

```
library(dplyr)
```

```
rladies_global %>%  
  filter(city == 'Bari')
```



DATA SCIENCE:

Cambia la tua vita con i dati

8 Marzo 2019

R-Ladies Bari

Chi sono le RLadies?



Community fondata dalla statista [Gabriela de Queiroz](#), San Francisco, 2012



OUR MISSION

"Promotes gender diversity
in the R community via
meetups and mentorship in a
friendly and safe
environment"

<https://rladies.org/>

Dove siamo oggi?



141

R-Ladies Cities



45

R-Ladies Countries



141

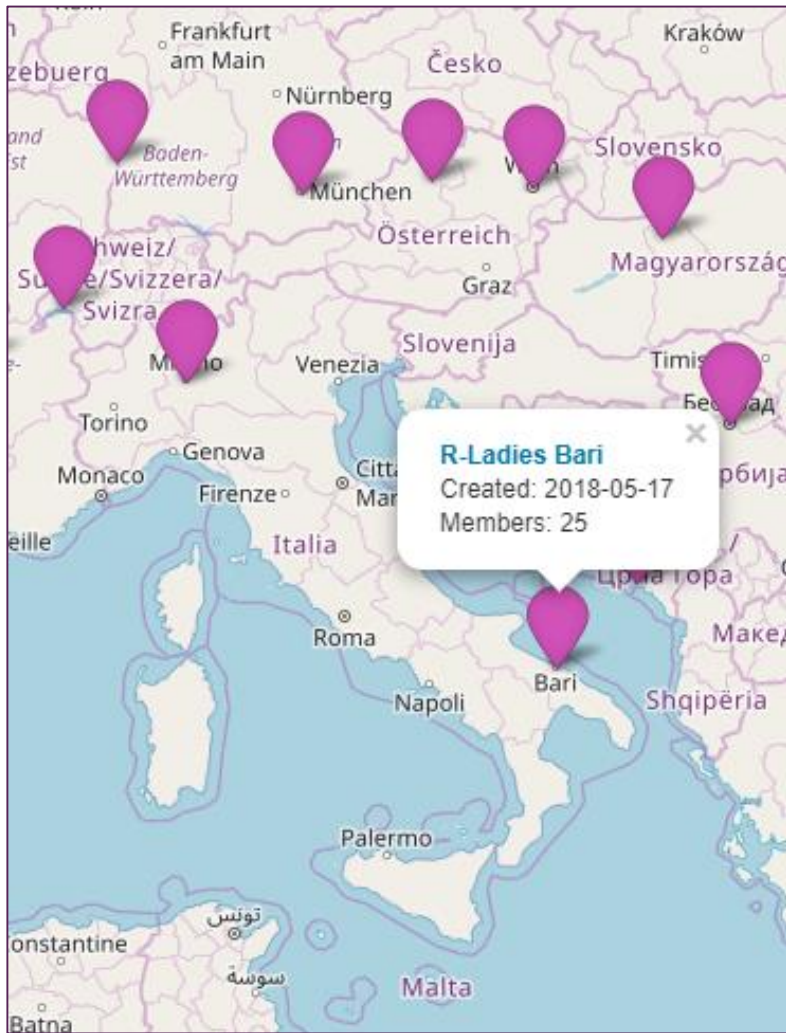
R-Ladies groups on meetup.com



41286

R-Ladies members on meetup.com





[ShinyApp](https://shinyapp.io/)

<https://gqueiroz.shinyapps.io/rshinylady/>



R-LADIES BARI

Founders of R-Ladies Bari

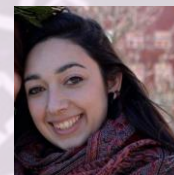


Giovanna de Vincenzo



- Data scientist @Infocamere
- Laurea in **Economia** @UniBA
- Master **Data Science** @TorVergata
- Interessi: **machine learning, analytics, data viz e innovazione digitale**

Francesca Lamontanara



- Data scientist @Infocamere
- Laurea in **Statistica** @UniBA
- Master **Data Science** @PoliBA & UniBA
- Interessi: **text mining & NLP, time series analytics e stochastic models**

La nostra mission è semplice



Perché siamo qui

- Women leadership
- Community globale
- Passione
- Punto di riferimento
- Nuove idee

I nostri obiettivi

- Meetup mensili
- Facilitare l'apprendimento del linguaggio R
- Incoraggiare le donne a programmare e al mondo scientifico

**More
Women
leader**

**More
Women
developer**

**More
Women
speaker**

| Eventi passati



Follow us



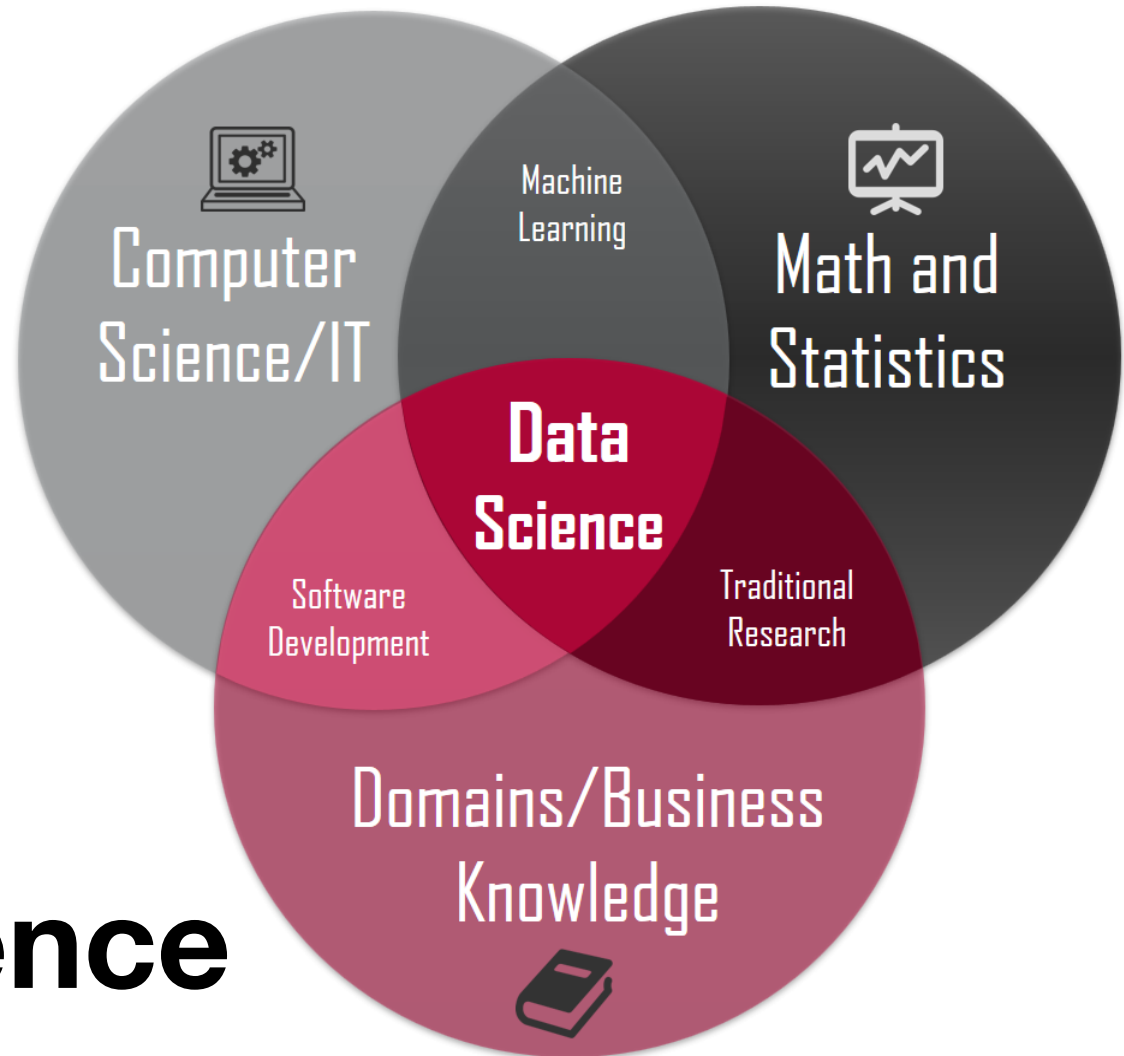
3 dicembre '18
Kickoff



29 Gennaio '19
**'Primi passi in
R Studio'**



5 Marzo '19
**'Manipolare i dati
con Tidyverse'**



Data Science

Millions of data



Donald J. Trump
@realDonaldTrump

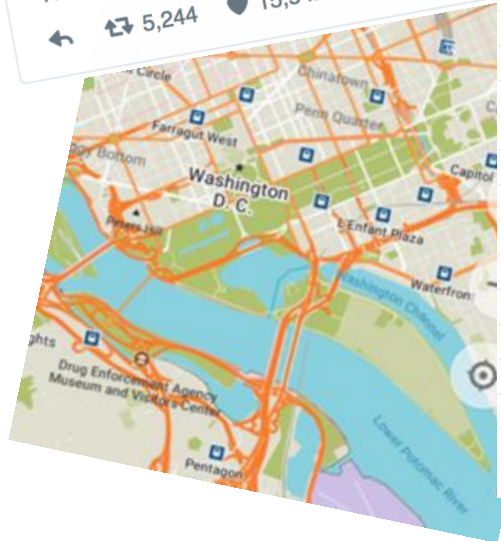
The United States must greatly strengthen and
nuclear capability until such time as the world c
senses regarding nukes

11:50 AM - 22 Dec 2016

5,244 15,542



	A	B	C	D	E
1	Salesperson	Region	Account	Order Amount	Month
2	Albertson, Kathy	East	29386	\$925.00	January
3	Albertson, Kathy	East	74830	\$875.00	February
4	Albertson, Kathy	East	90099	\$500.00	February
5	Albertson, Kathy	East	74830	\$350.00	March
6	Brennan, Michael	West	82853	\$400.00	January
7	Brennan, Michael	West	72949	\$850.00	January
8	Brennan, Michael	West	90044	\$1,500.00	January
9	Brennan, Michael	West	82853	\$550.00	February
10	Brennan, Michael	West	72949	\$400.00	March
11	Davis, William	South	55223	\$235.00	February
12	Davis, William	South	10354	\$850.00	January
13	Davis, William	South	50192	\$600.00	March
14	Davis, William	South	27589	\$250.00	January
15	Dumlao, Richard	West	67275	\$400.00	January



Ambiti di applicazione



“ *The sexy job in the next 10 years will be statisticians.*

People think I'm joking, but who would've guessed that computer engineers would've been the sexy job of the 1990s?

Hal Varian, chief economist at Google

|| Data Scientist Moderno

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau





Data Store



mongoDB



MySQL

Joins are what
RDBMS's
do for a living

Transform



Spark



If you change
the way
you look at things,
the things you
look at change.

Model



scikit
learn

pandas



Spark

AI usually
beats
natural stupidity

Visualize



Dont trust everything
you see,
even
salt looks like sugar

Other Tools



kafka



Luigi

I have enough tools,
said no
data scientist
ever

Data Science workflow

Un mondo di strumenti



python™



Cassandra



mongoDB



membase



riak

The background of the slide is a dark blue field filled with a complex, glowing network of thin, golden-yellow lines. These lines connect numerous small, circular nodes of the same color, creating a dense, web-like pattern that resembles a data network or a molecular structure. The nodes are more concentrated in some areas, forming clusters, while other areas are sparser. The overall effect is one of dynamic connectivity and digital complexity.

R for Data Science



Cos'è R



free software
social
easy

powerful
visualization tool

based on
packages

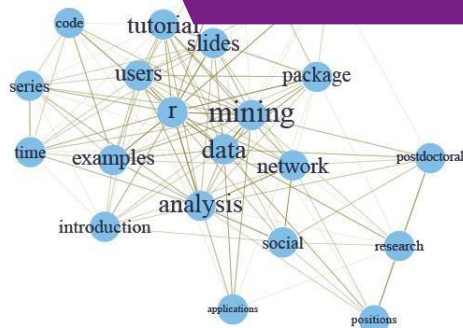
statistical
focus

RStudio IDE

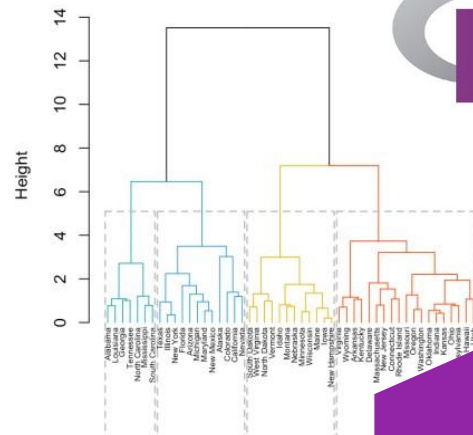
Visualization



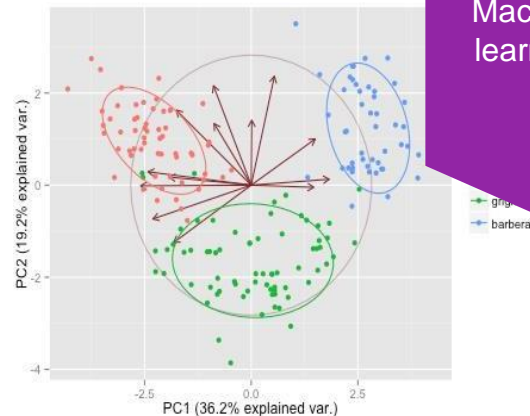
Text mining



Cluster Dendrogram



Machine learning



Alcuni esempi #R

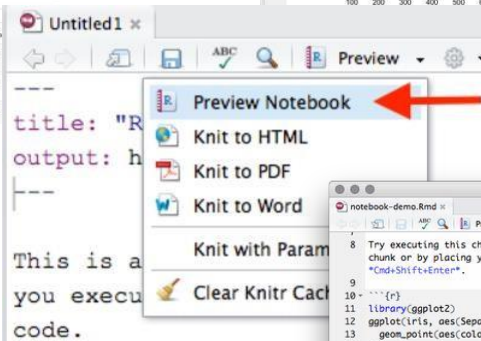
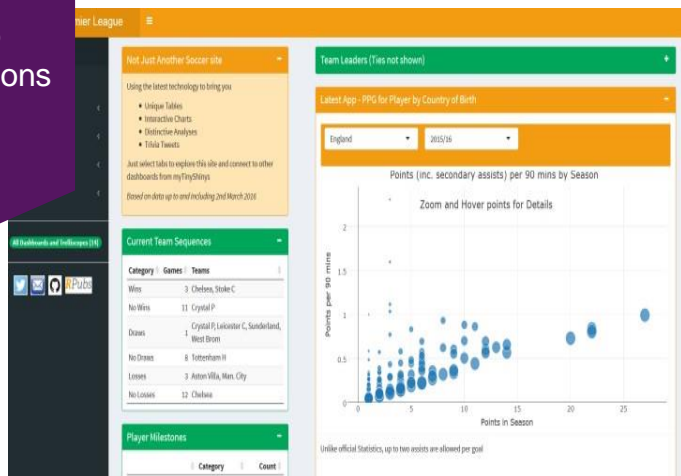


Real-time Dallas Police Calls

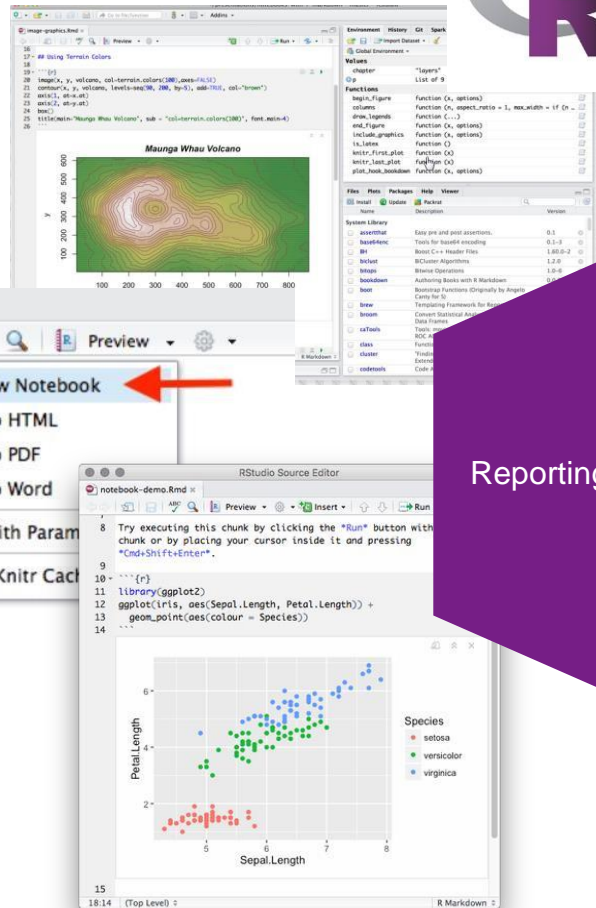
Showing the 83 calls the Dallas Police are responding to as of 5/23/2016 7:35:53 PM

Incident	Nature	Priority	DateTime	UnitNum	Block	Street	Beat	Reps
16-100468	QADP Out Arrow Shooter Foot	3	May 23 2016 12:24PM	D175	4800	Bryan St	136	2013
16-100469	QADP Out Armed Proponent Foot	3	May 23 2016 7:00PM	C355	3000	South Blvd	343	2113
16-100503	48A - COT w/Amulance	1	May 23 2016 6:00PM	C723	3800	E Overton Rd	757	4591

Web applications



Reporting



Coding in R

```
arrangements.R
6 #' @useDynLib arrangements
7 "_PACKAGE"
8
9 Arrangements <- R6::R6Class(
10   c("Arrangements", "iter", "abstractiter"),
11   private = list(
12     state = NULL
13   ),
14   public = list(
15     nextElem = function() {
16       out <- self$getNext()
17       is.null(out) && stop("StopIteration", call. = FALSE)
18       out
19     }
20   )
21 )
```

16:26 nextElem() R Script

Console Terminal Jobs

~/Dropbox/R/arrangements/

Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

Environment History Connections Build Git

New Connection

Connection	Status
------------	--------

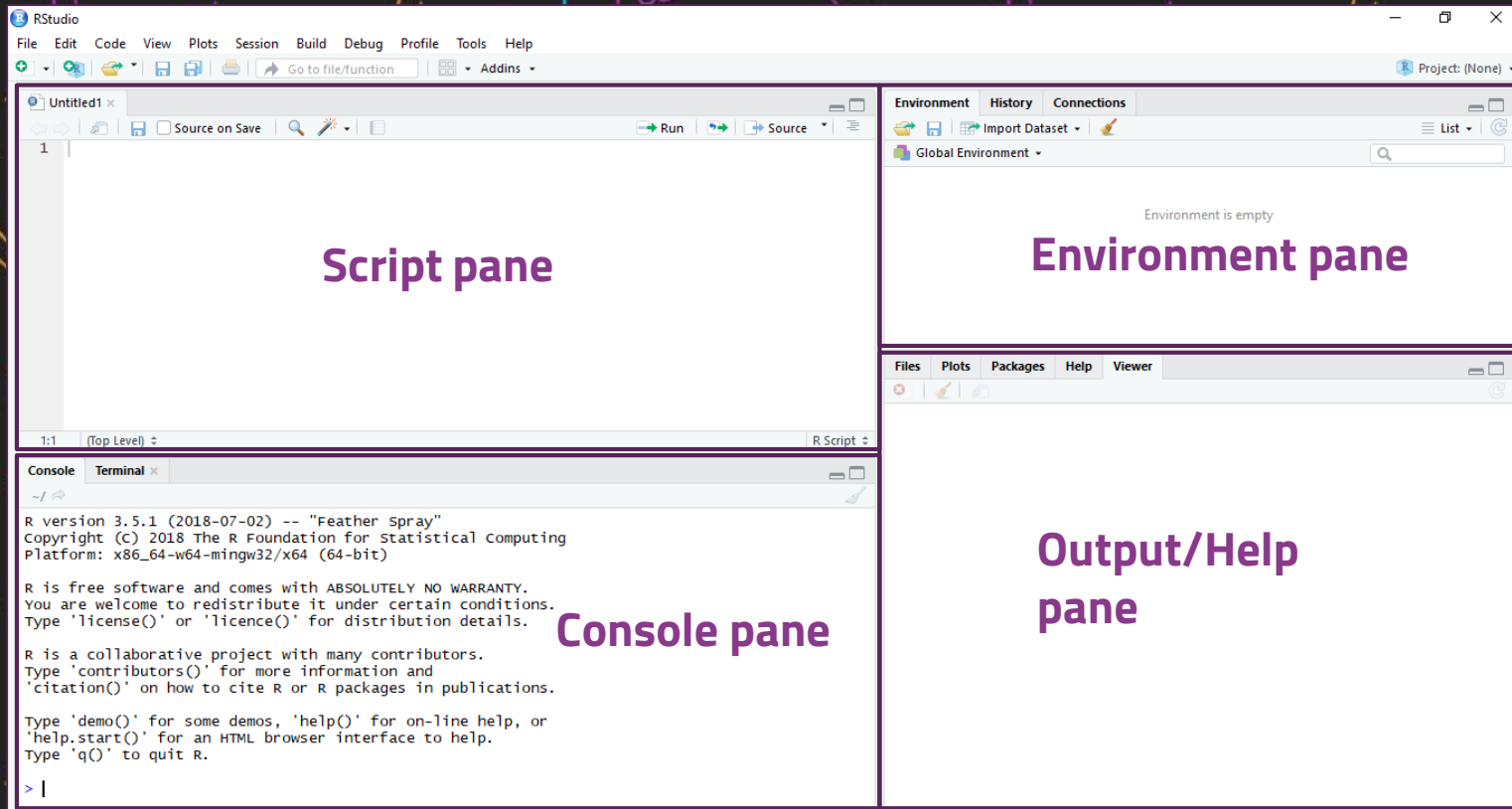
Files Plots Packages Help Viewer

New Folder Delete Rename More

Home > Dropbox > R > arrangements > R

Name	Size	Modified
..		
arrangements.R	465 B	Aug 9, 2018, 2:32 PM
combinations.R	5.4 KB	Aug 13, 2018, 10:59 PM
partitions.R	4.8 KB	Aug 13, 2018, 11:22 PM
permutations.R	5.7 KB	Aug 13, 2018, 11:13 PM
utils.R	783 B	Aug 23, 2018, 1:48 AM

What RStudio IDE offers





Cosa faremo?



Network dei
termini

WordCloud

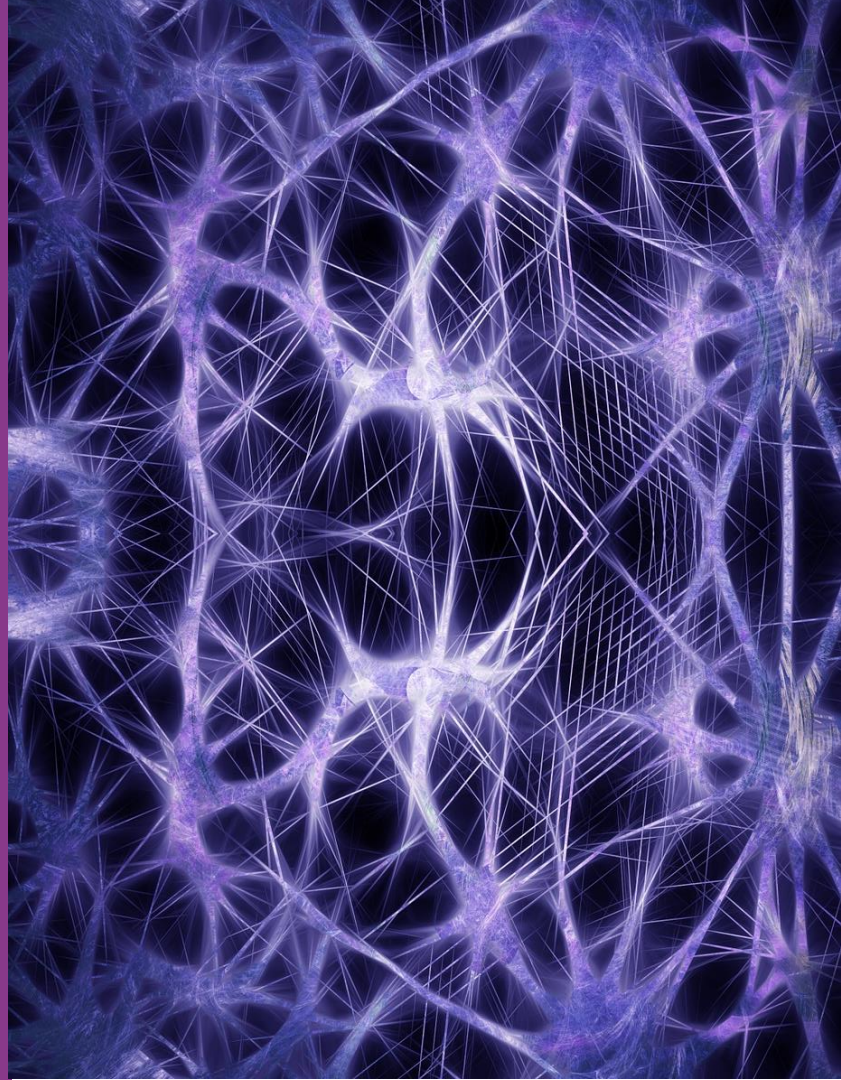
WordCloud



letteralmente
“nuvola di parole”

rappresentazione
grafica delle
parole più frequenti
in un testo

dimensioni e
frequenza delle parole
sono direttamente
proporzionali



Network di termini



dimensioni del
link e **frequenza**
delle relazioni sono
direttamente
proporzionali

raccomandazione
grafica delle
relazioni tra i termini

permette di
evidenziare
relazioni tra **gruppi
di parole**

Preparazione del testo

TESTO

TOKENIZZAZIONE

RIMOZIONE
PUNTEGGIATURA & STOP
WORDS

LEMMATIZZAZIONE

OUTPUT

[1] "THE BOY WHO LIVED Mr.
and Mrs. Dursley, of number f
our, Privet Drive, were proud
to say that they were perfec
tly normal, thank you very mu
ch. They were the last people
you'd expect to be involved
in anything strange or myster
ious, because they just didn'
t hold with such nonsense. M
r. Dursley was the director o
f a firm called Grunnings, wh
ich made drills. He was a big

Pre-processing

[1]	"BOY"	"LIVED"
[3]	"Dursley"	"Privet"
[5]	"Drive"	"proud"
[7]	"perfectly"	"normal"
[9]	"people"	"expect"
[11]	"involved"	"strange"
[13]	"mysterious"	"hold"
[15]	"nonsense"	"Dursley"
[17]	"director"	"firm"
[19]	"called"	"Grunnings"
[21]	"drills"	"beefy"
[23]	"neck"	"mustache"
[25]	"Dursley"	"thin"

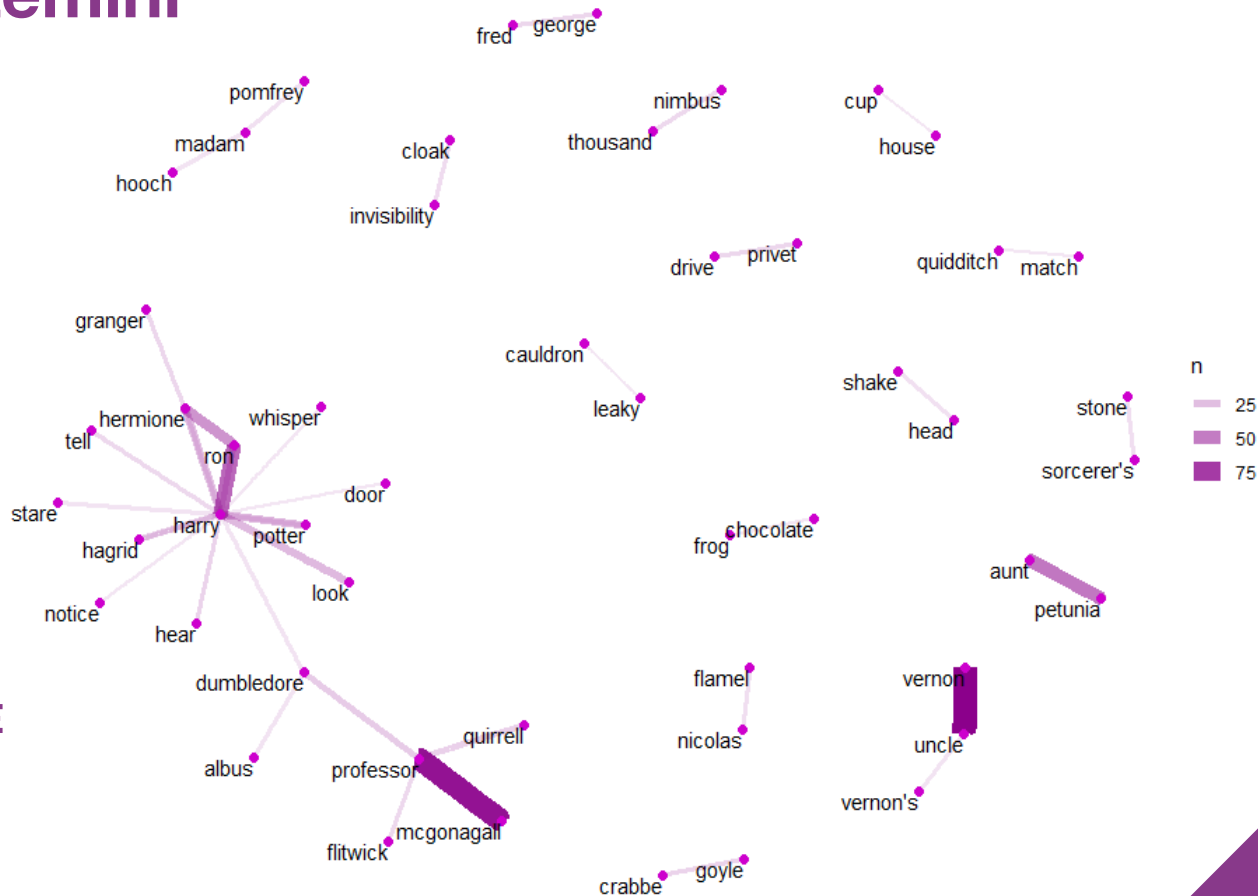


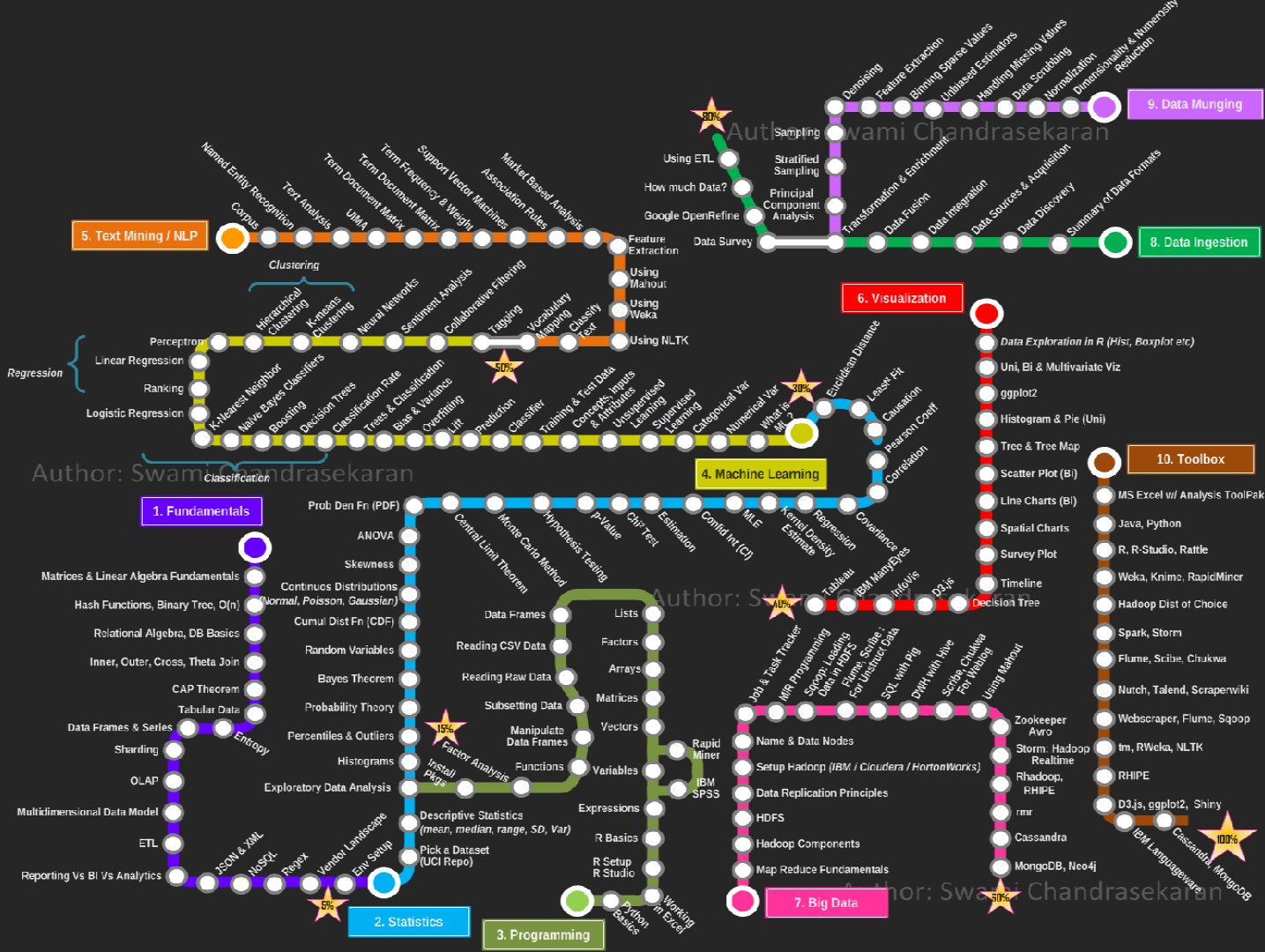
Network dei temini

CREAZIONE
BIGRAMMI

OUTPUT

TABELLA DI
FREQUENZE





Be ready
for the
journey!

Good Luck!