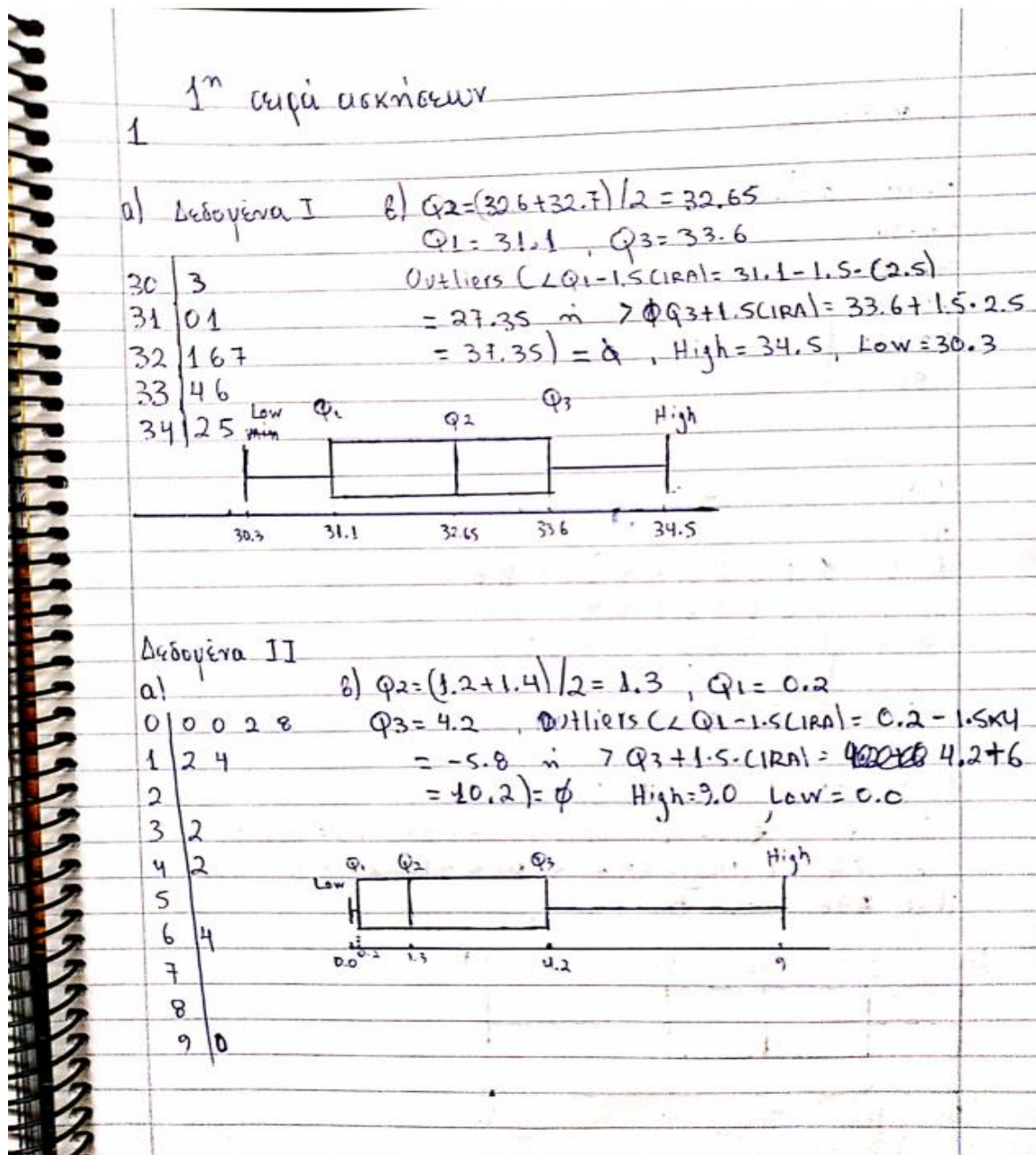


ΑΣΚΗΣΗ 1

a)



Δεδομένα 111

a)

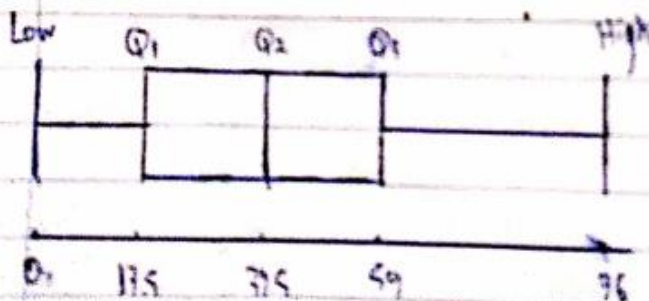
0-1	0 1 6 8 0 3 5 6 7 7 8 8
2-3	0 0 1 5 6 0 5 9
4-5	0 1 3 4 6 8 2 4 8 9 9
6-7	0 6
8-9	1 6 7 8 9 4 6

b) $Q_2 = 39.5$ $Q_1 = 17.5$ $Q_3 = 59$

Outliers $(Q_1 - 1.5 \cdot IQR) = 17.5 - 1.5 \cdot 41.5 = -44.75$

$(Q_3 + 1.5 \cdot IQR) = 59 + 1.5 \cdot 41.5 = 121.25 = \phi$

High = 96, Low = 0



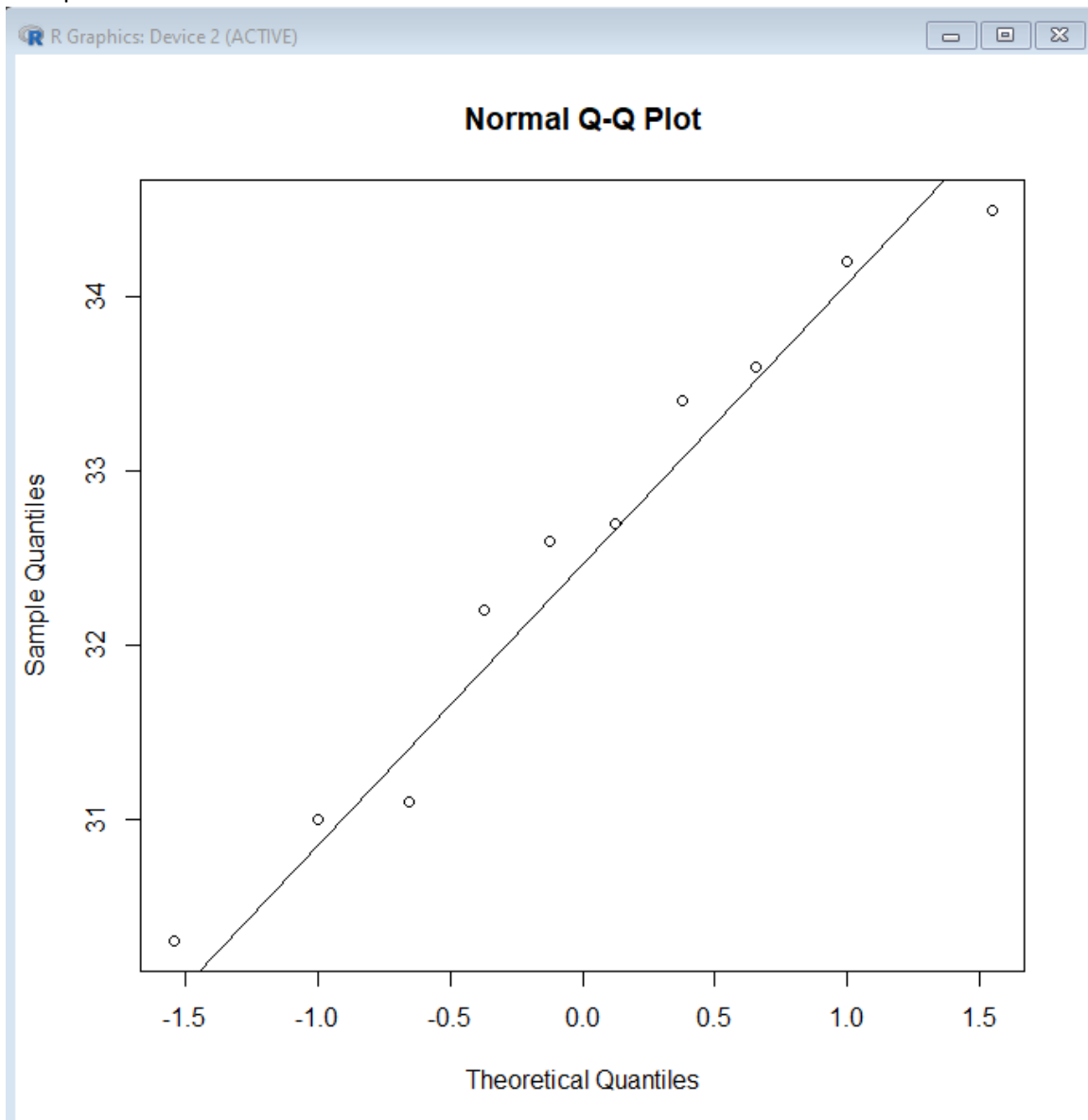
b) Για τα δεδομένα 1 παρατηρούμε ότι είναι αρκετά ομοιόμορφα κατανομημένα (τυπική αποκλίση=1.41 και μέση τιμή=32.55) άρα αυτές οι 2 τιμές αρκούν για την περιγραφή τους

Για τα δεδομένα 2 παρατηρούμε ότι οι μικρότερες τιμές είναι κατανομημένες «κοντά» σε σχέση με τις μεγαλύτερες τιμές (τυπική αποκλίση=3.059121 και μέση τιμή=2.64) άρα θα επιλέξουμε την αναπαράσταση με τους 5 αριθμούς

Για τα δεδομένα 3 τα πράγματα είναι πιο απλά. Έχουμε μεγαλύτερο εύρος τιμών και αρκετά μεγαλύτερη απόκλιση ((τυπική αποκλίση=28.26754 και μέση τιμή=41.15) άρα και εδώ θα επιλέξουμε την αναπαράσταση με τους 5 αριθμούς

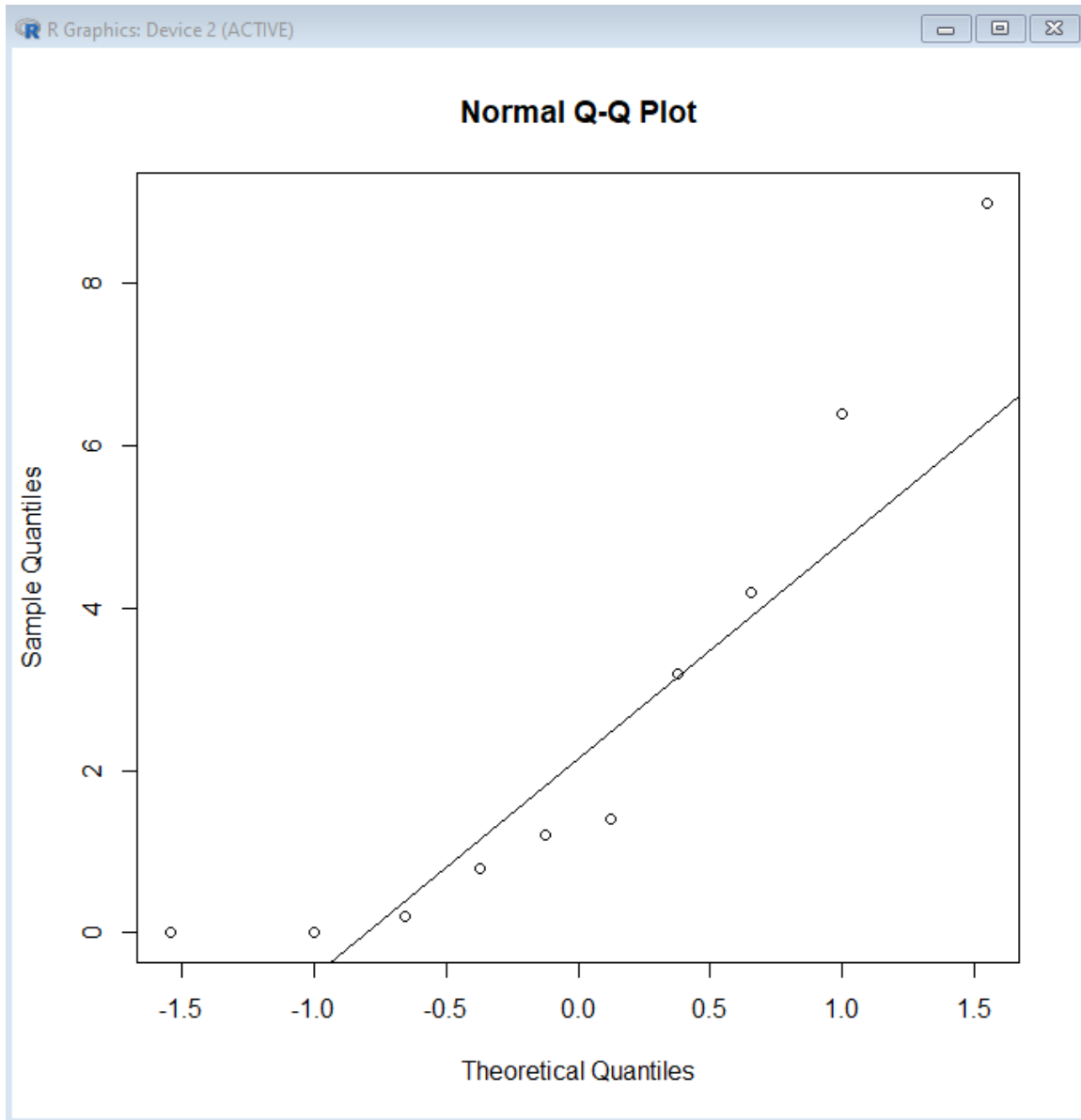
γ)Κάναμε σύγκριση των κατανομών με την κανονική με normal quantile plot(Διαφάνεια 1 σελίδα 34 ο «αλγόριθμος»).Με την χρήση της R έχουμε:

Δεδομένα 1:



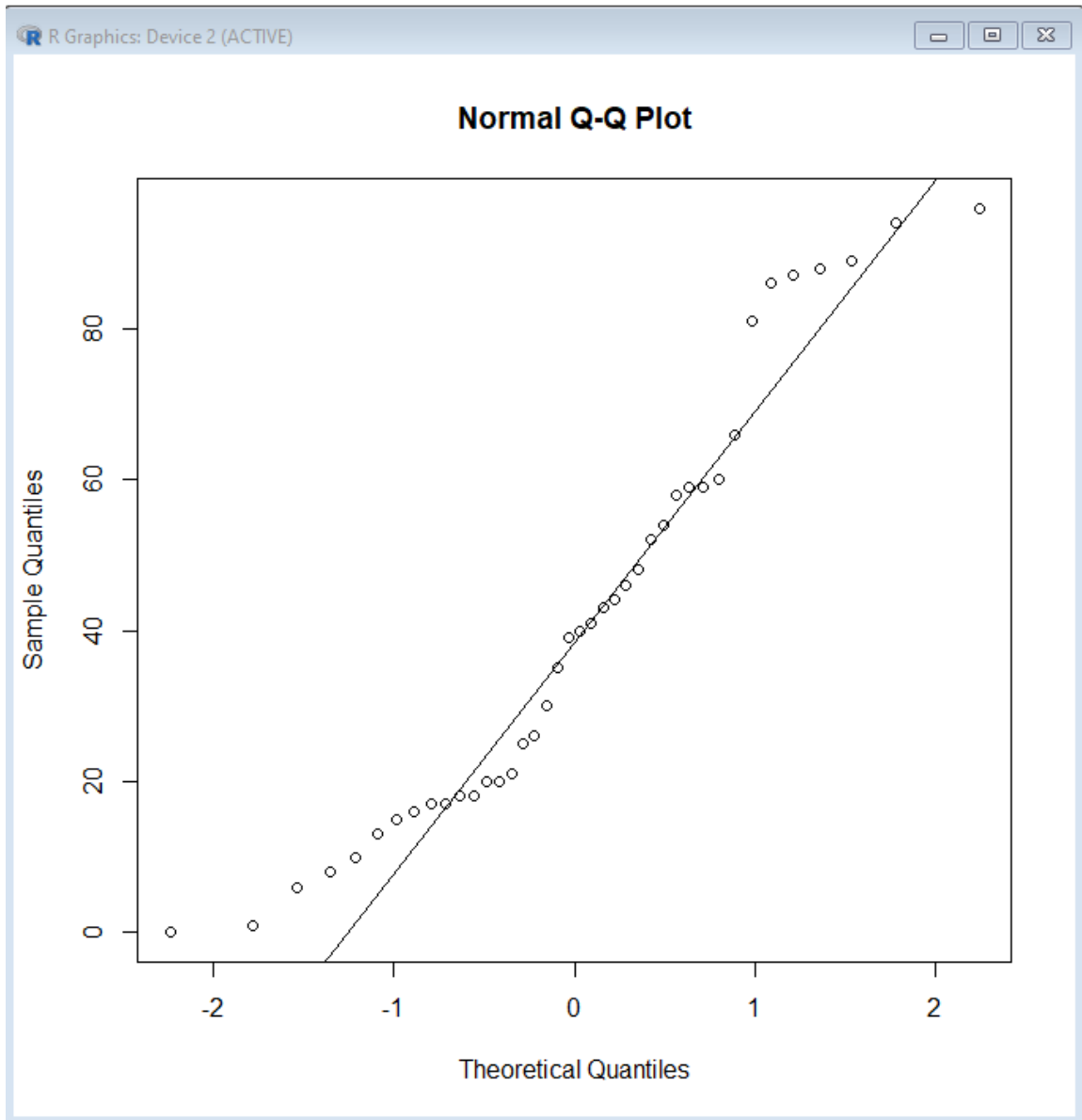
Παρατηρούμε μια συγγραμμικότητα των σημείων άρα τα δεδομένα 1 θα προσέγγιζαν αρκετά «καλά» την καμπύλη πυκνότητας της κανονικής κατανομής

Δεδομένα 2:



Εδώ παρατηρούμε ότι η σχέση $x=y$ (δηλαδή να έχω συγγραμμικά σημεία) είναι πολύ ασθενής άρα τα δεδομένα 2 δεν προσεγγίζουν την καμπύλη πυκνότητας της κανονικής κατανομής

Δεδομένα 3:



Εδώ πιστεύουμε ότι τα δεδομένα 3 προσεγγίζουν την καμπύλη πυκνότητας της κανονικής κατανομής. Στο διάγραμμα κάτω αριστερά έχουμε αρκετά ατυπικά σημεία που αποκλίνουν από την ευθεία αλλά προς στο κέντρο έχουμε μία ικανοποιητική συγγραμμικότητα των σημείων και για αυτό τον λόγο τείνουμε προς την απάντηση που ειπώθηκε πιο πάνω.

ΑΣΚΗΣΗ 2

a)

Τα δεδομένα προέρχονται από το πρωτάθλημα του NBA και αφορούν τους συνολικούς μέσους όρους όλων των παικτών, δηλαδή μέσος όρος ηλικίας, πόντων, κλπ για τις σεζόν από 1979-1980

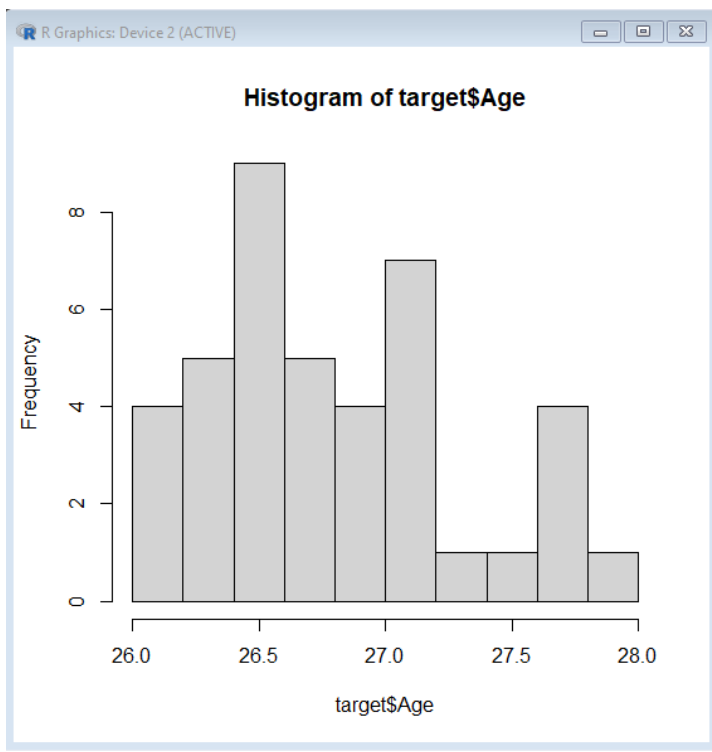
μέχρι τη σεζόν 2019-2020. Περιέχονται δηλαδή δεδομένα από 40 σεζόν. Τα βρήκαμε στο site <https://www.basketball-reference.com/>

b)

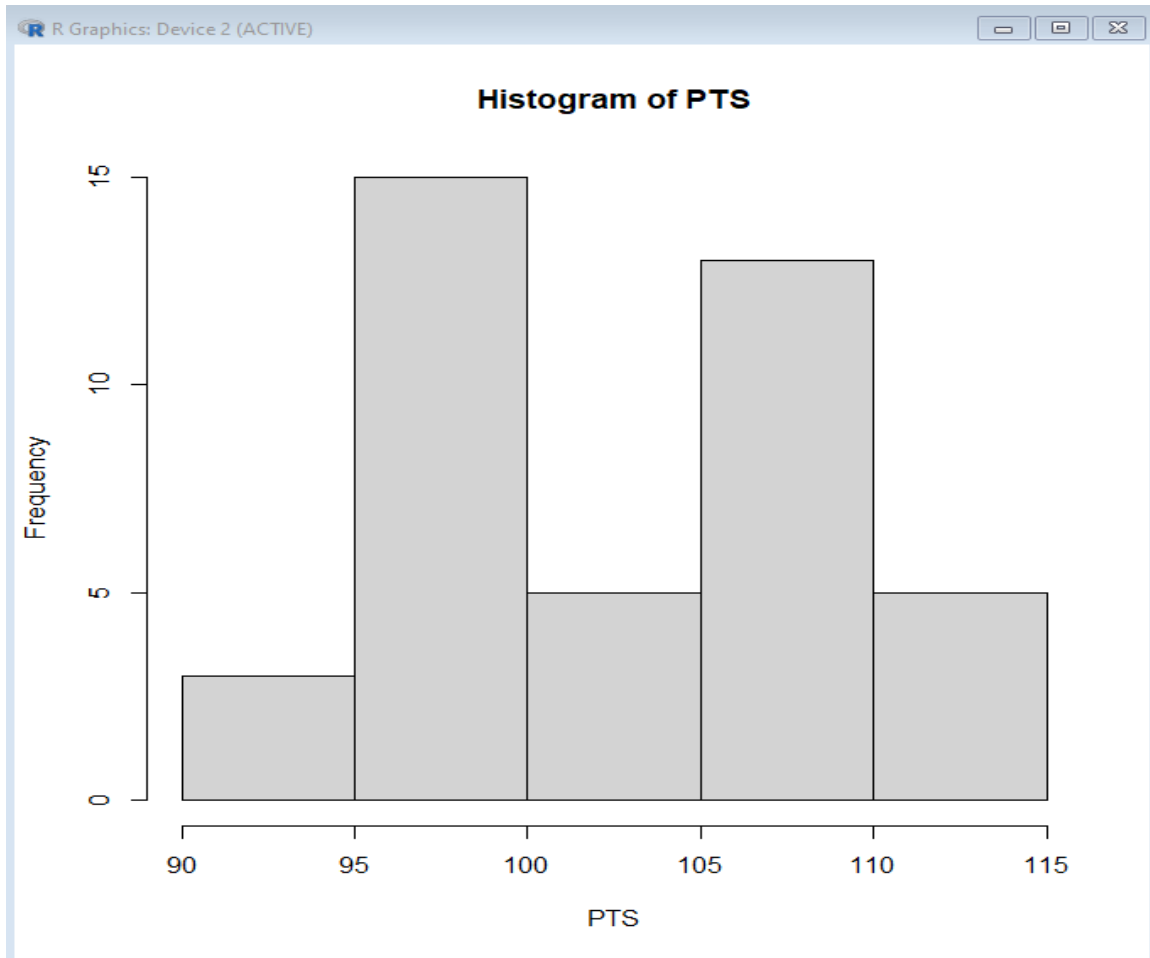
Κατηγορηματική μεταβλητή είναι η Season η οποία δείχνει σε ποια σεζόν του πρωταθλήματος αναφέρονται τα δεδομένα

Ποσοτικές μεταβλητές είναι όλες οι υπόλοιπες όπως Age που δείχνει τον μέσο όρο ηλικίας των παικτών σε μια σεζόν, 3P που δείχνει τον μέσο όρο τριπόντων σε μια σεζόν, G που δείχνει τον αριθμό παιχνιδιών ανά σεζόν, PTS που δείχνει τον μέσο όρο πόντων ομάδας ανά σεζόν

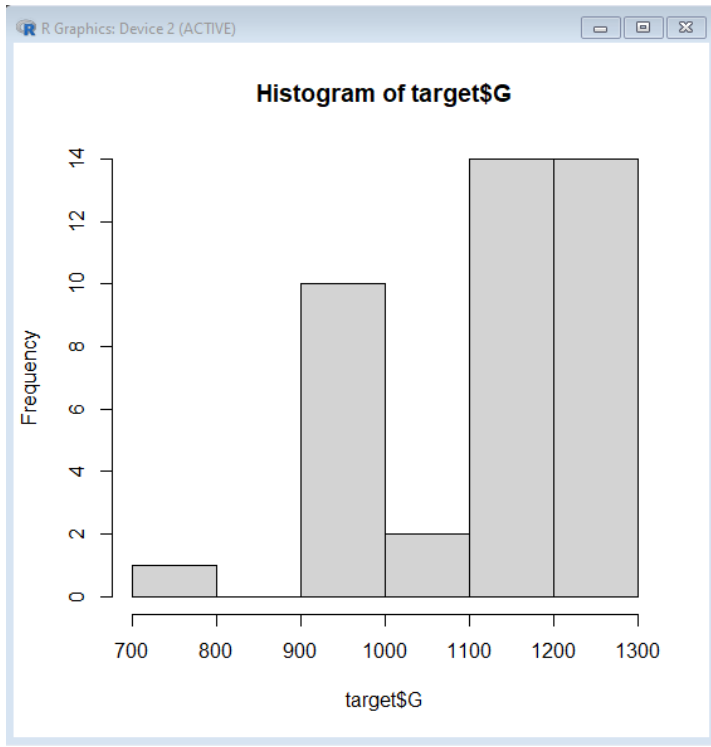
c)



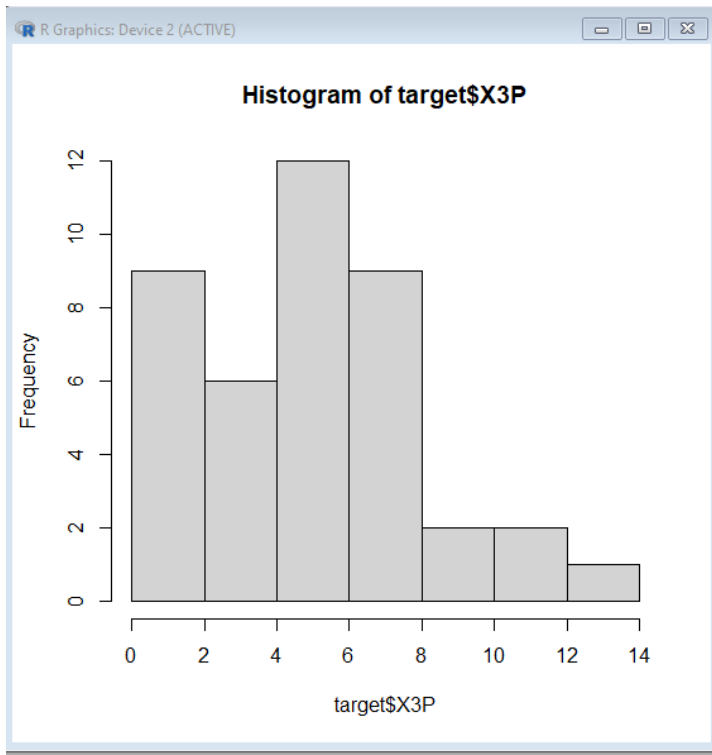
Στο παραπάνω ιστόγραμμα παρουσιάζεται η κατανομή της ποσοτικής μεταβλητής Age. Βλέπουμε πως ο μέσος όρος της ηλικίας των παικτών σε κάθε σεζόν βρίσκεται ανάμεσα στις τιμές 26 και 28 αυτό γίνεται καθώς οι παίκτες με ηλικία κάτω από 25 και πάνω από 30 είναι λίγοι σχετικά με τους παίκτες ηλικίας 25-30. Βλέπουμε τρία ατυπικά σημεία στα 27.3, 27.6 και 27.8 αυτά υπάρχουν καθώς για τις τρεις αυτές σεζόν η ηλικία των παικτών ήταν μεγαλύτερη από τη συνηθισμένη και αυτό μπορεί να οφείλεται σε έλλειψη ποιοτικών νέων παικτών από τα κολλέγια.



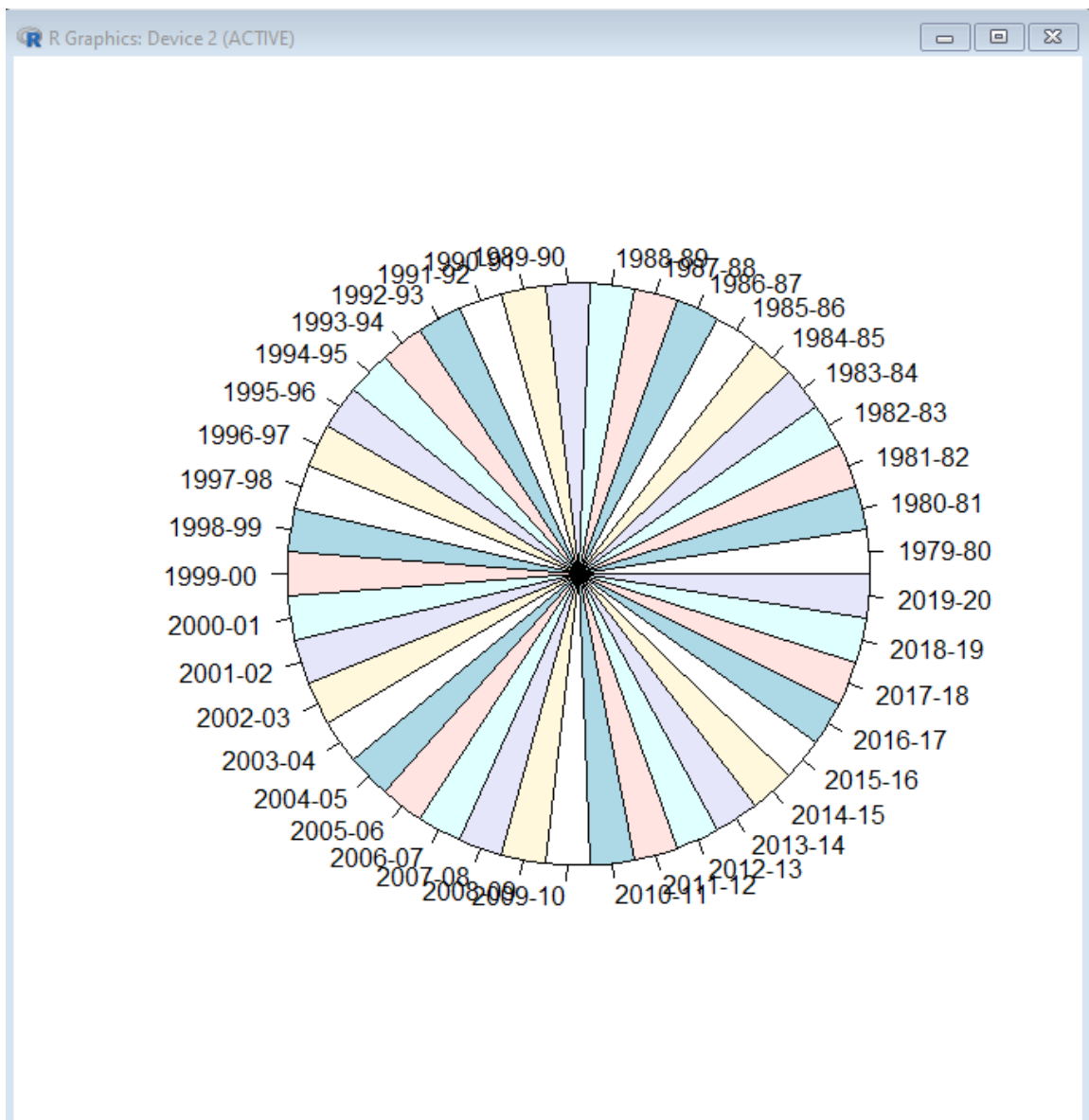
Σε αυτό το ιστόγραμμα βλέπουμε την κατανομή για την ποσοτική μεταβλητή PTS που δείχνει τον αριθμό μέσων πόντων ανά παιχνίδι που βάζει κάθε ομάδα. Παρατηρούμε ότι έχουμε αρκετές σεζόν με 95-100 και 105 με 110 πόντους ανά παιχνίδι και λιγότερες με 110+ και 95-. Λογικό συμπέρασμα μιας και οι περισσότερες ομάδες συνήθως στο NBA σε αυτές τις τιμές κυμαίνονται.



Σε αυτό το ιστόγραμμα βλέπουμε την κατανομή για την ποσοτική μεταβλητή G που δείχνει τον αριθμό παιχνιδιών σε κάθε περίοδο. Στον μεγαλύτερο αριθμό σεζόν τα παιχνίδια κυμαίνονται από 900-1000 και 1100-1300 αυτό οφείλεται πιθανώς στην αλλαγή των κανονισμών δηλαδή από κάποια σεζόν και μετά η διοίκηση του πρωταθλήματος έβαλε περισσότερες ομάδες στην διοργάνωση και έτσι αναγκαστικά αυξήθηκε ο αριθμός των παιχνιδιών. Υπάρχουν δύο ατυπικά σημεία στις τιμές 700-800 και 1000-1100 τα οποία μπορεί να οφείλονται σε κάποιο γεγονός που μπορεί να επίσπευσε την ολοκλήρωση του πρωταθλήματος όπως φυγή κάποιας ομάδας από την διοργάνωση.



Στο ιστόγραμμα αυτό παρουσιάζεται ο μέσος όρος εύστοχων τριπόντων σε κάθε παιχνίδι της σεζόν. Στις περισσότερες σεζόν ο μέσος όρος εύστοχων τριπόντων ανα παιχνίδι βρίσκεται στις τιμές 0-8. Αυτό οφείλεται στο γεγονός ότι στις παλαιότερες σεζόν οι παίκτες δεν έπαιρναν τόσες προσπάθειες για τρίποντα συνεπώς δεν υπήρχαν και τόσα εύστοχα σουτ. Αντίθετα τις τελευταίες πέντε σεζόν παρατηρείται αύξηση των εύστοχων τριπόντων, κάτι που αιτιολογεί και τα ατυπικά σημεία στις τιμές 8-14. Αυτή η συμπεριφορά της μεταβλητής οφείλεται στην αλλαγή του τρόπου παιχνιδιού τα τελευταία χρόνια.



Εδώ παρουσιάζεται η κατηγορηματική μεταβλητή Season. Λογικό κάθε Season να πιάνει «ομοιόμορφο» κομμάτι της πίτας με τις άλλες αφού επειδή μιλάμε για season και όχι για κάποια απάντηση «ναι» ή «όχι» σε κάποιο γκάλοπ άρα έχουμε την εμφάνιση της ίδιας μεταβλητής ΜΟΝΟ 1 ΦΟΡΑ.

d)

Age

Μέση τιμή: 26.85366

Τυπική απόκλιση: 0.4873898

Min: 26.10

Q1:26.50

m:26.70

Q3:27.20

Max:27.92

Πιο καταλληλά είναι η μέση τιμή και η τυπική απόκλιση καθώς οι τιμές είναι συμπυκνωμένες.

G

Μέση τιμή: 1112.634

Τυπική απόκλιση: 131.3284

Min: 725

Q1: 990

m: 1189

Q3: 1230

Max: 1230

Πιο κατάλληλη είναι η σύνοψη των πέντε αριθμών καθώς οι τιμές βρίσκονται σε μεγάλο εύρος.

3P

Μέση τιμή: 4.865854

Τυπική απόκλιση: 3.100372

Min: 0.5

Q1: 2.2

m: 5.2

Q3: 6.5

Max: 12.2

Πιο κατάλληλη είναι η σύνοψη των πέντε αριθμών καθώς οι τιμές βρίσκονται σε μεγάλο εύρος.

PTS

Μέση τιμή:102.6

Τυπική απόκλιση: 5.8096

Min: 91.6

Q1: 97.5

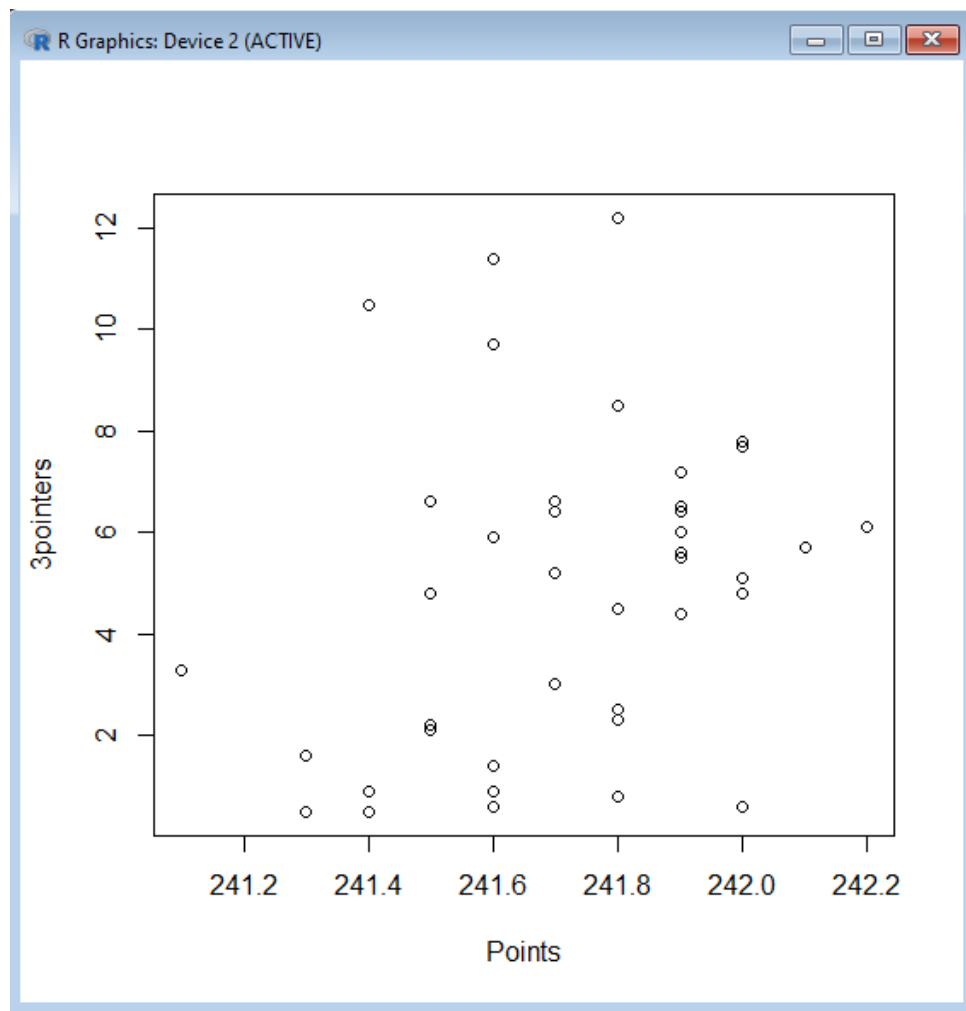
m: 101.4

Q3: 108.2

Max: 111.8

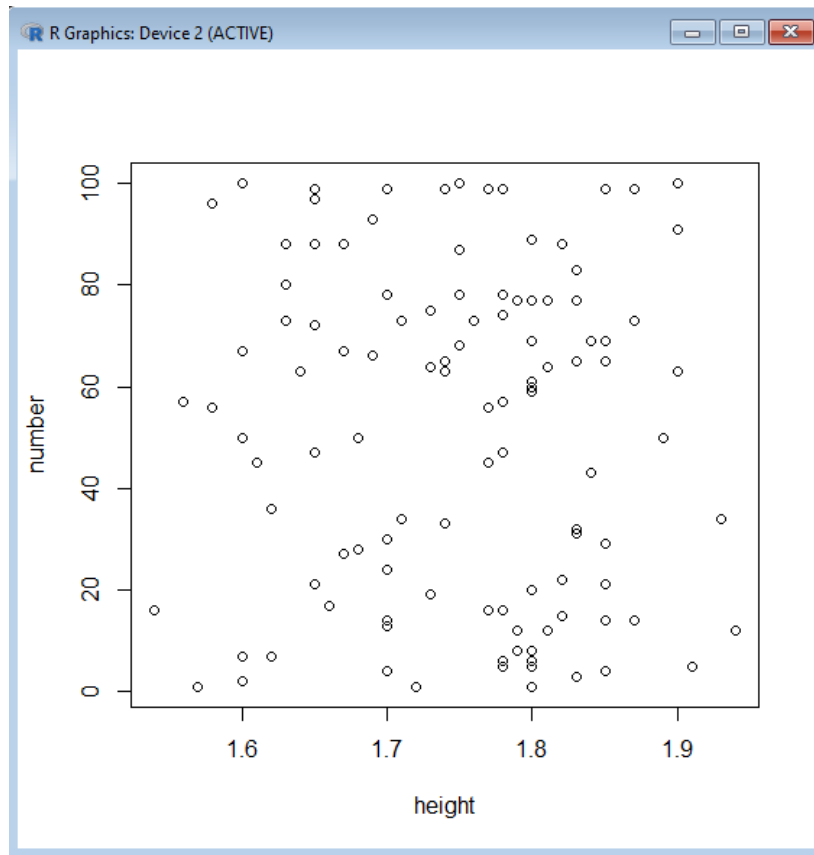
Πιο κατάλληλη είναι η σύνοψη των πέντε αριθμών καθώς οι τιμές βρίσκονται σε μεγάλο εύρος.

e)



Οι δύο μεταβλητές Αριθμός πόντων και αριθμός εύστοχων τριπόντων έχουν γραμμική αύξουσα σχέση καθώς είναι φυσικό όσα περισσότερα καλάθια τριών πόντων υπάρχουν σε ένα παιχνίδι τόσο μεγαλύτερος είναι και ο αριθμός των πόντων. Η σχέση είναι μέτρια προς ισχυρή καθώς βλέπουμε πως υπάρχουν και ορισμένα ατυπικά σημεία. Η συσχέτιση των δυο μεταβλητών είναι αιτιακή καθώς η μία επηρεάζει την τιμή της άλλης.

ΑΣΚΗΣΗ 3



Η σχέση ανάμεσα στο ύψος και τον τυχαίο αριθμό από το 0 μέχρι το 100 δεν είναι γραμμική. Παρατηρούμε ότι υπάρχουν δύο ομάδες, μια κάτω δεξιά ανάμεσα στις τιμές 1.75 και 1.85 για το ύψος και τις τιμές 0 και 20 για τον τυχαίο αριθμό. Η δεύτερη ομάδα είναι στο πάνω μέρος του διαγράμματος ανάμεσα στις τιμές 1.7 και 1.8 για το ύψος και 60 και 85 για τον τυχαίο αριθμό. Ατυπικά σημεία δεν υπάρχουν σε αυτή τη σχέση. Η σχέση είναι αρκετά ανίσχυρη.

Συντελεστής συσχέτισης είναι : -0.01213928, που δηλώνει μια ασθενής φθίνουσα γραμμική σχέση.

Γραμμική παλινδρόμηση: $\text{height} = 1.754\text{e}+00 - 9.467\text{e}-05 \cdot \text{number}$

Γραμμική παλινδρόμηση ελάχιστων τετραγώνων στην R:

