

# Ανάλυση Δεδομένων Διαμαντιών

Τσολακίδης

3/4/2024

## Table of Contents

|   |   |
|---|---|
| Ασκηση 1. Ανάλυση Δεδομένων Διαμαντιών .....                                | 1 |
| Φόρτωση του Dataset .....   | 1 |
| Ασκηση 2. Ανάλυση του dataset 'airquality' .....                            | 2 |
| Ποιά είναι η μέση τιμή της θερμοκρασίας για τη δεδομένη περίοδο; .....      | 2 |
| Ποιά ημέρα ήταν η θερμότερη; .....  | 2 |
| Ποιά είχε τον πολύ αέρα; .....  | 3 |
| Ποιές ημέρες η θερμοκρασία ήταν μεγαλύτερη από 90 βαθμούς Fahrenheit? ..... | 3 |

## Ασκηση 1. Ανάλυση Δεδομένων Διαμαντιών

Σε αυτή την ανάλυση θα μελετήσουμε το dataset διαμαντιών που περιέχεται στο πακέτο ggplot2.

### Φόρτωση του Dataset

```
# Φορτώνουμε τη βιβλιοθήκη ggplot2 για τη δημιουργία γραφημάτων
library(ggplot2)

# Αποθηκεύουμε τις στήλες 'carat' και 'price' σε ξεχωριστές μεταβλητές
carat <- diamonds$carat
price <- diamonds$price

# Υπολογίζουμε τη συσχέτιση μεταξύ των μεταβλητών 'carat' και 'price'
correlation <- cor(carat, price)

# Εκτυπώνουμε τη συσχέτιση
print(paste("Correlation between carat and price: ", correlation))

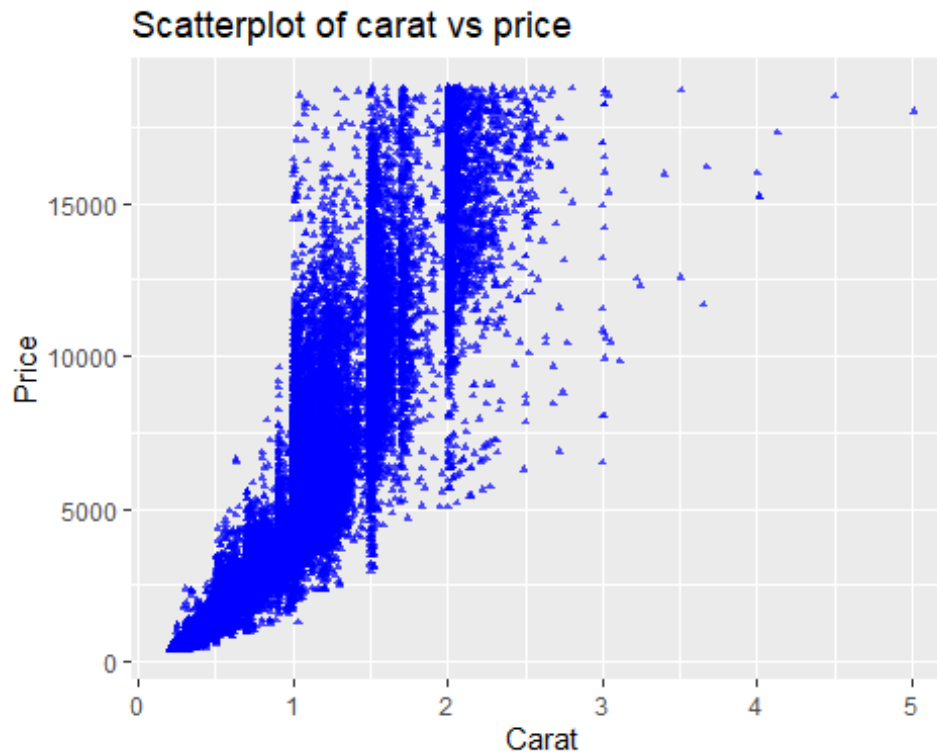
## [1] "Correlation between carat and price: 0.921591301193477"

# Δημιουργούμε ένα scatterplot των μεταβλητών 'carat' και 'price' με χρήση
της ggplot2
ggplot(diamonds, aes(x = carat, y = price)) +
  geom_point(color = "blue", # Χρώμα σημείων
             shape = 17,     # Τύπος σημείων (τετράγωνα)
```

```

size = 1,          # Μέγεθος σημείων
alpha = 0.7) +     # Διαφάνεια σημείων
labs(x = "Carat", y = "Price", title = "Scatterplot of carat vs price") #
Ορίζουμε τις ετικέτες των αξόνων και τον τίτλο του γραφήματος

```



Στις περισσότερες περιπτώσεις όσο αυξάνονται τα καράτια αυξάνεται και η τιμή, χωρίς αυτό να είναι απόλυτο

“Be alone, that is the secret of invention; be alone, that is when ideas are born.”  
- Nikola Tesla

## Ασκηση 2. Ανάλυση του dataset ‘airquality’

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

```

mean_temp <- mean(airquality$Temp)
cat("Η Μέση τιμή θερμοκρασίας είναι:", mean_temp, "Fahrenheit \n")

```

## Η Μέση τιμή θερμοκρασίας είναι: 77.88235 Fahrenheit

Ποιά ημέρα ήταν η θερμότερη;

```

hotterst_day <- airquality$Day[which.max(airquality$Temp)]
hotterst_month <- airquality$Month[which.max(airquality$Temp)]
cat("Η ημέρα με την υψηλότερη θερμοκρασία ήταν η ", hotterst_day, "/",
hotterst_month, "\n")

```

## Η ημέρα με την υψηλότερη θερμοκρασία ήταν η 28 / 8

Ποιά είχε τον πολύ αέρα;

```
windiest_day <- airquality$Day[which.max(airquality$Wind)]
windiest_month <- airquality$Month[which.max(airquality$Wind)]
cat("Η ημέρα με τον περισσότερο αέρα ήταν η", windiest_day, "/",
windiest_month, "\n")
```

## Η ημέρα με τον περισσότερο αέρα ήταν η 17 / 6

Ποιές ημέρες η θερμοκρασία ήταν μεγαλύτερη από 90 βαθμούς Fahrenheit?

```
days_over_90 <- airquality$Day[which(airquality$Temp > 90)]
days_over_90_month <- airquality$Month[which(airquality$Temp > 90)]
for (i in 1:length(days_over_90)){
  cat(days_over_90[i], "/", days_over_90_month[i], ", ")
}
```

## 11 / 6 , 12 / 6 , 8 / 7 , 9 / 7 , 14 / 7 , 10 / 8 , 28 / 8 , 29 / 8 , 30 / 8 , 31 / 8 , 1 / 9 , 2 / 9 , 3 / 9 , 4 / 9 ,

*# Get basic statistics about temperature*

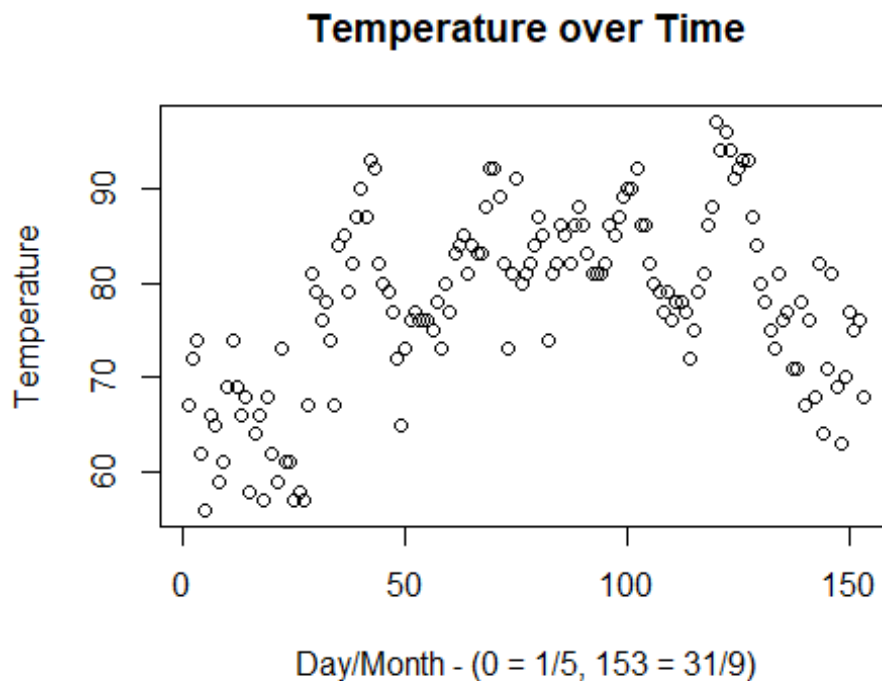
```
temp_length <- length(airquality$Temp)
```

*# Create a sequence of days for the time series plot*

```
days <- seq(1, temp_length)
```

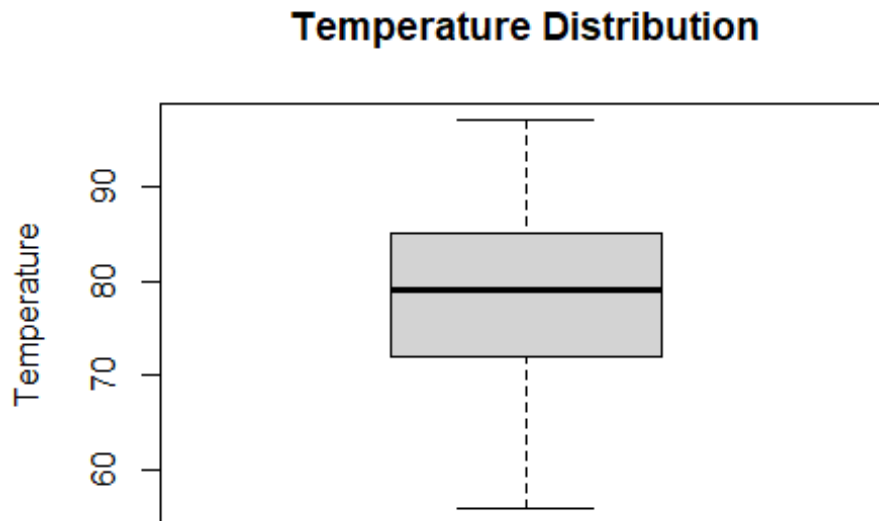
*# Plot the time series of temperature*

```
plot(days, airquality$Temp, xlab = "Day/Month - (0 = 1/5, 153 = 31/9)", ylab = "Temperature", main = "Temperature over Time")
```



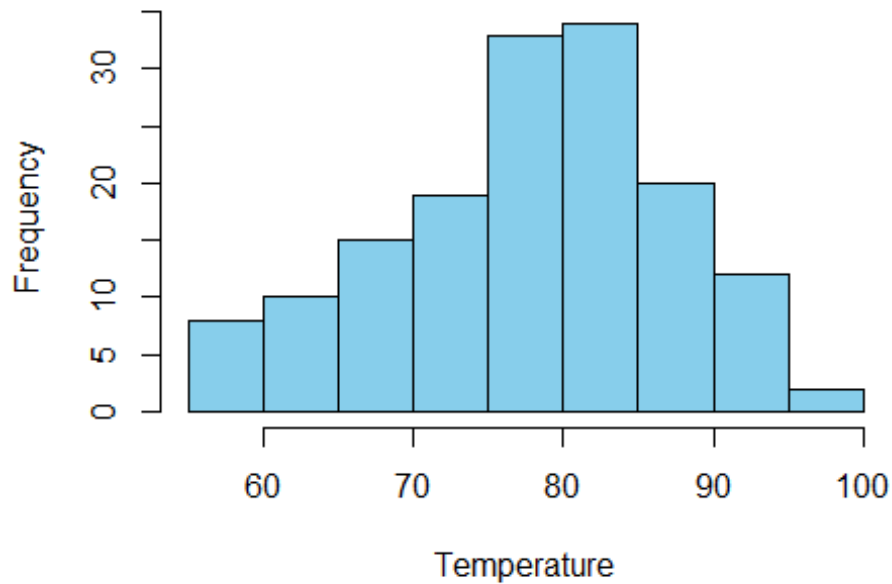
Οι υψηλότερες θερμοκρασίες είναι τους μήνες 7-8

```
# Create a boxplot of temperature  
boxplot(airquality$Temp, main = "Temperature Distribution", ylab =  
"Temperature")
```



```
# Create a histogram of temperature  
hist(airquality$Temp, main = "Temperature Histogram", xlab = "Temperature",  
ylab = "Frequency", col = "skyblue")
```

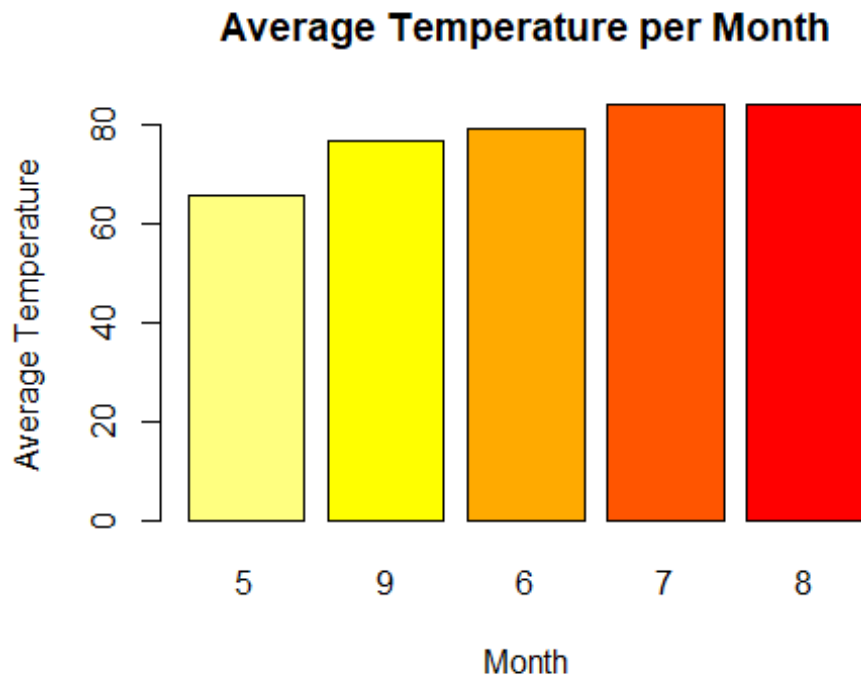
## Temperature Histogram



```
# calculate the mean per month
avg_temp_per_month <- tapply(airquality$Temp, airquality$Month, mean)

# Sort average temperatures and months
sorted_avg_temp <- avg_temp_per_month[order(avg_temp_per_month)]

# Create a bar chart of average temperature per month
barplot(sorted_avg_temp, main = "Average Temperature per Month",
  xlab = "Month", ylab = "Average Temperature",
  col = rev(heat.colors(length(sorted_avg_temp))))
```



*Ο 8ος μήνας φαίνεται και ο θερμότερος*

“Life is like riding a bicycle. To keep your balance, you must keep moving.”  
- Albert Einstein