

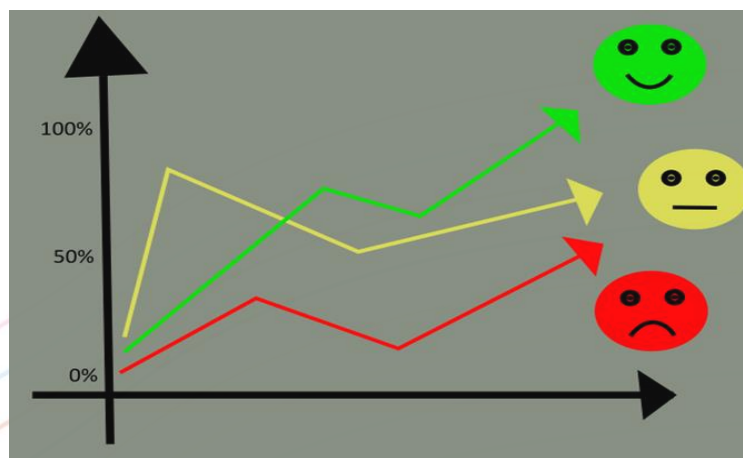


crea

Consiglio per la ricerca in agricoltura
e l'analisi dell'economia agraria

Centro di politiche e Bioeconomia
Ufficio di Statistica

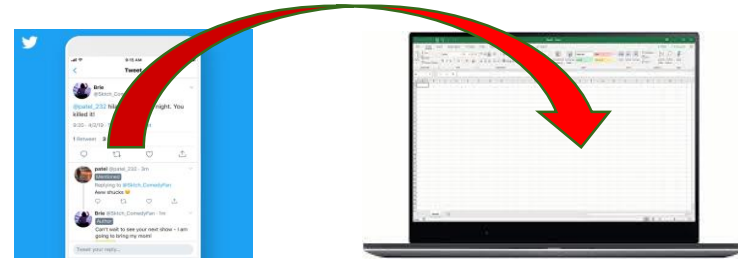
Come fare Sentiment con R(*studio*)



Giuliano Gabrieli

Roma, 19 febbraio 2019

1) Estrazione dati da Twitter



2) Pulizia del testo

Olivi veneti a rischio disseccamento. Progetto regionale di monitoraggio con @UniPadova <https://t.co/a28VyV8Oyn> di @AgricolturaIT #Veneto #venetoAgricoltura

Olivi veneti 🇮🇹 rischio disseccamento🇮🇹
Progetto regionale 🇮🇹 monitoraggio 🇮🇹
~~@UnifaVda - <https://t.me/a20V1V00yt> -~~
~~@agricultura27 -~~ ~~#Veneto~~
~~#VenetoAgricoltura~~

**olivi veneti rischio
disseccamento progetto regionale
monitoraggio**

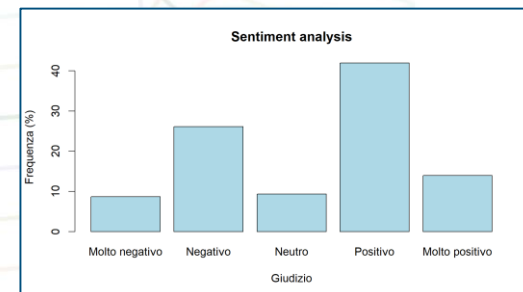
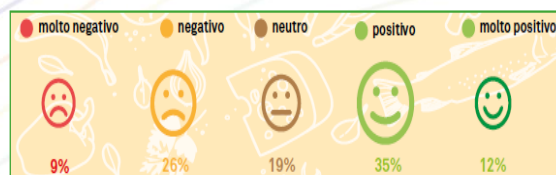
3) Sentiment analysis



WORDCLOUD



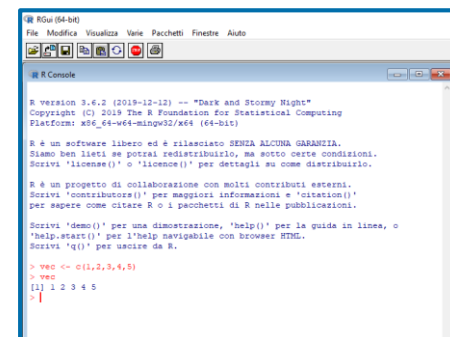
4) Rappresentazione dei risultati





- **R** è un software open source dedicato all'elaborazione statistica dei dati e alla grafica

<https://cran.r-project.org>



R console

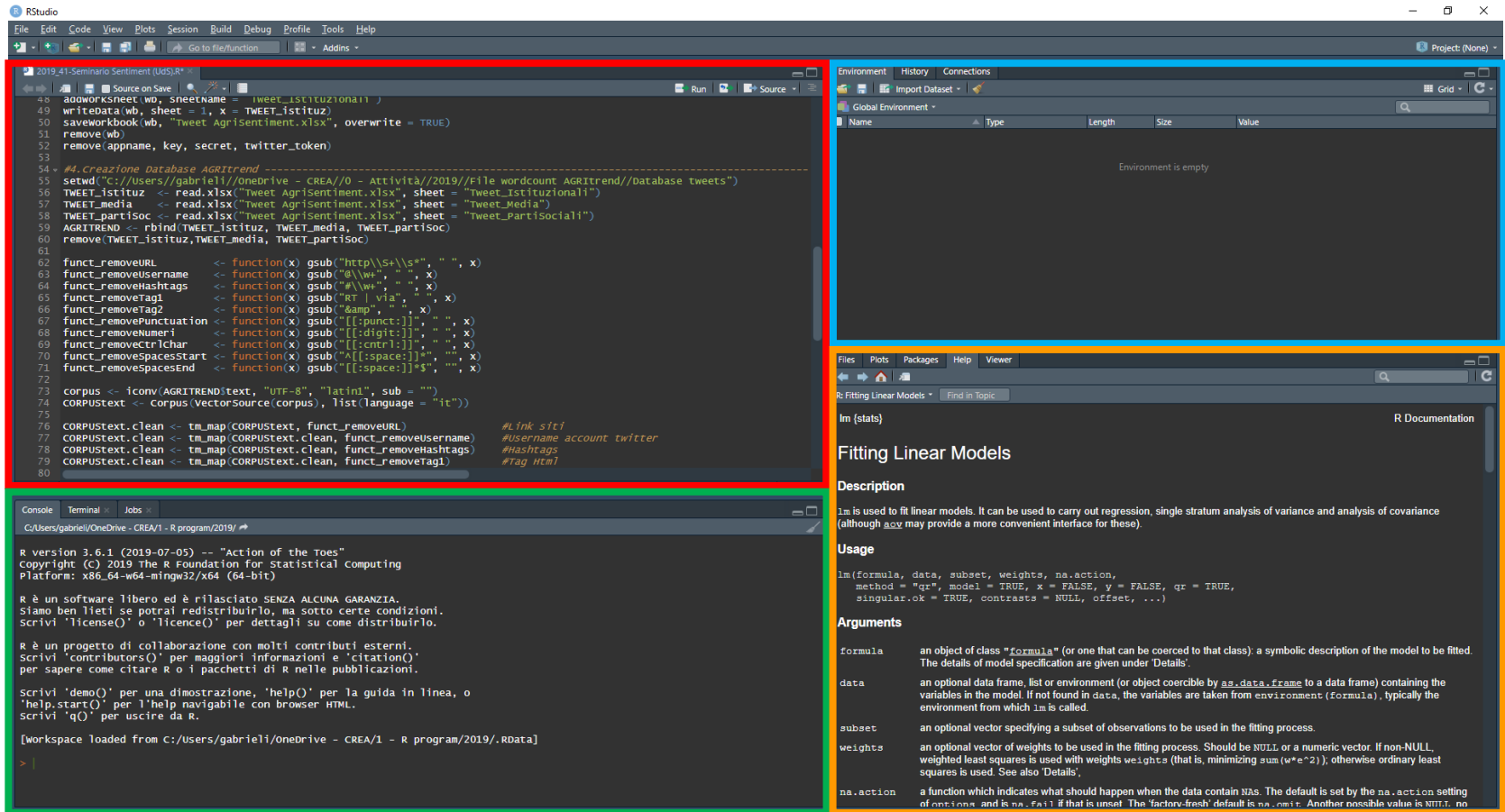


- **Rstudio**, un ambiente di sviluppo integrato (IDE: integrated development environment) open source, che introduce numerose funzionalità in R

<https://rstudio.com>

Editor

Environment, History



The screenshot displays the RStudio interface with four main panes:

- Editor:** Contains R code for data manipulation and analysis. The code includes:


```

48 addWorksheet(wb, sheetname = "tweet_istituzionale")
49 writeData(wb, sheet = 1, x = TWEET_istituz)
50 saveWorkbook(wb, "Tweet_AgrISentiment.xlsx", overwrite = TRUE)
51 remove(wb)
52 remove(appname, key, secret, twitter_token)
53
54 #4.Creazione Database AGRIrentrend -----
55 setwd("C:/Users/gabriele/OneDrive - CREA/0 - Attività/2019/File wordcount AGRIrentrend/Database tweets")
56 TWEET_istituz <- read.xlsx("Tweet_AgrISentiment.xlsx", sheet = "Tweet_istituzionale")
57 TWEET_media <- read.xlsx("Tweet_AgrISentiment.xlsx", sheet = "Tweet_Media")
58 TWEET_partisoc <- read.xlsx("Tweet_AgrISentiment.xlsx", sheet = "Tweet_Partisociali")
59 AGRIrentrend <- rbind(TWEET_istituz, TWEET_media, TWEET_partisoc)
60 remove(TWEET_istituz, TWEET_media, TWEET_partisoc)
61
62 funct_removeURL <- function(x) gsub("http[s+]?://.*", "", x)
63 funct_removeUsername <- function(x) gsub("@\\w+", "", x)
64 funct_removeHashtags <- function(x) gsub("#\\w+", "", x)
65 funct_removeTag1 <- function(x) gsub("RT|via", "", x)
66 funct_removeTag2 <- function(x) gsub("&#", "", x)
67 funct_removePunctuation <- function(x) gsub("[[:punct:]]", "", x)
68 funct_removeNumeri <- function(x) gsub("[[:digit:]]", "", x)
69 funct_removeCtrlChar <- function(x) gsub("[[:cntrl:]]", "", x)
70 funct_removeSpacesStart <- function(x) gsub("^[:space:]", "", x)
71 funct_removeSpacesEnd <- function(x) gsub("[:space:]+$", "", x)
72
73 corpus <- iconv(AGRIrentrend$text, "UTF-8", "latin1", sub = "")
74 CORPUS$text <- corpus(VectorSource(corpus), list(Language = "it"))
75
76 CORPUS$text.clean <- tm_map(CORPUS$text, funct_removeURL) #Link siti
77 CORPUS$text.clean <- tm_map(CORPUS$text.clean, funct_removeUsername) #Username account twitter
78 CORPUS$text.clean <- tm_map(CORPUS$text.clean, funct_removeHashtags) #Hashtags
79 CORPUS$text.clean <- tm_map(CORPUS$text.clean, funct_removeTag1) #Tag HTML
80

```
- Environment:** Shows the Global Environment with a table of variables. The table is empty, indicating no objects are currently in the environment.
- History:** Shows the execution history of the R script, including the same code as the Editor pane.
- Console:** Displays the R version (3.6.1) and the workspace loaded from the file "C:/Users/gabriele/OneDrive - CREA/1 - R program/2019/.Rdata".
- Files, Plots, Packages, Help, Viewer:** The top bar contains tabs for these functions. The Help tab is active, showing the documentation for the `lm` function.

Console

Files, Plot, Packages, Help

Controllo errori e auto-completamento del testo durante la scrittura del codice (Funzioni, Oggetti)

Editor

The screenshot shows the RStudio Editor with the following code:

```

60
61 table(DBsoglie%Giudizio)
62
63 tab
64
65 table {base}
66 tableGrob {gridExtra}
67 tabulate {base}
68
69
70 table(f1_)
71
72 F1_ALL.GEN.neg
73 F1_ALL.GEN.pos
74 F1_ALL.neg
75 F1_ALL.POL.neg
76 F1_ALL.POL.pos
77 F1_ALL.pos
78
79 F1_Score {MLmetrics}
80 F1_TEST.GEN.neg

```

Annotations and features shown:

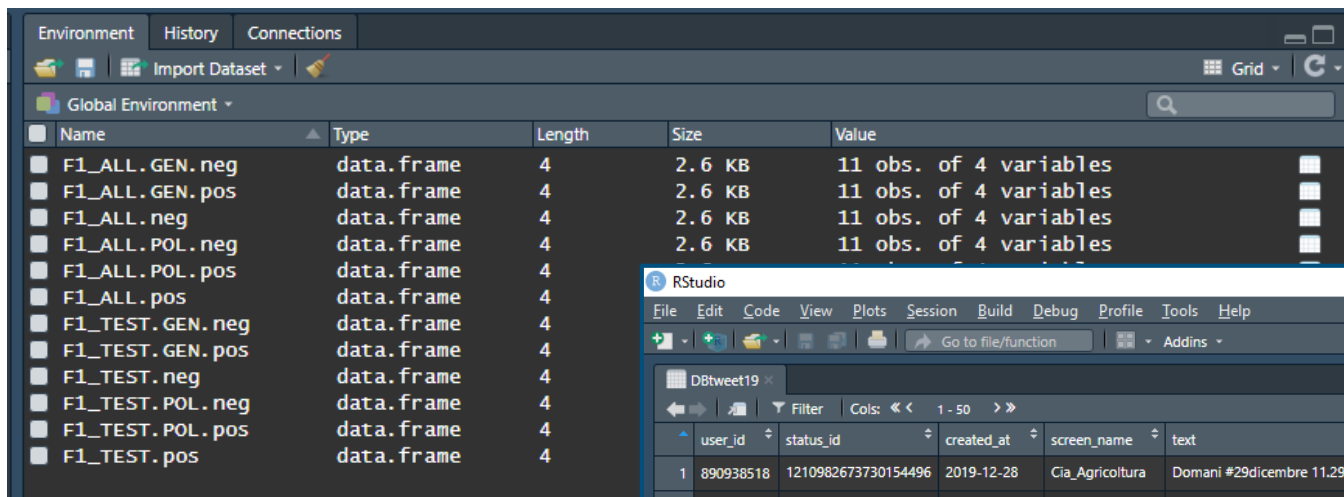
- ERRORI**: A green arrow points to a green squiggly line under the dollar sign (\$) in line 61, indicating a syntax error.
- FUNZIONI**: A yellow arrow points to the auto-completion dropdown menu for the word "tab" in line 63. The menu lists "table", "tableGrob", and "tabulate".
- OGGETTI (vettori, dataframe, liste,)**: An orange arrow points to the auto-completion dropdown menu for the argument "f1_" in line 70. The menu lists various objects like "F1_ALL.GEN.neg", "F1_ALL.GEN.pos", etc.

The console at the bottom shows the file path: `G:/Users/gabrieli/OneDrive - CREA/0 - Attività/2020/Punto a1 - Sentipolc (Test MAL)/Output data/`.

Pieno controllo del nostro Workspace attraverso un'unica schermata:

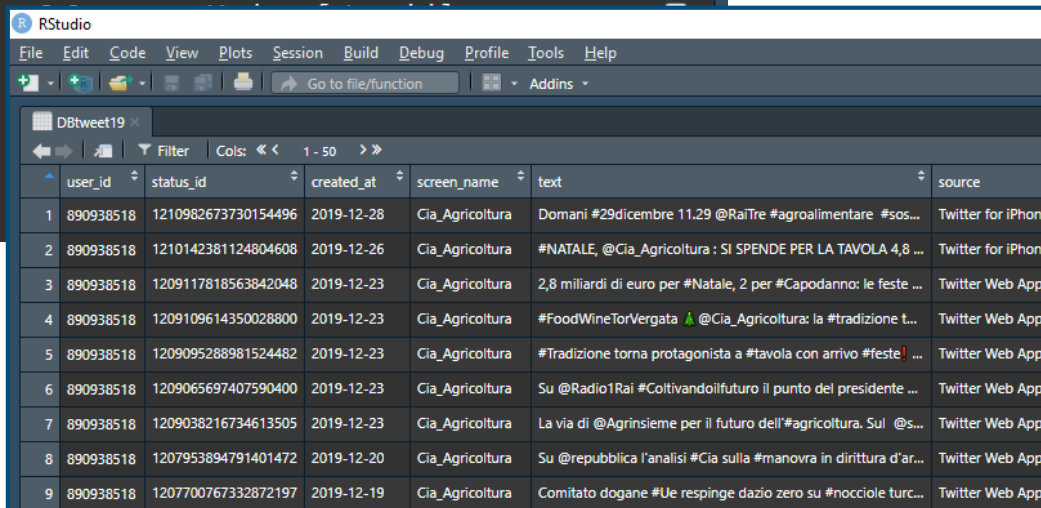
1) Oggetti creati (Vettori, dataframe, liste, funzioni, ...)

Environment, History



Name	Type	Length	Size	Value
F1_ALL.GEN.neg	data.frame	4	2.6 KB	11 obs. of 4 variables
F1_ALL.GEN.pos	data.frame	4	2.6 KB	11 obs. of 4 variables
F1_ALL.neg	data.frame	4	2.6 KB	11 obs. of 4 variables
F1_ALL.POL.neg	data.frame	4	2.6 KB	11 obs. of 4 variables
F1_ALL.POL.pos	data.frame	4	2.6 KB	11 obs. of 4 variables
F1_ALL.pos	data.frame	4	2.6 KB	11 obs. of 4 variables
F1_TEST.GEN.neg	data.frame	4	2.6 KB	11 obs. of 4 variables
F1_TEST.GEN.pos	data.frame	4	2.6 KB	11 obs. of 4 variables
F1_TEST.neg	data.frame	4	2.6 KB	11 obs. of 4 variables
F1_TEST.POL.neg	data.frame	4	2.6 KB	11 obs. of 4 variables
F1_TEST.POL.pos	data.frame	4	2.6 KB	11 obs. of 4 variables
F1_TEST.pos	data.frame	4	2.6 KB	11 obs. of 4 variables

Editor



	user_id	status_id	created_at	screen_name	text	source
1	890938518	1210982673730154496	2019-12-28	Cia_Agricoltura	Domani #29dicembre 11.29 @RaiTre #agroalimentare #sos...	Twitter for iPhon
2	890938518	1210142381124804608	2019-12-26	Cia_Agricoltura	#NATALE, @Cia_Agricoltura : SI SPENDE PER LA TAVOLA 4,8 ...	Twitter for iPhon
3	890938518	1209117818563842048	2019-12-23	Cia_Agricoltura	2,8 miliardi di euro per #Natale, 2 per #Capodanno: le feste ...	Twitter Web App
4	890938518	1209109614350028800	2019-12-23	Cia_Agricoltura	#FoodWineTorVergata 📍 @Cia_Agricoltura: la #tradizione t...	Twitter Web App
5	890938518	1209095288981524482	2019-12-23	Cia_Agricoltura	#Tradizione torna protagonista a #tavola con arrivo #feste ! ...	Twitter Web App
6	890938518	1209065697407590400	2019-12-23	Cia_Agricoltura	Su @Radio1Rai #Coltivandoilfuturo il punto del presidente ...	Twitter Web App
7	890938518	1209038216734613505	2019-12-23	Cia_Agricoltura	La via di @Agrinsieme per il futuro dell'agricoltura. Sul @s...	Twitter Web App
8	890938518	1207953894791401472	2019-12-20	Cia_Agricoltura	Su @repubblica l'analisi #Cia sulla #manovra in dirittura d'ar...	Twitter Web App
9	890938518	1207700767332872197	2019-12-19	Cia_Agricoltura	Comitato dogane #Ue respinge dazio zero su #nocciole turc...	Twitter Web App

2) Files, Grafici, packages, Help

Plots

Files

Help

Packages

Fitting Linear Models

Description

`lm` is used to fit linear models. It can be used to carry out regression, single stratum analysis of variance and analysis of (although `glm` may provide a more convenient interface for these).

Usage

```
lm(formula, data, subset, weights, na.action,
  method = "qr", model = TRUE, x = FALSE, y = FALSE, qr = TRUE,
  singular.ok = TRUE, contrasts = NULL, offset, ...)
```

Arguments

formula an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model. The details of model specification are given under 'Details'.

data an optional data frame, list or environment (or object coercible by `as.data.frame` to a data frame) containing the variables in the model. If not found in `data`, the variables are taken from `environment(formula)`, then from the environment from which `lm` is called.

subset an optional vector specifying a subset of observations to be used in the fitting process.

weights an optional vector of weights to be used in the fitting process. Should be `NULL` or a numeric vector. If `NULL`, weighted least squares is used with weights `weights` (that is, minimizing $\sum(w \cdot e^2)$); otherwise ordinary least squares is used. See also 'Details'.

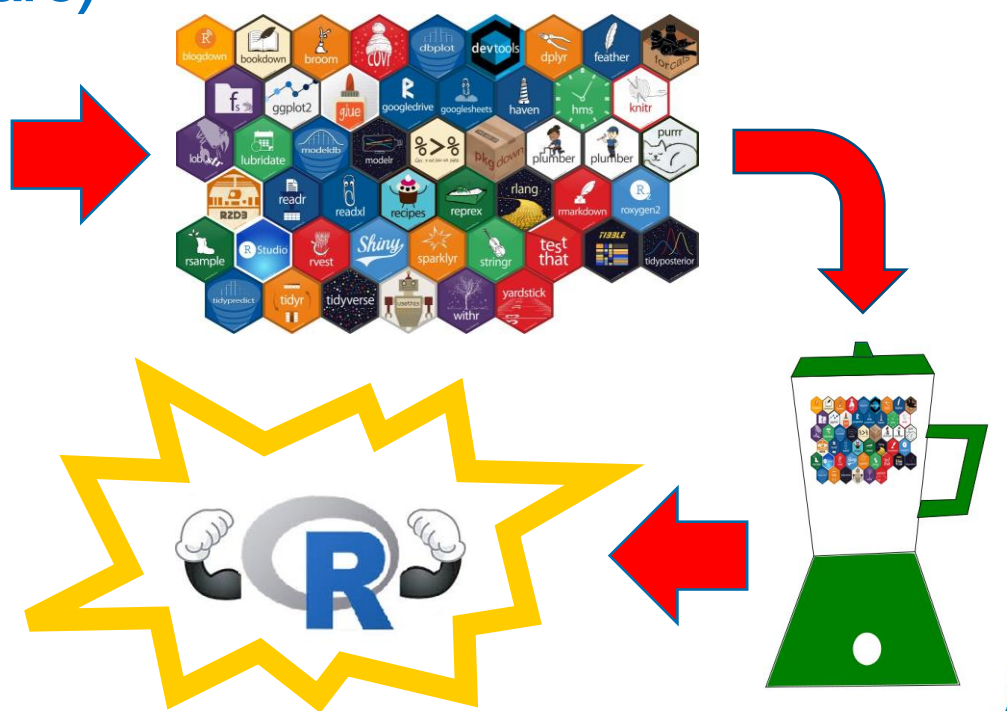
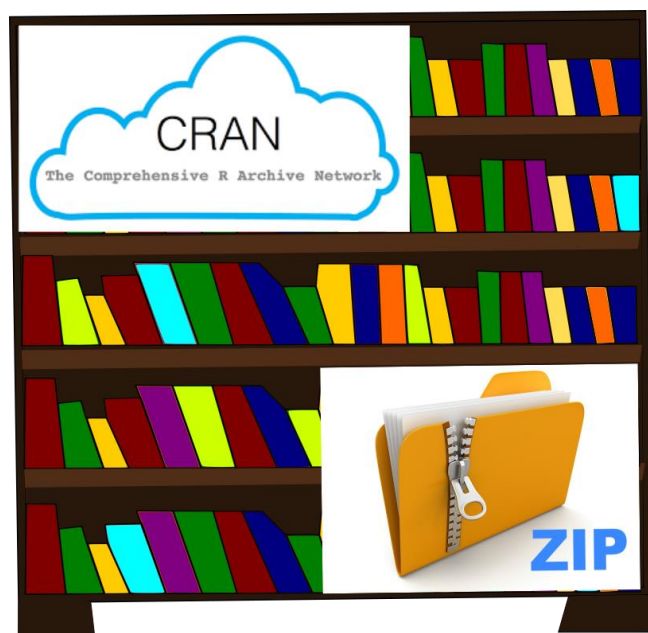
na.action a function which indicates what should happen when the data contain `NA`s. The default is set by the `na.action` of `options`, and is `na.fail` if that is unset. The 'factor-by-factor' default is `na.omit`. Another possible

Name	Description	Version
askpass	Safe Password Entry for R, Git, and SSH	1.1
assertthat	Easy Pre and Post Assertions	0.2.1
backports	Reimplementations of Functions Introduced Since R-3.0.0	1.1.5
BH	Boost C++ Header Files	1.69.0-1
bit	A Class for Vectors of 1-Bit Booleans	1.1-14
bit64	A S3 Class for Vectors of 64bit Integers	0.9-7
bitops	Bitwise Operations	1.0-6
brew	Templating Framework for Report Generation	1.0-6
callr	Call R from R	3.3.2
cellranger	Translate Spreadsheet Cell Ranges to Rows and Columns	1.1.0
cli	Helpers for Developing Command Line Interfaces	1.1.0
clipr	Read and Write from the System Clipboard	0.7.0
clisymbols	Unicode Symbols at the R Prompt	1.2.0
coda	Output Analysis and Diagnostics for MCMC	0.19-3
colorspace	A Toolbox for Manipulating and Assessing Colors and Palettes	1.4-1
commonmark	High Performance CommonMark and Github Markdown Rendering in R	1.7
corpus	Text Corpus Analysis	0.10.0
covr	Test Coverage for Packages	3.3.2
crayon	Colored Terminal Output	1.3.4
crosstalk	Inter-Widget Interactivity for HTML Widgets	1.0.0
curl	A Modern and Flexible Web Client for R	4.2
data.table	Extension of 'data.frame'	1.12.6
DBI	R Database Interface	1.0.0

Files, Plots, Packages, Help

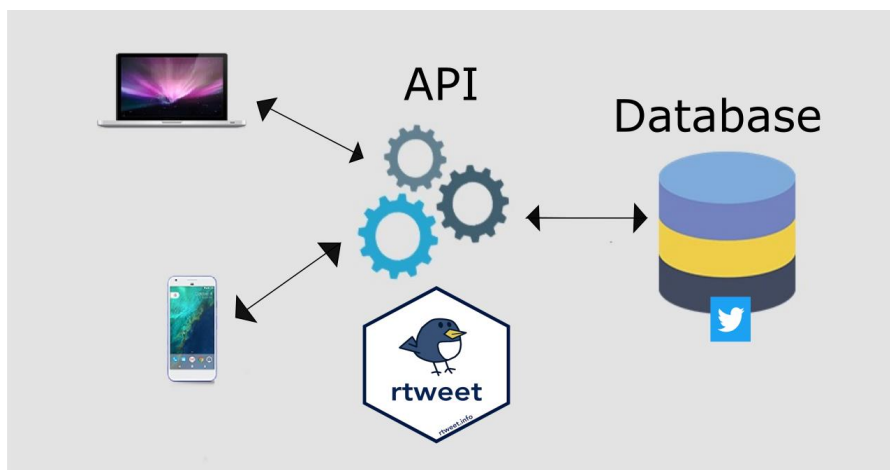
Possibilità di utilizzare pacchetti specifici, affidabili e scaricabili gratuitamente:

- ❑ Repository ufficiale di R, CRAN
- ❑ File zip locale (es. Github, piattaforma per lo sviluppo collaborativo di software)



☐ rtweet	→	Estrazione dati da Twitter (Limite di 45.000 tweets ogni 15 minuti): ☐ <i>get_timeline</i>: fino a 3.200 tweets pubblicati da un utente; ☐ <i>search_tweets</i>: fino a 18.000 tweets pubblicati negli ultimi 6-9 giorni.
☐ openxlsx - readr	→	Importazione/esportazione di file: ☐ <i>openxlsx</i>, file Excel (.xlsx), interfaccia di alto livello per scrittura, stile e modifica fogli di lavoro ☐ <i>readr</i>, file di testo (.txt, .csv, .tsv)
☐ tm	→	Applicazioni di text mining, di pulizia del testo (<i>tm_map</i>): ☐ <i>Funzioni standard</i> (punteggiatura, spazi bianchi, numeri, ...) ☐ <i>Funzioni definite dall'utente</i>
☐ stringr	→	Manipolazione stringhe di caratteri (<i>str_split</i> , <i>unlist</i> , ...)
☐ wordcloud	→	Generare un wordcloud
☐ corpus	→	Applicazioni su corpus (<i>text_tokens</i> , individuazione dei lemmi composti presenti nel formario affettivo)

- Twitter offre un accesso ai suoi dati attraverso le API (Application Programming Interfaces), ovvero interfacce di programmazione (regole) con cui diverse applicazioni, o software, comunicano tra loro



- Per utilizzare le API di Twitter, occorre ottenere le chiavi di accesso da Twitter attraverso la creazione di un'applicazione twitter (Twitter App)

Tutorial

<https://cran.r-project.org/web/packages/rtweet/vignettes/auth.html>

apps.twitter.com

Creare Nuova App

Informazioni necessarie:

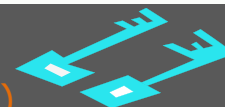
- Nome dell'app
- Descrizione dell'app
- Website URL (link pagina Twitter personale)
Callback URL (<http://127.0.0.1:1410>)
- Utilizzo dell'app

Cosa serve?

- Profilo twitter attivo

Si ottengono:

- Consumer key (API Key)
- Consumer secret (API Secret key)



```
appname <- «Tweet_Agricoltura»  
key      <- "kVuKkhl8FVNHTz68FTCYF"  
secret   <- "pjFkS2wEYGHhg7ZBGa1FxMM5XY3HRFz"  
  
twitter_token <- create_token(app = appname,  
                              consumer_key = key,  
                              consumer_secret = secret)
```



library (rtweet)



1) Estrazione Tweet

"Oggi pomeriggio con @Cia_Agricoltura abbiamo parlato della #manovra finanziaria grazie alla quale l'agricoltura ha ritrovato un ruolo importante in Italia [↓] <https://t.co/M5co0ltQNm>"

2) Pulizia del testo

oggi pomeriggio abbiamo parlato finanziaria grazie agricoltura
ritrovato ruolo importante italia

3) Lista dei tweet

```
Tweets <- {(oggi, pomeriggio, abbiamo, parlato, finanziaria, grazie,  
agricoltura, ritrovato, ruolo, importante, italia), .....}
```

4) Lessico affettivo MAL (148.867 termini)

```
MAL <- {(a_basso_livello, -0.75), (a_basso_reddito, -0.625), (abate,  
0.47), (abati, 0.47), (abbacchi, -0.125), .....}
```

oggi pomeriggio abbiamo
parlato finanziaria grazie
agricoltura ritrovato ruolo
importante italia

123 Retweet 578 Mi piace



Id	Token
.....	
2	agricoltura
.....	
4	grazie
5	importante
6	italia
.....	
9	ritrovato
.....	

Token	Score
.....	
agricoltura	0,1145
.....	
grazie	0,5427
.....	
.....	
importante	1,1335
.....	
parlato	-0,3529
.....	
ritrovato	0,0744
.....	

Morphologically-inflected
Affective
Lexicon



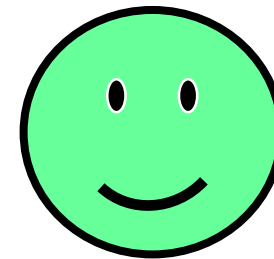
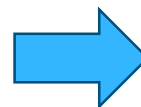
5) Calcolo del punteggio

Id	Token	Score	Id	Token	Score
1	abbiamo	0,1287	7	parlato	-0,3529
2	agricoltura	0,1145	8	pomeriggio	NA
3	finanziaria	0,25	9	ritrovato	0,0744
4	grazie	0,5427	10	ruolo	NA
5	importante	1,1335	11	oggi	0,1988
6	italia	NA			

Token	#	Ratio
Numero	11	72,7%
Con punteggio	8	

Punteggio totale

$$\sum p_i = 2,0897$$



Giudizio "Positivo"

Twitter

- ❑ **Twitter:** <https://twitter.com/login?lang=it>
- ❑ **Twitter Apps - Twitter Developer:** apps.twitter.com

Software

- ❑ **The Comprehensive R Archive Network:** <https://cran.r-project.org>
- ❑ **Rstudio:** <https://rstudio.com>

Package

- ❑ **"rtweet":** <https://cran.r-project.org/web/packages/rtweet/rtweet.pdf>
- ❑ **"readr":** <https://cran.r-project.org/web/packages/readr/readr.pdf>
- ❑ **"openxlsx":** <https://cran.r-project.org/web/packages/openxlsx/openxlsx.pdf>
- ❑ **"tm":** <https://cran.r-project.org/web/packages/tm/tm.pdf>
- ❑ **"stringr":** <https://cran.r-project.org/web/packages/stringr/stringr.pdf>
- ❑ **"wordcloud":** <https://cran.r-project.org/web/packages/corpus/corpus.pdf>
- ❑ **"corpus":** <https://cran.r-project.org/web/packages/corpus/corpus.pdf>

Grazie!

giuliano.gabrieli@crea.gov.it