⑂ jmportilla / **awesome-datascience**
forked from bulutyazilim/awesome-datascience

◉ Watch   3   ★ Star   4   ⑂ Fork   1,285

<> Code   ⑁ Pull requests 0   ▥ Projects 0   Insights ▾

📝 An awesome Data Science repository to learn and apply for real world problems.

◔ **205** commits	⑂ **1** branch	◌ **0** releases	⚌ **26** contributors

Branch: **master** ▾   New pull request

Find file   Clone or download

This branch is 59 commits behind bulutyazilim:master.

⑁ Pull request   ▣ Compare

|	**hmert** Merge pull request #66 from demidovakatya/master …	Latest commit cc7c6fd on Mar 19, 2016
🖹 .travis.yml	fix #60	2 years ago
🖹 Algorithms.md	some Data mining algorithms added	2 years ago
🖹 DataScienceMethodology.md	Data Science Methodology document added	a year ago
🖹 LICENSE	Initial commit	3 years ago
🖹 README.md	Fix images in the Infographic section	a year ago
🖹 contributing.md	Create contributing.md	2 years ago
🖹 githubGroups.md	create github group	2 years ago

📖 **README.md**

Awesome Data Science 👓 awesome

An open source Data Science repository to learn and apply towards solving real world problems.

Table of contents

- Motivation
- Infographic
- What is Data Science?
- Colleges
- MOOC's
- Data Sets
- Bloggers
- Facebook Accounts
- Twitter Accounts
- YouTube Videos & Channels
- Toolboxes - Environment
- Journals, Publications and Magazines
- Presentations
- Other Awesome Lists

Motivation
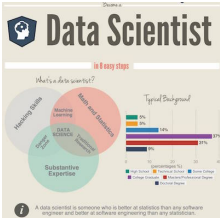
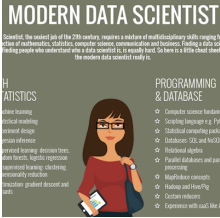*This part is for dummies who are new to Data Science*

This is a shortcut path to start studying **Data Science**. Just follow the steps to answer the questions, "What is Data Science and what should I study to learn Data Science?"
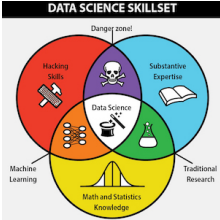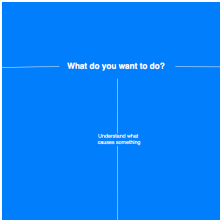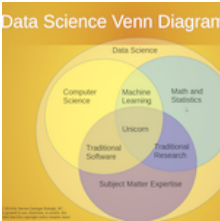
First of all, Data Science is one of the hottest topics on the Computer and Internet farmland nowadays. People have gathered data from applications and systems until today and now is the time to analyze them. The next steps are producing suggestions from the data and creating predictions about the future. Here you can find the biggest question for **Data Science** and hundreds of answers from experts. Our favorite data scientist is Clare Corthell. She is an expert in data-related systems and a hacker, and has been working on a company as a data scientist. Clare's blog. This web site helps you to understand the exact way to study as a professional data scientist.

Secondly, Our favorite programming language is *Python* nowadays for #DataScience. Python's - Pandas library has full functionality for collecting and analyzing data. We use Anaconda to play with data and to create applications.

This is the Guide to begin a **Data Science** project.

Infographic

Preview	Description
	A visual guide to Becoming a Data Scientist in 8 Steps by DataCamp (img)
	Mindmap on required skills (img)
	Swami Chandrasekaran made a Curriculum via Metro map.
	by @kzawadz via twitter, MarketingDistillery.com
	And a male version, from another article by MarketingDistillery.com

Preview	Description
	By [Data Science Central](#)
	From [this article](#) by Berkeley Science Review.
	Data Science Wars: R vs Python
	How to select statistical or machine learning techniques
	The Data Science Industry: Who Does What
	Data Science Venn Diagram

What is Data Science?

- [What is Data Science @ O'reilly](#)
- [What is Data Science @ Quora](#)
- [The sexiest job of 21st century](#)
- [What is data science](#)
- [What is a data scientist](#)
- [Wikipedia](#)
- [a very short history of #datascience](#)

- An Introduction to Data Science, PDF.
- Data Science Methodology by John Rollins PhD

COLLEGES

- A list of colleges and universities offering degrees in data science.
- Data Science Degree @ Berkeley
- Data Science Degree @ UVA
- Data Science Degree @ Wisconsin

MOOC's

- Google Making Sense of Data
- Coursera Introduction to Data Science
- Data Science - 9 Steps Courses, A Specialization on Coursera
- Data Mining - 5 Steps Courses, A Specialization on Coursera
- Machine Learning – 5 Steps Courses, A Specialization on Coursera
- CS 109 Data Science
- Schoolofdata
- OpenIntro
- Data science MOOC
- CS 171 Visualization
- Process Mining: Data science in Action
- Oxford Deep Learning
- Oxford Deep Learning - video
- Oxford Machine Learning
- UBC Machine Learning - video
- Data Science Specialization
- [Coursera Big Data Specialization] (https://www.coursera.org/specializations/big-data)
- Data Science and Analytics in Context by Edx
- Big Data University by IBM
- Udacity - Deep Learning

Data Sets

- Academic Torrents
- hadoopilluminated.com
- data.gov - The home of the U.S. Government's open data
- United States Census Bureau
- freebase.com
- usgovxml.com
- enigma.io - Navigate the world of public data - Quickly search and analyze billions of public records published by governments, companies and organizations.
- datahub.io
- aws.amazon.com/datasets
- databib.org
- datacite.org
- quandl.com - Get the data you need in the form you want; instant download, API or direct to your app.
- figshare.com

- [GeoLite Legacy Downloadable Databases](#)
- [Quora's Big Datasets Answer](#)
- [Public Big Data Sets](#)
- [Houston Data Portal](#)
- [Kaggle Data Sources](#)
- [Kaggle Datasets](#)
- [A Deep Catalog of Human Genetic Variation](#)
- [A community-curated database of well-known people, places, and things](#)
- [Google Public Data](#)
- [World Bank Data](#)
- [NYC Taxi data](#)
- [Open Data Philly](#) Connecting people with data for Philadelphia
- [A list of useful sources](#) A blog post includes many data set databases
- [grouplens.org](#) Sample movie (with ratings), book and wiki datasets
- [UC Irvine Machine Learning Repository](#) - contains data sets good for machine learning
- [research-quality data sets](#) by [Hilary Mason](#)
- [National Climatic Data Center - NOAA](#)
- [ClimateData.us](#) (related: [U.S. Climate Resilience Toolkit](#))
- [r/datasets](#)
- [MapLight](#) - provides a variety of data free of charge for uses that are freely available to the general public. Click on a data set below to learn more
- [GHDx](#) - Institute for Health Metrics and Evaluation - a catalog of health and demographic datasets from around the world and including IHME results
- [St. Louis Federal Reserve Economic Data - FRED](#)
- [Dept. of Politics @ New York University](#)
- [Open Data Sources](#)
- [UNICEF Statistics and Monitoring](#)
- [UNICEF Data](#)
- [undata](#)
- [NASA SocioEconomic Data and Applications Center - SEDAC](#)
- [The GDELT Project](#)
- [Sweden, Statistics](#)
- [Github free data source list](#)
- [StackExchange Data Explorer](#) - an open source tool for running arbitrary queries against public data from the Stack Exchange network.
- [San Fransisco Government Open Data](#)
- [IBM Blog abour open data](#)
- [Google Public Data](#)
- [Open data Index](#)

Bloggers

- [Wes McKinney](#) - Wes McKinney Blog.
- [Matthew Russell](#) - Mining The Social Web.
- [Greg Reda](#) - Greg Reda Personal Blog
- [Kevin Davenport](#) - Kevin Davenport Personal Blog
- [Julia Evans](#) - Recurse Center alumna
- [Hakan Kardas](#) - Personal Web Page
- [Sean J. Taylor](#) - Personal Web Page

- [Drew Conway](#) - Personal Web Page
- [Hilary Mason](#) - Personal Web Page
- [Noah Iliinsky](#) - Personal Blog
- [Matt Harrison](#) - Personal Blog
- [Data Science Renee](#) Documenting my path from "SQL Data Analyst pursuing an Engineering Master's Degree" to "Data Scientist"
- [Vamshi Ambati](#) - AllThings Data Sciene
- [Prash Chan](#) - Tech Blog on Master Data Management And Every Buzz Surrounding It
- [Clare Corthell](#) - The Open Source Data Science Masters
- [Paul Miller](#) Based in the UK and working globally, Cloud of Data's consultancy services help clients understand the implications of taking data and more to the Cloud.
- [Data Science London](#) Data Science London is a non-profit organization dedicated to the free, open, dissemination of data science. We are the largest data science community in Europe. We are more than 3,190 data scientists and data geeks in our community.
- [Datawrangling](#) by Peter Skomoroch. MACHINE LEARNING, DATA MINING, AND MORE
- [John Myles White](#) Personal Blog
- [Quora Data Science](#) - Data Science Questions and Answers from experts
- [Siah](#) a PhD student at Berkeley
- [Data Science Report](#) MDS, Inc. Helps Build Careers in Data Science, Advanced Analytics, Big Data Architecture, and High Performance Software Engineering
- [Louis Dorard](#) a technology guy with a penchant for the web and for data, big and small
- [Machine Learning Mastery](#) about helping professional programmers to confidently apply machine learning algorithms to address complex problems.
- [Daniel Forsyth](#) - Personal Blog
- [Data Science Weekly](#) - Weekly News Blog
- [Revolution Analytics](#) - Data Science Blog
- [R Bloggers](#) - R Bloggers
- [The Practical Quant](#) Big data
- [Micheal Le Gal](#) a data enthusiast who gets hooked on solving intriguing problems and crafting beautiful stories and visualizations with data. Over the past 5 years, He haas applied statistics to solve problems in government, brain sciences, and most recently, retail.
- [Datascope Anayltics](#) data-driven consulting and design
- [Yet Another Data Blog](#) Yet Another Data Blog
- [Spenczar](#) a data scientist at *Twitch*. I handle the whole data pipeline, from tracking to model-building to reporting.
- [KD Nuggets](#) Data Mining, Analytics, Big Data, Data, Science not a blog a portal
- [Meta Brown](#) - Personal Blog
- [Data Scientist](#) is building the data scientist culture.
- [WhatSTheBigData](#) is some of, all of, or much more than the above and this blog explores its impact on information technology, the business world, government agencies, and our lives.
- [Mic Farris](#) Focusing on science, datascience, business, technology, and channeling inner geekness!
- [Tevfik Kosar](#) - Magnus Notitia
- [New Data Scientist](#) How a Social Scientist Jumps into the World of Big Data
- [Harvard Data Science](#) - Thoughts on Statistical Computing and Visualization
- [Data Science 101](#) - Learning To Be A Data Scientist
- [Kaggle Past Solutions](#)
- [DataScientistJourney](#)
- [NYC Taxi Visualization Blog](#)
- [Learning Lover](#)
- [Huge Trello List of Great Data Science Resources](#)

- Dataists
- Data-Mania
- Data-Magnum
- Map Reduce Blog
- FastML Blog
- P-value - Musings on data science, machine learning and stats.
- datascopeanalytics
- Digital transformation
- datascientistjourney
- Data Mania Blog
- The File Drawer - Chris Said's science blog
- Emilio Ferrara's web page
- DataNews
- Reddit TextMining
- Periscopic
- Hilary Parker
- Data Stories
- Data Science Lab
- Meaning of
- Adventures in Data Land
- DATA MINERS BLOG
- Dataclysm
- FlowingData - Visualization and Statistics
- Calculated Risk
- Applied Data Labs - content and news about data-driven business.
- O'reilly Learning Blog
- Dominodatalab
- i am trask - A Machine Learning Craftsmanship Blog
- Vademecum of Practical Data Science - Handbook and recipes for data-driven solutions of real-world problems

Facebook Accounts

- Data
- Big Data Scientist
- Data Science 101
- Data Science Day
- Data Science Academy
- Facebook Data Science Page
- Data Science London
- Data Science Technology and Corporation
- Data Science - Closed Group
- Center for Data Science
- Big data hadoop NOSQL Hive Hbase
- Analytics, Data Mining, Predictive Modeling, Artificial Intelligence
- Big Data Analytics using R
- Big Data Analytics with R and Hadoop
- Big Data Learnings
- Big Data, Data Science, Data Mining & Statistics

- [BigData/Hadoop Expert](#)
- [Data Mining / Machine Learning / AI](#)
- [Data Mining/Big Data - Social Network Ana](#)
- [Vademecum of Practical Data Science](#)
- [Veri Bilimi Istanbul](#)

Twitter Accounts

- [Big Data Combine](#) - Rapid-fire, live tryouts for data scientists seeking to monetize their models as trading strategies
- [Big Data Mania](#) - Data Viz Wiz | Data Journalist | Growth Hacker | Author of Data Science for Dummies (2015)
- [Big Data Science](#) - Big Data, Data Science, Predictive Modeling, Business Analytics, Hadoop, Decision and Operations Research.
- [Charlie Greenbacker](#) - Director of Data Science at @ExploreAltamira
- [Chris Said](#) - Data scientist at Twitter
- [Clare Corthell](#) - Dev, Design, Data Science @mattermark #hackerei
- [DADI Charles-Abner](#) - #datascientist @Ekimetrics. , #machinelearning #dataviz #DynamicCharts #Hadoop #R #Python #NLP #Bitcoin #dataenthousiast
- [Data Science Central](#) - Data Science Central is the industry's single resource for Big Data practitioners.
- [Data Science London](#) Data Science. Big Data. Data Hacks. Data Junkies. Data Startups. Open Data
- [Data Science Renee](#) - Documenting my path from SQL Data Analyst pursuing an Engineering Master's Degree to Data Scientist
- [Data Science Report](#) - Mission is to help guide & advance careers in Data Science & Analytics
- [Data Science Tips](#) - Tips and Tricks for Data Scientists around the world! #datascience #bigdata
- [Data Vizzard](#) - DataViz, Security, Military
- [DataScienceX](#)
- [deeplearning4j](#) -
- [DJ Patil](#) - White House Data Chief, VP @ RelateIQ.
- [Domino Data Lab](#)
- [Drew Conway](#) - Data nerd, hacker, student of conflict.
- [Emilio Ferrara](#) - #Networks, #MachineLearning and #DataScience. I work on #Social Media. Postdoc at @IndianaUniv
- [Erin Bartolo](#) - Running with #BigData--enjoying a love/hate relationship with its hype. @iSchoolSU #DataScience Program Mgr.
- [Greg Reda](#) Working @ *GrubHub* about data and pandas
- [Gregory Piatetsky](#) - KDnuggets President, Analytics/Big Data/Data Mining/Data Science expert, KDD & SIGKDD co-founder, was Chief Scientist at 2 startups, part-time philosopher.
- [Gregory Piatetsky](#) - KDnuggets President, Analytics/Big Data/Data Mining/Data Science expert, KDD & SIGKDD co-founder, was Chief Scientist at 2 startups, part-time philosopher.
- [Hakan Kardas](#) - Data Scientist
- [Hilary Mason](#) - Data Scientist in Residence at @accel.
- [Jeff Hammerbacher](#) ReTweeting about data science
- [John Myles White](#) Scientist at Facebook and Julia developer. Author of Machine Learning for Hackers and Bandit Algorithms for Website Optimization. Tweets reflect my views only.
- [Juan Miguel Lavista](#) - Principal Data Scientist @ Microsoft Data Science Team
- [Julia Evans](#) - Hacker - Pandas - Data Analyze
- [Kenneth Cukier](#) - The Economist's Data Editor and co-author of Big Data (http://big-data-book.com).
- [Kevin Davenport](#) - Organizer of http://sddatascience.com
- [Kim Rees](#) - Interactive data visualization and tools. Data flaneur.
- [Kirk Borne](#) - DataScientist, PhD Astrophysicist, Top #BigData Influencer.
- [Linda Regber](#) - Data story teller, visualizations.

- [Luis Rei](#) - PhD Student. Programming, Mobile, Web. Artificial Intelligence, Intelligent Robotics Machine Learning, Data Mining, Natural Language Processing, Data Science.
- [Machine Learning](#) - Live Content Curated by top 1K Machine Learning Experts
- [Mark Stevenson](#) - Data Analytics Recruitment Specialist at Salt (@SaltJobs) | Analytics - Insight - Big Data - Datascience
- [Matt Harrison](#) - Opinions of full-stack Python guy, author, instructor, currently playing Data Scientist. Occasional fathering, husbanding, ult|goalt-imate, organic gardening.
- [Matthew Russell](#) - Mining the Social Web.
- [Mert Nuhoğlu](#) Data Scientist at BizQualify, Developer
- [Monica Rogati](#) - Data @ Jawbone. Turned data into stories & products at LinkedIn. Text mining, applied machine learning, recommender systems. Ex-gamer, ex-machine coder; namer.
- [Noah Iliinsky](#) - Visualization & interaction designer. Practical cyclist. Author of vis books: http://www.oreillynet.com/pub/au/4419
- [Paul Miller](#) - Cloud Computing/ Big Data/ Open Data Analyst & Consultant. Writer, Speaker & Moderator. Gigaom Research Analyst.
- [Peter Skomoroch](#) - Creating intelligent systems to automate tasks & improve decisions. Entrepreneur, ex Principal Data Scientist @LinkedIn. Machine Learning, ProductRei, Networks
- [Prash Chan](#) - Solution Architect @ IBM, Master Data Management, Data Quality & Data Governance Blogger. Data Science, Hadoop, Big Data & Cloud.
- [Quora Data Science](#) Quora's data science topic
- [R-Bloggers](#) - Tweet blog posts from the R blogosphere, data science conferences and (!) open jobs for data scientists.
- [Rand Hindi](#)
- [Randy Olson](#) - Computer scientist researching artificial intelligence. Data tinkerer. Community leader for @DataIsBeautiful. #OpenScience advocate.
- [Recep Erol](#) - Data Science geek @ UALR
- [Ryan Orban](#) - Data scientist, genetic origamist, hardware aficionado
- [Sean J. Taylor](#) - Social Scientist. Hacker. Facebook Data Science Team. Keywords: Experiments, Causal Inference, Statistics, Machine Learning, Economics.
- [Silvia K. Spiva](#) - #DataScience at Cisco
- [Spencer Nelson](#) - Data nerd
- [Talha Oz](#) - Enjoys ABM, SNA, DM, ML, NLP, HI, Python, Java. Top percentile kaggler/data scientist
- [Tasos Skarlatidis](#) - Complex Event Processing, Big Data, Artificial Intelligence and Machine Learning. Passionate about programming and open-source.
- [Terry Timko](#) - InfoGov; Bigdata; Data as a Service; Data Science; Open, Social & Business Data Convergence
- [TextDataMiningReddit](#)
- [Tony Baer](#) - IT analyst with Ovum covering Big Data & data management with some systems engineering thrown in.
- [Tony Ojeda](#) - Data Scientist | Author | Entrepreneur. Co-founder @DataCommunityDC. Founder @DistrictDataLab. #DataScience #BigData #DataDC
- [Vamshi Ambati](#) - Data Science @ PayPal. #NLP, #machinelearning; PhD, Carnegie Mellon alumni (Blog: http://allthingsds.wordpress.com)
- [Wes McKinney](#) - Pandas (Python Data Analysis library).
- [WileyEd](#) - Senior Manager - @Seagate Big Data Analytics | @McKinsey Alum | #BigData + #Analytics Evangelist | #Hadoop, #Cloud, #Digital, & #R Enthusiast
- [WNYC Data News Team](#) - The data news crew at @WNYC. Practicing data-driven journalism, making it visual and showing our work. @SkymindIO's open-source deep learning for the JVM. Integrates with Hadoop, Spark. Distributed GPU/CPUs | http://nd4j.org | http://www.skymind.io

Youtube Videos & Channels

- [What is machine learning?](#)
- [Andrew Ng: Deep Learning, Self-Taught Learning and Unsupervised Feature Learning](#)

- [Deep Learning: Intelligence from Big Data](#)
- [Interview with Google's AI and Deep Learning 'Godfather' Geoffrey Hinton](#)
- [Introduction to Deep Learning with Python](#)
- [What is machine learning, and how does it work?](#)
- [Data School](#) - Data Science Education
- [Neural Nets for Newbies by Melanie Warrick (May 2015)](#)
- [Google DeepMind co-founder Shane Legg - Machine Super Intelligence](#)

Toolboxes - Environment

- [Datalab from Google](#) easily explore, visualize, analyze, and transform data using familiar languages, such as Python and SQL, interactively.
- [Hortonworks Sandbox](#) is a personal, portable Hadoop environment that comes with a dozen interactive Hadoop tutorials.
- [R](#) is a free software environment for statistical computing and graphics.
- [RStudio](#) IDE – powerful user interface for R. It's free and open source, works onWindows, Mac, and Linux.
- [Python - Pandas - Anaconda](#) Completely free enterprise-ready Python distribution for large-scale data processing, predictive analytics, and scientific computing
- [Scikit-Learn](#) Machine Learning in Python
- [Data Science Toolbox](#) - Coursera Course
- [Data Science Toolbox](#) - Blog
- [Wolfram Data Science Platform](#) Take numerical, textual, image, GIS or other data and give it the Wolfram treatment, carrying out a full spectrum of data science analysis and visualization and automatically generating rich interactive reports—all powered by the revolutionary knowledge-based Wolfram Language.
- [Sense Data Science Development Paltform](#) A New Cloud Platform for Data Science and Big Data Analytics Collaborate on, scale, and deploy data analysis and advanced analytics projects radically faster. Use the most powerful tools — R, Python, JavaScript, Redshift, Hive, Impala, Hadoop, and more — supercharged and integrated in the cloud.
- [Mortardata](#) Solutions, code, and devops for high-scale data science.
- [Variance](#) Build powerful data visualizations for the web without writing JavaScript
- [Kite Development Kit](#) The Kite Software Development Kit (Apache License, Version 2.0), or Kite for short, is a set of libraries, tools, examples, and documentation focused on making it easier to build systems on top of the Hadoop ecosystem.
- [Domino Data Labs](#) Run, scale, share, and deploy your models — without any infrastructure or setup.
- [Apache Flink](#) A platform for efficient, distributed, general-purpose data processing.
- [Apache Hama](#) Apache Hama is an Apache Top-Level open source project, allowing you to do advanced analytics beyond MapReduce.
- [Weka](#) Weka is a collection of machine learning algorithms for data mining tasks.
- [Octave](#) GNU Octave is a high-level interpreted language, primarily intended for numerical computations.(Free Matlab)
- [Apache Spark](#) Lightning-fast cluster computing
- [Caffe](#) Deep Learning Framework
- [Torch](#) A SCIENTIFIC COMPUTING FRAMEWORK FOR LUAJIT
- [Nervana's python based Deep Learning Framework](#)
- [Aerosolve](#) - A machine learning package built for humans.
- [Intel framework](#) - Intel® Deep Learning Framework
- [Datawrapper](#) – An open source data visualization platform helping everyone to create simple, correct and embeddable charts. Also at [github.com](#)
- [Tensor Flow](#) - TensorFlow is an Open Source Software Library for Machine Intelligence
- [Natural Language Toolkit](#)
- [nlp-toolkit for node.js](#)
- [Julia](#) – high-level, high-performance dynamic programming language for technical computing.
- [IJulia](#) – a Julia-language backend combined with the Jupyter interactive environment.

Visualization Tools - Environments

- addepar
- amcharts
- anychart
- capsidea
- cartodb
- Cube
- d3plus
- Data-Driven Documents(D3js)
- datahero
- dygraphs
- ECharts
- exhibit
- Gatherplot
- gephi
- ggplot2
- Glue
- Google Chart Gallery
- highcarts
- import.io
- jqplot
- Matplotlib
- nvd3
- Opendata-tools - list of open source data visualization tools
- Openrefine
- plot.ly
- raw
- rcharts
- techanjs
- tenxer
- Timeline
- variancecharts
- vida
- Wolframalpha
- Wrangler
- r2d3
- NetworkX - High-productivity software for complex networks

Journals, Publications and Magazines

- ICML - International Conference on Machine Learning
- epjdatascience
- Journal of Data Science - an international journal devoted to applications of statistical methods at large
- Big Data Research
- Journal of Big Data
- Big Data & Society
- Data Science Journal

Presentations

- How to Become a Data Scientist
- Introduction to Data Science
- Intro to Data Science for Enterprise Big Data
- How to Interview a Data Scientist
- How to Share Data with a Statistician
- The Science of a Great Career in Data Science
- What Does a Data Scientist Do?
- Building Data Start-Ups: Fast, Big, and Focused
- How to win data science competitions with Deep Learning

Other Awesome Lists

- Other amazingly awesome lists can be found in the awesome-awesomeness list.
- Awesome Machine Learning A curated list of awesome Machine Learning frameworks, libraries and software.
- lists
- awesome-dataviz
- awesome-python
- Data Science IPython Notebooks.
- awesome-r
- awesome-datasets. – An awesome list of high-quality open datasets in public domains