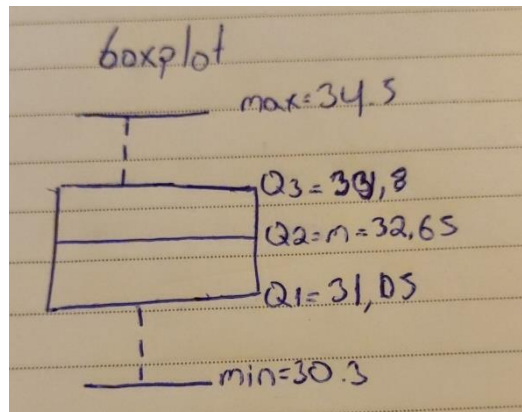


## ΑΣΚΗΣΗ 1

### Α Ερώτημα

Δεδομένα I

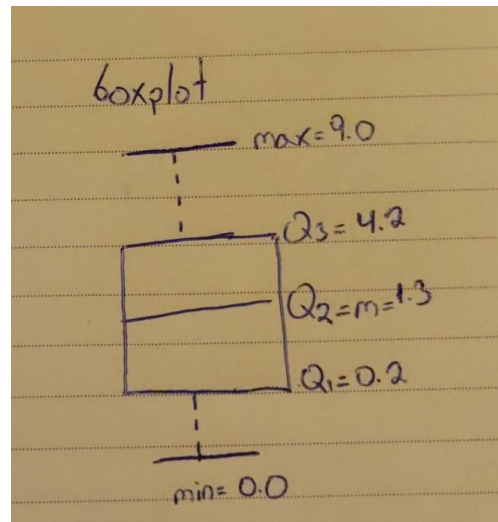
30	3
31	0 1
32	1 6 7
33	4 6
34	2 5



Δεδομένα II

stemplot

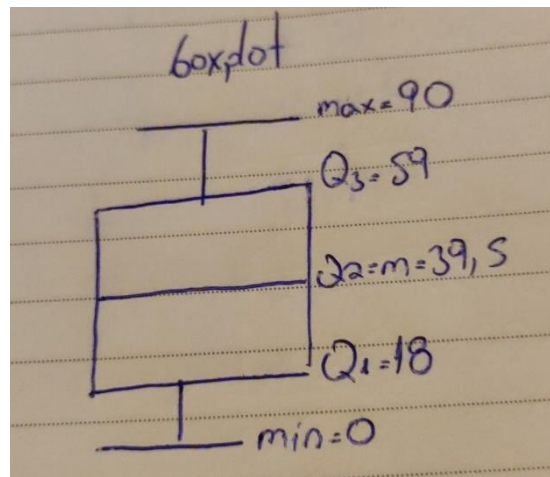
0	0 0 2 8
1	2 4
3	2
4	2
6	4
9	0



Δεδομένα III

stemplot

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



Δεν βρέθηκαν τυπικές τιμές χρησιμοποιώντας τους τύπους:

UPPER Fence:  $Q_3 + 1.5(Q_3 - Q_1)$

LOWER Fence:  $Q_3 - 1.5(Q_3 - Q_1)$  δηλαδή (+ή -) το  $1.5 * IQR$

## **Β Ερώτημα**

- Το σύνολο των δεδομένων I μπορούν να αναπαρασταθούν επαρκώς και με τους δύο τρόπους αλλά εμείς θεωρήσαμε βέλτιστο αυτόν του συνδυασμού της **Μέσης Τιμής και Τυπικής Απόκλισης** αφού τα Δεδομένα I είναι ομοιόμορφα κατανεμημένα με:

$$\text{Μέση Τιμή: } \mu = \frac{325.5}{10} = 32.55$$

$$\text{Τυπική Απόκλιση } \sigma = 1.41$$

- Το σύνολο Δεδομένων II αναπαριστάτε βέλτιστα με την **Σύνοψη των 5 αριθμών** καθώς με αυτό είναι ευδιάκριτο ότι τα στοιχεία έχουν μια κλίση προς τις χαμηλότερες τιμές δημιουργώντας έτσι αραίωση στις υψηλές και μια συσσώρευση στις χαμηλές κάτι που φαίνεται από την τιμή της Διαμέσου.

$$\left. \begin{array}{l} \text{Μέση Τιμή: } \mu = \frac{26.4}{10} = 2.65 \\ \text{Τυπική Απόκλιση } \sigma = 3.05 \end{array} \right\} \mu < \sigma$$

- Το σύνολο Δεδομένων III ερμηνεύονται και με τους δύο τρόπους καλά αλλά πιστεύουμε βέλτιστα με την **Σύνοψη των 5 αριθμών** αφού είναι πιο περιγραφική και ειδικά με την αναπαράσταση boxplot φαίνεται ξεκάθαρα η πραγματική κατανομή των τιμών.

$$\text{Μέση Τιμή: } \mu = \frac{1646}{40} = 41.15$$

$$\text{Τυπική Απόκλιση } \sigma = 28.26$$

## **Γ Ερώτημα**

### **Normal Quantile Plot: Δεδομένων I**

Βήμα 1: Διάταξη των Δεδομένων I σε αύξουσα σειρά

$$X_1 = 30$$

$$X_2 = 31$$

$$X_3 = 31$$

$$X_4 = 32$$

$$X_5 = 33$$

$$X_6 = 33$$

$$X_7 = 33$$

$$X_8 = 34$$

$$X_9 = 34$$

$$X_{10} = 35$$

Βήμα 2: Εύρεση των επί της εκατό τιμών αριστερά κάθε τιμής (δηλαδή τα ποσοστιμότητα):

$$X_1 = 0\%$$

$$X_2 = 10\%$$

$$X_3 = 20\%$$

$$X_4 = 30\%$$

$$X_5 = 40\%$$

$$X_6 = 50\%$$

$$X_7 = 60\%$$

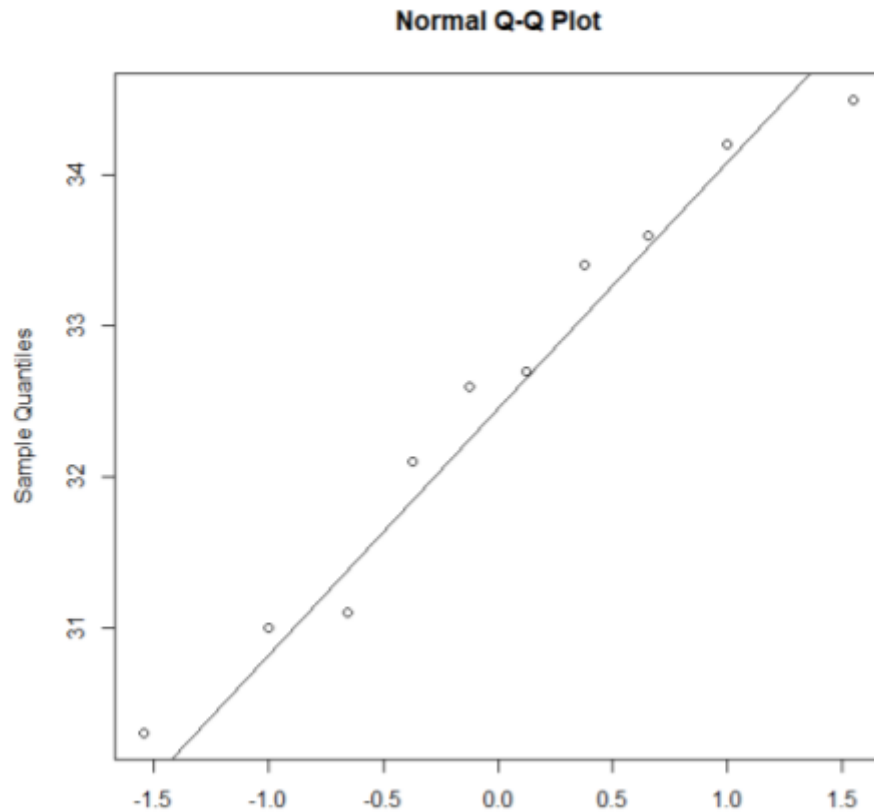
$$X_8 = 70\%$$

$$X_9 = 80\%$$

$$X_{10} = 90\%$$

Βήμα 3: Υπολογισμός των  $X_1 < X_2 < \dots < X_{10}$  όπου  $X_1 = P_1, X_2 = P_2, \dots, X_{10} = P_{10}$  ποσοστημότητα κανονικής κατανομής.

Βήμα 4: Τοποθετούμε τις τιμές των δεδομένων στον κάθετο άξονα του y και τις τιμές των ποσοστημορίων της κανονικής κατανομής στον οριζόντιο άξονα X.



Από το σχήμα των σημείων στο επίπεδο μπορούμε να συμπεράνουμε πως έχοντας ίσως παραπάνω παρατηρήσεις δηλαδή πλήθος δεδομένων θα φαινόταν ακόμα πιο ισχυρή η σχέση με τον άξονα  $x = y$ . **Με άλλα λόγια είναι ξεκάθαρο πως τα Δεδομένα I θα προσέγγιζαν επαρκώς την καμπύλη πυκνότητας της κανονικής κατανομής.**

### Normal Quantile Plot: Δεδομένων II

Βήμα 1: Διάταξη των Δεδομένων I σε αύξουσα σειρά

$$X_1 = 0$$

$$X_2 = 0$$

$$X_3 = 0$$

$$X_4 = 1$$

$$X_5 = 1$$

$$X_6 = 1$$

$$X_7 = 3$$

$$X_8 = 4$$

$$X_9 = 6$$

$$X_{10} = 9$$

Βήμα 2: Εύρεση των επί της εκατό τιμών αριστερά κάθε τιμής (δηλαδή τα ποσοστημόρια):

$$X_1 = 0\%$$

$$X_2 = 10\%$$

$$X_3 = 20\%$$

$$X_4 = 30\%$$

$$X_5 = 40\%$$

$$X_6 = 50\%$$

$$X_7 = 60\%$$

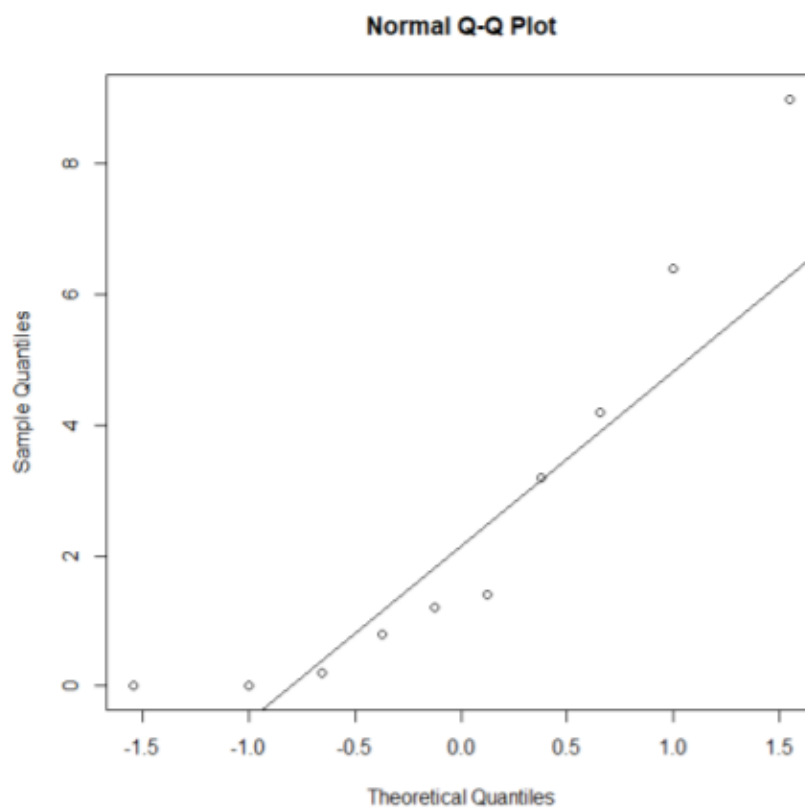
$$X_8 = 70\%$$

$$X_9 = 80\%$$

$$X_{10} = 90\%$$

Βήμα 3: Υπολογισμός των  $X_1 < X_2 < \dots < X_{10}$  όπου  $X_1 = P_1, X_2 = P_2, \dots, X_{10} = P_{10}$  ποσοστημόρια κανονικής κατανομής.

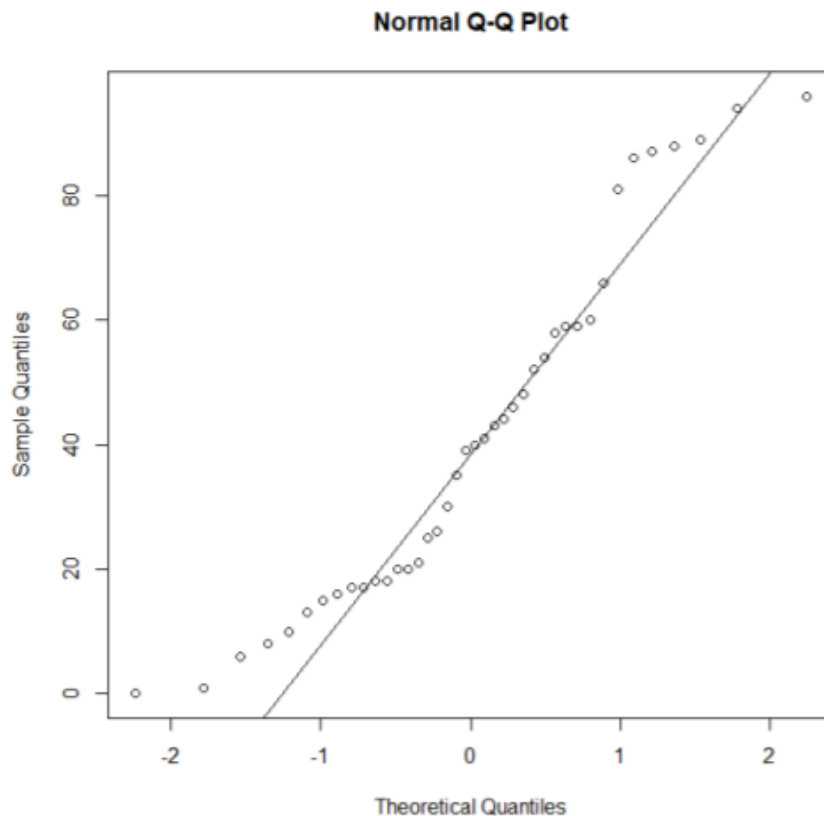
Βήμα 4: Τοποθετούμε τις τιμές των δεδομένων στον κάθετο άξονα του y και τις τιμές των ποσοστημορίων της κανονικής κατανομής στον οριζόντιο άξονα X.



Κατά την γνώμη μας τα δεδομένα αυτά **δεν προσεγγίζουν την καμπύλη πυκνότητας της κανονικής κατανομής επαρκώς με το συγκεκριμένο πλήθος δεδομένων**. Μπορεί όμως με τις διπλάσιες ακόμα μόνο παρατηρήσεις να μπορούσαμε να δούμε μια κάπως πιο ισχυρή σχέση.

### Normal Quantile Plot: Δεδομένων III

Για τα Δεδομένα III ακολουθώντας τον ίδιο αλγόριθμο μπορούμε να παράγουμε το ακόλουθο Normal-Quantile plot:



Με μία πρώτη ματιά στο Normal-Quantile plot των Δεδομένων III θα μπορούσε κάποιος να πει πως το σχήμα των στοιχείων προσεγγίζουν την μορφή μια ευθείας γραμμής, όμως παρατηρώντας τα λίγο πιο προσεκτικά μπορούμε να δούμε πως στις τιμές στο πρώτο τέταρτο του άξονα x υπάρχουν όχι απλά ατυπικές τιμές αλλά pattern από στοιχεία τα οποία αποκλίνουν από την ευθεία. Ακόμα φαίνεται το σχήμα των στοιχείων να έχει μια ελικοειδή μορφή που καθιστά ακόμα δυσκολότερη την ανάλυση, μιας και οι ελικοειδείς μορφές γενικά ακολουθούν την φορά ευθείας όμως με ένα μικρό εμβαδόν (ενός υποθετικού ελικοειδές σχήματος) να επικαλύπτεται με την ευθεία αυτή, μιας και το μεγαλύτερο αποκλίνει δημιουργώντας τα ημικύκλια. **Συνεπώς κατά την γνώμη μας τα Δεδομένα III δεν προσεγγίζουν την καμπύλη πυκνότητας της κανονικής κατανομής επαρκώς.**

## ΑΣΚΗΣΗ 2

### Α Ερώτημα

Τα στατιστικά δεδομένα προέρχονται από την ιστοσελίδα (<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>) η οποία ανηκει στην πολη της Νεας Υορκης και αποθηκευει και διαδιδει τα δεδομενα των διαδρομων ταξι της πολης κάθε μηνια. Όπως εξηγείται καλύτερα και στον οδηγό χρηστών στο [trip record user guide.pdf](#)

« The TLC & Data The New York City Taxi and Limousine Commission (TLC), created in 1971, is the agency responsible for licensing and regulating New York City's medallion (yellow) taxis, street hail livery (green) taxis, for-hire vehicles (FHVs), commuter vans, and paratransit vehicles. The TLC collects trip record information for each taxi and for-hire vehicle trip completed by our licensed drivers and vehicles. We receive taxi trip data from the technology service providers (TSPs) that provide electronic metering in each cab, and FHV trip data from the app, community livery, black car, or luxury limousine company, or base, who dispatched the trip. In each trip record dataset, one row represents a single trip made by a TLC-licensed vehicle ».

Τα δεδομένα που χρησιμοποιούμε στην εργασία μας είναι από τον Απρίλη του 2020 και αφορούν τα ταξί με χρώμα πρασινο (ή αλλιώς boro cab) και όπως εξηγείται καλύτερα στο pdf

«Green taxis, also known as boro taxis and street-hail liveries, were introduced in August of 2013 to improve taxi service and availability in the boroughs. Green taxis may respond to street hails, but only in the areas indicated in green on the map (i.e. above W 110 St/E 96th St in Manhattan and in the boroughs). Records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. As with the yellow taxi data, these records were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology service providers. The trip data was not created by the TLC, and TLC cannot guarantee their accuracy».

## **B Ερώτημα**

Τα δεδομένα μας αποτελούνται από 35612 εγγραφές και 20 μεταβλητές: 10 Κατηγορικές και 10 ποσοτικές και υπάρχει ακριβής περιγραφή τους στο [data dictionary trip records green.pdf](#).

Για παραδειγμα η μεταβλητή **VendorID** είναι κατηγορική και δηλώνει τον παροχο της κάθε εγγραφής, δηλαδή αυτόν που έστειλε τα δεδομένα στην TLC (New York City Taxi and Limousine Commission) και αποτελείται από 2 εταιρείες, οι οποίες περνούν ως τιμές στο dataset: 1= Creative Mobile Technologies, LLC; Και 2= VeriFone Inc.

Ακόμη, υπάρχουν και μεταβλητές όπως οι **lpep\_pickup\_datetime** και **lpep\_dropoff\_datetime** οι οποίες είναι κατηγορικές και δηλώνουν την ημερομηνία και την ώρα που έγινε η παραλαβή και η παραδοση αντίστοιχα.

Επειτα υπάρχει και η **Passenger\_count** η οποία είναι ποσοτική επίσης και δηλώνει πόσους επιβάτες είχε το ταξί.

Υπάρχει επίσης και η μεταβλητή **Payment\_type** η οποία είναι κατηγορική και δηλώνει το τρόπο πληρωμής της διαδρομής δηλαδή: 1= Credit card 2= Cash 3= No charge 4= Dispute 5= Unknown 6= Voided trip.

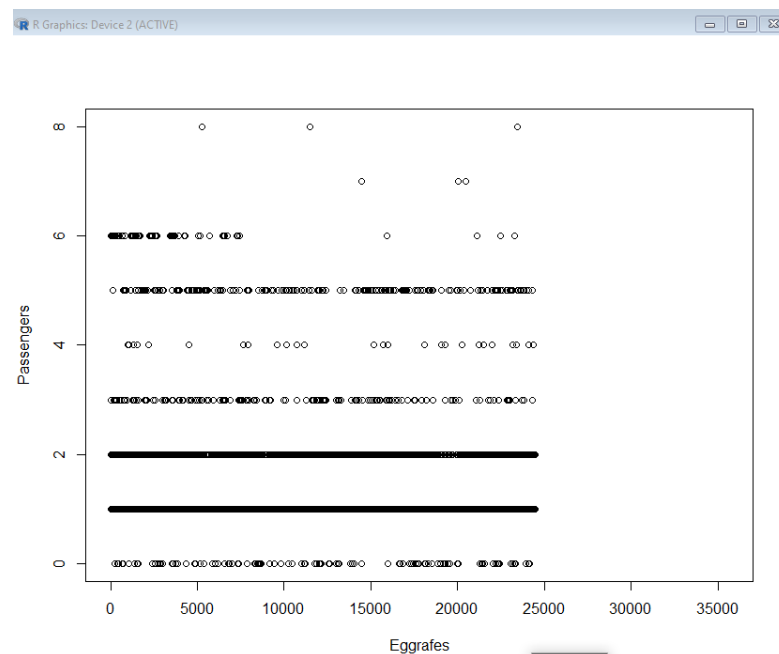
Επίσης, υπάρχουν και μεταβλητές όπως οι **Total\_amount** και **Tip\_amount** οι οποίες είναι ποσοτικές και δηλώνουν το ποσό που χρεώθηκε για τη διαδρομή και το ποσό φιλοδώρηματος αντίστοιχα.

Τελος, υπάρχουν και οι ποσοτικές μεταβλητές **PULocationID** και **DOLocationID** οι οποίες είναι κατηγορικές και δηλώνουν τις περιοχές της παραλαβής και της παραδοσης. Και άλλες.

**Σημείωση:** Λογω του ογκού των δεδομένων ρυθμίσαμε τη προβολή μόνο των 15000 εισοδών με την εντολή: `options(max.print = 15000)`.

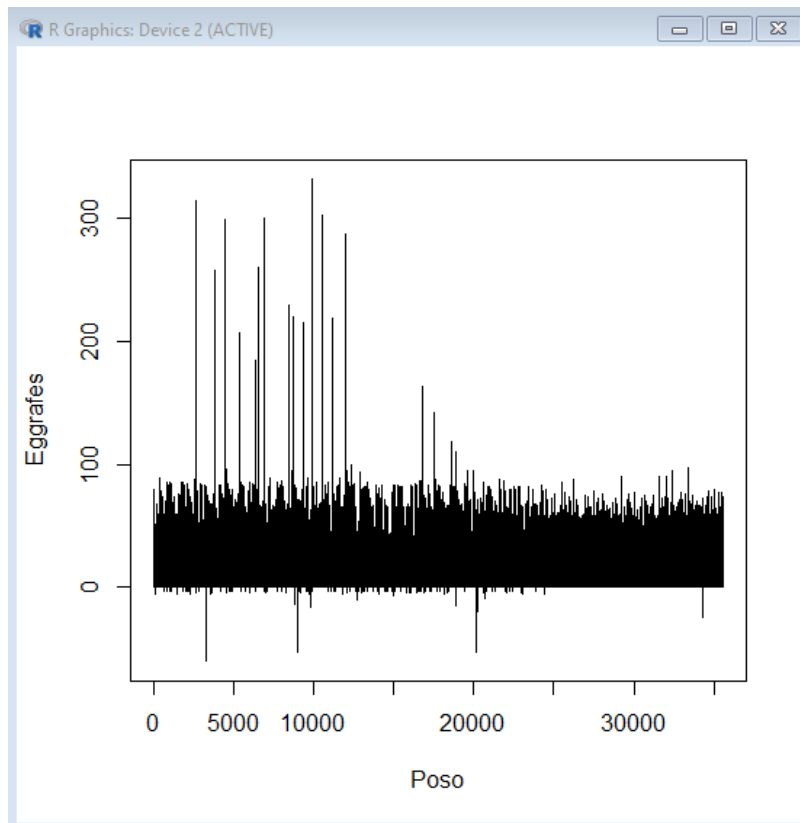
### Γ Ερώτημα

Η κατανομή της μεταβλητής `passenger_count` αποτυπώνεται στη παρακάτω γραφική παραστάση:

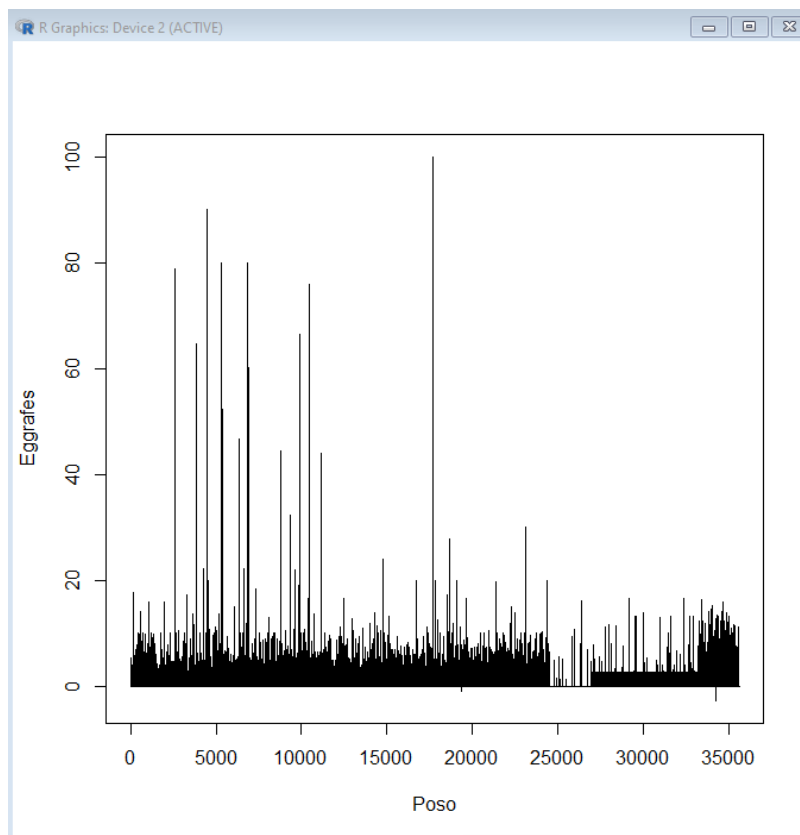


Παρατηρούμε ότι υπάρχουν ατυπικές τιμές στο 0, καθώς δε γίνεται να υπάρχουν 0 επιβατές. Επίσης παρατηρούμε ότι εμφανίζονται συχνότερα είτε 1 είτε 2 επιβατές.

Η κατανομή της μεταβλητής `fare_amount` (The time-and-distance fare calculated by the meter) αποτυπώνεται στη παρακάτω γραφική παραστάση:



Παρατηρούμε μερικά outliers όπου εμφανίζονται αρνητικές τιμές.  
Η κατανομή της μεταβλητής `tip_amount` αποτυπώνεται στη παρακάτω γραφική παρασταση:



Παρατηρούμε ότι οι περισσότεροι επιβάτες δεν δίνουν πάνω από 20 δολλάρια φιλοδώρημα.

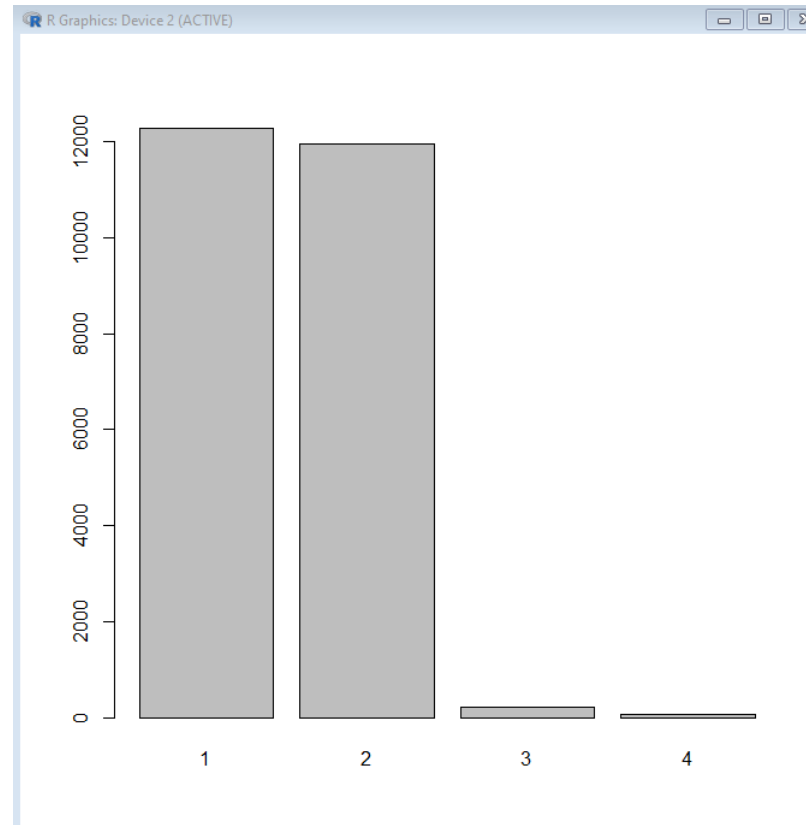


Η κατανομή της κατηγορικής μεταβλητής Payment\_type (A numeric code signifying how the passenger paid for the trip. 1= Credit card 2= Cash 3= No charge 4= Dispute 5= Unknown 6= Voided trip)

δίνεται από το παρακάτω barplot για :

payment\_type

1	2	3	4
12274	11954	216	58

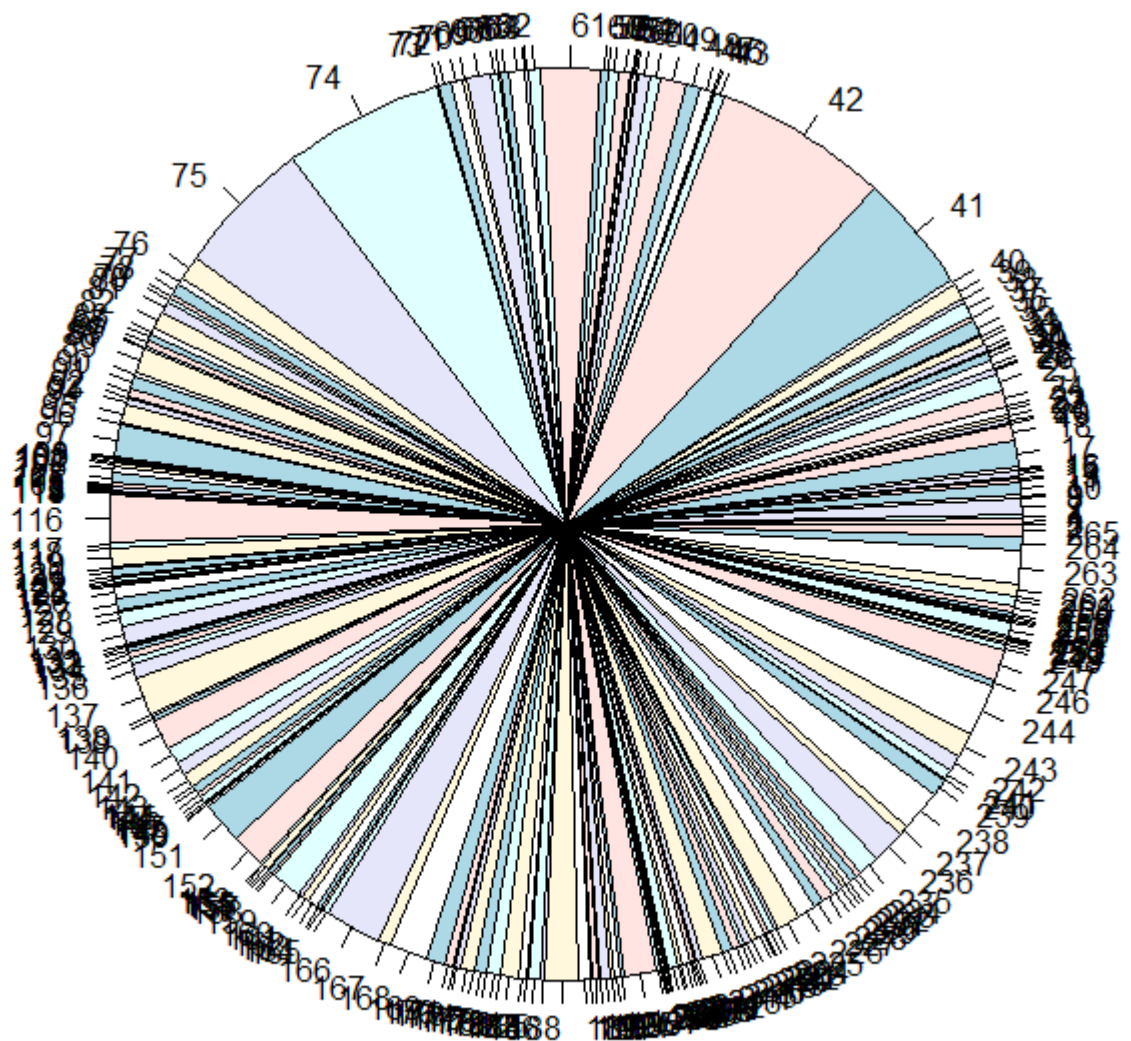


Η κατανομή της κατηγορικής μεταβλητής DOLocationID δηλαδή της περιοχής που κατεβήκε ο επιβατής δίνεται από το παρακάτω πίε για :

Κάθε αριθμός αντιστοιχεί σε διαφορετική γειτονία και οι αντιστοιχισμοί τους δίνονται από το αρχείο taxi+\_zone\_lookup.csv όπου οι συχνότερες περιοχές είναι :

41	"Manhattan"	"Central Harlem"	"Boro Zone"
42	"Manhattan"	"Central Harlem North"	"Boro Zone"
74	"Manhattan"	"East Harlem North"	"Boro Zone"
75	"Manhattan"	"East Harlem South"	"Boro Zone"
116	"Manhattan"	"Hamilton Heights"	"Boro Zone"
166	"Manhattan"	"Morningside Heights"	"Boro Zone"
244	"Manhattan"	"Washington Heights South"	"Boro Zone"
263	"Manhattan"	"Yorkville West"	"Yellow Zone"

Οι οποίες έχουν συχνότητα εμφάνισης πάνω από 600 φορές για τις 35000 εγγραφές με συχνότερη τη 42 όπου εμφανίζεται 2195 φορές.



### Δ Ερώτημα

Ποσοτικές μεταβλητές έχουμε τις εξής:

passenger\_count

```
> fivenum(passenger_count)
[1] 0 1 1 1 8
> sd(passenger_count)
[1] NA
> m <- passenger_count[!is.na(passenger_count)]
> sd(m)
[1] 0.5727962
> mean(m)
[1] 1.119745
> |
```

Καταλληλότεροι αριθμοί για να απααραστήσουν τη μεταβλητή είναι η μεση τιμή και η τυπική αποκλιση καθώς καταλαβαίνουμε καλύτερα που κυμαίνονται οι αριθμοί αφού οι 5 αριθμοί μας δείχνουν πολύ λίγες πληροφορίες.

## trip\_distance

```
> fivenum(trip_distance)
[1] 0.00 1.14 2.41 5.71 55027.35
> sd(trip_distance)
[1] 403.3147
> mean(trip_distance)
[1] 9.390396
> |
```

Η συνοψη των 5 αριθμων είναι καταλληλοτερη για να αναπαραστησει τη μεταβλητη καθως υπαρχουν υψηλες ατυπικες τιμες.

## fare\_amount

```
> fivenum(fare_amount)
[1] -60.00 7.00 11.16 20.50 331.50
> sd(fare_amount)
[1] 15.25576
> mean(fare_amount)
[1] 16.59169
> |
```

---

Η συνοψη των 5 αριθμων είναι καταλληλοτερη για να αναπαραστησει τη μεταβλητη καθως υπαρχουν υψηλες ατυπικες τιμες.

## Extra

```
> fivenum(extra)
[1] -1.0 0.0 0.0 0.0 4.5
> sd(extra)
[1] 0.4869793
> mean(extra)
[1] 0.2235623
> |
```

Καταλληλοτεροι αριθμοι για να απααραστησουν τη μεταβλητη είναι η μεση τιμη και η τυπικη αποκλιση καθως καταλαβαινουμε καλυτερα που κυμαινονται οι αριθμοι καθως οι 5 αριθμοι δινουν λιγες πληροφοριες.

## mta\_tax

```
> fivenum(mta_tax)
[1] -0.5 0.0 0.5 0.5 0.5
> sd(mta_tax)
[1] 0.2397702
> mean(mta_tax)
[1] 0.3314894
> |
```

---

Καταλληλότεροι αριθμοί για να απαπαραστήσουν τη μεταβλητή είναι η μέση τιμή και η τυπική αποκλίση καθώς καταλαβαίνουμε καλύτερα που κυμαίνονται οι αριθμοί καθώς οι 5 αριθμοί δίνουν λίγες πληροφορίες.

## tip\_amount

```
> fivenum( tip_amount )  
[1] -2.75  0.00  0.00  2.75 100.00  
> sd( tip_amount )  
[1] 2.26926  
> mean( tip_amount )  
[1] 1.324754  
> |
```

Καταλληλότεροι αριθμοί για να απαπαραστήσουν τη μεταβλητή είναι η μέση τιμή και η τυπική αποκλίση καθώς καταλαβαίνουμε καλύτερα που κυμαίνονται οι αριθμοί καθώς οι 5 αριθμοί δίνουν λίγες πληροφορίες.

## tolls\_amount

```
> fivenum( tolls_amount )  
[1] 0.00 0.00 0.00 0.00 28.34  
> sd( tolls_amount )  
[1] 1.953908  
> mean( tolls_amount )  
[1] 0.5721726  
> |
```

Καταλληλότεροι αριθμοί για να απαπαραστήσουν τη μεταβλητή είναι η μέση τιμή και η τυπική αποκλίση καθώς καταλαβαίνουμε καλύτερα που κυμαίνονται οι αριθμοί καθώς οι 5 αριθμοί δίνουν λίγες πληροφορίες.

## improvement\_surcharge

```
## ----  
> fivenum(improvement_surcharge)  
[1] -0.3 0.3 0.3 0.3 0.3  
> sd(improvement_surcharge)  
[1] 0.04240906  
> mean(improvement_surcharge)  
[1] 0.2959901  
> |
```

Καταλληλότεροι αριθμοί για να απαπαραστήσουν τη μεταβλητή είναι η μέση τιμή και η τυπική αποκλίση καθώς καταλαβαίνουμε καλύτερα που κυμαίνονται οι αριθμοί καθώς οι 5 αριθμοί δίνουν λίγες πληροφορίες.

## total\_amount

```
> fivenum(total_amount)
[1] -60.300  8.800 13.860 24.315 398.760
> sd(total_amount)
[1] 17.12488
> mean(total_amount)
[1] 19.68444
> |
```

Η συνοψη των 5 αριθμων είναι καταλληλοτερη για να αναπαραστησει τη μεταβλητη καθώς υπάρχουν υψηλες ατυπικες τιμες.

## congestion\_surcharge

```
> fivenum(congestion_surcharge)
[1] 0.00 0.00 0.00 0.00 2.75
> m <- congestion_surcharge[!is.na(congestion_surcharge)]
> sd(congestion_surcharge)
[1] NA
> sd(m)
[1] 1.007563
> mean(m)
[1] 0.4396274
> |
```

Καταλληλοτεροι αριθμοι για να απααραστησουν τη μεταβλητη είναι η μεση τιμη και η τυπικη αποκλιση καθώς καταλαβαινουμε καλυτερα που κυμαινονται οι αριθμοι καθώς οι 5 αριθμοι δινουν λιγες πληροφοριες.

## Ε Ερώτημα

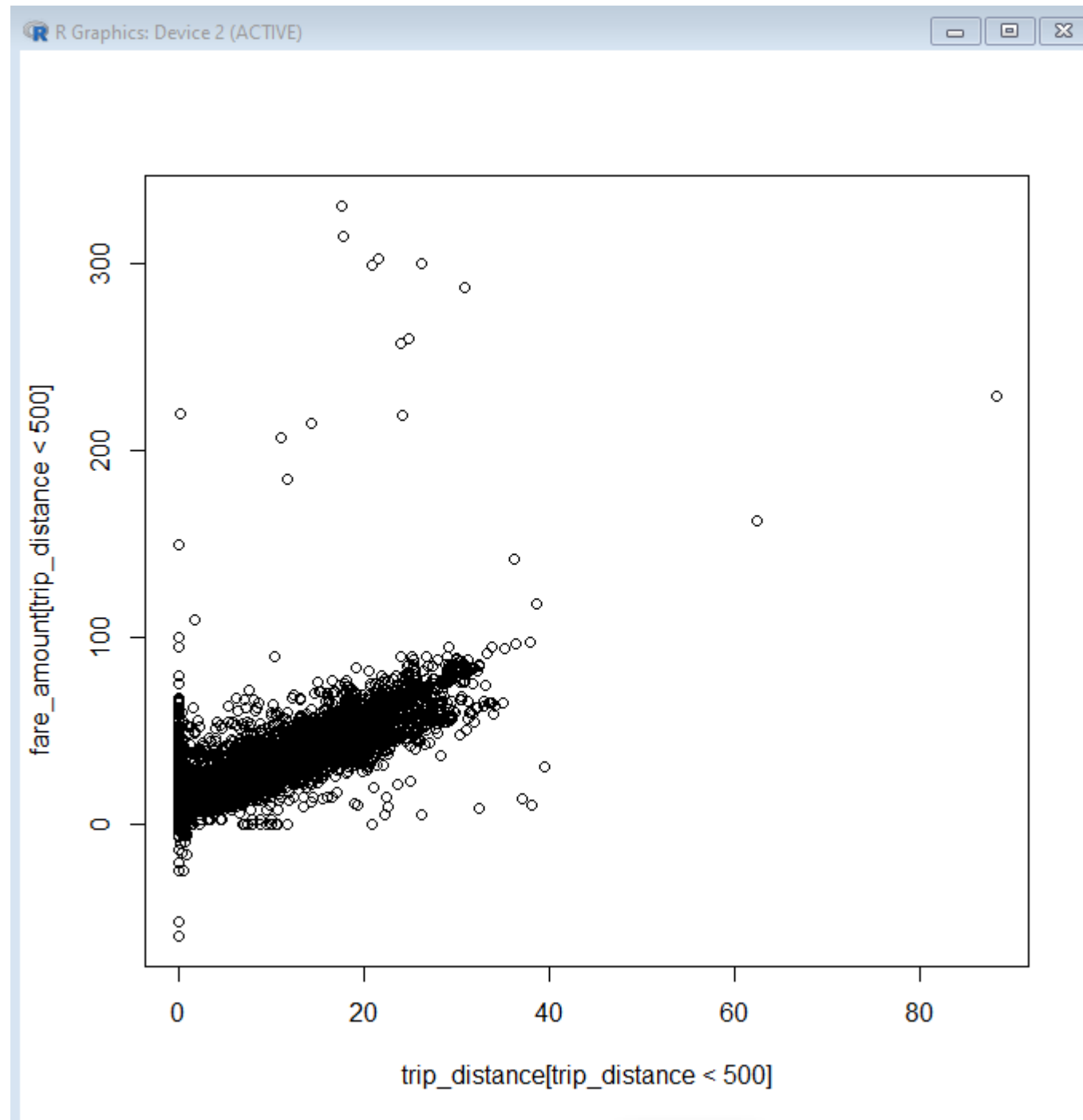
Διερευνουμε τη σχεση του fare\_amount(The time-and-distance fare calculated by the meter) και του trip\_distance και αμα δε λαβουμε υποψη τις ατυπικες τιμες παρατηρουμε ότι υπαρχει συσχετιση αναμεσα στις δυο μεταβλητες.

```
> cor(fare_amount[trip_distance<500],trip_distance[trip_distance<500])
[1] 0.863572
> |
```

Αυτό έχει λογικη εξηγηση αφού στο fare\_amount αποτυπωνεται το κοστος λαμβανοντας υποψη 2 μεταβλητες: τη διαρκεια και την αποσταση του ταξιδιου. Επομενως, η σχεση τους είναι αιτιατη αλλα δεν είναι απολυτα συνδεδεμενες και αυτό το παρατηρουμε με τις ατυπικες τιμες και αυτό συμβαινει γιατι υπαρχει και παραγοντας του χρονου για τον υπολογισμο του κοστους.

**Σημειωση:** Αφαιρουμε τις τιμες για τις οποιες το trip\_distance<500 για να φαινεται το σχημα κανονικα.

Παρακατω Βλεπουμε τη σχεση τους στο σχημα.



### ΑΣΚΗΣΗ 3

#### Α Ερώτημα

Load height in variable x:

```
> x <- data$height
> x
 [1] 1.85 1.54 1.80 1.83 1.73 1.59 1.73 1.58 1.42 1.80 1.70 1.83 1.70 1.83 1.78
[16] 1.94 1.74 1.92 1.90 1.83 1.67 1.60 1.71 1.70 1.80 1.55 1.64 1.78 1.69 1.85
[31] 1.53 1.83 1.73 1.80 1.80 1.64 1.60 1.64 1.55 1.81 1.86 1.82 1.83 1.80 1.90
[46] 1.80 1.78 1.85 1.67 1.86 1.77 1.73 1.80 1.75 1.73 1.83 1.73 1.78 1.82 1.70
[61] 1.79 1.72 1.80 1.74 1.84 1.85 1.72 1.73 1.73 1.75 1.83 1.85 1.80 1.84 1.78
[76] 1.70 1.80 1.85 1.82 1.62 1.77 1.83 1.80 1.77 1.72 1.66 NA 1.60 1.82 1.63
[91] 1.85 1.58 1.79 1.69 1.86 1.76 1.65 2.06 1.90 1.62 1.90 1.75 1.84 1.83 1.74
[106] 1.69 1.82 1.84 1.74 1.98 1.69 1.73 1.56 1.88 1.75 1.83 1.80 1.80 1.88 1.75
[121] 1.68 1.68 1.60
> |
```

Load weight in variable y:

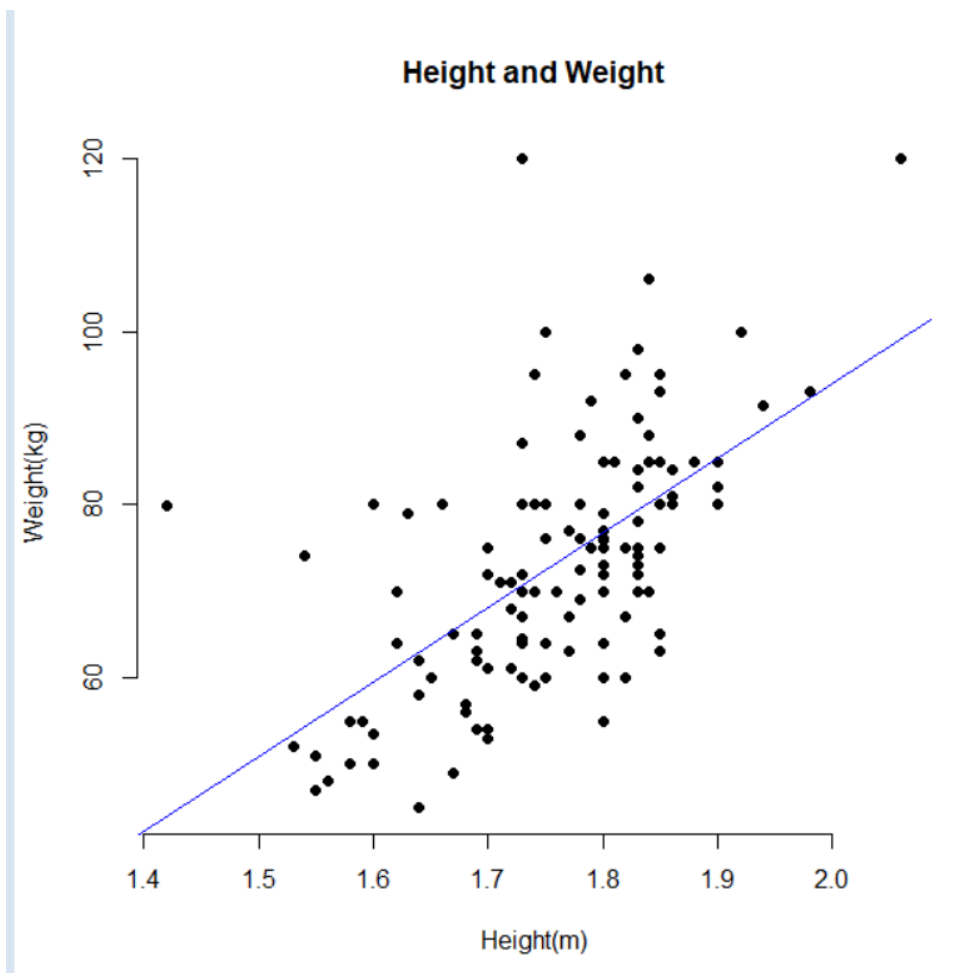
```
> y <- data$weight
> y
 [1] 85.0 74.0 64.0 90.0 120.0 55.0 64.0 55.0 79.8 55.0 75.0 75.0
[13] 54.0 78.0 76.0 91.5 59.0 100.0 85.0 98.0 49.0 50.0 71.0 53.0
[25] 85.0 47.0 62.0 88.0 54.0 80.0 52.0 70.0 70.0 64.0 79.0 58.0
[37] 53.4 45.0 51.0 85.0 80.0 95.0 74.0 70.0 82.0 75.8 69.0 93.0
[49] 65.0 84.0 77.0 80.0    NA 76.0 64.5 72.0 60.0 80.0 60.0 61.0
[61] 75.0 68.0 76.0 95.0 88.0 75.0 61.0 67.0 72.0 64.0 74.0 63.0
[73] 73.0 85.0 72.5 72.0 77.0 65.0 95.0 64.0 67.0 82.0 75.0 63.0
[85] 71.0 80.0    NA 50.0 67.0 79.0 95.0 50.0 92.0 62.0 81.0 70.0
[97] 60.0 120.0 80.0 70.0 80.0 80.0 106.0 73.0 80.0 63.0 75.0 70.0
[109] 70.0 93.0 65.0 87.0 48.0 85.0 100.0 84.0 72.0 60.0 85.0 60.0
[121] 57.0 56.0 80.0
> |
```

Make the scatter plot:

```
plot(x,y, main = "Height and Weight", xlab = "Height(m)", ylab= "Weight(kg)", pch= 16, frame=FALSE)
|
```

Apply Linear Regression:

```
> abline(lm(y~x, data= mtcars),col ="blue")
> |
```



Ως προς την μορφή των δεδομένων, από την παραπάνω κλίμακα φαίνεται να μην μπορεί να απεικονιστεί με σχήμα καθαρής ευθείας αφού υπάρχουν αρκετά outliers.

Ως προς την κατεύθυνση της ευθείας, είναι ξεκάθαρο πως η θεωρητική ευθεία που παριστάνουν έχει μια κλίση που τείνει τις 45ο .

Τέλος η δύναμη της σχέσης των δύο μεταβλητών ( $\chi$  = ύψος και  $y$ = βάρος) είναι θετική, καθώς η συν διακύμανση τους: `> cov(x,y)` δηλαδή κάτι θετικό αλλά και πάλι κάτι πιο κοντά στο 2 παρά στο 1.

```
[1] 1.9031
```

Η δύναμη της σχέσης δεν είναι θετικά ισχυρή.

## Β Ερώτημα

Ο συντελεστής συσχέτισης αφού φορτώσαμε τα δεδομένα στην R.

```
> cor(x,y)
[1] 0.5893806
```

( $\chi$  = ύψος και  $y$ = βάρος). Όπως αναφέρθηκε και παραπάνω η γραμμική παλινδρόμηση έχει εκτελεστεί παραπάνω μαζί με την απάντηση του ερωτήματος Α.