

# “Τα Big Data στο χώρο της Υγείας”

Εργαστήριο Βιοϊατρικής Τεχνολογίας

Εαρινό Εξάμηνο 2023

## Εργασία για το σπίτι

Στη συγκεκριμένα άσκηση καλείσθε να πραγματοποιήσετε κάποιες στατιστικές αναλύσεις δεδομένων με τη χρήση της γλώσσας R στο ολοκληρωμένο περιβάλλον ανάπτυξης RStudio. Απαραίτητες προϋποθέσεις για την πραγματοποίηση της άσκησης είναι οι εξής:

- Εγκατάσταση της R στο λειτουργικό σας σύστημα (<https://cran.r-project.org/>)
- Εγκατάσταση του RStudio (<https://rstudio.com/products/rstudio/download/>)

Αφού έχετε ολοκληρώσει τα παραπάνω, θα πρέπει να “φορτώσετε” το προ-εγκατεστημένο πακέτο “usingR” (<https://cran.r-project.org/web/packages/UsingR/>), το οποίο περιέχει διάφορα σύνολα δεδομένων (datasets). Για τη φόρτωση του πακέτου θα πρέπει να εκτελέσετε την παρακάτω εντολή στο παράθυρο κονσόλας / εντολών:

```
library('UsingR')
```

Από το συγκεκριμένο πακέτο θα εγκαταστήσετε το σύνολο δεδομένων wellbeing. Το σύνολο δεδομένων αυτό αποτελεί ουσιαστικά ένα πλαίσιο δεδομένων (data frame) με στοιχεία από 22 Ευρωπαϊκές χώρες σχετικά με το τι κάνει τους ανθρώπους ευτυχισμένους βάσει 10 διαφορετικών παραμέτρων. Για να φορτώσετε και να δείτε το σύνολο δεδομένων wellbeing θα πρέπει να εκτελέσετε τις παρακάτω εντολές:

```
data(wellbeing)  
View(wellbeing)
```

Το σύνολο δεδομένων **wellbeing** περιέχει **12 μεταβλητές** σε κάθε σειρά του πλαισίου δεδομένων. Η πρώτη μεταβλητή / στήλη αφορά στη χώρα (**Country**), η δεύτερη μεταβλητή / στήλη περιέχει το δείκτη ευεξίας της κάθε χώρας (**Well.being**) και οι υπόλοιπες 10 αφορούν σε διάφορες παραμέτρους. Αναλυτικότερα αυτές είναι:

0: **GDP**: Ακαθάριστο Εγχώριο Προϊόν (US \$)

1: **Equality**: Δείκτης ισότητας

2: **Food.consumption**: Κατανάλωση φαγητού (kCal / ημέρα / άτομο)

3: **Alcohol.consumption**: Κατανάλωση αλκοόλ (λίτρα καθ. αλκοόλ / έτος / άτομο)

4: **Energy.consumption**: Κατανάλωση ενέργειας (kWh / έτος / άτομο)

5: **Family**: Γονιμότητα (αριθμός παιδιών / γυναίκα)

6: **Working.hours**: Ώρες εργασίας (ΜΟ ωρών εργασίας / εβδ. / άτομο)

7: **Work.income**: Εργασιακό εισόδημα (ωρομίσθιο σε US \$ / άτομο)

8: **Health.spending**: Δαπάνες υγειονομικής περίθαλψης (% των κρατικών δαπανών)

9: **Military.spending**: Στρατιωτικές δαπάνες (% του ΑΕΠ)

Με βάση το τελευταίο ψηφίο του Αριθμού Μητρώου (AM) σας αντιστοιχεί στον κάθε ένα από εσάς μία από τις ανωτέρω 10 μεταβλητές, η οποία εξελέγη να ονομάζεται **μεταβλητή X**. Αφού βρείτε ποια είναι αυτή, θα πραγματοποιήσετε τους υπολογισμούς / υλοποιήσεις που σας ζητούνται παρακάτω. Π.χ. για έναν φοιτητή με AM 03110002 αντιστοιχεί η μεταβλητή 2: **Food.consumption** κ.ο.κ.

## Ζητούμενα εργασίας

- 1) Υπολογίστε το μέσο όρο της μεταβλητής  $X$  για όλες τις Ευρωπαϊκές χώρες του συνόλου δεδομένων.
- 2) Εντοπίστε την Ευρωπαϊκή χώρα με τη μεγαλύτερη τιμή της μεταβλητής  $X$ .
- 3) Ταξινομήστε όλες τις Ευρωπαϊκές χώρες (σε φθίνουσα σειρά) με βάση τη μεταβλητή  $X$ .  
*Σύσταση: αποθηκεύστε το νέο πλαίσιο δεδομένων σε μία νέα μεταβλητή wellbeing2.*
- 4) Εμφανίστε τα ονόματα των 3 πρώτων Ευρωπαϊκών χωρών βάσει της ταξινόμησης του προηγούμενου ερωτήματος.
- 5) Απεικονίστε γραφικά το δείκτη ευεξίας (Well.being) των Ευρωπαϊκών χωρών συναρτήσει της μεταβλητής  $X$ .
- 6) Απομονώστε τις Ευρωπαϊκές χώρες οι οποίες έχουν τιμή μεταβλητής  $X$  μεγαλύτερη από το μέσο όρο της μεταβλητής  $X$  των 22 χωρών.  
*Σύσταση: αποθηκεύστε το νέο πλαίσιο δεδομένων σε μία νέα μεταβλητή wellbeing3.*
- 7) Απεικονίστε γραφικά το δείκτη ευεξίας (Well.being) των απομονωμένων Ευρωπαϊκών χωρών που υπολογίσατε στο προηγούμενο ερώτημα. Η απεικόνιση θα πρέπει να είναι στη μορφή ραβδογράμματος με φθίνουσα σειρά.

*Προσοχή: Στο αρχικό πλαίσιο δεδομένων wellbeing κάποιες αριθμητικές τιμές λείπουν και σημειώνονται με το σύμβολο NA (Not Available). Εάν η μεταβλητή  $X$  περιέχει τέτοιες τιμές προνοήστε για τη σωστή διαχείρισή τους. Για περισσότερες πληροφορίες μπορείτε να συμβουλευτείτε το σύνδεσμο: <https://www.statmethods.net/input/missingdata.html>*

## Παραδοτέο

Το **τελικό παραδοτέο** της εργασίας θα είναι ένα **αρχείο R script** (.R) με όνομα της μορφής *ErwnimoOnomaAM.R*, π.χ. *ParadopoulosGeorgios03110002.R*. Στο παραδοτέο αρχείο R script θα πρέπει αρχικά να φορτώνετε το σύνολο δεδομένων wellbeing και στη συνέχεια να υπολογίζετε κάθε ένα από τα 7 ζητούμενα της εργασίας. Ο κώδικας θα πρέπει να περιέχει πολύ συνοπτικά σχόλια / τίτλους για το κάθε ζητούμενο.

