A Micro Project report on

Google Data Analysis Using R

Submitted to the CMR Institute of Technology in partial fulfilment of the requirement for the award of the Laboratory of

DATA MINING AND DATA ANALYTICS

of

III-B.Tech. I-Semester

in

Computer Science and Engineering Department

Submitted by

Arun Kumar Gupta	**(20R01A0505)**
G Deepak	**(20R01A0520)**
Joyjit Paul	**(20R01A0528)**
M Sujith Reddy	**(20R01A0538)**
P Bharath	**(20R01A0544)**
T Varun Kumar	**(20R01A0555)**

Under the Guidance of

DR.Y. SUCHARITHA

Associate professor, Department of Computer Science and Engineering

CMR INSTITUTE OF TECHNOLOGY
(UGC AUTONOMOUS)
(Approved by AICTE, Affiliated to JNTU, Kukatpally, Hyderabad)
Kandlakoya, Medchal Road, Hyderabad

2022-2023

CMR INSTITUTE OF TECHNOLOGY

(UGC AUTONOMOUS)

(Approved by AICTE, Affiliated to JNTU, Kukatpally, Hyderabad)

Kandlakoya, Medchal Road, Hyderabad.

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that a Micro Project entitled with: "**Google Data Analysis Using R**" is being

Submitted By

Arun Kumar Gupta	**(20R01A0505)**
G Deepak	**(20R01A0520)**
Joyjit Paul	**(20R01A0528)**
M Sujith Reddy	**(20R01A0538)**
P Bharath	**(20R01A0544)**
T Varun Kumar	**(20R01A0555)**

In partial fulfilment of the requirement for award of the Data Mining and Data Analytics of III-B. Tech I- Semester in CSE towards a record of a Bonafide work carried out under our guidance and supervision.

Signature of Faculty
(Dr. Y. Sucharitha)

Signature of HOD
(Mr. A. Prakash)

Course Coordinator

ACKNOWLEDGEMENT

We are extremely grateful to **Dr. M. Janga Reddy**, **Director**, **Dr. B. Satyanarayana**, **Principal** and **Mr. A. Prakash Head of Department**, Dept of Computer Science and Engineering, CMR Institute of Technology for their inspiration and valuable guidance during entire duration.

We are extremely thankful to our Data Mining and Data Analytics Lab faculty in-charge Dr. Y. Sucharitha Associate professor, Computer Science and Engineering department, CMR Institute of Technology for his constant guidance, encouragement and moral support throughout the project.

We express our thanks to all staff members and friends for all the help and coordination extended in bringing out this Project successfully in time.

Finally, we are very much thankful to our parents and relatives who guided directly or indirectly for successful completion of the project.

Arun Kumar Gupta	(20R01A0505)
G Deepak	(20R01A0520)
Joyjit Paul	(20R01A0528)
M Sujith Reddy	(20R01A0538)
P Bharath	(20R01A0544)
T Varun Kumar	(20R01A0555)

CONTENTS

INTRODUCTION

R Programming Language

R is an open-source programming language that is widely used as a statistical software and data analysis tool. R generally comes with the Command-line interface. R is available across widely used platforms like Windows, Linux, and macOS. Also, the R programming language is the latest cutting-edge tool.

Why Use R?

- It is a great resource for data analysis, data visualization, data science and machine learning
- It provides many statistical techniques (such as statistical tests, classification, clustering and data reduction)
- It is easy to draw graphs in R, like pie charts, histograms, box plot, scatter plot, etc++
- It works on different platforms (Windows, Mac, Linux)
- It is open-source and free
- It has many packages that can be used to solve different problems

What is Google Analytics?

Google Analytics is a web analytics service offered by Google that measures website traffic

and creates analysis reports. Google Analytics service comes under Google Marketing Platform brand. Google launched Google Analytics on November 14, 2005. Google Analytics is used to track website activity, such as the duration of each session, pages reached per session, the bounce rate of individuals using the site, and the source of the traffic.

Google Analytics is a primarily used free web analytics tool. It provides in-depth insight into your website and business's online performance. It can be integrated with Google Ads to launch online campaigns to promote and sell their products and increase traffic on your website. It offers a wealth of data that companies can use to evaluate their website performance. It helps them plan for an effective digital marketing strategy and change tactics to achieve the best results.

Google Analytics can be used for both websites and mobile apps. It analyzes website data and creates customized reports as per business needs.

GOOGLE TRENDS DATA ANALYSIS IN R

For most of us, Google Web Search and the other main Google products are our weapons of choice whenever we need to find anything on the internet or in the real world. Whether it is to get up-to-the-minute news of the Covid pandemic, the latest scores of your favorite sport, or to find out how that tasty dish is made, Google is our #1 source of information right now.

One way of using some of the information Google has on us in our favor is with Google Trends. It makes it easier to discover trends and analyze the behavior of our customers and users in general. **Google Trends** is one of the best tools for knowledge discovery and to show in real-time (or almost) how relevant a subject is — at least in terms of Web searches and public interest.

PROJECT PREREQUISITES

For this project I choose RStudio Desktop in order to prepare, process, clean, analyze and create the visualizations. The data set was too large to be processed in Ms Excel, google spreadsheets and RStudio Cloud.

SOURCE CODE

To find the trending tweets:

```
library(rtweet)

library(dplyr)

library(ggplot2)

library(twitteR)

library(tidyverse)

library(stringr)

library(tidytext)

consumer_key <-"HJFjGj4jePPbVlmHRwqUUA2hR"

consumer_secret <-"iMwL77EecWmwJcjZ2IJcACtbKeqrU330pNARSrJDTznLblfyA1"

access_token<-"1332340495473803265-3zFRSVUyjLTj6N2VMaIG3PNjA8bcbd"

access_secret <-"J80cCRB5jqQp0tdeOTzjNC0r3k9X9U9zcVVmllJT4fVhg"

twitter_token = rtweet::create_token(app ='DMDA',

                        consumer_key ="HJFjGj4jePPbVlmHRwqUUA2hR",

                        consumer_secret
="iMwL77EecWmwJcjZ2IJcACtbKeqrU330pNARSrJDTznLblfyA1",access_token="13323
40495473803265-3zFRSVUyjLTj6N2VMaIG3PNjA8bcbd",
```

```
access_secret="J80cCRB5jqQp0tdeOTzjNC0r3k9X9U9zcVVmllJT4fVhg")

# TESTING LOCATION FILTER

View(trends_available() %>% filter(countryCode=="IN"))

# TESTING WOEID (Where On Earth IDentifier)

trending_tweets<-get_trends(woeid =23424848)

View(trending_tweets)

top_tweet<-head(trending_tweets$trend,1)

View(top_tweet)

tweets<-search_tweets(top_tweet,n=100,include_rts = FALSE,`-filter` = "replies",

                      lang = "en")

View(tweets)

# Frequency of Tweets time series graph

ts_plot(tweets, "hours") + labs(x = NULL, y = NULL,

title = "Frequency of tweets with a #ClimateEmergency hashtag",

subtitle = paste0(format(min(tweets$created_at), "%d %B %Y"), " to ", format(max(tweets$created_at),"%d %B %Y")),

caption = "Data collected from Twitter's REST API via rtweet") +theme_minimal()


#To Find Most Frequent Words used in Tweets


words <- tweets %>%mutate(text = str_remove_all(text, “&amp;|&lt;|&gt;”),

text = str_remove_all(text, “\\s?(f|ht)(tp)(s?)(://)([^\\.]*)[\\.|/](\\S*)”),

text = str_remove_all(text, “[^\x01-\x7F]”)) %>%

unnest_tokens(word, text, token = “tweets”) %>%

filter(!word %in% stop_words$word,
```

```
!word %in% str_remove_all(stop_words$word, “”),

str_detect(word, “[a-z]”), !str_detect(word, “^#”), !str_detect(word, “@\\S+”)) %>%

count(word, sort = TRUE)

library(wordcloud)

words %>% with(wordcloud(word, n, random.order = FALSE, max.words = 100, colors = “#F29545”))

tweets %>% unnest_tokens(mentions, text, “tweets”, to_lower = FALSE) %>%

filter(str_detect(mentions, “^@”)) %>%  count(mentions, sort = TRUE) %>%top_n(10)

mytext<-sub(“.”, “”,top_tweet)

print(mytext)
```

Analyzing “IRAN” Keyword using Google Trends Data

```
library(gtrendsR)

library(dplyr)

library(ggplot2)

library(ggforce)

library(spData)

library(ggrepel)

library(tmap)

library(lubridate)


trends <- gtrends(keyword = c(“Iran”),time=”now 7-d”)


View(trends)
```

```
View(trends$interest_over_time)

View(trends$interest_by_country)

View(trends$interest_by_region)

View(trends$interest_by_city)

View(trends$interest_by_dma)

View(trends$related_topics)

View(trends$related_queries)
```

ANALYSIS PART

```
print(summary(trends$related_queries))

print(summary(trends))

print(summary(trends$related_topics))

print(summary(trends$interest_over_time))

print(summary(trends$interest_by_city))

print(summary(trends$interest_by_region))

print(summary(trends$interest_by_dma))

#TimeSeries Analysis

trends_timeseries<- ts(trends$interest_over_time,start = decimal_date(dmy("23-11-2022")),
                       frequency =0.5)

# Printing the timeseries data.

Print(trends_timeseries)

plot(trends_timeseries,xlab ="Weekly Data",
     ylab ="count",
     main ="Iran search trends",
```

```
    col.main =”darkgreen”)

dev.off()

iot<-trends$interest_over_time

iot %>% +top_n(5,hits) %>% + arrange(desc(hits))

iot%>% ggplot(aes(x=date,y = hits,group=keyword ,color = keyword))  +theme_bw()+

labs(title = “Google Web searches for ‘Iran’ in last 7 Days”,

caption = “Obs: 3/22 was the day with the most searches”,x= NULL, y = “Interest”)+

ggforce::facet_zoom(xlim = c(as.POSIXct(as.Date(“2022-11-25”)),as.POSIXct(as.Date(“2022-11-25”)))) +geom_smooth(span=0.1,se=FALSE) + geom_vline(xintercept = as.POSIXct(as.Date(“2022-11-25”)),color = “red”, lwd = 0.5,linetype=”dashed”)+theme(legend.position = “none”) +geom_point(color=”black”)+

geom_label_repel(data = subset(iot2020, hits == 100),aes(label = as.character(date)),

size = 5, box.padding = unit(0.35, “lines”), point.padding = unit(0.3, “lines”))

trends$related_queries %>%filter(related_queries==”top”) %>%

mutate(value=factor(value,levels=rev(as.character(value))),subject=as.numeric(subject)) %>%top_n(10,value) %>%

ggplot(aes(x=value,y=subject,fill=”red”)) +

geom_bar(stat=’identity’,show.legend = F) +

coord_flip() + labs(title=”Queries most related with ‘Iran’”)
```

countries <- spData::world %>%

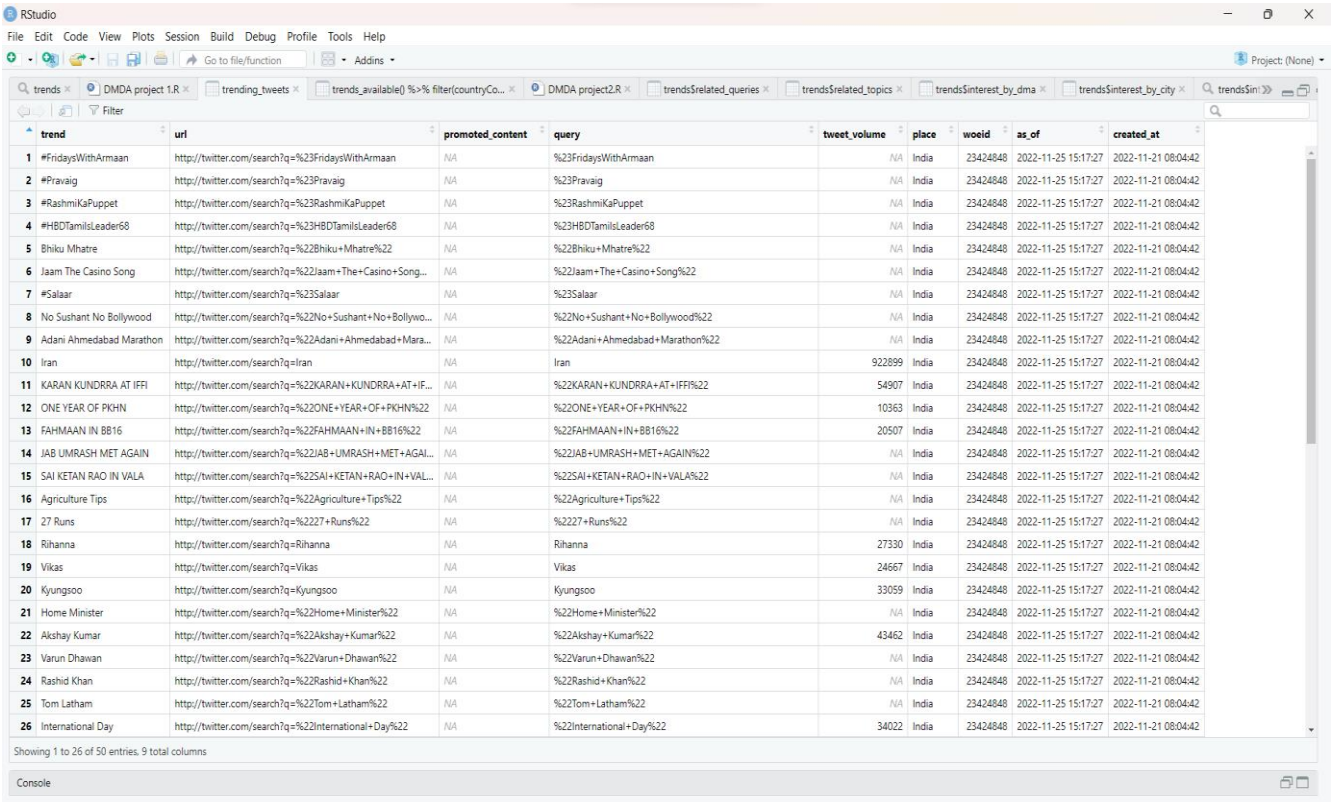
left_join(y=trends$interest_by_country,by = c(“name_long” = “location”),keep=T)

tm_shape(countries) +tm_fill(“hits”, title = “Iran Keyword Search Interest”,legend.reverse = T,id = “name_long”,

popup.vars=c(Name = “name_long”,Search.Interest = “hits”,Population = “pop”,Life.Expectancy = “lifeExp”, GDP.per.capita = “gdpPercap”))

RESULT:

1.Trending Tweets



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

trends | DMDA project 1.R | trending_tweets | trends_available() %>% filter(countryCo... | DMDA project2.R | trends$related_queries | trends$related_topics | trends$interest_by_dma | trends$interest_by_city | trends$int

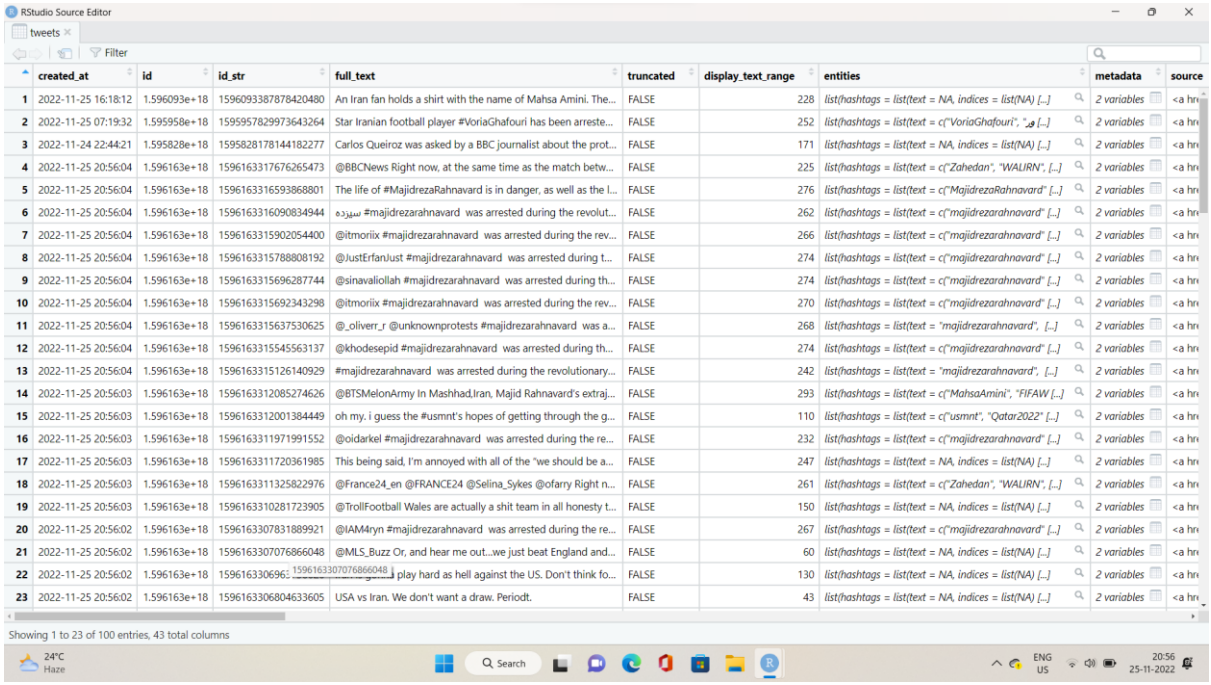
Filter

	trend	url	promoted_content	query	tweet_volume	place	woeid	as_of	created_at
1	#FridaysWithArmaan	http://twitter.com/search?q=%23FridaysWithArmaan	NA	%23FridaysWithArmaan	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
2	#Pravaig	http://twitter.com/search?q=%23Pravaig	NA	%23Pravaig	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
3	#RashmiKaPuppet	http://twitter.com/search?q=%23RashmiKaPuppet	NA	%23RashmiKaPuppet	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
4	#HBDTamilsLeader68	http://twitter.com/search?q=%23HBDTamilsLeader68	NA	%23HBDTamilsLeader68	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
5	Bhiku Mhatre	http://twitter.com/search?q=%22Bhiku+Mhatre%22	NA	%22Bhiku+Mhatre%22	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
6	Jaam The Casino Song	http://twitter.com/search?q=%22Jaam+The+Casino+Song...	NA	%22Jaam+The+Casino+Song%22	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
7	#Salaar	http://twitter.com/search?q=%23Salaar	NA	%23Salaar	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
8	No Sushant No Bollywood	http://twitter.com/search?q=%22No+Sushant+No+Bollywo...	NA	%22No+Sushant+No+Bollywood%22	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
9	Adani Ahmedabad Marathon	http://twitter.com/search?q=%22Adani+Ahmedabad+Mara...	NA	%22Adani+Ahmedabad+Marathon%22	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
10	Iran	http://twitter.com/search?q=Iran	NA	Iran	922899	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
11	KARAN KUNDRRA AT IFFI	http://twitter.com/search?q=%22KARAN+KUNDRRA+AT+IF...	NA	%22KARAN+KUNDRRA+AT+IFFI%22	54907	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
12	ONE YEAR OF PKHN	http://twitter.com/search?q=%22ONE+YEAR+OF+PKHN%22	NA	%22ONE+YEAR+OF+PKHN%22	10363	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
13	FAHMAAN IN BB16	http://twitter.com/search?q=%22FAHMAAN+IN+BB16%22	NA	%22FAHMAAN+IN+BB16%22	20507	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
14	JAB UMRASH MET AGAIN	http://twitter.com/search?q=%22JAB+UMRASH+MET+AGAI...	NA	%22JAB+UMRASH+MET+AGAIN%22	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
15	SAI KETAN RAO IN VALA	http://twitter.com/search?q=%22SAI+KETAN+RAO+IN+VAL...	NA	%22SAI+KETAN+RAO+IN+VALA%22	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
16	Agriculture Tips	http://twitter.com/search?q=%22Agriculture+Tips%22	NA	%22Agriculture+Tips%22	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
17	27 Runs	http://twitter.com/search?q=%2227+Runs%22	NA	%2227+Runs%22	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
18	Rihanna	http://twitter.com/search?q=Rihanna	NA	Rihanna	27330	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
19	Vikas	http://twitter.com/search?q=Vikas	NA	Vikas	24667	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
20	Kyungsoo	http://twitter.com/search?q=Kyungsoo	NA	Kyungsoo	33059	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
21	Home Minister	http://twitter.com/search?q=%22Home+Minister%22	NA	%22Home+Minister%22	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
22	Akshay Kumar	http://twitter.com/search?q=%22Akshay+Kumar%22	NA	%22Akshay+Kumar%22	43462	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
23	Varun Dhawan	http://twitter.com/search?q=%22Varun+Dhawan%22	NA	%22Varun+Dhawan%22	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
24	Rashid Khan	http://twitter.com/search?q=%22Rashid+Khan%22	NA	%22Rashid+Khan%22	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
25	Tom Latham	http://twitter.com/search?q=%22Tom+Latham%22	NA	%22Tom+Latham%22	NA	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42
26	International Day	http://twitter.com/search?q=%22International+Day%22	NA	%22International+Day%22	34022	India	23424848	2022-11-25 15:17:27	2022-11-21 08:04:42

Showing 1 to 26 of 50 entries, 9 total columns

Console

2. Full text of Tweets



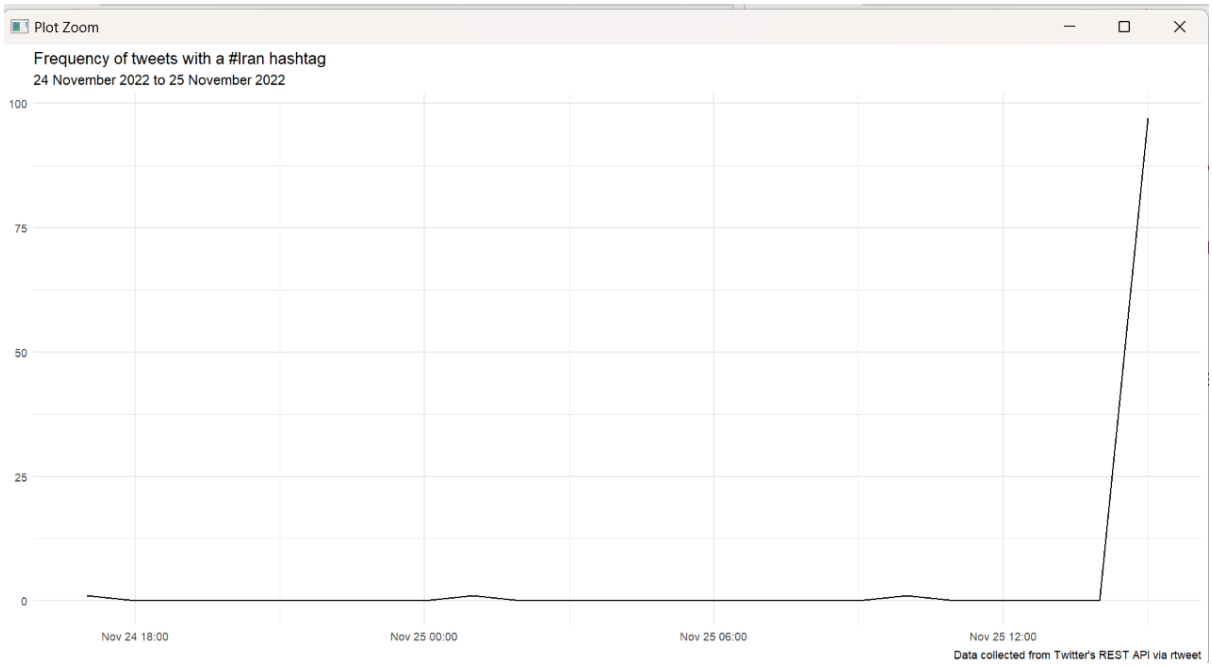
RStudio Source Editor — tweets

Filter

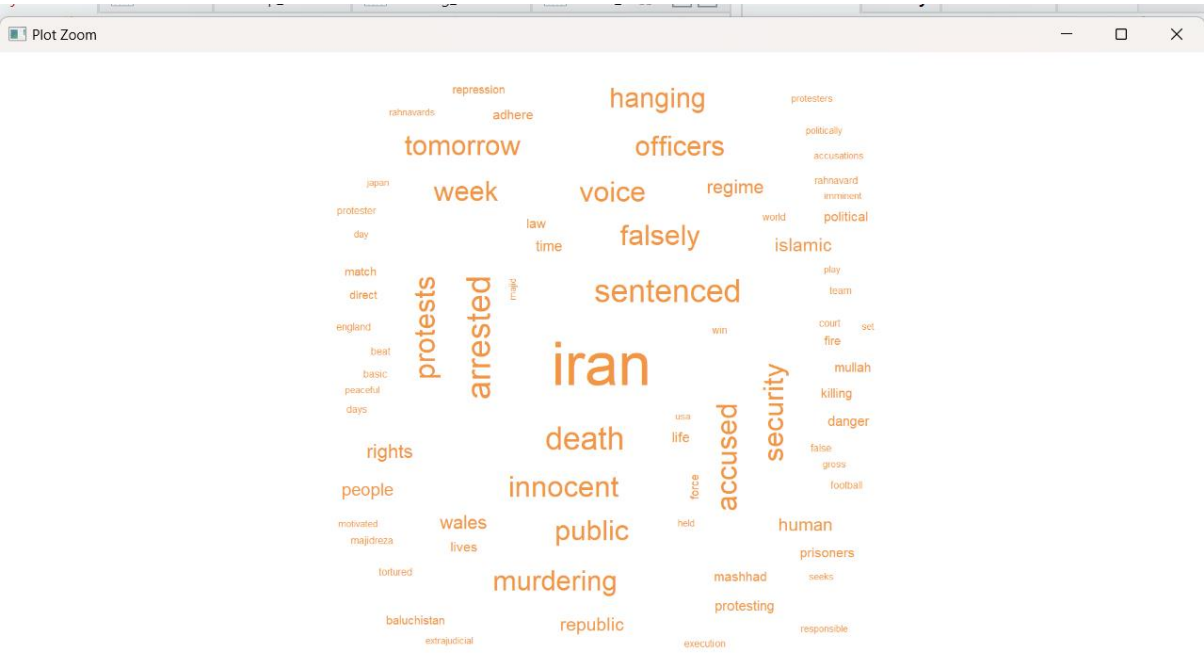
	created_at	id	id_str	full_text	truncated	display_text_range	entities	metadata	source
1	2022-11-25 16:18:12	1.596093e+18	1596093387878420480	An Iran fan holds a shirt with the name of Mahsa Amini. The...	FALSE	228	list(hashtags = list(text = NA, indices = list(NA) [...]	2 variables	<a hr
2	2022-11-25 07:19:32	1.595958e+18	1595957829973643264	Star Iranian football player #VoriaGhafouri has been arreste...	FALSE	252	list(hashtags = list(text = c("VoriaGhafouri", "و [...]	2 variables	<a hr
3	2022-11-24 22:44:21	1.595828e+18	1595828178144182277	Carlos Queiroz was asked by a BBC journalist about the prot...	FALSE	171	list(hashtags = list(text = NA, indices = list(NA) [...]	2 variables	<a hr
4	2022-11-25 20:56:04	1.596163e+18	1596163317676265473	@BBCNews Right now, at the same time as the match betw...	FALSE	225	list(hashtags = list(text = c("Zahedan", "WALIRN", [...]	2 variables	<a hr
5	2022-11-25 20:56:04	1.596163e+18	1596163316593868801	The life of #MajidrezaRahnavard is in danger, as well as the l...	FALSE	276	list(hashtags = list(text = c("MajidrezaRahnavard" [...]	2 variables	<a hr
6	2022-11-25 20:56:04	1.596163e+18	1596163316090834944	سیزده #majidrezarahnavard was arrested during the revolut...	FALSE	262	list(hashtags = list(text = c("majidrezarahnavard" [...]	2 variables	<a hr
7	2022-11-25 20:56:04	1.596163e+18	1596163315902054400	@itmoriix #majidrezarahnavard was arrested during the rev...	FALSE	266	list(hashtags = list(text = c("majidrezarahnavard" [...]	2 variables	<a hr
8	2022-11-25 20:56:04	1.596163e+18	1596163315788808192	@JustErfanJust #majidrezarahnavard was arrested during t...	FALSE	274	list(hashtags = list(text = c("majidrezarahnavard" [...]	2 variables	<a hr
9	2022-11-25 20:56:04	1.596163e+18	1596163315696287744	@sinavaliollah #majidrezarahnavard was arrested during th...	FALSE	274	list(hashtags = list(text = c("majidrezarahnavard" [...]	2 variables	<a hr
10	2022-11-25 20:56:04	1.596163e+18	1596163315692343298	@itmoriix #majidrezarahnavard was arrested during the rev...	FALSE	270	list(hashtags = list(text = c("majidrezarahnavard" [...]	2 variables	<a hr
11	2022-11-25 20:56:04	1.596163e+18	1596163315637530625	@_oliverr_r @unknownprotests #majidrezarahnavard was a...	FALSE	268	list(hashtags = list(text = "majidrezarahnavard", [...]	2 variables	<a hr
12	2022-11-25 20:56:04	1.596163e+18	1596163315545563137	@khodesepid #majidrezarahnavard was arrested during th...	FALSE	274	list(hashtags = list(text = c("majidrezarahnavard" [...]	2 variables	<a hr
13	2022-11-25 20:56:04	1.596163e+18	1596163315126140929	#majidrezarahnavard was arrested during the revolutionary...	FALSE	242	list(hashtags = list(text = "majidrezarahnavard", [...]	2 variables	<a hr
14	2022-11-25 20:56:03	1.596163e+18	1596163312085274626	@BTSMelonArmy In Mashhad,Iran, Majid Rahnavard's extraj...	FALSE	293	list(hashtags = list(text = c("MahsaAmini", "FIFAW [...]	2 variables	<a hr
15	2022-11-25 20:56:03	1.596163e+18	1596163312001384449	oh my. i guess the #usmnt's hopes of getting through the g...	FALSE	110	list(hashtags = list(text = c("usmnt", "Qatar2022" [...]	2 variables	<a hr
16	2022-11-25 20:56:03	1.596163e+18	1596163311971991552	@oidarkel #majidrezarahnavard was arrested during the re...	FALSE	232	list(hashtags = list(text = c("majidrezarahnavard" [...]	2 variables	<a hr
17	2022-11-25 20:56:03	1.596163e+18	1596163311720361985	This being said, I'm annoyed with all of the "we should be a...	FALSE	247	list(hashtags = list(text = NA, indices = list(NA) [...]	2 variables	<a hr
18	2022-11-25 20:56:03	1.596163e+18	1596163311325822976	@France24_en @FRANCE24 @Selina_Sykes @ofarry Right n...	FALSE	261	list(hashtags = list(text = c("Zahedan", "WALIRN", [...]	2 variables	<a hr
19	2022-11-25 20:56:03	1.596163e+18	1596163310281723905	@TrollFootball Wales are actually a shit team in all honesty t...	FALSE	150	list(hashtags = list(text = NA, indices = list(NA) [...]	2 variables	<a hr
20	2022-11-25 20:56:02	1.596163e+18	1596163307831889921	@IAM4ryn #majidrezarahnavard was arrested during the re...	FALSE	267	list(hashtags = list(text = c("majidrezarahnavard" [...]	2 variables	<a hr
21	2022-11-25 20:56:02	1.596163e+18	1596163307076866048	@MLS_Buzz Or, and hear me out...we just beat England and...	FALSE	60	list(hashtags = list(text = NA, indices = list(NA) [...]	2 variables	<a hr
22	2022-11-25 20:56:02	1.596163e+18	1596163306963307076866048	play hard as hell against the US. Don't think fo...	FALSE	130	list(hashtags = list(text = NA, indices = list(NA) [...]	2 variables	<a hr
23	2022-11-25 20:56:02	1.596163e+18	1596163306804633605	USA vs Iran. We don't want a draw. Periodt.	FALSE	43	list(hashtags = list(text = NA, indices = list(NA) [...]	2 variables	<a hr

Showing 1 to 23 of 100 entries, 43 total columns

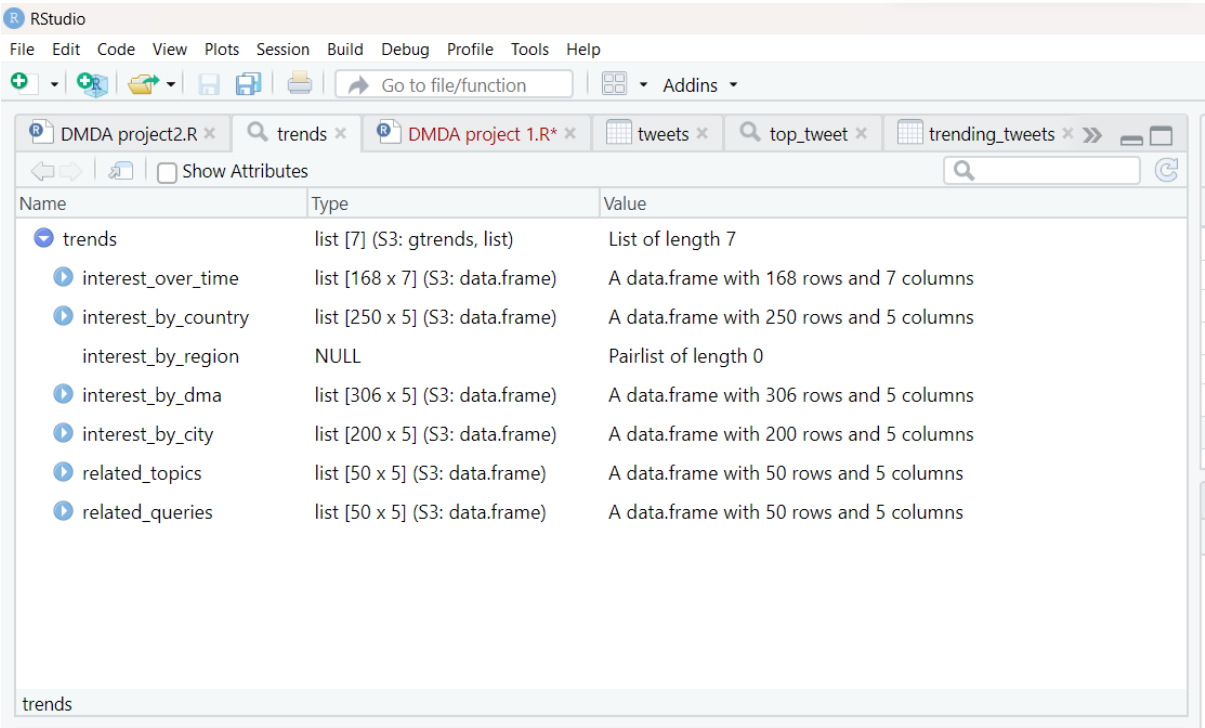
3. Frequency of Tweets with #Iran hashtag



4. Most Frequent Words used with #Iran



5. Google Iran Trends Data set



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

DMDA project2.R | trends$interest_over_time | trends | DMDA project 1.R* | tweets | top

Filter

	date	hits	keyword	geo	time	gprop	category
1	2022-11-18 16:00:00	1	Iran	world	now 7-d	web	0
2	2022-11-18 17:00:00	1	Iran	world	now 7-d	web	0
3	2022-11-18 18:00:00	1	Iran	world	now 7-d	web	0
4	2022-11-18 19:00:00	1	Iran	world	now 7-d	web	0
5	2022-11-18 20:00:00	1	Iran	world	now 7-d	web	0
6	2022-11-18 21:00:00	1	Iran	world	now 7-d	web	0
7	2022-11-18 22:00:00	1	Iran	world	now 7-d	web	0
8	2022-11-18 23:00:00	1	Iran	world	now 7-d	web	0
9	2022-11-19 00:00:00	1	Iran	world	now 7-d	web	0
10	2022-11-19 01:00:00	1	Iran	world	now 7-d	web	0
11	2022-11-19 02:00:00	1	Iran	world	now 7-d	web	0
12	2022-11-19 03:00:00	1	Iran	world	now 7-d	web	0

Showing 1 to 12 of 168 entries, 7 total columns

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

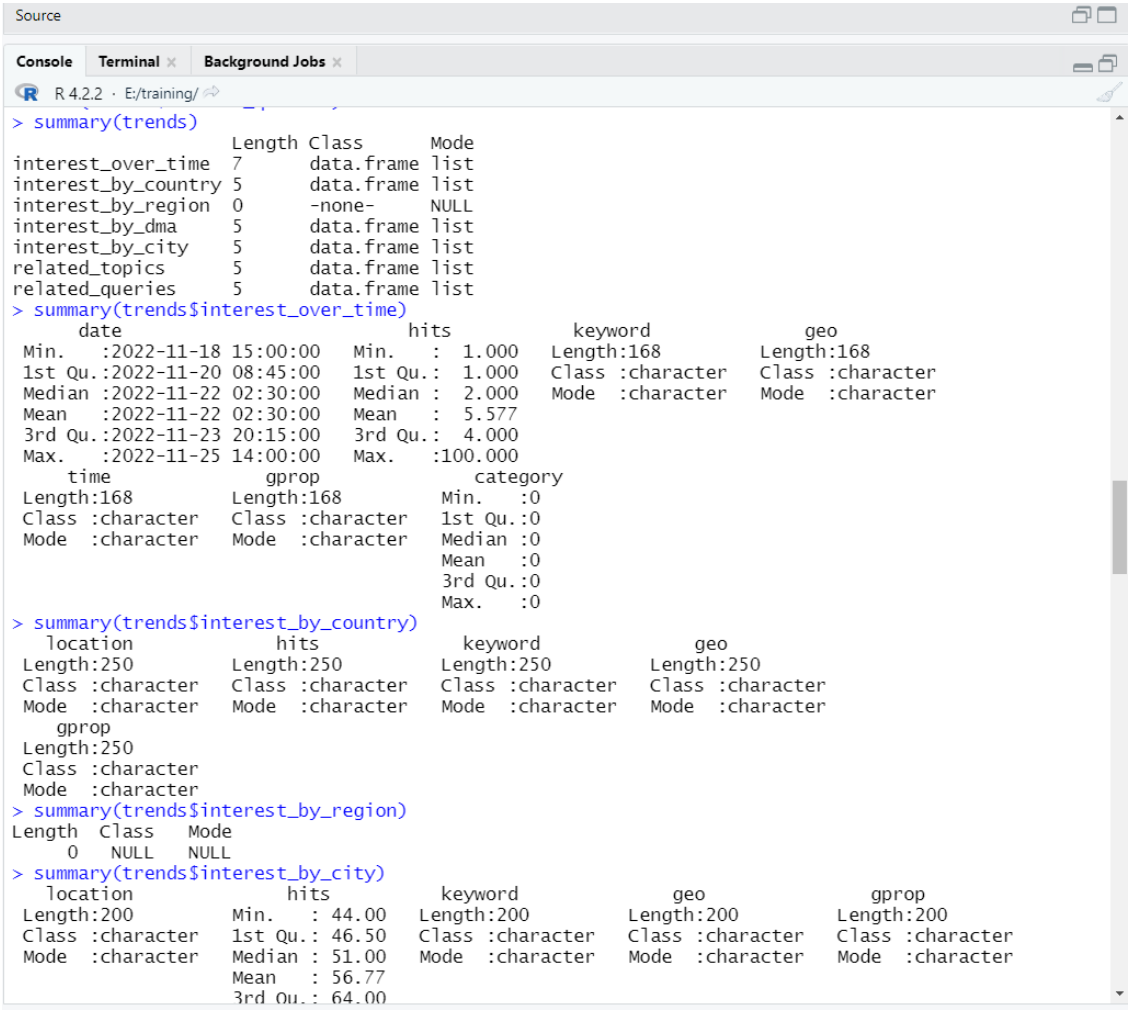
DMDA project2.R | trends$interest_by_country | trends$interest_over_

Filter

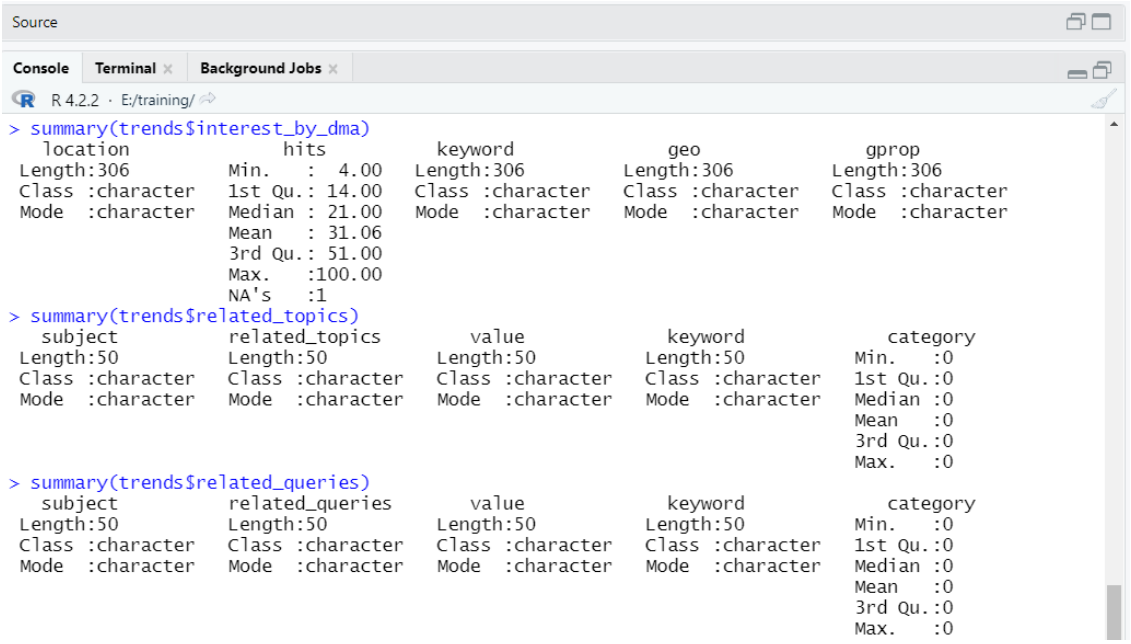
	location	hits	keyword	geo	gprop
1	Timor-Leste		Iran	world	web
2	Solomon Islands		Iran	world	web
3	Qatar	100	Iran	world	web
4	Uganda	96	Iran	world	web
5	Vanuatu		Iran	world	web
6	Haiti	83	Iran	world	web
7	Zimbabwe	80	Iran	world	web
8	Gibraltar		Iran	world	web
9	Central African Republic		Iran	world	web
10	Somalia		Iran	world	web
11	United Kingdom	71	Iran	world	web
12	Maldives		Iran	world	web

Showing 1 to 12 of 250 entries, 5 total columns

6. Summary of Trends



```
Source
Console  Terminal  Background Jobs
R 4.2.2 · E:/training/
> summary(trends)
                    Length Class      Mode
interest_over_time  7      data.frame list
interest_by_country 5      data.frame list
interest_by_region  0      -none-     NULL
interest_by_dma     5      data.frame list
interest_by_city    5      data.frame list
related_topics      5      data.frame list
related_queries     5      data.frame list
> summary(trends$interest_over_time)
      date                       hits           keyword             geo
 Min.   :2022-11-18 15:00:00   Min.   :  1.000   Length:168         Length:168
 1st Qu.:2022-11-20 08:45:00   1st Qu.:  1.000   Class :character   Class :character
 Median :2022-11-22 02:30:00   Median :  2.000   Mode  :character   Mode  :character
 Mean   :2022-11-22 02:30:00   Mean   :  5.577
 3rd Qu.:2022-11-23 20:15:00   3rd Qu.:  4.000
 Max.   :2022-11-25 14:00:00   Max.   :100.000
     time              gprop              category
 Length:168         Length:168         Min.   :0
 Class :character   Class :character   1st Qu.:0
 Mode  :character   Mode  :character   Median :0
                                       Mean   :0
                                       3rd Qu.:0
                                       Max.   :0
> summary(trends$interest_by_country)
   location            hits            keyword             geo
 Length:250         Length:250         Length:250         Length:250
 Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character
    gprop
 Length:250
 Class :character
 Mode  :character
> summary(trends$interest_by_region)
Length  Class   Mode
     0   NULL   NULL
> summary(trends$interest_by_city)
   location            hits            keyword             geo                gprop
 Length:200         Min.   : 44.00   Length:200         Length:200         Length:200
 Class :character   1st Qu.: 46.50   Class :character   Class :character   Class :character
 Mode  :character   Median : 51.00   Mode  :character   Mode  :character   Mode  :character
                    Mean   : 56.77
                    3rd Qu.: 64.00
```



```
Source
Console  Terminal  Background Jobs
R 4.2.2 · E:/training/
> summary(trends$interest_by_dma)
   location            hits            keyword             geo                gprop
 Length:306         Min.   :  4.00   Length:306         Length:306         Length:306
 Class :character   1st Qu.: 14.00   Class :character   Class :character   Class :character
 Mode  :character   Median : 21.00   Mode  :character   Mode  :character   Mode  :character
                    Mean   : 31.06
                    3rd Qu.: 51.00
                    Max.   :100.00
                    NA's   :1
> summary(trends$related_topics)
   subject          related_topics        value             keyword            category
 Length:50          Length:50          Length:50          Length:50          Min.   :0
 Class :character   Class :character   Class :character   Class :character   1st Qu.:0
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :0
                                                                            Mean   :0
                                                                            3rd Qu.:0
                                                                            Max.   :0
> summary(trends$related_queries)
   subject          related_queries       value             keyword            category
 Length:50          Length:50          Length:50          Length:50          Min.   :0
 Class :character   Class :character   Class :character   Class :character   1st Qu.:0
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :0
                                                                            Mean   :0
                                                                            3rd Qu.:0
                                                                            Max.   :0
```

7. TimeSeries Data

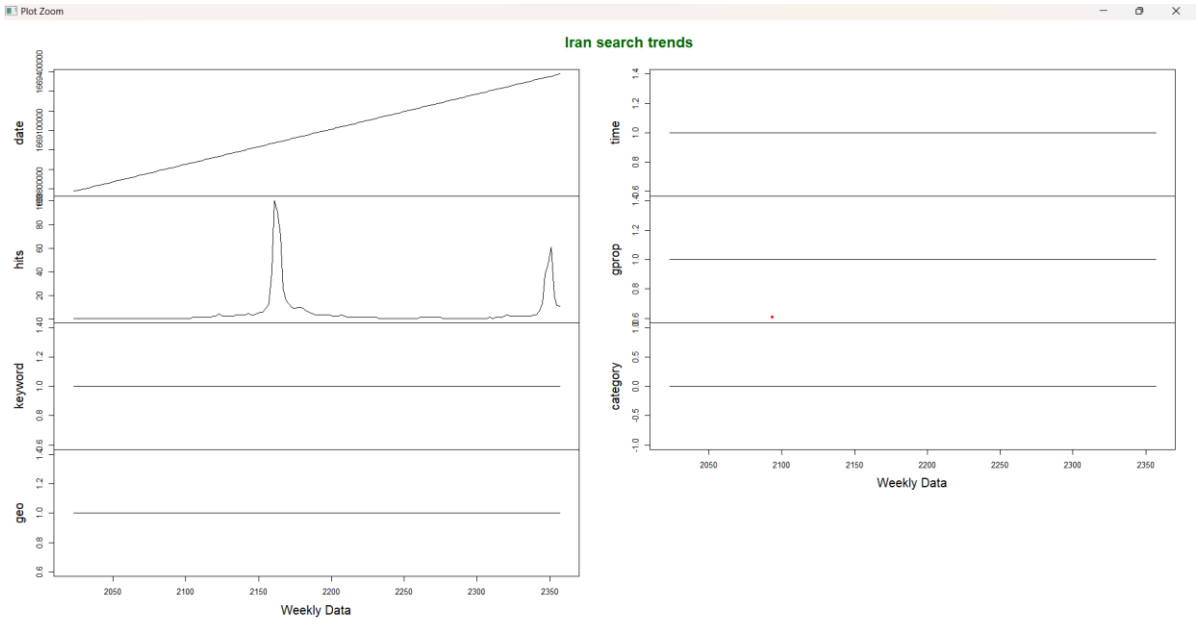
Source

Console | Terminal | Background Jobs

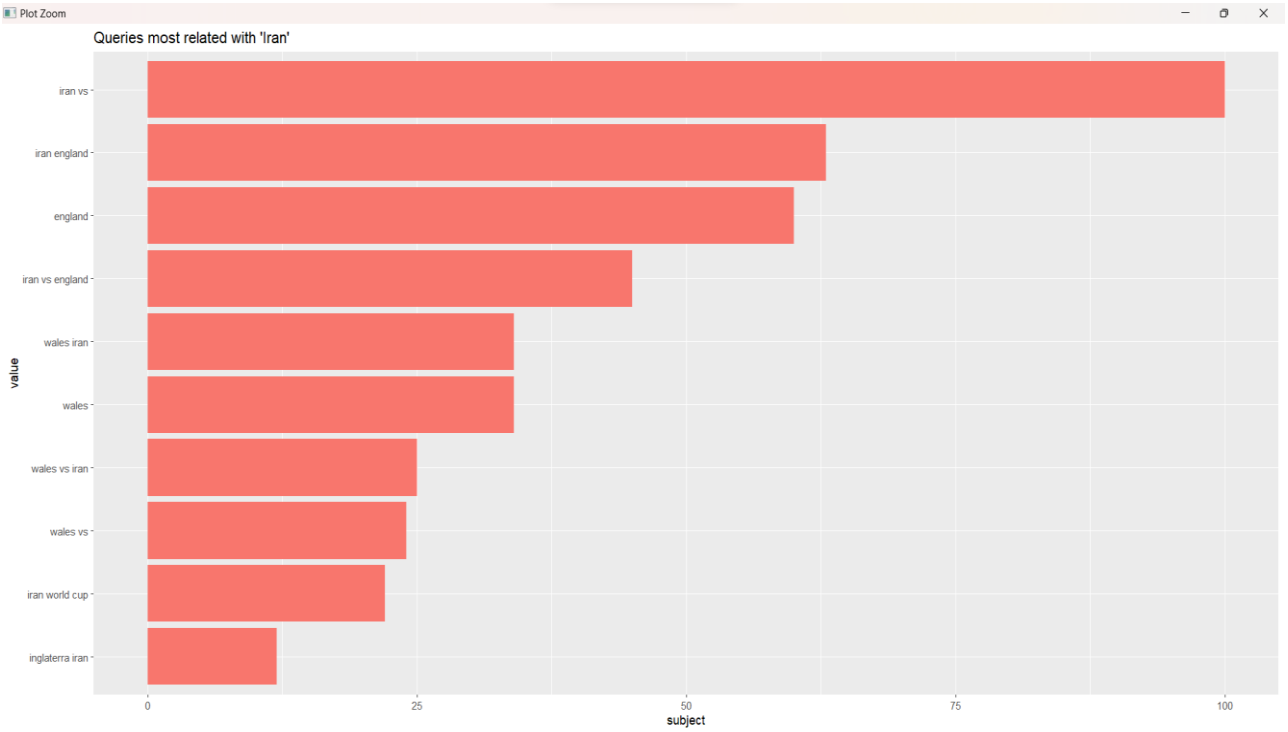
R 4.2.2 · E:/training/

```
> trends_timeseries<- ts(trends$interest_over_time,start = decimal_date(dmy("23-11-2022")),
+                        frequency =0.5)
> # Printing the timeseries data.
> print(trends_timeseries)
Time Series:
Start = 2022.89315068493
End = 2356.89315068493
Frequency = 0.5
             date hits keyword geo time gprop category
2022.893 1668783600    1       1   1    1     1        0
2024.893 1668787200    1       1   1    1     1        0
2026.893 1668790800    1       1   1    1     1        0
2028.893 1668794400    1       1   1    1     1        0
2030.893 1668798000    1       1   1    1     1        0
2032.893 1668801600    1       1   1    1     1        0
2034.893 1668805200    1       1   1    1     1        0
2036.893 1668808800    1       1   1    1     1        0
2038.893 1668812400    1       1   1    1     1        0
2040.893 1668816000    1       1   1    1     1        0
2042.893 1668819600    1       1   1    1     1        0
2044.893 1668823200    1       1   1    1     1        0
2046.893 1668826800    1       1   1    1     1        0
2048.893 1668830400    1       1   1    1     1        0
2050.893 1668834000    1       1   1    1     1        0
2052.893 1668837600    1       1   1    1     1        0
2054.893 1668841200    1       1   1    1     1        0
2056.893 1668844800    1       1   1    1     1        0
2058.893 1668848400    1       1   1    1     1        0
2060.893 1668852000    1       1   1    1     1        0
2062.893 1668855600    1       1   1    1     1        0
2064.893 1668859200    1       1   1    1     1        0
2066.893 1668862800    1       1   1    1     1        0
2068.893 1668866400    1       1   1    1     1        0
2070.893 1668870000    1       1   1    1     1        0
2072.893 1668873600    1       1   1    1     1        0
2074.893 1668877200    1       1   1    1     1        0
2076.893 1668880800    1       1   1    1     1        0
2078.893 1668884400    1       1   1    1     1        0
2080.893 1668888000    1       1   1    1     1        0
2082.893 1668891600    1       1   1    1     1        0
2084.893 1668895200    2       1   1    1     1        0
2086.893 1668898800    1       1   1    1     1        0
2088.893 1668902400    1       1   1    1     1        0
```

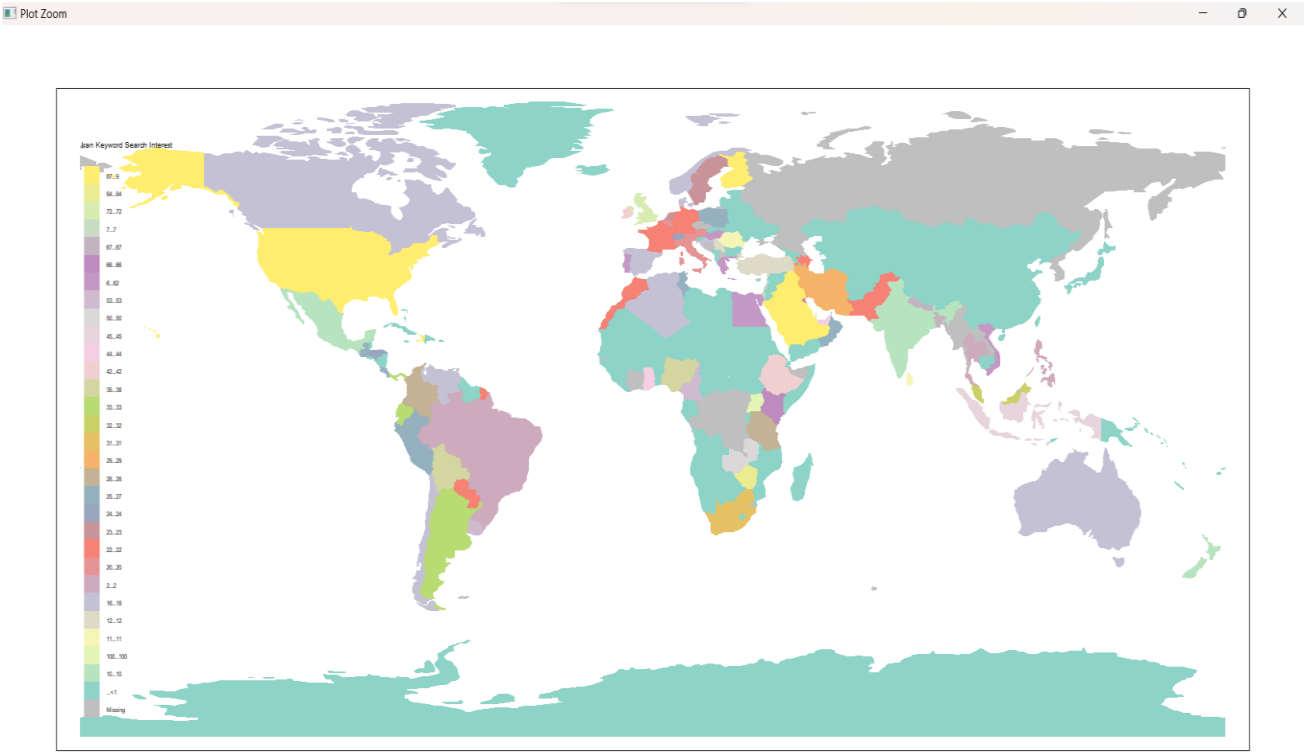
8. Plotting Time series of Iran Search Trends



8. Queries Most related with Iran



9. Iran Keyword Search Interest



CONCLUSION:

The R programming language was designed to work with data at all stages of the data analysis process. Here, in this project we have tried to analyze data regarding trending topics in the twitter and found a trending topic as Iran. Using google trends we have gathered the data regarding Iran and its various trends according to region, date, and its number of hits per day in various regions of the world.

Finally, we have analyzed the data using Timeseries concept and have visualized the number of hits, queries related to Iran and its search interes.

REFERENCE:

- **https://rafabelokurows.medium.com/google-trends-data-analysis-in-r-5b353022802c**
- **https://medium.com/@traffordDataLab/exploring-tweets-in-r-54f6011a193d**