Descriptive Statistics

Teresa Graffi

23/8/2021

Dataset

Creating an object containing last 2000 tweets from the timelines of each of the three selected accounts (Giorgia Meloni, Carlo Calenda, Matteo Renzi):

```
tlMCR <- get_timeline(c("GiorgiaMeloni", "CarloCalenda", "matteorenzi"), 2000, token=auth)
head(tlMCR)
 > # A tibble: 6 x 90
 >   user_id  status_id   created_at          screen_name  text             source
 >   <chr>    <chr>       <dttm>              <chr>        <chr>            <chr>
 > 1 1305370~ 1433071752~ 2021-09-01 14:18:25 GiorgiaMelo~ "Arrivati nella~ Twitte~
 > 2 1305370~ 1433066336~ 2021-09-01 13:56:54 GiorgiaMelo~ ""Se è un sogno~ Twitte~
 > 3 1305370~ 1433036480~ 2021-09-01 11:58:16 GiorgiaMelo~ "Mesi da scerif~ Twitte~
 > 4 1305370~ 1432994186~ 2021-09-01 09:10:12 GiorgiaMelo~ ""Mi picchiano ~ Twitte~
 > 5 1305370~ 1432702669~ 2021-08-31 13:51:49 GiorgiaMelo~ "I miei miglior~ Twitte~
 > 6 1305370~ 1432669186~ 2021-08-31 11:38:46 GiorgiaMelo~ "Raffiche di sb~ Twitte~
 > # ... with 84 more variables: display_text_width <dbl>,
 > #   reply_to_status_id <chr>, reply_to_user_id <chr>,
 > #   reply_to_screen_name <chr>, is_quote <lgl>, is_retweet <lgl>,
 > #   favorite_count <int>, retweet_count <int>, quote_count <int>,
 > #   reply_count <int>, hashtags <list>, symbols <list>, urls_url <list>,
 > #   urls_t.co <list>, urls_expanded_url <list>, media_url <list>,
 > #   media_t.co <list>, media_expanded_url <list>, media_type <list>,
 > #   ext_media_url <list>, ext_media_t.co <list>, ext_media_expanded_url <list>,
 > #   ext_media_type <chr>, mentions_user_id <list>, mentions_screen_name <list>,
 > #   lang <chr>, quoted_status_id <chr>, quoted_text <chr>,
 > #   quoted_created_at <dttm>, quoted_source <chr>, quoted_favorite_count <int>,
 > #   quoted_retweet_count <int>, quoted_user_id <chr>, quoted_screen_name <chr>,
 > #   quoted_name <chr>, quoted_followers_count <int>,
 > #   quoted_friends_count <int>, quoted_statuses_count <int>,
 > #   quoted_location <chr>, quoted_description <chr>, quoted_verified <lgl>,
 > #   retweet_status_id <chr>, retweet_text <chr>, retweet_created_at <dttm>,
 > #   retweet_source <chr>, retweet_favorite_count <int>,
 > #   retweet_retweet_count <int>, retweet_user_id <chr>,
 > #   retweet_screen_name <chr>, retweet_name <chr>,
 > #   retweet_followers_count <int>, retweet_friends_count <int>,
 > #   retweet_statuses_count <int>, retweet_location <chr>,
 > #   retweet_description <chr>, retweet_verified <lgl>, place_url <chr>,
 > #   place_name <chr>, place_full_name <chr>, place_type <chr>, country <chr>,
 > #   country_code <chr>, geo_coords <list>, coords_coords <list>,
 > #   bbox_coords <list>, status_url <chr>, name <chr>, location <chr>,
 > #   description <chr>, url <chr>, protected <lgl>, followers_count <int>,
```

```
> #   friends_count <int>, listed_count <int>, statuses_count <int>,
> #   favourites_count <int>, account_created_at <dttm>, verified <lgl>,
> #   profile_url <chr>, profile_expanded_url <chr>, account_lang <lgl>,
> #   profile_banner_url <chr>, profile_background_url <chr>,
> #   profile_image_url <chr>
```

Temporal view

Temporal band in which the 2000 tweets have been created:

```
by(tlMCR$created_at,tlMCR$screen_name,summary)
 > tlMCR$screen_name: CarloCalenda
 >                  Min.               1st Qu.                Median
 > "2021-06-08 09:10:30" "2021-06-27 17:37:41" "2021-07-21 14:46:05"
 >                  Mean               3rd Qu.                  Max.
 > "2021-07-21 15:18:16" "2021-08-15 09:36:41" "2021-09-01 16:44:25"
 > ------------------------------------------------------------
 > tlMCR$screen_name: GiorgiaMeloni
 >                  Min.               1st Qu.                Median
 > "2020-08-13 06:52:14" "2020-11-27 06:53:19" "2021-03-07 16:25:28"
 >                  Mean               3rd Qu.                  Max.
 > "2021-02-27 01:55:45" "2021-05-27 10:25:45" "2021-09-01 14:18:25"
 > ------------------------------------------------------------
 > tlMCR$screen_name: matteorenzi
 >                  Min.               1st Qu.                Median
 > "2020-03-19 22:40:52" "2020-06-15 16:46:35" "2020-11-23 14:43:05"
 >                  Mean               3rd Qu.                  Max.
 > "2020-11-15 02:05:21" "2021-03-24 08:24:18" "2021-09-01 17:27:35"
```

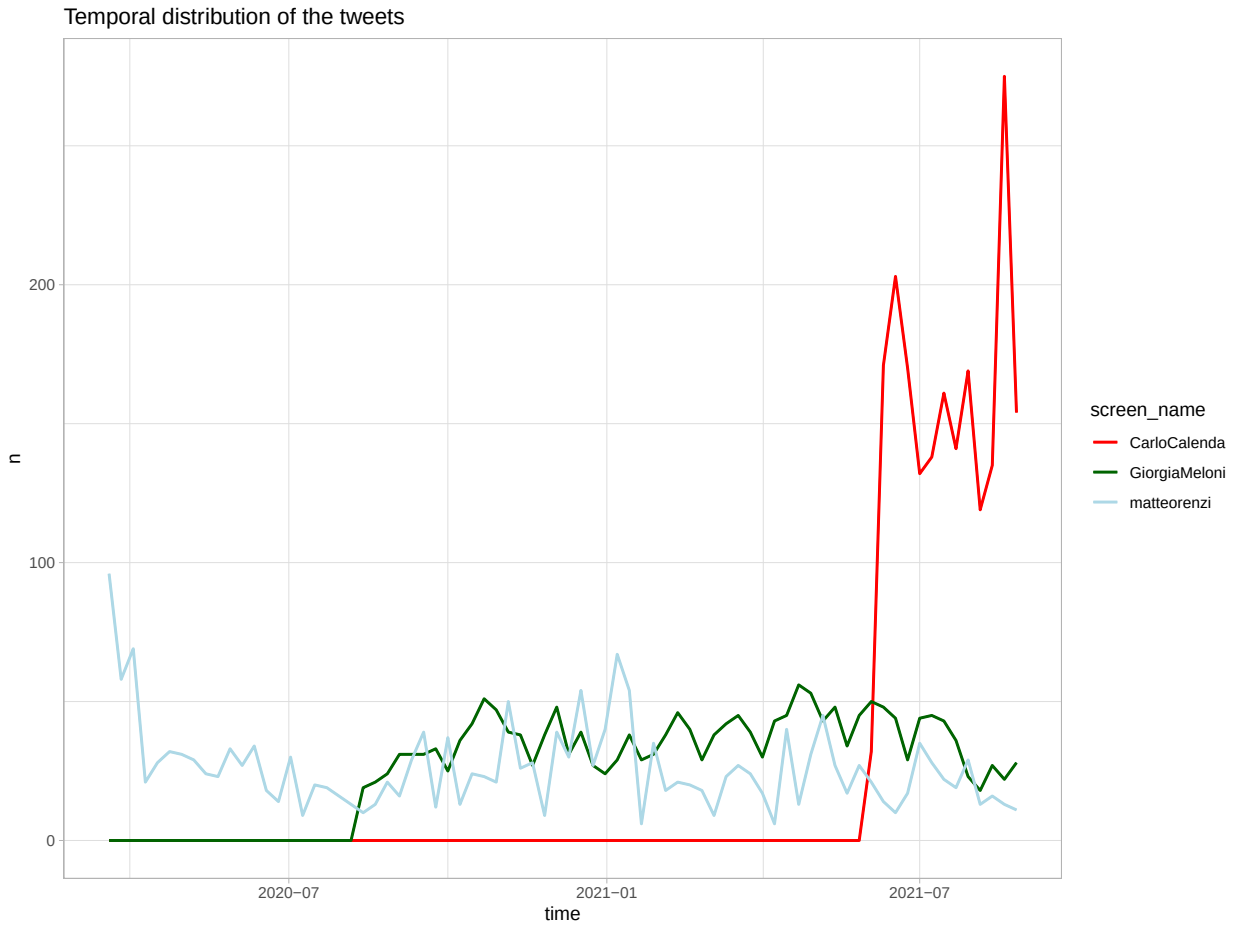
M has been posting 2000 tweets in 1 year C posted 2000 tweets in just 3 months R posted 2000 tweets in almost 1 year and an half

Restricting the timelines to study the those tweets published starting from 2020-01-01:

```
tlMCRr <- tlMCR[as.Date(tlMCR$created_at) >= "2020-01-01",]
table(tlMCRr$screen_name)
 >
 >  CarloCalenda GiorgiaMeloni   matteorenzi
 >          2000          2000          1998
```

To have also a temporal point of view of when the tweets have been published, from the tweets extracted from the timelines a graph as been plotted considering the weeks starting from 2020-01-01. What is remarkable is that M. Renzi is the first one of the three in publishing tweets, while G. Meloni started during the second part of the year 2020 and C. Calenda joined the other two users just during the past few months, but the number of tweets published by him is significantly higher than the ones of the other two users.

```
tlMCRr %>%
  dplyr::group_by(screen_name) %>%
  ts_plot("weeks") +
  theme_light() +
  scale_color_manual(values=c("red", "darkgreen", "lightblue")) +
  geom_line(size=0.8) +
  labs(title="Temporal distribution of the tweets")
```



Variables of interest

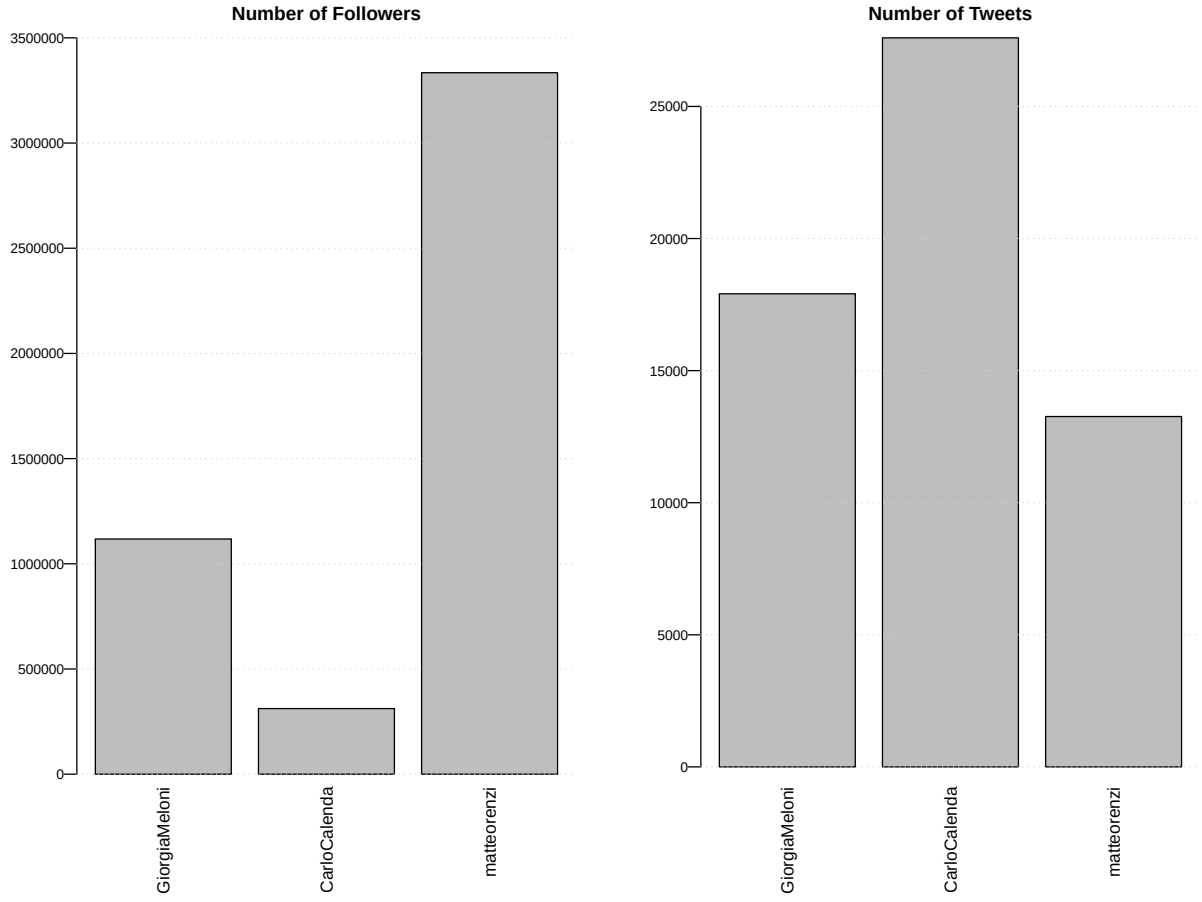
To plot some graphs considering number of followers and tweet production of the three accounts, some vectors have been created, containing the selected users (vsUsr) and the most interesting variables:

```
vsUsr <- c("GiorgiaMeloni", "CarloCalenda", "matteorenzi") #Vector with Selected Users
sUsrs <- lookup_users(users = vsUsr, token=auth)
sUsrs[,c("user_id","screen_name","location","account_created_at","followers_count",
         "friends_count","statuses_count","favourites_count")]
 > # A tibble: 3 x 8
 >   user_id screen_name location account_created_at  followers_count friends_count
 >   <chr>   <chr>       <chr>    <dttm>                         <int>         <int>
 > 1 130537~ GiorgiaMel~ "Italia" 2010-04-07 15:43:26         1117884           922
 > 2 241606~ CarloCalen~ ""       2014-03-28 15:27:38          311820           863
 > 3 187628~ matteorenzi "Italy"  2009-01-08 13:15:21         3334343           970
 > # ... with 2 more variables: statuses_count <int>, favourites_count <int>
```

Setting the margin limits and plotting a bar histogram that shows the number of followers, and another one for the number of tweets:

```
par(mar=c(6,4,2,1))
par(mgp=c(1.5,0.5,0))
```

```
par(mfrow=c(1,2))

barplot(sUsrs$followers_count, names.arg = sUsrs$screen_name, las=2,
        main="Number of Followers",
        ylim = c(0,3500000), cex.axis = 0.7, cex.names = 0.8, cex.main=0.9) +
        grid(NA,NULL)
 > numeric(0)

barplot(sUsrs$statuses_count,names.arg = sUsrs$screen_name,las=2,
        main="Number of Tweets",
        cex.axis = 0.7,cex.names = 0.8, cex.main=0.9)+
        grid(NA,NULL)
```

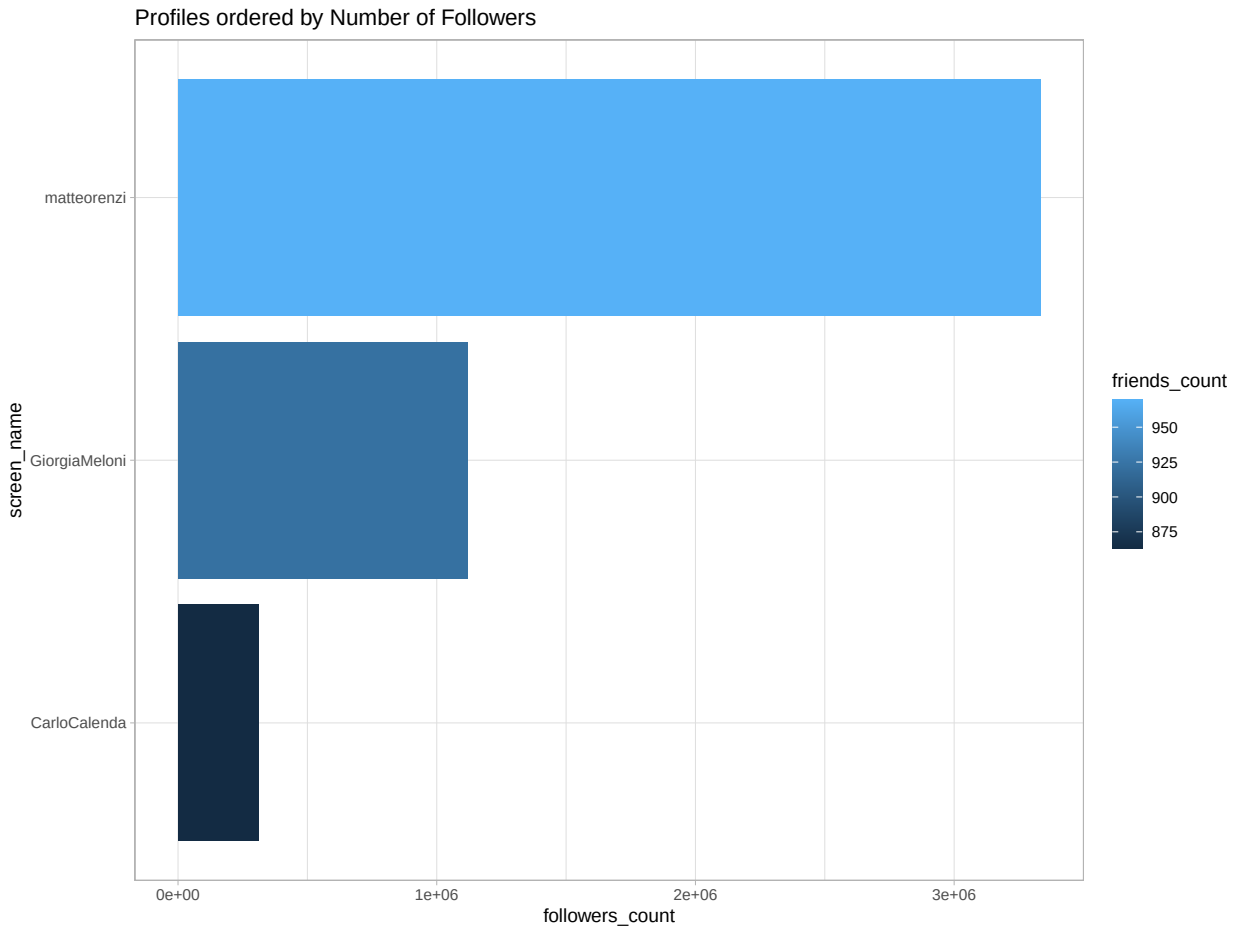



```
 > numeric(0)
```

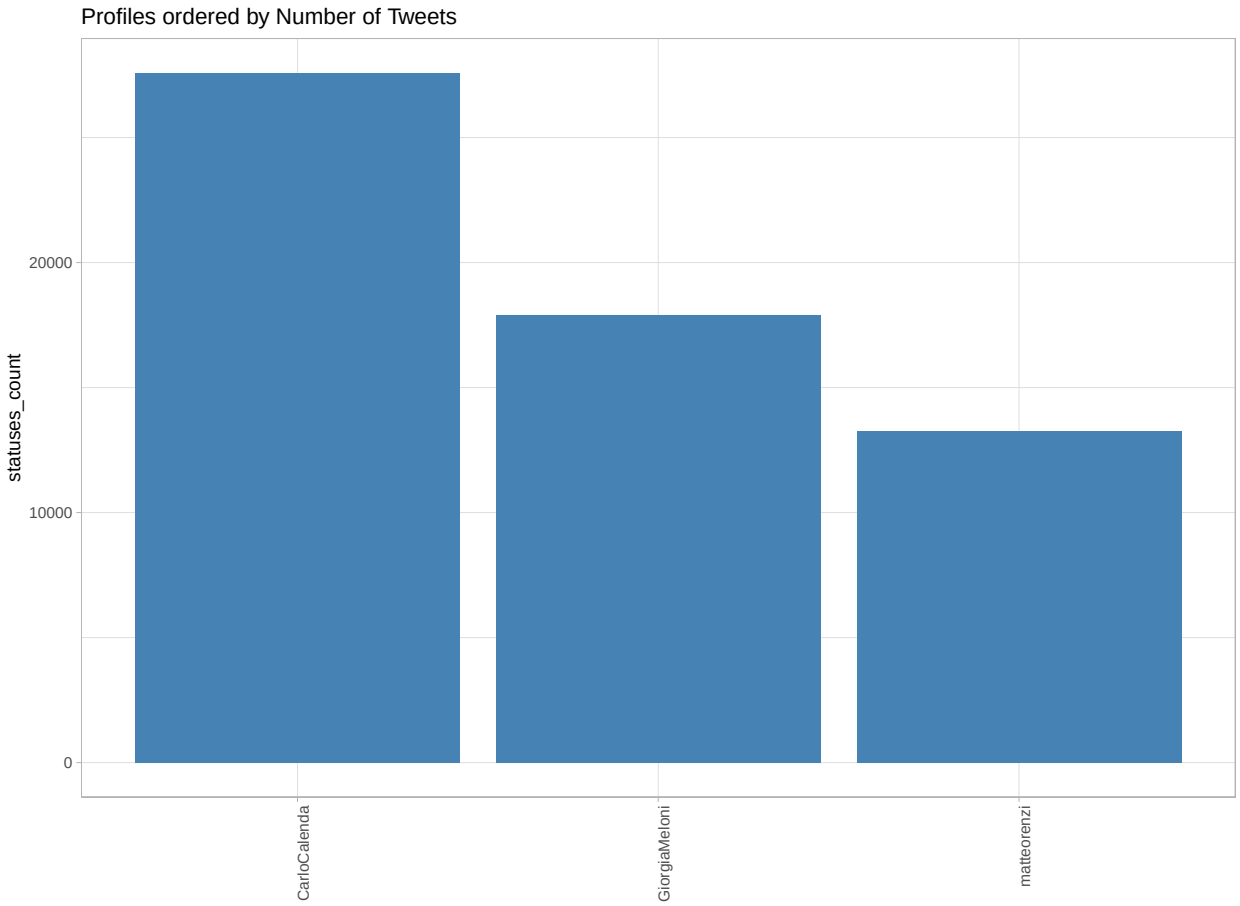
The user with the highest number of followers is Renzi with more than 3 milion followers, while Calenda can't reach half a milion (330.000) and Meloni slightly exceeds 1 milion followers. Whereas for what concerns the number of tweets, Calenda published the higher number of tweets (more than 25.000), Meloni exceed 15.000 and Renzi reaches almost 15.000.

Using the library {ggplot2} we can show some further information about these profiles. The first plot shows on the x-axis the number of followers and on the y-axis the name of three users, while the color of the bars represents the friends count that is the number of users that the selected users are following.

```
library(ggplot2)
ggplot(data=sUsrs,
       aes(x=screen_name,
           y=followers_count,
           fill=friends_count)) +
  geom_bar(stat="identity") +
  coord_flip() +
  theme_light() +
  ggtitle("Profiles ordered by Number of Followers")
```




```
ggplot(data=sUsrs,
       aes(x=screen_name,
           y=statuses_count)) +
  geom_bar(stat="identity", fill="steelblue") +
  theme_light() +
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1)) +
  ggtitle("Profiles ordered by Number of Tweets") +
  xlab("")
```

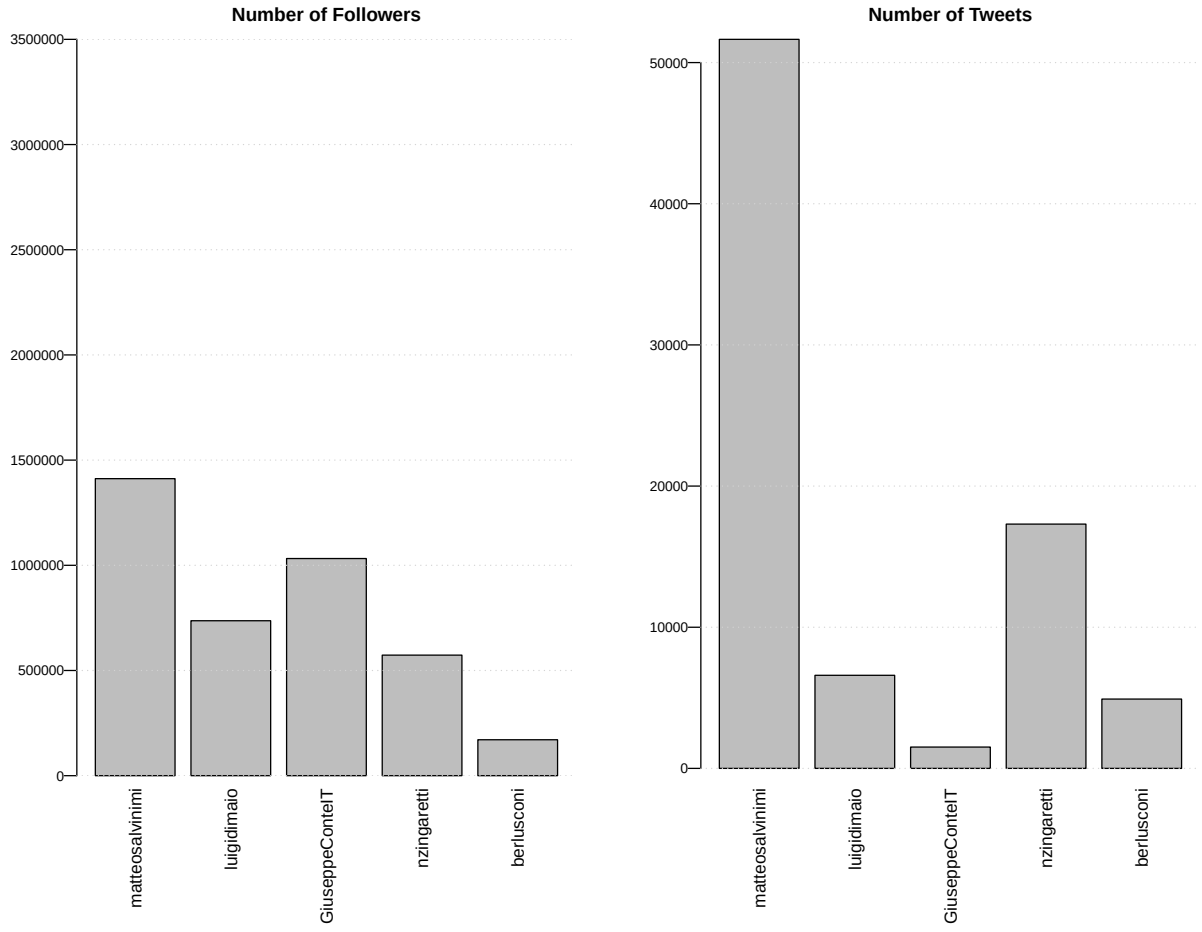


A look to the statistics of other users

Accounts of the 5 leaders of the main Italian parties To get a set of the actual context of politicians using Twitter, the same features of before have been plotted but regarding other Italians politicians (Matteo Salvini, Luigi di Maio, Giuseppe Conte, Nicola Zingaretti, Silvio Berlusconi):

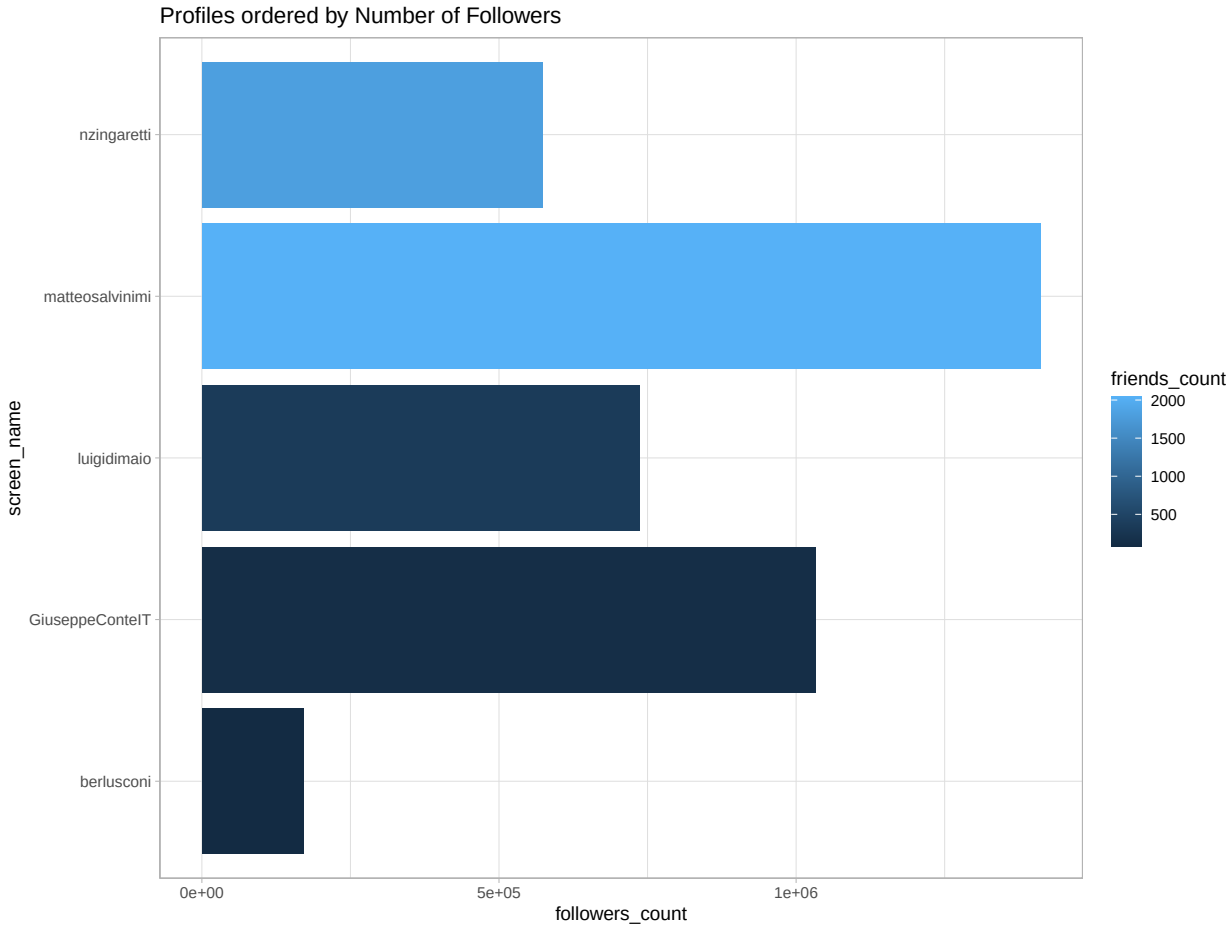
```
#Grafici con numero di follower, produzione di tweet per altri politici
vUsr <- c("matteosalvinimi", "luigidimaio", "GiuseppeConteIT", "nzingaretti", "berlusconi")
Usrs <- lookup_users(users = vUsr, token = auth)
Usrs[,c("user_id","screen_name","location","account_created_at","followers_count",
        "friends_count","statuses_count","favourites_count")]
 > # A tibble: 5 x 8
 >   user_id screen_name location account_created_at  followers_count friends_count
 >   <chr>   <chr>       <chr>    <dttm>                         <int>         <int>
 > 1 270839~ matteosalv~ ""       2011-03-23 10:32:56         1411850          2040
 > 2 480627~ luigidimaio "Pomigl~ 2009-06-17 18:34:46          736539           351
 > 3 999578~ GiuseppeCo~ ""       2018-05-24 09:09:30         1032240           136
 > 4 403544~ nzingaretti ""       2011-11-02 17:05:45          573009          1802
 > 5 920277~ berlusconi  ""       2017-10-17 13:15:10          170757            81
 > # ... with 2 more variables: statuses_count <int>, favourites_count <int>

par(mar=c(6,4,2,1))
par(mgp=c(1.5,0.5,0))
```

```
par(mfrow=c(1,2))
#istogramma a barre per numero di followers
barplot(Usrs$followers_count, names.arg = Usrs$screen_name, las=2,
        main="Number of Followers",
        ylim = c(0,3500000), cex.axis = 0.7, cex.names = 0.8, cex.main=0.9) +
  grid(NA,NULL)
 > numeric(0)

#istogramma a barre per numero di tweet
barplot(Usrs$statuses_count,names.arg = Usrs$screen_name,las=2,
        main="Number of Tweets",
        cex.axis = 0.7,cex.names = 0.8, cex.main=0.9)+
  grid(NA,NULL)
```

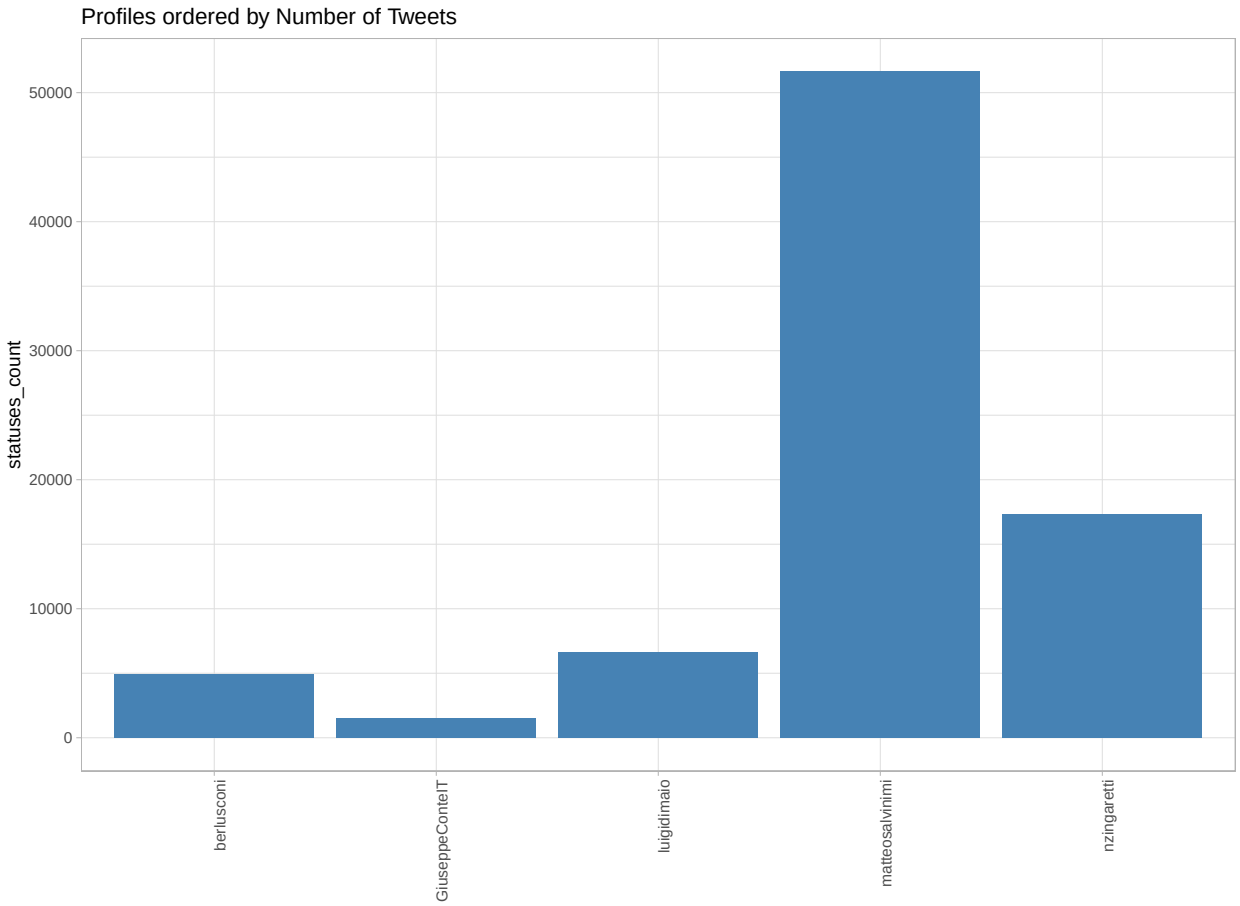



```
 > numeric(0)

library(ggplot2)
ggplot(data=Usrs,
       aes(x=screen_name,
           y=followers_count,
           fill=friends_count)) +
  geom_bar(stat="identity") +
  coord_flip() +
```

```
  theme_light() +
  ggtitle("Profiles ordered by Number of Followers")
```




```
ggplot(data=Usrs,
       aes(x=screen_name,
           y=statuses_count)) +
  geom_bar(stat="identity", fill="steelblue") +
  theme_light() +
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1)) +
  ggtitle("Profiles ordered by Number of Tweets") +
  xlab("")
```



Main features

> **Followers** It is also networthy to see whether the three selected accounts have common followers or not. To do so 10.000 followers for each user (GM, CC, MR) have beeen selected; these lists can be joined together to verify how many users (user_id) follow either one, two or all three accounts.

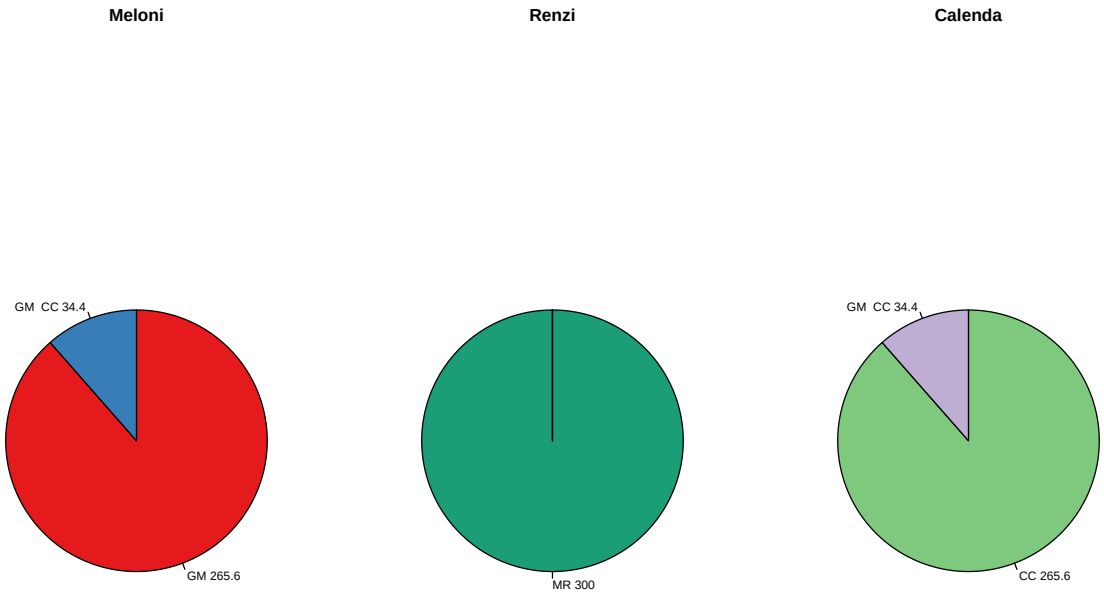
```
fGM <- get_followers(user = sUsrs$user_id[1],n = 3000, token = auth)
fCC <- get_followers(user = sUsrs$user_id[2],n = 3000, token = auth)
fMR <- get_followers(user = sUsrs$user_id[5],n = 3000, token = auth)

fGM$user <- "GM"
fMR$user <- "MR"
fCC$user <- "CC"

# creo un data frame dall'abbinamento delle liste di Meloni e Renzi
fGMMR <- merge(fGM, fMR, by = "user_id", all = T)
# modifico i nomi delle ultime due colonne
colnames(fGMMR)[2:3] <- c("user.GM", "user.MR")
# abbino il nuovo data frame con la lista dei follower di Calenda
fGMMRCC <- merge(fGMMR, fCC, by = "user_id", all = T)
colnames(fGMMRCC)[4] <- "user.CC"
```

```
# sostituisco gli NA con uno spazio
fGMMRCC[is.na(fGMMRCC)] <- ""
# creo un campo con l'accorpamento delle sigle
fGMMRCC$who <- paste(fGMMRCC$user.GM,fGMMRCC$user.MR,fGMMRCC$user.CC,sep=" ")
# elimino gli spazi inutili dal campo who
fGMMRCC$who <- gsub(" "," ",fGMMRCC$who)
fGMMRCC$who <- gsub("^\\s+|\\s+$", "", fGMMRCC$who)

cbind(n=addmargins(table(fGMMRCC$who)),
      percent=addmargins(prop.table(table(fGMMRCC$who)))*100)
 >           n    percent
 > CC     2656  30.683919
 > GM     2656  30.683919
 > GM  CC  344   3.974122
 > MR     3000  34.658041
 > Sum    8656 100.000000
```

From the table it can be seen that just Meloni and Calenda have some followers in common. Still it is important to remember that the 3000 followers are just a sample and it is small compared to the total number of followers for each account so it may be that extracting different times the samples, the result could change a little bit.

It is possible to plot these results using a chart pie:

```
library(RColorBrewer)

par(mfrow = c(1,3))
tbGM <- table(fGMMRCC[fGMMRCC$user.GM != "","who"])
pie(tbGM,clockwise = T,labels = paste(names(tbGM),tbGM/1000*100),cex=0.9,radius = 0.9,
    main="Meloni",col=brewer.pal(4,"Set1"))
tbMR <- table(fGMMRCC[fGMMRCC$user.MR != "","who"])
pie(tbMR,clockwise = T,labels = paste(names(tbMR),tbMR/1000*100),cex=0.9,radius = 0.9,
    main="Renzi",col=brewer.pal(4,"Dark2"))
tbCC <- table(fGMMRCC[fGMMRCC$user.CC != "","who"])
pie(tbCC,clockwise = T,labels = paste(names(tbCC),tbCC/1000*100),cex=0.9,radius = 0.9,
    main="Calenda",col = brewer.pal(4,"Accent"))
```



Now that some information regarding the followers of the users has been collected, it is possible to see also the content of the tweets published from the three selected accounts. Creating an object called "tl3" we save last 1000 tweets published for each of the three accounts, without taking into consideration retweets.

```
tl3 <- get_timeline(user = sUsrs$user_id[c(1,2,3)], n = 1000, include_rts = F,
                    token = auth)
table(tl3$screen_name)
 >
 >  CarloCalenda GiorgiaMeloni   matteorenzi
 >           716           854           615
```

> Retweets, Quotes and Hashtags for each account For each account 2000 tweets have been collected, then the temporal band of the creation of the tweets is shown as well as the number tweets that are retweets or quotes (False stands for non retweets, True means that tweet is a retweet, the same applies for quotes). The last two tables display the 10 most used hashtags of the account and the 10 most used mentions, this passage is done by transforming the content of the list containing the hashtags in a vector using the function unlist, then a frequency table for the hashtags is created and the result is ordered in decreasing order. Here follows the results for these procedures applied on the profile of Giorgia Meloni.

```
#MELONI
tlGM <- get_timeline("GiorgiaMeloni", 2000, token = auth)
summary(tlGM$created_at)
 >                  Min.               1st Qu.                Median
```

```
 > "2020-08-13 06:52:14" "2020-11-27 06:53:19" "2021-03-07 16:25:28"
 >                   Mean                3rd Qu.                   Max.
 > "2021-02-27 01:55:45" "2021-05-27 10:25:45" "2021-09-01 14:18:25"
table(tlGM$is_retweet)
 >
 > FALSE  TRUE
 >  1706   294
table(tlGM$is_quote)
 >
 > FALSE  TRUE
 >  1987    13

sort(table(unlist(tlGM$hashtags)),decreasing = T)[1:10]
 >
 >  FratellidItalia           Meloni  BastaCoprifuoco     BloccoNavale
 >              136               74               28               17
 >     BastaSbarchi QuartaRepubblica    StaseraItalia      portaaporta
 >               15               13               13               12
 >       coprifuoco          COVID19
 >               11               11
sort(table(unlist(tlGM$mentions_screen_name)),decreasing = T)[1:10]
 >
 > FratellidItalia   GiorgiaMeloni        ECRparty      BrunoVespa        Corriere
 >             280             157              26              11              11
 >      AcquaroliF   GuidoCrosetto          vox_es        LaStampa   Santi_ABASCAL
 >              10              10               9               8               8
```

The same operations are performed on the other two profiles, in the following lines the results for Calenda profile are shown.

```
#CALENDA
tlCC <- get_timeline("CarloCalenda", 2000,token = auth)
summary(tlCC$created_at)
 >                   Min.                1st Qu.                  Median
 > "2021-06-08 09:10:30" "2021-06-27 17:37:41" "2021-07-21 14:46:05"
 >                   Mean                3rd Qu.                   Max.
 > "2021-07-21 15:18:16" "2021-08-15 09:36:41" "2021-09-01 16:44:25"
table(tlCC$is_retweet)
 >
 > FALSE  TRUE
 >  1431   569
table(tlCC$is_quote)
 >
 > FALSE  TRUE
 >  1488   512

sort(table(unlist(tlCC$hashtags)),decreasing = T)[1:10]
 >
 >   RomaSulSerio CalendaSindaco           Roma         Azione       Michetti
 >            106             37             27             18              9
 >          Raggi      Pride2021        rifiuti         Draghi        Calenda
 >              7              6              6              5              4
sort(table(unlist(tlCC$mentions_screen_name)),decreasing = T)[1:10]
 >
```

```
>    CarloCalenda  CalendaSindaco       Azione_it   virginiaraggi gualtierieurope
>             163             137              80              47              43
>   Enrico__Costa  MatteoRichetti  EnricoMichetti    diarioromano     elevisconti
>              28              25              22              17              15
```

For what concerns Matteo Renzi's account the following is the data collected:

```
#RENZI
tlMR <- get_timeline("matteorenzi", 2000, token = auth)
summary(tlMR$created_at)
>                  Min.               1st Qu.                Median
> "2020-03-19 22:40:52" "2020-06-15 16:46:35" "2020-11-23 14:43:05"
>                  Mean               3rd Qu.                  Max.
> "2020-11-15 02:05:21" "2021-03-24 08:24:18" "2021-09-01 17:27:35"
table(tlMR$is_retweet)
>
> FALSE  TRUE
>  1150   848
table(tlMR$is_quote)
>
> FALSE  TRUE
>  1963    35

sort(table(unlist(tlMR$hashtags)),decreasing = T)[1:10]
>
>           RenziRep  LaMossaDelCavallo              Enews     ControCorrente
>                 57                 47                 42                 39
>         ItaliaViva          RenziCorr VisioneNonRimpasto      RenziCorriere
>                 30                 25                 24                 19
>                Mes            RenziQN
>                 18                 13
sort(table(unlist(tlMR$mentions_screen_name)),decreasing = T)[1:10]
>
>      ItaliaViva TeresaBellanova        JoeBiden        lisanoja  RobertoBurioni
>             137              50              45              40              31
>    elenabonetti      sandrogozi     matteorenzi  raffaellapaita        marattin
>              29              27              26              26              25
```

To better visualize the differences for the three accounts, the past results obtained individually for each user can be combined together as follows:

```
print(c("RETWEETS",table(c(tlMR$screen_name, tlMR$is_retweet)),
       table(c(tlGM$screen_name, tlGM$is_retweet )),
       table(c(tlCC$screen_name, tlCC$is_retweet ))))
>                      FALSE   matteorenzi          TRUE         FALSE
>    "RETWEETS"        "1150"        "1998"         "848"        "1706"
> GiorgiaMeloni          TRUE  CarloCalenda         FALSE          TRUE
>        "2000"         "294"        "2000"        "1431"         "569"

print(c("QUOTES", table(c(tlMR$is_quote, tlMR$screen_name)),
       table(c(tlGM$is_quote, tlGM$screen_name)),
       table(c(tlCC$is_quote, tlCC$screen_name))))
>                      FALSE   matteorenzi          TRUE         FALSE
```

```
 >      "QUOTES"        "1963"        "1998"          "35"        "1987"
 > GiorgiaMeloni          TRUE  CarloCalenda         FALSE          TRUE
 >        "2000"          "13"        "2000"        "1488"         "512"

print(c("HASHTAGS", sort(table(unlist(tlMR$hashtags)),decreasing = T)[1:5],
        sort(table(unlist(tlGM$hashtags)),decreasing = T)[1:5],
      sort(table(unlist(tlCC$hashtags)),decreasing = T)[1:5]))
 >                          RenziRep LaMossaDelCavallo             Enews
 >        "HASHTAGS"             "57"              "47"             "42"
 >    ControCorrente        ItaliaViva   FratellidItalia           Meloni
 >              "39"             "30"             "136"             "74"
 >   BastaCoprifuoco      BloccoNavale      BastaSbarchi      RomaSulSerio
 >              "28"             "17"              "15"             "106"
 >    CalendaSindaco             Roma            Azione          Michetti
 >              "37"             "27"              "18"              "9"

print(c("MENTIONS",sort(table(unlist(tlMR$mentions_screen_name)),decreasing = T)[1:5],
        sort(table(unlist(tlGM$mentions_screen_name)),decreasing = T)[1:5],
        sort(table(unlist(tlCC$mentions_screen_name)),decreasing = T)[1:5]))
 >                     ItaliaViva TeresaBellanova        JoeBiden        lisanoja
 >      "MENTIONS"           "137"            "50"            "45"            "40"
 >  RobertoBurioni FratellidItalia   GiorgiaMeloni        ECRparty      BrunoVespa
 >            "31"           "280"           "157"            "26"            "11"
 >        Corriere    CarloCalenda  CalendaSindaco       Azione_it   virginiaraggi
 >            "11"           "163"           "137"            "80"            "47"
 > gualtierieurope
 >            "43"
```

> **Hashtags** An other interesting thing to see is the possibility of having some common hashtags between the two users, here it can be seen that the three accounts have few common hashtags.

```
vsUsr <- c("GiorgiaMeloni","CarloCalenda","matteorenzi")
lstHS <- list()
nHS <- numeric()
uHS <- numeric()
for(i in 1:3){
  tmp <- tlMCRr[tlMCRr$screen_name==vsUsr[i],"hashtags"]
  tmp <- unlist(tmp,use.names = F)
  tmp <- na.omit(tmp)
  nHS[i] <- length(tmp)
  uHS[i] <- length(unique(tmp))
  lstHS[[i]] <- tmp
}
names(lstHS) <- vsUsr
names(nHS) <- vsUsr
names(uHS) <- vsUsr
cbind(Hashtag.Tot=nHS,Hashtag.Univoci=uHS)
 >               Hashtag.Tot Hashtag.Univoci
 > GiorgiaMeloni        1130             501
 > CarloCalenda          449             191
 > matteorenzi          1348             576

library(purrr)
```

```
 >
 > Attaching package: 'purrr'
 > The following object is masked from 'package:ndjson':
 >
 >     flatten
 > The following object is masked from 'package:rtweet':
 >
 >     flatten
Reduce(intersect,list(lstHS[[1]],lstHS[[2]],lstHS[[3]]))
 >  [1] "Conte"         "BebeVio"       "Afghanistan"   "Meloni"
 >  [5] "Tg2Post"       "Azzurri"       "Covid19"       "StaseraItalia"
 >  [9] "PNRR"          "Draghi"        "scuola"        "Senato"
 > [13] "M5S"          "Salvini"
```

It is also possible to exclude one of the three accounts to see whether considering just two of them they have more common hashtags. The first result stands for the hashtags used by both G. Meloni and M. Renzi; The second one is the one obtained excluding M. Renzi, thus just G. Meloni and C. Calenda; The last one is the number of hashtags used by C. Calenda and M. Renzi. We can see that there is a noticeable difference between the number of common hashtags between G. Meloni and M. Renzi with respect to the two combinations in which C. Calenda is present, indeed it is more than twice of the numbers obtained from the other combinations.

```
Reduce(intersect,list(lstHS[[1]],lstHS[[3]]))
 >  [1] "Conte"                 "BebeVio"               "Afghanistan"
 >  [4] "terremoto"             "Biden"                 "Calabria"
 >  [7] "Meloni"                "Sardegna"              "Arcuri"
 > [10] "Taekwondo"             "Tg2Post"               "PaoloBorsellino"
 > [13] "19luglio"              "Germania"              "Belgio"
 > [16] "Azzurri"               "VivoAzzurro"           "EURO2020"
 > [19] "ForzaAzzurri"          "Covid19"               "Mugello"
 > [22] "AstraZeneca"           "WillyMonteiro"         "lariachetira"
 > [25] "Carabinieri"           "portaaporta"           "QuartaRepubblica"
 > [28] "StaseraItalia"         "CarlaFracci"           "Capaci"
 > [31] "GiovanniFalcone"       "Battiato"              "mezzorainpiù"
 > [34] "BuonaDomenica"         "PortaaPorta"           "PrimoMaggio"
 > [37] "ZonaBianca"            "PNRR"                  "RecoveryPlan"
 > [40] "Marche"                "Draghi"                "scuola"
 > [43] "16marzo"               "AldoMoro"              "Sanremo2021"
 > [46] "Covid"                 "Congo"                 "LunaRossaPradaPirelli"
 > [49] "AmericasCup"           "BuongiornoATutti"      "Camera"
 > [52] "fiducia"               "Foibe"                 "10febbraio"
 > [55] "Giornatadellamemoria"  "Senato"                "Grillo"
 > [58] "COVID19"               "redditodicittadinanza" "cashback"
 > [61] "PaoloRossi"            "Report"                "patrimoniale"
 > [64] "imprese"               "RecoveryFund"          "M5S"
 > [67] "Maradona"              "leggedibilancio"       "Covid_19"
 > [70] "clickday"              "Vienna"                "lockdown"
 > [73] "Nizza"                 "DecretiSicurezza"      "coronavirus"
 > [76] "Mose"                  "Venezia"               "INPS"
 > [79] "Toscana"               "primogiornodiscuola"   "Puglia"
 > [82] "Bari"                  "Salvini"               "COVID<U+30FC>19"

Reduce(intersect,list(lstHS[[1]],lstHS[[2]]))
```

```
 >  [1] "Speranza"           "Conte"              "BebeVio"
 >  [4] "Afghanistan"        "Kabul"              "Tokyo2020"
 >  [7] "Meloni"             "greenpass"          "Tg2Post"
 > [10] "Azzurri"            "ItaliaInghilterra"  "Euro2020Final"
 > [13] "Wimbledon"          "BerrettiniDjokovic" "Berrettini"
 > [16] "Euro2020"           "Spinazzola"         "RaffaellaCarrà"
 > [19] "Covid19"            "Michetti"           "Ardea"
 > [22] "Matone"             "StaseraItalia"      "Anni20"
 > [25] "PNRR"               "Draghi"             "scuola"
 > [28] "politica"           "Senato"             "M5S"
 > [31] "11settembre"        "Salvini"
```

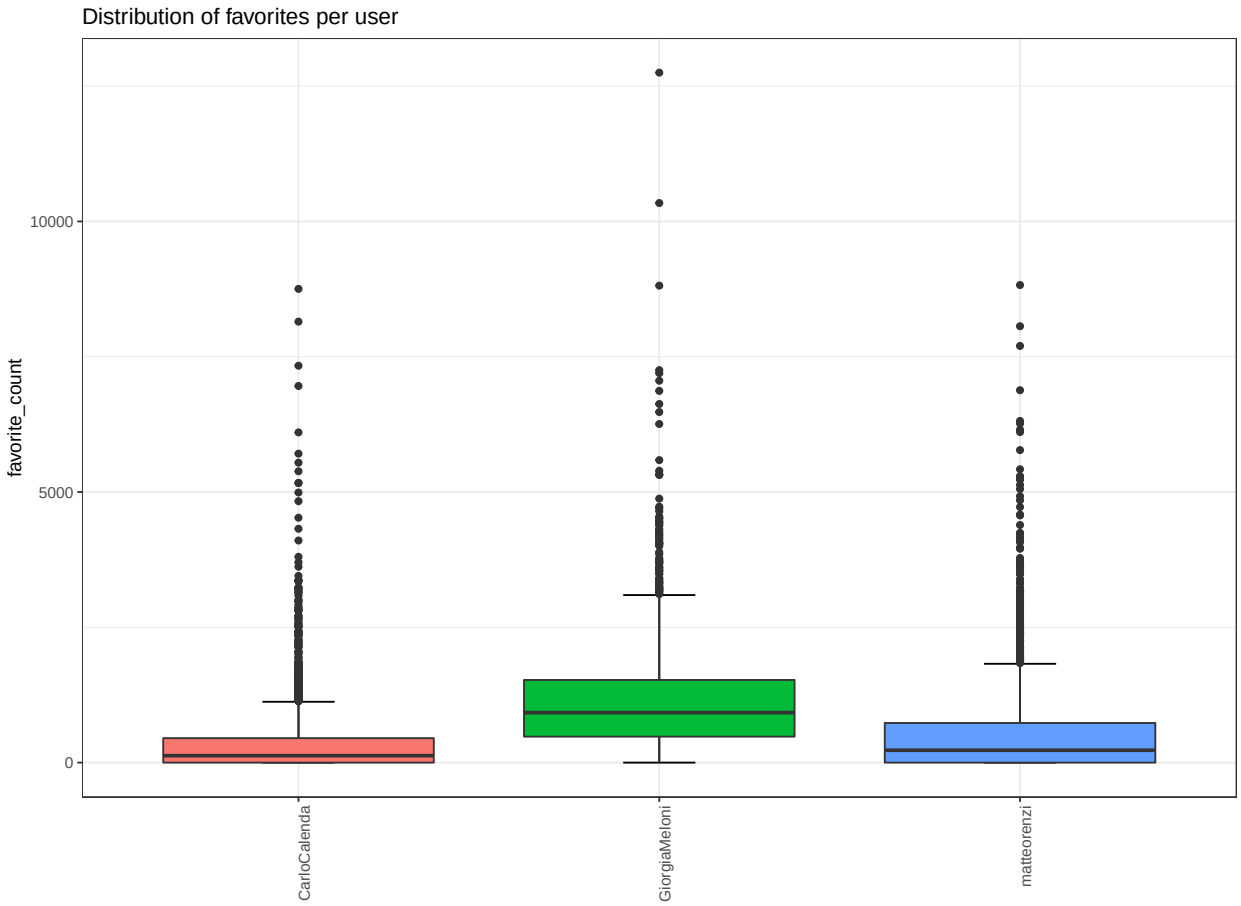
```
Reduce(intersect,list(lstHS[[2]],lstHS[[3]]))
 >  [1] "RomaSulSerio"   "NoVax"          "Taverna"        "Conte"
 >  [5] "Controcorrente" "BebeVio"        "Raggi"          "Roma"
 >  [9] "Afghanistan"    "vaccini"        "scuola"         "Draghi"
 > [13] "Meloni"         "Mattarella"     "ddlZan"         "Gualtieri"
 > [17] "siamoInOnda"    "Salvini"        "InOnda"         "Travaglio"
 > [21] "Senato"         "Quota100"       "Azzurri"        "Milano"
 > [25] "cultura"        "Covid19"        "DdlZan"         "M5S"
 > [29] "StaseraItalia"  "PNRR"           "Tg2Post"        "DDLZan"
 > [33] "DeLuca"         "Emiliano"       "UE"             "G7"
 > [37] "turismo"        "antiEuro"
```

> **Favorites** An important feature that can be observed from the three profiles is the number of favorites given by the users to the tweets of the selected accounts. Using the object created previously in which are collected 2000 tweets for each account starting from 2020-01-01. In the following table it can be seen how many times those tweets have been indicated as favorites by other users.

```
library(dplyr)
tbFav <- tlMCRr %>%
  group_by(screen_name) %>%
  summarise(n=n(),min=min(favorite_count),max=max(favorite_count),
            totFavoriti=sum(favorite_count),
            media=mean(favorite_count))
tbFav
 > # A tibble: 3 x 6
 >   screen_name       n   min   max totFavoriti media
 >   <chr>         <int> <int> <int>       <int> <dbl>
 > 1 CarloCalenda   2000     0  8754      804176  402.
 > 2 GiorgiaMeloni  2000     0 12750     2261266 1131.
 > 3 matteorenzi    1998     0  8825     1157837  579.
```

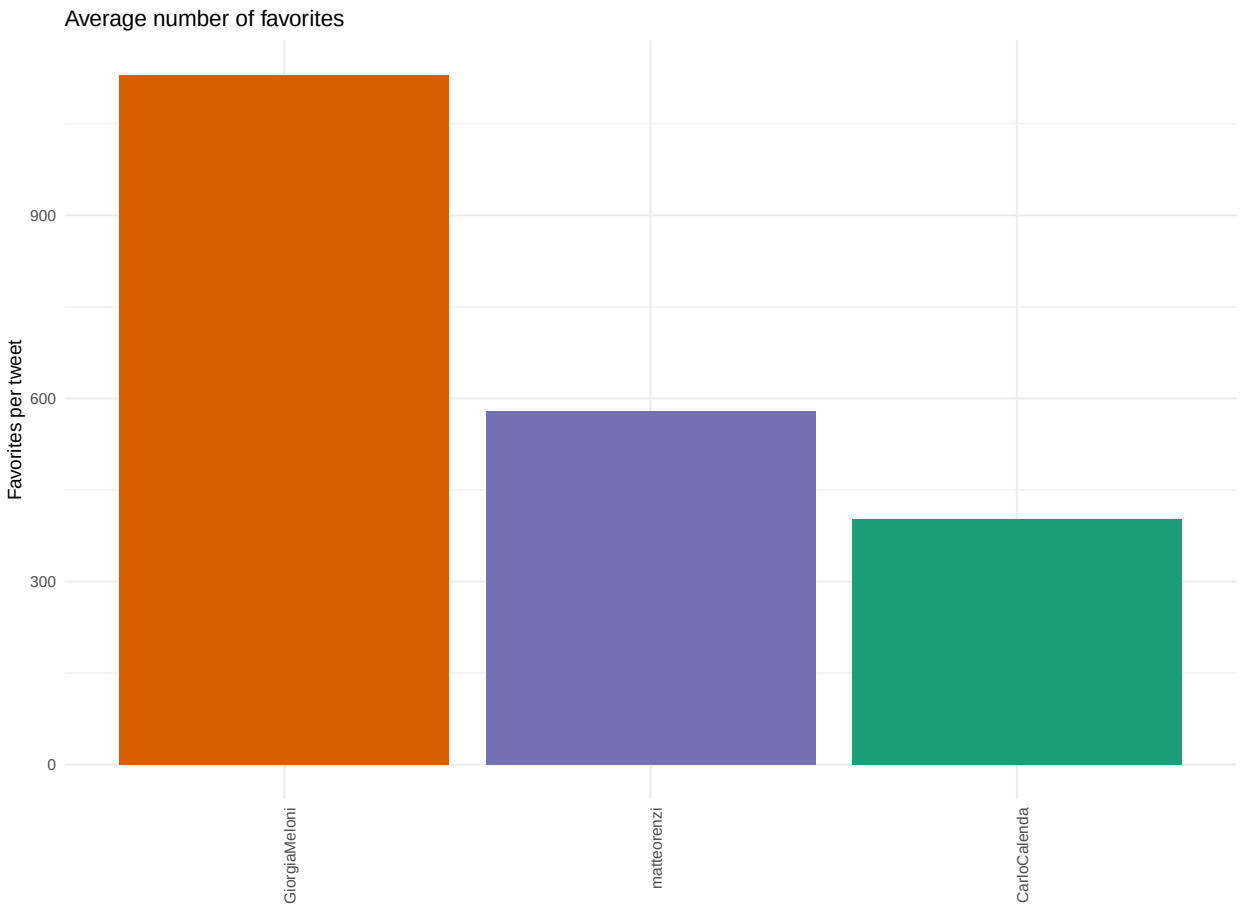
The following plot shows the distribution of the number of favorites for each account. The account reaching a higher number is the one of G. Meloni, whereas the other two profiles have a similar distribution.

```
ggplot(tlMCRr, aes(x=screen_name, y=favorite_count,fill = screen_name)) +
  stat_boxplot(geom = "errorbar", width = 0.2) +
  geom_boxplot() +
  theme_bw()+xlab("") + ggtitle("Distribution of favorites per user") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(legend.position = "none")
```



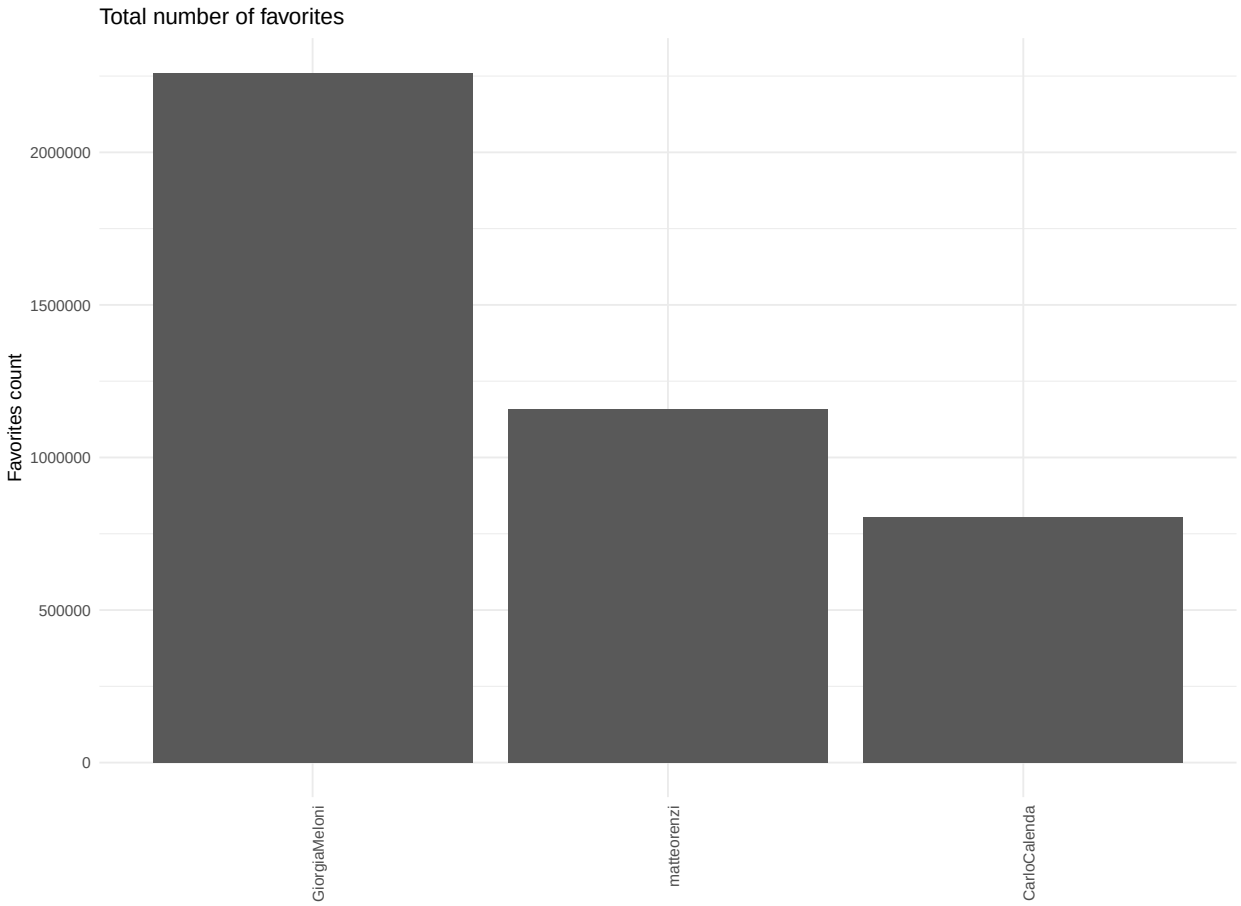
Indeed, it is possible to plot the average number of favorites per account that shows a behavior similar to the one of the distribution.

```
ggplot(tbFav, aes(x=reorder(screen_name,-media), y=media,
                  fill=screen_name))+
  geom_bar(stat="identity")+
  theme_minimal()+xlab("")+ylab("Favorites per tweet")+
  ggtitle("Average number of favorites") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  scale_fill_brewer(palette = "Dark2") + theme(legend.position = "none")
```



The graph representing the total number of tweets has the same features of the previous plots shown before, confirming that G. Meloni has a high number of favorites for many tweets, and the other two profile have a number of favorites that does not differ significantly. It is also remarkable that C. Calenda has a number of followers extremely small with respect to M. Renzi (3K wrt 3M).

```
ggplot(tbFav, aes(x=reorder(screen_name,-totFavoriti), y=totFavoriti))+
  geom_bar(stat="identity")+
  theme_minimal()+xlab("")+ylab("Favorites count")+
  ggtitle("Total number of favorites") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  scale_fill_brewer(palette = "Dark2") + theme(legend.position = "none")
```



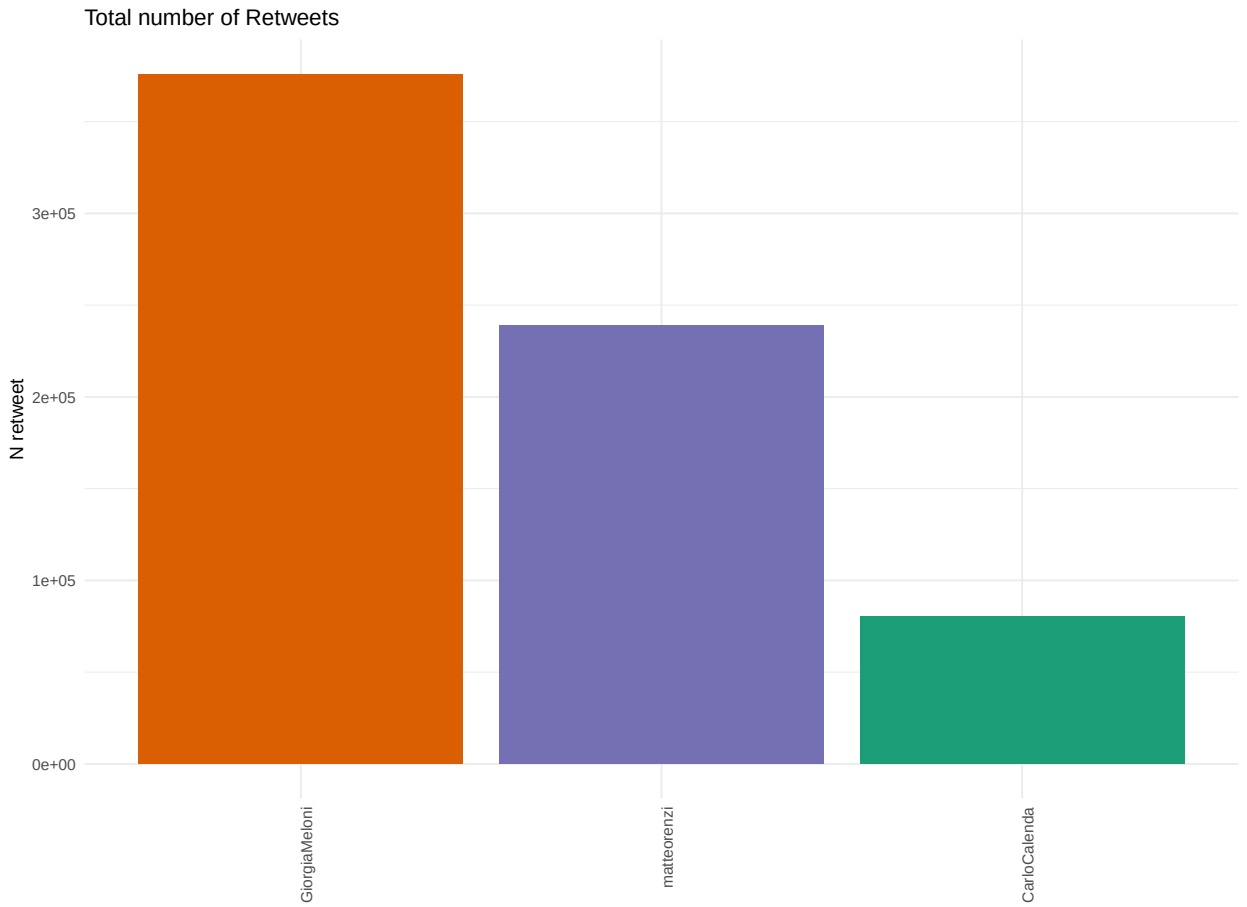
> **Retweets** Retweets are a considerable indicator to understand the differences between different Twitter's profiles. In this case it can be seen how many times the tweets of the three selected accounts have been retweeted by other users. As it can be seen from the following table, the number of times Calenda's tweets have been retweeted is very low with respect to the other two politicians.

```
tbRT <- tlMCRr[tlMCRr$is_retweet==F,] %>%
group_by(screen_name) %>%
summarise(n=n(),min=min(retweet_count),max=max(retweet_count),
            totFavoriti=sum(retweet_count),
            media=mean(retweet_count))
tbRT
 > # A tibble: 3 x 6
 >   screen_name       n   min   max totFavoriti media
 >   <chr>         <int> <int> <int>       <int> <dbl>
 > 1 CarloCalenda   1431     0  1082       80192  56.0
 > 2 GiorgiaMeloni  1706     6  2032      376007 220.
 > 3 matteorenzi    1150     0  1724      238973 208.
```

In this plot it can be seen the total number of retweets for each account.

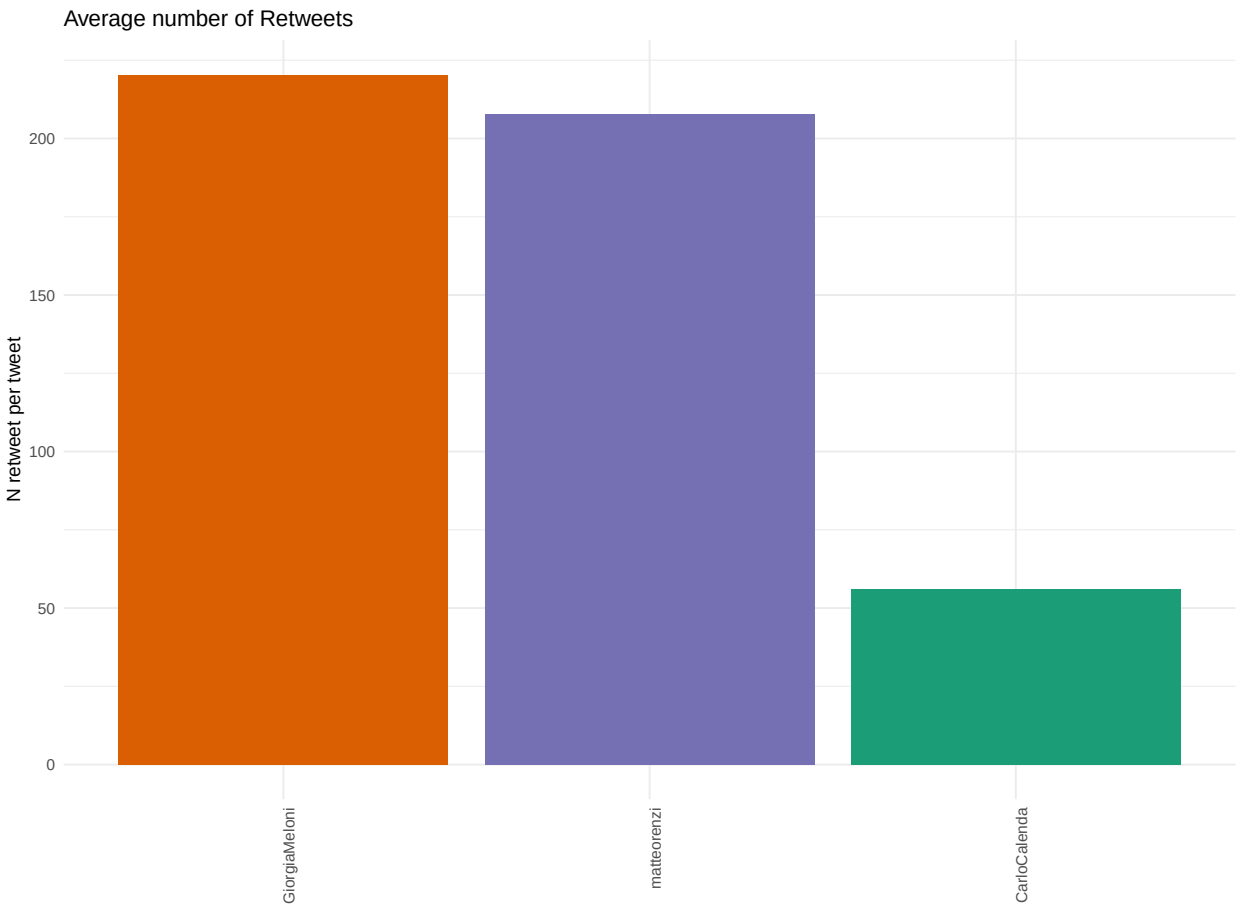
```
ggplot(tbRT, aes(x=reorder(screen_name, -totFavoriti), y=totFavoriti,
                 fill=screen_name)) +
```

```
  geom_bar(stat="identity")+
  theme_minimal()+xlab("")+ylab("N retweet")+
  ggtitle("Total number of Retweets") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  scale_fill_brewer(palette = "Dark2") + theme(legend.position = "none")
```



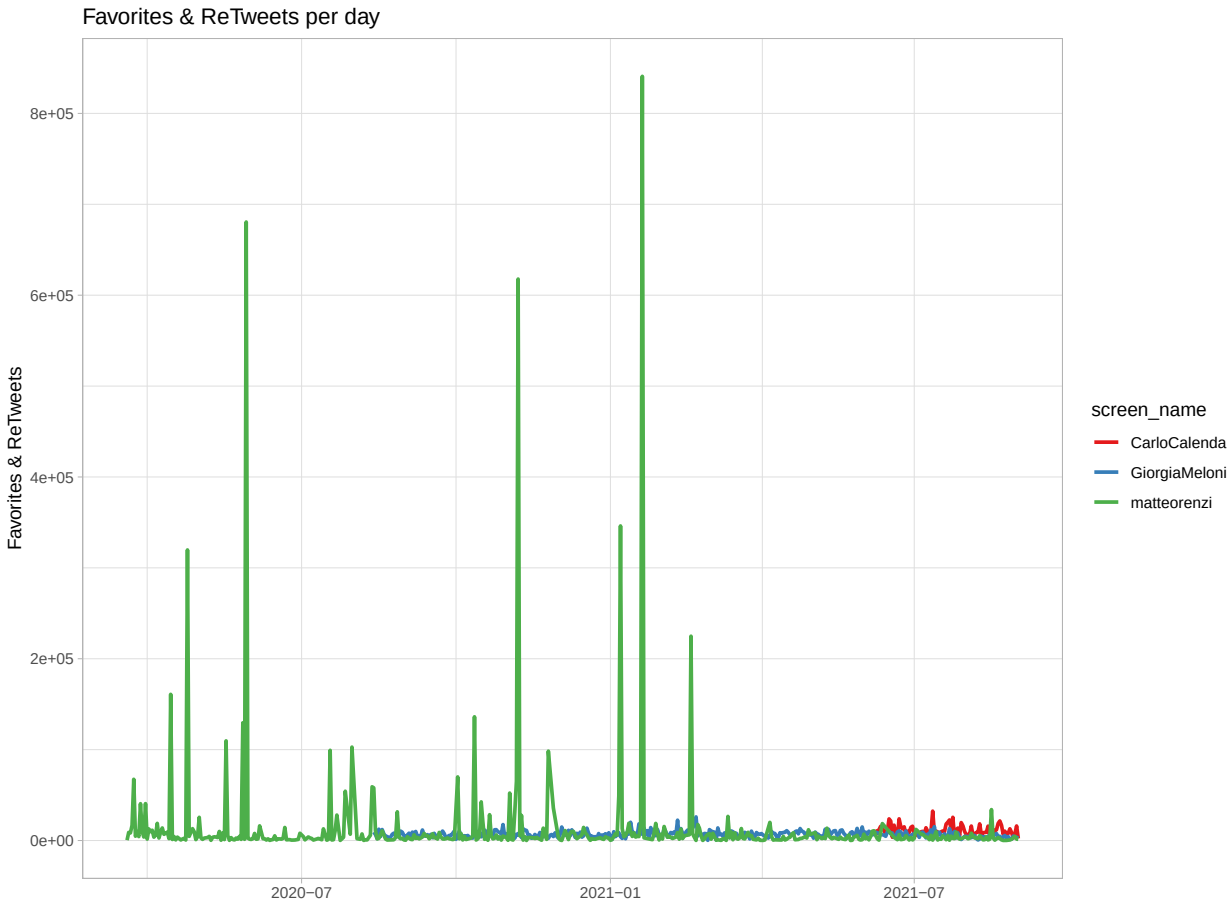
For what concerns the average number of retweets the shape of the graph changes with respect to the one representing the total number of retweets. Indeed, the average number is similar for G. Meloni and M. Renzi meaning that M. Renzi seems to receive a constant number of tweets with respect to M. Meloni's tweets.

```
ggplot(tbRT, aes(x=reorder(screen_name, -media), y=media,
                 fill=screen_name)) +
  geom_bar(stat="identity")+
  theme_minimal()+xlab("")+ylab("N retweet per tweet")+
  ggtitle("Average number of Retweets") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  scale_fill_brewer(palette = "Dark2") + theme(legend.position = "none")
```



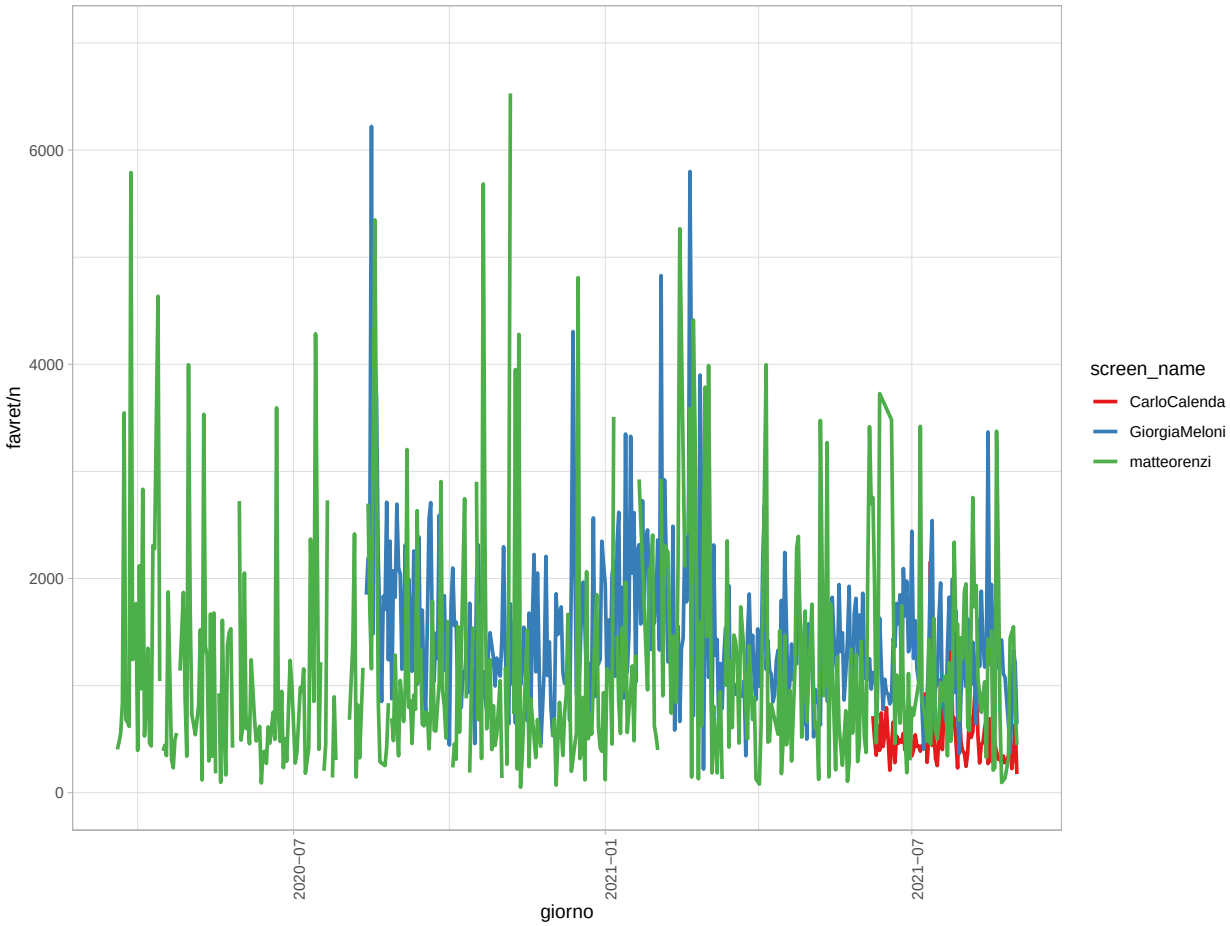
> **Favorites & Retweets** Putting together favorites and retweets received by the three accounts we can build an interesting plot that shows how these two indicators have been changing during past few months. Indeed, M. Renzi has reached higher numbers but paying attention to the right part of the graph it can be seen that C. Calenda was not so active in publishing as the other two users until 2021-06.

```
tab2 <- tlMCRr[,c("screen_name","created_at","favorite_count","retweet_count")] %>%
  group_by(screen_name,giorno=as.Date(created_at)) %>%
  summarise(favoriti=sum(favorite_count),retweet=sum(retweet_count),n=n())
 > `summarise()` has grouped output by 'screen_name'. You can override using the `.groups` argument.

tab2$favret <- tab2$favoriti+tab2$retweet

ggplot(tab2, aes(x=giorno, y=favret, group=screen_name)) +
  geom_line(aes(color=screen_name),size= 1)+
  scale_color_manual(values=brewer.pal(9, "Set1")[c(1:5,9)])+
  theme_light()+ggtitle("Favorites & ReTweets per day") + xlab("") +
  ylab("Favorites & ReTweets")
```



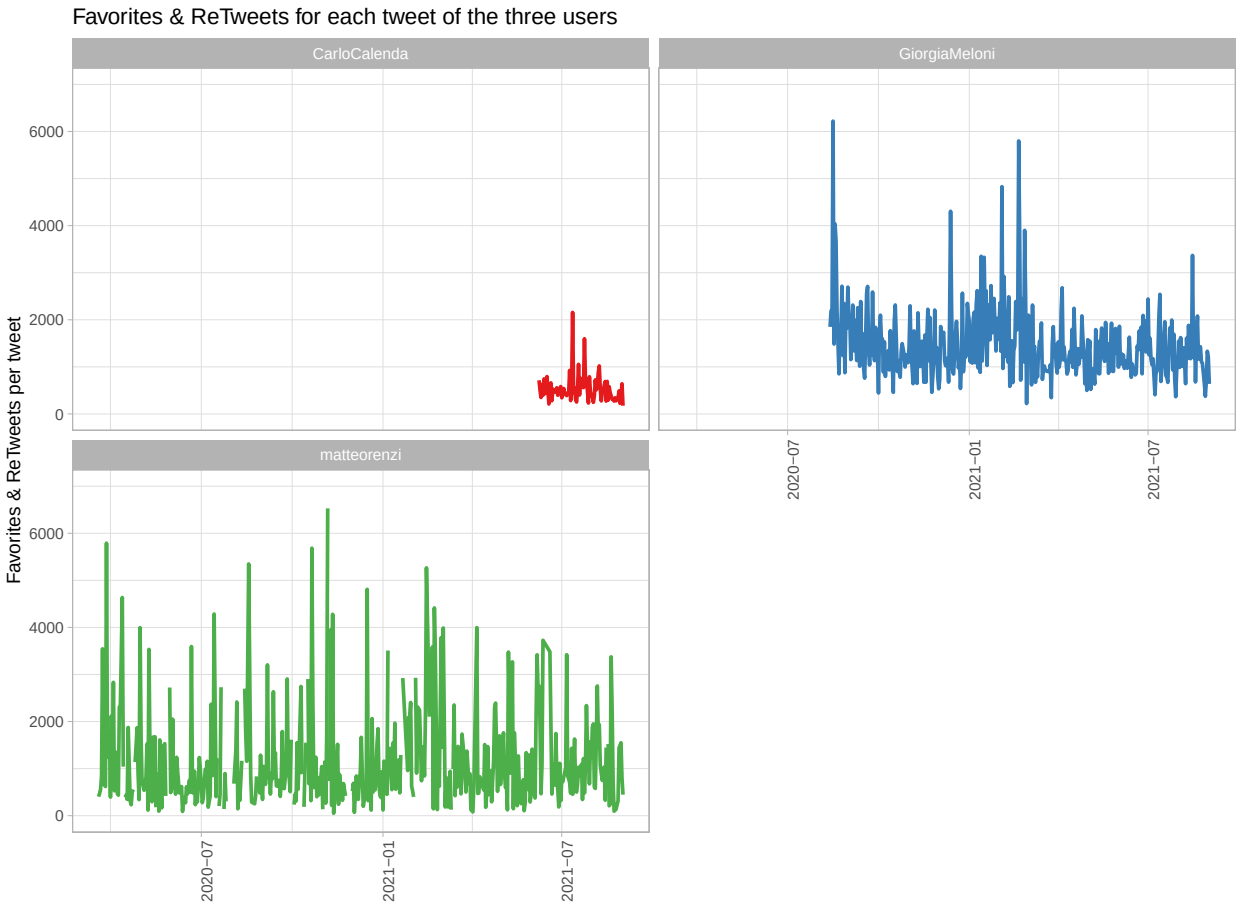
Here a different way to visualize the same result of before.

```
p1 <- ggplot(tab2, aes(x=giorno, y=favret/n)) +
  geom_line(aes(color=screen_name),size= 1)+
  ylim(0,7000) +
  scale_color_manual(values=brewer.pal(9, "Set1")[c(1:5,9)])+
  theme_light() + theme(axis.text.x = element_text(angle = 90, hjust = 1))
p1
```



To have a cleaner view for the three profiles we can show them individually as follows.

```
p1 + facet_wrap( ~ screen_name, nrow = 2) +
  theme(legend.position = "none") +
  ggtitle("Favorites & ReTweets for each tweet of the three users") +
  xlab("") + ylab("Favorites & ReTweets per tweet")
```



> **Type of tweet** From the 2000 tweets collected from the three timelines of the accounts we can also divide the tweets to see how many of them are retweets, both from an absolute point of view and from a relative point of view(%).

```
addmargins(table(tlMCRr$screen_name,tlMCRr$is_retweet))
 >
 >                FALSE TRUE  Sum
 >   CarloCalenda  1431  569 2000
 >   GiorgiaMeloni 1706  294 2000
 >   matteorenzi   1150  848 1998
 >   Sum           4287 1711 5998

round(addmargins(prop.table(addmargins(table(tlMCRr$screen_name,tlMCRr$is_retweet),1),1),2)*100,1)
 >
 >                FALSE  TRUE   Sum
 >   CarloCalenda  71.6  28.4 100.0
 >   GiorgiaMeloni 85.3  14.7 100.0
 >   matteorenzi   57.6  42.4 100.0
 >   Sum           71.5  28.5 100.0
```

Plotting these results it can be seen that G. Meloni is the account that publishes more original tweets while M. Renzi is the one publishing more retweets (almost reaching 50%)

```r
tb <- as.data.frame(table(tlMCRr$is_retweet,tlMCRr$screen_name))
ggplot(tb,aes(x=Var2,y=Freq, fill=Var1))+
  geom_bar(position = "fill", stat = "identity") +
  scale_fill_brewer(palette="Dark2") +
  theme_minimal() + xlab("")+ylab("")+
  ggtitle("Distribution of Retweets") +
  labs(fill = "is_retweet")
```