

Lab 1

Μέλη Ομάδας:

- Κωνσταντίνος Κανελλής - kkanellis@uth.gr (AEM: 1763)
- Ελευθέριος Παναγιώτης Λούκας - eloukas@uth.gr (AEM: 2029)
- Κλάιντι Μποντούρρι - kbontourri@uth.gr (AEM: 2020)

Συμπεράσματα απο το Lab1

Με την εντολή `summary()` μπορούμε να πάρουμε γρήγορα κάποια στατιστικά όπως το `max`, `min`, `mean` για να έχουμε μια γενική εικόνα των δεδομένων μας. Ένα παράδειγμα τέτοιας χρήσης θα ήταν αν κάποιος ενδιαφέρεται να μάθει πόσες χρηματικές μονάδες είναι το μέσο κόστος `Room and Board` για ένα φοιτητή. Στην παρακάτω εικόνα φαίνεται το `output` απο το `summary(college)`:

```
> summary(college)
Private      Apps      Accept      Enroll      Top10perc      Top25perc      F.Undergrad      P.Undergrad
No :212   Min.   : 81   Min.   : 72   Min.   : 35   Min.   : 1.00   Min.   : 9.0   Min.   : 139   Min.   : 1.0
Yes:565   1st Qu.: 776   1st Qu.: 604   1st Qu.: 242   1st Qu.:15.00   1st Qu.: 41.0   1st Qu.: 992   1st Qu.: 95.0
          Median : 1558   Median : 1110   Median : 434   Median :23.00   Median : 54.0   Median : 1707   Median : 353.0
          Mean   : 3002   Mean   : 2019   Mean   : 780   Mean   :27.56   Mean   : 55.8   Mean   : 3700   Mean   : 855.3
          3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.:35.00   3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.: 967.0
          Max.   :48094   Max.   :26330   Max.   :6392   Max.   :96.00   Max.   :100.0   Max.   :31643   Max.   :21836.0

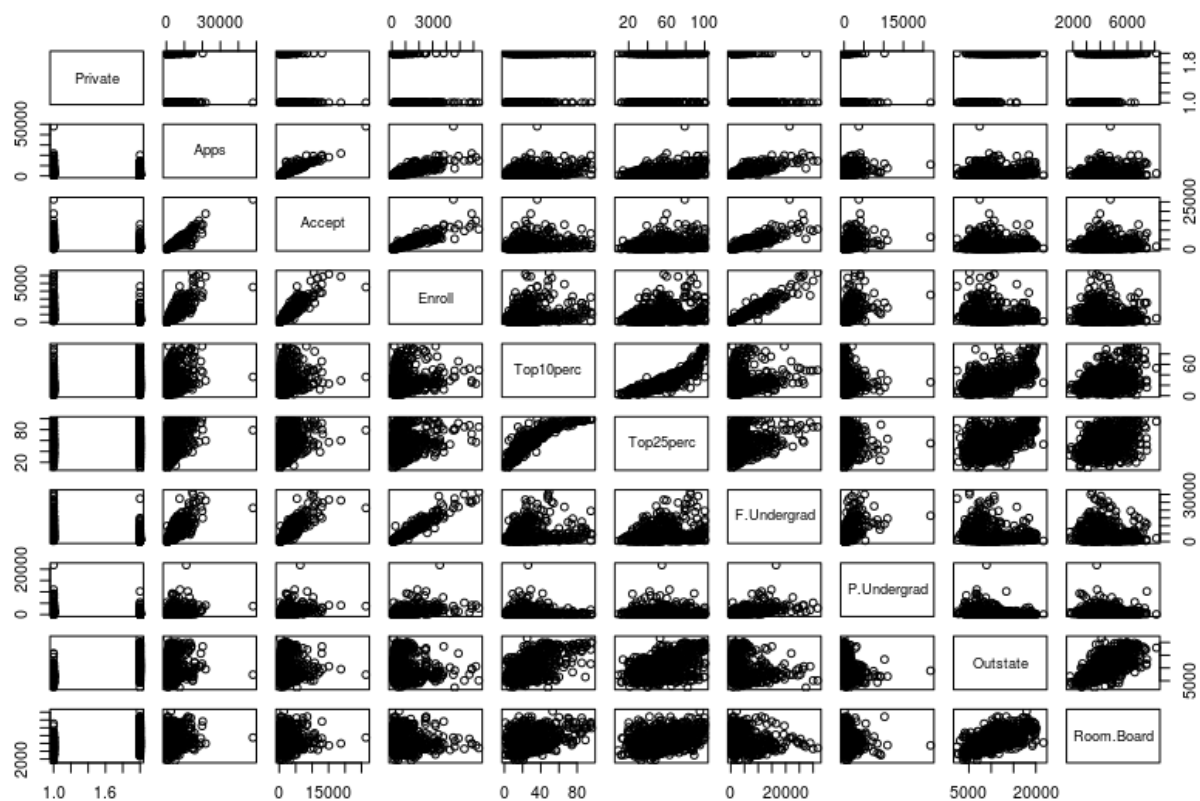
Outstate      Room.Board      Books      Personal      PhD      Terminal      S.F.Ratio      perc.alumni
Min.   : 2340   Min.   :1780   Min.   : 96.0   Min.   : 250   Min.   : 8.00   Min.   : 24.0   Min.   : 2.50   Min.   : 0.00
1st Qu.: 7320   1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850   1st Qu.: 62.00   1st Qu.: 71.0   1st Qu.:11.50   1st Qu.:13.00
Median : 9990   Median :4200   Median : 500.0   Median :1200   Median : 75.00   Median : 82.0   Median :13.60   Median :21.00
Mean   :10441   Mean   :4358   Mean   : 549.4   Mean   :1341   Mean   : 72.66   Mean   : 79.7   Mean   :14.09   Mean   :22.74
3rd Qu.:12925   3rd Qu.:5050   3rd Qu.: 600.0   3rd Qu.:1700   3rd Qu.: 85.00   3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00
Max.   :21700   Max.   :8124   Max.   :2340.0   Max.   :6800   Max.   :103.00   Max.   :100.0   Max.   :39.80   Max.   :64.00

Expend      Grad.Rate
Min.   : 3186   Min.   : 10.00
1st Qu.: 6751   1st Qu.: 53.00
Median : 8377   Median : 65.00
Mean   : 9660   Mean   : 65.46
3rd Qu.:10830   3rd Qu.: 78.00
Max.   :56233   Max.   :118.00
> |
```

Lab 1

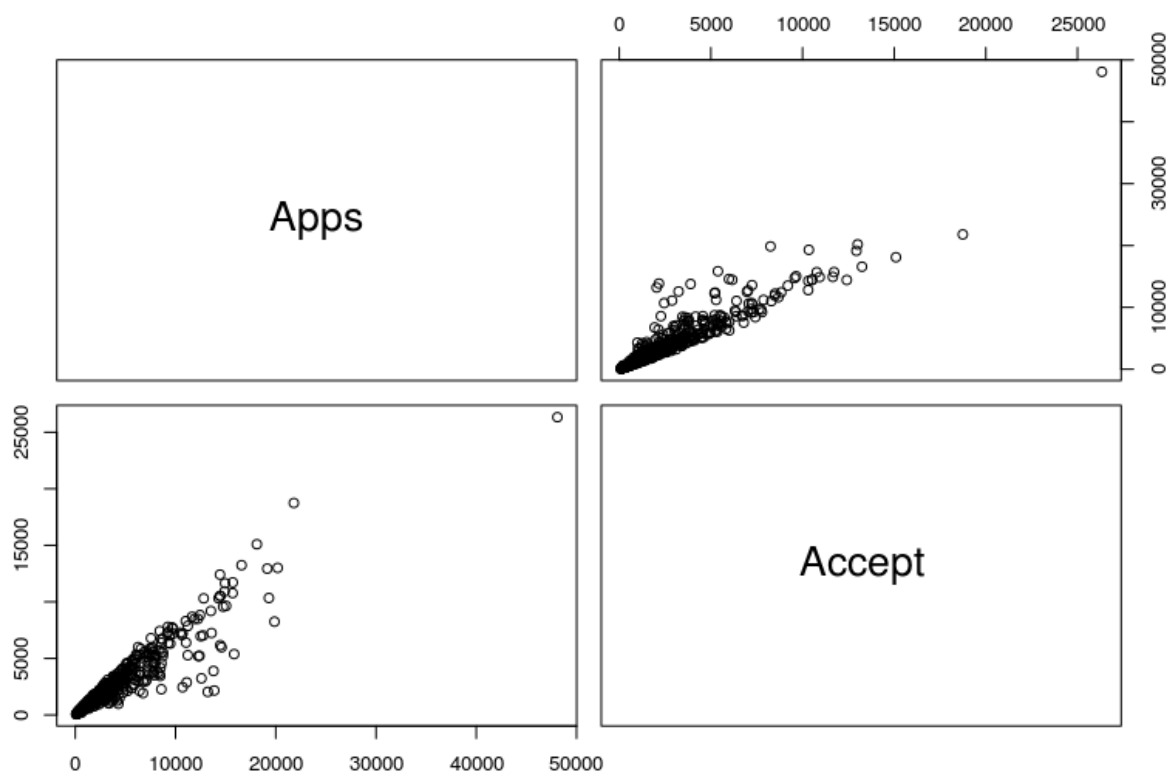
Έπειτα, με το scatterplot προσπαθούμε να δούμε πως συμπεριφέρονται “ζευγάρια” δεδομένων όπως για παράδειγμα μεταξύ Apps και Accept students.

Στην παρακάτω εικόνα είναι το scatterplot απο 10 πρώτες μεταβλητές των δεδομένων.



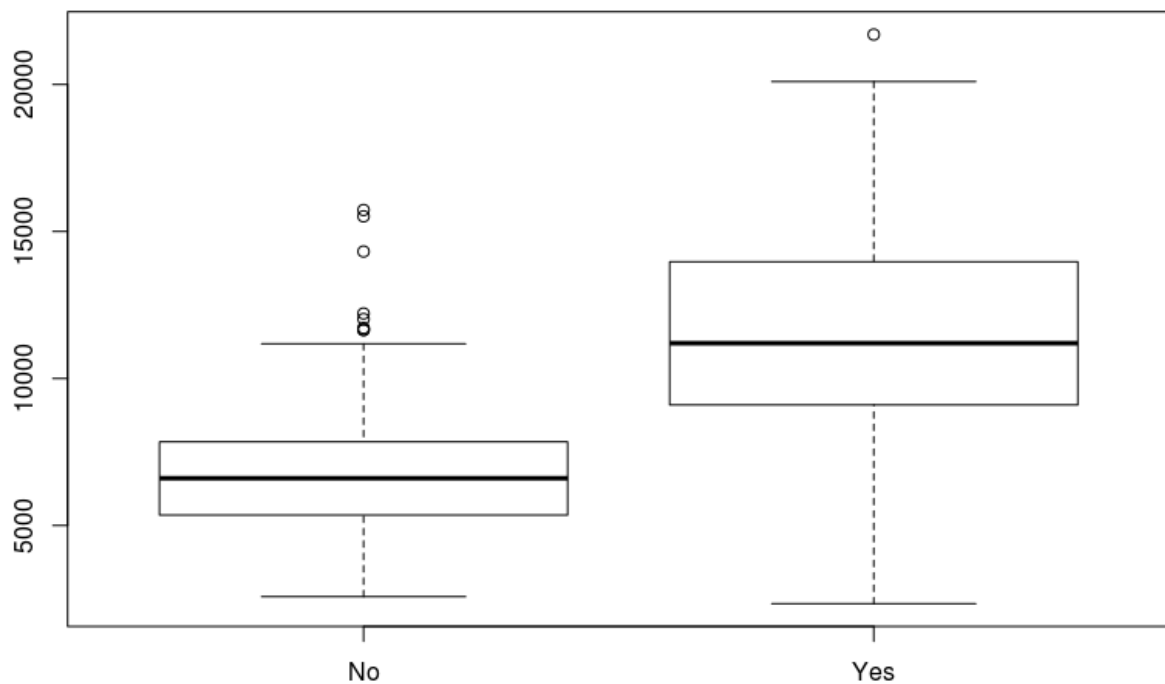
Lab 1

Πιο ειδικά, για παράδειγμα, παρατηρούμε ότι σε γενικές γραμμές, η σχέση απο κάθε πανεπιστήμιο μεταξύ των αριθμών των αιτήσεων (Apps) και των αριθμό των αποδοχών (Accept) είναι γραμμική. Δηλαδή όσο μεγαλύτερος ο αριθμός των αιτήσεων, τόσο μεγαλύτερος και ο αριθμός των εισακτέων φοιτητών. (Κάτω αριστερό γράφημα)



Lab 1

Με το boxplot μπορούμε να συγκρίνουμε 2 μεταβλητές όπως για παράδειγμα το Outstate με το Private. Δηλαδή μπορούμε να δούμε τι γίνεται με τα εξωτερικά δίδακτρα αν ένα πανεπιστήμιο είναι ιδιωτικό ή όχι.

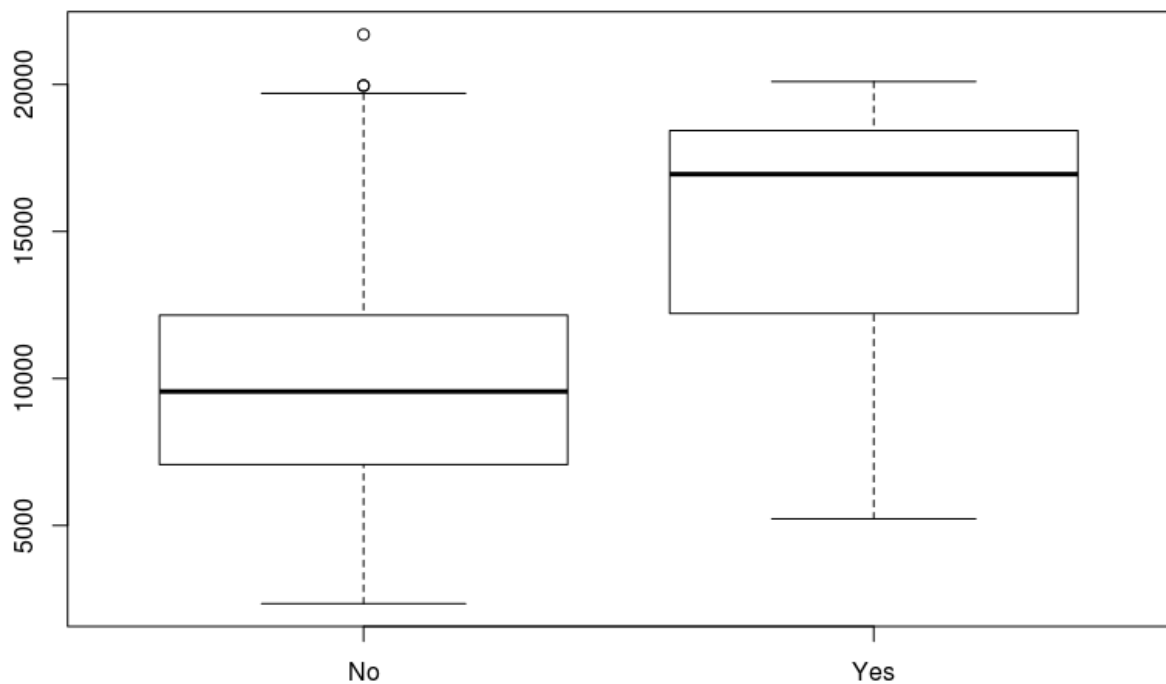


Εδώ παρατηρούμε ότι στα μη ιδιωτικά πανεπιστήμια τα εξωτερικά έξοδα είναι εμφανώς λιγότερα από αυτά των ιδιωτικών (με μικρές εξαιρέσεις). Επίσης με το boxplot μπορούμε να δούμε ποιά είναι συμμετρικά. (Αν η bold γραμμή είναι στο κέντρο του ορθογωνίου ή όχι.)

Lab 1

Στην συνέχεια ζητήθηκε να συγκρίνουμε το Outstate με το αν ένα πανεπιστήμιο είναι Elite ή όχι.
(Elite χαρακτηρίζεται όποιο πανεπιστήμιο έχει $> 50\%$ των Top10perc)

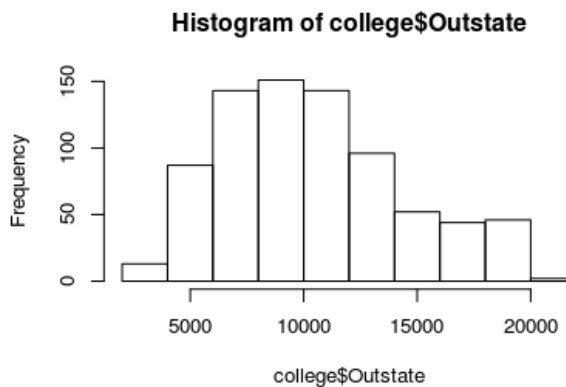
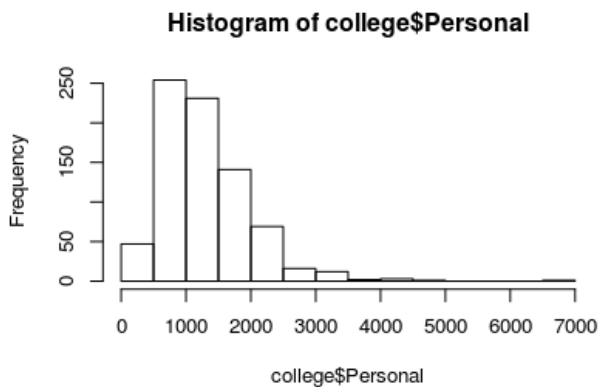
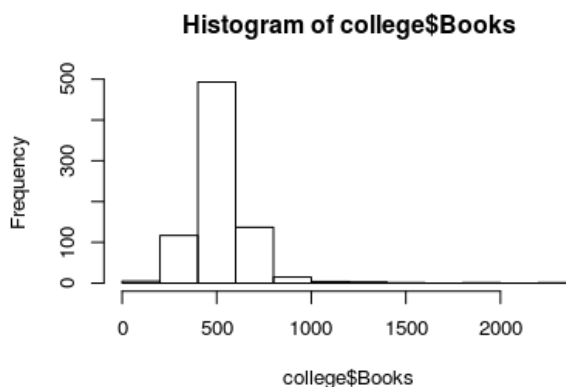
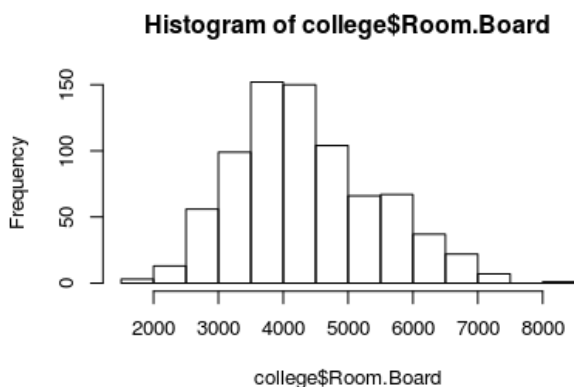
Με το `summary()` μπορούμε να δούμε πόσα πανεπιστήμια είναι Elite και με το `boxplot` κάποια χαρακτηριστικά όπως εξηγήθηκε παραπάνω.



Και εδώ είναι εμφανές πόσα περισσότερα είναι τα εξωτερικά δίδακτρα σε ένα Elite πανεπιστήμιο.

Lab 1

Τέλος, με τα histograms μπορούμε να πάρουμε μια εικόνα για το πώς κατανέμονται οι διάφορες τιμές των μεταβλητών. Για παράδειγμα, στην παρακάτω εικόνα υπάρχει ιστόγραμμα για τα έξοδα του φοιτητή (Room.Board, Books, Personal, Outstate).



Ένας φοιτητής, βλέποντας αυτό το γράφημα, μπορεί εύκολα να καταλάβει τα έξοδα που θα έχει στα περισσότερα πανεπιστήμια όσον αφορά, για παράδειγμα, τα βιβλία. Το πάνω δεξιό γράφημα μας πληροφορεί ότι σχεδόν σε όλα τα πανεπιστήμια (500 περίπου) θα πληρώσει γύρω στις 500 χρηματικές μονάδες. Αναλόγως ισχύει και στα υπόλοιπα γραφήματα.

Lab 1

Github repository: <https://github.com/eloukas/PSEs-and-Data-Science>

Μπορείτε να επικοινωνήσετε μαζί μας αν θέλετε να επιβεβαιώσετε τον κώδικα.
(το repo είναι private προσωρινά)

- | | | |
|--------------------------------|--|-------------|
| - Κωνσταντίνος Κανελλής | - kkanellis@uth.gr | (AEM: 1763) |
| - Ελευθέριος Παναγιώτης Λούκας | - eloukas@uth.gr | (AEM: 2029) |
| - Κλάιντι Μποντούρρι | - kbontourri@uth.gr | (AEM: 2020) |