179	0.0	8.14	0	0.5380	5.570	98.1	3.7979	4	307	21.0	376.57	21.02	13.6
22	0.85204	0.0	8.14	0	0.5380	5.965	89.2	4.0123	4	307	21.0	392.53	13.83	19.6
23	1.23247	0.0	8.14	0	0.5380	6.142	91.7	3.9769	4	307	21.0	396.90	18.72	15.2
24	0.98843	0.0	8.14	0	0.5380	5.813	100.0	4.0952	4	307	21.0	394.54	19.88	14.5
25	0.75026	0.0	8.14	0	0.5380	5.924	94.1	4.3996	4	307	21.0	394.33	16.30	15.6
26	0.84054	0.0	8.14	0	0.5380	5.599	85.7	4.4546	4	307	21.0	303.42	16.51	13.9
27	0.67191	0.0	8.14	0	0.5380	5.813	90.3	4.6820	4	307	21.0	376.88	14.81	16.6
28	0.95577	0.0	8.14	0	0.5380	6.047	88.8	4.4534	4	307	21.0	306.38	17.28	14.8
29	0.77299	0.0	8.14	0	0.5380	6.495	94.4	4.4547	4	307	21.0	387.94	12.80	18.4
30	1.00245	0.0	8.14	0	0.5380	6.674	87.3	4.2390	4	307	21.0	380.23	11.98	21.0
31	1.13081	0.0	8.14	0	0.5380	5.713	94.1	4.2330	4	307	21.0	360.17	22.60	12.7
32	1.35472	0.0	8.14	0	0.5380	6.072	100.0	4.1750	4	307	21.0	376.73	13.04	14.5
33	1.38799	0.0	8.14	0	0.5380	5.950	82.0	3.9900	4	307	21.0	232.60	27.71	13.2

Переглянула інформацію про дані:



Boston {MASS} R Documentation

Housing Values in Suburbs of Boston

Description

The `Boston` data frame has 506 rows and 14 columns.

Usage

```
Boston
```

Format

This data frame contains the following columns:

- `crim`
per capita crime rate by town.
- `zn`
proportion of residential land zoned for lots over 25,000 sq.ft.
- `indus`
proportion of non-retail business acres per town.

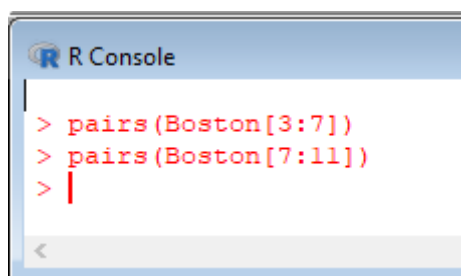
Переглянула розмірність даних Boston:



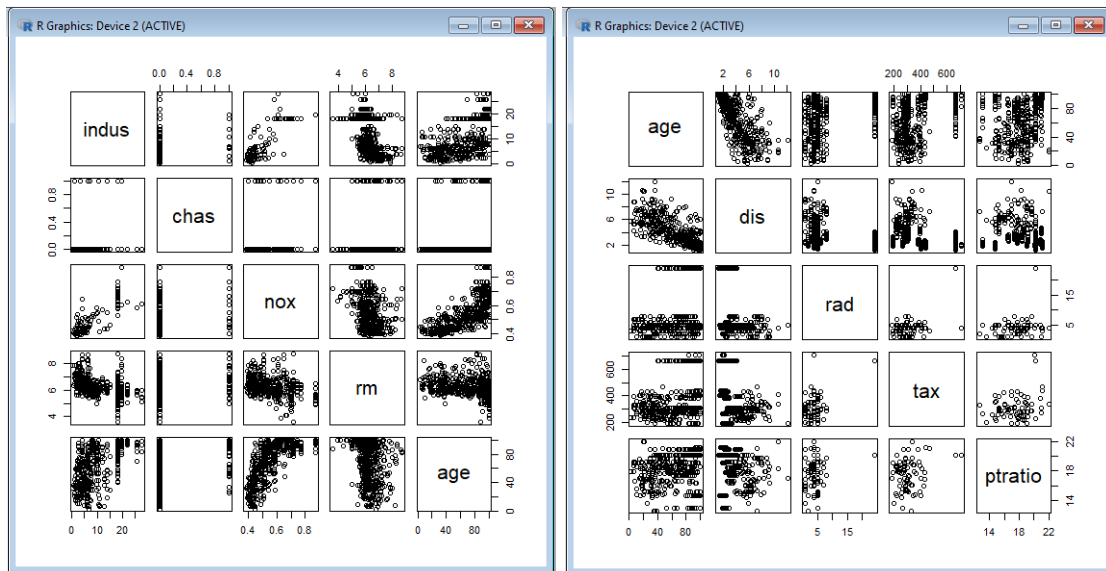
```
R Console
> dim(Boston)
[1] 506 14
> .
```

Дані містять 506 рядків та 14 стовпців.

3.2 Побудувала декілька попарних графіків показників



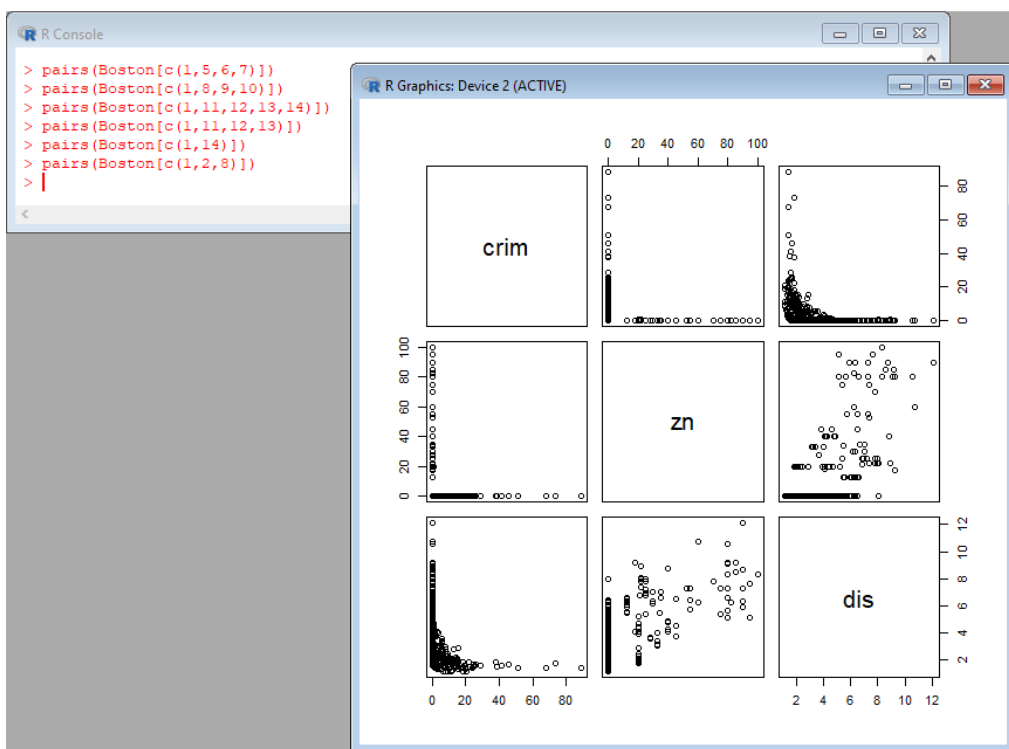
```
R Console
> pairs(Boston[3:7])
> pairs(Boston[7:11])
> |
```

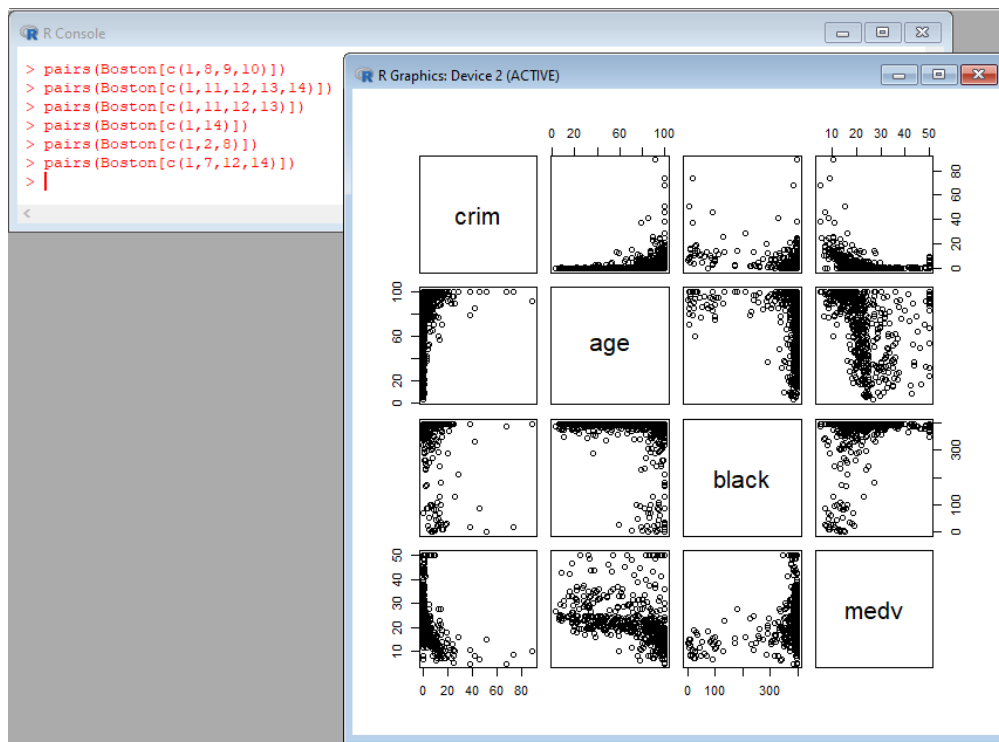


По цих графіках видно, що жодних чітко виражених залежностей між обраними показниками немає.

3.3 Чи пов'язаний якийсь із показників із рівнем злочинності на душу населення?

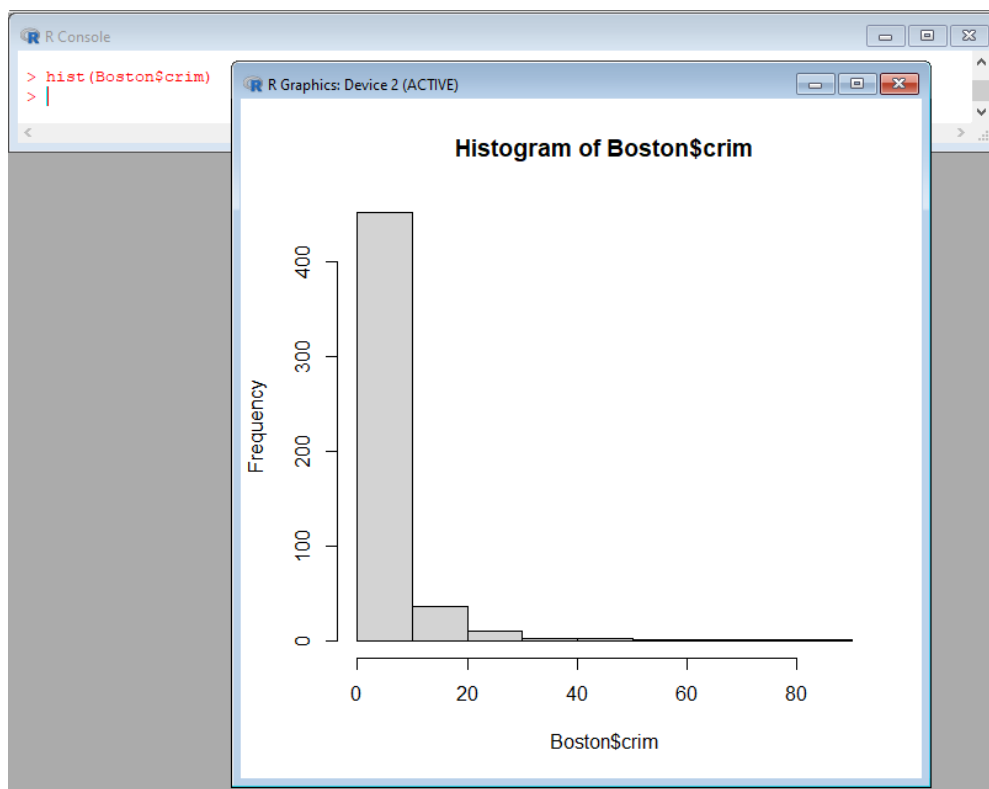
Для цього попарними графіками переглянула усі показники, та підсумувала відповідь у наступних забраженнях:





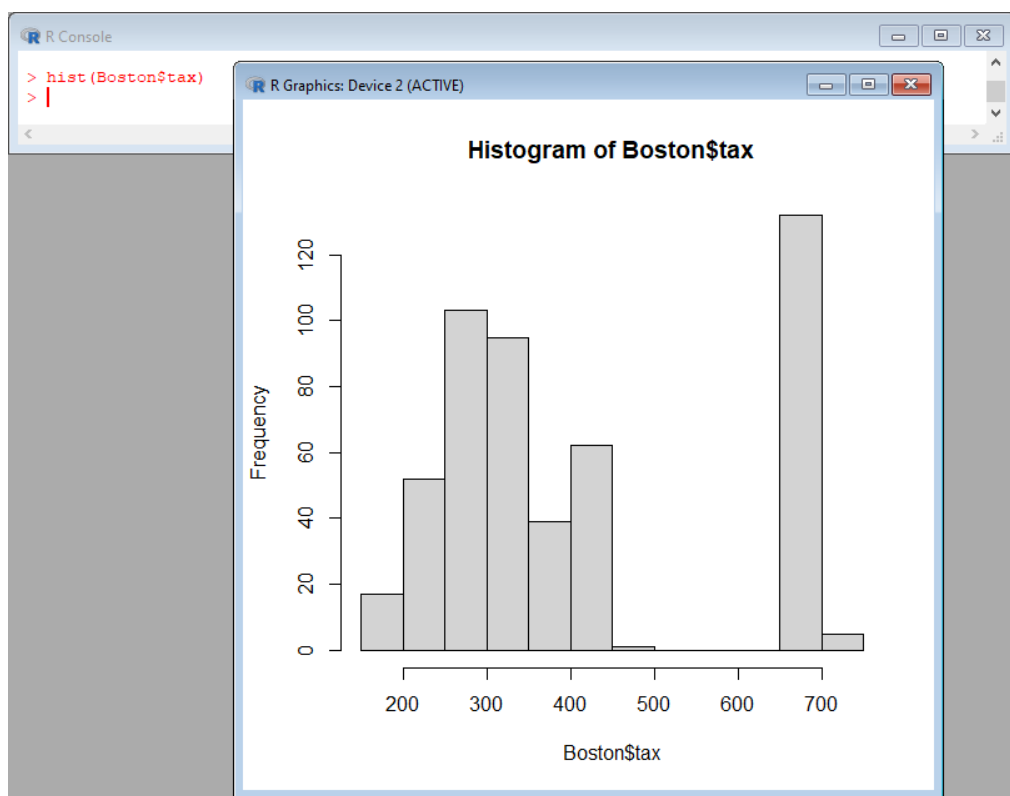
Чітко виражених залежностей немає, але в загальних рисах на першому зображенні видно, що чим менші показники *zn* та *dis* тим нижчий рівень злочинності, тоді як на другому зображенні видно, що чим менші показники *age*, *black* та *medv* тим вищий рівень злочинності.

3.4 Чи є в кварталах Бостона особливо високий рівень злочинності?



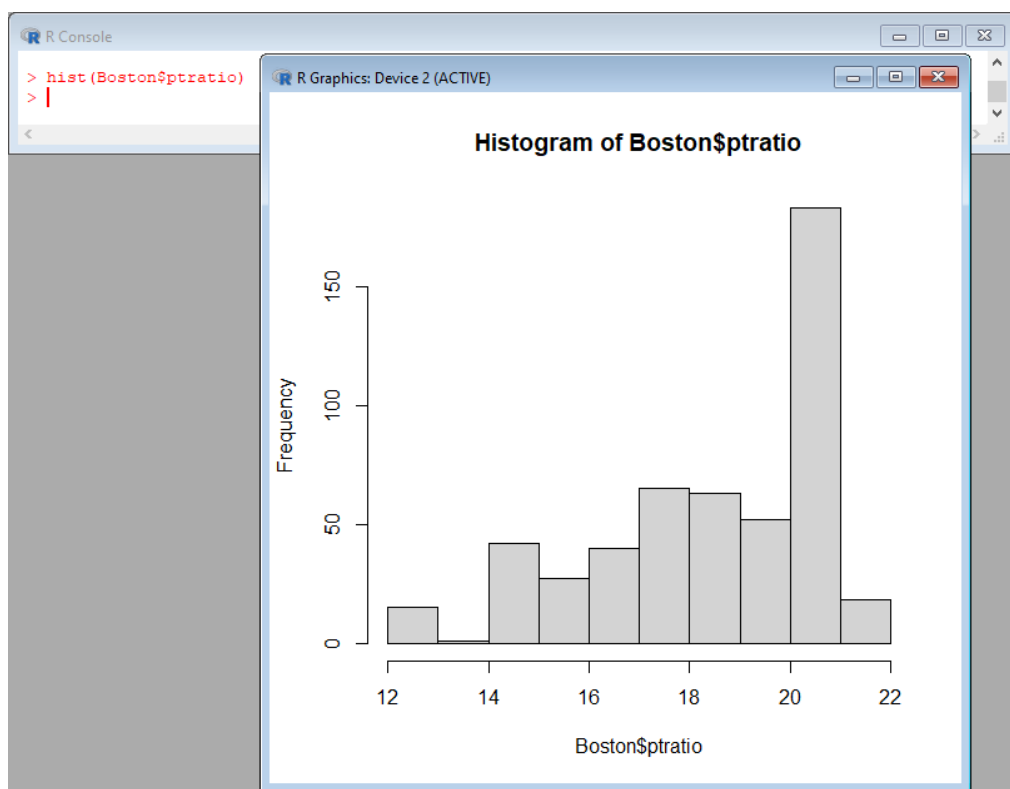
У більшості кварталів Бостона рівень злочинності низький (менше 20), але є й райони, де рівень злочинності особливо високий (більше 80).

Чи є в кварталах Бостона особливо високі податкові ставки?



Більша частина податкових ставок не досягає позначки в 450, але є квартали, у яких особливо високі – понад 650.

Чи є в кварталах Бостона особливо високі співвідношення учнів та вчителів?



Загалом співвідношення вчителів і учнів у школах не сильно відрізняються, але є досить багато районів, де воно 20-21.

3.5 Скільки кварталів в даній множині межують з річкою Charles?

```
R Console
> table(Boston$chas)

 0    1
471  35
> |
```

Межують з річкою Charles 35 кварталів (і не межують відповідно 471).

3.6 Обчислення медіани для відношення учні-вчителі для міста загалом:

```
R Console
> median(Boston$ptratio)
[1] 19.05
> |
```

3.6 Який квартал (квартали) міста має найменшу медіану кількості зайнятих помешкань?

Знаходжу найменшу медіану зайнятих помешкань:

```
R Console
> min_medv = min(Boston$medv)
> min_medv
[1] 5
> |
```

Цій медіані відповідають квартали 399 та 406. Значення інших показників цих кварталів:

```
R Console
> res=Boston[Boston$medv == min_medv,]
> res
      crim zn  indus chas   nox   rm age   dis rad tax ptratio  black lstat medv
399 38.3518  0  18.1    0 0.693 5.453 100 1.4896 24 666    20.2 396.90 30.59    5
406 67.9208  0  18.1    0 0.693 5.683 100 1.4254 24 666    20.2 384.97 22.98    5
> |
```

Використаю функцію `summary()`, щоб порівняти ці показники цих районів з показниками інших:


```
R Console
> summary(Boston)
      crim          zn          indus          chas          nox
Min.   : 0.00632   Min.   : 0.00   Min.   : 0.46   Min.   :0.00000   Min.   :0.3850
1st Qu.: 0.08205   1st Qu.: 0.00   1st Qu.: 5.19   1st Qu.:0.00000   1st Qu.:0.4490
Median : 0.25651   Median : 0.00   Median : 9.69   Median :0.00000   Median :0.5380
Mean   : 3.61352   Mean   : 11.36   Mean   :11.14   Mean   :0.06917   Mean   :0.5547
3rd Qu.: 3.67708   3rd Qu.: 12.50   3rd Qu.:18.10   3rd Qu.:0.00000   3rd Qu.:0.6240
Max.   :88.97620   Max.   :100.00   Max.   :27.74   Max.   :1.00000   Max.   :0.8710

      rm          age          dis          rad          tax
Min.   :3.561     Min.   : 2.90   Min.   : 1.130   Min.   : 1.000   Min.   :187.0
1st Qu.:5.886     1st Qu.: 45.02   1st Qu.: 2.100   1st Qu.: 4.000   1st Qu.:279.0
Median :6.208     Median : 77.50   Median : 3.207   Median : 5.000   Median :330.0
Mean   :6.285     Mean   : 68.57   Mean   : 3.795   Mean   : 9.549   Mean   :408.2
3rd Qu.:6.623     3rd Qu.: 94.08   3rd Qu.: 5.188   3rd Qu.:24.000   3rd Qu.:666.0
Max.   :8.780     Max.   :100.00   Max.   :12.127   Max.   :24.000   Max.   :711.0

      ptratio      black      lstat      medv
Min.   :12.60     Min.   : 0.32   Min.   : 1.73   Min.   : 5.00
1st Qu.:17.40     1st Qu.:375.38   1st Qu.: 6.95   1st Qu.:17.02
Median :19.05     Median :391.44   Median :11.36   Median :21.20
Mean   :18.46     Mean   :356.67   Mean   :12.65   Mean   :22.53
3rd Qu.:20.20     3rd Qu.:396.23   3rd Qu.:16.95   3rd Qu.:25.00
Max.   :22.00     Max.   :396.90   Max.   :37.97   Max.   :50.00
> |
```

Співставивши ці значення видно, що у районах 399 та 406 показники `crim`, `indus`, `nox`, `age`, `rad`, `tax`, `ptratio`, `black`, `lstat` - є більшими за середні, а показники `zn`, `chas`, `rm`, `dis`, `medv` – меншими за середні.