Εργασία 4

Μάριος Κιοράι

2024-04-02

Εργασία 2

Ερώτηση1: Να μελετήσετε τη μεταξύ τους συσχέτιση (cor) a)carat x price

```
library(ggplot2)
data(diamonds)
cor(diamonds$carat, diamonds$price, method = "pearson")
```

```
## [1] 0.9215913
```

Παρατηρούμε ότι ο αριθμός που μας δίνεται από την συνάρτηση της συσχέτισης είναι πολύ κοντά στον αριθμό 1. Αυτό σημαίνει ότι οι δύο αυτές μεταβλητές έχουν πολύ έντονη θετική συσχέτιση. Πιο συγκεκριμένα καθώς αυξάνεται το μέγεθος των καρατιών, αυξάνεται και η τιμή των διαμαντιών. **b)carat x depth**

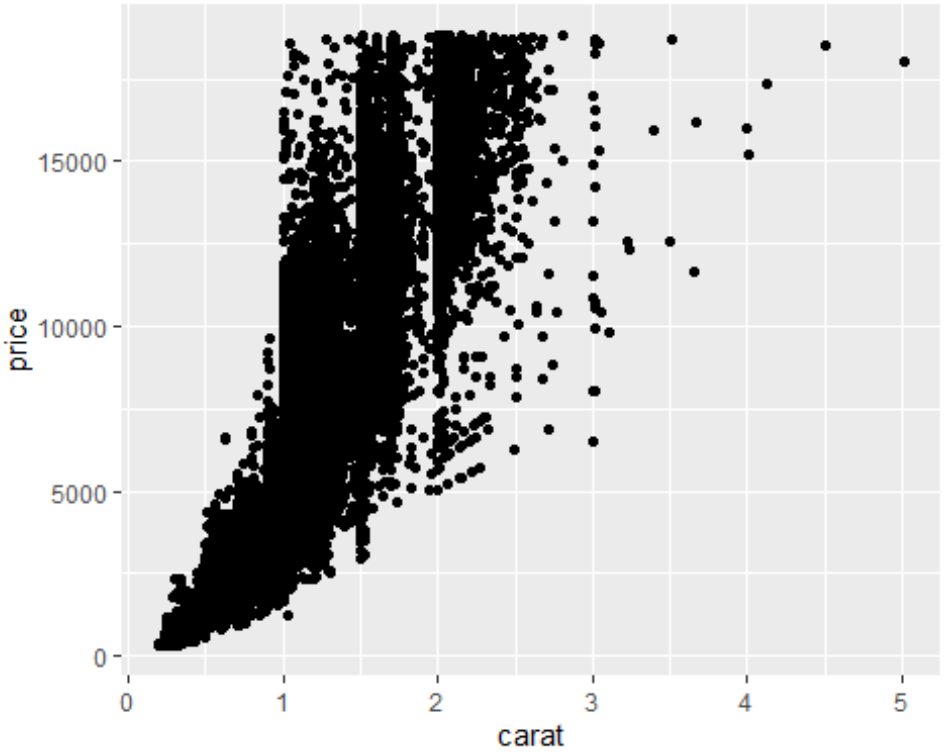
```
cor(diamonds$carat, diamonds$depth, method = "pearson")
```

```
## [1] 0.02822431
```

Παρατηρούμε ότι ο αριθμός που μας δίνεται από την συνάρτηση της συσχέτισης είναι πολύ κοντά στον αριθμό 0. Αυτό σημαίνει ότι οι δύο αυτές μεταβλητές έχουν πολύ αδύναμη συσχέτιση.

Ερώτηση2: Να δημιουργήσετε το αντίστοιχο διάγραμμα διασποράς (scatterplot) a)carat x price

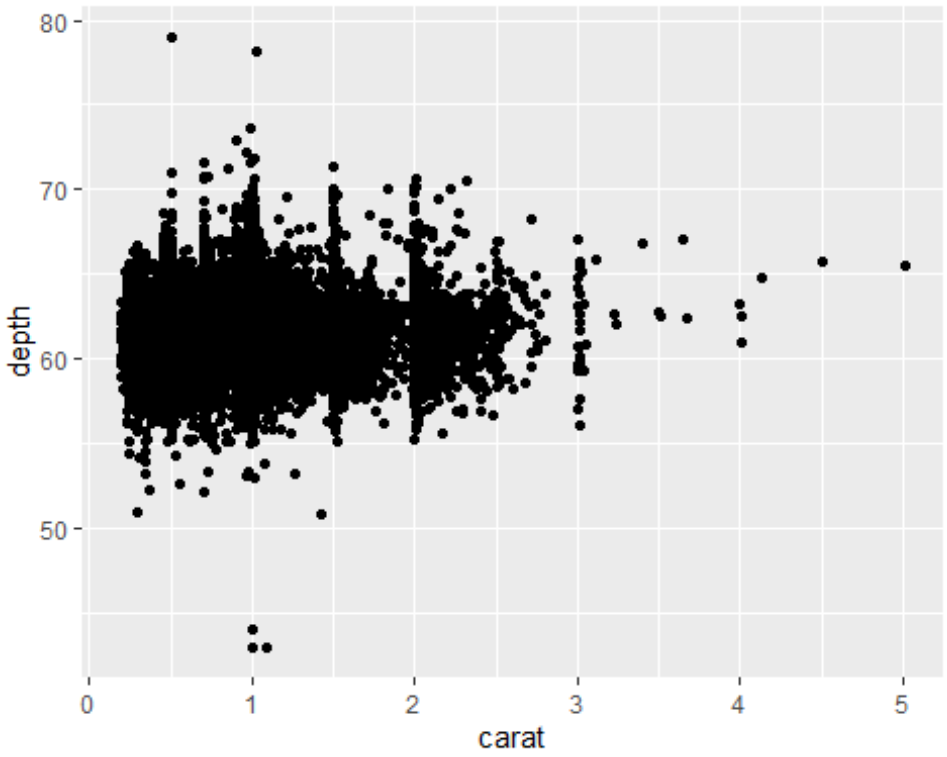
```
ggplot(diamonds, aes(x=carat,y=price))+geom_point()
```



- Καθώς το μέγεθος των καρατιών αυξάνεται η τιμή των διαμαντιών αυξάνεται επίσης.
- Τα δεδομένα όμως παρόλο που δείχνουν μια θετική τάση καθώς αυξάνεται η τιμή των καρατιών βλέπουμε οτι έχουμε μια πιο μεγάλη διασπορά στο διάγραμμα και αυτό συμβαίνει διότι μεγαλύτερα διαμάντια είναι πιο σπάνια να βρεθούν.
- Γι αυτόν ακριβώς τον λόγο έχουμε και αυτή την μεγάλη συσπείρωση όταν η τιμή των καρατιών είναι μικρή γιατί είναι πιο συνηθισμένα τα μικρότερα.

b)carat x depth

```
ggplot(diamonds, aes(x=carat,y=depth))+geom_point()
```



- Η σχέση μεταξύ καρατίου και βάθους δεν είναι γραμμική σε αντίθεση με πριν.
- Το ποσοστό βάθους συγκεντρώνεται κυρίως γύρω από μια συγκεκριμένη ζώνη (περίπου μεταξύ 60% και 70%), το οποίο αποτελεί συνηθισμένο εύρος για καλοκομμένα διαμάντια.
- Η συσπείρωση όταν η τιμή των καρατιών είναι μικρή γιατί είναι πιο συνηθισμένα τα μικρότερα διαμάντια και η διασπορά στο τέλος επείδη είναι πιο σπάνια.

Ερώτηση3: Να σχολιάσετε τα ευρήματά σας. Ο σχολιασμός βρίσκεται κάτω από κάθε ξεχωριστή γραμμή κώδικα

Εργασία 3

Ερώτηση1: Ποιά είναι η μέση τιμή της θερμοκρασίας για τη δεδομένη περίοδο;

```
mean(airquality$Temp)
```

```
## [1] 77.88235
```

Χρησιμοποιώντας την συνάρτηση mean εξάγουμε το παραπάνω αποτέλεσμα

Ερώτηση2: Ποιά ημέρα ήταν η θερμότερη;

```r
max(airquality$Temp)
```

```
## [1] 97
```

```r
hottest_day_index <- which.max(airquality$Temp)
hottest_day <- airquality$Day[hottest_day_index]
print(hottest_day)
```

```
## [1] 28
```

Εντοπίζουμε την μεγαλύτερη θερμοκρασία και στην συνέχεια την χρησιμοποιούμε για να βρούμε την ημέρα

Ερώτηση3: Ποιά είχε τον πολύ αέρα;

```r
max(airquality$Wind)
```

```
## [1] 20.7
```

```r
windiest_day_index <- which.max(airquality$Wind)
windiest_day <- airquality$Day[windiest_day_index]
print(windiest_day)
```

```
## [1] 17
```

Εντοπίζουμε την μεγαλύτερη τιμή αέρα και στην συνέχεια την χρησιμοποιούμε για να βρούμε την ημέρα

Ερώτηση4: Ποιές ημέρες η θερμοκρασία ήταν μεγαλύτερη από 90 βαθμούς Fahrenheit?

```r
more_than_90 <- subset(airquality, Temp > 90)
print(more_than_90)
```

```
##     Ozone Solar.R Wind Temp Month Day
## 42     NA     259 10.9   93     6  11
## 43     NA     250  9.2   92     6  12
## 69     97     267  6.3   92     7   8
## 70     97     272  5.7   92     7   9
## 75     NA     291 14.9   91     7  14
## 102    NA     222  8.6   92     8  10
## 120    76     203  9.7   97     8  28
## 121   118     225  2.3   94     8  29
## 122    84     237  6.3   96     8  30
## 123    85     188  6.3   94     8  31
## 124    96     167  6.9   91     9   1
## 125    78     197  5.1   92     9   2
## 126    73     183  2.8   93     9   3
## 127    91     189  4.6   93     9   4
```

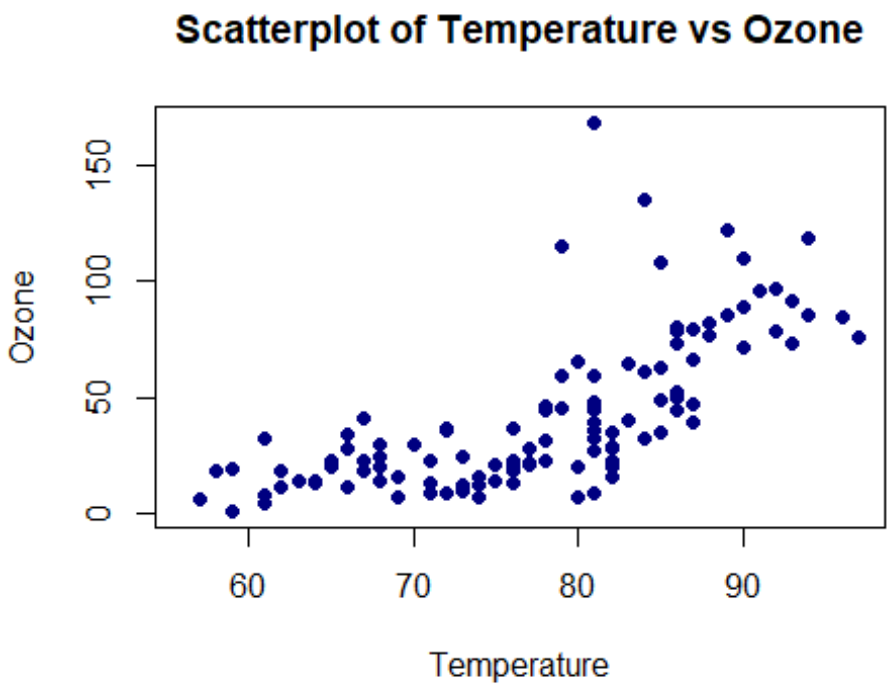
```r
print(more_than_90$Day)
```

```
##  [1] 11 12  8  9 14 10 28 29 30 31  1  2  3  4
```

Αποθηκεύουμε στην hotter_than_90 τις τιμές που είναι μεγαλύτερες του 90. Εμφανίζουμε τις αποθηκευμένες τιμές και επαληθεύουμε τις τιμές. Για τις τιμές αυτές εμφανίζουμε τις ημέρες.

Scaterplot

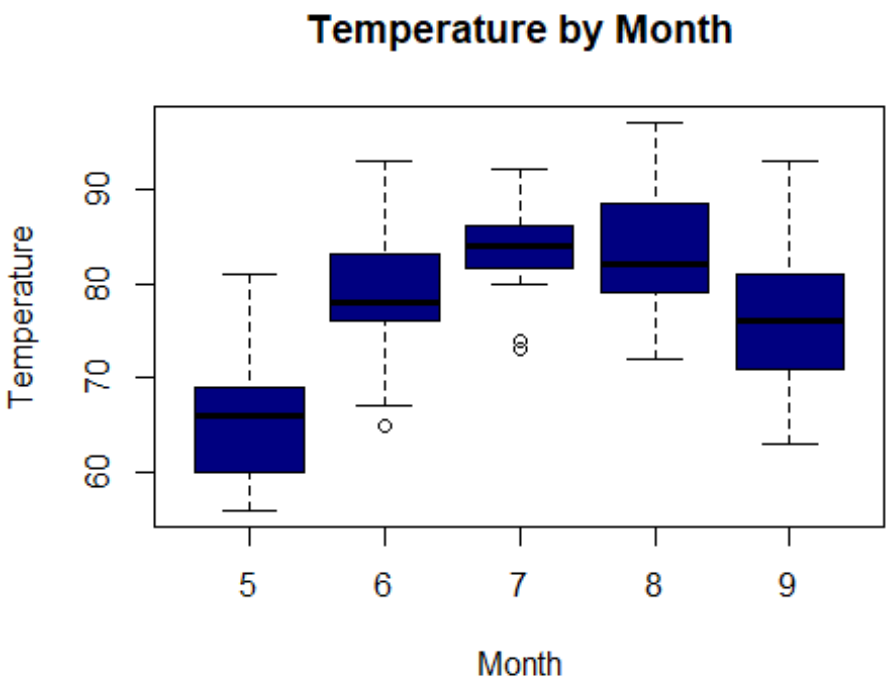
```
plot(x = airquality$Temp, y = airquality$Ozone, main = "Scatterplot of
Temperature vs Ozone", xlab = "Temperature", ylab = "Ozone",
pch = 16,
col = "navy")
```



Έχουμε θετική συσχέτιση στο διάγραμμα δηλαδή μια αύξηση των επιπέδων του όζοντος καθώς αυξάνεται η θερμοκρασία και όσο αυξάνεται η θερμοκρασία, έχουμε λάποιες τιμές που είναι πολύ μεγαλύτερες από το "φυσιολογικό".

Boxplot

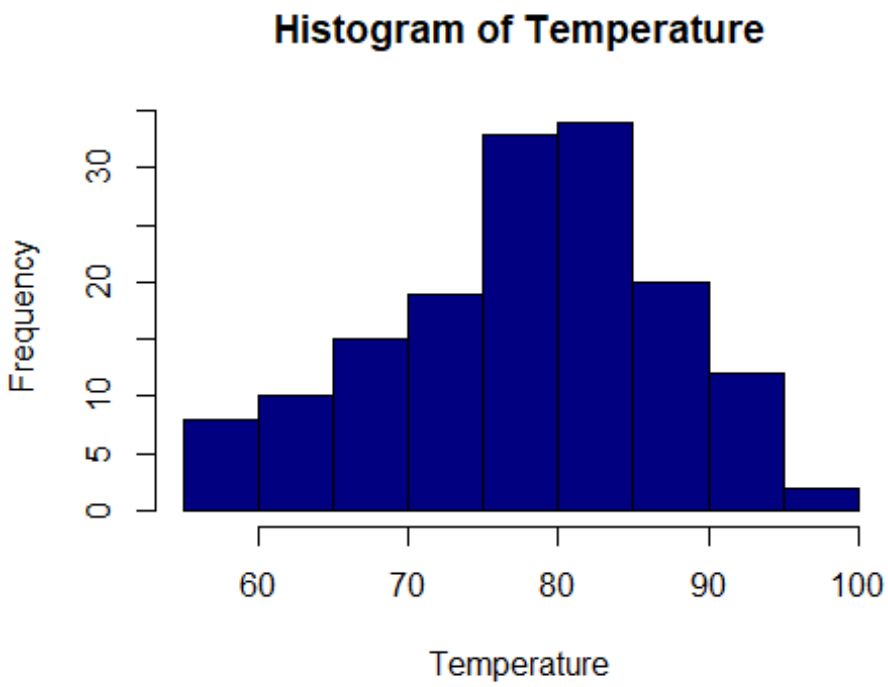
```
boxplot(Temp ~ Month,  data = airquality,  main = "Temperature by Month
", xlab = "Month", ylab ="Temperature",col="navy")
```



Παρατηρούμε στο διπλανό διάγραμμα τις θερμοκρασίες μεταξύ των 5 μηνών 5-9. Ο 5ος και ο 9ος μήνας δίχνουν ένα μεγαλύτερο διάστημα θερμοκρασιών(Μαιος,Σεπτεμβριος) και ο 8ος τις μεγαλυτερες θερμοκρασιες(Αυγουστος).

Histgram

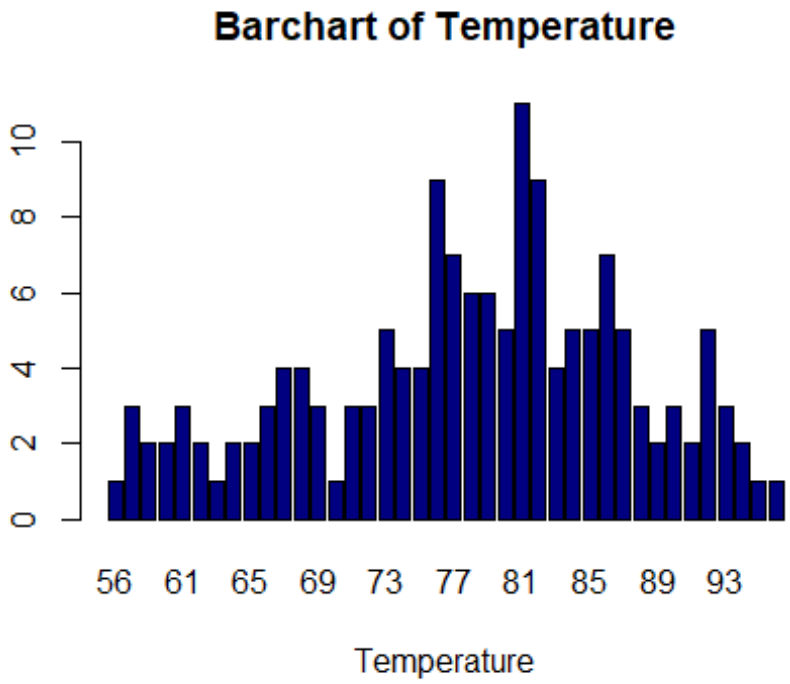
```
hist(airquality$Temp,
     main="Histogram of Temperature",
     xlab="Temperature",
     col="navy")
```



Βλέπουμε τις τιμές των θερμοκρασιών στον άξονα x και την συχνότητα με την οποια εμφανίζονται στον άξονα y και παρατηρούμε ότι οι περισότερες τιμές μαζεύονται γύρω από 70-90

Barchart

```
barplot(table(airquality$Tem),
 main="Barchart of Temperature",xlab="Temperature",col="navy")
```



Στο διπλανό διάγραμμα βλέπουμε στον άξονα x τις τιμές των θερμοκρασιών και στον άξονα y το πλήθος για κάθε θερμοκρασία