

Introduzione

Laboratorio di Data Science
in Economia
CLEBA



Roberto Benedetti

Dipartimento di Economia, email
benedett@unich.it

Buona creanza presentarsi. Chi sono ?

Titolo di studio:

Institution:

University of Rome “La Sapienza” 1988 Degree in Economic Statistics

University of Rome “La Sapienza” 1994 PhD in Methodological Statistics

Interessi di Ricerca:

Economic Statistics, Spatial Analysis and Survey Sampling

Esperienza Professionale:

	<u>Company Institution</u>	<u>Position</u>	<u>Description</u>
2001 ad oggi	University “G. D’Annunzio” of ChPescara	Full Professor	Econ. Statistics
1994 – 2001	Istat – National Statistical Institute	Dal 1994: Senior Researcher; Dal 2001 Research Director. Head of the Italian Agricultural Statistical Service	Met. & Agr.. Stat.
1989 - 1994	Telespazio S.p.A. per le comunicazioni spaziali	Researcher	Statistical analysis of remotely sensed data
1987 - 1989	SOMEA S.p.A. (Mathematical and Applied Economic society)	Researcher	I/O matrix estimation,

Alcune definizioni: Business analytics (BA)

Business analytics (BA) refers to all the methods and techniques that are used by an organization to measure performance. Business analytics are made up of **statistical methods** that can be applied to a specific project, process or product. Business analytics can also be used to evaluate an entire company. Business analytics are performed in order to identify weaknesses in existing processes and highlight meaningful data that will help an organization prepare for future growth and challenges.

The need for good business analytics has spurred the creation of business analytics software and enterprise platforms that mine an organization's data in order to automate some of these measures and pick out meaningful insights.

<https://www.techopedia.com/definition/344/business-analytics-ba>

Alcune definizioni: Data Science

- Data science is a broad field that refers to the collective processes, theories, concepts, tools and technologies that enable the review, analysis and extraction of valuable knowledge and information from raw data. It is geared toward helping individuals and organizations make better decisions from stored, consumed and managed data.
- Data science is formerly known as datalogy.

<https://www.techopedia.com/definition/30202/data-science>

La scienza dei dati è l'insieme di principi metodologici (basati sul metodo scientifico) e tecniche multidisciplinari volto a interpretare ed estrarre conoscenza dai dati attraverso la relativa fase di analisi da parte di un esperto (Data Scientist). I metodi della scienza dei dati (spesso associati al concetto di data mining) si basano su tecniche provenienti da varie discipline, principalmente da matematica, statistica, scienza dell'informazione, informatica e scienze sociali, in particolar modo nei seguenti sottodomini: basi di dati e visualizzazione dati o business intelligence, intelligenza artificiale o apprendimento automatico (vedi anche Big data). Il ruolo di scienziato dei dati è stato definito dalla «Harvard Business Review» come "la professione più attraente del XXI secolo" ed è considerato uno dei quattro ruoli chiave preposti all'utilizzo sistematico dei Big data nelle aziende.

https://it.wikipedia.org/wiki/Scienza_dei_dati

Alcune definizioni: Data Mining

- Il data mining (letteralmente dall'inglese estrazione di dati) è l'insieme di tecniche e metodologie che hanno per oggetto l'estrazione di informazioni utili da grandi quantità di dati (es. database, datawarehouse ecc...), attraverso metodi automatici o semi-automatici (es. machine learning) e l'utilizzo scientifico, aziendale/industriale o operativo delle stesse.
- https://it.wikipedia.org/wiki/Data_mining

Alcune definizioni: Big Data

- Il termine big data ("grandi [masse di] dati" in inglese), o megadati, indica genericamente una raccolta di dati così estesa in termini di volume, velocità e varietà da richiedere tecnologie e metodi analitici specifici per l'estrazione di valore o conoscenza. Il termine è utilizzato in riferimento alla capacità (propria della scienza dei dati) di analizzare ovvero estrapolare e mettere in relazione un'enorme mole di dati eterogenei, strutturati e non strutturati, allo scopo di scoprire i legami tra fenomeni diversi (ad esempio correlazioni) e prevedere quelli futuri. La disciplina può essere vista come un'evoluzione dei tradizionali metodi di business intelligence, allargata a moli di dati ancor più variegata e soprattutto più voluminose.
- https://it.wikipedia.org/wiki/Big_data

Alcune definizioni: Statistica

La statistica è una disciplina che ha come fine lo studio quantitativo e qualitativo di un particolare fenomeno collettivo in condizioni di incertezza o non determinismo, cioè di non completa conoscenza di esso o parte di esso.

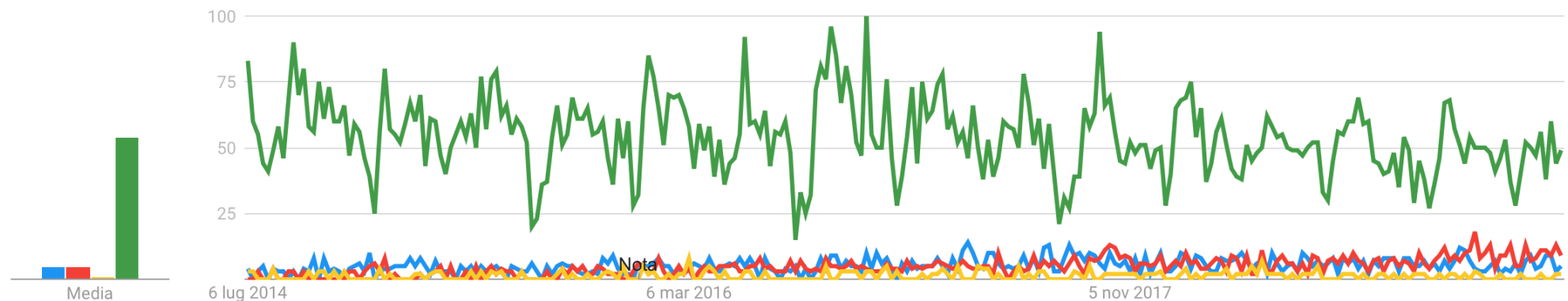
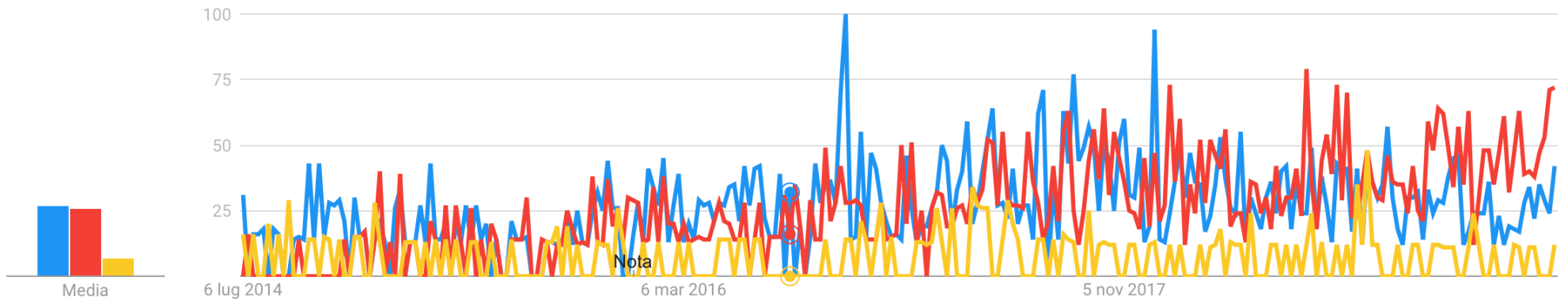
Essa è uno strumento del metodo scientifico[1] e si avvale della matematica e del metodo sperimentale per studiare i modi in cui un fenomeno collettivo può essere sintetizzato e compreso. Ciò avviene attraverso la raccolta e l'analisi delle informazioni relative al fenomeno studiato.

Con il termine statistica, nel linguaggio di tutti i giorni, si indicano anche semplicemente i risultati numerici (le statistiche richiamate nei telegiornali, ad esempio l'inflazione, il PIL ecc.) di un processo di sintesi dei dati osservati, cioè gli indici statistici..

- <https://it.wikipedia.org/wiki/Statistica>

Analizziamo i Trend (Italia)

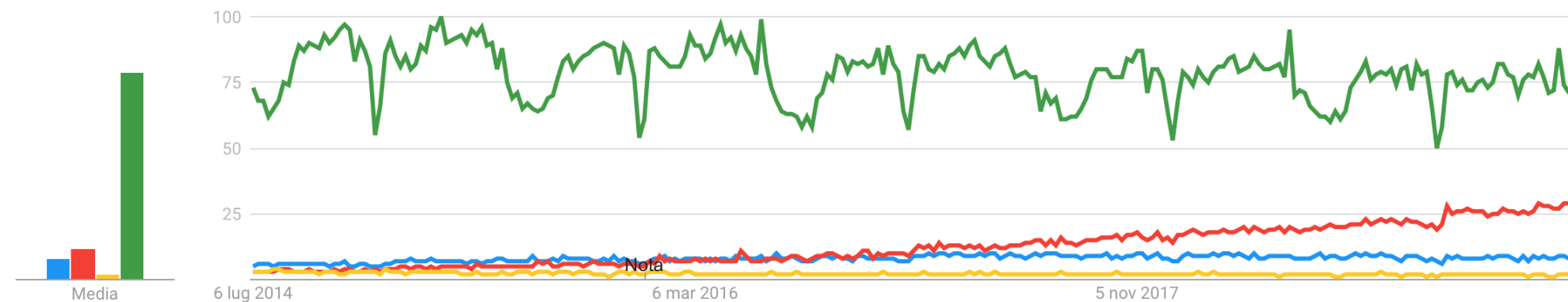
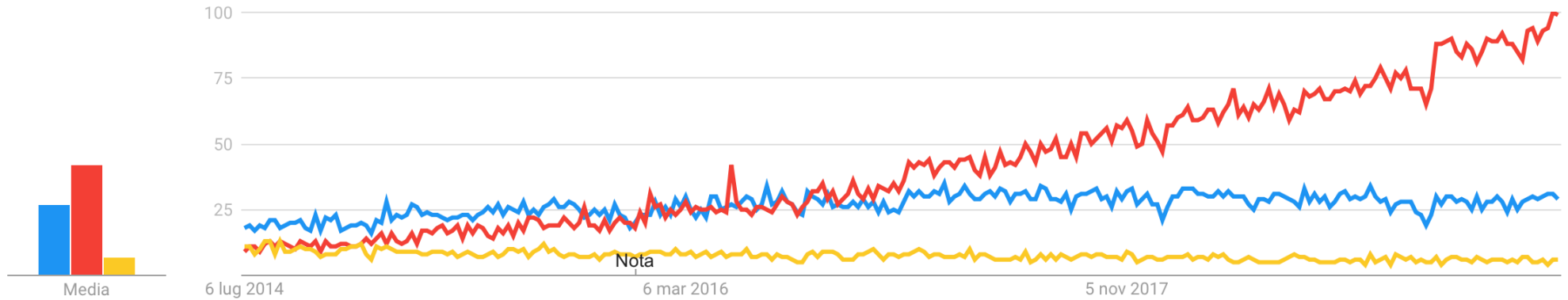
Big Data, Data Science, Data Mining



Big Data, Data Science, Data Mining, Statistica

Analizziamo i Trend (Mondo)

Big Data, Data Science, Data Mining



Big Data, Data Science, Data Mining, Statistica

Argomenti trattati

- Cluster Analysis
- Analisi statistica dei dati di Network
- Regressione (selezione regressori, stime robuste, regressione non lineare, regressione non parametrica, regressione quantilica etc)
- Analisi delle serie storiche

Dove trovare i dati

- Github

<https://github.com/awesomedata/awesome-public-datasets>

- Kaggle datasets

<https://www.kaggle.com/datasets>

- The Koblenz Network Collection

<http://konect.uni-koblenz.de>

- Katya Ognyanova Network datasets

<https://kateto.net/2016/05/network-datasets/>

- Stanford Large Network Dataset Collection

<http://snap.stanford.edu/data/index.html>

- Microsoft R Application Network

<https://mran.revolutionanalytics.com/documents/data>

- Awesome R

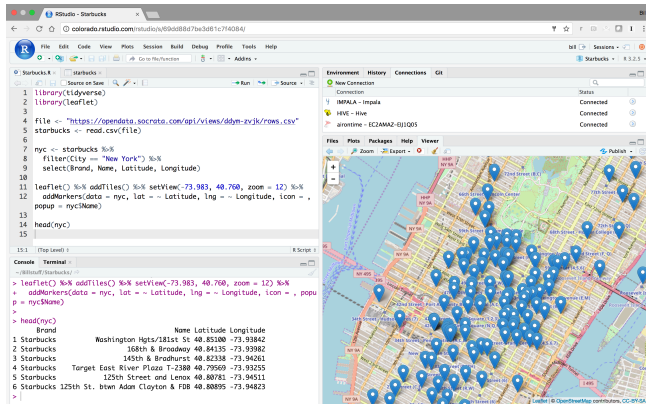
<https://awesome-r.com>



Che sistema utilizzare (quale IDE per R)

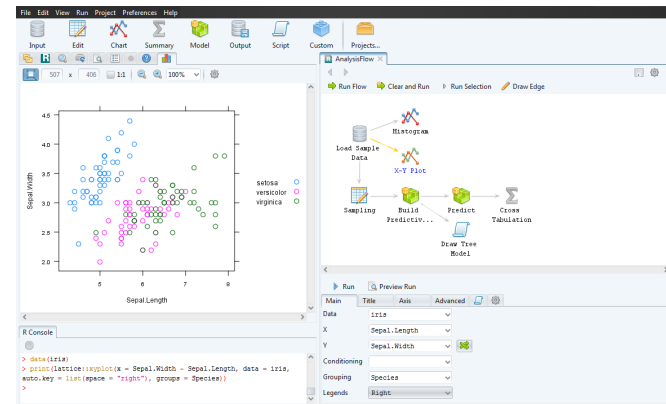
- RStudio

<https://rstudio.com/products/rstudio/>



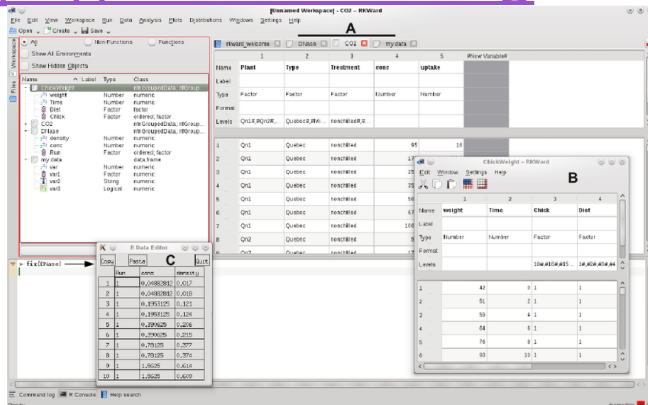
- R AnalyticFlow

<http://r.analyticflow.com/en/>



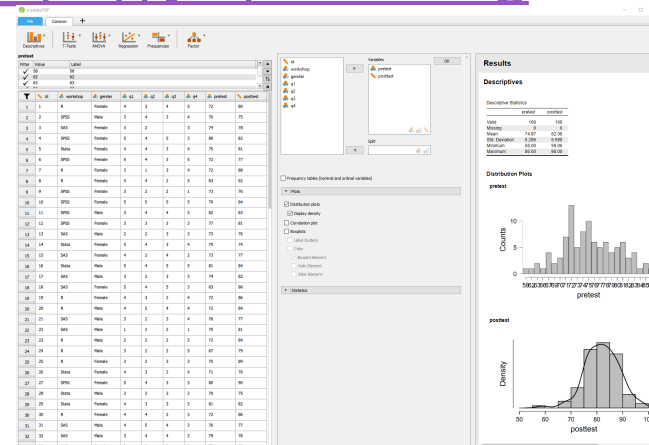
- RKWard

<https://rkward.kde.org>



- JASP

<https://jasp-stats.org>



Possiamo iniziare:

**3, 2, 1, ...lift off, we have a lift off ...tower cleared
(lancio Apollo 11)**

