Εθνικό Μετσόβιο Πολυτενχείο

Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών

Εργαστήριο Βϊοιατρικής Τεχνολογίας

3η Εργαστηριακή άσκηση «Ηλεκτρονικός φάκελος υγείας και Ηλεκτρονική Συνταγογράφηση»

Αργυρώ Τσίπη
03119950

Α Μέρος

Α.1) Για το πρώτο ερώτημα, μέσω των search_topic & search_query αναζητούμε άρθρα στο PubMed που περιέχουν τον όρο e-prescription.
Χρησιμοποίησα τη μεταβλητή retmax και την έθεσα ίση με 50 ώστε να ορίσω τον μέγιστο αριθμό άρθων που θέλω να αναζητήσω, καθώς και τις mindate & maxdate ισες με 201x με x = 0 (ΜΟ: 03119950) για τη χρονιά 2010.

```
> # a.1
> library(RISmed)
> search_topic<-'e-prescription'
> search_query<-EUtilsSummary(search_topic, retmax=50, mindate=2010, maxdate=2010)
> summary(search_query)
Query:
("electronic prescribing"[MeSH Terms] OR ("electronic"[All Fields] AND "prescribing"[All Fields]) OR "electronic
 prescribing"[All Fields] OR "e prescription"[All Fields]) AND 2010[EDAT] : 2010[EDAT]

Result count:  183
```

Βλέπουμε ότι ο κώδικας εκτελέστηκε και εκτυπώνει τον συνολικό αριθμό άρθρων του PubMed που δημοσιεύτηκαν το έτος 2010.

Α.2) Βρίσκουμε τα αναγνωριστικά των 50 αυτών άρθρων (article ID) μέσω της QueryId που εμφανίζονται στην οθόνη ως:

```
> # a.2
> # see the ids of our returned query
> QueryId(search_query)
 [1] "21190583" "21181080" "21179880" "21178789" "21174487" "21173070" "21172978" "21167008" "21157235"
[10] "21150230" "21144049" "21134929" "21131811" "21127317" "21122059" "21112243" "21109619" "21098354"
[19] "21097562" "21093337" "21087524" "21084362" "21078233" "21074470" "21069521" "21044357" "21049811"
[28] "21049774" "21049590" "21047299" "21043192" "21037044" "20979932" "20976098" "20966784" "20962131"
[37] "20961164" "20959581" "20946221" "20939864" "20927673" "20927290" "20925440" "20922345" "20887239"
[46] "20876411" "20876410" "20875257" "20858645" "20854392"
> # get actual data from PubMed
> records<-EUtilsGet(search_query)
> class(records)
[1] "Medline"
attr(,"package")
[1] "RISmed"
```

Παίρνοντας τα δεδομένα με την EUtilsGet, εξάγουμε τις πληροφορίες που μας ενδιαφέρουν και αναζητούμε τους τίτλους των δέκα πρώτων άρθρων με την ετικέτα ArticleTitle & τις περιλήψεις των δύο τελευταίων άρθρων με την ετικέτα AbsctractText .
Για να προσδιορίσω τα δέκα πρώτα άρθρα, εκτέλεσα την εντολή head με δευτέρο όρισμα τον αριθμό δέκα και αντίστοιχα για τις δύο τελευταίες περιλήψεις την εντολή tail με δεύτερο όρισμα τον αριθμό 2.

```
> # store it
> pubmed_data<-data.frame('Title'=ArticleTitle(records))
> head(pubmed_data,10)
                                                                                                                                                  Title
1                                                                              Ethnographic study of ICT-supported collaborative work routines in general practice.
2                                                                                                        Setting up an emergency stock for metabolic diseases.
3                                    [Safer drug administration with computerized drug records. Solution for many of the old-time problems--but new risks appear].
4                                         Changes in performance after implementation of a multifaceted electronic-health-record-based quality improvement system.
5                                           Physicians' decisions to prescribe antidepressant therapy in older patients with depression in a US managed care plan.
6                                                                                Use of acid-suppressive drugs and risk of pneumonia: a systematic review and meta-analysis.
7                                                                                                       Reducing insulin errors. Try electronic prescribing.
8  Enhancing participant safety through electronically generated medication order sets in a clinical research environment: a medical informatics initiative.
9                                                                                              The influence of payer mix on electronic prescribing by physicians.
10                                                      Rechallenge with platinum plus fluoropyrimidine +/- epirubicin in patients with oesophagogastric cancer.
> pubmed_data<-data.frame('Abstract'=AbstractText(records))
> tail(pubmed_data,2)
   Abstract
49       NA
50       NA
```

Όπως βλέπουμε εκτυπώθηκαν οι τίτλοι, ενώ οι περιλήψεις δεν υπήρχαν οπότε εκτυπώθηκε αριστερά ο αριθμός του άρθρου και δεξιά NA.

Β Μέρος

Β.1) Στο μέρος αυτό διαβάζουμε με την read.csv το αρχείο csv, το οποίο περιέχει δύο στήλες: code, name.

Β.2.α) Αρχικά αναζητούμε τον κωδικό T950 (αρχικό γράμμα επωνύμου και 3 τελευταία ψηφία ΑΜ: **Τ**σίπη 03119**950**).
Όμως, δεν υπήρχε τέτοιος κωδικός και όπως βλέπουμε εκτελώντας τις εντολές εκτύπωσε ότι **δεν υπάρχει**. (βλ. παρακάτω φωτογραφία)
Αντίστοιχα, για αρχικό γράμμα επωνύμου και 2 τελευταία ψηφία ΑΜ, δηλαδή για κωδικό T50, πάλι **δεν υπάρχει** και εκτύπωσε το ίδιο αποτέλεσμα με πριν.
Επέλεξα έναν κωδικό στην τύχη, όπως έλεγε η οδηγία στην εκφώνηση της άσκησης. Επομένως, για κωδικό **Z91030**, βρέθηκε περιγραφή/όνομα ασθένειας **Bee allergy status.**

```
> # b.1
> # read & import data from csv file
> data<-read.csv("icd10.csv", header = TRUE, sep = ";", stringsAsFactors=FALSE)
> # b.2.a
> # T950 (Tsipi, 03119950) doesn't exist, T50 doesn't exist, so i chose one randomly
> # In R, indices start at 1 and the first row with the names of the variables is not counted.
> a = data[data$code == 'T950', ]
> print(a)
[1] code name
<0 rows> (or 0-length row.names)
> a = data[data$code == 'T50', ]
> print(a)
[1] code name
<0 rows> (or 0-length row.names)
> a = data[data$code == 'Z91030', ]
> print(a)
        code               name
71542 Z91030 Bee allergy status
```

Β2.β.γ) Επαναλαμβάνουμε την ίδια διαδικασία δύο φορές, μία για αρχικό γράμμα ονόματος & 3 τελευταία ψηφία ΑΜ (κωδικό **A950**) και μία για έναν τυχάιο κωδικό, έστω **Z91040**.

```
> # b.2.b
> # A950 (Argyro, 03119950) 555 Sylvatic yellow fever
> # Remember that the first row of the CSV file is not counted as the first row because it has
 the names of the variables.
> a = data[data$code == 'A950', ]
> print(a)
    code                  name
555 A950 Sylvatic yellow fever
> # b.2.c
> a = data[data$code == 'Z91040', ]
> print(a)
        code                name
71544 Z91040 Latex allergy status
```

Παρατηρούμε ότι στον κωδικό A950 αντιστοιχεί η ασθένεια **Sylvatic yellow fever**, ενώ στον Z91040 η **Latex allergy status**.

Β.3) Για τις τρεις ασθένειες αυτές αναζητάμε τον αριθμό των abstracts που έχουν δημοσιευτεί στο PubMed τα τελευταία πέντε χρόνια με την EUtilsSummary & ορίσματα: το όνομα της ασθένειας, mindate, maxdate, και τα υπόλοιπα για esearch στο pubmed.

Βλέπουμε, ότι για την Bee allergy status υπάρχουν **7** abstracts, για την Sylvatic yellow fever **86** και για την Latex allergy status **11**.

```
> # b.3.a
> res_a <- EUtilsSummary("Bee allergy status", type="esearch", db="pubmed", datetype='pdat', mi
ndate=2017, maxdate=2022, retmax=500)
> QueryCount(res_a)
[1] 7
> # b.3.b
> res_b <- EUtilsSummary("Sylvatic yellow fever", type="esearch", db="pubmed", datetype='pdat',
 mindate=2017, maxdate=2022, retmax=500)
> QueryCount(res_b)
[1] 86
> # b.3.c
> res_c <- EUtilsSummary("Latex allergy status", type="esearch", db="pubmed", datetype='pdat',
 mindate=2017, maxdate=2022, retmax=500)
> QueryCount(res_c)
[1] 11
```

Β.4) Για να φτίαξω το **barplot** με τις 3 ασθένειες, αρχικά αναζήτησα για τα 3 τελευταία χρόνια πόσα άρθρα δημοσιεύτηκαν κάθε χρονιά ξεχωριστά για την κάθε ασθένεια ξεχωριστά.

```
> # finding the no. of articles for each year & each disease (2020-2022)
> res_a1 <- EUtilsSummary("Bee allergy status", type="esearch", db="pubmed", datetype='pdat', mindate=2020, maxdate=2020, retmax=500)
> QueryCount(res_a1)
[1] 3
> res_a2 <- EUtilsSummary("Bee allergy status", type="esearch", db="pubmed", datetype='pdat', mindate=2021, maxdate=2021, retmax=500)
> QueryCount(res_a2)
[1] 2
> res_a3 <- EUtilsSummary("Bee allergy status", type="esearch", db="pubmed", datetype='pdat', mindate=2022, maxdate=2022, retmax=500)
> QueryCount(res_a3)
[1] 0
> res_b1 <- EUtilsSummary("Sylvatic yellow fever", type="esearch", db="pubmed", datetype='pdat', mindate=2020, maxdate=2020, retmax=500)
> QueryCount(res_b1)
[1] 17
> res_b2 <- EUtilsSummary("Sylvatic yellow fever", type="esearch", db="pubmed", datetype='pdat', mindate=2021, maxdate=2021, retmax=500)
> QueryCount(res_b2)
[1] 12
> res_b3 <- EUtilsSummary("Sylvatic yellow fever", type="esearch", db="pubmed", datetype='pdat', mindate=2022, maxdate=2022, retmax=500)
> QueryCount(res_b3)
[1] 8
> res_c <- EUtilsSummary("Latex allergy status", type="esearch", db="pubmed", datetype='pdat', mindate=2020, maxdate=2020, retmax=500)
> QueryCount(res_c)
[1] 4
> res_c <- EUtilsSummary("Latex allergy status", type="esearch", db="pubmed", datetype='pdat', mindate=2021, maxdate=2021, retmax=500)
> QueryCount(res_c)
[1] 4
> res_c <- EUtilsSummary("Latex allergy status", type="esearch", db="pubmed", datetype='pdat', mindate=2022, maxdate=2022, retmax=500)
> QueryCount(res_c)
[1] 0
> # placing the no. of articles based on year in arrays
> max1.art = c(3, 17, 4)
> max2.art = c(2, 12, 4)
> max3.art = c(0,8,0)
> # creating a data frame with the previous arrays
> datafr<- data.frame(max1.art,max2.art, max3.art)
```

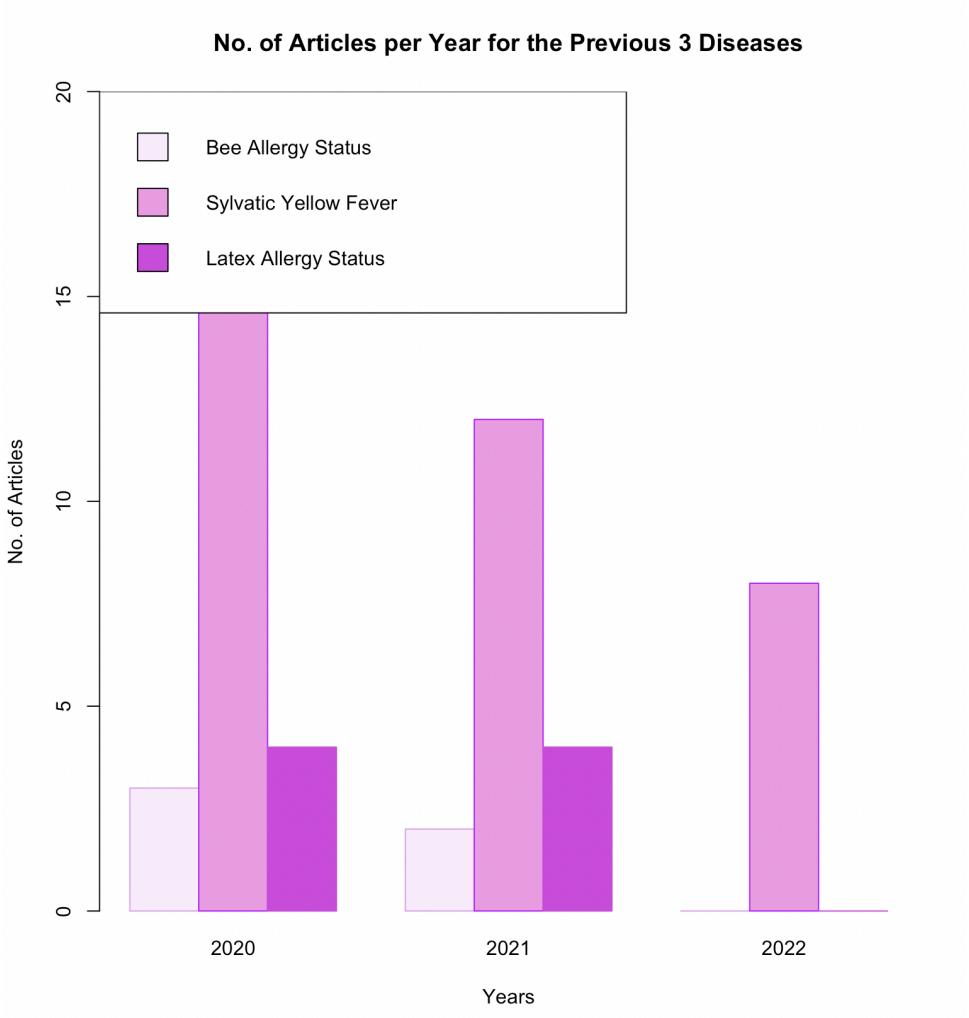
Έπειτα, έφτιαξα τρεις μονοδιάστατους **πίνακες** στους οποίους έβαλα τον αριθμό των άρθρων της αντίστοιχης χρονίας. Δηλαδή: το 2020 για την πρώτη ασθένεια δημοσιεύτηκαν 3 άρθρα, για την δεύτερη 12, για την τρίτη 4. Άρα ο πρώτος πίνακας είναι [3,17,4]. Το ίδιο έκανα και για τις χρονιές 2021 και 2022.

Έπειτα, δημιούργησα ένα **data framework** με τους πίνακες αυτούς.

Μέσω της εντολής barplot και legent, εκτύπωσα και στοίχησα το ακόλουθο διάγραμμα. Δίνοντας κάποια έξτρα ορίσματα άλλαξα χρώματα & έβαλα τίτλους στους άξονες.

(Η barplot ως πρώτο όρισμα χρειάζεται το data framework, γι'αυτό είχε ακολουθήσει η παραπάνω διαδικασία με τους πίνακες.)

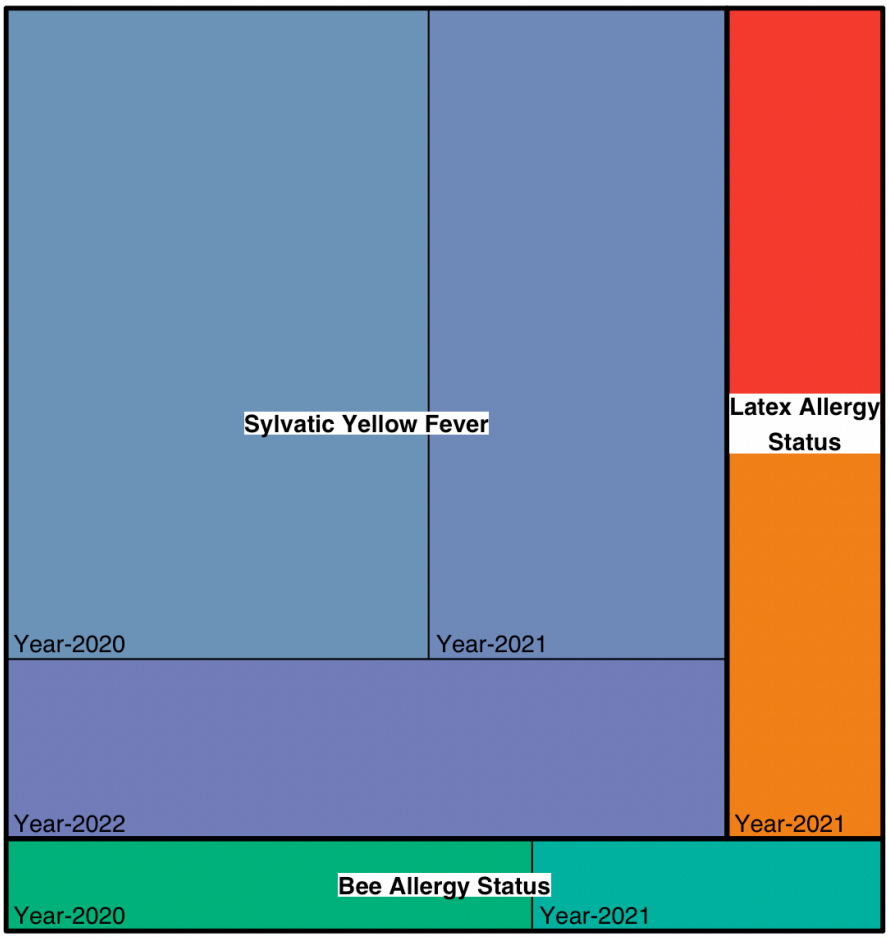
```
> barplot(as.matrix(datafr), main = "No. of Articles per Year for the Previous 3 Diseases", xla
b = "Years", names.arg = c("2020", "2021", "2022"), ylab = "No. of Articles", ylim = c(0,20), c
ol = c("#f7eefb","#dda0dd", "#ba55d3"), border = c("#cea6e1","purple", "#ba55d3"), beside = TRU
E)
> legend("topleft",
+        c("Bee Allergy Status", "Sylvatic Yellow Fever", "Latex Allergy Status"),
+        fill = c("#f7eefb","#dda0dd", "#ba55d3"),
+ )
```



Β.5 Για το **treemap**, χρειάστηκα το πακέτο ggplot2 -και προεραιτικά, αν θέλουμε να γίνει διαδραστικό, το πακέτο d3treeR. Στην παρακάτω φωτογραφία φαίνεται αναλυτικά ο κώδικας:

```
> # b.5
> #install.packages("treemap")
> library(treemap)
> group<-c(rep("Bee Allergy Status", 3), rep("Sylvatic Yellow Fever", 3), rep("Latex Allergy St
atus", 3))
> subgroup<-paste("Year", c(2020,2021,2022,2020,2021,2022,2020,2021,2022), sep = "-")
> value<- c(3,2,0,17,12,8,4,4,0)
> ndatafr<- data.frame(group, subgroup, value)
> # basic treemap
> p<-treemap(ndatafr, index = c("group", "subgroup"), vSize = "value", type = "index", palette
 = "Set2", bg.labels = c("white"), align.labels = list(c("center","center"), c("left", "botto
m")))
```

Δημιουργούμε ένα νέο data framework με κατηγορίες, υποκατηγορίες, τιμές και θέτουμε πώς θέλουμε να χωρίζονται σε "κουτάκια".
Η κατηγορία **group** περιέχει τα ονόματα των "εξωτερικών κουτιών", τα οποία χωρίζονται σε υποκατηγορίες **subgroups** "εσωτερικών κουτιιών".
Οι ονομασίες των groups έιναι τα ονόματα των ασθενειών, ενώ των subgroups είναι τα έτη 2020, 2021 και 2022. Δηλαδή τα groups χωρίζονται με βάση τις ασθένειες, ενώ τα subgroups με βάση τα έτη.
Έπειτα, χρειαζόμαστε κάποια **values**, κάποιες τιμές ουσιαστικά, που δίνουν πόσα άρθρα δημοσιεύτηκαν κάθε χρονιά. Δηλαδή έτσι κάθε subgroup "κουτί" θα έχει μία τιμή "μέσα του" που θα αναφέρεται στον αριθμό άρθρων για τη συγκεκριμένη χρονιά και ασθένεια.
Με αυτά, λοιπόν, group, subgroup & value, δημιουργούμε το data framework και το treemap.



Παρατηρούμε ότι δύο ασθένειες, ενώ τις θέσαμε όλες να έχουν 3 υποκατηγορίες (subgroups), έχουν μόνο δύο. Αυτό συμβάινει, διότι οι τιμές (values) -δηλ. τα άρθρα που δημοσιέυτηκαν την συγκεκριμένη χρονιά- είναι μηδέν, άρα δεν φαίνεται η υποκατηγορία στις ασθένειες που δεν έχουν άρθρα για κάποια χρονιά.

Τελικά , όπως ζήταγε η εκφώνηση : το treemap αποτελείται από 3 κατηγορίες (1 για την κάθε ασθένεια) και το κάθε group θα αποτελείται από 3 subgroups (1 για κάθε έτος).

Αν θέλαμε να το κάνουμε "διαδραστικό" το treemap θα μπορούσαμε μέσω του πακέτου-βιβλιοθήκης d3treeR να χρησιμοποιήσουμε την εντολή inter<- d3tree2(p, rootname="General"), όπου rootname γίνεται ο τίτλος του διαγράμματος.