

# Ανάλυση Συναισθήματος

Στα tweets του Donald Trump από το 2017 έως το 2021

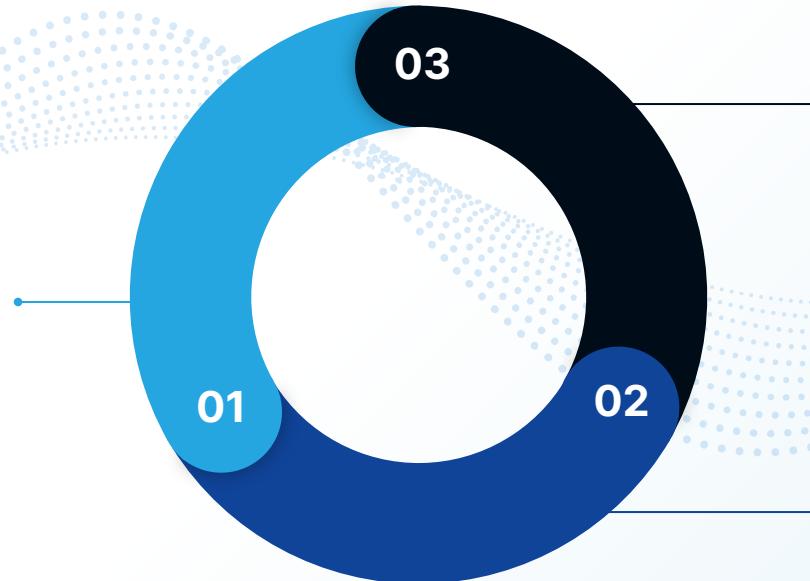
Διδάσκον: Κουτσούπιάς Νικόλαος

Φώτογλου Κωνσταντίνος, 4515, (ies1545)  
Μέθοδοι και Τεχνικές Έρευνας Πεδίου II,  
Πανεπιστήμιο Μακεδονίας, 2020-21

# Δομή της Εργασίας

## Πρώτο στάδιο

Αρχικά γίνεται μία σύντομη αναφορά στην Ανάλυση Συναισθήματος



## Τρίτο Στάδιο

Παρουσίαση μεθοδολογίας, κώδικα, και αποτελεσμάτων

## Δεύτερο Στάδιο

Παρουσίαση των βασικών πακέτων που χρησιμοποιήθηκαν στην ανάλυση

# 1. Ανάλυση Συναίσθηματος

Συνοπτική αναφορά στην θεωρία

# Ανάλυση Συναισθήματος

- "Ανάλυση Συναισθήματος είναι η μελέτη γνώμης, αισθημάτων, στάσεων και συναισθημάτων φυσικών προσώπων με εφαρμογές πληροφορικής " Bing Liu (2016)
- Ως τομέας έρευνας είναι σχετικά νέος. Πιο συγκεκριμένα ξεκίνησε στις αρχές του 21ου αιώνα από τους Bo Pang, Lillian Lee, Shivakumar Vaithyanathan

# Ανάλυση Συναισθήματος: Εφαρμογές

Επιχειρήσεις

Εξαγωγή πληροφοριών σχετικά με την αποδοχή των προϊόντων από τους καταναλωτές

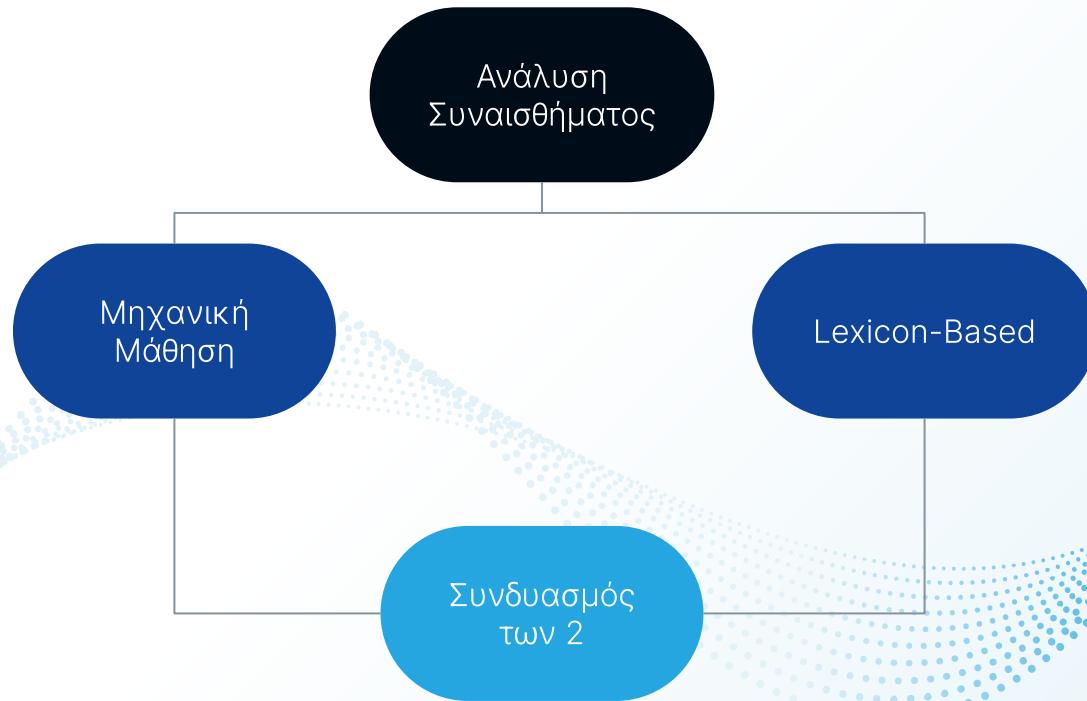
Πολιτική

Ανάλυση στάσης της κοινής γνώμης σε διάφορα ζητήματα

Οικονομία, Χρ/κή

Πρόβλεψη οικονομικών δεικτών, τιμών μετοχών, κ.α.

# Ανάλυση Συναισθήματος: Προσεγγίσεις



## 2.Παρουσίαση Πακέτων

Αναφορά στα πακέτα που χρησιμοποιήθηκαν

# Παρουσίαση Πακέτων: Κύρια Πακέτα

## ggplot2

Πακέτο για τη δημιουργία γραφικών.

## tm

Πακέτο με λειτουργίες εξόρυξης κειμένου

## syuzhet

Πακέτο με λειτουργίες εξαγωγής συναισθήματος από κείμενο χρησιμοποιώντας μία ποικιλία από λεξικά συναισθημάτων

# Παρουσίαση Πακέτων: Λοιπά Πακέτα

**lubridate**

Κωδικοποίηση ημ/νιών και  
ωρών

**scales**

Λειτουργίες κλίμακας,  
επεξεργασία αξόνων

**wordcloud**

Δημιουργία νεφών  
λέξεων

**stringr**

Πακέτο διαχείρισης  
ποιοτικών μεταβλητών

**reshape2**

Πακέτο αναδιάρθρωσης  
δεδομένων

**dplyr**

Πακέτο χειραγώησης  
δεδομένων

## 2.Παρουσίαση Ανάλυσης

Μεθοδολογία, κύρια, αποτελέσματα

## Trump Twitter Archive V2

Insights FAQ

insights FAQ

Twitter has permanently suspended Trump's account (January 8th, 2021)

Search for anything...

Search tips Retweet filters Deleted filters Date filters Device filters Export

56,571 tweets found

Latest 

1. Jan 8th 2021 - 10:44:28 AM EST  
84k 511k Show

To all of those who have asked, I will not be going to the Inauguration on January 20th.

2. Jan 8th 2021 - 9:46:38 AM EST  
109k 481k Show

The 75,000,000 great American Patriots who voted for me, AMERICA FIRST, and MAKE AMERICA GREAT AGAIN, will have a GIANT VOICE long into the future. They will not be disrespected or treated unfairly in any way, shape or form!!!

3. Jan 7th 2021 - 7:10:24 PM EST  
155k 629k Show

<https://t.co/csX07ZVWGe>

4. Jan 6th 2021 - 6:01:04 PM EST

These are the things and events that happen when a sacred landslide election victory is so unceremoniously & viciously stripped away from great patriots who have been badly

## Trump Twitter Archive V2

Insight

### FAQs

#### What's happening here?

This site checks Twitter every 60 seconds and records every Trump tweet into a database. Before the site launched in 2016, all available tweets were captured and added to the database for perpetuity.

#### Can I have the data?

Yes. You can download the full dataset via Google Drive either as a [CSV file](#) or [JSON file](#) (last updated 01/08/2020). You can credit the site however you'd like. Note that the dates are stored in UTC time.

Τα Δεδομένα

 f_tweets	16711 obs. of 9 variables	
 mySentiment	16711 obs. of 10 variables	
 tweets	56571 obs. of 9 variables	
 wordCorpus	Large SimpleCorpus (16711 elements, 2.8 MB)	

# Εισαγωγή των πακέτων στο R-Studio

```
#Εισαγωγή Πακέτων
library(ggplot2)
library(lubridate)
library(scales)
library(tm)
library(stringr)
library(wordcloud)
library(syuzhet)
library(reshape2)
library(dplyr)
```

# Εισαγωγή, κωδικοποίηση και φιλτράρισμα δεδομένων

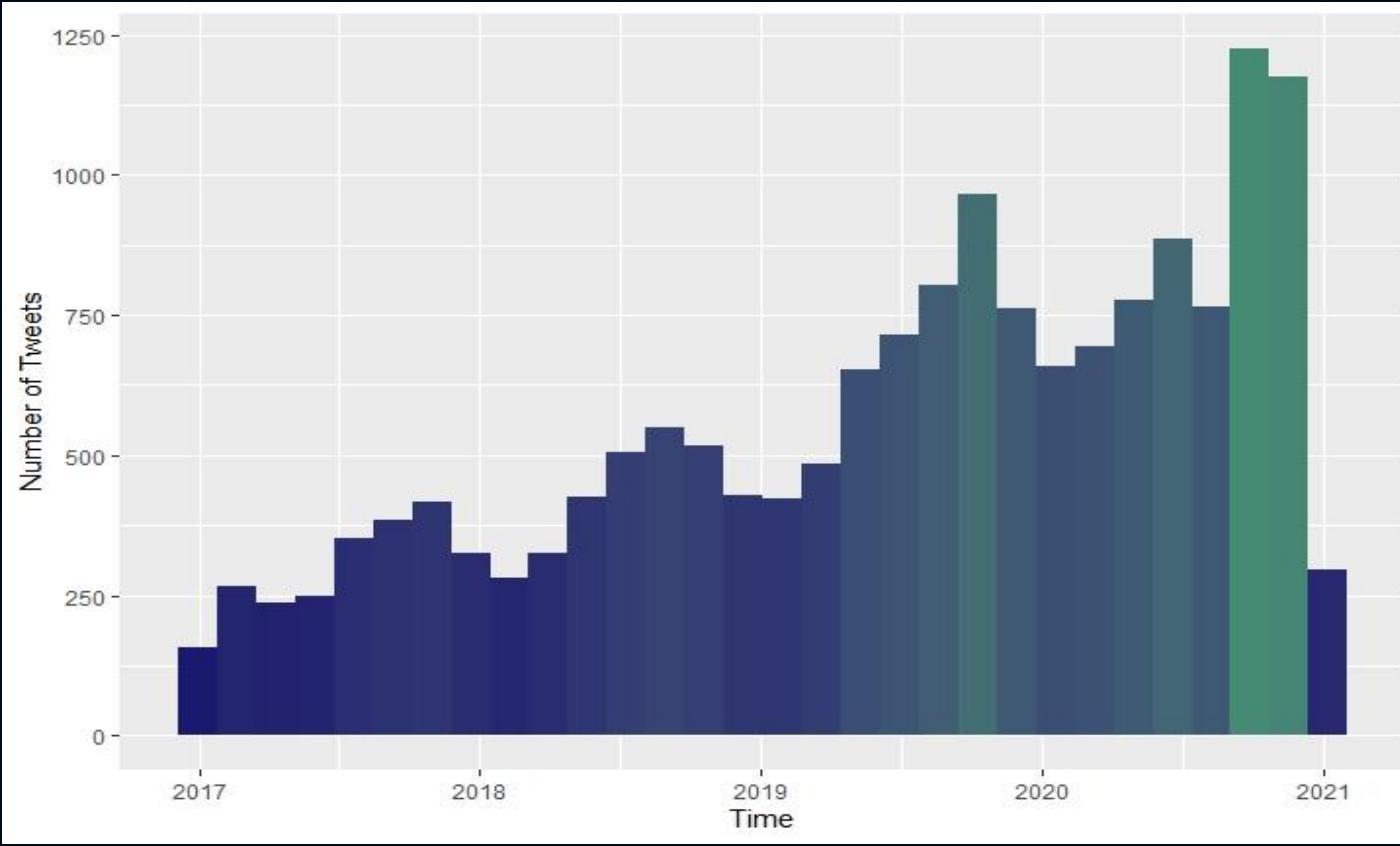
```
#Εισαγωγή και διαχείριση δεδομένων
tweets <- read.csv(choose.files(),
stringsAsFactors = FALSE)

tweets$date <- ymd_hms(tweets$date)
tweets$date <- with_tz(tweets$date,
"America/Chicago")

f_tweets <- tweets %>% filter(date >=
"2017-01-01", isRetweet == 'f')
```

# Εξερευνώντας τα δεδομένα

```
#Πράσινα συνολικών tweet ανά τον χρόνο
ggplot(data = f_tweets, aes(x = date)) +
  geom_histogram(aes(fill = ..count..)) +
  theme(legend.position = "none") +
  xlab("Time") + ylab("Number of tweets") +
  scale_fill_gradient(low = "midnightblue", high
= "aquamarine4")
```



# Καθαρισμός κειμένου και δημιουργία Corpus

```
#Καθαρισμός Κειμένου
nohandles <- str_replace_all(f_tweets$text,
"@\\w+", "")
wordCorpus <- Corpus(VectorSource(nohandles))
wordCorpus <- tm_map(wordCorpus,
removePunctuation)
wordCorpus <- tm_map(wordCorpus,
content_transformer(tolower))
wordCorpus <- tm_map(wordCorpus, removeWords,
stopwords("english"))
wordCorpus <- tm_map(wordCorpus, removeWords,
c("amp", "4yo", "3yo", "2yo"))
wordCorpus <- tm_map(wordCorpus, stripWhitespace)

removeURL <- function(x)
gsub('https[:alnum:]*' , '' , x)
wordCorpus <- tm_map(wordCorpus,
content_transformer(removeURL))
```

# Συνεχίζοντας την εξερεύνηση

```
#Δημιουργία wordcloud
pal <- brewer.pal(9,"BuPu")
pal <- pal[-(1:4)]

wordcloud(words = wordCorpus, scale=c(5,0.3),
max.words=100, random.order=FALSE,
          rot.per=0.35, use.r.layout=FALSE,
colors = pal)
```

democrat  
republican  
house  
can  
going  
back  
deal  
congress  
military  
biden  
job  
usa  
two got  
jobs  
usa  
american  
donâ€¢<sup>TMT</sup>  
working  
democrats  
states  
new  
thank  
total  
last  
must  
president  
many  
america  
american  
see  
ever  
wall  
north  
nothing  
party  
maga  
security  
first  
long  
day  
done  
border  
country  
economy  
dems  
state  
work  
republicans  
history  
china  
left  
Want  
media  
election  
years  
vote  
news  
great  
will  
just  
now  
big  
law  
getway  
fake  
people  
make  
time  
like  
strong  
russia  
world  
witch  
good  
one  
joe  
really  
complete  
made  
said  
even  
trump  
far  
look  
better  
bad  
today  
much  
also  
know  
hard  
well  
crime  
ever

# Δημιουργία πίνακα συναισθημάτων με το λεξικό NRC και προετοιμασία δεδομένων για ανάλυση

```
#Δημιουργία Πίνακα Συναισθημάτων
mySentiment <- get_nrc_sentiment(f_tweets$text)
fs_tweets <- cbind(f_tweets, mySentiment)

#Προετοιμασία δεδομένων
sentimentTotals <-
  data.frame(colSums(fs_tweets[,c(10:19)]))
names(sentimentTotals) <- "count"
sentimentTotals <- cbind("sentiment" =
  rownames(sentimentTotals), sentimentTotals)
rownames(sentimentTotals) <- NULL
```

	sentiment	count
1	anger	10131
2	anticipation	11094
3	disgust	6737
4	fear	10275
5	joy	8664
6	sadness	8572
7	surprise	6270
8	trust	18465
9	negative	19143
10	positive	24327

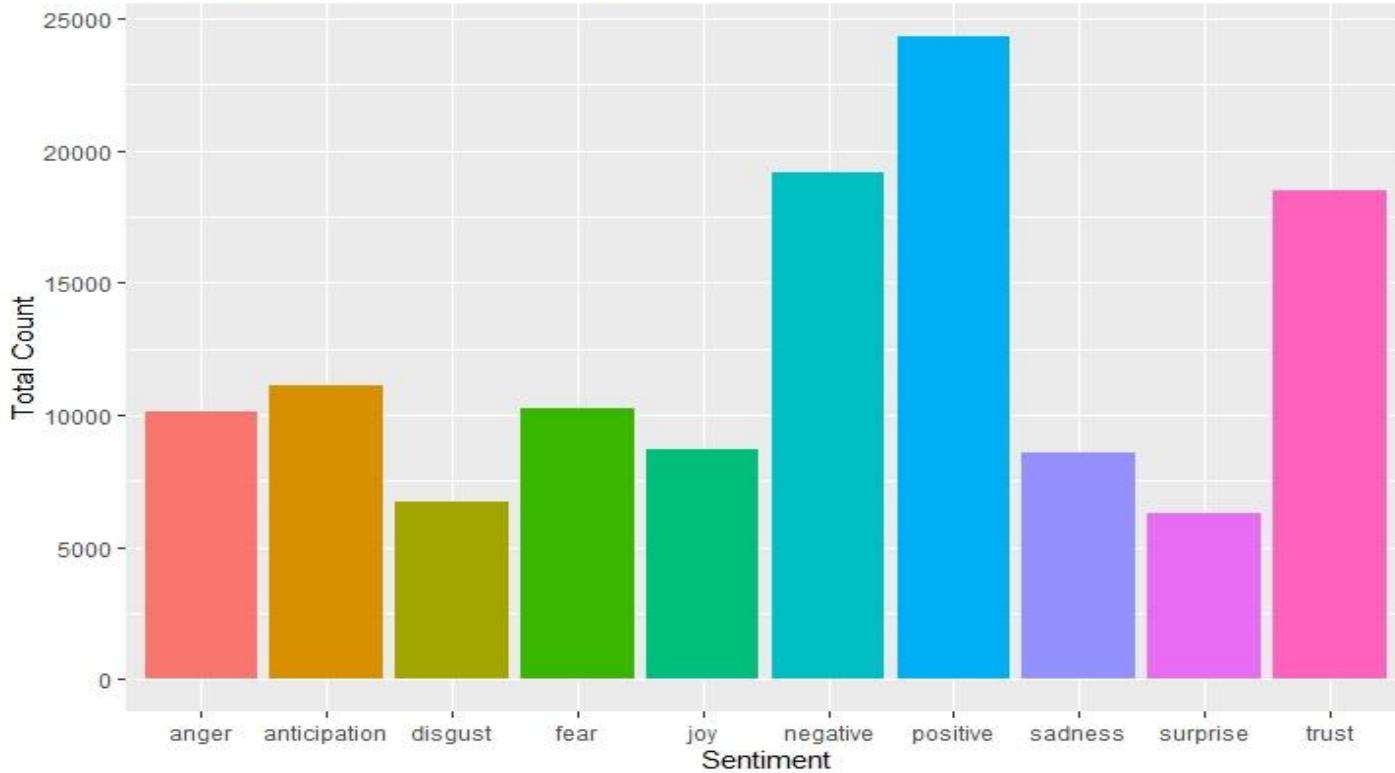
Filter

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive												
1	0	2	0	0	4	0	2	3	0	4												
2	3	2	1	3	0	2	1	3	4	5												
3	0	0	0	0	0	0	0	0	0	0												
4	0	0	0	0	0	0	0	0	0	0												
5	0	0	0	0	0	0	0	0	0	0												
6	0	0	0	0	0	0	0	0	0	0												
7	0	0	0	0	0	0	0	0	0	0												
8	0	0	0	0	0	0	0	0	0	0												
9	0	0	0	0	0	0	0	0	0	0												
10	0	0	0	0	0	0	0	0	0	0												
11	0	0	0	0	0	0	0	0	0	0												
12	0	1	0	1	1	1	1	1	1	1												
	id	text	isRetweet	isDeleted	device	favorites	retweets	date	isFlagged	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative				
13	0	1	0	0	1	1	73748	17404	2020-03-02 19:34:50	f	0	2	0	0	4	0	2	3				
14	0	0	0	0	0	2	1304875e+18	The Unsolicited ...	f	Twitter for iPhone	80527	23502	2020-09-12 15:10:58	f	3	2	1	3	0	2	1	3
15	0	0	0	0	0	3	1.223641e+18	Getting a little e...	f	Twitter for iPhone	285863	30209	2020-02-01 10:14:02	f	0	0	0	0	0	0	0	0
16	0	0	0	0	0	4	1.319502e+18	https://t.co/qwq...	f	Twitter for iPhone	130822	19127	2020-10-22 23:52:14	f	0	0	0	0	0	0	0	0
17	0	0	0	0	0	5	1.319501e+18	https://t.co/lEu...	f	Twitter for iPhone	153446	20275	2020-10-22 23:46:53	f	0	0	0	0	0	0	0	0
18	0	3	0	0	1	6	1.319501e+18	https://t.co/5Cr...	f	Twitter for iPhone	102150	14815	2020-10-22 23:46:49	f	0	0	0	0	0	0	0	0
19	1	0	1	1	0	7	1.319500e+18	https://t.co/tQC...	f	Twitter for iPhone	255804	35594	2020-10-22 23:46:45	f	0	0	0	0	0	0	0	0
20	0	0	0	0	0	8	1.215248e+18	Thank you Elise!...	f	Twitter for iPhone	48510	11608	2020-01-09 06:24:31	f	0	0	0	0	0	0	0	0
21	0	0	0	0	0	9	1.319491e+18	As per your requ...	f	Twitter for iPhone	253761	79855	2020-10-22 23:09:59	f	0	0	0	0	0	0	0	0
22	0	0	0	0	0	10	1.345959e+18	https://t.co/Nmd...	f	Twitter for iPhone	95300	28318	2021-01-03 23:02:51	f	0	0	0	0	0	0	0	0
						11	1.325885e+18	The threshold id...	f	Twitter for iPhone	493076	100609	2020-11-09 13:36:26	f	0	0	0	0	0	0	0	0
						12	1.315780e+18	â€œlâ€™m runn...	f	Twitter for iPhone	142084	32953	2020-10-12 17:22:39	f	0	1	0	1	1	1	0	3
						13	1.319684e+18	HUGE win today ...	f	Twitter for iPhone	215994	51830	2020-10-23 11:55:29	f	0	1	0	0	1	0	0	4
						14	1.319656e+18	Thank you Megy...	f	Twitter for iPhone	178163	24864	2020-10-23 10:04:10	f	0	0	0	0	0	0	0	0
						15	1.319511e+18	11 DAYS! #MAG...	f	Twitter for iPhone	197604	49800	2020-10-23 00:26:41	f	0	0	0	0	0	0	0	0
						16	1.325890e+18	Nevada is turnin...	f	Twitter for iPhone	363489	78378	2020-11-09 13:54:33	t	0	0	0	0	0	0	0	0
						17	1.345967e+18	https://t.co/Ezkc...	f	Twitter for iPhone	83455	21427	2021-01-03 23:34:06	f	0	0	0	0	0	0	0	0
						18	1.325891e+18	Wisconsin is look...	f	Twitter for iPhone	347994	61006	2020-11-09 14:02:19	f	0	3	0	0	1	0	1	1
						19	1.320096e+18	Nobody is showi...	f	Twitter for iPhone	195132	33268	2020-10-24 15:11:38	f	1	0	1	1	0	1	0	0
						20	1.345967e+18	https://t.co/2rJf...	f	Twitter for iPhone	69095	19024	2021-01-03 23:35:31	f	0	0	0	0	0	0	0	0
						21	1.345967e+18	https://t.co/uGfz...	f	Twitter for iPhone	99420	28624	2021-01-03 23:35:08	f	0	0	0	0	0	0	0	0
						22	1.345969e+18	https://t.co/PXN...	f	Twitter for iPhone	105704	27741	2021-01-03 23:44:45	f	0	0	0	0	0	0	0	0
						23	1.325895e+18	Pennsylvania pre...	f	Twitter for iPhone	479292	81458	2020-11-09 14:17:47	t	2	1	2	2	0	1	0	2
						24	1.320096e+18	Just landed in Oh...	f	Twitter for iPhone	139153	14000	2020-10-24 15:13:08	f	0	0	0	0	0	0	0	0
						25	1.325896e+18	Georgia will be a...	f	Twitter for iPhone	637719	94570	2020-11-09 14:21:43	f	0	0	0	0	0	0	0	0
						26	1.291061e+18	https://t.co/NX5z...	f	Twitter for iPhone	28711	8085	2020-08-05 12:20:01	f	0	0	0	0	0	0	0	0

# Δημιουργία γραφήματος συνολικής βαθμολογίας συναισθήματος για όλα τα tweets

```
#Πράσινα συνολικού σκορ συναισθήματος για όλα τα
tweets
ggplot(data = sentimentTotals, aes(x = sentiment,
y = count)) +
  geom_bar(aes(fill = sentiment), stat =
"identity") +
  theme(legend.position = "none") +
  xlab("Sentiment") + ylab("Total Count") +
  ggtitle("Total Sentiment Score for All Tweets")
```

### Total Sentiment Score for All Tweets



# Ας δούμε αν το συναίσθημα των tweet άλλαξε με την πάροδο του χρόνου

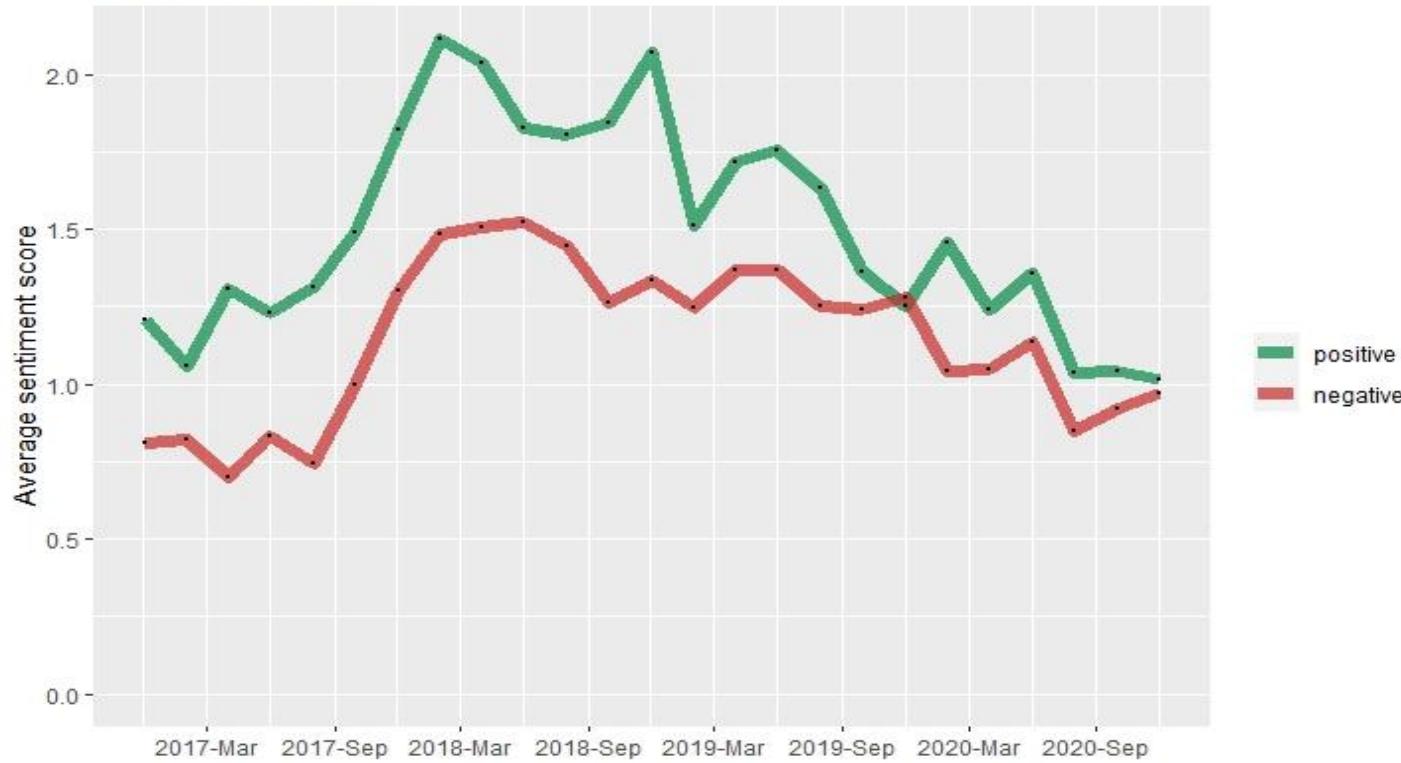
```
#Μέσοι όροι θετικών και αρνητικών βαθμολογιών  
posnegtime <- fs_tweets %>%  
  group_by(date = cut(date, breaks="2 months"))  
%>%  
  summarise(negative = mean(negative),  
            positive = mean(positive)) %>% melt  
names(posnegtime) <- c("date", "sentiment",  
"meanvalue")  
posnegtime$sentiment =  
factor(posnegtime$sentiment, levels(posnegtime$sen-  
timent)[c(2,1)])
```

	date	sentiment	meanvalue
18	2019-10-01	negative	1.2421566
19	2019-12-01	negative	1.2793946
20	2020-02-01	negative	1.0414201
21	2020-04-01	negative	1.0475687
22	2020-06-01	negative	1.1357063
23	2020-08-01	negative	0.8463902
24	2020-10-01	negative	0.9177989
25	2020-12-01	negative	0.9729207
26	2016-12-01	positive	1.2077295
27	2017-02-01	positive	1.0561404
28	2017-04-01	positive	1.3063380
29	2017-06-01	positive	1.2288557
30	2017-08-01	positive	1.3125000
31	2017-10-01	positive	1.4878543
32	2017-12-01	positive	1.8202247
33	2018-02-01	positive	2.1121212

# Δημιουργία γραφήματος με βάση τον πίνακα των μέσων όρων

```
#Δημιουργία γραφήματος Συναίσθημα ανά τον χρόνο
ggplot(data = posnegtime, aes(x = as.Date(date),
y = meanvalue, group = sentiment)) +
  geom_line(size = 2.5, alpha = 0.7, aes(color =
sentiment)) +
  geom_point(size = 0.5) +
  ylim(0, NA) +
  scale_colour_manual(values = c("springgreen4",
"firebrick3")) +
  theme(legend.title=element_blank(),
axis.title.x = element_blank()) +
  scale_x_date(breaks = date_breaks("6 months"),
            labels = date_format("%Y-%b")) +
  ylab("Average sentiment score") +
  ggtitle("Sentiment Over Time")
```

## Sentiment Over Time



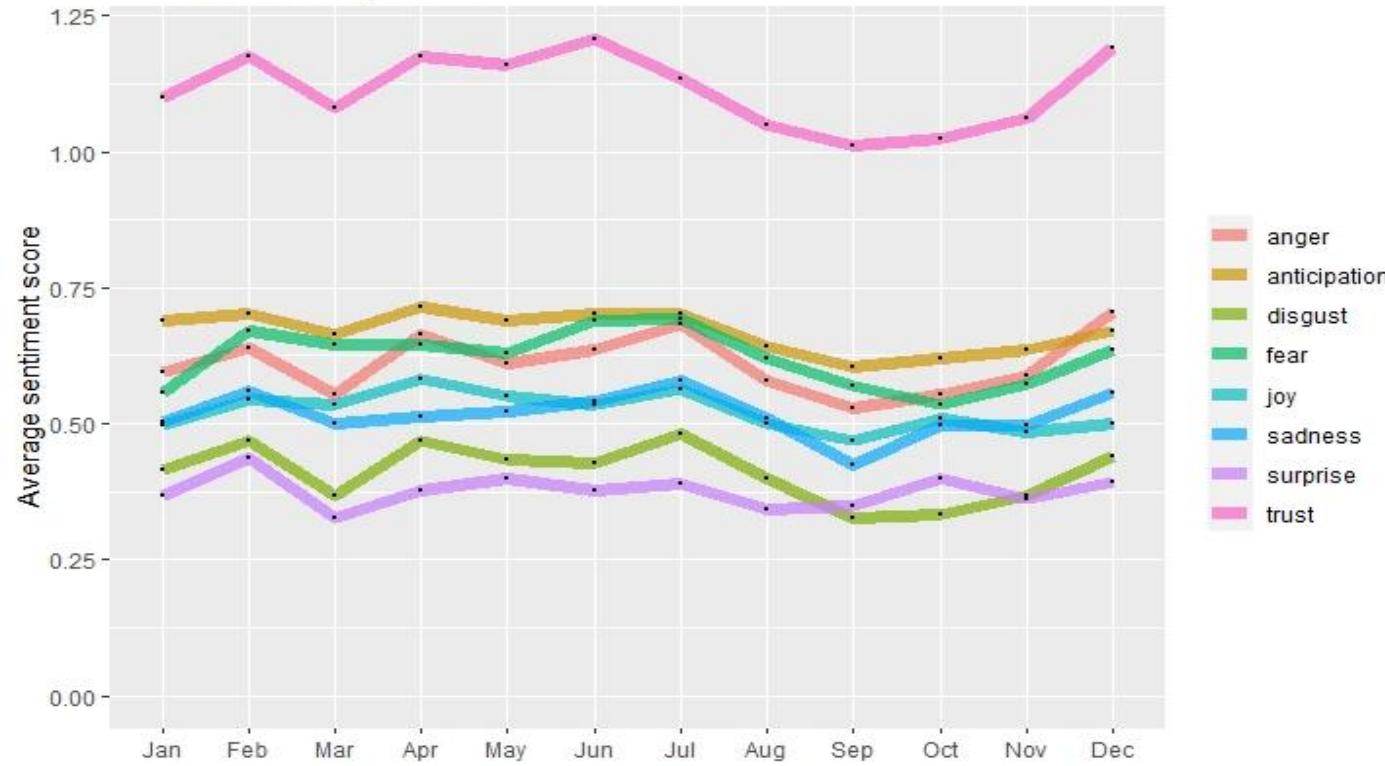
# Δημιουργία μέσων όρων για κάθε συναίσθημα του λεξικού NRC

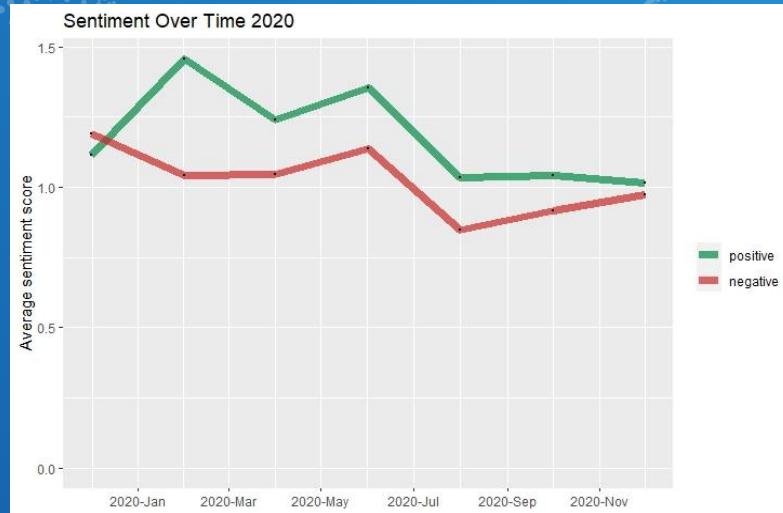
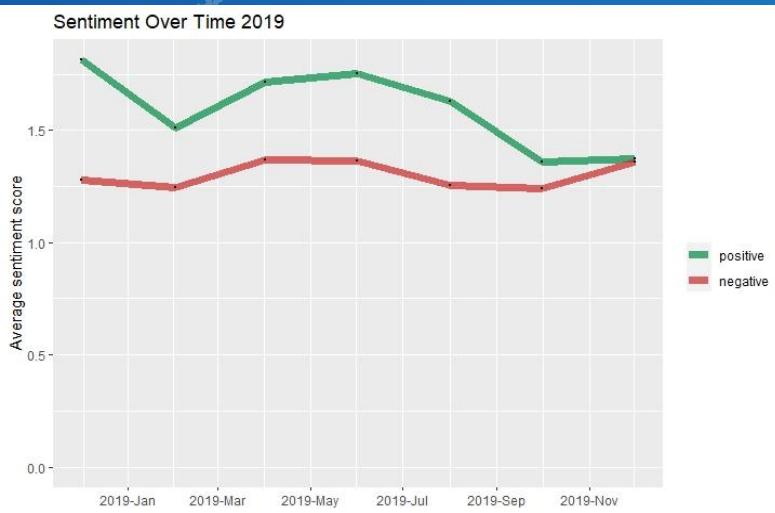
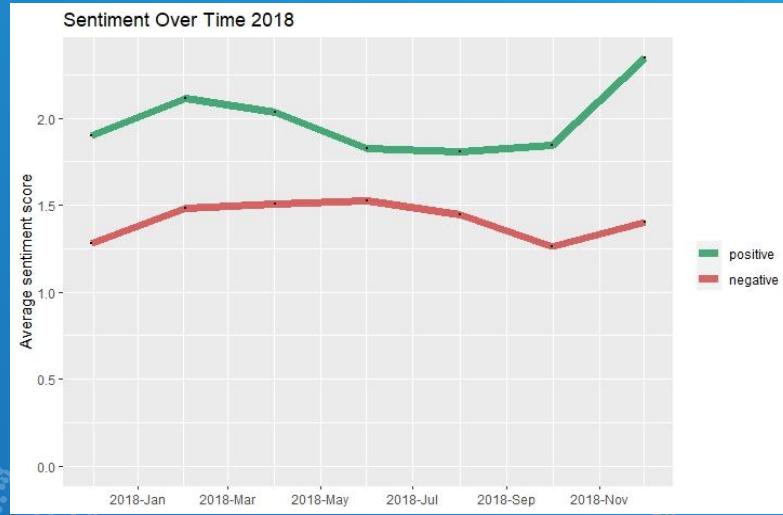
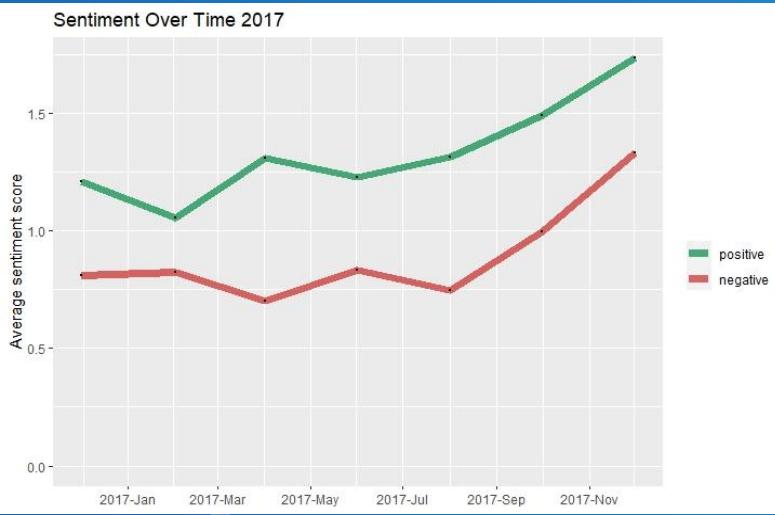
```
#Προετοιμασία για ανάλυση συναίσθημάτων μέσα σε
κάθε χρονιά
fs_tweets$month <- month(fs_tweets$date, label =
TRUE)
monthlysentiment <- fs_tweets %>% group_by(month)
%>%
  summarise(anger = mean(anger),
            anticipation = mean(anticipation),
            disgust = mean(disgust),
            fear = mean(fear),
            joy = mean(joy),
            sadness = mean(sadness),
            surprise = mean(surprise),
            trust = mean(trust)) %>% melt
names(monthlysentiment) <- c("month",
"sentiment", "meanvalue")
```

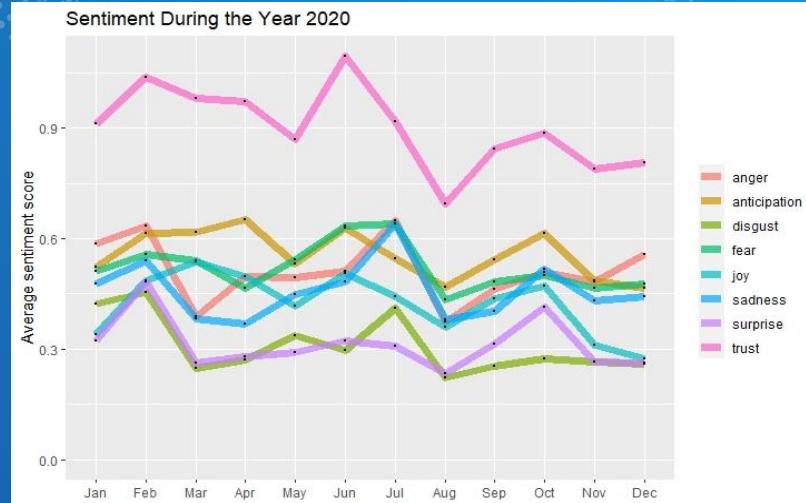
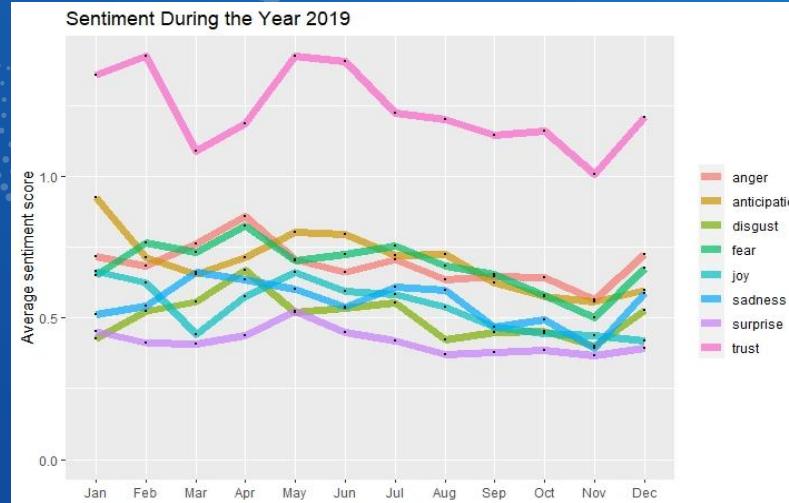
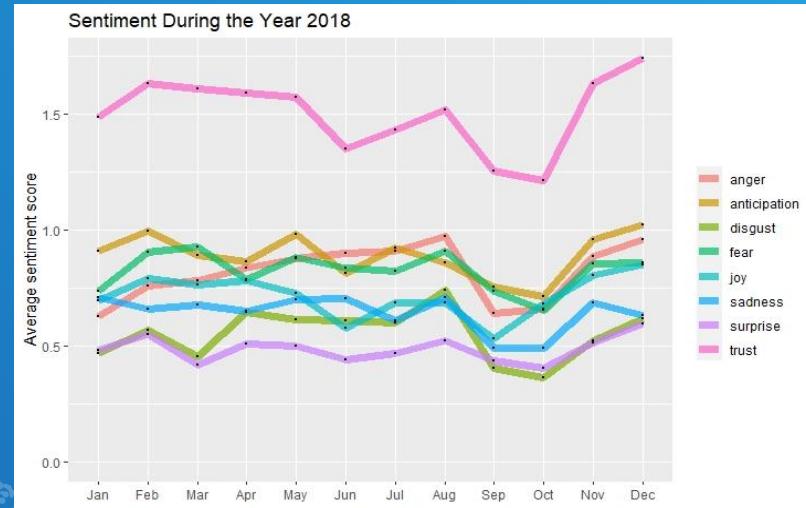
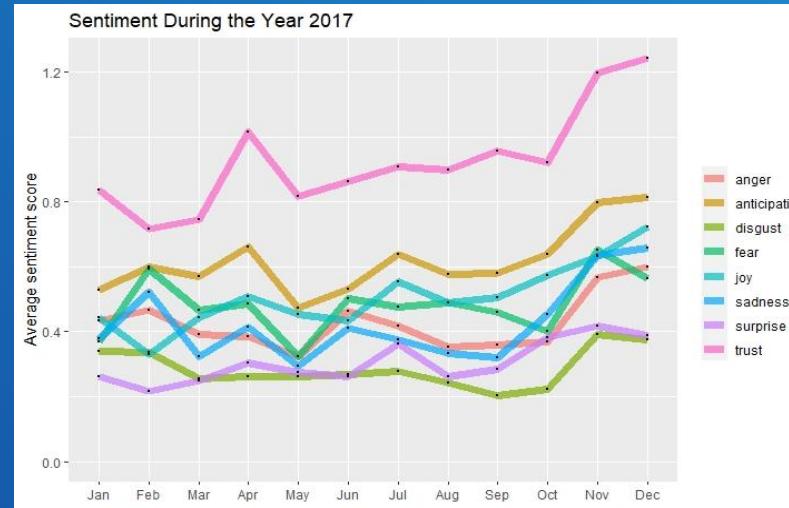
# Δημιουργία γραφήματος συναισθημάτων κατά τη διάρκεια ενός έτους

```
#Δημιουργία γραφήματος συναισθημάτων κατά τη διάρκεια ενός χρόνου
ggplot(data = monthlysentiment, aes(x = month, y = meanvalue, group = sentiment)) +
  geom_line(size = 2.5, alpha = 0.7, aes(color = sentiment)) +
  geom_point(size = 0.5) +
  ylim(0, NA) +
  theme(legend.title=element_blank(),
axis.title.x = element_blank()) +
  ylab("Average sentiment score") +
  ggtitle("Sentiment During the Year")
```

## Sentiment During the Year







### #Εισαγωγή Γαλετών

```
library(ggplot2)
library(lubridate)
library(scales)
library(tm)
library(stringr)
library(wordcloud)
library(syuzhet)
library(reshape2)
library(dplyr)
```

### #Εισαγωγή και διαχείριση δεδομένων

```
tweets <- read.csv(choose.files(), stringsAsFactors = FALSE)
tweets$date <- ymd_hms(tweets$date)
tweets$date <- with_tz(tweets$date, "America/Chicago")
f_tweets <- tweets %>% filter(date >= "2017-01-01" & date <= "2021-01-07", isRetweet == 'f')
```

### #Γραφικά συνολικών tweets

```
ggplot(data = f_tweets, aes(x = date)) +
  geom_histogram(aes(fill = ..count..)) +
  theme(legend.position = "none") +
  xlab("Time") + ylab("Number of Tweets") +
  scale_fill_gradient(low = "midnightblue", high = "aquamarine4")
```

### #Καθαρισμός Κειμένου

```
nohandles <- str_replace_all(f_tweets$text, "@\\w+", "")
wordCorpus <- Corpus(VectorSource(nohandles))
wordCorpus <- tm_map(wordCorpus, removePunctuation)
wordCorpus <- tm_map(wordCorpus, content_transformer(tolower))
wordCorpus <- tm_map(wordCorpus, removeWords, stopwords("english"))
wordCorpus <- tm_map(wordCorpus, removeWords, c("amp","4yo","3yo","2yo"))
wordCorpus <- tm_map(wordCorpus, stripWhitespace)

removeURL <- function(x) gsub('https[[:alnum:]]*', "", x)
wordCorpus <- tm_map(wordCorpus, content_transformer(removeURL))
```

### #Δημιουργία wordcloud

```
pal <- brewer.pal(9,"BuPu")
pal <- pal[-(1:4)]
```

```
wordcloud(words = wordCorpus, scale=c(5,0.3), max.words=100, random.order=FALSE,
          rot.per=0.35, use.r.layout=False, colors = pal)
```

### #Δημιουργία Πίνακα Συνολικών πόνω στα tweets

```
mySentiment <- get_nrc_sentiment(f_tweets$text)
fs_tweets <- cbind(f_tweets, mySentiment)
```

### #Σύνολα

```
sentimentTotals <- data.frame(colSums(fs_tweets[,c(10:19)]))
names(sentimentTotals) <- "count"
sentimentTotals <- cbind("sentiment" = rownames(sentimentTotals), sentimentTotals)
rownames(sentimentTotals) <- NULL
```

### #Γραφικά συνολικών αποτελεσμάτων για όλα τα tweets

```
ggplot(data = sentimentTotals, aes(x = sentiment, y = count)) +
  geom_bar(aes(fill = sentiment), stat = "identity") +
  theme(legend.position = "none") +
  xlab("Sentiment") + ylab("Total Count") + ggtitle("Total Sentiment Score for All Tweets")
```

### #Μέσοι χρονικών και εργατικών βαθμολογιών

```
posnegtime <- fs_tweets %>%
  group_by(date = cut(date, breaks="2 months")) %>%
  summarise(negative = mean(negative),
            positive = mean(positive)) %>% melt
names(posnegtime) <- c("date", "sentiment", "meanvalue")
posnegtime$sentiment = factor(posnegtime$sentiment,levels=posnegtime$sentiment)[c(2,1)])
```

### #Δημιουργία γραφικών Σύνολης για τον χρόνο

```
ggplot(data = posnegtime, aes(x = as.Date(date), y = meanvalue, group = sentiment)) +
  geom_line(size = 2.5, alpha = 0.7, aes(color = sentiment)) +
  geom_point(size = 0.5) +
  ylim(0, NA) +
  scale_colour_manual(values = c("springgreen4", "firebrick3")) +
  theme(legend.title=element_blank(), axis.title.x = element_blank()) +
  scale_x_date(breaks = date_breaks("2 months"),
               labels = date_format("%Y-%b")) +
  ylab("Average sentiment score") +
  ggtitle("Sentiment Over Time")
```

### #Προς ορθορία δεδομένων για την διάσταση συνολικών μέσω σε κάθε χρονιά

```
fs_tweets$month <- month(fs_tweets$date, label = TRUE)
monthlysentiment <- fs_tweets %>% group_by(month) %>%
  summarise(anger = mean(anger),
            anticipation = mean(anticipation),
            disgust = mean(disgust),
            fear = mean(fear),
            joy = mean(joy),
            sadness = mean(sadness),
            surprise = mean(surprise),
            trust = mean(trust)) %>% melt
names(monthlysentiment) <- c("month", "sentiment", "meanvalue")
```

### #Δημιουργία γραφικών συνολικών στη διάρκεια ενός χρόνου

```
ggplot(data = monthlysentiment, aes(x = month, y = meanvalue, group = sentiment)) +
  geom_line(size = 2.5, alpha = 0.7, aes(color = sentiment)) +
  geom_point(size = 0.5) +
  ylim(0, NA) +
  theme(legend.title=element_blank(), axis.title.x = element_blank()) +
  ylab("Average sentiment score") +
  ggtitle("Sentiment During the Year")
```

# Πηγές

- [Joy to the World, and also Anticipation, Disgust, Surprise... | Julia Silge](#)
- [Ten Thousand Tweets | Julia Silge](#)
- [\(1825\) R - Sentiment Analysis and Wordcloud with R from Twitter Data | Example using Apple Tweets - YouTube](#)
- [The Comprehensive R Archive Network \(r-project.org\)](#)
- [886.pdf \(ntua.gr\)](#)
- [Stack Overflow - Where Developers Learn, Share, & Build Careers](#)
- [ITA - Search \(thetrumparchive.com\)](#)

# Ευχαριστώ

## Για την προσοχή σας

“It is very hard for them to attack me on looks, because i am so good looking”

**Donald Trump**

August 7, 2015

