

```
library(dplyr)
```

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



**Kickoff event**

**July 20<sup>th</sup> 2017**

**19-20.30**



**Big thanks to  
our sponsor  
& host**

# Who are Rladies?



- Founded by statistician Gabriela de Queiroz
- ▷ San Francisco, Oct 2012
- ▶ To create a community of R-ladies users to encourage ladies to:
  - ▷ Join the coding community
  - ▷ Develop R packages
  - ▷ Increase the attendance/visibility of women as speakers at R conferences and meetings



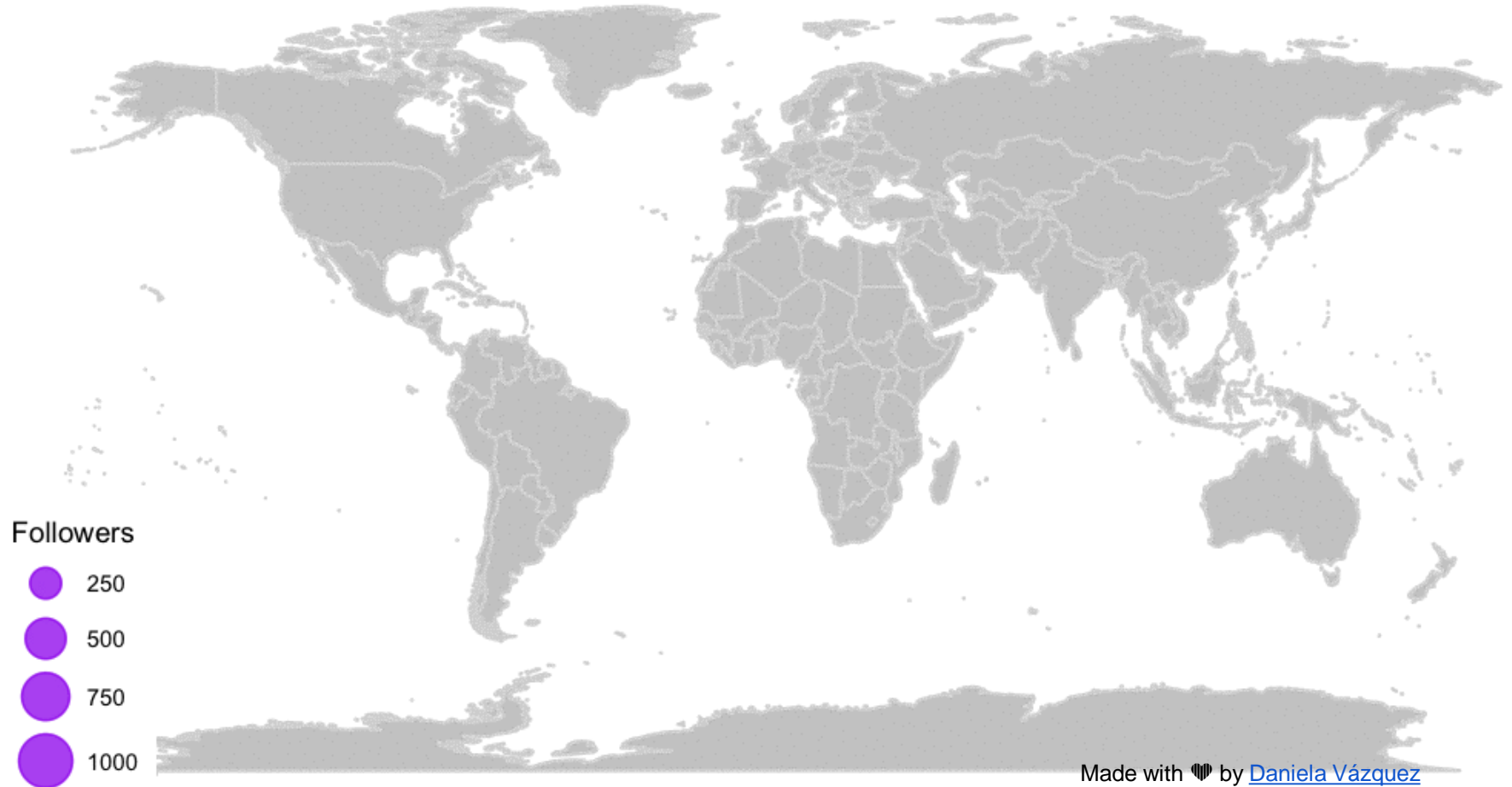
# Where are ladies now?

 **+20 COUNTRIES**

 **+70 CHAPTERS**

2011-09-01

## Twitter followers by each chapter





# 14%

R package developers are female

# 26%

Only make up data professionals

# 20%

Of all tech startups across the world are founded by women

- 1) <http://blog.revolutionanalytics.com/2016/09/all-the-r-ladies.html>
- 2) <https://www.betterbuys.com/bi/women-in-data-science/>
- 3) <http://www.dailyinfographic.com/33-must-know-facts-about-women-in-technology>



# Rladies Milan

- Founded in 30 june 2017
- 35 members





***Parvaneh shafiei***

- Senior Data scientist @ EY
- Studied in computer science in Polimi
- 3 years experience in data processing & analytics & machine learning
- My interests are machine learning, artificial intelligence and analytics





***Erica Yu***

- Studied architecture in Polimi
- Working as a freelance designer
- Passionate about analytics & data mining
- Intend to find a job related to data science



# Our mission is simple

## Why we are here

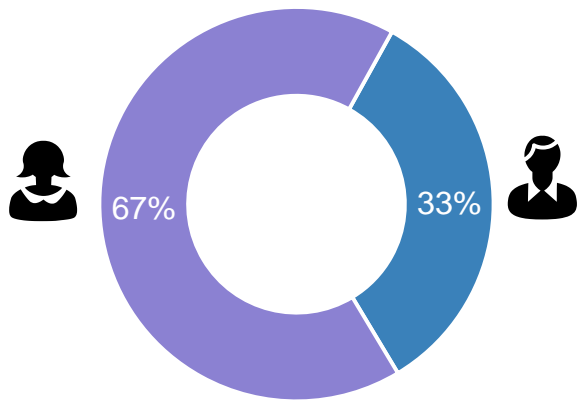
- Gather who are passionate about data science
- Improve R knowledge
- Know new people
- Launch new ideas

## Our goal

- Monthly workshops
- Monthly talks
- Monthly projects
- Facilitate the access to R community
- Grow network



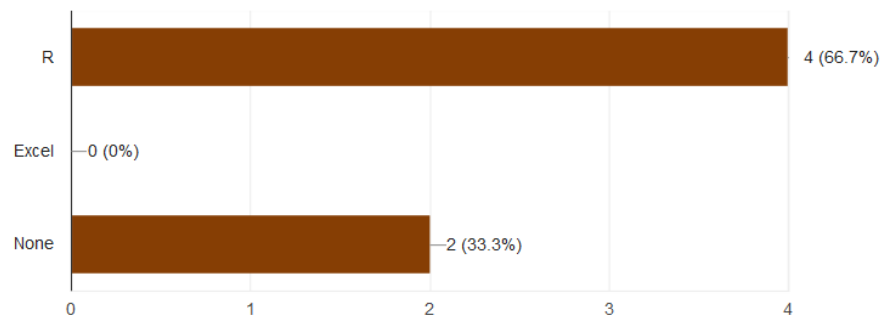
# About you



1 out of 2 participated in survey

Do you have experience in analytics by R or Excel?

6 responses

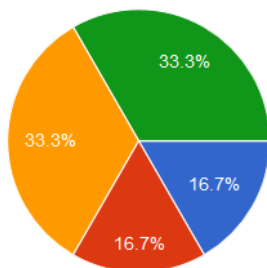


# About you



## What is your R level?

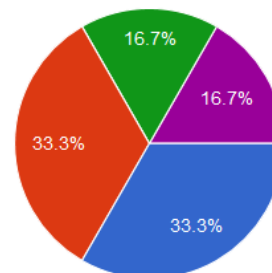
6 responses



- Beginner
- Intermediate
- Advanced
- I dont know R

## What is you job title?

6 responses



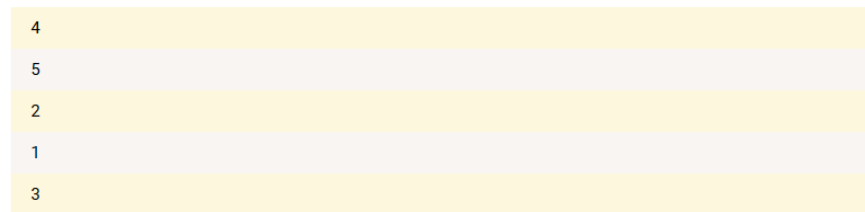
- Analyst
- Developer/Engineer
- Executive
- IT / Technical Manager
- Researcher / Academic
- Solutions Architect
- System Administrator
- Student

# About you



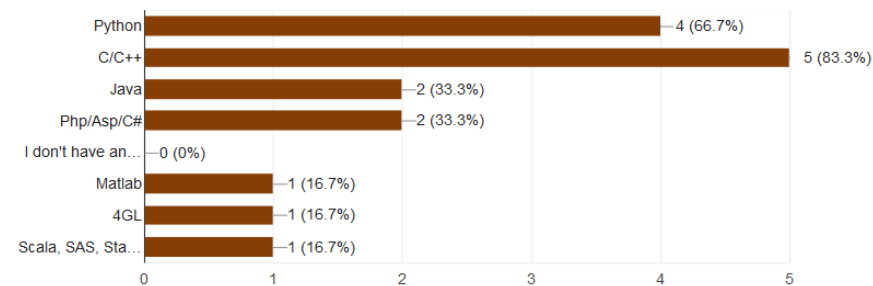
If you are working, please provide how many years

5 responses



Do you have programming experience with any of listed languages?

6 responses

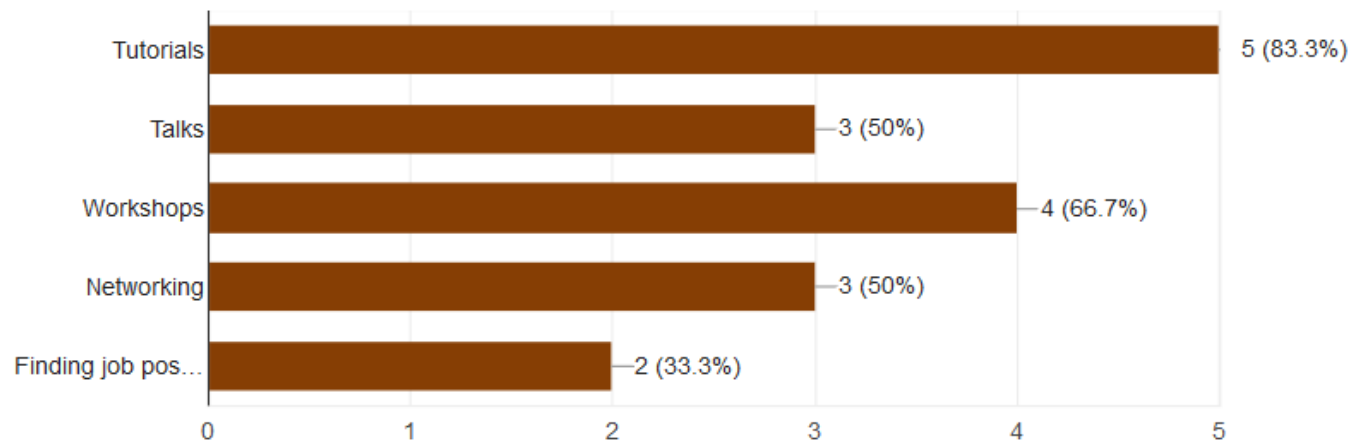


# About you

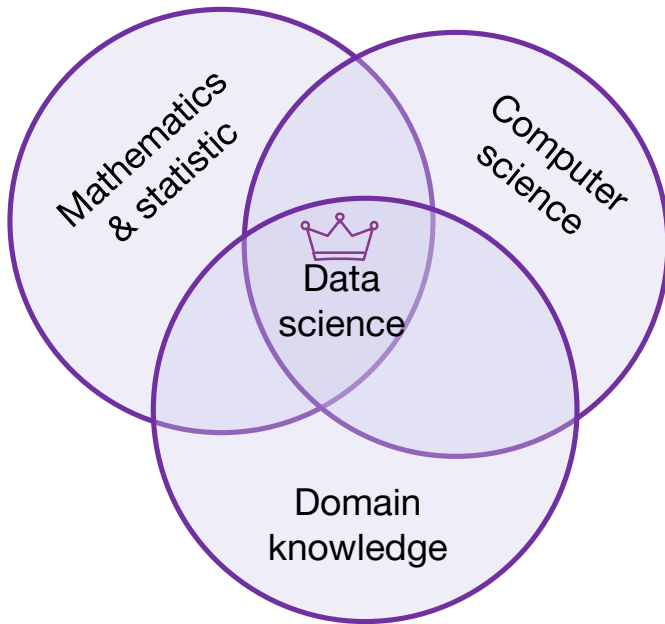


## What would you like to see in R-Ladies Milan?

6 responses



# What is data science?



## Data is the new oil!

- It is the art of analyzing and extracting patterns and trends from the data through various tools
- It is about discovering new insights that they are not visible just by raw numbers

# What is the workflow of data analytics?

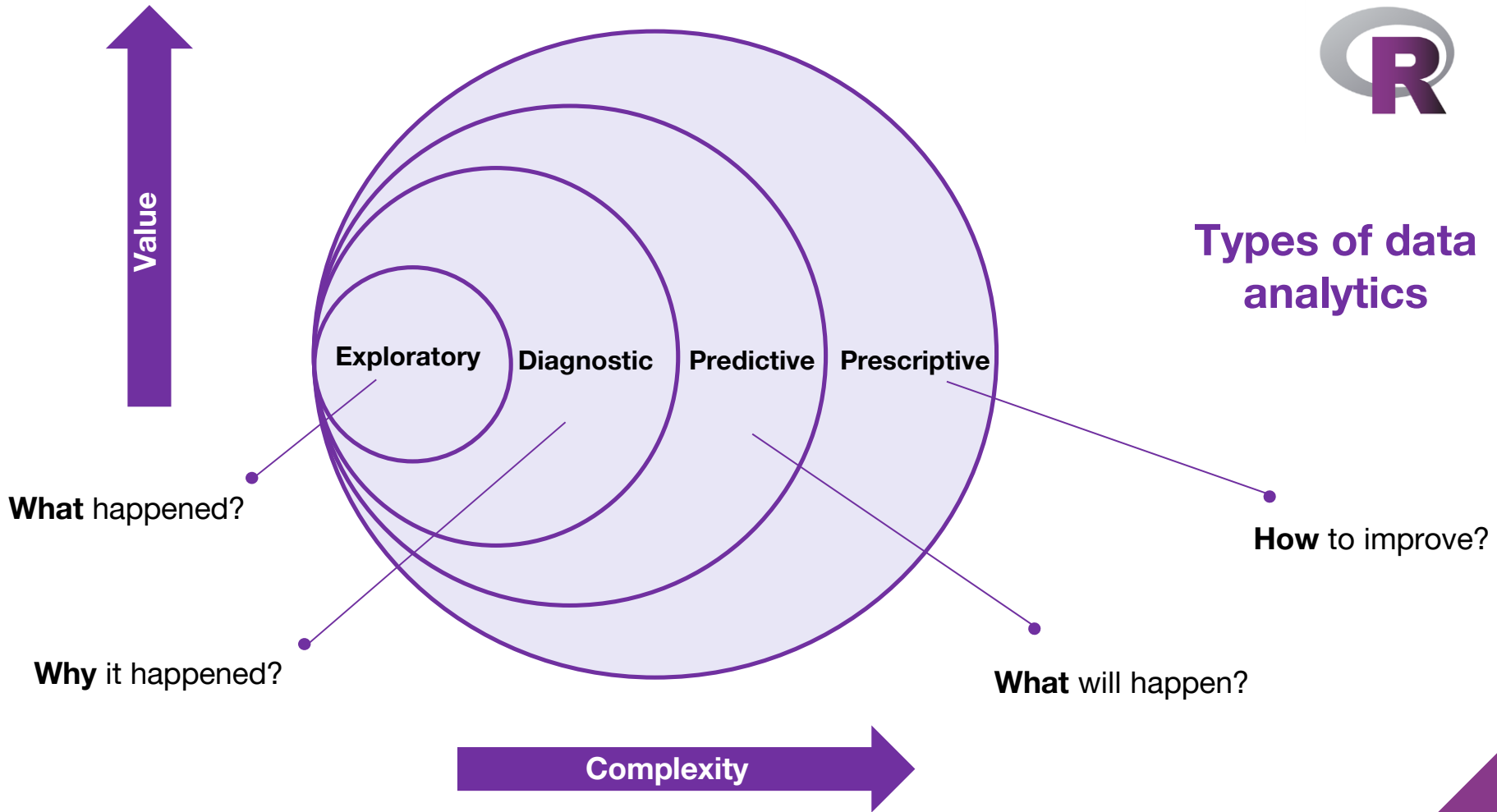


**Make invisible, visible!**





## Types of data analytics





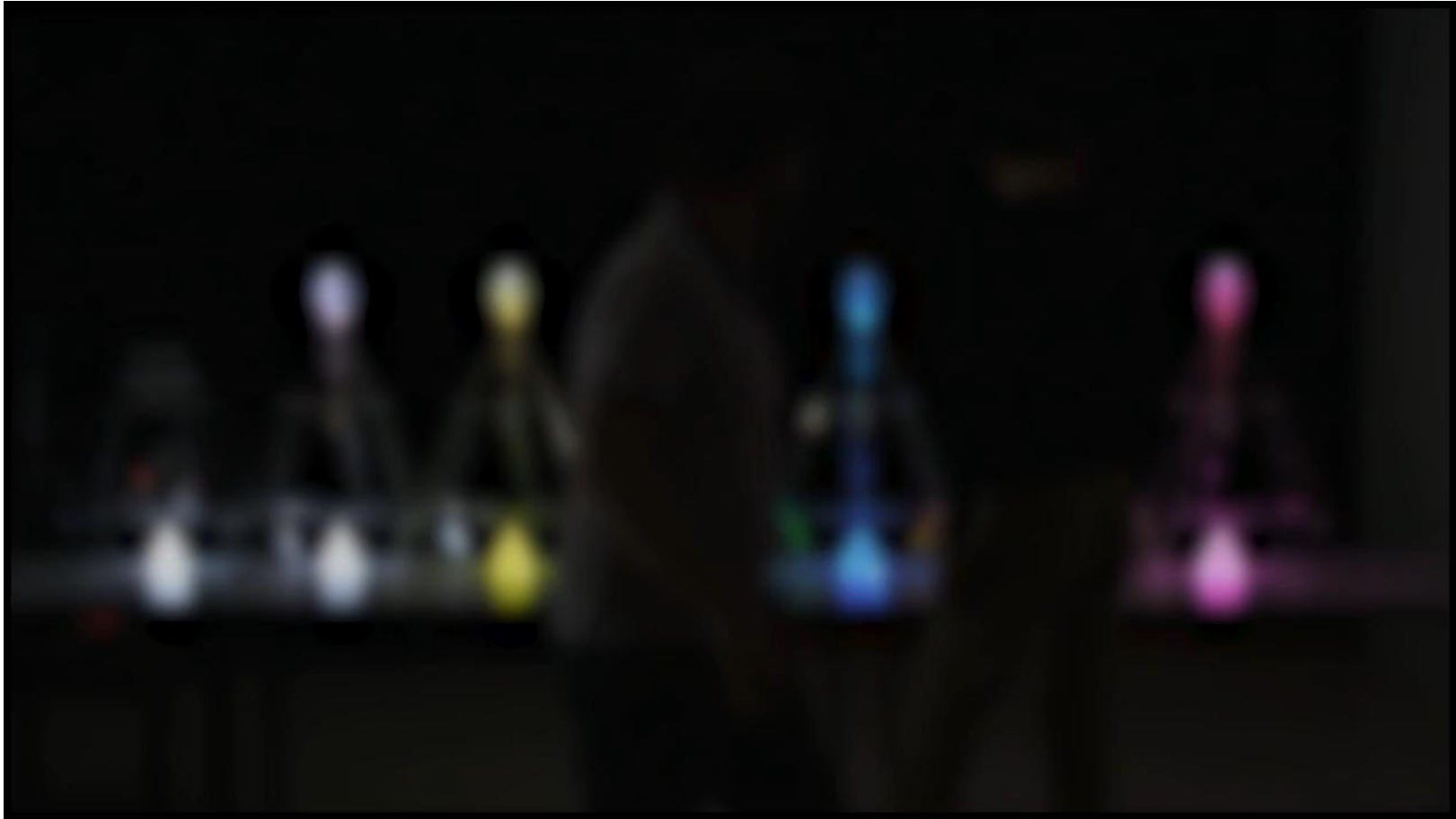
**Data  
Driven  
Sculpture**



**Data  
Dress**



**Data  
Cocktail**



“ *A data scientist is better at statistic than a software engineer and is better at software engineering than a statistician* ”



Josh Wills  
@josh\_wills

Follow

Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

9:55 AM - 3 May 2012

1,613 Retweets 1,254 Likes



## MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21st century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

### 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

### 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

### DOMAIN KNOWLEDGE & SOFT SKILLS

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

### 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 scientists

- Use analytics & technical skills to extract data



### Statistician

- Understand statistic and apply it on real problems



### Data engineer

- Responsible for architecture of data
- Ensure the flow of data within servers & applications





*R* is a free software environment for statistical computing, graphics, machine learning, etc.  
It consists of various packages which are focused for different objectives and use cases.

---

## Some example of R usages

---

Visualization

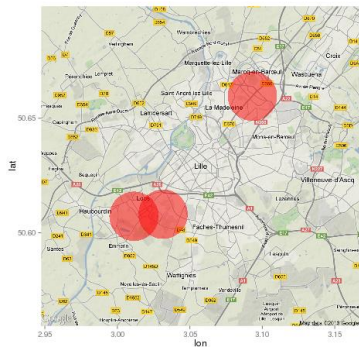
Text mining

Machine  
learning

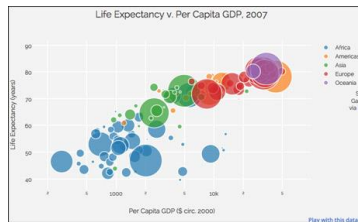
Web  
applications

Reporting

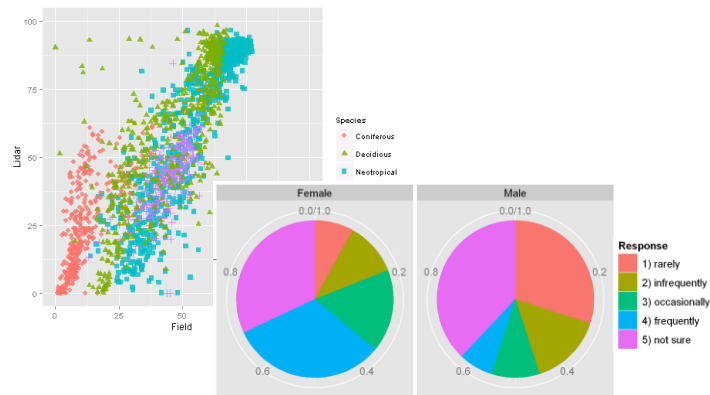
# Visualization



ggmap



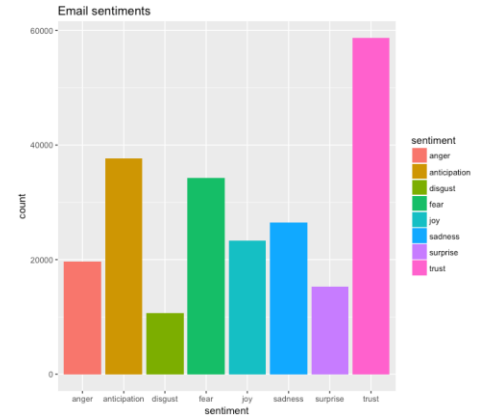
htmlwidgets for R



ggplot2

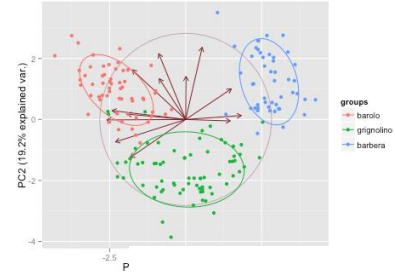
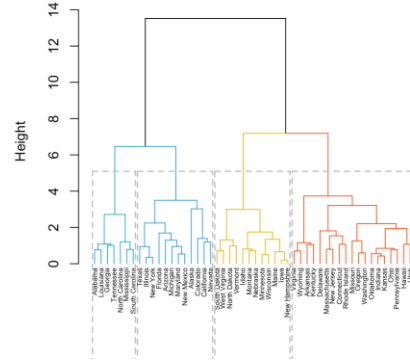


tm

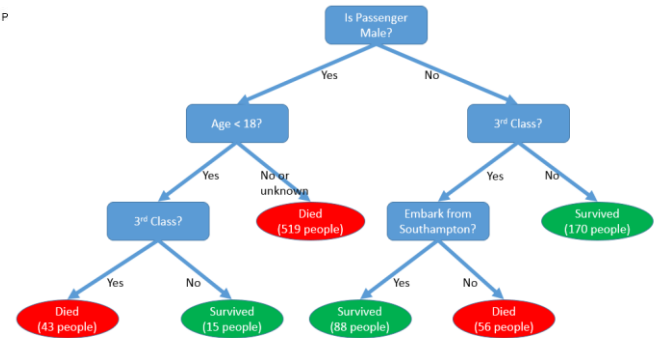


# Machine learning

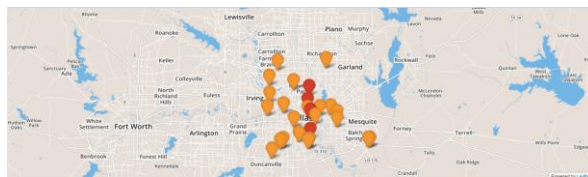
Cluster Dendrogram



the caret package



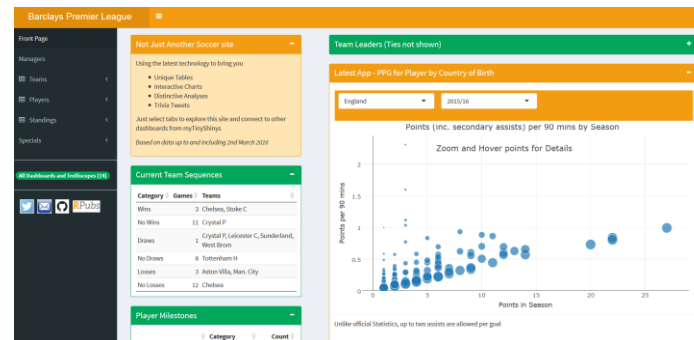
# Web applications



Real-time Dallas Police Calls

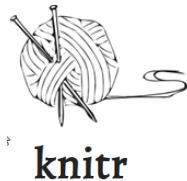
Showing the 40 calls the Dallas Police are responding to as of 5/25/2018 1:35:53 PM

Incident	Name	Priority	Date/Time	Latitude	Block	Street	Dist	Reporting/Info
18-035588	SHRP One Armed Prosser Foot	1	May 25 2018 12:24PM	32.75	4000	Bryan St	250	2102
18-035468	SHRP One Armed Prosser Foot	1	May 25 2018 1:08PM	32.85	3000	South Blvd	340	2113
18-035023	WV - COT WV/No/No/No	1	May 25 2018 1:08PM	32.71	2800	E Chilton Rd	717	4301





# Reporting



The screenshot displays the RStudio Source Editor with a notebook titled 'notebook-demo.Rmd'. The code chunk contains the following R code:

```
8 Try executing this chunk by clicking the "Run" button within the
9 chunk or by placing your cursor inside it and pressing
10 "Cmd+Shift+Enter".
11 library(ggplot2)
12 ggplot(iris, aes(Sepal.Length, Petal.Length)) +
13   geom_point(aes(colour = Species))
14
```

Below the code, a scatter plot of Petal.Length vs Sepal.Length is shown, with points colored by Species. The plot shows three distinct clusters of points (red, green, and blue) corresponding to the three species in the iris dataset.

Overlaid on the RStudio window is a 'Preview' menu for the notebook. The menu options are:

- Preview Notebook (highlighted with a red arrow)
- Knit to HTML
- Knit to PDF
- Knit to Word
- Knit with Parameters...
- Clear Knitr Cache...

The background of the menu shows a preview of the rendered HTML output, which includes the title 'title: "R"', the output 'output: h', and the text 'This is a you execu code.' followed by 'code.' and 'code.'.



# Why use R?

1. It is free!

- Free to install, use, update and clone

	SPSS	Matlab	Minitab	STATISTICA QC
Business	\$99,00 USD per user per month	€2,000 individual	€1,495 single license	€1,612,00 Single license

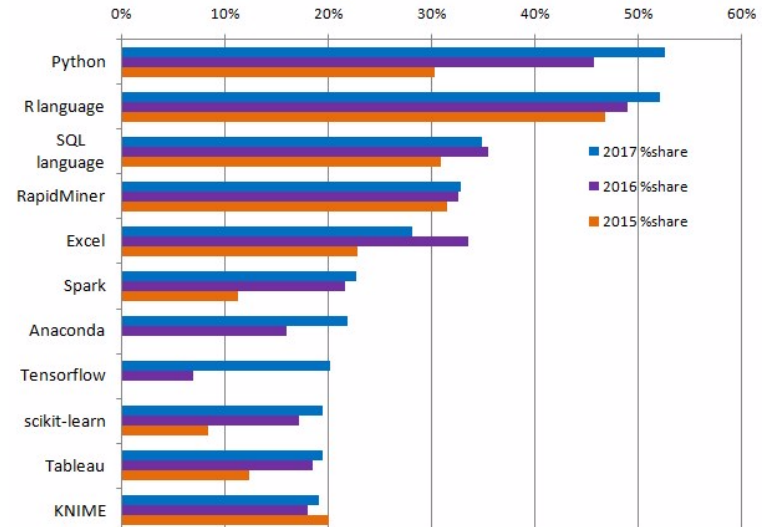


# Why use R?

2. It is popular

\*2017 KDnuggets survey

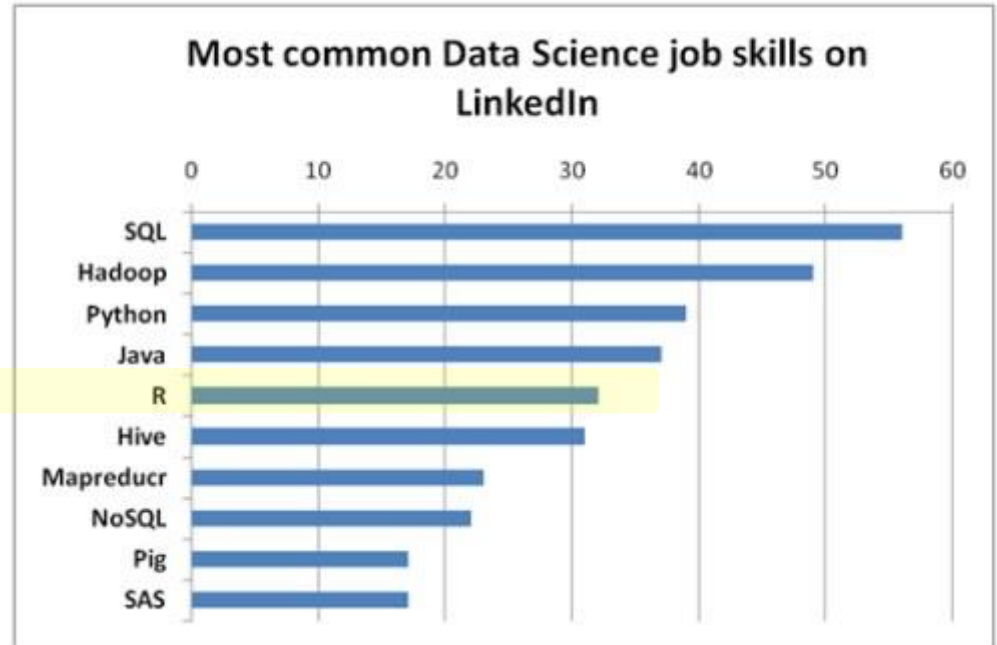
KDnuggets Analytics, Data Science, Machine Learning Software Poll, top tools share, 2015-2017





# Why use R?

3. It is among data science skills that it is necessary to learn



\*2016 CrowdFlower analysis on job position in linkedin



# Why use R?

3. Highest paid technical skill in US\*

AVERAGE SALARY FOR High Paying Skills and Experience		
SKILL	2013	YR/YR CHANGE
R	\$ 115,531	n/a
NoSQL	\$ 114,796	1.6%
MapReduce	\$ 114,396	n/a
PMBok	\$ 112,382	1.3%
Cassandra	\$ 112,382	n/a
Omnigraffle	\$ 111,039	0.3%
Pig	\$ 109,561	n/a
SOA (Service Oriented Architecture)	\$ 108,997	-0.5%
Hadoop	\$ 108,669	-5.6%
Mongo DB	\$ 107,825	-0.4%

\*2014 Dice Tech Salary Survey







# Why use R?

5. R is well supported by its community



stackoverflow



Revolutions

Daily news about using open source R for big data analysis, predictive modeling, data science, and visualization since 2008



# Resources to learn R - 1

The image shows the Udemy course page for 'R Programming A-Z™: R For Data Science With Real Exercises!'. The course is created by Kirill Eremenko and the SuperDataScience Team. It has a rating of 4.6 from 4,554 reviews and 27,362 students. The course is in English and was last updated in December 2016. The price is 10€, a 95% discount from the original price of 200€. A green 'ACQUISTA ORA' button is visible. Below the title, there is a section 'Cosa imparerò?' (What I will learn) with a list of topics: learning to program in R, core principles of programming, creating variables, using R Studio, creating vectors, and understanding data types. A video player thumbnail is also present.

udemy Categorie Cerca corsi Diventa un insegnante

Business Sviluppo IT & Software Produttività in ufficio Sviluppo personale Design Marketing Stile di vita Fotografia

Ottieni Udemy per la tua azienda Regala questo corso Lista dei desideri

## R Programming A-Z™: R For Data Science With Real Exercises!

Learn Programming In R And R Studio, Data Analytics, Data Science, Statistical Analysis, Packages, Functions, GGPLOT2

4.6 (4.554 valutazioni) 27.362 studenti iscritti

Creato da Kirill Eremenko, SuperDataScience Team Ultimo aggiornamento 12/2016 Inglese English

### Cosa imparerò?

- ✓ Learn to program in R at a good level
- ✓ Learn the core principles of programming
- ✓ Learn how to create variables
- ✓ Learn how to use R Studio
- ✓ Learn how to create vectors in R
- ✓ Learn about integer, double, logical, character and other types in R

Visualizza in anteprima questo corso

10 € 200 € 95% di sconto

A questo prezzo per 2 giorni!

ACQUISTA ORA

Aggiungi al carrello

Garanzia di rimborso entro 30 giorni



The image shows the Coursera course page for 'R Programming' by Johns Hopkins University. The course is part of the Data Science specialization. The page includes a navigation menu with links to Overview, Syllabus, FAQs, Creators, Pricing, and Ratings and Reviews. The 'About this course' section states that students will learn how to program in R and how to use R for effective data analysis. The 'Who is this class for?' section recommends some programming experience. The course is created by Johns Hopkins University. A blue 'Enroll' button is visible.

coursera Catalog Search catalog Fo

## R Programming

Home > Data Science > Data Analysis

**About this course:** In this course you will learn how to program in R and how to use R for effective data analysis. You will learn how to install and configure software necessary for a statistical programming environment and describe generic programming language concepts as they are implemented in a high-level statistical language. The course covers practical issues in statistical computing which includes

▼ More

**Who is this class for:** Some programming experience (in any language) is recommended.

**Created by:** Johns Hopkins University

JOHNS HOPKINS UNIVERSITY

Enroll

Starts Jul 17





# Resources to learn R - 2

 [Courses](#) ▾ [Programs](#) ▾ [Schools & Partners](#) [About](#) ▾  

[Home](#) > [All Subjects](#) > [Data Analysis & Statistics](#) > [Introduction to R for Data Science](#)



## Introduction to R for Data Science

Learn the R statistical programming language, the lingua franca of data science in this hands-on course.




### About this course

This course is part of the [Microsoft Professional Program Certificate in Data Science](#).

R is rapidly becoming the leading language in data science and statistics. Today, R is the tool of choice for data science professionals in every industry and field. Whether you are full-time number cruncher, or just the occasional data analyst, R will suit your needs.




 [Home](#) [Courses](#) [Tracks](#) [Pricing](#) [Business](#) [Community](#) [Sign in](#) [Create Free Account](#)

FREE COURSE

## Introduction to R

[Start Course For Free](#)



4 hours | 0 Videos | 62 Exercises | 445,778 Participants | 6200 XP

### Course Description

In this Introduction to R, you will master the basics of this beautiful open source language, including factors, lists and data frames. With the knowledge gained in this course, you will be ready to undertake your first very own data analysis. With over 2 million users worldwide R is rapidly becoming the leading programming language in statistics and data science. Every year, the number of R users grows by 40% and an increasing number of organizations are using it in their day-to-day activities. Leverage the power of R by completing this free R online course today!

This course is part of these tracks:

- Data Analyst with R
- Data Scientist with R
- R Developer
- R Programming





# Resources to learn R - 3

**Welcome to Kaggle Kernels**  
The best place to explore data science results and share your own work

```
library(kgboost)
library(Matrix)
input_dir = "
```

**Code**  
Skip the download. Kernels is preloaded with the most common data science languages and libraries.

**Learn**  
Gain exposure to new tools and techniques. The "hottest" kernels showcase the best work on Kaggle.

**Mentor**  
Give back by sharing what you know. You can answer questions and leave feedback on others' code and results.

[New Kernel](#)

**New to Data Science?**  
Get started with a tutorial on our most popular competition for beginners, *Titanic: Machine Learning from Disaster*.

[Learn more](#)

**Build a Model**  
Get the data & use whatever tools or methods you prefer to make predictions.

**Make a Submission**  
Upload your prediction file for real-time scoring & a spot on the leaderboard.

[Submit](#)

[Dismiss](#)

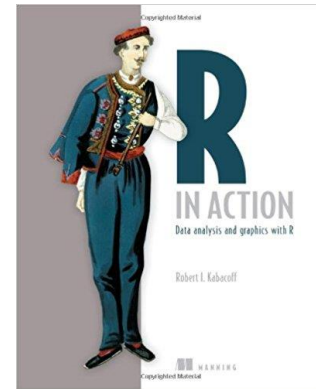
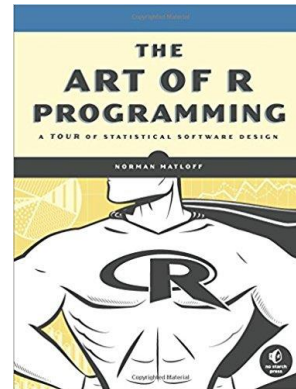
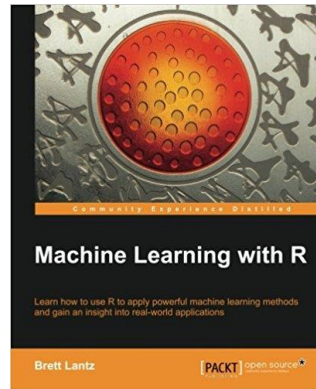
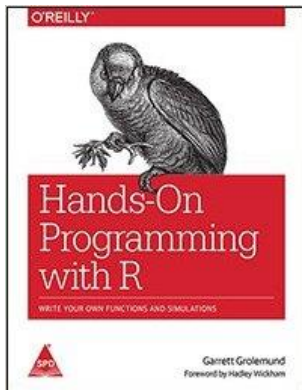
13 active competitions

Sort by: Prize

Active	All	Entered	All Categories	Search
	<b>Passenger Screening Algorithm Challenge</b> Improve the accuracy of the Department of Homeland Security's threat recognition algorithms <i>Featured</i> · 5 months to go	<b>\$1,500,000</b> 130 teams		
	<b>Zillow Prize: Zillow's Home Value Prediction (Zestimate)</b> Can you improve the algorithm that changed the world of real estate? <i>Featured</i> · 8 months to go	<b>\$1,200,000</b> 1,616 teams		

kaggle

# Resources to learn R - 4



Cookbook for R



RDocumentation



Be ready  
for the  
journey!

# Famous female data scientists (some examples)



## Hillary Mason

- Founder of FastForward Labs
- Fortune 40 under 40
- Top 100 most creative people in business by *Fast Company*



## Carla Gentry

- Data Scientist at Talent Analytics
- Experience over 20 years which includes working for Fortune 500 companies like Hershey, Kraft, Johnson & Johnson, Kellogg's and Firestone
- One of the most popular personalities in Big Data community to follow on Twitter



# Famous female data scientist (some examples)



**Diane Greene**

- Senior Vice President at Google
- Founder and CEO of VMWare
- Frontrunner in the advancement of cloud and data technology



**Daphne Koller**

- President and Co-Founder of Coursera
- Various awards such as ONR Young Investigator Award, ACM Infosys Award, MacArthur Foundation Fellowship, etc.
- Professor @ Stanford University for almost 18 years



## Follow us



<https://twitter.com/RladiesMilan>



<https://www.meetup.com/R-Ladies-Milan/>



<https://www.facebook.com/rladiesmilan/>



**We are  
looking  
forward for  
you ideas &  
feedbacks**

# Events



20 july  
Kickoff

September  
Workshop

November  
Project

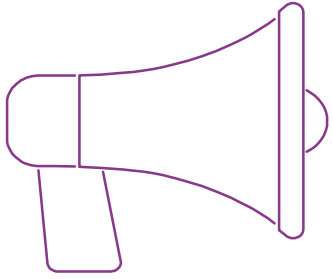


# How to get involved?

- Join our meetups
- Follow us in twitter & facebook

- Give a talk
- Lead a workshop
- Be a co-organizer

- Share your experience with others
- Ask your company to sponsor us



**We are looking  
always for  
sponsors and  
speakers**



**Print("Thank you")**

**MUCH R**

**SO LADIES**

**REALLY EVENT**

**WOW**

