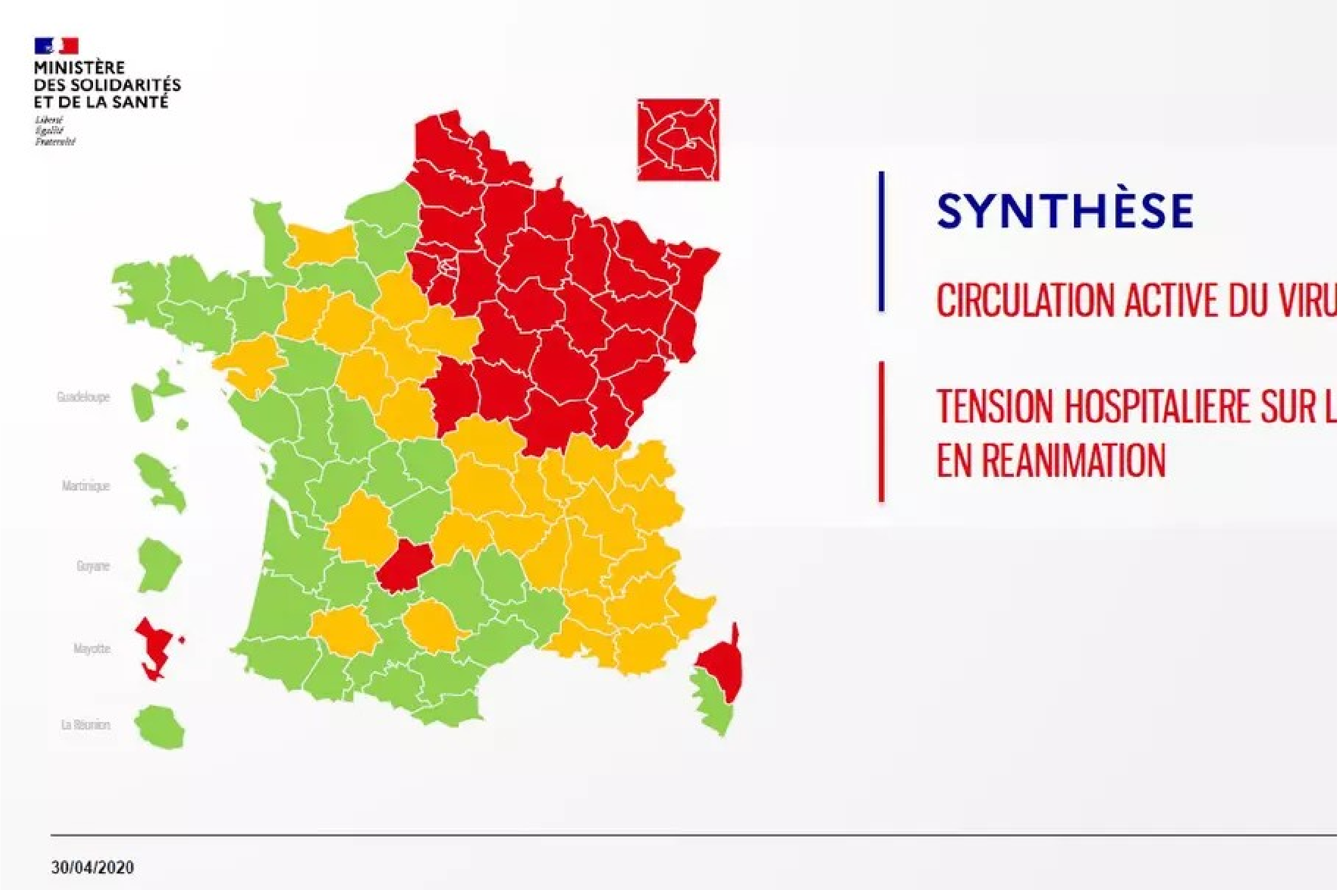
TwitterProjectt

Steve CHEMI

4/28/2020

Introduction

All the departments of France are affected differently by the epidemic. Yesterday the government communicated by a map of the territory in three colours: green, orange or red depending on the circulation of the virus and the number of hospitalizations in each department. Every day this map allows to visualize the evolution of the situation in view of the deconfinement. With 289 deaths yesterday in France, the decline of the epidemic is confirmed in hospital services. The results in Italy, Spain, USA and worldwide, map of the most affected departments, deconfinement measures, mask prices capped, this is the most debated subject in France, which is why we have chosen to conduct this study to find out what the French think about this pandemic.



MINISTÈRE DES SOLIDARITÉS ET DE LA SANTÉ
Liberté Égalité Fraternité

SYNTHÈSE

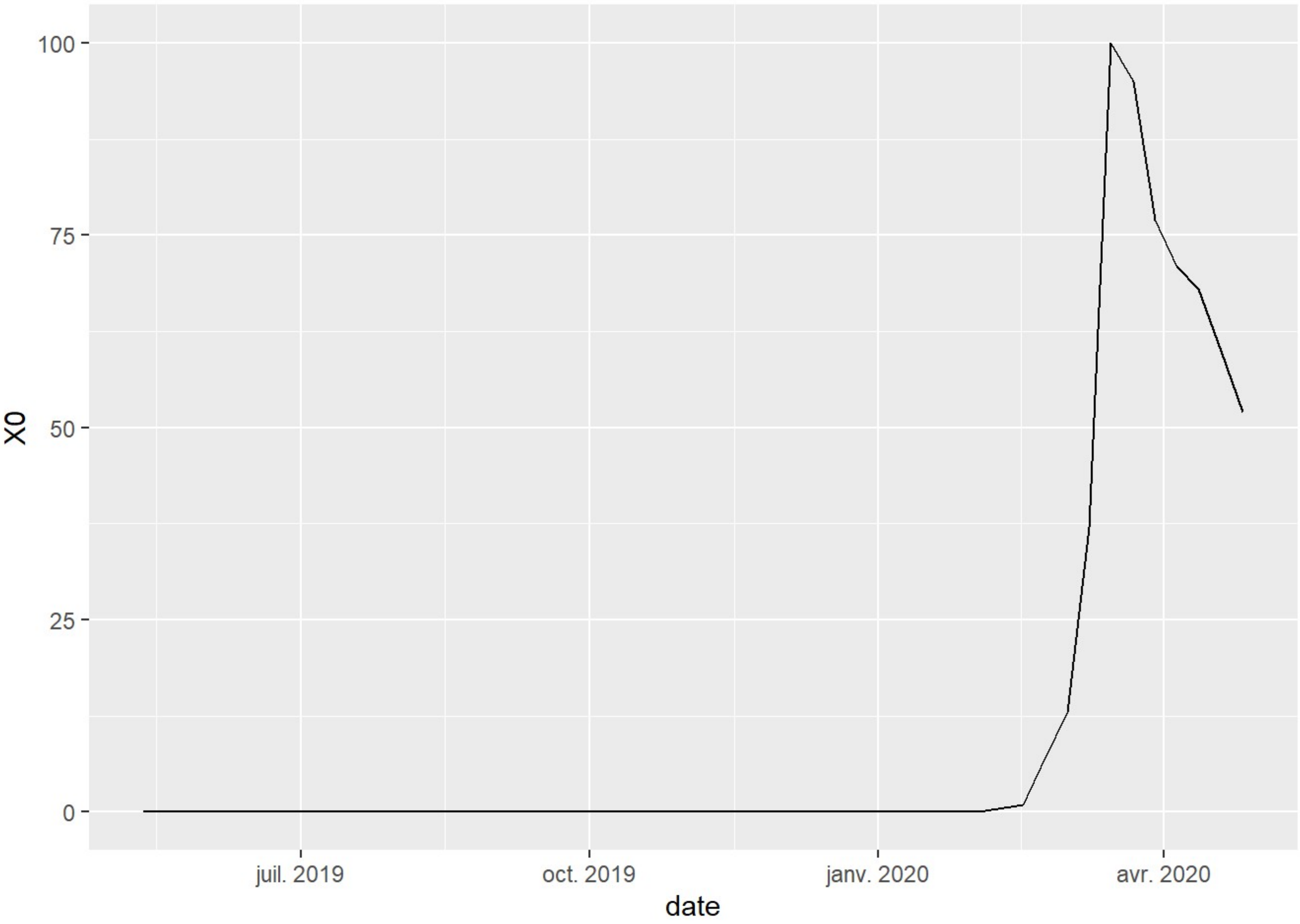
CIRCULATION ACTIVE DU VIRU

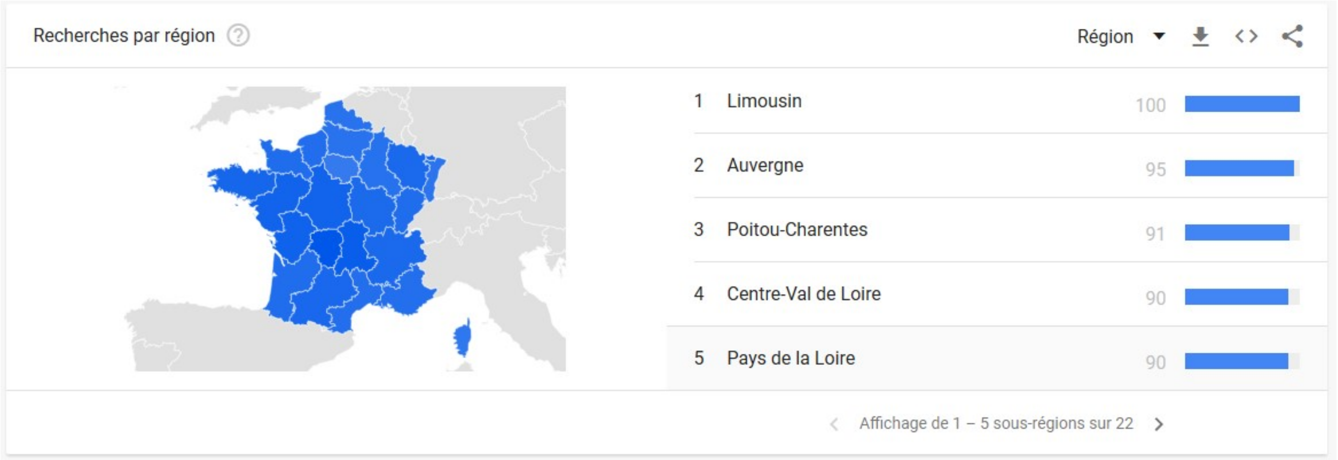
TENSION HOSPITALIERE SUR L
EN REANIMATION

30/04/2020

The evolution of googles recerches on the Netflix theme are given by the following timeline:

```
recherches = read.csv('C:/Users/steve/Downloads/multiTimeline.csv')
setDT(recherches)
recherches$date <- as.Date(recherches$X2019.05.05)
ggplot(recherches, aes(x=date, y=X0)) + geom_line()
```





Recherches par région

1	Limousin	100
2	Auvergne	95
3	Poitou-Charentes	91
4	Centre-Val de Loire	90
5	Pays de la Loire	90

Affichage de 1 – 5 sous-régions sur 22

Google trends data about covid19 in France

We thus note that the subject remains rather solicited by the Net surfers.

Decription:

Our project is about analysing people’s feeling about Covid-19. By using everyday social networks we could see that Covid was in all the debates. We wanted to have a different point of view about it, a golbal look over the situation. Here’s the list of the libraries we used to best exploit this Datastet

First step with the environment

Here, we imported our environment with all the Tweets we downloaded before. We used the notebook to have a better control on our chunks and to better comment each step of the process.

```
load("C:/Users/steve/OneDrive/Documents/R/NLP/EnvTwitt3004.RData")
```

Ploting a histogram to see the amount of data each day, each minute:

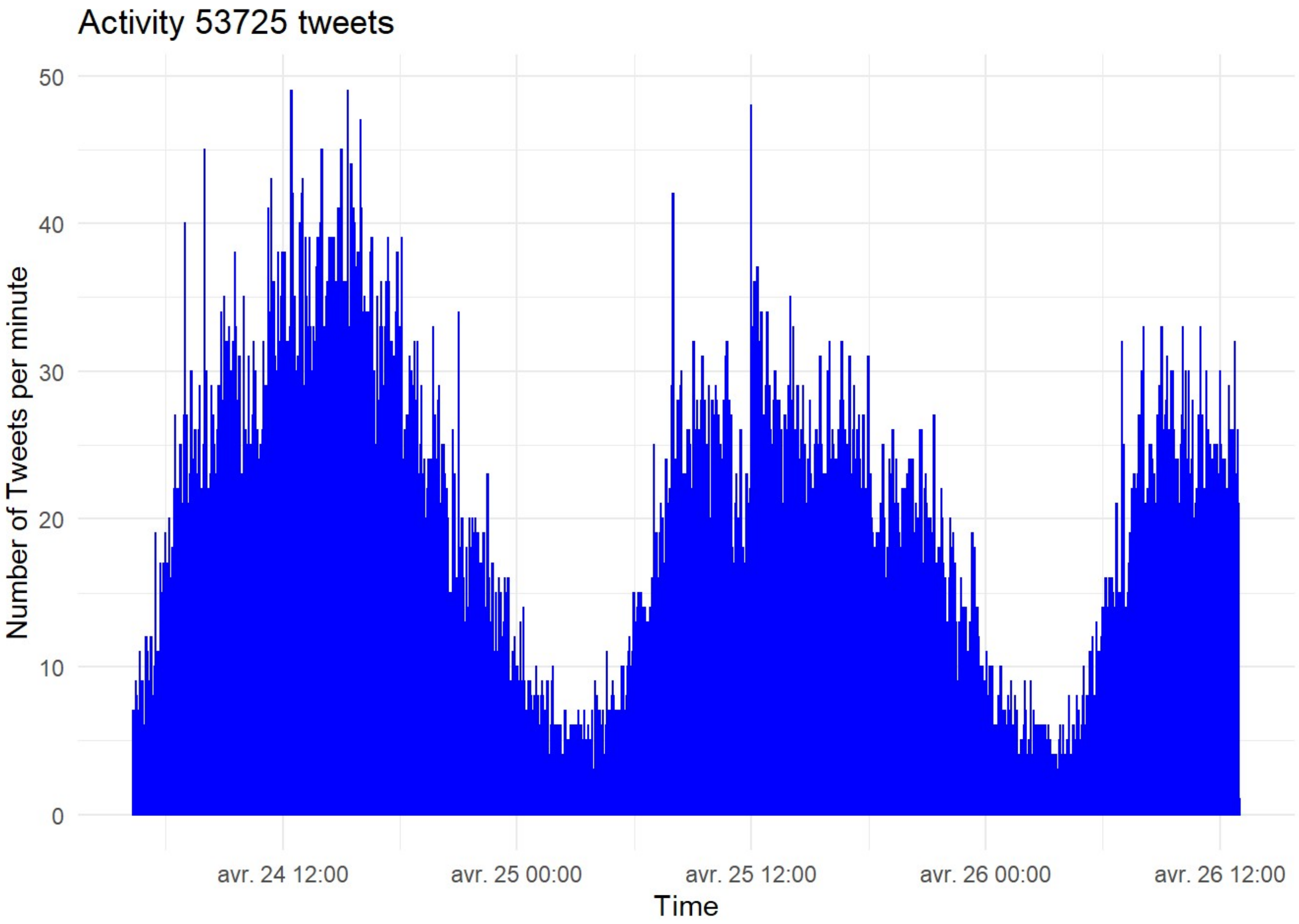
First important and useful step to clearly visualize the data distribution overtime (peak and off-peak hours).

```
ggplot(tweets.df, aes(x=created_at)) +
  geom_histogram(aes(y=..count..), #make histogram
                 binwidth=60, #each bar contains number of tweets during 60 s
                 colour="blue", #colour of frame of bars
                 fill="blue", #fill colour for bars
                 alpha=0.8) + # bars are semi transparant
  ggtitle(paste0("Activity ",number.of.tweets," tweets")) + #title
  scale_y_continuous(name="Number of Tweets per minute") +
  scale_x_datetime(name = "Time") +
  theme_minimal(base_family="Times New Roman")
```

```
## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): famille de
## police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): famille de
## police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): famille de
## police introuvable dans la base de données des polices Windows
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows
```

```
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x,
## x$y, : famille de police introuvable dans la base de données des polices
## Windows
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows
```



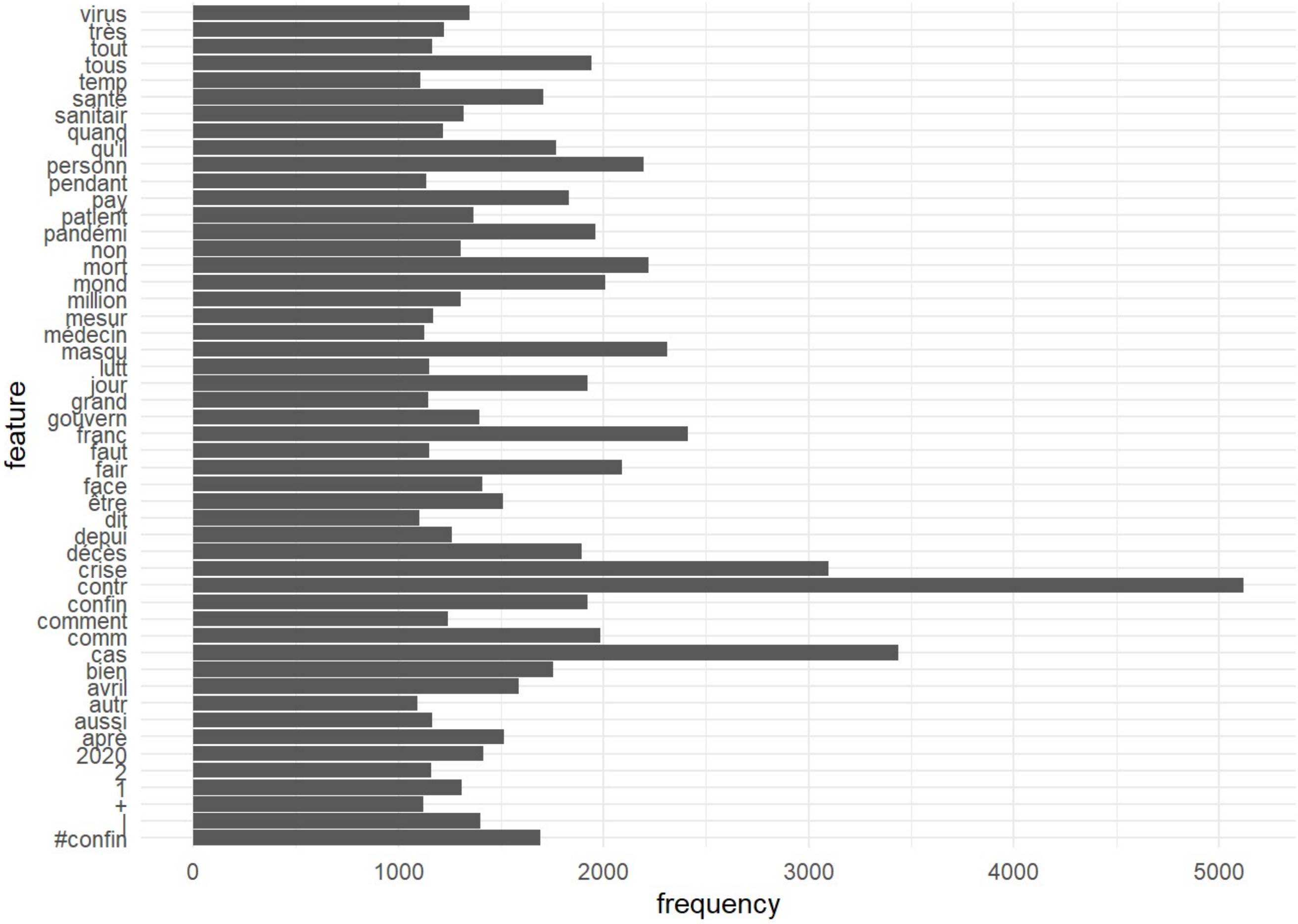
We can see that most of posts are shared between midday and 6PM (which is totally logical as it represents the peak hours of our daily life. The higher number of tweets we have is a bit less than 50 tweets/minute. The minimum we can notice is about a bit more than 5tweets/ minutes and it is on the middle of the night)

Word Frequency

On the previous step we plotted the data over time in a graph, now we want to better see the frequnce of words, not on time but for words themselves. It allows us to have a first sight on the feeling of people, what they think and express on internet.

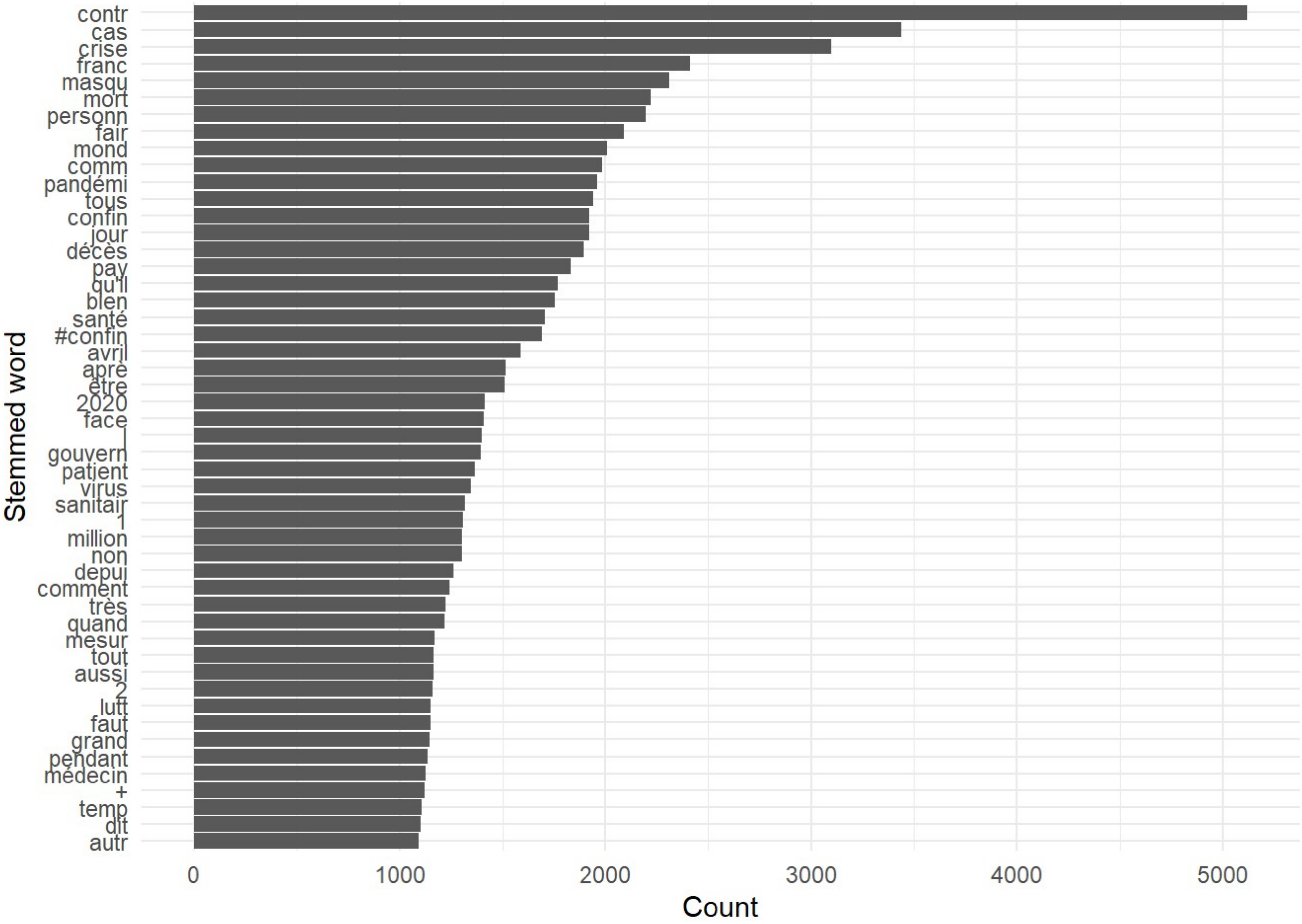
Most used words: contre (against) refering to everything against covid (hospitals, vaccine, quarantine etc) (more than 5K) cas (covid cases) about the virus evolution in France and over the world (alomst 3,5K) franc (beginning of France) all about Covid in France Masque (masks) people want informationon how to get masks, how to use and clean them Pandémie (pandemic) still about its evolution Gouvernement (government) people are also very interested about the governement decisions Décès (death) of course death toll is very important for people, horrible to say but it gives the rythm of the virus and its progression or regression

```
dfFreq <- textstat_frequency(dfmat_corp_twitter) %>% as.data.table
ggplot(dfFreq[1:50,], aes(x=feature, y=frequency)) +
  geom_col() +
  coord_flip() +
  theme_minimal()
```



```
ggplot(dfFreq[1:50,], aes(x=reorder(feature, -rank), y=frequency)) +
  geom_col() +
  coord_flip() +
  labs(x = "Stemmed word", y = "Count") +
  theme_minimal(base_family="Times New Roman")
```

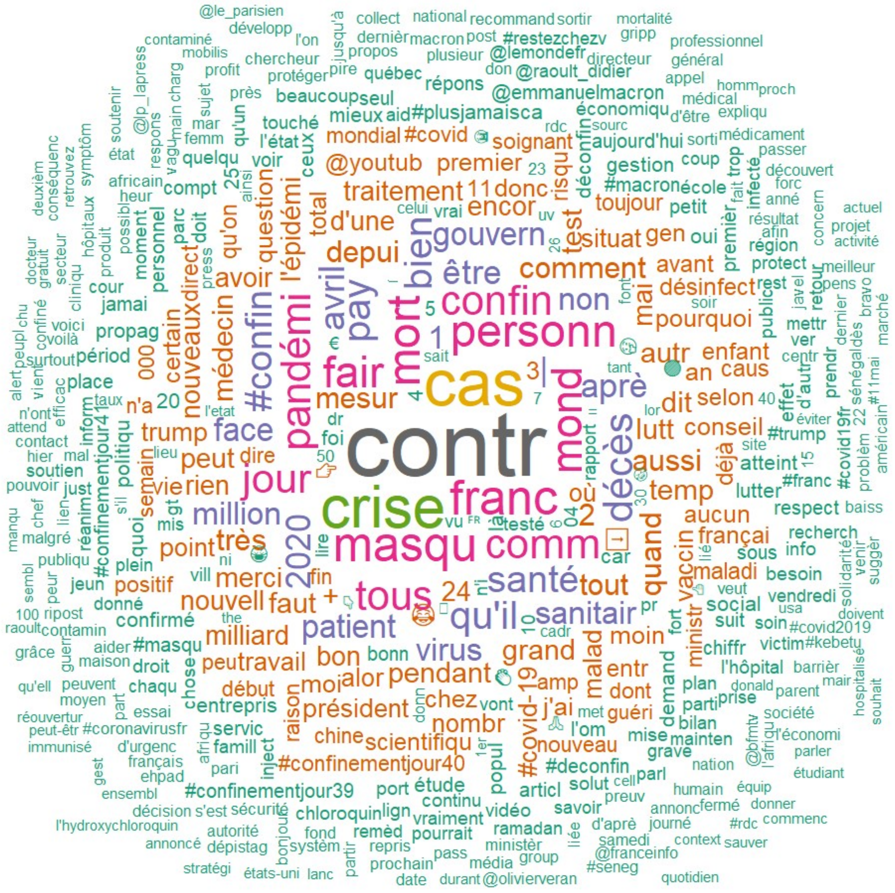
```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x,
## x$y, : famille de police introuvable dans la base de données des polices
## Windows
```



A more elegant but perhaps less useful way of showing the word-frequencies are with

It was relevant to propose a different data vizualisation about the word frequence, not only by using count but also playing on the size (contre is the most counted one, so the biggest here in the plot)

```
textplot_wordcloud(dfmat_corp_twitter, min_count = 6, random_order = FALSE,
                   rotation = .25,
                   color = RColorBrewer::brewer.pal(8, "Dark2"))
```

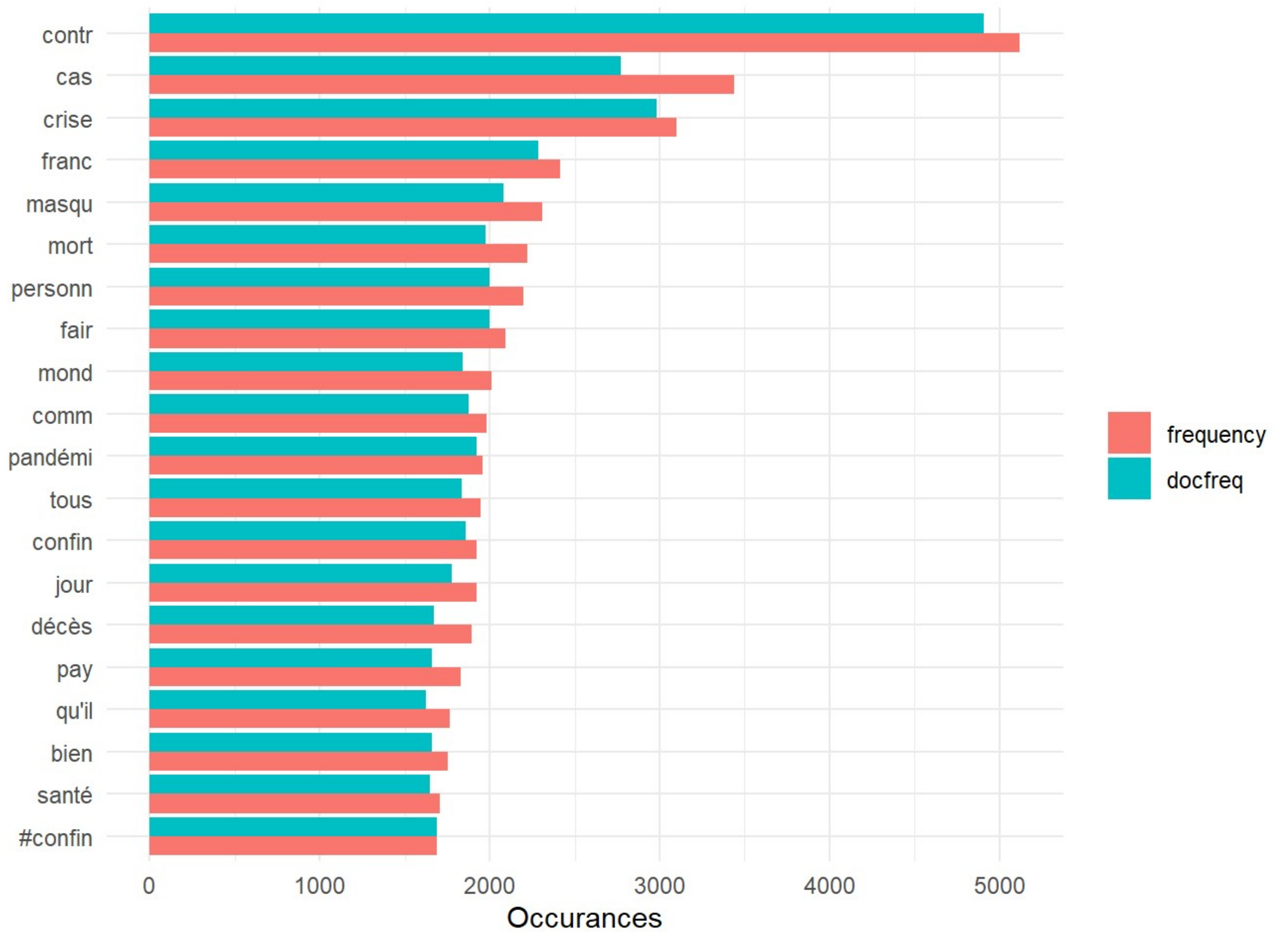


```
tweets.df[grepl("contr",text), list(text) ]
```

```
##
text
##    1:                                          Pour lutter contre la propagation
du virus COVID-19, un ensemble de mesures, informations et consignes officielles et
actualisées sont mis à la disposition des professionnels du secteur social et médic
o-social. https://t.co/RDf14co1x5 https://t.co/r3hkDomyLN
##    2:
Dominique Brown s’attaque à des propos tenus contre des aînés https://t.co/DJuOiQU6
PU https://t.co/83lDvrfJiy
##    3:
À Berlin, près d'un millier d'Allemands manifestent contre le confinement https://
t.co/JehyDzWyBe https://t.co/oTKPtHeKqw
##    4:
Dominique Brown s’attaque à des propos tenus contre des aînés https://t.co/InZsrDl3
aX https://t.co/pMJcYi4VCX
##    5:
À Berlin, près d'un millier d'Allemands manifestent contre le confinement https://
t.co/jhb5pHMROq https://t.co/wFPiEx0KmA
##   ---
## 5819:
Bruxelles prend prétexte du Covid-19 pour bloquer les mesures contre les inégalités
femme-homme https://t.co/o6YEwL7mQt via @humanite_fr
## 5820:
Avec le gouverneur de la banque centrale en action d pas de contacte . Évitez les c
ontacts pour la sécurité et lutter contre le covid 19 https://t.co/k498fRaHKc
## 5821:
Faire une insomnie en 2020, c'est apprendre que de l'autre cote de l'Atlantique, le
président des États-Unis vient de proposer de s'injecter de la javel pour lutter co
ntre le covid-19.\nÇa valait le coup vraiment
## 5822:
@rochkaborepf Merci Monsieur le Président. Nous ferons de ce Ramadan, une riposte s
pirituelle contre le COVID-19. Plein succès dans activité du jour.
## 5823: "C’est aussi une période d’opportunité pour la #RSE : celle de montrer déf
initivement que s’investir pour être une entreprise plus résiliente, qui anticipe m
ieux, qui contribue mieux, ça paie. Et notamment sur le long terme, ou face aux cri
ses "  #ODD https://t.co/ouMY2EKQEd https://t.co/gOQk548cEF
```

Here, “contre” refers to everything linked to the fight against something. Most of “contre” we notice are directly linked to CoronaVirus fight. (It could have been about many different subjects, but as we see it on social network and by analyzing tweets, every talk or post is about CoronaVirus fight.)

```
dfFreq_long_top20 = dfFreq[rank <= 20] %>%
   melt(id.vars = c("feature","group","rank"),
        measure.vars = c("frequency","docfreq")
)



ggplot(dfFreq_long_top20, aes(x=reorder(feature,-rank), y=value, fill = variable))
+
   geom_bar(position="dodge", stat="identity") +
   scale_x_discrete() +
   labs(x = "", y = "Occurances", fill = "") +
   coord_flip() +
   theme_minimal()
```

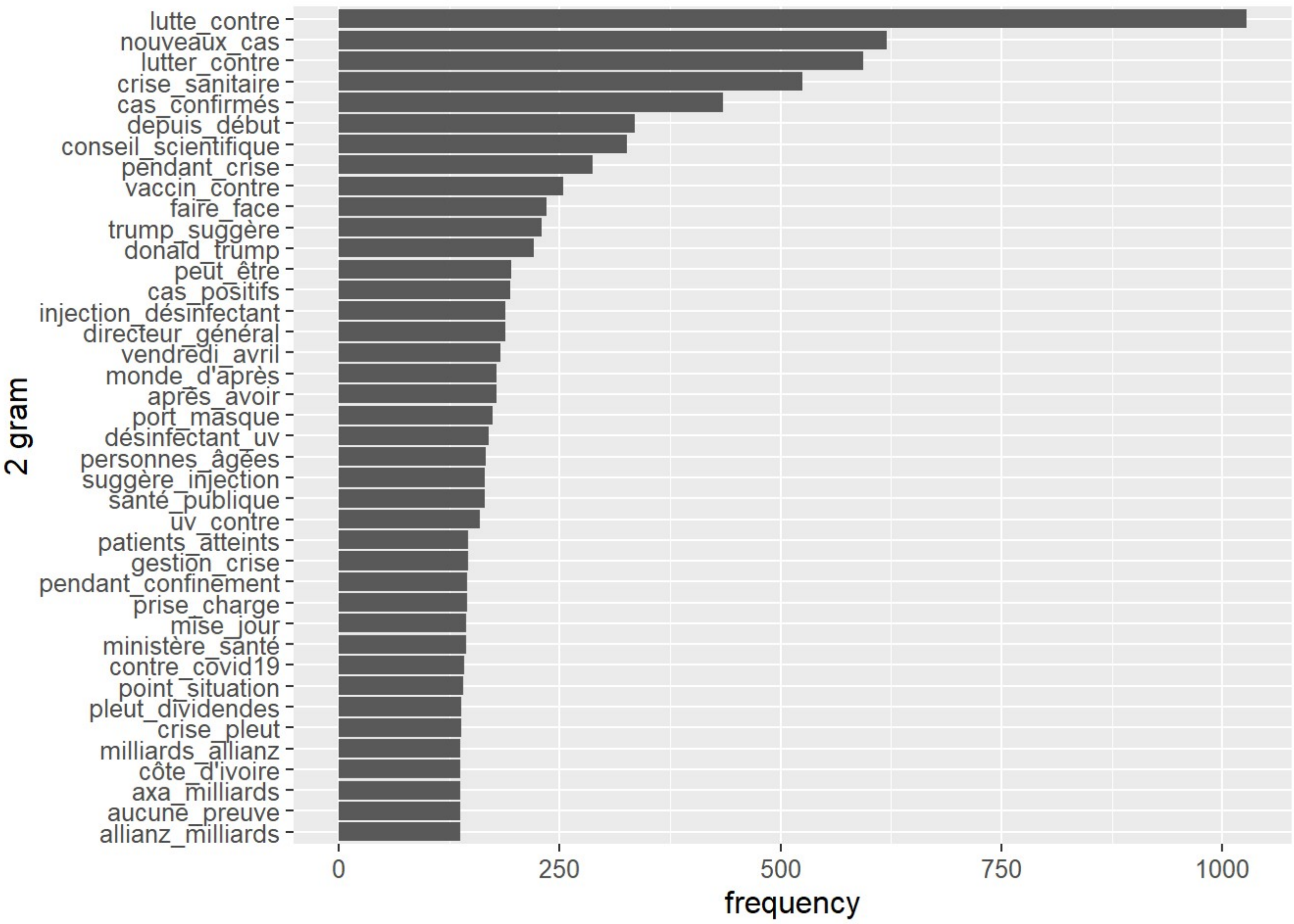


We have two different indexes here. In red, we have the frequence of each wordtweet (we have 5000 tweets) and in blue/green, the word contre appears almost 5000 times. We can observe not suchna big difference between the two, meaning that there isn't too much repetitions in the words.

```
TokensStemmed <- tokens_remove(tok_tweets, words.to.remove)

dfm2 <- dfm(tokens_ngrams(TokensStemmed,n=2))

dfFreq2 <- textstat_frequency(dfm2)

ggplot(dfFreq2[1:40,], aes(x=reorder(feature, frequency), y=frequency)) +
   geom_col() +
   coord_flip() +
   scale_x_discrete(name = "2 gram") +
   theme(text=element_text(size=12, family="Times New Roman"))
```

```
## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): famille de
## police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): famille de
## police introuvable dans la base de données des polices Windows
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows
```

```
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x,
## x$y, : famille de police introuvable dans la base de données des polices
## Windows
```

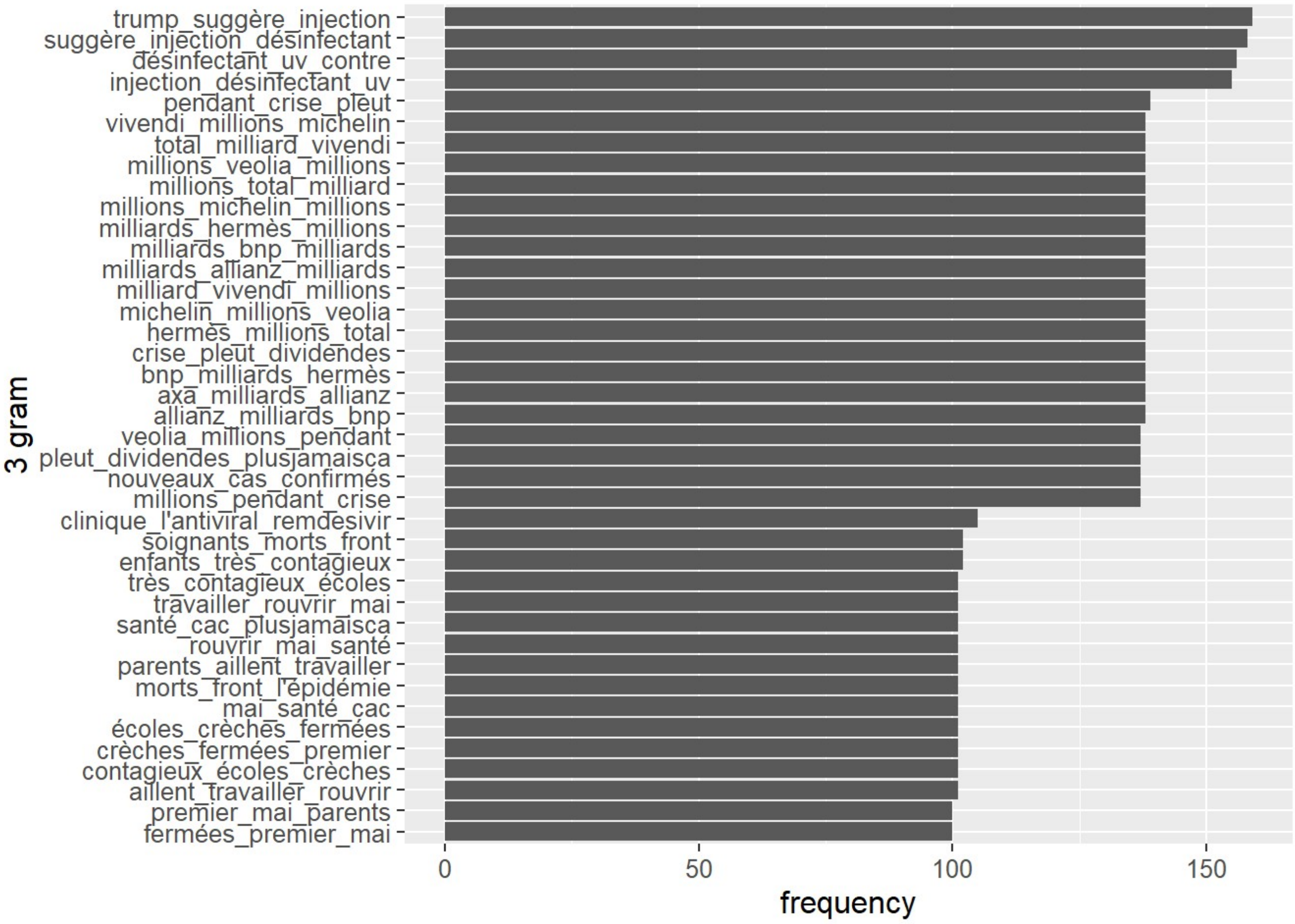


```
TokensStemmed <- tokens_remove(tok_tweets, words.to.remove)

dfm3 <- dfm(tokens_ngrams(TokensStemmed,n=3))

dfFreq3 <- textstat_frequency(dfm3)

ggplot(dfFreq3[1:40,], aes(x=reorder(feature, frequency), y=frequency)) +
   geom_col() +
   coord_flip() +
   scale_x_discrete(name = "3 gram") +
   theme(text=element_text(size=12, family="Times New Roman"))
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## famille de police introuvable dans la base de données des polices Windows
```

```
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x,
## x$y, : famille de police introuvable dans la base de données des polices
## Windows
```



2 gram graph: *The most relevant 2key words are clearly related to covid19 fight:* Lutte contre (fight against) Nouveaux cas (new cases) Crise sanitaire: (sanitary crisis) Cas confirmés: (confirmed cases) Conseil scientifique: (scientific counsil) Trump suggère: (Trump suggests) Vaccin contre: (vaccine against)

3 gram graph: *We can admit exactly the same thing:* Trump suggère injection (Trump suggests injection) Ecoles, crèches fermées (schools closed) Parents aillent travailler (parents going to work)

```
dtm <- convert(dfmat_corp_twitter, to = "topicmodels")
lda <- LDA(dtm, k = 6, control=list(seed=12))

terms(lda, 8) %>% utf8::utf8_print()
```

```
##      Topic 1    Topic 2      Topic 3   Topic 4   Topic 5    Topic 6
## [1,] "crise"    "contr"      "cas"     "grand"   "crise"    "contr"
## [2,] "million"  "cas"        "franc"   "jour"    "sanitair" "mort"
## [3,] "cas"      "+"          "personn" "mond"    "franc"    "comm"
## [4,] "milliard" "fair"       "contr"   "masqu"   "fair"     "mond"
## [5,] "confin"   "avril"      "face"    "contr"   "bien"     "pandémi"
## [6,] "tous"     "d'une"      "masqu"   "virus"   "mort"     "santé"
## [7,] "comm"     "\U0001f44f" "2020"    "#confin" "masqu"    "tous"
## [8,] "alor"     "qu'il"      "qu'il"   "amp"     "aprè"     "non"
```

Make a document feature matrix with 2-grams only. (To make one with 1-gram and 2-grams, use n = 1:2).

Topic 1: Clearly about the worldsize of the virus, the million cases we have, the quarantine we are all facing all over the world Topic 2: Covid situation during month of April, what we did and what is still left to do Topic 3: All about covid fighting in France, using masks, counting the cases Topic 4: Related to Topic 3, we have words related to the fight against the virus, its global size and the means we have to elimnate the virus (masks, quarantine) Topic 5: Covid consquences, we face a sanitary crisis and we think about the post-virus period we will face Topic 6: Related to Topic1: worldsize of the virus, the consequences on everybody.

```
dfm2 <- dfm(tokens_ngrams(TokensStemmed,n=2))
dfm2 <- convert(dfm2, to = "topicmodels")
lda2 <- LDA(dfm2, k = 6, control=list(seed=123))
terms(lda2, 8)
```

```
##      Topic 1             Topic 2                  Topic 3
## [1,] "pendant_crise"     "trump_suggère"          "nouveaux_cas"
## [2,] "crise_pleut"       "injection_désinfectant" "cas_confirmés"
## [3,] "pleut_dividendes"  "désinfectant_uv"        "lutte_contre"
## [4,] "axa_milliards"     "suggère_injection"      "cas_positifs"
## [5,] "milliards_allianz" "uv_contre"              "enfants_très"
## [6,] "allianz_milliards" "lutter_contre"          "très_contagieux"
## [7,] "milliards_bnp"     "chu_besançon"           "rouvrir_mai"
## [8,] "bnp_milliards"     "monde_d'après"          "mai_santé"
##      Topic 4                  Topic 5                    Topic 6
## [1,] "lutte_contre"           "lutte_contre"             "soignants_morts"
## [2,] "lutter_contre"          "conseil_scientifique"     "morts_front"
## [3,] "crise_sanitaire"        "directeur_général"        "hommage_soignants"
## [4,] "depuis_début"           "jean-philippe_ruggieri"   "front_l'épidémie"
## [5,] "l'antiviral_remdesivir" "général_nexity"           "crise_sanitaire"
## [6,] "clinique_l'antiviral"   "effets_indésirables"      "donald_trump"
## [7,] "essai_clinique"         "contre_paludisme"         "lutter_contre"
## [8,] "preuve_personnes"       "nouvelle_alerte"          "lutte_contre"
```

###Make a document feature matrix with 3-grams only. (To make one with 1-gram and 2-grams, use n = 1:2).

```
dfm3 <- dfm(tokens_ngrams(TokensStemmed,n=3))
dfm3 <- convert(dfm3, to = "topicmodels")
lda3 <- LDA(dfm3, k = 6, control=list(seed=123))
terms(lda3, 8)
```

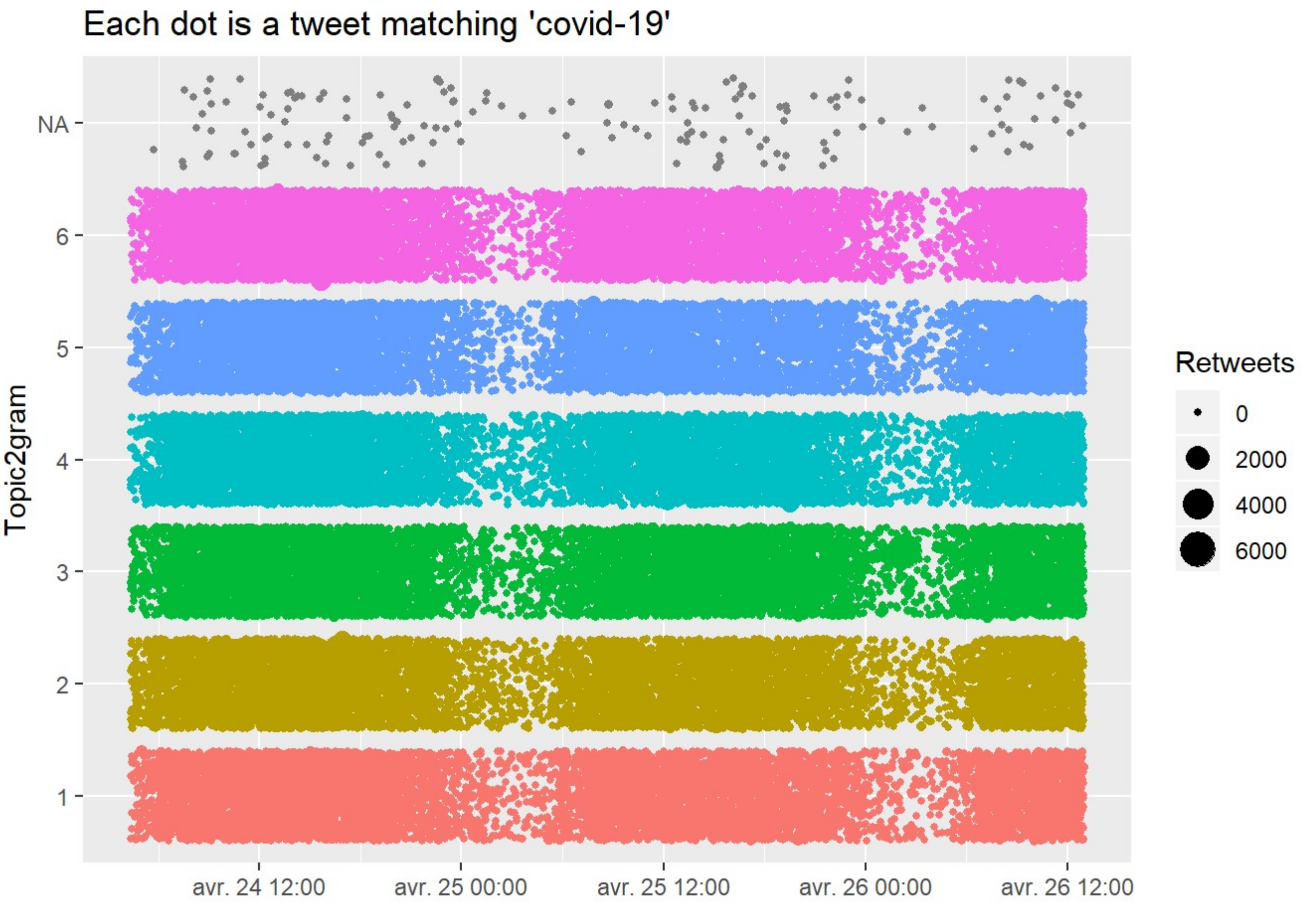
```
##      Topic 1
## [1,] "pendant_crise_pleut"
## [2,] "axa_milliards_allianz"
## [3,] "milliards_allianz_milliards"
## [4,] "allianz_milliards_bnp"
## [5,] "milliards_bnp_milliards"
## [6,] "bnp_milliards_hermès"
## [7,] "milliards_hermès_millions"
## [8,] "hermès_millions_total"
##      Topic 2
## [1,] "nouvelle_alerte_effets"
## [2,] "alerte_effets_indésirables"
## [3,] "effets_indésirables_l'hydroxychloroquine"
## [4,] "vers_surveillance_masse"
## [5,] "applis_vers_surveillance"
## [6,] "utile_lutter_contre"
## [7,] "autorités_santé_comme"
## [8,] "santé_comme_utile"
##      Topic 3                          Topic 4
## [1,] "directeur_général_nexity"       "clinique_l'antiviral_remdesivir"
## [2,] "effets_secondaires_graves"      "parcs_bondés_berlin"
## [3,] "option_l'is_option"             "bondés_berlin_contre"
## [4,] "l'is_option_l'intégration"      "berlin_contre_rues"
## [5,] "option_l'intégration_fiscale"   "contre_rues_désertes"
## [6,] "secondaires_graves_traitements" "rues_désertes_paris"
## [7,] "rétablir_l'isf_récupérer"       "delabrousse_découverte_mondiale"
## [8,] "l'isf_récupérer_milliards"      "chu_besançon_l'équipe"
##      Topic 5                          Topic 6
## [1,] "trump_suggère_injection"        "soignants_morts_front"
## [2,] "suggère_injection_désinfectant" "morts_front_l'épidémie"
## [3,] "désinfectant_uv_contre"         "hommage_soignants_morts"
## [4,] "injection_désinfectant_uv"      "entrevue_explosive_david"
## [5,] "enfants_très_contagieux"        "explosive_david_icke"
## [6,] "très_contagieux_écoles"         "david_icke_crise"
## [7,] "contagieux_écoles_crèches"      "icke_crise_partie"
## [8,] "écoles_crèches_fermées"         "culture_religieuse_l'iran"
```

We here have 6 topics list and we want to have a better view about the intesity in tweets about each topic.

We use a different graph to have a different view of the topics and their intensity

```
topicAssignment2grams =
   data.table(
      index = lda2 %>%
         topics %>%
         names %>%
         gsub("text","", .)
      %>% as.integer,
      topic = lda2 %>% topics
   )
tweets.df$Topic2gram = NA # creates a new col 'topic', assign it to NA
tweets.df$Topic2gram[topicAssignment2grams$index] = topicAssignment2grams$topic
tweets.df$Topic2gram = tweets.df$Topic2gram %>% as.factor
```

```
ggplot(tweets.df, aes(x=created_at, y=Topic2gram, col=Topic2gram)) +
   geom_jitter(aes(size = retweet_count)) +
   ggtitle(paste0("Each dot is a tweet matching '",query,"'")) +
   scale_y_discrete() +
   scale_x_datetime(name = "") +
   scale_color_discrete(guide = FALSE) +
   scale_size_continuous(name="Retweets")
```

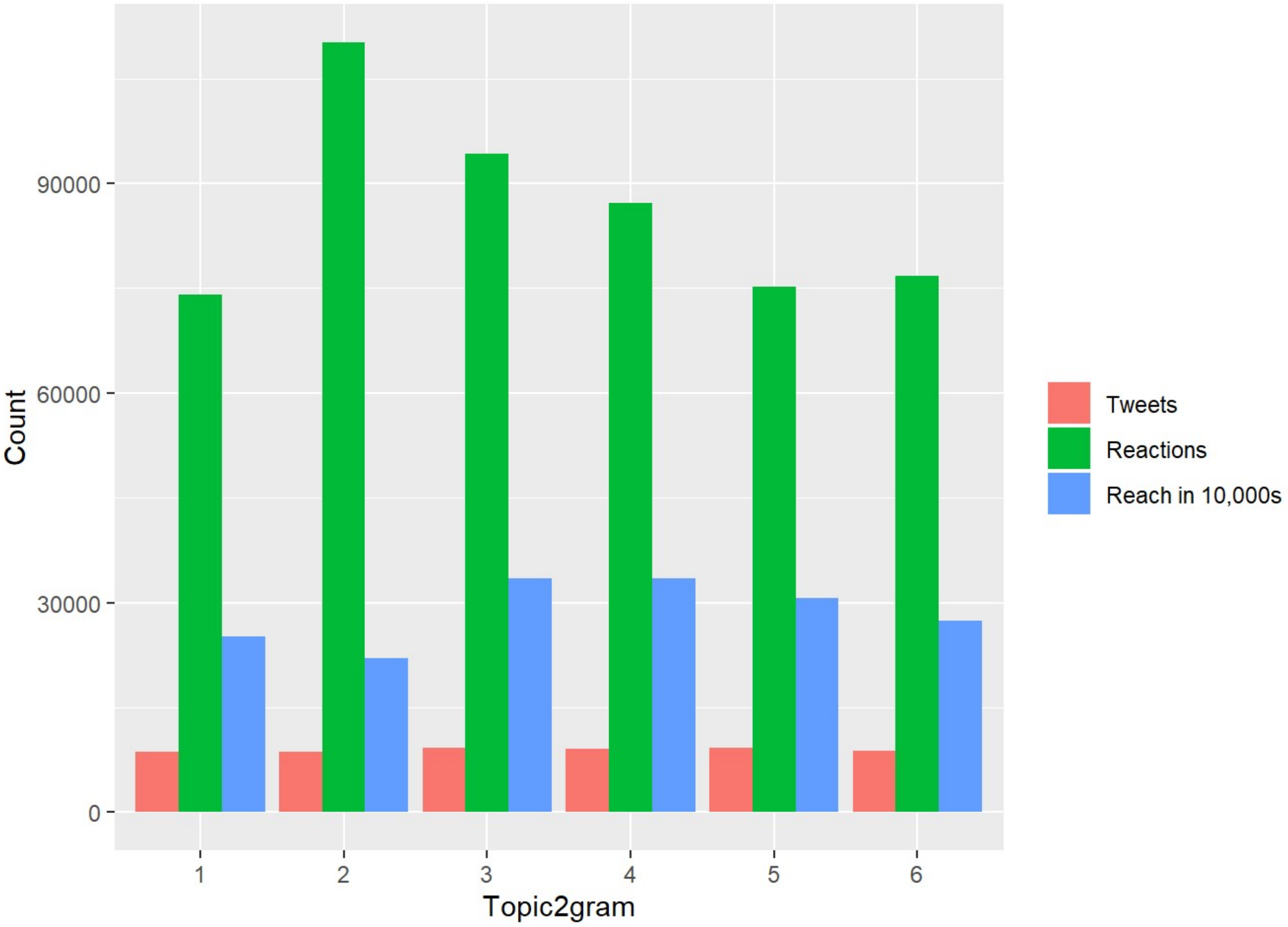


A big majority of the tweets have a Topic to be related to a topic except a few ones without topic related.

We use a different graph to have a different view of the topics and their intensity

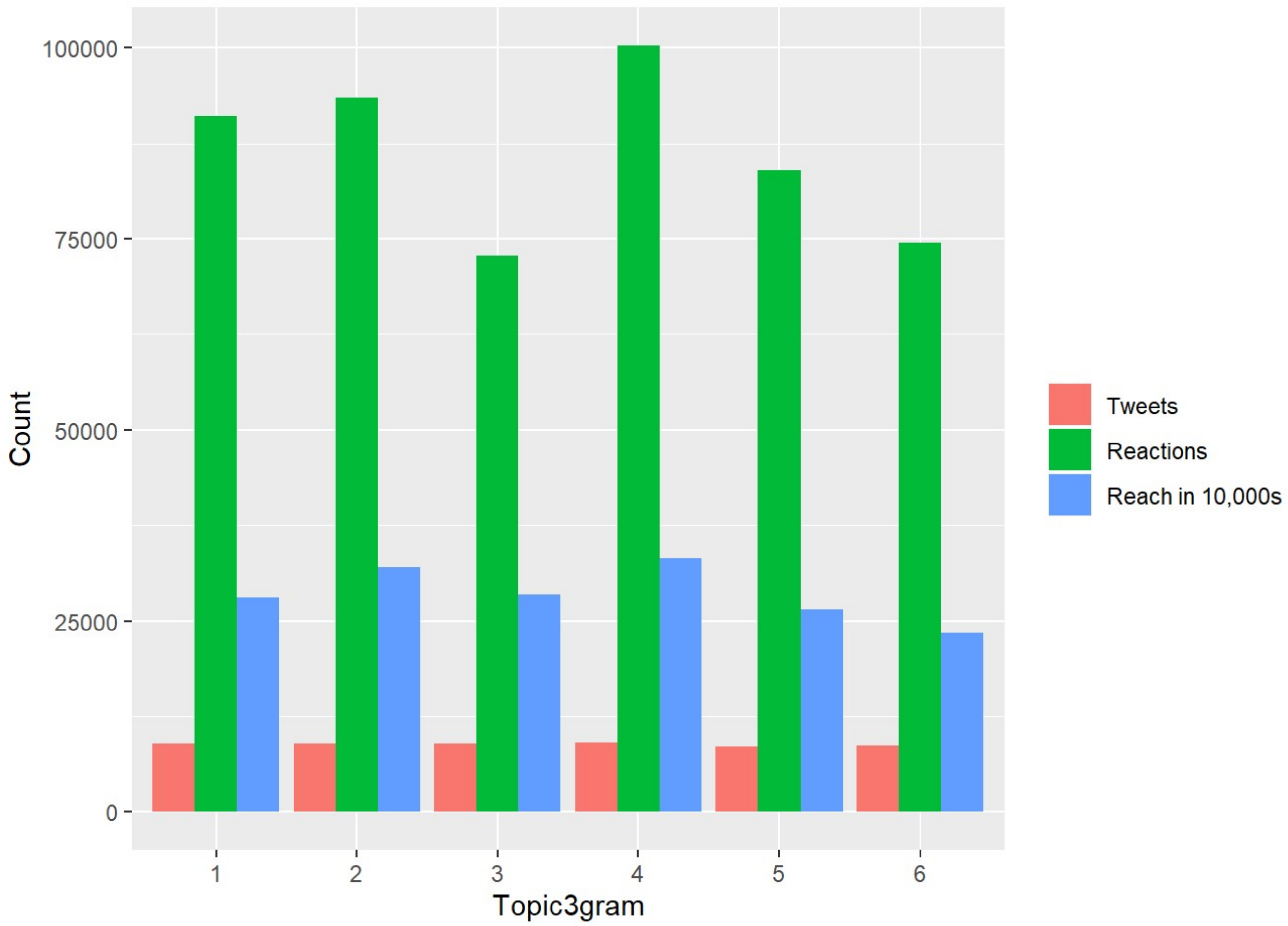
```
tweets.df[!is.na(Topic2gram),
          list(
             TotalTweets = .N,
             TotalReactions=sum(retweet_count, na.rm = TRUE) +
                sum(favorite_count, na.rm = TRUE)+
                sum(reply_count, na.rm = TRUE)+
                sum(quote_count, na.rm = TRUE),
             Reach = sum(followers_count)/10000
             ),
          by = Topic2gram] %>%
   melt(id.vars = "Topic2gram") %>%
   ggplot(aes(x = Topic2gram, y = value, fill=variable)) +
      geom_bar(position="dodge", stat="identity") +
      scale_fill_discrete(name= "", breaks=c("TotalTweets","TotalReactions","Reac
h"), labels = c("Tweets","Reactions","Reach in 10,000s")) +
      scale_y_continuous(name = "Count")
```

```
## Warning in melt.data.table(., id.vars = "Topic2gram"):
## 'measure.vars' [TotalTweets, TotalReactions, Reach] are not all of the same
## type. By order of hierarchy, the molten data value column will be of type
## 'double'. All measure variables not of type 'double' will be coerced too.
## Check DETAILS in ?melt.data.table for more on coercion.
```



```
topicAssignment3grams =
   data.table(
      index = lda3 %>%
         topics %>%
         names %>%
         gsub("text","", .)
      %>% as.integer,
      topic = lda3 %>% topics
   )
tweets.df$Topic3gram = NA # creates a new col ‘topic’, assign it to NA
tweets.df$Topic3gram[topicAssignment3grams$index] = topicAssignment3grams$topic
tweets.df$Topic3gram = tweets.df$Topic3gram %>% as.factor
tweets.df[!is.na(Topic3gram),
          list(
             TotalTweets = .N,
             TotalReactions=sum(retweet_count, na.rm = TRUE) +
                sum(favorite_count, na.rm = TRUE)+
                sum(reply_count, na.rm = TRUE)+
                sum(quote_count, na.rm = TRUE),
             Reach = sum(followers_count)/10000
             ),
          by = Topic3gram] %>%
   melt(id.vars = "Topic3gram") %>%
   ggplot(aes(x = Topic3gram, y = value, fill=variable)) +
      geom_bar(position="dodge", stat="identity") +
      scale_fill_discrete(name= "", breaks=c("TotalTweets","TotalReactions","Reac
h"), labels = c("Tweets","Reactions","Reach in 10,000s")) +
      scale_y_continuous(name = "Count")
```

```
## Warning in melt.data.table(., id.vars = "Topic3gram"):
## 'measure.vars' [TotalTweets, TotalReactions, Reach] are not all of the same
## type. By order of hierarchy, the molten data value column will be of type
## 'double'. All measure variables not of type 'double' will be coerced too.
## Check DETAILS in ?melt.data.table for more on coercion.
```



2 Gram: This graph gives us a vizualisation about the Tweets topic and their intensity (number of tweets posted:red/ the reactions to those: green/ and the number of people who saw the post: blue).

We notice that Topic 2 and Topic 6 are the ones who provocated most of the reactions even tho, there are not the ones with more tweets but lots of view and so many reactions (more than 9K fot both of them)

3 gram:here the most active ones Topic 2 and 3. Important to notice that Topic 2 is very very active in 2 grams and 3 grams, making people react to it a lot -> people were very worried about the situation during April and lots of people give their opinion and information.

We are now going to focus on the sentiment analysis to have a better view about the influence of each word in its topic and globally

```
df <- tweets.df[,.(created_at,text,Topic2gram)]
```

We are creating a Dataframe of only tweets about Topic2grams including the creation date. ## We are now going to create cuts every 5 minutes

```
df$roundTime <- as.POSIXct(cut(df$created_at, breaks = "5 mins"))
df$text[1]
```

```
## [1] "@BFMTV Sourtout pas par des incapables comme vous qui ont rien rien prévu #
Covid_19"
```

```
df$text[1] %>% get_sentences
```

```
## [[1]]
## [1] "@BFMTV Sourtout pas par des incapables comme vous qui ont rien rien prévu #
Covid_19"
##
## attr(,"class")
## [1] "get_sentences"           "get_sentences_character"
## [3] "list"
```

```
df$text[1] %>% get_sentences %>% sentiment
```

```
##    element_id sentence_id word_count sentiment
## 1:          1           1         15         0
```

We created a new column in our data frame, where we group every tweet grouped by a 5 minute rythm (breaks = 5 mins) We also tried to compute the sentiment of all those tweets and we got 0 because the language of our tweets was French and the package was in English.

To solve this language problem, we downloaded a french dictonary file helping us to give a score to every word we haveA.

We will read the FrenchAdj.csv downloaded from Campus

```
fr_keys <-read.csv("C:/Users/steve/OneDrive/Documents/R/NLP/frenchAdj.csv")
head(fr_keys)
```

```
##   X     x     y
## 1 1  tout  0.28
## 2 2 petit  0.18
## 3 3 grand  0.50
## 4 4  seul  0.46
## 5 5 autre -0.04
## 6 6 mÃªme -0.06
```

Now, as in the lecture, we remove the Index X

```
fr_keys <- fr_keys[,2:3]
```

```
fr_keys <- as_key(fr_keys)
```

```
## Warning in as_key(fr_keys): Column 1 was a factor...
## Converting to character.
## Warning in as_key(fr_keys): One or more terms in the first column appear as term
s in the comparison.
##   I found the following dubious fellas:
##
##    * certain
##    * immense
##
## These terms have been removed.
```

```
head(fr_keys)
```

```
##              x     y
## 1: abandonnÃ© -0.20
## 2:     abject -1.00
## 3: abominable -1.00
## 4:     absent  0.78
## 5:     absolu  0.20
## 6:    absurde  0.64
```

```
head(fr_keys)
```

```
##              x     y
## 1: abandonnÃ© -0.20
## 2:     abject -1.00
## 3: abominable -1.00
## 4:     absent  0.78
## 5:     absolu  0.20
## 6:    absurde  0.64
```

Our dictionary is now fully adapted to french tweet and to sentiment library format.

```
df$roundTime <- as.POSIXct(cut(df$created_at, breaks = "5 mins"))
df$text[1]
```

```
## [1] "@BFMTV Sourtout pas par des incapables comme vous qui ont rien rien prévu #
Covid_19"
```

```
df$text[1] %>% get_sentences
```

```
## [[1]]
## [1] "@BFMTV Sourtout pas par des incapables comme vous qui ont rien rien prévu #
Covid_19"
##
## attr(,"class")
## [1] "get_sentences"           "get_sentences_character"
## [3] "list"
```

```
df$text[1] %>% get_sentences %>% sentiment(polarity_dt = fr_keys)
```

```
##    element_id sentence_id word_count  sentiment
## 1:          1           1         15 0.03098387
```

Here with have the sentiment analysis about the 1st post. We can admit that it contains sentiment in it (positive or negative) but a feelig is clearly expressed in it.

```
sentiment_by_tweet =
   df[,
      list(text %>% get_sentences %>% sentiment_by(polarity_dt = fr_keys),
           Topic2gram)]
```

Unfortunately, we couldn't have the sentiment analysis due to programming issues. However, we got a lot of relevant information about the people's feelings.

conclusion

At the end of our analysis, we were able to see that covid 19 in France is at the heart of people's concerns. The fight against this disease, the search for a vaccine and the government's action plan are the main issues surrounding the virus. In addition, the French are also interested in the response that other countries, notably the United States (what does Donald Trump think?) have undertaken against the virus. Finally, the last real concern we have observed is the period of deconfinement and how it will be set up operationally.