

В следващите задачи разглеждаме 'survey' от пакета 'MASS'.

Зад.1 Намерете:

а) броя на мъжете;

```
> table( Sex )
```

```
118
```

б) броя на мъжете пушачи;

```
> t = table( Sex, Smoke )
```

```
> sum( t[2, -2] )
```

```
28
```

в) средната височина на мъжете;

```
> mean( Height[ Sex == 'Male' ], na.rm = T )
```

```
178.8
```

г) височината и пола на 6-те най-млади студента.

```
> i = order( Age )
```

```
> Height[ i[ 1 : 6 ] ]
```

```
NA NA NA 160.00 172.00 170.18
```

```
> Sex[ i[ 1 : 6 ] ]
```

```
Male Male Female Female Female Female
```

Зад.2 Въз основа на данните пресметнете вероятностите:

а) Случайно избран човек да се окаже редовен пушач.

```
> prop.table( table( Smoke ) )
```

```
0.072
```

б) Случайно избран човек да се окаже редовно пушещ мъж.

```
> prop.table( table( Sex, Smoke ) )
```

```
0.051
```

в) Случайно избран мъж да се окаже редовен пушач.

```
> prop.table( table( Sex, Smoke ), 1 )
```

```
0.102
```

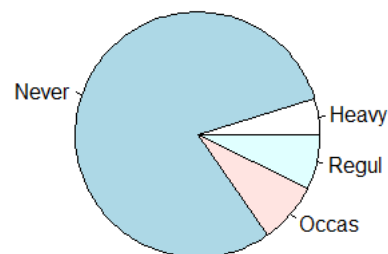
г) Случайно избран редовен пушач да се окаже мъж.

```
> prop.table( table( Sex, Smoke ), 2 )
```

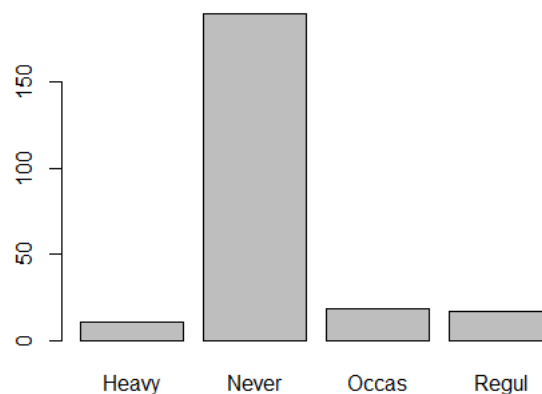
```
0.705
```

Зад.3 Представете графично данните за тютюнопушенето на студентите. А също и за тютюнопушенето в зависимост от пола.

```
> pie( table( Smoke ) )
```



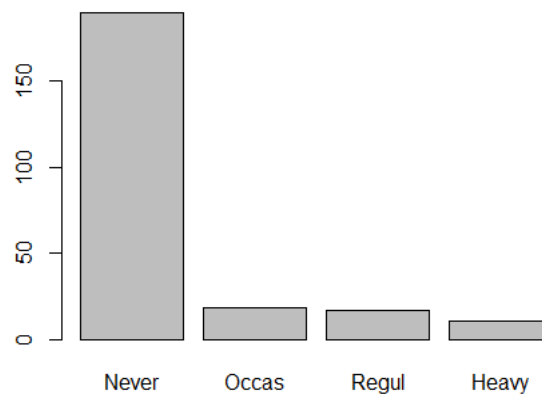
```
> barplot( table( Smoke ) )
```



И от двете графики ясно се вижда, че непушачите са най-много.

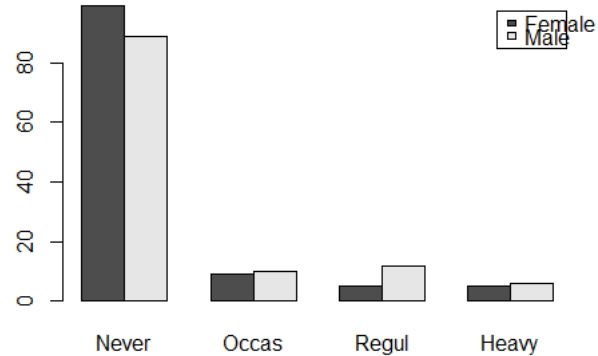
На втората графика нивата на фактора не са подходящо подредени, ще ги разменим за да има градация. За целта ще създадем нов фактор.

```
> SMK = factor( Smoke, levels = levels(Smoke)[ c(2, 3, 4, 1) ] )  
> tb = table( SMK )  
> barplot( tb )
```



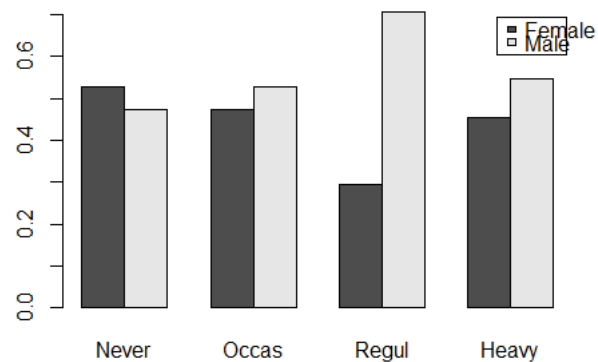
Ще представим на една графика тютюнопушенето и пола.

```
> t = table( Sex , SMK )  
> barplot(t, beside = T, legend = T)
```



Ще покажем процентното отношение на мъжете и жените във всяка от групите.

```
> pt = prop.table( t, 2 )  
> barplot(pt, beside = T, legend = T)
```



От графиката изглежда, че мъжете пушат повече от жените.

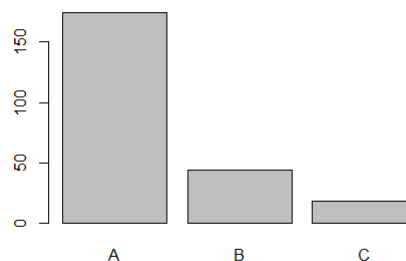
Зад.4 Разделете студентите според възрастта им на три групи: А - до 20г, В - от 20 до 25 и С - над 25. Представете графично. Направете таблица за разпределението на пушачите в различните възрасти, представете графично.

```
> group = cut( Age, breaks = c(0, 20, 25, 100), labels = c('A', 'B', 'C'))
```

```
> table( group )
```

```
group
  A   B   C
174  44  19
```

```
> barplot( table( group))
```



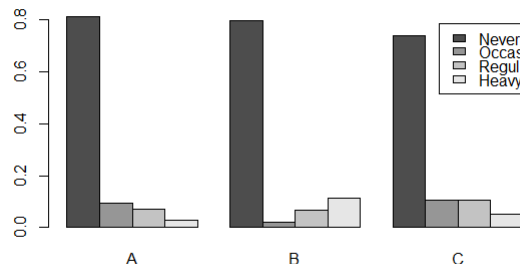
Ще потърсим връзката между възрастта и тютюнопушенето.

```
> tb = table(SMK, group )
```

```
      group
SMK    A   B   C
Never 140  35  14
Occas  16   1   2
Regul  12   3   2
Heavy   5   5   1
```

```
> prt看 = prop.table(tb, 2)
```

```
> barplot( prt看 , beside = T, legend = T)
```



С нарастване на възрастта пушачите изглежда се увеличават.