



Ακρωνύμιο πρότασης:

DataPopEU

Τίτλος πρότασης:

**Καινοτόμες μέθοδοι και δεδομένα υψηλής ποιότητας για τη μελέτη του
λαϊκισμού και του Ευρωσκεπτικισμού**

**Έκθεση σχετικά με τη διαδικασία ανάλυσης των δεδομένων
του Twitter (R code report)**

Ιούλιος 2022



ΔΕΔΟΜΕΝΑ

Στόχος της παρούσας έρευνας ήταν να ανιχνεύσει και να αναλύσει το λαϊκισμό και τον ευρωσκεπτικισμό όπως αυτοί αποτυπώνονται στα tweets των ελλήνων υποψηφίων βουλευτών και στα tweets των ελλήνων χρηστών του Twitter.

Τα tweets των υποψηφίων βουλευτών αφορούν τους υποψήφιους των βουλευτικών εκλογών του Ιουλίου του 2019 και του (...μήνας...) του 2015. Η συλλογή των tweets για τους υποψήφιους του 2019 ξεκίνησε στις 26-05-2019, την περίοδο πριν από τις εκλογές και ολοκληρώθηκε στις 4-08-2021 (συλλογή 2015;). Από την πλευρά των χρηστών η συλλογή των δεδομένων έγινε με τρεις τρόπους: 1) Με τις λέξεις κλειδιά «λαο» και «λαϊκ» 2) με τις λέξεις κλειδιά «ευρω» και «ευρωπ» και 3) με τη λέξη κλειδί «εκλογ». Για την πρώτη και δεύτερη κατηγορία τα δεδομένα συλλέχθηκαν την περίοδο 22-06 με 20-08 του 2019. Για την τρίτη, συλλέχθηκαν την περίοδο 22-06 με 07-07 του 2019. Ο κώδικας για τη συλλογή όλων των παραπάνω παρουσιάστηκε σε προηγούμενη αναφορά.

Το βασικό μας ερευνητικό ερώτημα είναι να συγκρίνουμε τα tweets των υποψηφίων βουλευτών του 2019 με αυτά των απλών χρηστών του Twitter όσον αφορά το λαϊκισμό και τον ευρωσκεπτικισμό, όπως επίσης και τα tweets των υποψηφίων του 2019 με τα tweets των υποψηφίων του 2015.

Τέλος, τα κόμματα τα οποία θα εξεταστούν παρακάτω είναι μόνο η ΝΔ, ο ΣΥΡΙΖΑ και το ΠΑΣΟΚ – ΚΙΝΑΛ, καθώς αυτά είναι τα μόνα κοινά μεταξύ των δύο εκλογικών αναμετρήσεων που μελετάμε. Το ΚΚΕ, μολονότι είναι κοινό, εξαιρέθηκε της συλλογής και ανάλυσης των δεδομένων, καθώς η κομματική γραμμή δεν επιτρέπει στους υποψήφιους του τη διεξαγωγή προεκλογικής καμπάνιας στα social media.

ΜΕΘΟΔΟΛΟΓΙΑ

Πριν την ανάλυση συναισθήματος που θα παρουσιαστεί παρακάτω, έγινε καθάρισμα των δεδομένων. Το καθάρισμα σαν διαδικασία περιλαμβάνει την αφαίρεση όλου του «θορύβου», όπως οι πιο συχνές λέξεις («stopwords») (π.χ. άρθρα, αντωνυμίες κ.λπ.), τα σημεία στίξης, τα κεφαλαία γράμματα και οι τόνοι. Όλα τα παραπάνω έγιναν και για τα δεδομένα της παρούσας έρευνας. Για την αφαίρεση των stopwords χρησιμοποιήσαμε λίστα που περιέχει όλες τις ελληνικές stopwords¹.

Η ανάλυση συναισθήματος είναι μία τεχνική για την ανίχνευση συναισθημάτων σε όγκους κειμένων (π.χ. θυμός, φόβος, χαρά κ.λπ.), αλλά και πόλωσης (polarity). Η πόλωση συνήθως μετριέται σε μια κλίμακα από -1 (αρνητικό) έως 1 (θετικό) και χρησιμοποιείται για να φανεί αν το γενικό συναίσθημα ενός κειμένου (ή ενός συνόλου κειμένων) είναι θετικό ή αρνητικό. Η ανάλυση συναισθήματος ανήκει στις λεγόμενες λεξικές μεθόδους (dictionary methods) (Grimmer & Stewart, 2013). Υπάρχουν λεξικά, όπου σε κάθε λέξη αποδίδεται ένα σκορ συναισθήματος (polarity score), συνήθως από -1

¹ <https://github.com/stopwords-iso/stopwords-el>

έως 1. Στη συνέχεια, οι λέξεις του λεξικού καταμετρούνται στο κείμενο του οποίου ο ερευνητής θέλει να υπολογίσει το συναίσθημα και υπολογίζεται ένα μέσο σκορ συναισθήματος με βάση το σκορ της κάθε λέξης.

Στην παρούσα έρευνα, ο λόγος για τον οποίο επιλέχθηκε η ανάλυση συναισθήματος για την ανίχνευση του λαϊκισμού και του ευρωσκεπτικισμού είναι γιατί έγινε η υπόθεση ότι σε ένα tweet στο οποίο υπάρχει το σχήμα του λαϊκισμού (ένας «αγνός» λαός εναντίον μιας «εχθρικής» ελίτ) θα είναι ενδεχομένως πιο πιθανό να υπάρχει αρνητισμός λόγω της εχθρότητας προς τις/την ελίτ που επιβουλεύονται/εται με εχθρικό τρόπο το λαό. Το ίδιο ισχύει και για τον εντοπισμό του ευρωσκεπτικισμού. Δηλαδή, ένα tweet το οποίο εμπριέχει ευρωσκεπτικισμό είναι πιο πιθανό να χαρακτηρίζεται από αρνητισμό απέναντι στην Ευρωπαϊκή Ένωση. Συνεπώς, η βασική μας παραδοχή είναι ότι περισσότερος αρνητισμός συνεπάγεται περισσότερο λαϊκισμό/ευρωσκεπτικισμό και περισσότερος θετικισμός το αντίθετο. Προφανώς η εν λόγω παραδοχή γίνεται σε ένα πιο γενικό επίπεδο και κατά την διάρκεια των αναλύσεων θα «χαλαρώνει» για να είμαστε σε θέση να εντοπίζουμε ευκολότερα τις περιπτώσεις εκείνες που δεν ακολουθούν τον «κανόνα». Τέλος, να σημειωθεί ότι δεν έγινε ανάλυση συναισθήματος για ύπαρξη επιμέρους συναισθημάτων, αλλά για να υπολογιστεί η γενικότερη πόλωση (polarity) των κειμένων.

Για την πραγματοποίηση της ανάλυσης συναισθήματος χρησιμοποιήθηκε ένα λεξικό για την ελληνική γλώσσα². Το συγκεκριμένο λεξικό συντάχθηκε από τέσσερις αξιολογητές και ο καθένας κωδικοποίησε κάθε λέξη με μία από τις ακόλουθες κατηγορίες: “POS”, “NEG”, “N/A”, δηλαδή θετικό, αρνητικό και ουδέτερο, αντίστοιχα. Προκειμένου να γίνει ευκολότερη η ανάλυση, μετατρέψαμε τις κατηγορίες σε αριθμούς -1,1 και 0 αντίστοιχα και βγάλαμε το μέσο όρο των τεσσάρων αξιολογητών για κάθε λέξη. Στη συνέχεια, ενώσαμε τις λέξεις του λεξικού με τις λέξεις των άρθρων που συλλέξαμε³. Οι λέξεις στο λεξικό είναι περιοριστικές γιατί είναι μόνο στο πρώτο πρόσωπο της οριστικής ενεστώτα αν είναι ρήμα (π.χ. κάνω) και μόνο στο αρσενικό γένος αν είναι επίθετο (π.χ. καλός) και μόνο στον ενικό αν είναι ουσιαστικό (π.χ. το πράγμα). Συνεπώς, για να γίνει η ένωση πιο αποτελεσματικά χρησιμοποιήσαμε δύο μεθόδους.

Η πρώτη είναι η μέθοδος “Adaptive Fuzzy String Matching” των Kaufman και Klevs (Kaufman & Klevs, 2021). Η μέθοδος αυτή δημιουργήθηκε για τη λύση του προβλήματος της ένωσης δύο βάσεων δεδομένων με βάση μόνο μια κοινή μεταβλητή η οποία όμως είναι ατελής και πολύ διαφορετική μεταξύ των δύο βάσεων, ένα κοινό πρόβλημα στο πεδίο της ανάλυσης δεδομένων. Για παράδειγμα, μπορεί σε μία βάση η κοινή μεταβλητή να περιέχει ονοματεπώνυμο με ολόκληρο το επώνυμο, ενώ στην άλλη να περιέχει ονοματεπώνυμο με μόνο το πρώτο γράμμα του επώνυμου. Συνεπώς, ο εν λόγω αλγόριθμος είναι σε θέση να αποφασίσει ότι η παρατήρηση «Σταύρος Μ.» είναι κοινή με την παρατήρηση «Σταύρος Μάρκου» ή ότι η παρατήρηση «JP Morgan» είναι κοινή με την παρατήρηση «JPM». Ο τρόπος που αυτό επιτυγχάνεται είναι μέσω της

² <https://github.com/MKLab-ITI/greek-sentiment-lexicon>

³ Για να πραγματοποιηθεί αυτό, μετατρέψαμε τις λέξεις των άρθρων σε πινακοειδή μορφή έτσι ώστε κάθε γραμμή να είναι και μία λέξη (one-word-per-row format).

επιστράτευσης μιας σειράς δεικτών όπως cosine similarity, Jaccard κ.λπ. και της εισαγωγής τους σε ένα μοντέλο μηχανικής μάθησης για την πρόβλεψη της πιθανότητας δύο λέξεις να ταιριάζουν ή όχι. Ο συγκεκριμένος αλγόριθμος είναι διαθέσιμος στην R μέσω της βιβλιοθήκης “stringmatch”. Παρακάτω παραθέτουμε τον κώδικα που συγγράψαμε για την ένωση του λεξικού με τα tweets που συλλέξαμε, χρησιμοποιώντας την μέθοδο των Kaufman και Levs:

```
library(stringmatch)

matched_words = data.frame()

for (i in 1:122405) {

  stringmatch <- as.data.frame(stringmatch(text_cleaned$Words[i], sentiments$Term))

  stringmatch <- stringmatch %>% filter((osa <= 1) & (cosine <= 0.25) & (jw <= 0.20) &
(pred >= 0.90) & (jaccard <= 0.20))

  matched_words = bind_rows(matched_words, stringmatch)

}
```

Να επισημανθεί ότι ο συγκεκριμένος κώδικας παίρνει ώρα να εκτελεστεί λόγω των πολλών (μοναδικών) λέξεων που πρέπει να ενωθούν, οπότε χρειάζεται πολλούς πόρους (επεξεργαστή και μνήμη RAM). Ωστόσο επισυνάπτουμε ένα αρχείο με όνομα “matched_words”, το οποίο περιέχει το αποτέλεσμα του παραπάνω κώδικα για να μη χρειαστεί κάποιος να τον τρέξει ξανά.

Αφού ενώθηκαν οι λέξεις, υπολογίσαμε το μέσο όρο των σκορ των λέξεων για κάθε tweet. Το αποτέλεσμα ήταν τα άρθρα να πάρουν ένα μέσο σκορ σε μια κλίμακα από -3,25 έως 3,25.

Η δεύτερη μέθοδος που χρησιμοποιήσαμε για την αποτελεσματικότερη ένωση των λέξεων των tweets με το λεξικό ήταν η ριζοποίηση (stemming). Ο μοναδικός αλγόριθμος για την ριζοποίηση των λέξεων στην ελληνική γλώσσα είναι του Νταή (Ntais, 2006), ο οποίος εφαρμόζει τον γνωστό αλγόριθμο του Porter για την ριζοποίηση (Porter, 1980) στην ελληνική γλώσσα. Ο κώδικας για την εφαρμογή του αλγόριθμου που έχει συγγράψει είναι διαθέσιμη μόνο στη γλώσσα προγραμματισμού JavaScript. Οι Ανδρεάδης και Μακρόπουλος (βιβλιογραφία), έχουν μεταφέρει τον αλγόριθμο στη γλώσσα προγραμματισμού R με τη δημιουργία μιας συνάρτησης με το όνομα «greekstemmer», και έτσι εφαρμόσαμε τον αλγόριθμο με τον κώδικα στην R, καθώς αυτή είναι η γλώσσα προγραμματισμού που χρησιμοποιούμε κατ’ εξοχήν στα πλαίσια του εν λόγω ερευνητικού προγράμματος. Παρακάτω επισυνάπτουμε τον κώδικα για την ριζοποίηση στην R:

```
greekstemmer<-function(x){
```

```

orix<-x
#step2a *
#regexr("^(.+?)(ΑΔΕΣ|ΑΔΩΝ)$", c("ΜΠΑΜΠΑΔΕΣ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexr("^(.+?)(ΑΔΕΣ|ΑΔΩΝ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-
regexr("(OK|MAM|MAN|ΜΠΑΜΠ|ΠΑΤΕΡ|ΓΙΑΓΙ|ΝΤΑΝΤ|ΚΥΡ|ΘΕΙ|ΠΕΘΕΡ)$",
x[res!=-1], perl=TRUE)
x[res!=-1][res2==-1]<-paste0(x[res!=-1][res2==-1], "ΑΔ")
#step2b *
#regexr("^(.+?)(ΕΔΕΣ|ΕΔΩΝ)$", c("ΚΕΦΤΕΔΕΣ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexr("^(.+?)(ΕΔΕΣ|ΕΔΩΝ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-regexr("(ΟΠ|ΙΠ|ΕΜΠ|ΥΠ|ΓΗΠ|ΔΑΠ|ΚΡΑΣΠ|ΜΙΑ)$", x[res!=-1], perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΕΔ")
#step2c *
#regexr("^(.+?)(ΟΥΔΕΣ|ΟΥΔΩΝ)$", c("ΠΑΛΕΞΟΥΔΕΣ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexr("^(.+?)(ΟΥΔΕΣ|ΟΥΔΩΝ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-
regexr("(ΑΡΚ|ΚΑΛΙΑΚ|ΠΕΤΑΛ|ΛΙΧ|ΠΛΕΞ|ΣΚ|Σ|ΦΛ|ΦΡ|ΒΕΛ|ΛΟΥΛ|ΧΝ|ΣΠ|ΤΡΑΓ|
ΦΕ)$", x[res!=-1], perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΟΥΔ")
#step2d *
#regexr("^(.+?)(ΕΩΣ|ΕΩΝ)$", c("ΒΑΣΙΛΕΩΣ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexr("^(.+?)(ΕΩΣ|ΕΩΝ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))

```

```

x[res!=-1]
res2<-regexpr("^(Θ|Δ|ΕΛ|ΓΑΛ|Ν|Π|ΙΔ|ΠΑΡ)$", x[res!=-1], perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "E")
#step3 *
#regexpr("^(.+?)(ΙΑ|ΙΟΥ|ΙΩΝ)$", c("ΠΑΙΔΙΑ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexpr("^(.+?)(ΙΑ|ΙΟΥ|ΙΩΝ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-regexpr("(Α|Ε|Η|Ι|Ο|Υ|Ω)$", x[res!=-1], perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "I")
#step4 *
#regexpr("^(.+?)(ΙΚΑ|ΙΚΟ|ΙΚΟΥ|ΙΚΩΝ)$", c("ΚΙΚΙΡΙΚΟΥ", "ΑΔΕΛΦΕΣ"),
perl=TRUE)
## ADDED ΘΕΤ IN RES2
res<-regexpr("^(.+?)(ΙΚΑ|ΙΚΟ|ΙΚΟΥ|ΙΚΩΝ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-
regexpr("^(ΑΛ|ΑΔ|ΕΝΔ|ΑΜΑΝ|ΑΜΜΟΧΑΛ|ΗΘ|ΑΝΗΘ|ΑΝΤΙΔ|ΦΥΣ|ΒΡΩΜ|ΓΕΡ|ΕΞ
ΩΔ|ΘΕΤ|ΚΑΛΠ|ΚΑΛΛΙΝ|ΚΑΤΑΔ|ΜΟΥΛ|ΜΠΑΝ|ΜΠΙΑΓΙΑΤ|ΜΠΟΛ|ΜΠΟΣ|ΝΙΤ|ΞΙΚ
|ΣΥΝΟΜΗΛ|ΠΕΤΣ|ΠΙΤΣ|ΠΙΚΑΝΤ|ΠΛΙΑΤΣ|ΠΟΣΤΕΛΝ|ΠΡΩΤΟΔ|ΣΕΡΤ|ΣΥΝΑΔ|ΤΣΑ
Μ|ΥΠΟΔ|ΦΙΛΟΝ|ΦΥΛΟΔ|ΧΑΣ)$", x[res!=-1], perl=TRUE)
res3<-regexpr("(Α|Ε|Η|Ι|Ο|Υ|Ω)$", x[res!=-1], perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "IK")
x[res!=-1][res3!=-1]<-paste0(x[res!=-1][res3!=-1], "IK")
#step5a1 *
#regexpr("^(.+?)(ΑΓΑΜΕ|ΗΣΑΜΕ|ΟΥΣΑΜΕ|ΗΚΑΜΕ|ΗΘΗΚΑΜΕ)$",
c("ΑΓΑΠΗΘΗΚΑΜΕ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexpr("^(.+?)(ΑΓΑΜΕ|ΗΣΑΜΕ|ΟΥΣΑΜΕ|ΗΚΑΜΕ|ΗΘΗΚΑΜΕ)$", x,
perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))

```

```

x[res!=-1]
#step5a2 *
#regexpr("^(.+?)(AME)$", c("ΤΡΑΓΟΥΔΑΜΕ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexpr("^(.+?)(AME)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-
regexpr("^(ΑΝΑΠ|ΑΠΟΘ|ΑΠΟΚ|ΑΠΟΣΤ|ΒΟΥΒ|ΞΕΘ|ΟΥΛ|ΠΕΘ|ΠΙΚΡ|ΠΟΤ|ΣΙΧ|Χ)$", x[res!=-1], perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΑΜ")
#step5b1 *
#regexpr("^(.+?)(ΑΓΑΝΕ|ΗΣΑΝΕ|ΟΥΣΑΝΕ|ΙΟΝΤΑΝΕ|ΙΟΤΑΝΕ|ΙΟΥΝΤΑΝΕ|ΟΝΤΑΝΕ|ΟΤΑΝΕ|ΟΥΝΤΑΝΕ|ΗΚΑΝΕ|ΗΘΗΚΑΝΕ)$", c("ΑΓΑΠΗΘΗΚΑΝΕ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-
regexpr("^(.+?)(ΑΓΑΝΕ|ΗΣΑΝΕ|ΟΥΣΑΝΕ|ΙΟΝΤΑΝΕ|ΙΟΤΑΝΕ|ΙΟΥΝΤΑΝΕ|ΟΝΤΑΝΕ|ΟΤΑΝΕ|ΟΥΝΤΑΝΕ|ΗΚΑΝΕ|ΗΘΗΚΑΝΕ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-regexpr("^(ΤΡ|ΤΣ)$", x[res!=-1], perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΑΓΑΝ")
#step5b2 *
#regexpr("^(.+?)(ΑΝΕ)$", c("ΤΡΑΓΟΥΔΑΝΕ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexpr("^(.+?)(ΑΝΕ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-regexpr("(Α|Ε|Η|Ι|Ο|Υ|Ω)$", x[res!=-1], perl=TRUE)
res3<-
regexpr("^(ΒΕΤΕΡ|ΒΟΥΛΚ|ΒΡΑΧΜ|Γ|ΔΡΑΔΟΥΜ|Θ|ΚΑΛΠΟΥΖ|ΚΑΣΤΕΛ|ΚΟΡΜΟΡ|ΛΑΟΠΛ|ΜΩΑΜΕΘ|Μ|ΜΟΥΣΟΥΛΜ|Ν|ΟΥΛ|Π|ΠΕΛΕΚ|ΠΛ|ΠΟΛΙΣ|ΠΟΡΤΟΛ|ΣΑΡΑΚΑΤΣ|ΣΟΥΛΤ|ΤΣΑΡΛΑΤ|ΟΡΦ|ΤΣΙΓΓ|ΤΣΟΠ|ΦΩΤΟΣΤΕΦ|Χ|ΨΥΧΟΠΛ|ΑΓ|ΟΡΦ|ΓΑΛ

```

```
|ΓΕΡ|ΔΕΚ|ΔΙΠΛ|ΑΜΕΡΙΚΑΝ|ΟΥΡ|ΠΙΘ|ΠΟΥΡΙΤ|Σ|ΖΩΝΤ|ΙΚ|ΚΑΣΤ|ΚΟΠ|ΛΙΧ|ΛΟΥΘ  
ΗΡ|ΜΑΙΝΤ|ΜΕΛ|ΣΙΓ|ΣΠ|ΣΤΕΓ|ΤΡΑΓ|ΤΣΑΓ|Φ|ΕΡ|ΑΔΑΠ|ΑΘΙΓΓ|ΑΜΗΧ|ΑΝΙΚ|ΑΝΟΡ  
Γ|ΑΠΗΓ|ΑΠΙΘ|ΑΤΣΙΓΓ|ΒΑΣ|ΒΑΣΚ|ΒΑΘΥΓΑΛ|ΒΙΟΜΗΧ|ΒΡΑΧΥΚ|ΔΙΑΤ|ΔΙΑΦ|ΕΝΟ  
ΡΓ|ΘΥΣ|ΚΑΙΠΝΟΒΙΟΜΗΧ|ΚΑΤΑΓΑΛ|ΚΛΙΒ|ΚΟΙΛΑΡΦ|ΛΙΒ|ΜΕΓΛΟΒΙΟΜΗΧ|ΜΙΚΡ  
ΟΒΙΟΜΗΧ|ΝΤΑΒ|ΞΗΡΟΚΛΙΒ|ΟΛΙΓΟΔΑΜ|ΟΛΟΓΑΛ|ΠΕΝΤΑΡΦ|ΠΕΡΗΦ|ΠΕΡΙΤΡ|Π  
ΛΑΤ|ΠΟΛΥΔΑΠ|ΠΟΛΥΜΗΧ|ΣΤΕΦ|ΤΑΒ|ΤΕΤ|ΥΠΕΡΗΦ|ΥΠΟΚΟΠ|ΧΑΜΗΛΟΔΑΠ|  
ΨΗΛΟΤΑΒ)$", x[res!=-1], perl=TRUE)
```

```
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "AN")
```

```
x[res!=-1][res3!=-1]<-paste0(x[res!=-1][res3!=-1], "AN")
```

```
#step5cA *
```

```
#regexpr("^(.+?)(ΗΣΕΤΕ)$", c("ΤΡΑΓΟΥΔΗΣΕΤΕ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
```

```
res<-regexpr("^(.+?)(ΗΣΕΤΕ)$", x, perl=TRUE)
```

```
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
```

```
x[res!=-1]
```

```
#step5cB *
```

```
#regexpr("^(.+?)(ΕΤΕ)$", c("ΠΑΙΖΕΤΕ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
```

```
res<-regexpr("^(.+?)(ΕΤΕ)$", x, perl=TRUE)
```

```
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
```

```
x[res!=-1]
```

```
res2<-regexpr("(A|E|H|I|O|Ω)$", x[res!=-1], perl=TRUE)
```

```
res3<-
```

```
regexpr("(ΟΔ|ΑΙΡ|ΦΟΡ|ΤΑΘ|ΔΙΑΘ|ΣΧ|ΕΝΔ|ΕΥΡ|ΤΙΘ|ΥΠΕΡΘ|ΡΑΘ|ΕΝΘ|ΡΟΘ|ΣΘ|ΠΥΡ  
|ΑΙΝ|ΣΥΝΔ|ΣΥΝ|ΣΥΝΘ|ΧΩΡ|ΠΙΟΝ|ΒΡ|ΚΑΘ|ΕΥΘ|ΕΚΘ|ΝΕΤ|ΡΟΝ|ΑΡΚ|ΒΑΡ|ΒΟΛ|ΩΦ  
ΕΛ)$", x[res!=-1], perl=TRUE)
```

```
res4<-
```

```
regexpr("^(ΑΒΑΡ|ΒΕΝ|ΕΝΑΡ|ΑΒΡ|ΑΔ|ΑΘ|ΑΝ|ΑΠΛ|ΒΑΡΟΝ|ΝΤΡ|ΣΚ|ΚΟΠ|ΜΠΟΡ|ΝΙ  
Φ|ΠΑΓ|ΠΑΡΑΚΑΛ|ΣΕΡΠ|ΣΚΕΛ|ΣΥΡΦ|ΤΟΚ|ΥΔ|ΕΜ|ΘΑΡΡ|Θ)$", x[res!=-1],  
perl=TRUE)
```

```
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ET")
```

```
x[res!=-1][res3!=-1]<-paste0(x[res!=-1][res3!=-1], "ET")
```

```
x[res!=-1][res4!=-1]<-paste0(x[res!=-1][res4!=-1], "ET")
```

```
#step5d *
```



```
#regexpr("^(.+?)(ΟΝΤΑΣ|ΩΝΤΑΣ)$", c("ΠΟΚΑΧΟΝΤΑΣ", "ΑΔΕΛΦΕΣ"),  
perl=TRUE)
```

```
res<-regexpr("^(.+?)(ΟΝΤΑΣ|ΩΝΤΑΣ)$", x, perl=TRUE)
```

```
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
```

```
x[res!=-1]
```

```
res2<-regexpr("^(ΑΡΧ)$", x[res!=-1], perl=TRUE)
```

```
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΟΝΤ")
```

```
res3<-regexpr("(ΚΡΕ)$", x[res!=-1], perl=TRUE)
```

```
x[res!=-1][res3!=-1]<-paste0(x[res!=-1][res3!=-1], "ΩΝΤ")
```

```
#step5e *
```

```
#regexpr("^(.+?)(ΟΜΑΣΤΕ|ΙΟΜΑΣΤΕ)$", c("ΑΓΑΠΙΟΜΑΣΤΕ", "ΑΔΕΛΦΕΣ"),  
perl=TRUE)
```

```
res<-regexpr("^(.+?)(ΟΜΑΣΤΕ|ΙΟΜΑΣΤΕ)$", x, perl=TRUE)
```

```
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
```

```
x[res!=-1]
```

```
res2<-regexpr("^(ΟΝ)$", x[res!=-1], perl=TRUE)
```

```
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΟΜΑΣΤ")
```

```
#step5fA *
```

```
#regexpr("^(.+?)(ΙΕΣΤΕ)$", c("ΑΓΑΠΙΕΣΤΕ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
```

```
res<-regexpr("^(.+?)(ΙΕΣΤΕ)$", x, perl=TRUE)
```

```
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
```

```
x[res!=-1]
```

```
res2<-regexpr("^(Π|ΑΠ|ΣΥΜΠ|ΑΣΥΜΠ|ΑΚΑΤΑΠ|ΑΜΕΤΑΜΦ)$", x[res!=-1],  
perl=TRUE)
```

```
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΙΕΣΤ")
```

```
#step5fB *
```

```
#regexpr("^(.+?)(ΕΣΤΕ)$", c("ΠΙΕΣΤΕ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
```

```
res<-regexpr("^(.+?)(ΕΣΤΕ)$", x, perl=TRUE)
```

```
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
```

```

x[res!=-1]
res2<-regexpr("^(ΑΛ|ΑΡ|ΕΚΤΕΛ|Ζ|Μ|Ξ|ΠΑΡΑΚΑΛ|ΑΡ|ΠΡΟ|ΝΙΣ)$", x[res!=-1],
perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΕΣΤ")
#step5gA *
#regexpr("^(.+?)(ΗΘΗΚΑ|ΗΘΗΚΕΣ|ΗΘΗΚΕ)$", c("ΑΓΑΠΗΘΗΚΑ", "ΑΔΕΛΦΕΣ"),
perl=TRUE)
res<-regexpr("^(.+?)(ΗΘΗΚΑ|ΗΘΗΚΕΣ|ΗΘΗΚΕ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
#step5gB *
#regexpr("^(.+?)(ΗΚΑ|ΗΚΕΣ|ΗΚΕ)$", c("ΑΝΕΒΗΚΑ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexpr("^(.+?)(ΗΚΑ|ΗΚΕΣ|ΗΚΕ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-regexpr("(ΣΚΩΛ|ΣΚΟΥΛ|ΝΑΡΘ|ΣΦ|ΟΘ|ΠΙΘ)$", x[res!=-1], perl=TRUE)
res3<-regexpr("^(ΔΙΑΘ|Θ|ΠΑΡΑΚΑΤΑΘ|ΠΡΟΣΘ|ΣΥΝΘ|)$", x[res!=-1], perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΗΚ")
x[res!=-1][res3!=-1]<-paste0(x[res!=-1][res3!=-1], "ΗΚ")
#step5h *
#regexpr("^(.+?)(ΟΥΣΑ|ΟΥΣΕΣ|ΟΥΣΕ)$", c("ΠΗΔΟΥΣΑ", "ΑΔΕΛΦΕΣ"),
perl=TRUE)
res<-regexpr("^(.+?)(ΟΥΣΑ|ΟΥΣΕΣ|ΟΥΣΕ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-
regexpr("^(ΦΑΡΜΑΚ|ΧΑΔ|ΑΓΚ|ΑΝΑΡΡ|ΒΡΟΜ|ΕΚΛΙΠ|ΛΑΜΠΙΔ|ΛΕΧ|Μ|ΠΑΤ|Ρ|Λ|Μ
ΕΔ|ΜΕΣΑΖ|ΥΠΟΤΕΙΝ|ΑΜ|ΑΙΘ|ΑΝΗΚ|ΔΕΣΠΟΖ|ΕΝΔΙΑΦΕΡ|ΔΕ|ΔΕΥΤΕΡΕΥ|ΚΑΘΑΡ
ΕΥ|ΠΛΕ|ΤΣΑ)$", x[res!=-1], perl=TRUE)

```

```

res3<-
regexpr("(ΠΟΔΑΡ|ΒΛΕΠ|ΠΑΝΤΑΧ|ΦΡΥΔ|ΜΑΝΤΙΑ|ΜΑΛΛ|ΚΥΜΑΤ|ΛΑΧ|ΛΗΓ|ΦΑΓ
|ΟΜ|ΠΡΩΤ)$", x[res!==-1], perl=TRUE)

x[res!==-1][res2!==-1]<-paste0(x[res!==-1][res2!==-1], "ΟΥΣ")
x[res!==-1][res3!==-1]<-paste0(x[res!==-1][res3!==-1], "ΟΥΣ")
#step5i *
#regexpr("^(.+?)(ΑΓΑ|ΑΓΕΣ|ΑΓΕ)$", c("ΠΕΤΑΓΑ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexpr("^(.+?)(ΑΓΑ|ΑΓΕΣ|ΑΓΕ)$", x, perl=TRUE)
x[res!==-1]<-substr(x[res!==-1], 1, as.integer(attributes(res)$capture.start[res!==-1,2]-1))
x[res!==-1]
res2<-regexpr("^(ΨΟΦ|ΝΑΥΛΟΧ)$", x[res!==-1], perl=TRUE)
res3<-regexpr("(ΚΟΛΛ)$", x[res!==-1], perl=TRUE)
res4<-
regexpr("^(ΑΒΑΣΤ|ΠΟΛΥΦ|ΑΔΗΦ|ΠΑΜΦ|Ρ|ΑΣΠ|ΑΦ|ΑΜΑΛ|ΑΜΑΛΛΙ|ΑΝΥΣΤ|ΑΠ
ΕΡ|ΑΣΠΑΡ|ΑΧΑΡ|ΔΕΡΒΕΝ|ΔΡΟΣΟΠ|ΞΕΦ|ΝΕΟΠ|ΝΟΜΟΤ|ΟΛΟΠ|ΟΜΟΤ|ΠΡΟΣΤ|ΠΡ
ΟΣΩΠ|ΟΠ|ΣΥΜΠ|ΣΥΝΤ|Τ|ΥΠΟΤ|ΧΑΡ|ΑΕΙΠ|ΑΙΜΟΣΤ|ΑΝΥΠ|ΑΠΟΤ|ΑΡΤΙΠ|ΔΙΑΤ|Ε
Ν|ΕΠΙΤ|ΚΡΟΚΑΛΟΠ|ΣΙΔΗΡΟΠ|Λ|ΝΑΥ|ΟΥΛΑΜ|ΟΥΡ|Π|ΤΡ|Μ)$", x[res!==-1],
perl=TRUE)
res5<-regexpr("(ΟΦ|ΠΕΛ|ΧΟΡΤ|ΛΛ|ΣΦ|ΡΠ|ΦΡ|ΠΡ|ΛΟΧ|ΣΜΗΝ)$", x[res!==-1],
perl=TRUE)
#needs correction
x[res!==-1][res5!==-1][res2=-1]<-paste0(x[res!==-1][res5!==-1][res2=-1], "ΑΓ")
x[res!==-1][res4!==-1][res3=-1]<-paste0(x[res!==-1][res4!==-1][res3=-1], "ΑΓ")
#step5j *
#regexpr("^(.+?)(ΗΣΕ|ΗΣΟΥ|ΗΣΑ)$", c("ΑΦΗΣΕ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexpr("^(.+?)(ΗΣΕ|ΗΣΟΥ|ΗΣΑ)$", x, perl=TRUE)
x[res!==-1]<-substr(x[res!==-1], 1, as.integer(attributes(res)$capture.start[res!==-1,2]-1))
x[res!==-1]
res2<-regexpr("^(Ν|ΧΕΡΣΟΝ|ΔΩΔΕΚΑΝ|ΕΡΗΜΟΝ|ΜΕΓΑΛΟΝ|ΕΠΙΤΑΝ)$", x[res!==-
1], perl=TRUE)
x[res!==-1][res2!==-1]<-paste0(x[res!==-1][res2!==-1], "ΗΣ")

```

```

#step5k *
#regexpr("^(.+?)(ΗΣΤΕ)$", c("ΚΟΙΝΟΧΡΗΣΤΕ", "ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexpr("^(.+?)(ΗΣΤΕ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-
regexpr("^(ΑΣΒ|ΣΒ|ΑΧΡ|ΧΡ|ΑΠΛ|ΑΕΙΜΝ|ΔΥΣΧΡ|ΕΥΧΡ|ΚΟΙΝΟΧΡ|ΠΑΛΙΜΨ)$",
x[res!=-1], perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΗΣΤ")
#step5l *
#regexpr("^(.+?)(ΟΥΝΕ|ΗΣΟΥΝΕ|ΗΘΟΥΝΕ)$", c("ΑΓΑΠΗΘΟΥΝΕ", "ΑΔΕΛΦΕΣ"),
perl=TRUE)
res<-regexpr("^(.+?)(ΟΥΝΕ|ΗΣΟΥΝΕ|ΗΘΟΥΝΕ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-regexpr("^(Ν|Ρ|ΣΠΙ|ΣΤΡΑΒΟΜΟΥΤΣ|ΚΑΚΟΜΟΥΤΣ|ΕΞΩΝ)$", x[res!=-1],
perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΟΥΝ")
#step5l *
#regexpr("^(.+?)(ΟΥΜΕ|ΗΣΟΥΜΕ|ΗΘΟΥΜΕ)$", c("ΑΓΑΠΗΘΟΥΜΕ",
"ΑΔΕΛΦΕΣ"), perl=TRUE)
res<-regexpr("^(.+?)(ΟΥΜΕ|ΗΣΟΥΜΕ|ΗΘΟΥΜΕ)$", x, perl=TRUE)
x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-regexpr("^(ΠΑΡΑΣΟΥΣ|Φ|Χ|ΩΡΙΟΠΛ|ΑΖ|ΑΛΛΟΣΟΥΣ|ΑΣΟΥΣ)$", x[res!=-1],
perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΟΥΜ")
#step6A NA TO ΔΟΥΜΕ
#regexpr("^(.+?)(ΜΑΤΑ|ΜΑΤΩΝ|ΜΑΤΟΣ)$", c("ΠΡΟΒΛΗΜΑΤΑ", "ΑΔΕΛΦΕΣ"),
perl=TRUE)
res<-regexpr("^(.+?)(ΜΑΤΑ|ΜΑΤΩΝ|ΜΑΤΟΣ)$", x, perl=TRUE)

```

```

x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
res2<-regexpr("^(.+?)(ΜΑΤΑ|ΜΑΤΩΝ|ΜΑΤΟΣ)$", x[res!=-1], perl=TRUE)
x[res!=-1][res2!=-1]<-paste0(x[res!=-1][res2!=-1], "ΜΑ")
#step6b  ΝΑ ΤΟ ΔΟΥΜΕ

#regexpr("^(.+?)(Α|ΑΓΑΤΕ|ΑΓΑΝ|ΑΕΙ|ΑΜΑΙ|ΑΝ|ΑΣ|ΑΣΑΙ|ΑΤΑΙ|ΑΩ|Ε|ΕΙ|ΕΙΣ|ΕΙΤΕ|ΕΣΑΙ|ΕΣ|ΕΤΑΙ|Ι|ΙΕΜΑΙ|ΙΕΜΑΣΤΕ|ΙΕΤΑΙ|ΙΕΣΑΙ|ΙΕΣΑΣΤΕ|ΙΟΜΑΣΤΑΝ|ΙΟΜΟΥΝ|ΙΟΜΟΥΝΑ|ΙΟΝΤΑΝ|ΙΟΝΤΟΥΣΑΝ|ΙΟΣΑΣΤΑΝ|ΙΟΣΑΣΤΕ|ΙΟΣΟΥΝ|ΙΟΣΟΥΝΑ|ΙΟΤΑΝ|ΙΟΥΜΑ|ΙΟΥΜΑΣΤΕ|ΙΟΥΝΤΑΙ|ΙΟΥΝΤΑΝ|Η|ΗΔΕΣ|ΗΔΩΝ|ΗΘΕΙ|ΗΘΕΙΣ|ΗΘΕΙΤΕ|ΗΘΗΚΑΤΕ|ΗΘΗΚΑΝ|ΗΘΟΥΝ|ΗΘΩ|ΗΚΑΤΕ|ΗΚΑΝ|ΗΣ|ΗΣΑΝ|ΗΣΑΤΕ|ΗΣΕΙ|ΗΣΕΣ|ΗΣΟΥΝ|ΗΣΩ|Ο|ΟΙ|ΟΜΑΙ|ΟΜΑΣΤΑΝ|ΟΜΟΥΝ|ΟΜΟΟΥΝΑ|ΟΝΤΑΙ|ΟΝΤΑΝ|ΟΝΤΟΥΣΑΝ|ΟΣ|ΟΣΑΣΤΑΝ|ΟΣΑΣΤΕ|ΟΣΟΥΝ|ΟΣΟΥΝΑ|ΟΤΑΝ|ΟΥ|ΟΥΜΑΙ|ΟΥΜΑΣΤΕ|ΟΥΝ|ΟΥΝΤΑΙ|ΟΥΝΤΑΝ|ΟΥΣ|ΟΥΣΑΝ|ΟΥΣΑΤΕ|Υ|ΥΣ|Ω|ΩΝ)$",      c("ΠΡΟΒΛΗΜΑΤΑ", "ΑΔΕΛΦΕΣ"), perl=TRUE)

res<-
regexpr("^(.+?)(Α|ΑΓΑΤΕ|ΑΓΑΝ|ΑΕΙ|ΑΜΑΙ|ΑΝ|ΑΣ|ΑΣΑΙ|ΑΤΑΙ|ΑΩ|Ε|ΕΙ|ΕΙΣ|ΕΙΤΕ|ΕΣΑΙ|ΕΣ|ΕΤΑΙ|Ι|ΙΕΜΑΙ|ΙΕΜΑΣΤΕ|ΙΕΤΑΙ|ΙΕΣΑΙ|ΙΕΣΑΣΤΕ|ΙΟΜΑΣΤΑΝ|ΙΟΜΟΥΝ|ΙΟΜΟΥΝΑ|ΙΟΝΤΑΝ|ΙΟΝΤΟΥΣΑΝ|ΙΟΣΑΣΤΑΝ|ΙΟΣΑΣΤΕ|ΙΟΣΟΥΝ|ΙΟΣΟΥΝΑ|ΙΟΤΑΝ|ΙΟΥΜΑ|ΙΟΥΜΑΣΤΕ|ΙΟΥΝΤΑΙ|ΙΟΥΝΤΑΝ|Η|ΗΔΕΣ|ΗΔΩΝ|ΗΘΕΙ|ΗΘΕΙΣ|ΗΘΕΙΤΕ|ΗΘΗΚΑΤΕ|ΗΘΗΚΑΝ|ΗΘΟΥΝ|ΗΘΩ|ΗΚΑΤΕ|ΗΚΑΝ|ΗΣ|ΗΣΑΝ|ΗΣΑΤΕ|ΗΣΕΙ|ΗΣΕΣ|ΗΣΟΥΝ|ΗΣΩ|Ο|ΟΙ|ΟΜΑΙ|ΟΜΑΣΤΑΝ|ΟΜΟΥΝ|ΟΜΟΟΥΝΑ|ΟΝΤΑΙ|ΟΝΤΑΝ|ΟΝΤΟΥΣΑΝ|ΟΣ|ΟΣΑΣΤΑΝ|ΟΣΑΣΤΕ|ΟΣΟΥΝ|ΟΣΟΥΝΑ|ΟΤΑΝ|ΟΥ|ΟΥΜΑΙ|ΟΥΜΑΣΤΕ|ΟΥΝ|ΟΥΝΤΑΙ|ΟΥΝΤΑΝ|ΟΥΣ|ΟΥΣΑΝ|ΟΥΣΑΤΕ|Υ|ΥΣ|Ω|ΩΝ)$", x, perl=TRUE)

x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
#step7 parathetika  *

#regexpr("^(.+?)(ΕΣΤΕΡ|ΕΣΤΑΤ|ΟΤΕΡ|ΟΤΑΤ|ΥΤΕΡ|ΥΤΑΤ|ΩΤΕΡ|ΩΤΑΤ)$",
c("ΠΙΟΤΕΡ", "ΑΔΕΛΦΕΣ"), perl=TRUE)

res<-regexpr("^(.+?)(ΕΣΤΕΡ|ΕΣΤΑΤ|ΟΤΕΡ|ΟΤΑΤ|ΥΤΕΡ|ΥΤΑΤ|ΩΤΕΡ|ΩΤΑΤ)$",      x,
perl=TRUE)

x[res!=-1]<-substr(x[res!=-1], 1, as.integer(attributes(res)$capture.start[res!=-1,2]-1))
x[res!=-1]
return(x)
}

```

Φαίνεται ότι από άποψη εξοικονόμησης υπολογιστικών πόρων η μέθοδος της ριζοποίησης είναι προτιμότερη (η μέθοδος stringmatch μπορεί να χρειαστεί μέχρι και 2 ώρες για να εκτελεστεί από τον υπολογιστή). Ωστόσο, από άποψη ακρίβειας φαίνεται ότι η μέθοδος stringmatch είναι λίγο περισσότερη ακριβής. Για να δώσουμε ένα παράδειγμα, θα πάρουμε δύο από τις πιο θετικές λέξεις του λεξικού (3.25), τις λέξεις «χαρά» και «χάρη». Οι λέξεις των tweets που ενώθηκαν με το λεξικό με τη μέθοδο stringmatch είναι οι λέξεις χάρης, χάρη, χαρά. Αντίστοιχα οι λέξεις των tweets που ενώθηκαν με τη μέθοδο της ριζοποίησης είναι αρκετά περισσότερες καθώς εντοπίζονται σωστά λέξεις οι οποίες προέρχονται από τις λέξεις χαρά και χάρη, όπως χαρείς, χαρείτε, χάρηκαν, χαρούν, χαρές κ.ο.κ.. Ωστόσο, ενώθηκαν εξίσου πολλές λέξεις οι οποίες δεν προέρχονται από τις λέξεις χαρά και χάρη όπως οι λέξεις χάρος, χάρου, χαράσουμε, χάρι κ.ο.κ. Αυτό έχει ως αποτέλεσμα πολλές λέξεις στα tweets οι οποίες έχουν αρνητική σημασία, όπως π.χ. η λέξη χάρος, να αποκτούν πολύ θετικό σκόρ (3.25) μόνο και μόνο επειδή έχουν την ίδια ρίζα με την αντίστοιχη θετική λέξη του λεξικού.

Αντίστοιχα λάθη συμβαίνουν προφανώς και σε λέξεις που ενώνονται με τη μέθοδο stringmatch. Για παράδειγμα, υπάρχουν δύο πολύ αρνητικές λέξεις στο λεξικό, η λέξη «μισός» και η λέξη «μισώ». Σε αυτή την περίπτωση με τη μέθοδο stringmatch οι λάθος λέξεις που ενώνονται είναι οι λέξεις μισό και μισοί. Ωστόσο, με την μέθοδο της ριζοποίησης οι λάθος λέξεις που ενώνονται είναι περισσότερες. Εκτός από το μισό και μισοί ενώνονται και οι λέξεις μισές, μισή, μισής, μισά, μισού. Και σε αυτή την περίπτωση φαίνεται, δηλαδή, τα λάθη της δεύτερης μεθόδου να είναι περισσότερα. Τα λάθη της μεθόδου stringmatch φαίνεται να προέρχονται κυρίως από το γεγονός ότι μπορεί να αποτύχει να διακρίνει την αλλαγή που επέρχεται σε μία λέξη από την προσθαφαίρεση 1 ή 2 γραμμάτων σε αυτή. Πρακτικά, τα λάθη αυτής της μεθόδου θα μπορούσαν να περιοριστούν (σίγουρα όχι να εξαλειφτούν) αν δοκιμαστούν διαφορετικές πιο «καλές» τιμές στους δείκτες jaccard, cosine κ.λπ ή αν ανεβάσουμε κι άλλο τις προβλεφθείσες πιθανότητες του μοντέλου. Από την άλλη τα λάθη που προέρχονται από τη μέθοδο της ριζοποίησης δεν μπορούν εύκολα να διορθωθούν, καθώς οι ρίζες των λέξεων είναι αμετάβλητες και συνεπώς σχεδόν πάντα θα υπάρχουν λέξεις με κοινή ρίζα οι οποίες όμως εκφράζουν εντελώς διαφορετικό συναίσθημα.

ΑΝΑΛΥΣΗ

ΥΠΟΨΗΦΙΟΙ 2019

Τα tweets που συλλέχθηκαν από τους λογαριασμούς των υποψηφίων βουλευτών συνδυάστηκαν με δεδομένα που συλλέχθηκαν από την έρευνα υποψηφίων βουλευτών του 2019, προκειμένου τα πρώτα να εμπλουτιστούν με καινούργιες μεταβλητές που αφορούν ιδιότητες των ελλήνων βουλευτών όπως κόμμα, ημερομηνία πρώτης εκλογής κ.λπ. οι οποίες δεν ήταν διαθέσιμες στα δεδομένα που συλλέχθηκαν μέσω του Twitter API. Παρακάτω παρουσιάζεται η συχνότητα των tweets ανά κόμμα.

Κόμμα	Συχνότητα
-------	-----------

ΚΙΝΑΛ	30517
ΝΔ	119629
ΣΥΡΙΖΑ	51861

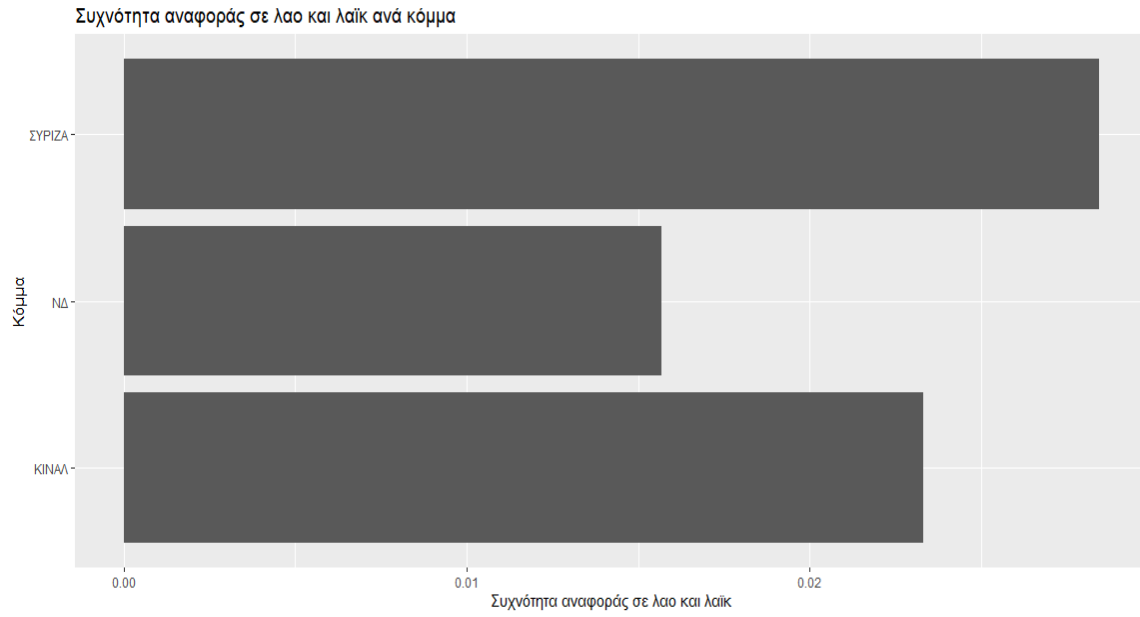
Στη συνέχεια για την αποτελεσματική ανίχνευση του λαϊκισμού και του ευρωσκεπτικισμού προσπαθήσαμε να εντοπίσουμε tweets που περιείχαν τις λέξεις «λαο», «λαϊκ», «ευρω», «ευρωπ». Επίσης, για την αποτελεσματικότερη ανίχνευση του λαϊκισμού χρησιμοποιήσαμε και λέξεις που υποδηλώνουν εχθρούς του λαού όπως «τραπεζ», «καναλαρχ», «διαφθορ», «διεφθαρμ» κ.λπ. Μετά τον εντοπισμό όλων των παραπάνω λέξεων χωρίσαμε τα tweets σε αυτά που αποτελούν retweets και σε αυτά που είναι πρωτότυπα. Οι περισσότερες από τις αναλύσεις που θα παρουσιαστούν παρακάτω έγιναν στο corpus των πρωτότυπων tweets, καθώς υποθέσαμε ότι τα πρωτότυπα tweets είναι αυτά που αποτυπώνουν καλύτερα τον λαϊκισμό και τον ευρωσκεπτικισμό όπως αυτοί εκφράζονται πρωτογενώς από τον εκάστοτε υποψήφιο και άρα είναι πιο πιθανόν να δώσουν αμερόληπτα αποτελέσματα, ενώ τα retweets αποτελούν περισσότερο ένδειξη της απήχησης του λαϊκισμού και του ευρωσκεπτικισμού, αποτελώντας κυρίως ένα μέσο «ανακύκλωσής» του στην σφαίρα του Twitter.

Τέλος, πριν περάσουμε στις αναλύσεις να σημειωθεί ότι όλα στατιστικά μεγέθη που υπολογίστηκαν για τους υποψηφίους (και του 2015 και του 2019), και τα οποία θα παρουσιαστούν παρακάτω, υπολογίστηκαν σταθμίζοντας με βάση την κατανομή των εκλογικών ποσοστών των κομμάτων στις εθνικές εκλογές (το 2015 με βάση τα ποσοστά του Ιανουαρίου), προκειμένου η πολιτική ελίτ του Twitter να είναι περισσότερο αντιπροσωπευτική του συνόλου της πολιτικής ελίτ σε εθνικό επίπεδο.

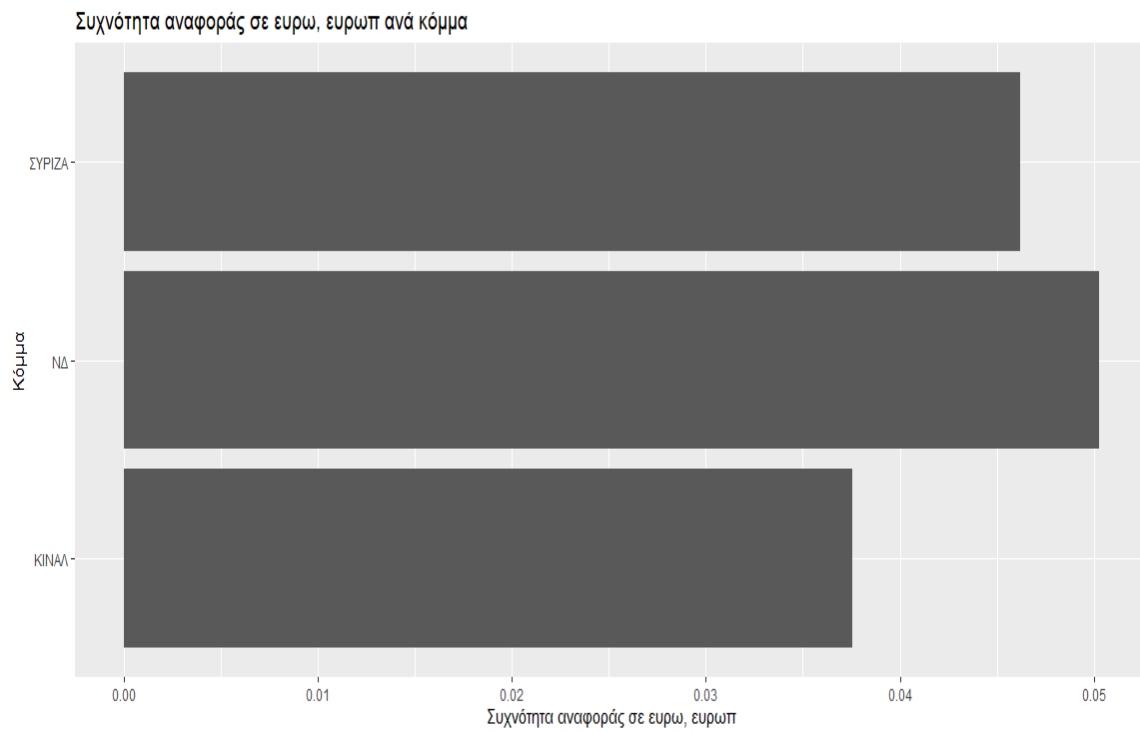
Τούτων δοθέντων, τα tweets που αναφέρουν τις ρίζες «λαο» και «λαϊκ» ως προς το σύνολο των tweets αποτελούν ένα ποσοστό 2.2 %. Αντίστοιχα, τα tweets που αναφέρουν τις ρίζες «ευρω», «ευρωπ» αποτελούν ένα ποσοστό 4.6% ως προς το σύνολο των tweets. Τέλος τα tweets που αναφέρουν τους εχθρούς του λαού ανέρχονται σε ένα ποσοστό 2.5%, ως προς το σύνολο.

Στα παρακάτω γραφήματα (1,2 και 3), παραθέτουμε τη συχνότητα αναφοράς των παραπάνω λέξεων ανά κόμμα.

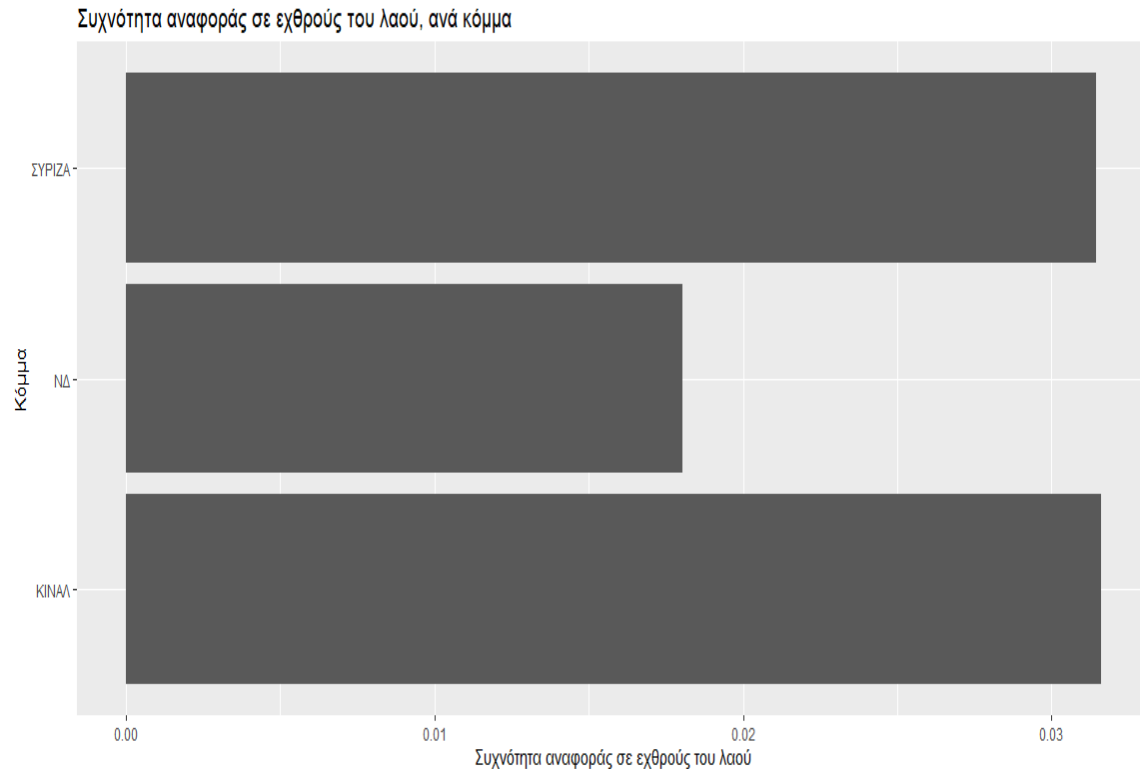
ΓΡΑΦΗΜΑ 1



ΓΡΑΦΗΜΑ 2



ΓΡΑΦΗΜΑ 3



Βλεπουμε ότι οι αναφορές στο λαό είναι πιο συχνές για το ΣΥΡΙΖΑ, οι αναφορές στην Ευρώπη συχνότερες για τη ΝΔ και οι αναφορές στους εχθρούς του λαού συχνότερες για τον ΣΥΡΙΖΑ και το ΚΙΝΑΛ. Αντίστοιχα, λιγότερες αναφορές στην Ευρώπη φαίνεται να έχει το ΚΙΝΑΛ, και στο λαό και στους εχθρούς του λαού η ΝΔ.

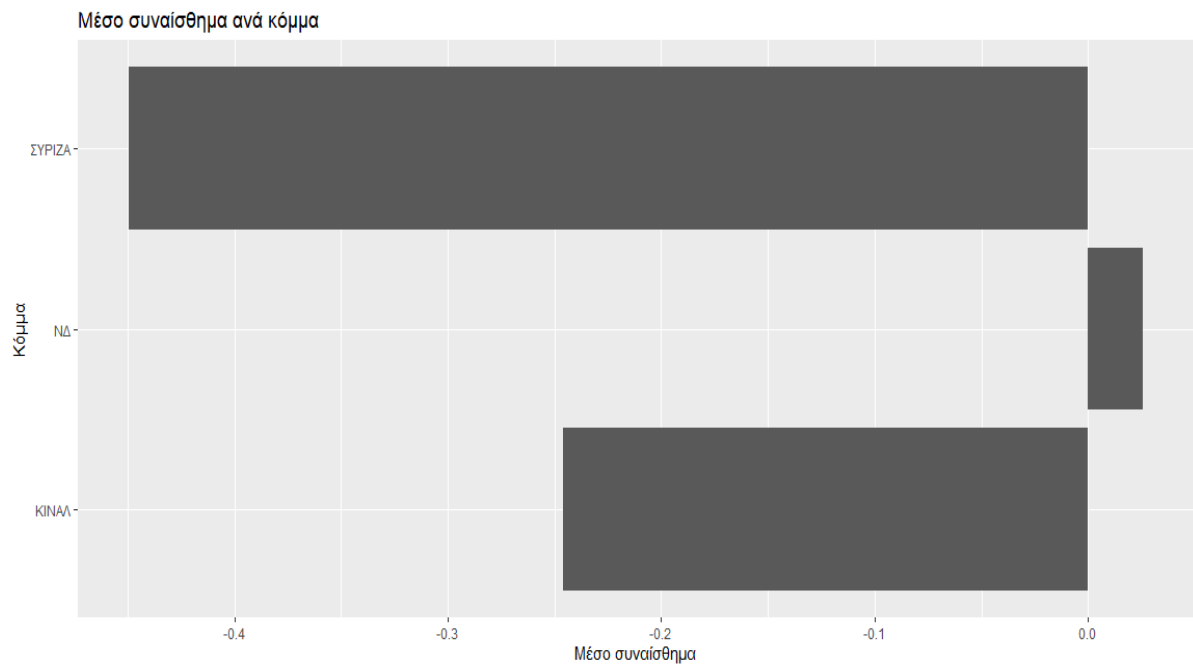
Περνώντας τώρα στην ανάλυση συναισθήματος υπολογίσαμε ότι το μέσο συναίσθημα των tweets είναι -0.22 για τους υποψηφίους του 2019. Στον πίνακα 2 βλέπουμε κατά πόσο αλλάζει το μέσο συναίσθημα των tweets όταν γίνεται αναφορά στο λαό στην Ευρώπη και στους εχθρούς του λαού αντίστοιχα.

ΠΙΝΑΚΑΣ 2: Αναφορές σε λαό, Ευρώπη, εχθρούς λαού και συναίσθημα.

	Λαός	Ευρώπη	Εχθροί Λαού
Μη αναφορά	-0.220	-0.199	-0.214
Αναφορά	-0.300	-0.593	-0.478

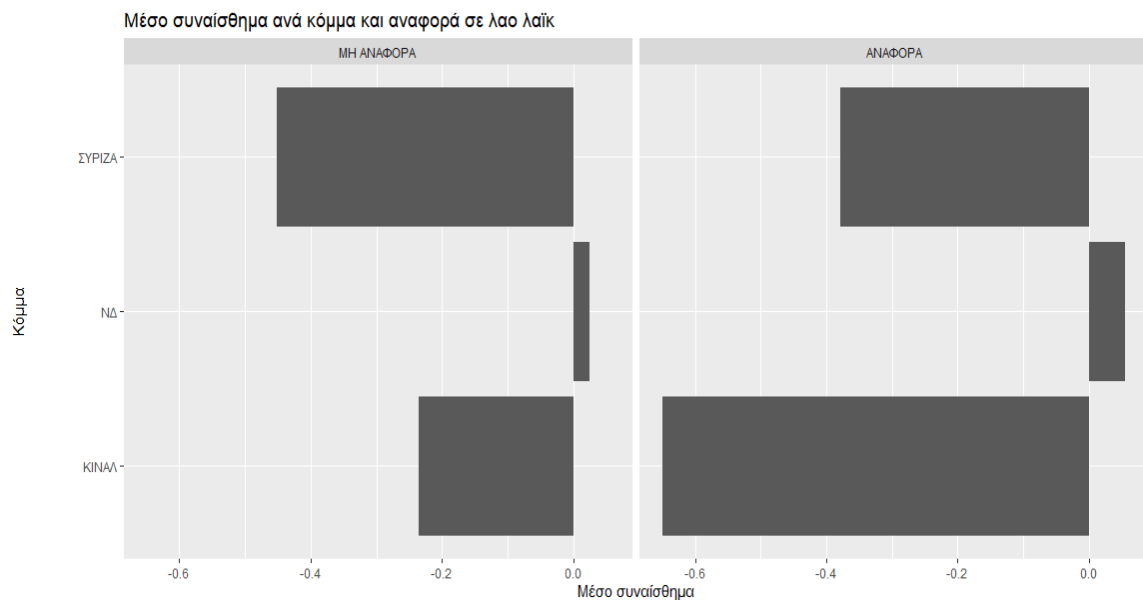
Παρακάτω στο Γράφημα 4 παραθέτουμε το μέσο συναίσθημα των tweets ομαδοποιημένο ανά κόμμα.

ΓΡΑΦΗΜΑ 4

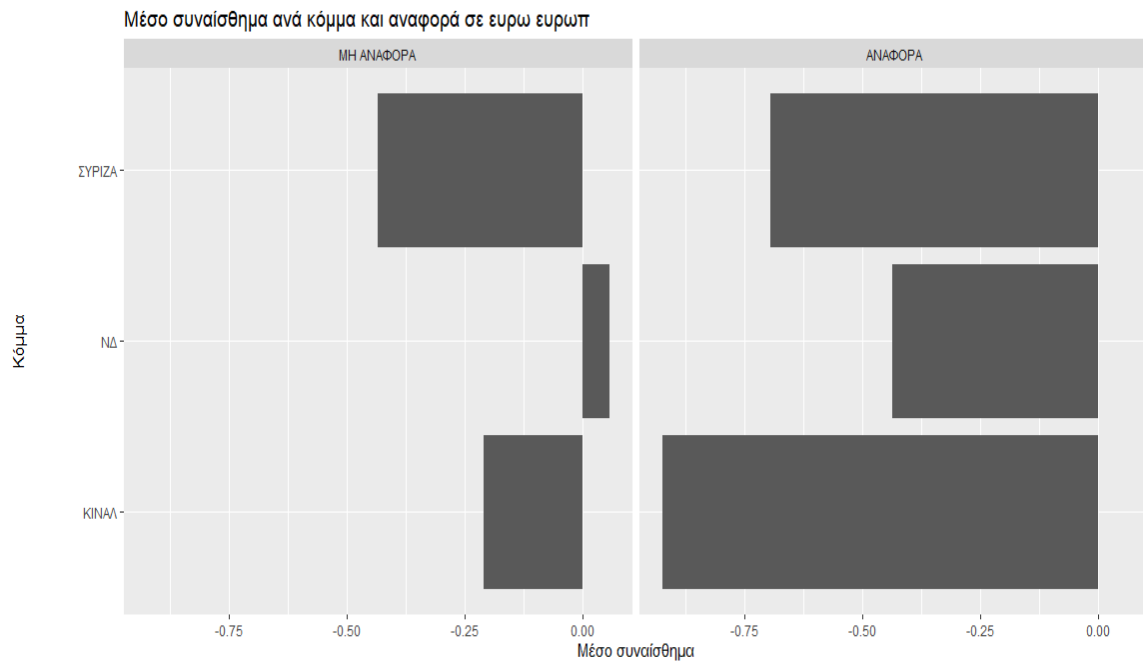


Βλέπουμε ότι η ΝΔ είχε κατά μέσο όρο θετικότερα συναισθήματα μεταξύ των κομμάτων και ο ΣΥΡΙΖΑ τα πιο αρνητικά με το ΚΙΝΑΛ να είναι κάπου ενδιάμεσα. Στη συνέχεια, στα γραφήματα 5-7 παραθέτουμε το μέσο συναίσθημα των tweets ανά κόμμα για επιμέρους αναφορές σε λαό, Ευρώπη και εχθρούς λαού.

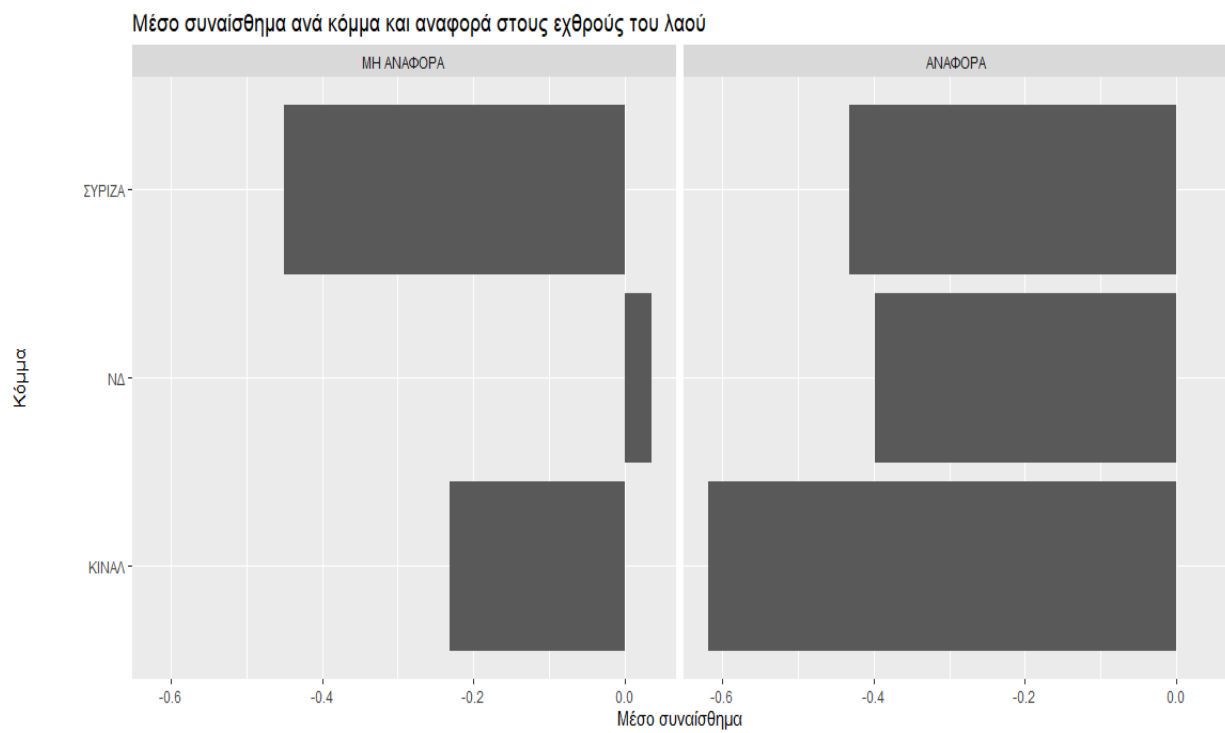
ΓΡΑΦΗΜΑ 5



ΓΡΑΦΗΜΑ 6



ΓΡΑΦΗΜΑ 7



Όπως είπαμε και στο κεφάλαιο της μεθοδολογίας αναμένουμε περισσότερο αρνητικά tweets να συνεπάγονται και περισσότερο λαϊκισμό και ο ευρωσκεπτικισμό. Συνεπώς αυτό που θα περιμέναμε από τα παραπάνω τρία γραφήματα θα ήταν το πιο αρνητικό κόμμα να ήταν ο ΣΥΡΙΖΑ, ως το πιο λαϊκιστικό/ευρωσκεπτικιστικό κόμμα, μετά το ΚΙΝΑΛ και μετά η ΝΔ ως το λιγότερο λαϊκιστικό/ευρωσκεπτικιστικό κόμμα. Η εικόνα είναι όπως την αναμένουμε για το ΣΥΡΙΖΑ και τη ΝΔ σε όλα τα γραφήματα, αλλά όχι για το ΚΙΝΑΛ. Κατ' αρχάς όταν αναφέρεται στα tweets ο λαός (Γράφημα 5) ο αρνητισμός του ΚΙΝΑΛ φαίνεται να ανεβαίνει δυσανάλογα περισσότερο από τα άλλα κόμματα. Έτσι, στο γράφημα 5 παρατηρούμε ότι όταν γίνονται αναφορές στο λαό αυξάνεται ο αρνητισμός κυρίως του ΚΙΝΑΛ. Ο αρνητισμός αυτός φαίνεται να πηγάζει από τα tweets της προέδρου του κόμματος, Φώφης Γεννηματά, η οποία έχει μεγαλύτερη συχνότητα αρνητικών tweets που αναφέρονται στο λαό, σε σύγκριση με τους υπόλοιπους υποψηφίους του κόμματος. Ο αρνητισμός των tweet της Γεννηματά θεματολογικά φαίνεται να αφορά μία αντίθεση προς το λαϊκισμό του ΣΥΡΙΖΑ, αλλά και μια κριτική προς τον υπάρχοντα δικομματισμό ΣΥΡΙΖΑ-ΝΔ., και όχι τόσο πολύ λαϊκιστικό λόγο.

Παρόμοια κατανομή έχουμε και στο γράφημα 6 όπου όταν αναφέρεται η Ευρώπη στα tweets ο αρνητισμός του ΚΙΝΑΛ φαίνεται μεγαλώνει περισσότερο από τα άλλα κόμματα. Και σε αυτή την περίπτωση ο αρνητισμός δεν υποδεικνύει ενδεχομένως ευρωσκεπτικισμό, αλλά προέρχεται από τα περισσότερα αρνητικά tweets της Φώφης Γεννηματά και επιπλέον της Ζέφης Δημαδάμα. Θεματολογικά τα συγκεκριμένα tweets αναφέρονται στις διπλωματικές εντάσεις που είχε η χώρα μας με την Τουρκία μετά την εισβολή της δεύτερης στην περιοχή του Έβρου. Ο αρνητισμός στα συγκεκριμένα tweets προέρχεται από την έντονη αντίθεση προς την επιθετική στάση της Τουρκίας και όχι τόσο από αντίθεση στην Ευρωπαϊκή Ένωση. Η Ευρωπαϊκή Ένωση αναφέρεται σταθερά στα συγκεκριμένα tweets, επειδή επισημαίνεται η ανάγκη να παρέμβει αυτή προκειμένου να λυθεί το ζήτημα σε ευρωπαϊκό επίπεδο.

Τέλος, παρόμοια κατανομή με τις δύο προηγούμενες φαίνεται να έχουμε και στο γράφημα 7 όπου το ΚΙΝΑΛ έρχεται πάλι πρώτο σε αρνητισμό όταν αναφέρονται οι εχθροί του λαού. Ο αρνητισμός στα συγκεκριμένα tweets φαίνεται να προέρχεται από τον υποψήφιο Μιχάλη Κατρίνη. Τα αρνητικά tweets του συγκεκριμένου υποψηφίου φαίνεται πράγματι να απηχούν λαϊκισμό. Ο βασικότερος εχθρός του λαού ο οποίος αναφέρεται φαίνεται να είναι οι τράπεζες και καταλογίζεται ευθύνη στην κυβέρνηση ότι τις προστατεύει. Επίσης, υπάρχουν και ορισμένα tweets του ίδιου όπου κάνει λόγο για εξυπηρέτηση οικονομικών ελίτ από την κυβέρνηση. Τέλος, υπάρχουν και πολλά που αναφέρουν τη λέξη «σύστημα». Ωστόσο η εν λόγω λέξη δεν έχει τη σημασία του συστήματος ως εχθρού, αλλά αναφέρεται στο σύστημα υγείας εν μέσω πανδημίας και ο αρνητισμός φαίνεται ότι προέρχεται από μια κριτική προς την κυβέρνηση η οποία κατηγορείται ότι δεν προνοεί για το ΕΣΥ.

Στη συνέχεια θα προσπαθήσουμε να εξετάσουμε τον αντίκτυπο που έχουν τα λαϊκιστικά και ευρωσκεπτικιστικά tweets των υποψηφίων στο Twitter, μέσω των favorites και των retweets. Κατ' αρχάς, όπως βλέπουμε και στους πίνακες που ακολουθούν (Πίνακας 3 και

4) τα retweets και τα favorites φαίνεται να αυξάνονται όταν αναφέρεται ο λαός, η Ευρώπη και οι εχθροί του λαού.

ΠΙΝΑΚΑΣ 3: Μέσος όρος retweets ανά αναφορά σε λαό, Ευρώπη και εχθρούς του λαού

	Λαός	Ευρώπη	Εχθροί Λαού
Μη αναφορά	14.9	14.9	15.0
Αναφορά	20.6	17.8	18.4

ΠΙΝΑΚΑΣ 4: Μέσος όρος favorites ανά αναφορά σε λαό, Ευρώπη και εχθρούς του λαού

	Λαός	Ευρώπη	Εχθροί Λαού
Μη αναφορά	69.9	70.0	70.3
Αναφορά	98.4	80.8	79.7

Απο τη στιγμή που θεωρούμε το βαθμό αρνητισμού πιθανή ένδειξη για το λαϊκισμό/ευρωσκεπτικισμό, όπως έχουμε πει, παρακάτω στον πίνακα 5 εξετάζουμε το μέσο όρο retweets και favorites για κάθε συναισθηματική κατηγορία έτσι ώστε να δούμε κατά πόσο είναι μεγαλύτερη η απήχηση των αρνητικών tweets.

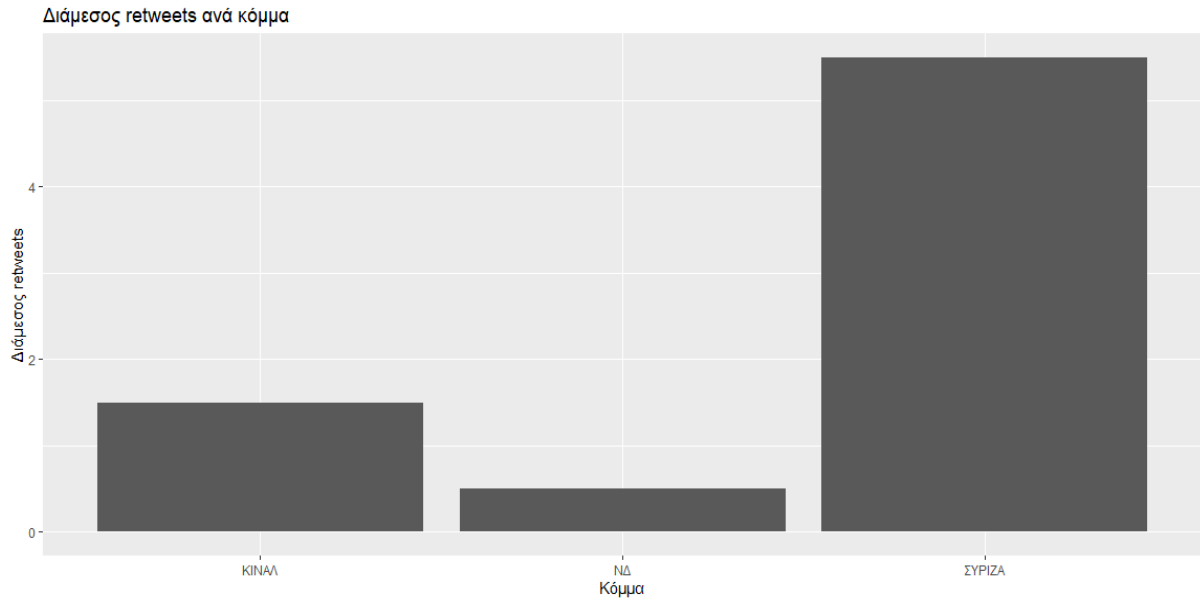
ΠΙΝΑΚΑΣ 5: Μέσος όρος retweets και favorites ανά συναισθηματική κατηγορία tweets

Συναισθηματική Κατηγορία Tweets	Retweets	Favorites
Αρνητικά	19.4	89.1
Ουδέτερα	15.1	71.5
Θετικά	16.2	86.8

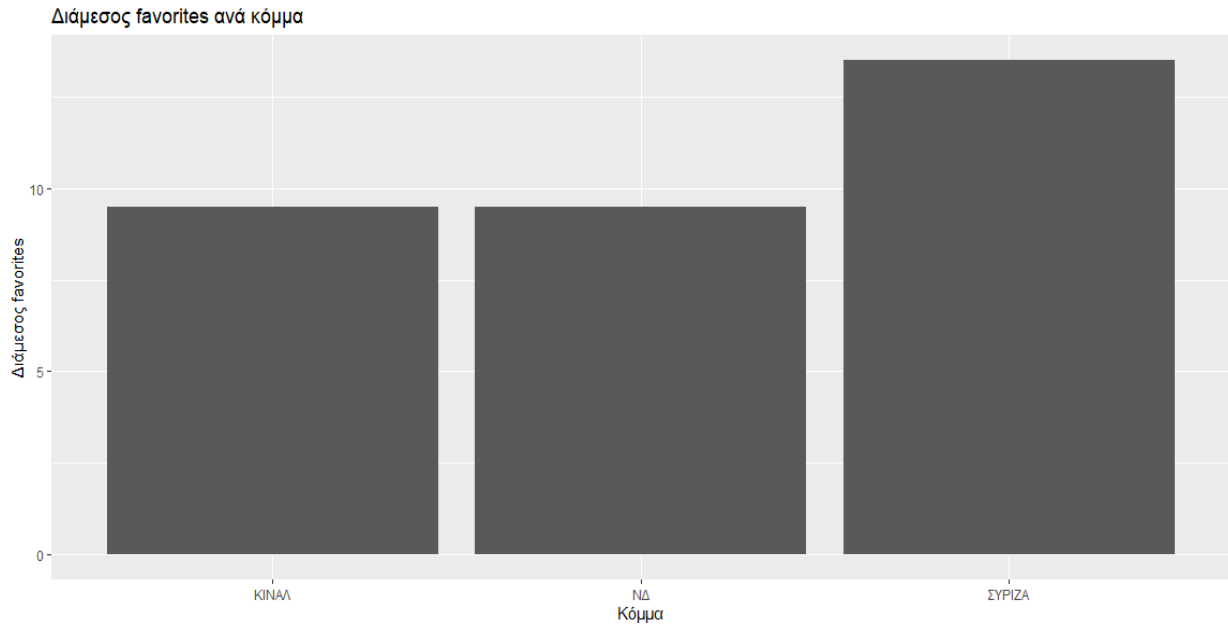
Με βάση τους πίνακες 3,4 και 5 μπορούμε να συμπεράνουμε ότι retweets και τα favorites αυξάνονται όταν αναφέρονται οι λέξεις λαός, Ευρώπη και οι εχθροί του λαού, ενώ ταυτόχρονα τα αρνητικά συναισθηματικά tweets έχουν περισσότερα retweets και favorites από τα θετικά και τα ουδέτερα. Συνεπώς, υπάρχει απήχηση του λαϊκιστικού και ευρωσκεπτικιστικού λόγου των υποψηφίων του 2019.

Στα επόμενα δύο γραφήματα (Γράφημα 8 και 9) βλέπουμε τη διάμεσο των retweets και των favorites ανά κόμμα. Ο λόγος που επιλέξαμε τη διάμεσο αντί για το μέσο όρο είναι γιατί υπάρχουν ορισμένες ακραίες τιμές retweets και favorites, οι οποίες προέρχονται κυρίως από το λογαριασμό του πρωθυπουργού Κυριάκου Μητσοτάκη και από το λογαριασμό του προέδρου του ΜΕΡΑ 25, Γιάννη Βαρουφάκη.

ΓΡΑΦΗΜΑ 8

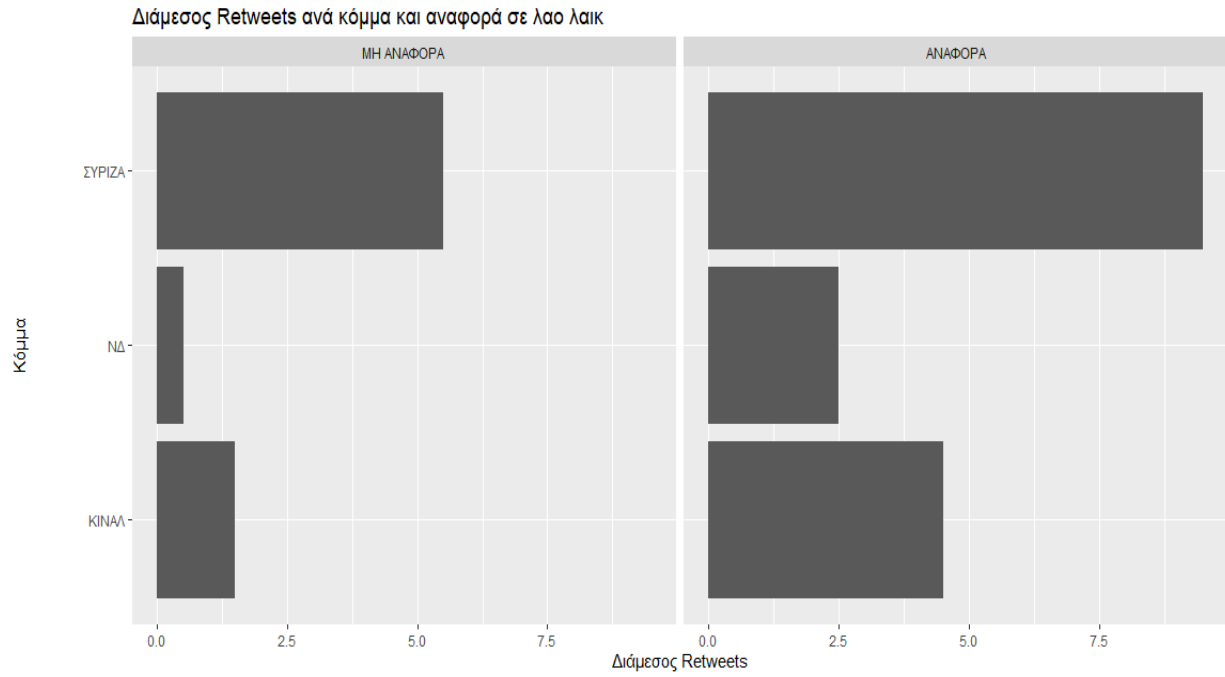


ΓΡΑΦΗΜΑ 9



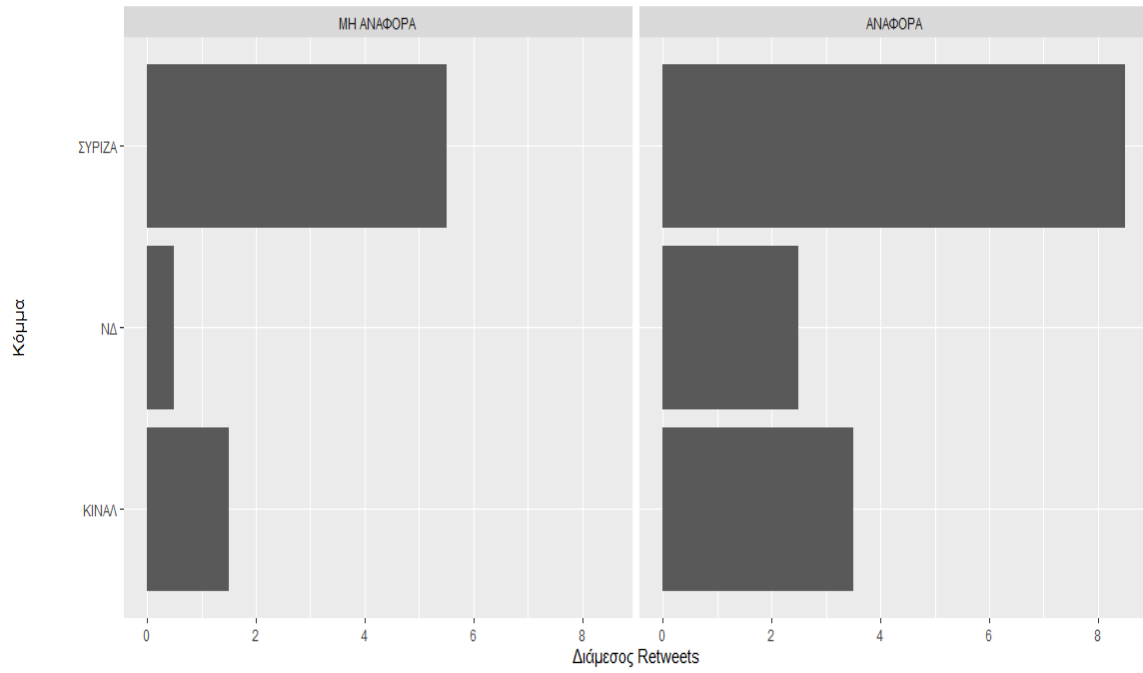
Στα επόμενα 6 διαγράμματα που ακολουθούν (Γράφημα 10-15), αποτυπώνουμε τα retweets και τα favorites ανά κόμμα και αναφορά στο λαό, στην Ευρώπη και στους εχθρούς του λαού.

ΓΡΑΦΗΜΑ 10



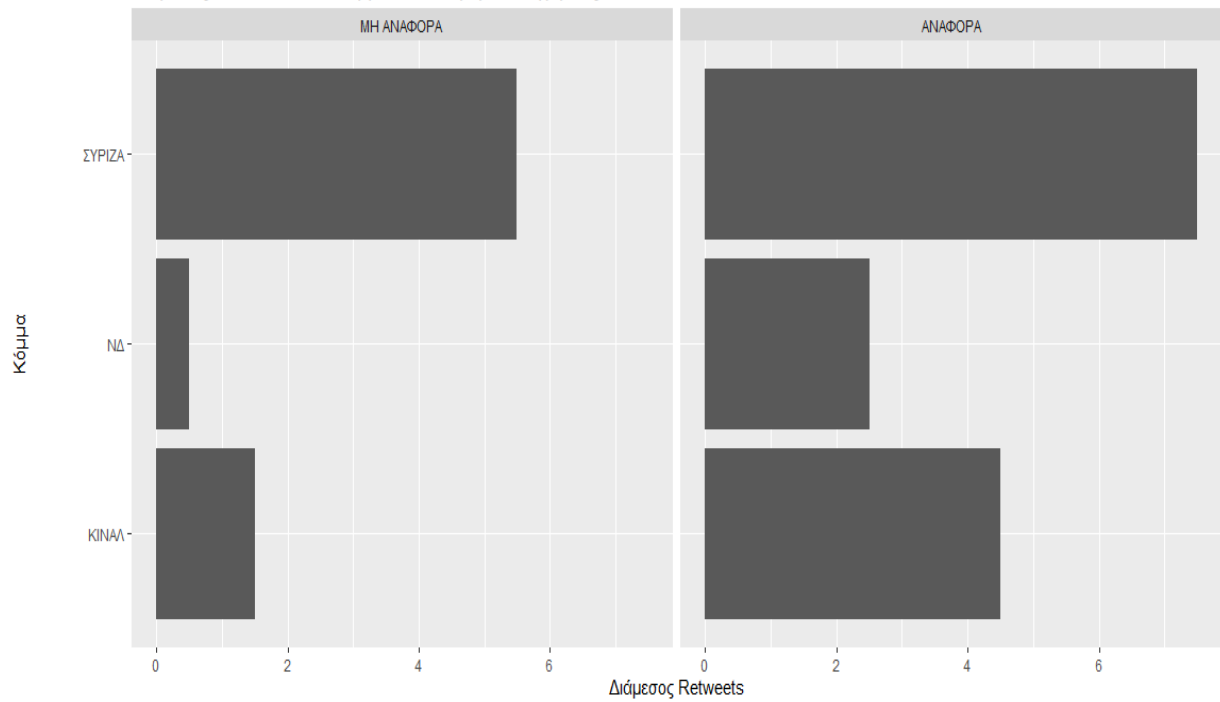
ΓΡΑΦΗΜΑ 11

Διάμεσος Retweets ανά κόμμα και αναφορά σε ευρω ευρωπ



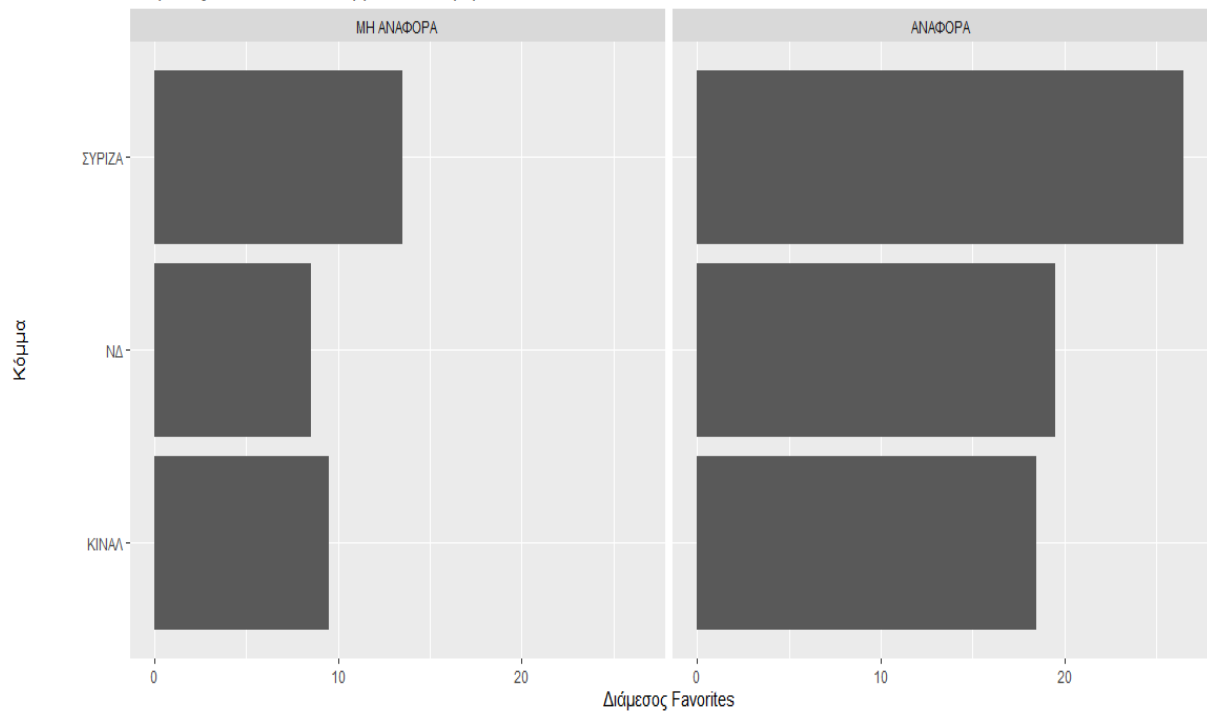
ΓΡΑΦΗΜΑ 12

Διάμεσος Retweets ανά κόμμα και αναφορά σε εχθρούς του λαού



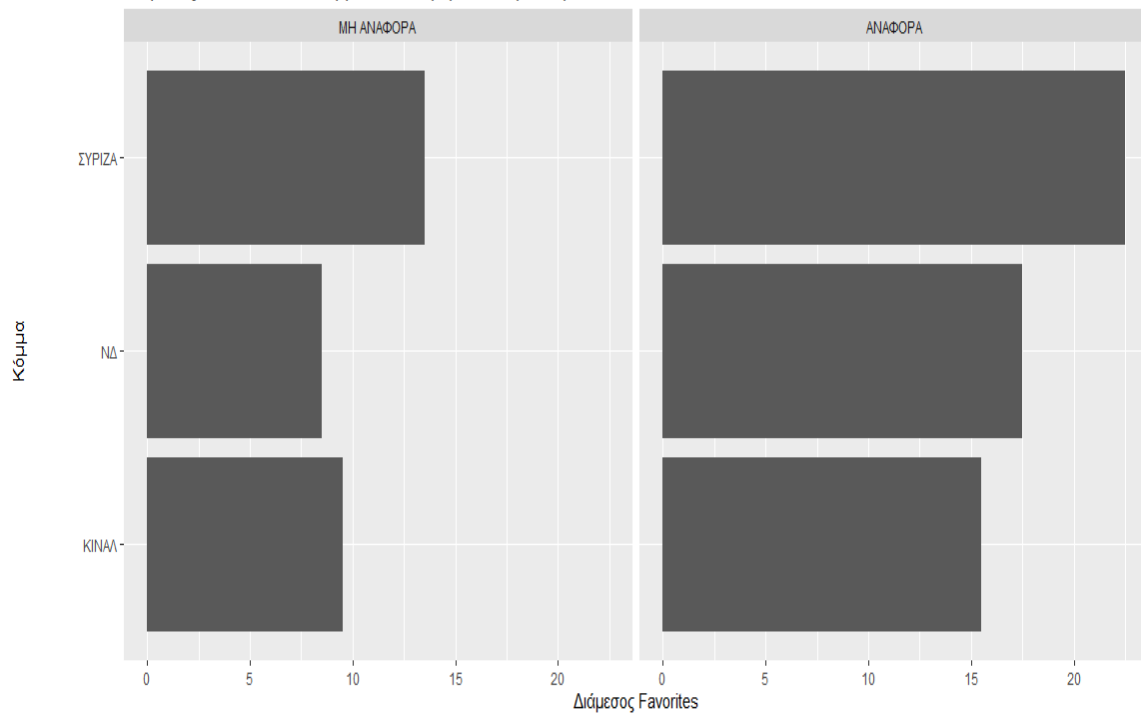
ΓΡΑΦΗΜΑ 13

Διάμεσος Favorites ανά κόμμα και αναφορά σε λαο λαικ



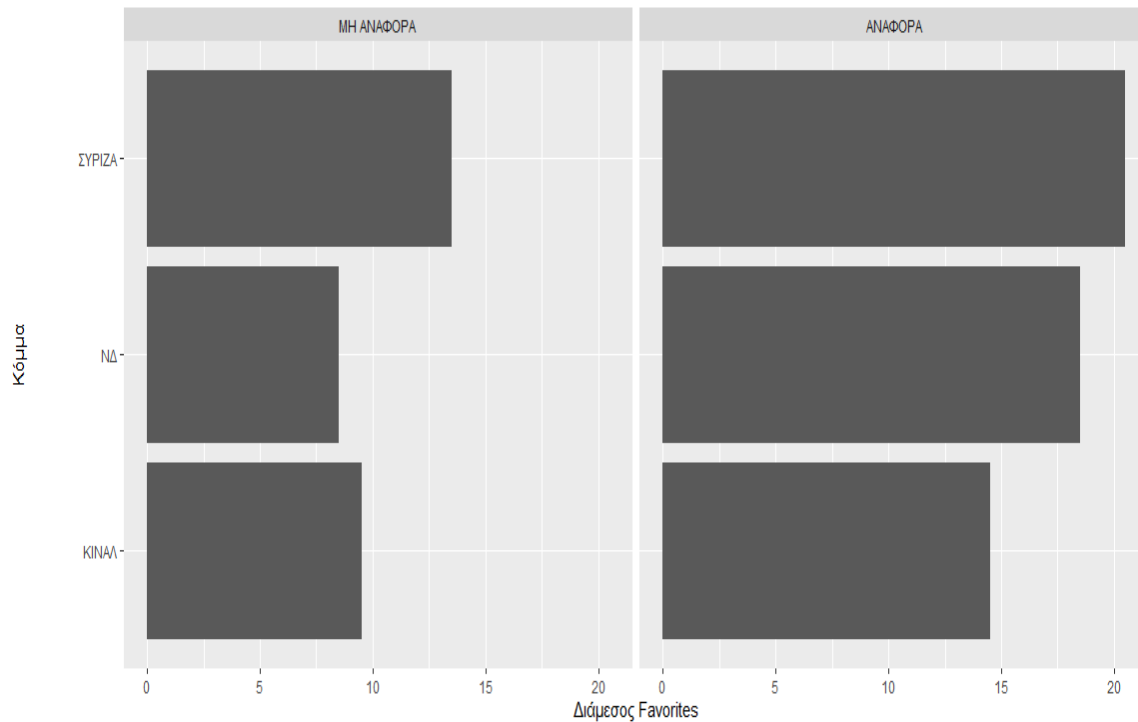
ΓΡΑΦΗΜΑ 14

Διάμεσος Favorites ανά κόμμα και αναφορά σε ευρω ευρωπ



ΓΡΑΦΗΜΑ 15

Διάμεσος Favorites ανά κόμμα και αναφορά σε εχθρούς του λαού



Για τα retweets (Γραφήματα 10-12) η εικόνα που βλέπουμε είναι η αναμενόμενη. Όταν υπάρχει αναφορά σε λαό, Ευρώπη και εχθρούς του λαού η διάμεσος των retweets ανεβαίνει περισσότερο για τον ΣΥΡΙΖΑ, λιγότερο για το ΚΙΝΑΛ και ακόμα λιγότερο για τη ΝΔ. Η εικόνα είναι λίγο διαφορετική στα favorites όπου η ΝΔ έρχεται δεύτερη σε σειρά και στις τρεις περιπτώσεις. Αυτό συμβαίνει γιατί, στην περίπτωση των αναφορών στο λαό οι υποψήφιοι Κωνσταντίνος Μπογδάνος και Άδωνις Γεωργιάδης της ΝΔ έχουν μεγαλύτερο αριθμό favorites από τους υπόλοιπους υποψηφίους. Όσον αφορά τα tweets με αναφορά στο λαό του Γεωργιάδη αυτά απηχούν κάποιου είδους ιδιότυπο εθνικολαϊκισμό, κυρίως αντιδιαστέλοντας τον λαό απέναντι στο ΣΥΡΙΖΑ και την πρώην κυβέρνηση ΣΥΡΙΖΑ οι οποίοι παρουσιάζονται να θέλουν το «κακό» του λαού.

Στην περίπτωση των αναφορών στην Ευρώπη και στους εχθρούς του λαού η ΝΔ έρχεται δεύτερη στα favorites, λόγω των tweets με πολλά favorites του Κυριάκου Μητσοτάκη, που αναφέρουν τις κατηγορίες αυτές, Βέβαια, τα εν λόγω tweets δεν είναι λαϊκιστικά ή ευρωσκεπτικιστικά. Οι αναφορές στην Ευρώπη γίνονται κυρίως στο πλαίσιο της επιθετικής στάσης της Τουρκίας και οι αναφορές στους εχθρούς του λαού γίνονται κυρίως μέσω της λέξης «σύστημα», όπου εν προκειμένω αναφέρεται στο σύστημα υγείας ενόψει της πανδημικής κρίσης του κορωνοϊού.

Τέλος, επιπρόσθετα, να σημειώσουμε ότι οι πέντε υποψήφιοι με τα περισσότερα retweets κατά μέσο όρο είναι ο Αντώνης Σαμαράς, ο λογαριασμός του πρωθυπουργού, ο Ευκλείδης Τσακαλώτος, ο ατομικός λογαριασμός του Κυριάκου Μητσοτάκη και ο Κωνσταντίνος Κυρανάκης. Οι πέντε υποψήφιοι με τα περισσότερα favorites είναι οι ίδιοι

με τη διαφορά ότι από την πεντάδα βγαίνει ο Τσακαλώτος και μπαίνει ο Αλέξης Τσίπρας.

ΣΥΓΚΡΙΣΗ ΥΠΟΨΗΦΙΩΝ ΒΟΥΛΕΥΤΩΝ ΚΑΙ ΧΡΗΣΤΩΝ ΤΟΥ TWITTER (2019).

Όπως είπαμε σε προηγούμενη ενότητα, τα tweets που συλλέγησαν από τους χρήστες του Twitter είναι χωρισμένα τρεις κατηγορίες: σε αυτά που συλλέγησαν με βάση τις αναφορές σε λαό, στην Ευρώπη και στις εκλογές. Η σύγκριση όσον αφορά τις αναφορές στο λαό, την Ευρώπη και τους εχθρούς του λαού θα γίνει μεταξύ των tweets των υποψηφίων και των tweets των χρηστών που αναφέρονται σε εκλογές. Άλλες μεμονωμένες αναλύσεις θα γίνουν μεταξύ των χρηστών που αναφέρονται σε λαό και των υποψηφίων και μεταξύ των χρηστών που αναφέρονται στην Ευρώπη και των υποψηφίων. Επίσης, να σημειωθεί ότι δεν έχουμε συλλέξει tweets χρηστών με βάση αναφορές στους εχθρούς του λαού.

Στον πίνακα που ακολουθεί (Πίνακας 6) βλέπουμε τη σχετική συχνότητα των αναφορών σε λαό, Ευρώπη και εχθρούς του λαού από τους υποψήφιους και από τους χρήστες⁴.

ΠΙΝΑΚΑΣ 6: Σχετική συχνότητα αναφορών σε λαό, Ευρώπη και εχθρούς του λαού. Υποψήφιοι και χρήστες (2019).

	Υποψήφιοι	Χρήστες
Αναφορά στο λαό	2.2%	2.9%
Αναφορά στην Ευρώπη	4.6%	2.1%
Αναφορά στους εχθρούς του λαού	2.5%	2.3%

Βλέπουμε ότι οι χρήστες αναφέρονται κάπως περισσότερο από τους υποψήφιους στο λαό, ενώ οι υποψήφιοι αναφέρονται κάπως περισσότερο από τους χρήστες στους εχθρούς του λαού και αρκετά περισσότερο στην Ευρώπη.

Επίσης, όσον αφορά το συνολικό μέσο συναίσθημα, αυτό είναι -0.53 για τους χρήστες και -0.22 για τους υποψήφιους, όπως είδαμε σε προηγούμενη ενότητα.

Στον παρακάτω πίνακα (Πίνακας 7) βλέπουμε την αλλαγή στο συναίσθημα όταν αναφέρεται ο λαός, η Ευρώπη και οι εχθροί του λαού, για τους υποψήφιους και τους χρήστες⁵

⁴ Τους χρήστες που αναφέρουν «εκλογ».

⁵ Τους χρήστες που αναφέρουν «εκλογ».

ΠΙΝΑΚΑΣ 7: Μέσο συναίσθημα ανά αναφορά σε λαό, Ευρώπη και εχθρούς του λαού από τους υποψήφιους και τους χρήστες (2019).

	ΥΠΟΨΗΦΙΟΙ			ΧΡΗΣΤΕΣ		
	Λαός	Ευρώπη	Εχθροί Λαού	Λαός	Ευρώπη	Εχθροί Λαού
Μη αναφορά	-0.220	-0.199	-0.214	-0.523	-0.519	-0.520
Αναφορά	-0.300	-0.593	-0.478	-0.568	-0.728	-0.721

ΠΙΝΑΚΑΣ 8: Ποσοστό (%) tweets ανά συναισθηματική κατηγορία για τους υποψήφιους και τους χρήστες (2019).

	ΥΠΟΨΗΦΙΟΙ			ΧΡΗΣΤΕΣ		
	Λαός	Ευρώπη	Εχθροί Λαού	Λαός ⁶	Ευρώπη ⁷	Εχθροί Λαού ⁸
Αρνητικό Συναίσθημα	51.4	55.4	55.9	55.2	76.7	61.5
Ουδέτερο Συναίσθημα	16.7	19.3	17.6	15.8	10.1	14.2
Θετικό συναίσθημα	31.9	25.2	26.6	29.0	13.2	24.3

⁶ Χρήστες με αναφορά σε λαο

⁷ Χρήστες με αναφορά σε Ευρώπη

⁸ Χρήστες με αναφορά σε εκλογές. Ο λόγος που επιλέχθηκε η συγκεκριμένη κατηγορία εδώ είναι επειδή δεν συλλέξαμε δεδομένα με λέξεις κλειδιά εχθρούς του λαού. Συνεπώς, για τη συγκεκριμένη στήλη υπολογίζουμε τα ποσοστά των συναισθημάτων με βάση το υποσύνολο των χρηστών που αναφέρει εχθρούς του λαού από το σύνολο των χρηστών που αναφέρουν «εκλογ».

Με βάση τον πίνακα 7 παρατηρούμε ότι και στην περίπτωση των υποψηφίων και στην περίπτωση των χρηστών το αρνητικό συναίσθημα αυξάνεται περισσότερο όταν γίνεται αναφορά στην Ευρώπη και στους εχθρούς του λαού συγκριτικά με το όταν γίνεται αναφορά στο λαό. Η μόνη διαφορά είναι ότι οι τιμές των συναισθημάτων είναι λίγο πιο αρνητικές για τους χρήστες, αλλά αυτό οφείλεται στο γεγονός ότι οι χρήστες έχουν γενικότερα περισσότερο αρνητικά tweets από τους υποψήφιους, όπως είδαμε και προηγουμένως.

Στον πίνακα 8 παρατηρούμε ότι το ποσοστό των tweets που έχουν αρνητικό συναίσθημα είναι μεγαλύτερο για όλες τις αναφορές (λαό, Ευρώπη, εχθρούς λαού) και για τους υποψήφιους και για τους χρήστες. Επίσης, βλέπουμε μία παρόμοια εικόνα με τον πίνακα 7, ότι δηλαδή το ποσοστό των tweets με αρνητικό συναίσθημα είναι μεγαλύτερο για τις αναφορές στην Ευρώπη και στους εχθρούς του λαού σε σχέση με τις αναφορές στο λαό, και για τους υποψηφίους και για τους χρήστες. Και σε αυτή την περίπτωση, όπως και πριν, το ποσοστό των αρνητικών tweets των χρηστών είναι μεγαλύτερο, αλλά αυτό οφείλεται στο γεγονός ότι οι χρήστες έχουν γενικότερα περισσότερο αρνητικά tweets από τους υποψηφίους.

Συνοπτικά, θα μπορούσαμε να πούμε ότι οι υποψήφιοι και οι χρήστες συγκλίνουν όσον αφορά το συναίσθημα των tweets. Εκφράζουν αρνητισμό όταν αναφέρονται στο λαό, στην Ευρώπη και στους εχθρούς του λαού, και δη παραπάνω αρνητισμό όταν αναφέρονται στην Ευρώπη και στους εχθρούς του λαού.

Στη συνέχεια, επιχειρήσαμε να συγκρίνουμε τις κατανομές των συναισθημάτων των tweets των χρηστών και των υποψηφίων, όταν αναφέρεται λαός, Ευρώπη και εχθροί του λαού αντίστοιχα. Για το σκοπό αυτό χρησιμοποιήσαμε το επαγωγικό στατιστικό κριτήριο Kolmogorov-Smirnov. Το συγκεκριμένο κριτήριο χρησιμοποιείται στην περίπτωσή μας για τον έλεγχο της μηδενικής υπόθεσης ότι οι δύο κατανομές συναισθημάτων (υποψήφιοι και χρήστες) προέρχονται από τον ίδιο πληθυσμό. Για το σκοπό αυτό συγκρίνονται μεταξύ τους οι δύο εμπειρικές αθροιστικές κατανομές των κατανομών προς εξέταση και η τιμή του κριτηρίου είναι το supremum της διαφοράς τους. Το συγκεκριμένο κριτήριο έχει χρησιμοποιηθεί για τη μελέτη της σύγκλισης των θέσεων υποψηφίων και ψηφοφόρων σε μελέτη των Ανδρεάδη και Καρτσουνίδου (Kartsounidou & Andreadis, 2020). Παρακάτω στον πίνακα 9 βλέπουμε τις τιμές του κριτηρίου και το επίπεδο σημαντικότητας τους για κάθε μία από τις συγκρίσεις προς εξέταση.

ΠΙΝΑΚΑΣ 9: Τιμές κριτηρίου Kolmogorov-Smirnov για τη σύγκριση των κατανομών των συναισθημάτων των tweets υποψηφίων και χρηστών (2019), ανά αναφορές σε λαό, Ευρώπη και εχθρούς του λαού.

Αναφορά σε λαό ⁹	Αναφορά στην Ευρώπη ¹⁰	Αναφορά στους εχθρούς του λαού ¹¹
D = 0.041102*	D = 0.25155***	D = 0.067025

*** p<0.001 ** p<0.01 *p<0.05

Αυτό που βλέπουμε είναι ότι οι κατανομές των συναισθημάτων μεταξύ υποψηφίων και χρηστών είναι περισσότερο κοινές στην περίπτωση των αναφορών στο λαό και στους εχθρούς του λαού και λιγότερο κοινές στην περίπτωση των αναφορών στην Ευρώπη¹².

Από τη στιγμή που υποψήφιοι και χρήστες εκφράζουν κοινά συναισθήματα όταν αναφέρονται στο λαό και στους εχθρούς του λαού αυτό σημαίνει ότι είναι πολύ πιθανό να αντιλαμβάνονται το λαϊκισμό από κοινού (όπου ο λαϊκισμός είναι υπαρκτός). Από την άλλη, όταν αναφέρονται στην Ευρώπη φαίνεται τα συναισθήματα να είναι κάπως περισσότερο διαφορετικά, ενδεχομένως γιατί οι αναφορές των υποψηφίων στην Ευρώπη να είναι πιο «εκλεπτυσμένες» και με περισσότερα τεχνικά στοιχεία σε σχέση με τις αναφορές των χρηστών που ενδέχεται να είναι πιο απλοϊκές.

ΣΥΓΚΡΙΣΗ ΥΠΟΨΗΦΙΩΝ 2015 ΜΕ ΤΟΥΣ ΥΠΟΨΗΦΙΟΥΣ ΤΟΥ 2019

Παρακάτω προχωράμε στην ανάλυση των υποψηφίων του 2015 και στη σύγκρισή τους με τους υποψηφίους του 2019.

Τα tweets που συλλέχθηκαν από τους λογαριασμούς των υποψηφίων βουλευτών του 2015 συνδυάστηκαν με δεδομένα που δημιουργήσαμε προκειμένου τα πρώτα να εμπλουτιστούν με καινούργιες μεταβλητές που αφορούν ιδιότητες των ελλήνων βουλευτών όπως κόμμα, κ.λπ. οι οποίες δεν ήταν διαθέσιμες στα δεδομένα που συλλέχθηκαν μέσω του Twitter API. Παρακάτω παρουσιάζεται η συχνότητα των tweets ανά κόμμα.

⁹ Σύγκριση χρηστών με αναφορά στο λαό και υποψηφίων

¹⁰ Σύγκριση χρηστών με αναφορά στην Ευρώπη και υποψηφίων

¹¹ Σύγκριση χρηστών με αναφορά σε εκλογές που αναφέρουν εχθρούς του λαού και υποψηφίων, καθώς δε συλλέξαμε, όπως προειπώθηκε, tweets με αναφορές στους εχθρούς του λαού.

¹² Η τιμή όσον αφορά τους εχθρούς του λαού ενδέχεται να μην είναι στατιστικά σημαντική επειδή το ποσοστό των αναφορών στους εχθρούς του λαού μεταξύ των χρηστών που αναφέρουν εκλογές είναι σχετικά μικρό, άρα και ο αριθμός των παρατηρήσεων (tweets) σχετικά μικρός. (βλ. Υποσημείωση 11)

Κόμμα	Συχνότητα
ΠΑΣΟΚ	78984
ΝΔ	135127
ΣΥΡΙΖΑ	86661

Όπως με τους υποψήφιους του 2019, έτσι και εδώ χωρίσαμε τα tweets σε αυτά που αποτελούν retweets και σε αυτά που είναι πρωτότυπα. Επίσης, όλα τα στατιστικά μεγέθη που θα παρουσιαστούν παρακάτω υπολογίστηκαν και αυτά σταθμίζοντας με βάση την κατανομή των εκλογικών ποσοστών των κομμάτων στις εθνικές εκλογές του Ιανουαρίου του 2015, προκειμένου η πολιτική ελίτ του Twitter να είναι περισσότερο ανιπροσωπευτική του συνόλου της πολιτικής ελίτ σε εθνικό επίπεδο.

Τούτων δοθέντων, προχωράμε τώρα στις συγκρίσεις στον πίνακα 10 των σχετικών συχνοτήτων των αναφορών σε λαό., Ευρώπη και εχθρούς του λαού, μεταξύ των υποψηφίων του 2019 και του 2015

ΠΙΝΑΚΑΣ 10: Σχετικές συχνότητες (%) αναφορών σε λαό, Ευρώπη και εχθρούς του λαού. Υποψήφιοι 2019 και 2015.

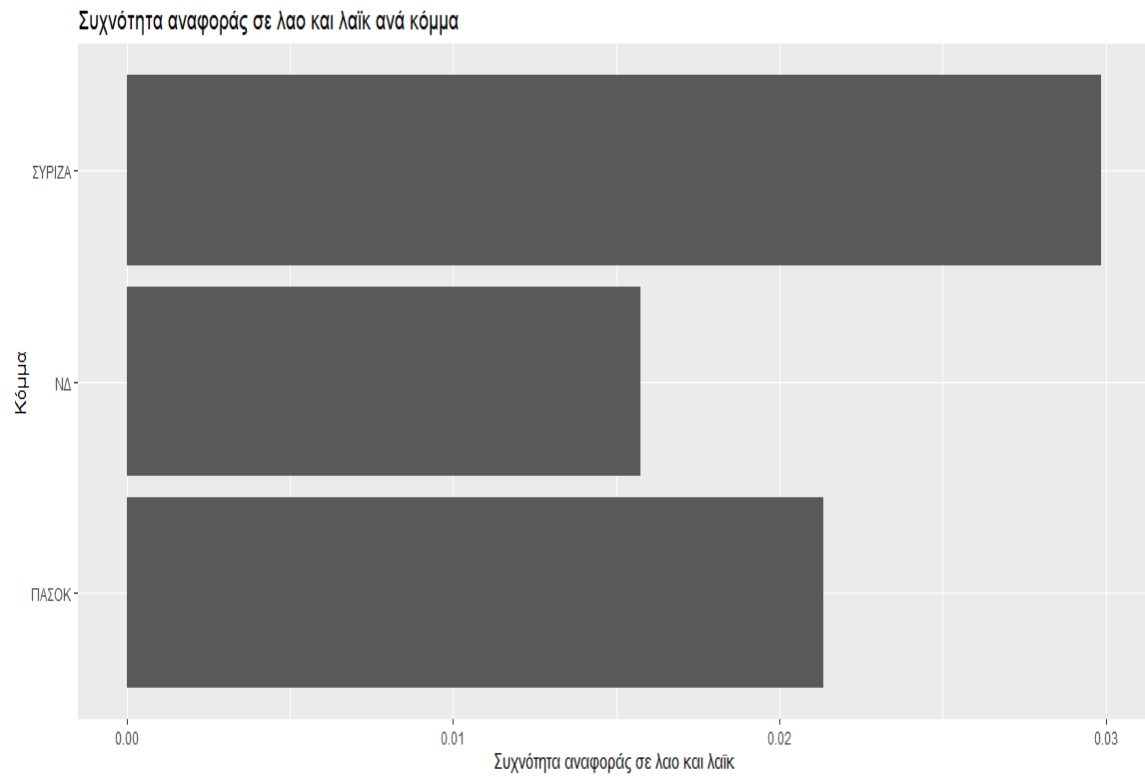
	Αναφορά σε λαό	Αναφορά σε Ευρώπη	Αναφορά σε εχθρούς του λαού
Υποψήφιοι 2015	2.5	3.3	1.7
Υποψήφιοι 2019	2.2	4.6	2.5

Βλέπουμε ότι και οι δύο ομάδες υποψηφίων αναφέρονται συγκριτικά περισσότερο στην Ευρώπη συγκριτικά με τον λαό και τους εχθρούς του λαού.

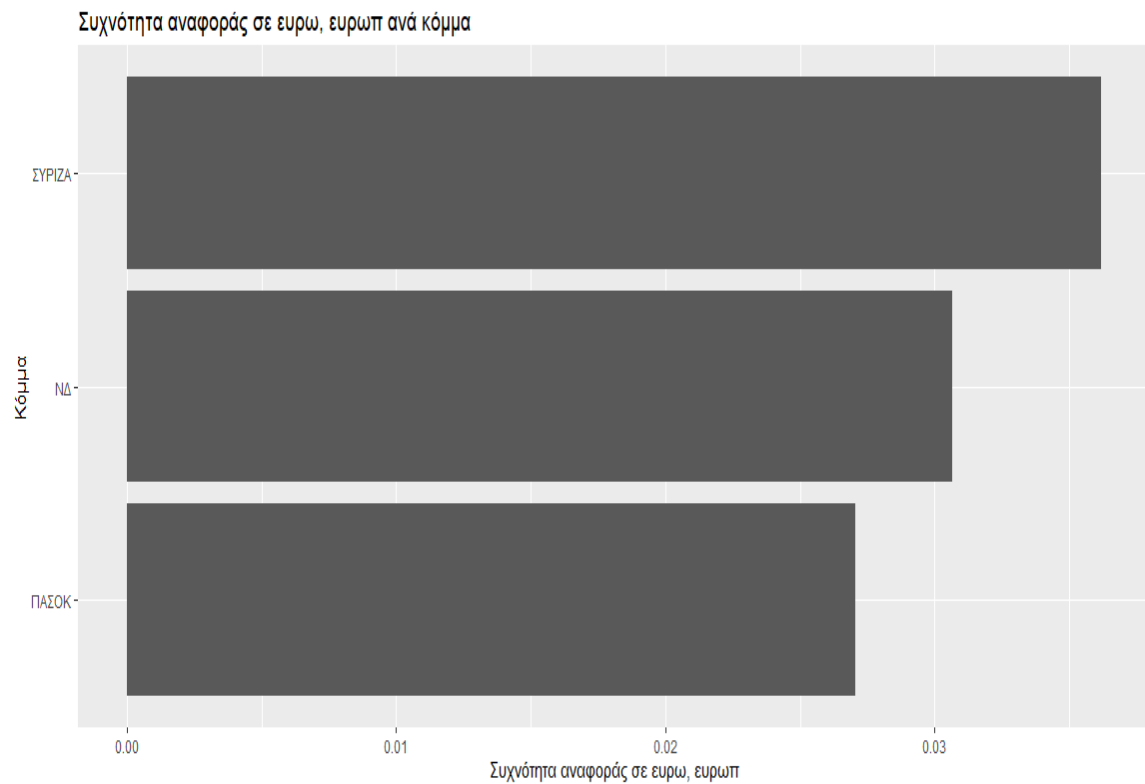
Στα παρακάτω γραφήματα (16,17 και 18), παραθέτουμε τη συχνότητα αναφοράς των παραπάνω λέξεων ανά κόμμα για τους υποψηφίους του 2015.

Οι κατανομές των σχετικών συχνοτήτων των αναφορών στο λαό μεταξύ των υποψηφίων του 2015 και του 2019 (Γραφήματα 16 και 1) είναι παρόμοιες. Οι σχετικές συχνότητες των αναφορών στην Ευρώπη είναι ελαφρώς διαφορετικές (Γραφήματα 17 και 2). Οι υποψήφιοι της ΝΔ έχουν περισσότερες αναφορές από τους υποψηφίους του ΣΥΡΙΖΑ, στην Ευρώπη το 2019, ενώ λιγότερες το 2015. Το ΠΑΣΟΚ – ΚΙΝΑΛ παραμένει και στις δύο περιπτώσεις στην τρίτη θέση. Τέλος, οι σχετικές συχνότητες των αναφορών στους εχθρούς του λαού (Γραφήματα 18 και 3) είναι και αυτές ελαφρώς διαφορετικές. Το ΚΙΝΑΛ είναι πρώτο στις αναφορές στους εχθρούς του λαού το 2019 (με πολύ μικρή διαφορά από τον ΣΥΡΙΖΑ που είναι δεύτερος), ενώ δεύτερο το 2015 (και πρώτο ο ΣΥΡΙΖΑ). Η ΝΔ παραμένει στην τρίτη θέση και στις δύο περιπτώσεις.

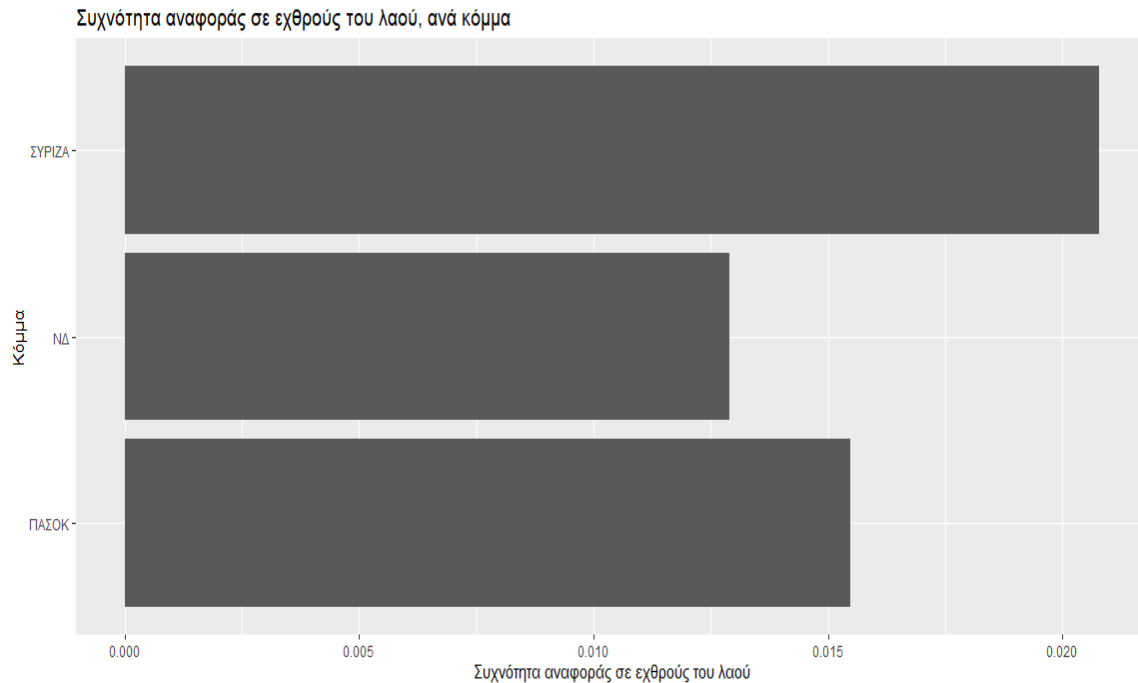
ΓΡΑΦΗΜΑ 16



ΓΡΑΦΗΜΑ 17



ΓΡΑΦΗΜΑ 18



Περνώντας τώρα στην ανάλυση συναισθήματος υπολογίσαμε ότι το μέσο συναίσθημα των tweets των υποψηφίων είναι -0.31 , ελαφρώς πιο αρνητικό από το μέσο συναίσθημα των υποψηφίων του 2019 το οποίο είναι -0.22 . Γεγονός λογικό αν σκεφτούμε ότι η περίοδος του 2015 ήταν μία έντονα πολωτική περίοδος (με την αποκορύφωση της πόλωσης να είναι το δημοψήφισμα), και επίσης μια περίοδος εν μέσω οικονομικής κρίσης. Στον πίνακα 11 βλέπουμε κατά πόσο αλλάζει το μέσο συναίσθημα των tweets όταν γίνεται αναφορά στο λαό, στην Ευρώπη και στους εχθρούς του λαού αντίστοιχα, από τους υποψήφιους του 2019 και του 2015 αντίστοιχα.

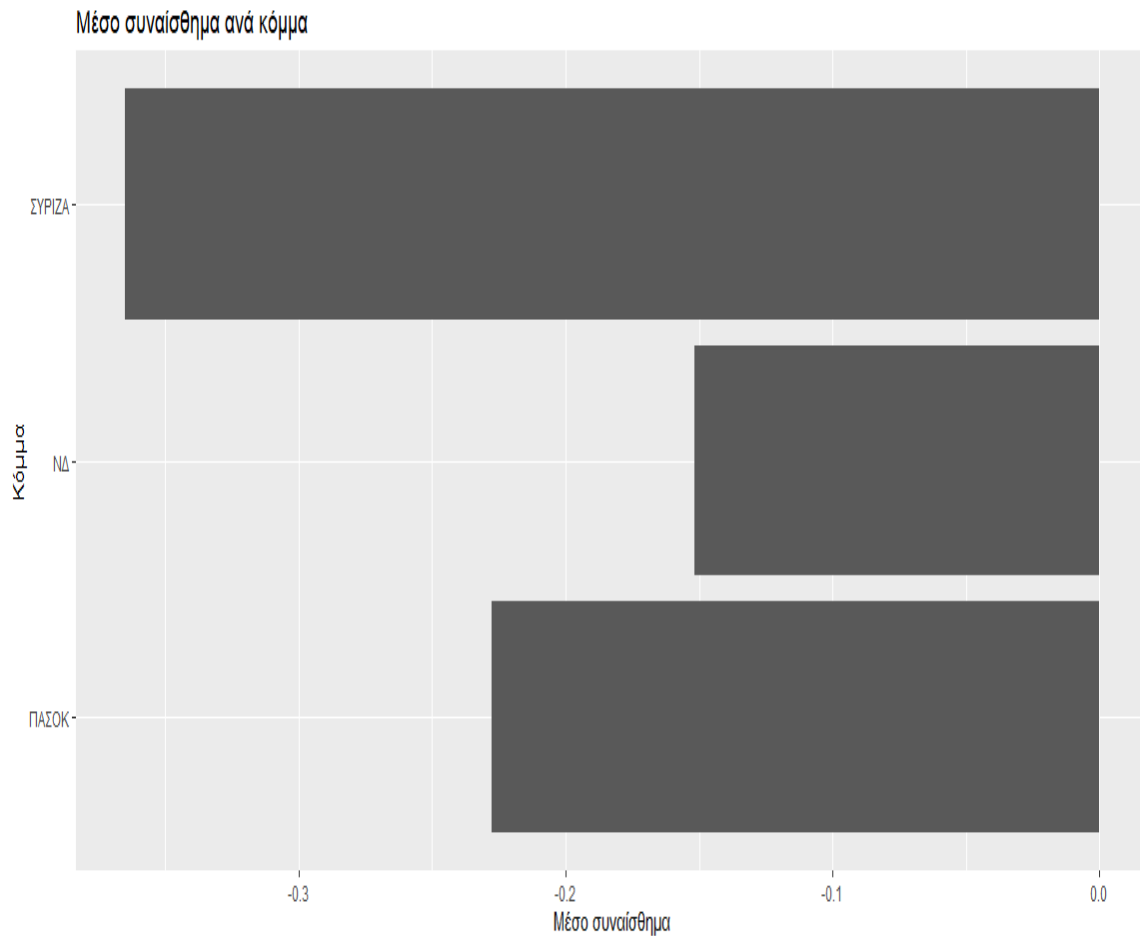
Από τον πίνακα 11 βλέπουμε ότι και οι δύο ομάδες υποψηφίων εκφράζονται δυσανάλογα πιο αρνητικά όταν αναφέρονται στην Ευρώπη και στους εχθρούς του λαού, απ' ότι όταν αναφέρονται στο λαό.

Στη συνέχεια στο γράφημα 19 απεικονίζουμε το μέσο συναίσθημα ανά κόμμα για τους υποψήφιους του 2015. Η εικόνα που βλέπουμε στον εν λόγω γράφημα είναι παρόμοια με την εικόνα του 2019 (Γράφημα 4). Τα πιο αρνητικά tweets κατά μέσο όρο είναι του ΣΥΡΙΖΑ ακολουθούμενα από αυτά του ΠΑΣΟΚ – ΚΙΝΑΛ και έπειτα από αυτά της ΝΔ. Η μόνη διαφορά είναι ότι η ΝΔ είναι ελαφρώς πιο αρνητική το 2015 σε σχέση με το 2019 και αυτό σχετίζεται μάλλον με το γεγονός ότι το συνολικό συναίσθημα του 2015 ήταν πιο αρνητικό από αυτό του 2019 όπως είδαμε παραπάνω.

ΠΙΝΑΚΑΣ 11: Αναφορές σε λαό, Ευρώπη, εχθρούς λαού και συναίσθημα. Υποψήφιοι 2019 και 2015.

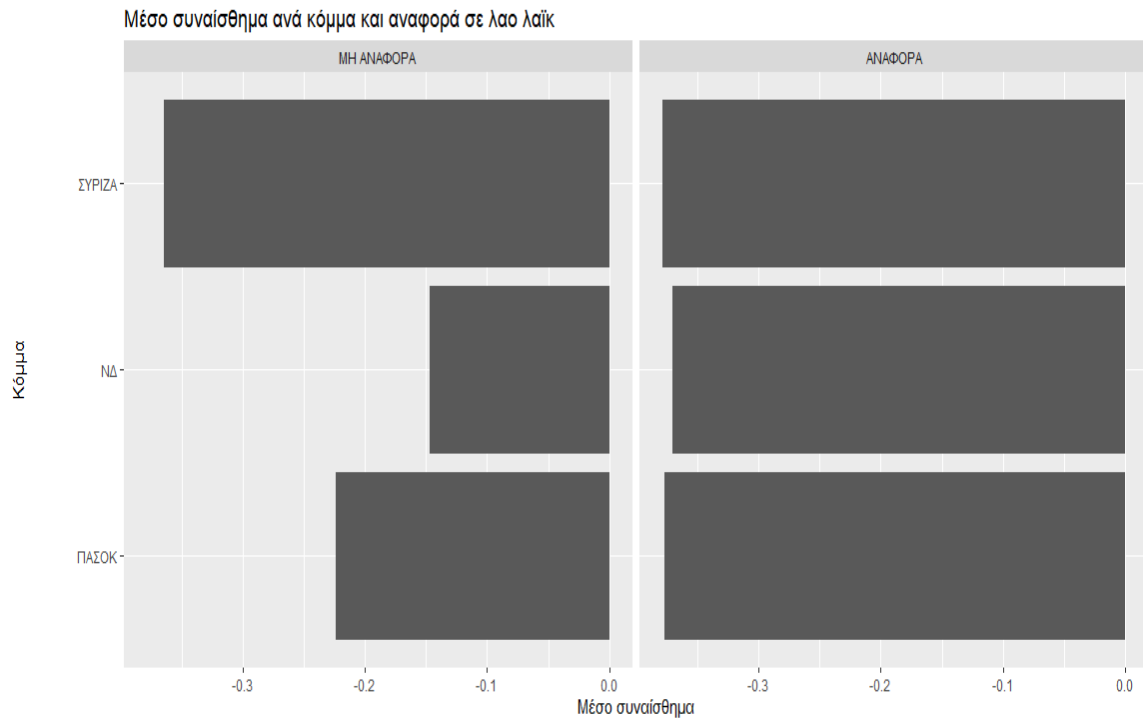
	ΥΠΟΨΗΦΙΟΙ 2015			ΥΠΟΨΗΦΙΟΙ 2019		
	Λαός	Ευρώπη	Εχθροί Λαού	Λαός	Ευρώπη	Εχθροί λαού
Μη αναφορά	-0.316	-0.296	-0.302	-0.220	-0.199	-0.214
Αναφορά	-0.084	-0.600	-0.530	-0.300	-0.593	-0.478

ΓΡΑΦΗΜΑ 19

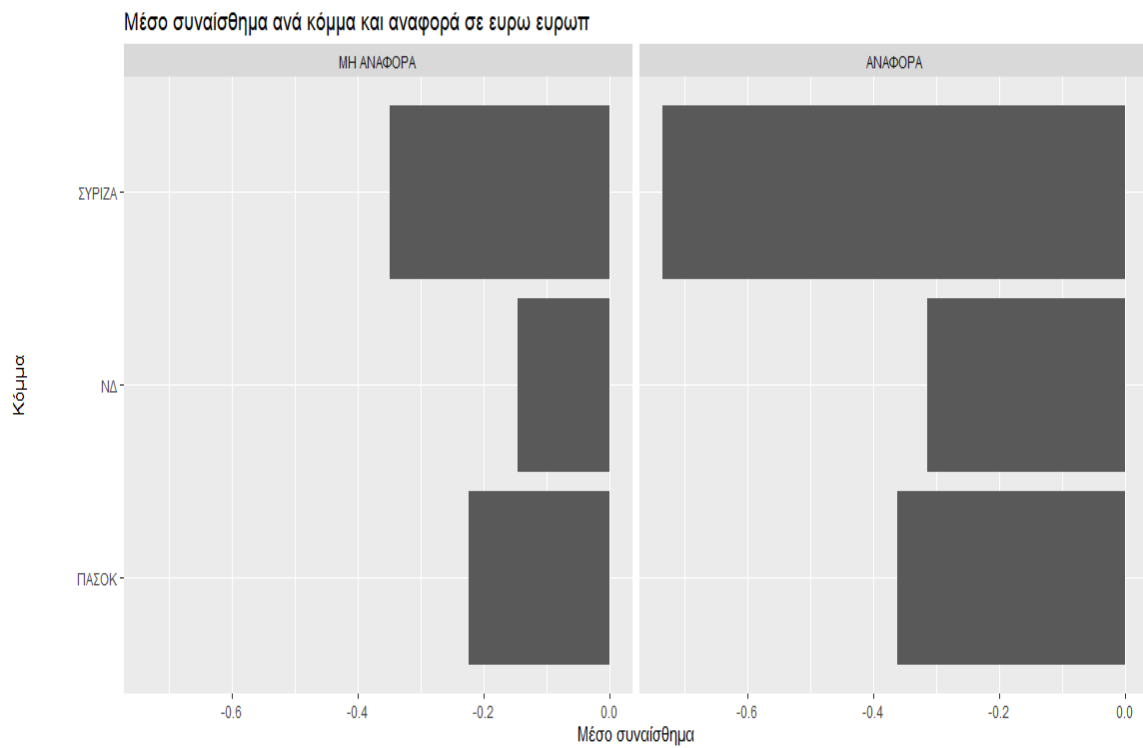


Στα επόμενα τρία γραφήματα (20-22) παρουσιάζουμε το μέσο συναίσθημα ανά κόμμα για τις περιπτώσεις των αναφορών σε λαό, Ευρώπη και εχθρούς του λαού.

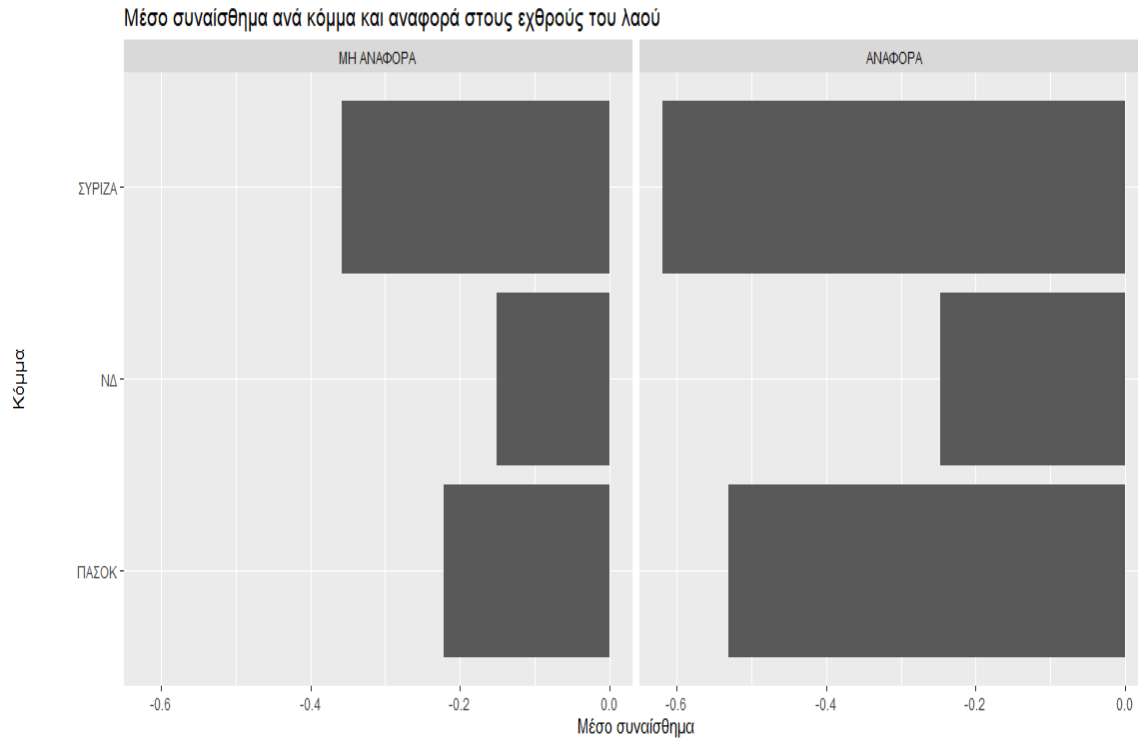
ΓΡΑΦΗΜΑ 20



ΓΡΑΦΗΜΑ 21



ΓΡΑΦΗΜΑ 22



Συγκρίνοντας το γράφημα 20 και το γράφημα 5 διαπιστώνουμε ότι όταν αναφέρεται ο λαός, ο αρνητισμός αυξάνεται για όλα τα κόμματα του 2015, ενώ το 2019 αυξάνεται μόνο για ΣΥΡΙΖΑ και ΚΙΝΑΛ (κυρίως το δεύτερο). Και στις δύο περιπτώσεις ο μεγάλος αρνητισμός του ΚΙΝΑΛ προκύπτει κυρίως από τα tweets της Φώφης Γεννηματά (το 2015 ως απλή υποψήφιος και το 2019 ως πρόεδρος του κόμματος). Και στην περίπτωση του 2015 ο αρνητισμός των tweet της Γεννηματά θεματολογικά φαίνεται να αφορά μία αντίθεση προς το λαϊκισμό του ΣΥΡΙΖΑ, αλλά και μια κριτική προς τον υπάρχοντα δικομματισμό ΣΥΡΙΖΑ-ΝΔ.

Συγκρίνοντας τώρα τα γραφήματα 21 και 6 βλέπουμε ότι η κατανομή του μέσου συναίσθηματος των κομμάτων του 2015 είναι όπως την περιμέναμε¹³. Ο πιο αρνητικός όταν γίνεται αναφορά στην Ευρώπη είναι ο ΣΥΡΙΖΑ, ακολουθούμενος από το ΠΑΣΟΚ και έπειτα τη ΝΔ. Στην περίπτωση του 2019, όπως είδαμε και σε προηγούμενο κεφάλαιο, το ΚΙΝΑΛ είναι περισσότερο αρνητικό λόγω των tweets της Γεννηματά και της Δημαδάμα τα οποία αναφέρονται αρνητικά στην επιθετικότητα της Τουρκίας στον Έβρο και τονίζοντας, παράλληλα, την ανάγκη της Ευρωπαϊκής Ένωσης να παρέμβει διπλωματικά.

¹³ Δεδομένου όπως είπαμε και στην εισαγωγή ότι περισσότερος αρνητισμός ενδεχομένως συνεπάγεται και περισσότερο λαϊκισμό/ευρωσκεπτικισμό, μια κατανομή όπως αυτή του γραφήματος 21 είναι κατανοητή αν λάβουμε υπόψη ότι ο ΣΥΡΙΖΑ είναι το περισσότερο λαϊκιστικό/ευρωσκεπτικιστικό κόμμα και η ΝΔ το λιγότερο με το ΠΑΣΟΚ-ΚΙΝΑΛ να είναι κάπου ενδιάμεσα.

Τέλος, συγκρίνοντας το γράφημα 22 με το γράφημα 7 διαπιστώνουμε ότι και στην περίπτωση των αναφορών στους εχθρούς του λαού η κατανομή των κομμάτων για το 2015 είναι όπως την αναμένουμε. Για το 2019 έγινε η επίσημανση ότι το ΚΙΝΑΛ φαίνεται περισσότερο αρνητικό απ'όσο θα αναμέναμε λόγω των tweets του υποψήφιου βουλευτή Μιγάλη Κατρίνη.

Στη συνέχεια εξετάζουμε τον αντίκτυπο που είχαν τα λαϊκιστικά και ευρωσκεπτικιστικά tweets των υποψηφίων του 2015 στο Twitter, μέσω των favorites και των retweets και θα τα συγκρίνουμε με τα αντίστοιχα των υποψηφίων του 2019.

ΠΙΝΑΚΑΣ 12: Μέσος όρος retweets ανά αναφορά σε λαό, Ευρώπη και εχθρούς του λαού. Υποψήφιοι 2015 και 2019.

	ΥΠΟΨΗΦΙΟΙ 2019			ΥΠΟΨΗΦΙΟΙ 2015		
	Λαός	Ευρώπη	Εχθροί Λαού	Λαός	Ευρώπη	Εχθροί λαού
Μη αναφορά	14.9	14.9	15.0	5.7	5.7	5.7
Αναφορά	20.6	17.8	18.4	7.7	7.5	7.1

ΠΙΝΑΚΑΣ 13: Μέσος όρος favorites ανά αναφορά σε λαό, Ευρώπη και εχθρούς του λαού. Υποψήφιοι 2015 και 2019.

	ΥΠΟΨΗΦΙΟΙ 2019			ΥΠΟΨΗΦΙΟΙ 2015		
	Λαός	Ευρώπη	Εχθροί Λαού	Λαός	Ευρώπη	Εχθροί Λαού
Μη αναφορά	69.9	70.0	70.3	9.3	9.3	9.4
Αναφορά	98.4	80.8	79.7	15.2	14.2	13.0

Από τους πίνακες 12 και 13 παρατηρούμε ότι ο μέσος όρος των retweets και των favorites αυξάνεται όταν υπάρχει αναφορά στο λαό, στην Ευρώπη και στους εχθρούς του λαού. Το ίδιο ισχύει και για τους υποψηφίους του 2015 και για τους υποψηφίους του 2019.

Παρακάτω στον πίνακα 14 εξετάζουμε το μέσο όρο retweets και favorites για κάθε συναισθηματική κατηγορία έτσι ώστε να δούμε κατά πόσο είναι μεγαλύτερη η απήχηση των αρνητικών tweets, κάνοντας ταυτόχρονα σύγκριση μεταξύ των υποψηφίων του 2019 και των υποψηφίων του 2015.

ΠΙΝΑΚΑΣ 14: Μέσος όρος retweets και favorites ανά συναισθηματική κατηγορία tweets. Υποψήφιοι 2019 και 2015.

Συναισθηματική Κατηγορία Tweets	ΥΠΟΨΗΦΙΟΙ 2019	ΥΠΟΨΗΦΙΟΙ 2019	ΥΠΟΨΗΦΙΟΙ 2015	ΥΠΟΨΗΦΙΟΙ 2015
	Retweets	Favorites	Retweets	Favorites
Αρνητικά	19.4	89.1	4.5	9.7
Ουδέτερα	15.1	71.5	4.1	8.9
Θετικά	16.2	86.8	4.1	9.6

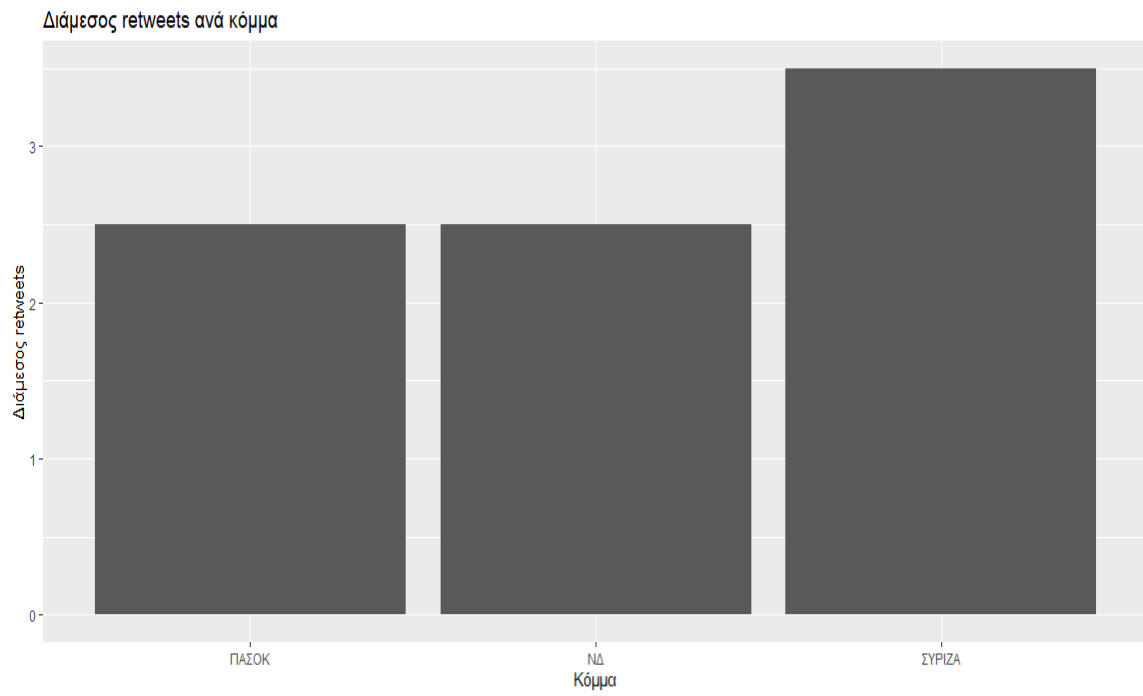
Από τον πίνακα 14 βλέπουμε ότι τα αρνητικά συναισθήματα έχουν ελαφρώς περισσότερα retweets και favorites τόσο των υποψηφίων του 2019 όσο και των υποψηφίων του 2015. Συνεπώς, δεδομένου ότι τα αρνητικά tweets συνεπάγονται περισσότερο λαϊκισμό/ευρωσκεπτικισμό μπορούμε να πούμε ότι υπάρχει απήχηση του λαϊκιστικού και ευρωσκεπτικιστικού λόγου και των υποψηφίων του 2019 και των υποψηφίων του 2015.

Στα επόμενα δύο γραφήματα (Γράφημα 23 και 24) βλέπουμε τη διάμεσο των retweets ανά κόμμα και τον μέσο όρο των favorites ανά κόμμα για τους υποψηφίους του 2015. Ο λόγος που επιλέξαμε τη διάμεσο αντί για το μέσο όρο στην περίπτωση των retweets είναι γιατί υπάρχουν ορισμένες ακραίες τιμές retweets, οι οποίες προέρχονται κυρίως από το λογαριασμό του βουλευτή του ΣΥΡΙΖΑ Γιάννη Βαρουφάκη. Επίσης, χρειάζεται να σημειωθεί ότι για όλους τους υπολογισμούς σχετικά με τα retweets που ακολουθούν αφαιρέσαμε τα tweets εκείνα που είχαν 0 retweets. Ο λόγος για αυτό είναι το γεγονός ότι ο αριθμός των μηδενικών retweets είναι πολύ μεγάλος (64% των tweets έχει 0 retweets) με αποτέλεσμα αυτά να «τραβάνε» τη διάμεσο προς το μέρος τους με αποτέλεσμα η διάμεσος κάθε κόμματος να λαμβάνει την τιμή 0 και έτσι να μην μπορεί να γίνει περαιτέρω ανάλυση.

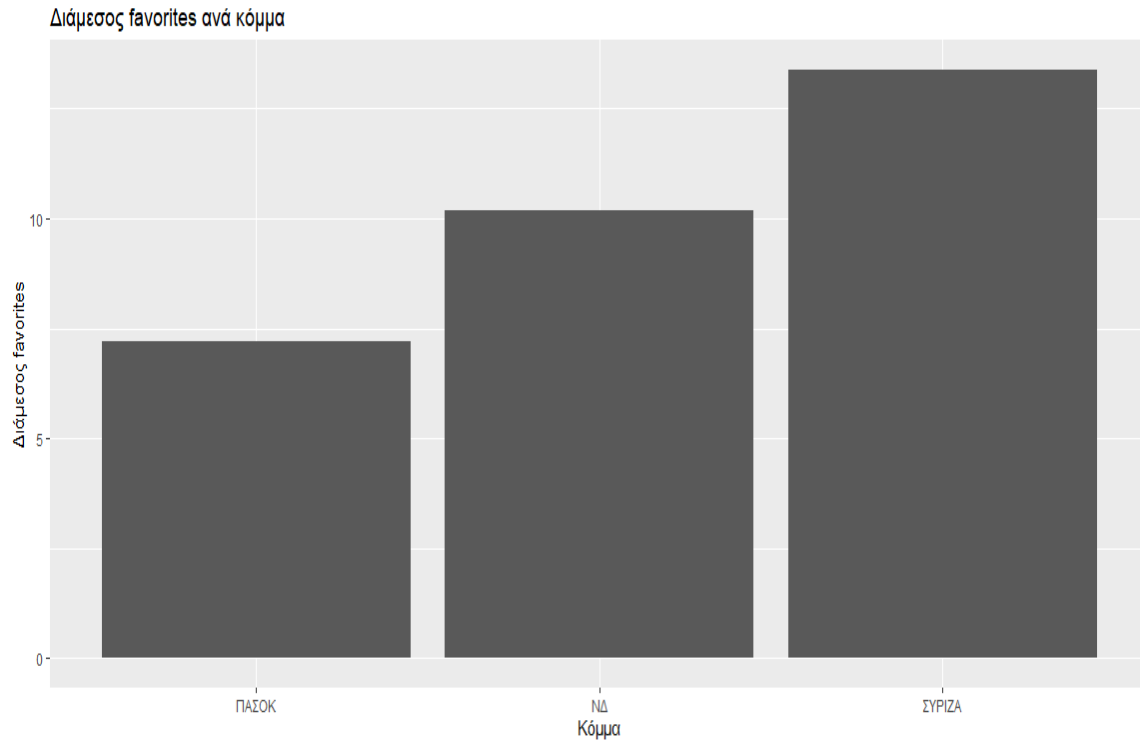
Από το γράφημα 23 και το γράφημα 24 βλέπουμε ότι οι κατανομές των retweets και των favorites ανά κόμμα είναι ίδιες με τις αντίστοιχες κατανομές του 2015. Ο ΣΥΡΙΖΑ είναι πρώτος σε retweets και favorites και στις δύο περιπτώσεις.

Στα επόμενα 6 γραφήματα (25-30) αποτυπώνουμε τα retweets και τα favorites ανά κόμμα και αναφορά στο λαό, στην Ευρώπη και στους εχθρούς του λαού για τους υποψηφίους του 2015.

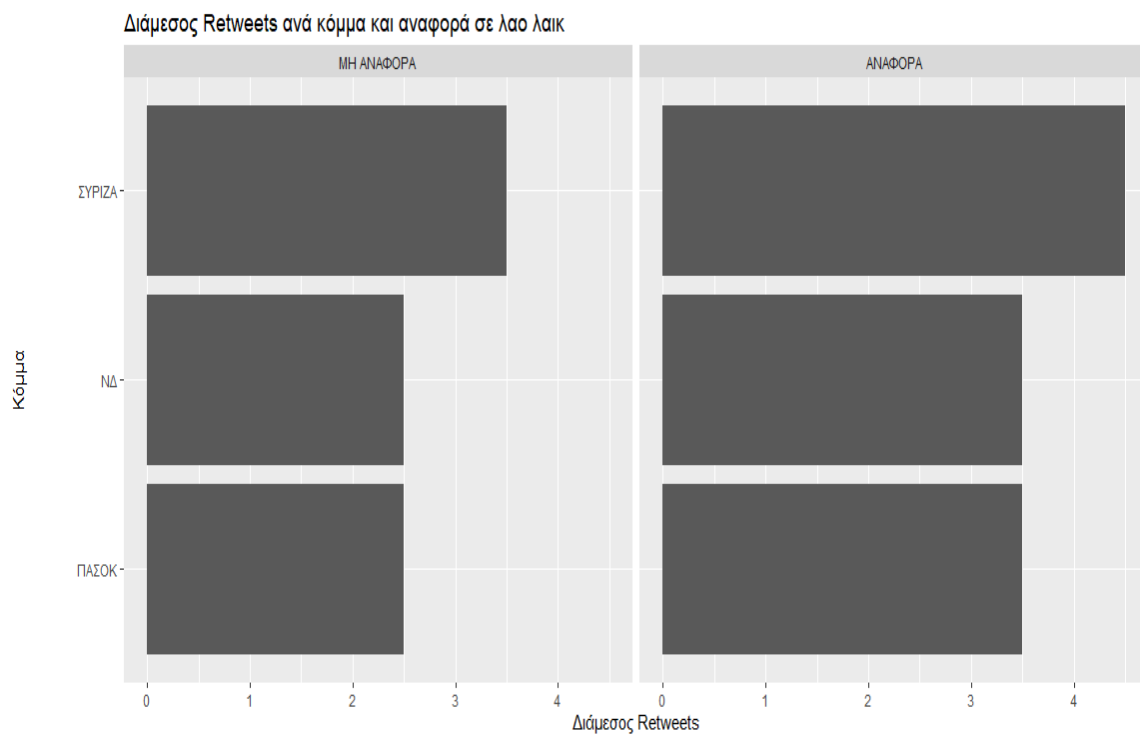
ΓΡΑΦΗΜΑ 23



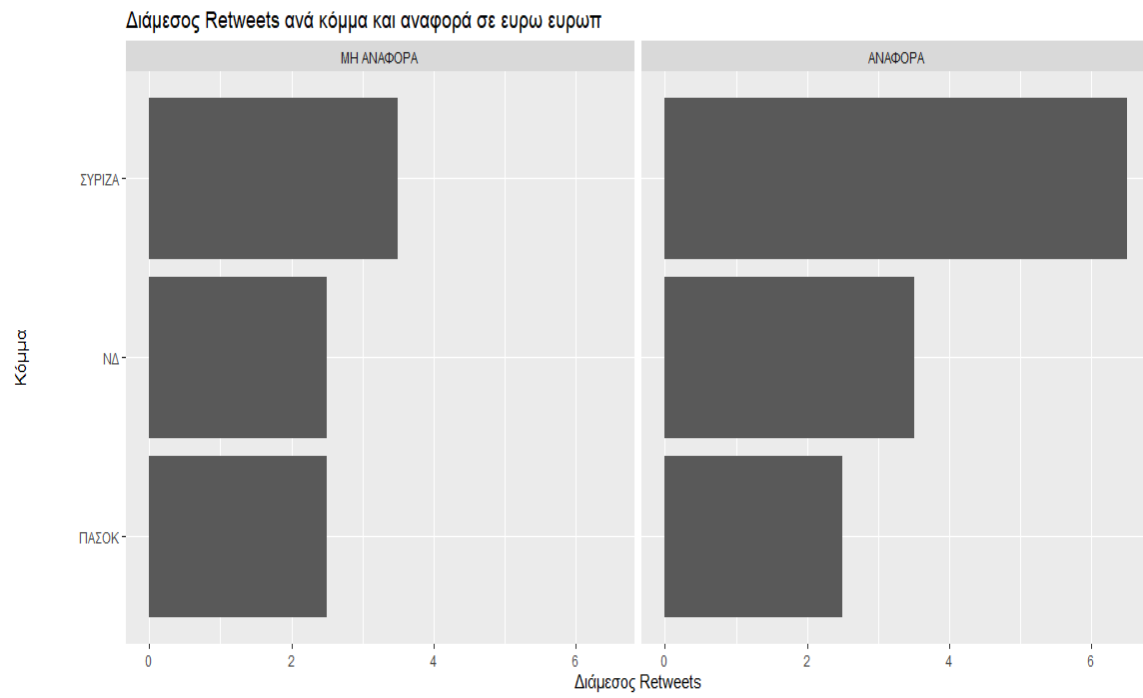
ΓΡΑΦΗΜΑ 24



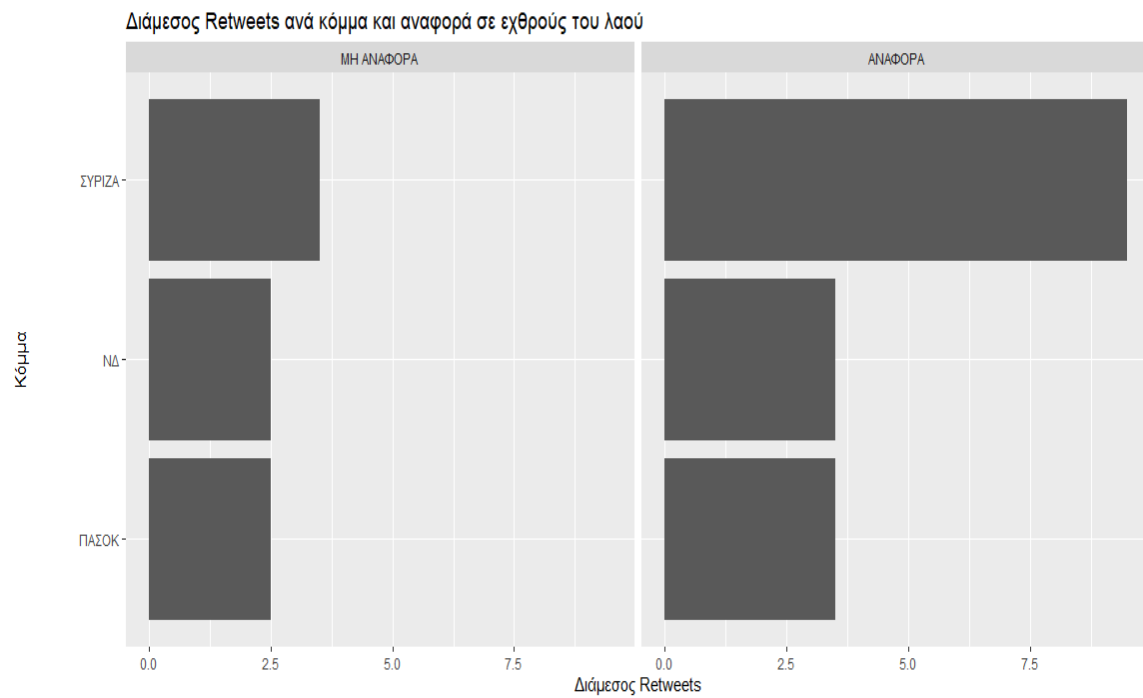
ΓΡΑΦΗΜΑ 25



ΓΡΑΦΗΜΑ 26

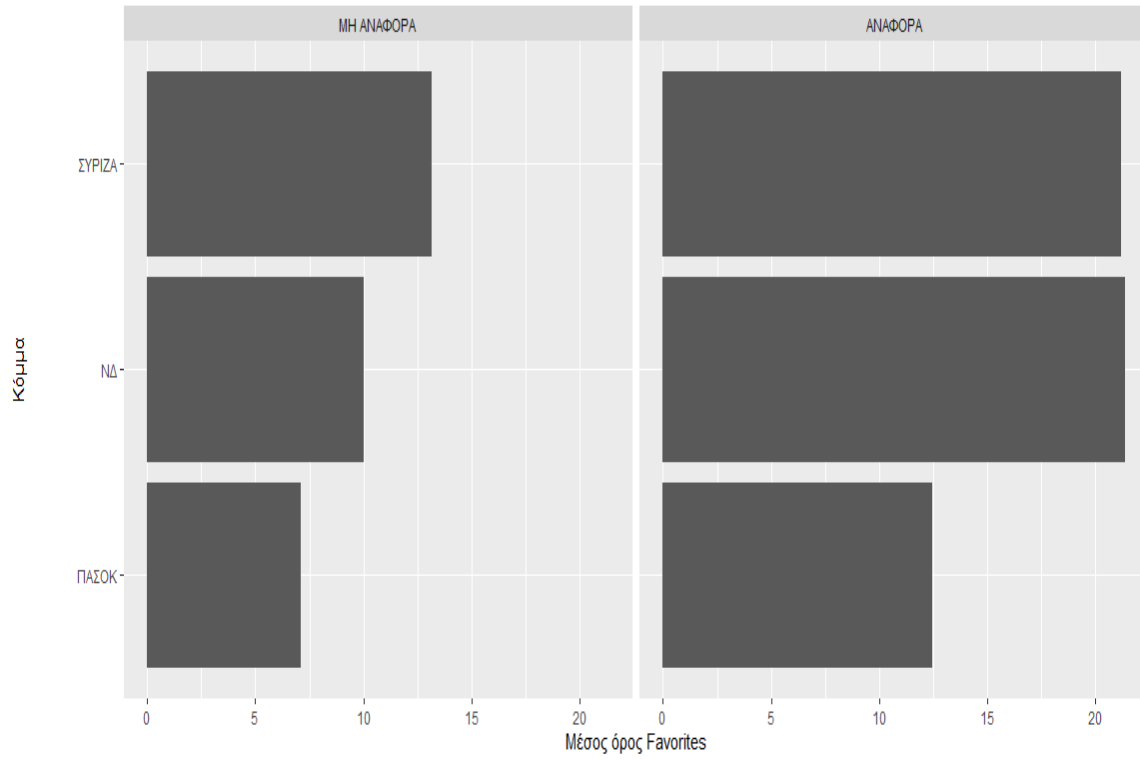


ΓΡΑΦΗΜΑ 27



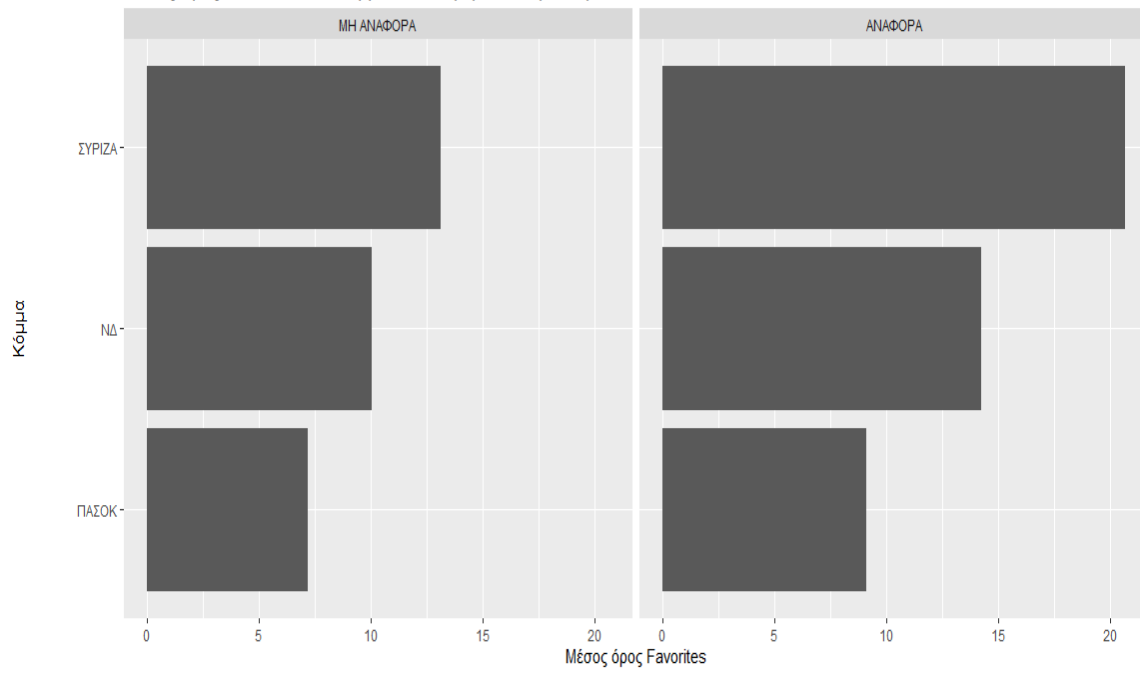
ΓΡΑΦΗΜΑ 28

Μέσος όρος Favorites ανά κόμμα και αναφορά σε λαο λαικ

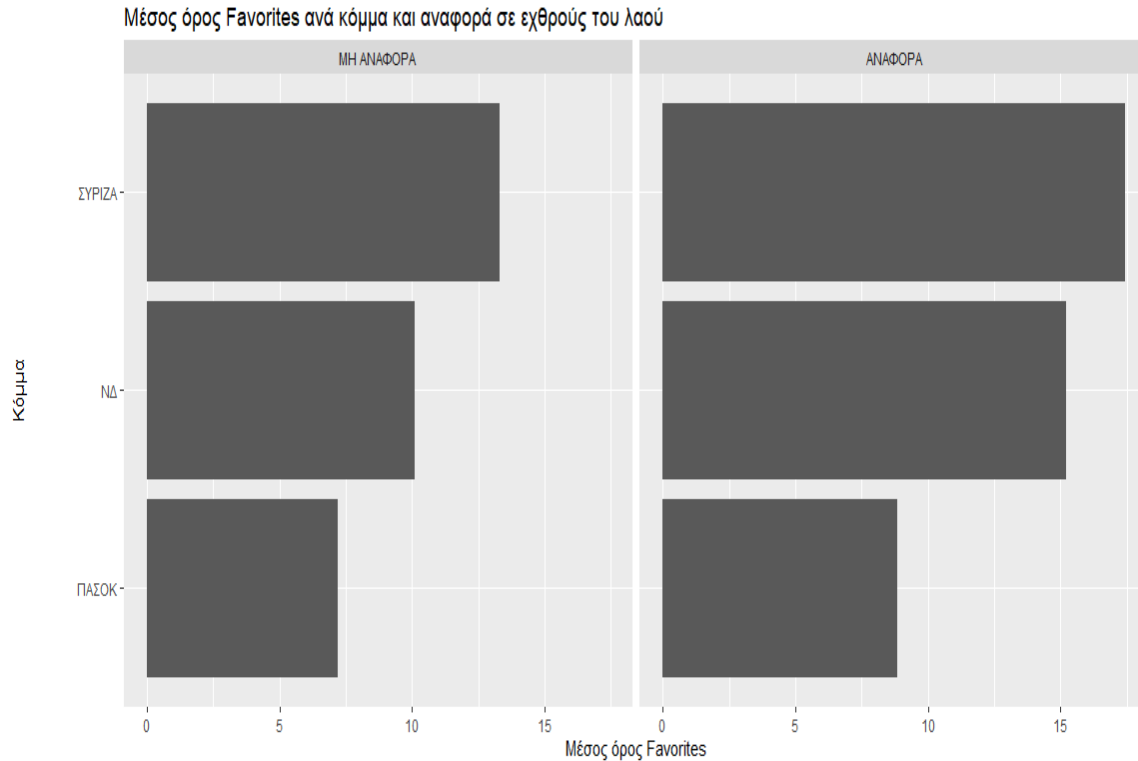


ΓΡΑΦΗΜΑ 29

Μέσος όρος Favorites ανά κόμμα και αναφορά σε ευρω ευρωπ



ΓΡΑΦΗΜΑ 30



Για τα retweets (Γραφήματα 25-27) βλέπουμε ότι όταν υπάρχει αναφορά σε λαό, Ευρώπη και εχθρούς του λαού η διάμεσος των retweets ανεβαίνει περισσότερο για τον ΣΥΡΙΖΑ και λιγότερο για το ΠΑΣΟΚ και τη ΝΔ. Η εικόνα είναι κάπως διαφορετική μόνο στα favorites όταν υπάρχει αναφορά στο λαό όπου η ΝΔ έρχεται πρώτη. Σε αυτή την περίπτωση αυτό οφείλεται στα tweets του προέδρου της ΝΔ Κυριάκου Μητσοτάκη τα οποία αναφέρονται στο λαό. Ωστόσο, τα tweets του συγκεκριμένου φαίνεται στην πλειοψηφία τους να αναφέρουν το λαϊκισμό και να εκδηλώνουν μια αρνητική στάση απέναντί του. Με άλλα λόγια, φαίνεται ο αντιλαϊκισμός του συγκεκριμένου υποψηφίου να συγκεντρώνει αρκετά favorites.

Επίσης ενδιαφέρον έχει το γεγονός ότι η ΝΔ έρχεται δεύτερη σε όλες τις περιπτώσεις (πλην την περίπτωση των favorites σε αναφορές στο λαό). Το γεγονός αυτό οφείλεται σε δύο παράγοντες. Πρώτον, ότι το 2015 ο Κυριάκος Μητσοτάκης είναι ο υποψήφιος της ΝΔ με τα περισσότερα retweets και favorites των tweets με αναφορά στο λαό (τα οποία είναι στην πλειοψηφία τους αντιλαϊκιστικά). Δεύτερον, ότι το 2015, πάλι ο Κυριάκος Μητσοτάκης είναι ο υποψήφιος της ΝΔ με τα περισσότερα retweets και favorites στις αναφορές στην Ευρώπη και στους εχθρούς του λαού. Τα συγκεκριμένα tweets του εν λόγω υποψηφίου δεν φαίνεται να είναι λαϊκιστικά ή ευρωσκεπτικιστικά, αλλά αναφέρονται πολύ στην Ευρώπη και τις τράπεζες, καθώς ήταν μια περίοδος όπου η διαπραγματεύσεις με τους Ευρωπαίους εταίρους ήταν το πιο έντονο πολιτικό διακύβευμα της επικαιρότητας, όπως επίσης και η περίοδος των capital controls, και το συνακόλουθο κλείσιμο των τραπεζών.

Όσον αφορά γενικότερα τα retweets και τα favorites ανά κόμμα και αναφορά σε λαό, Ευρώπη και εχθρούς του λαού τα γραφήματα για το 2015 (25-30) φαίνεται να παρουσιάζουν πολύ παρόμοια εικόνα με εκείνη του 2019 (γραφήματα 10-15). Η διαφορά φαίνεται να είναι μόνο στα retweets, όπου η ΝΔ έρχεται δεύτερη θέση σε όλες τις αναφορές (λαό, Ευρώπη, εχθρούς λαού) το 2015, αλλά τρίτη θέση στις αντίστοιχες αναφορές το 2019. Ο λόγος για αυτό επισημάνθηκε αμέσως παραπάνω και έχει να κάνει με το γεγονός ότι κυρίως ο Κυριάκος Μητσοτάκης συγκεντρώνει πολλά retweets (και favorites) στα tweets που αναφέρεται συχνά, συνήθως με αντιλαϊκιστικό τόνο, στο λαό στην Ευρώπη και στους εχθρούς του λαού.

Τέλος, να αναφέρουμε ότι οι πρώτοι πέντε υποψήφιοι του 2015 με τα περισσότερα retweets κατά μέσο όρο είναι με τη σειρά ο Γιάννης Βαρουφάκης, ο Αντώνης Σαμαράς, ο Νίκος Φίλης, ο Κυριάκος Μητσοτάκης και ο Αλέξης Τσίπρας. Όσον αφορά τα favorites, οι πρώτοι υποψήφιοι είναι παρόμοιοι με τη διαφορά ότι από την πεντάδα βγαίνει ο Νίκος Φίλης και μπαίνει ο Άδωνις Γεωργιάδης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3), 267–297. <https://doi.org/10.1093/pan/mps028>
- Kaufman, A. R., & Klevs, A. (2021). Adaptive Fuzzy String Matching: How to Merge Datasets with Only One (Messy) Identifying Field. *Political Analysis*, 1–7. <https://doi.org/10.1017/pan.2021.38>
- Ntais, G. (2006). *Development of a Stemmer for the Greek Language*. 46.
- Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14(3), 130–137. <https://doi.org/10.1108/eb046814>

Kartsounidou, E. & Andreadis, I. (2020), *Exploring Congruence in Greece (2009-2019)*.

European Consortium for Political Research (ECPR).