Introduction to R



Available Statistical Packages



Available Statistical Packages

Proprietary

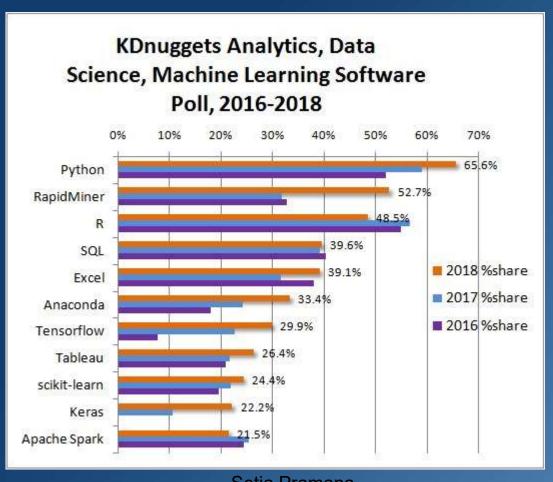
- Excel
- SPSS
- MINITAB
- SAS
- Stata
- Statistica
- Many more

Free Software

- LibreOffice Calc
- R
- CS Pro
- WinBugs
- Epilnfo
- Many more......

Setia Pramana

Statistical Software Used



What is R?

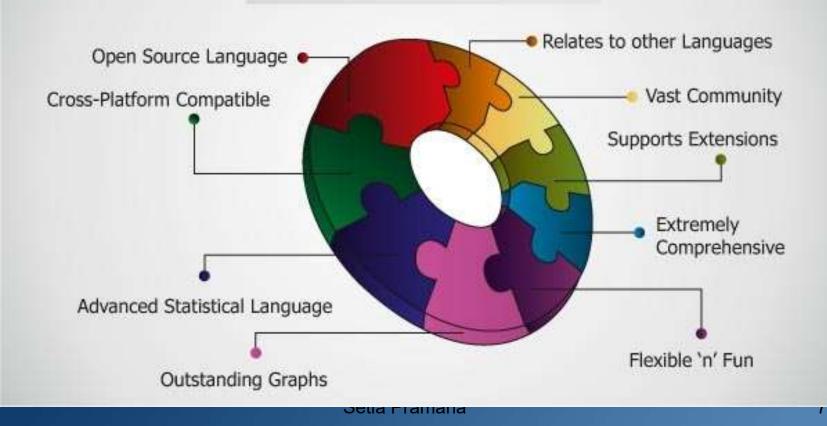
- A language and environment for statistical computing and graphics.
- An integrated suite of software facilities for data manipulation, calculation and graphical display.
- First appeared in 1996 by Prof. Ross Ihaka and Robert Gentleman of the University of Auckland, NZ.
- GNU software -> Free. Similar like S language.
- Open source, maintained and developed by a community of developers.
- Works in Windows, Unix, MacOs

R Includes

- Effective data handling and storage facility,
- A suite of operators for calculations on arrays, in particular matrices
- A large, coherent, integrated collection of intermediate tools for data analysis,
- Graphical facilities for data analysis and display either onscreen or on hardcopy
- Well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

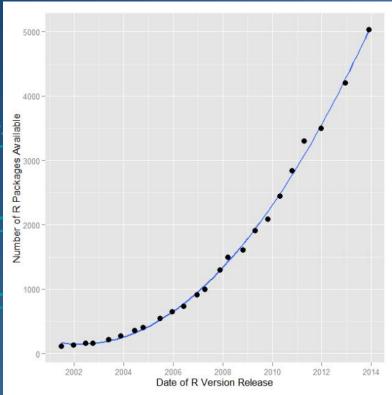
http://www.r-project.org/

R Programming Language



Why R?

- It is not only statistical software but also a language
- 10.000 add-on packages → lots of pre-prepared packages (http://cran.r project.org/web/packages/)
- With many applications http://cran.r-project.org/web/views/,
 http://www.revolutionanalytics.com/r-language-features-applications-and-extensions#thirdparty.
- Access to powerful, cutting-edge analytics



Why R?

 Flexible (complex or standard statistical practices, bayesian modelling, GIS map building, building interactive web applications, building interactive tests, etc.)

We can make our own package and publish it

Great Graphics and data visualization

Can be used

 Well Support resources-we

And many me



na pablion it					
Th	Dan Lin Jin Shkedy Daniel Yekutieli Oharumika Amaratunga Luc Bijnem - Editors	al ne 2010 ion			
Contents Editorial	Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R	3			
Contributed IsoGene: An MCMC for Ge Mapping and tanvinom: A neuralnet: Tra glmperm: A I Models Online Reproc Fingerprint Two-sided Ex Book Revie	Order-Restricted Analysis of Microarray Data © Springer	bs 5 13 18 25 30 39 39 44 53 . 9			
A Beginner's		59			

Why R?

- Can be integrated with other languages (C/C++, Java).
- R can interact with many data sources and other statistical packages (SAS, Stata, SPSS, and Minitab).
- For the high performance computing task \rightarrow multiple cores, either on a single machine or across a network.



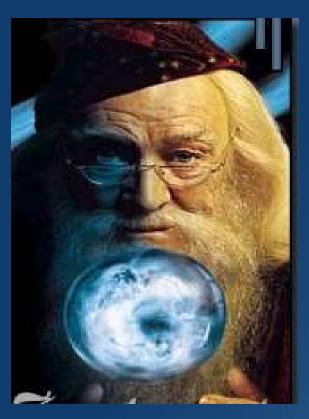


Setia Pramana 11

But.....

- R has no warranty
- Command Line Interface: difficult for some users.
- Users must learn a new way of thinking about data and data analysis sequence
- That's all I guess

Learning R





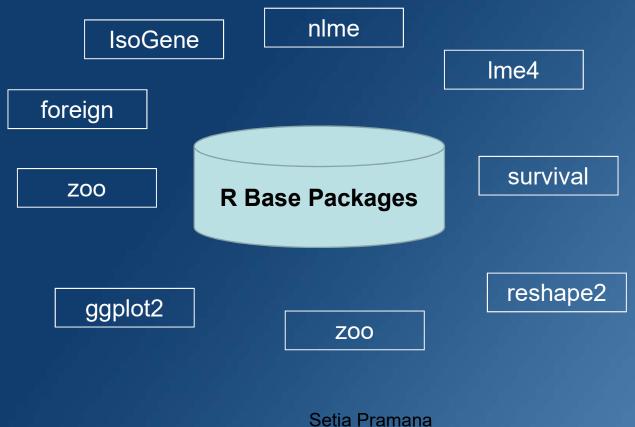
Setia Pramana



Companies that use R for Analytics



R Library/packages



15

My R Packages

- IsoGene
- IsoGeneGUI
- nea
- neaGUI
- biclustGUI
- OCRME
- More detail: http://setiopramono.wordpress.com/rprogramming/

Cutting Edge Technologies

Data Mining

- Decision trees: rpart, party
- Random forest: randomForest, party
- SVM: e1071, kernlab
- Neural networks: nnet, neuralnet, RSNNS
- Performance evaluation: ROCR
- Data Mining GUI <u>rattle</u>
- etc...
- http://www.rdatamining.com/

Social Media

- Text mining: tm
- Topic modelling: topic models, lda
- Word cloud: wordcloud
- Facebook: RFacebook
- Twitter data access: twitteR
- Social Network: sna, igraph, RSiena

(http://www.jstatsoft.org/v24/i06/paper)

http://www.r-bloggers.com/an-example-of-social-network-analysis-with-r-using-package-igraph/

Parallel Computing

- snow (Simple Network of Workstations) & snowfall for development of parallel R programs.
- multicore parallel processing of R code on machines with multiple cores or CPUs
- More: http://cran.r-

 project.org/web/views/HighPerformanceComputing.html

Setia Pramana

Big Data

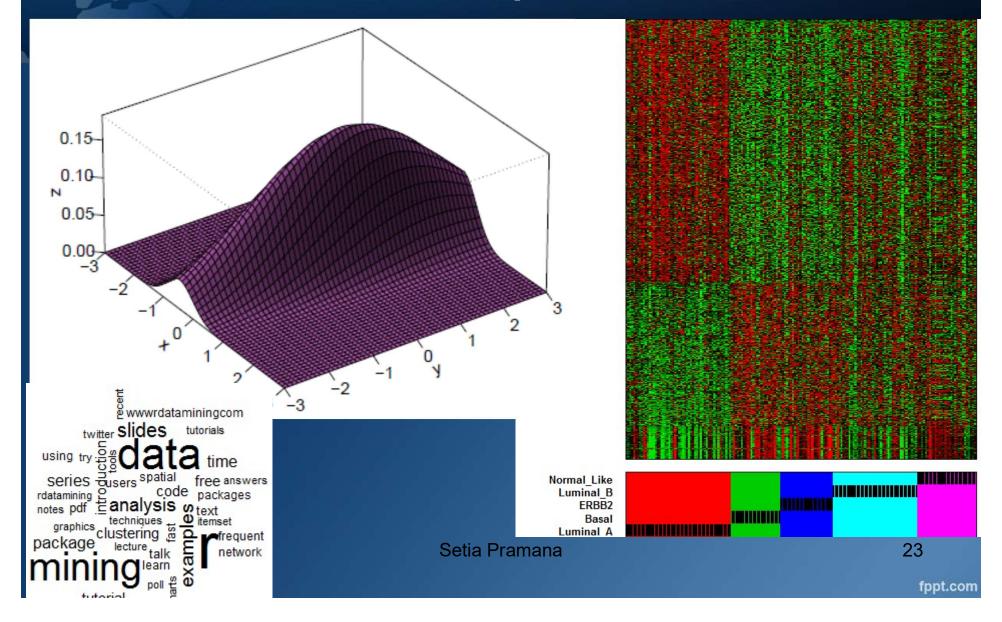
- RHadoop a collection of five R packages that allow users to manage and analyze data with Hadoop, developed by Revolution Analytics
- RHIPE an R and Hadoop Integrated Programming Environment
- More.....

R Graphics and Visualization

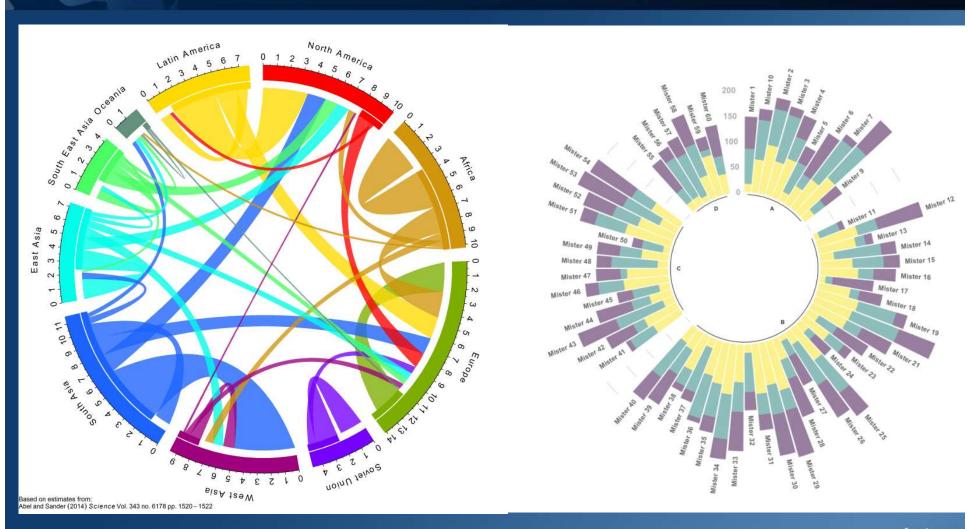
- R provides wide range graphics and visualizations
- Basic Plots: bar plots, basic 3D plots, heatmap.,etc
- Geographic Maps
- Projection Maps
- Social Network Graphs
- Animated graphics and movies (animation)
- Motion Charts (GoogleViz)
- Interactive Graphics (rggobi)
- Image format: BMP, JPEG, PDF, PNG etc...
- More: https://www.r-graph-gallery.com/

Setia Pramana 22

R Graphics



R Graphics



R Graphical User Interfaces

- R uses Command line interface and it is preferred for advanced users → allows direct control, more accurate, flexible and the analysis is reproducible.
- Requires good knowledge of the language → difficult for beginners or less frequent users.
- R provides tools for building GUIs → RGUI

R GUI Projects

- Integrated development environment (IDE)/Script Editors aimed to provide feature-rich environments to edit R scripts and code: Rstudio (www.rstudio.com), and architect (www.Openanalytics.eu)
- Web based application: the Rweb (Banfield, 1999),
 R.Net (www.u.arizona.edu/ ryckman/Net.php),
 or gWidgetsWWW (Verzani, 2012).

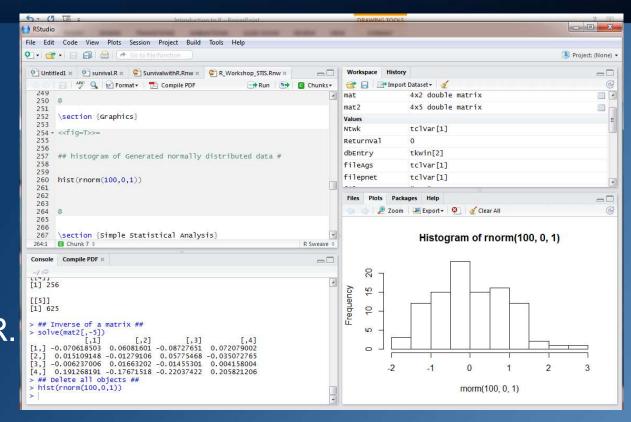
R GUI Projects

- Python: OpenMeta-Analyst (Wallace et al, 2012)
- Java: JGR (Java GUI for R), Deducer (Fellows, 2012), and Glotaran (Snellenburg, 2012).
- Php: R-php (http://dssm.unipa.it/R-php/)
- Other extensions connect R to graphical toolboxes for developing menus and dialog boxes: Tcltk, Gtk.

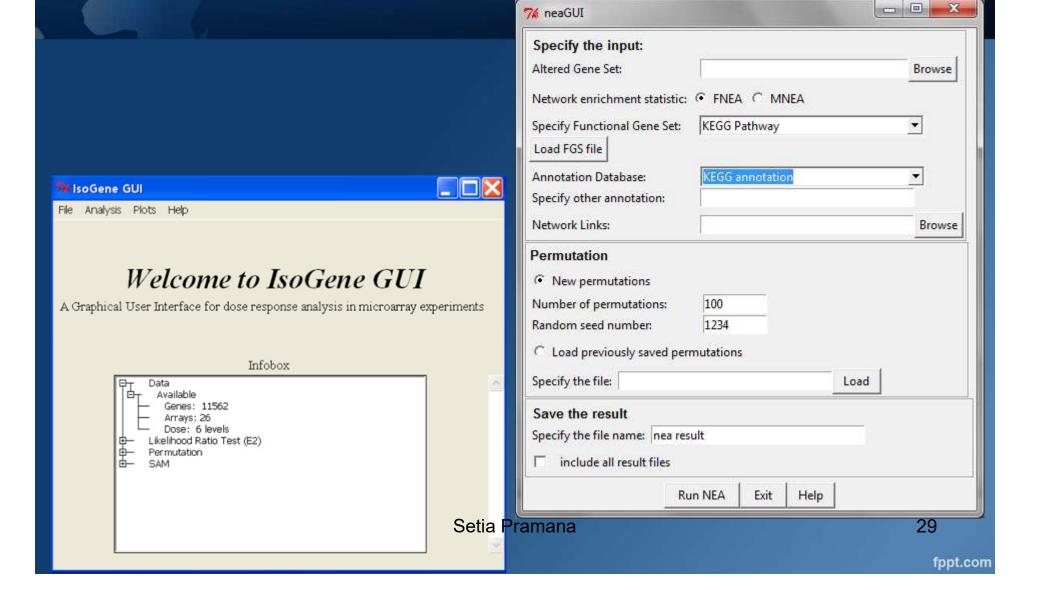
R Studio

- Download from Rstudio.com
- Powerfull IDE

 (Integrated
 Development
 Environment) for R.

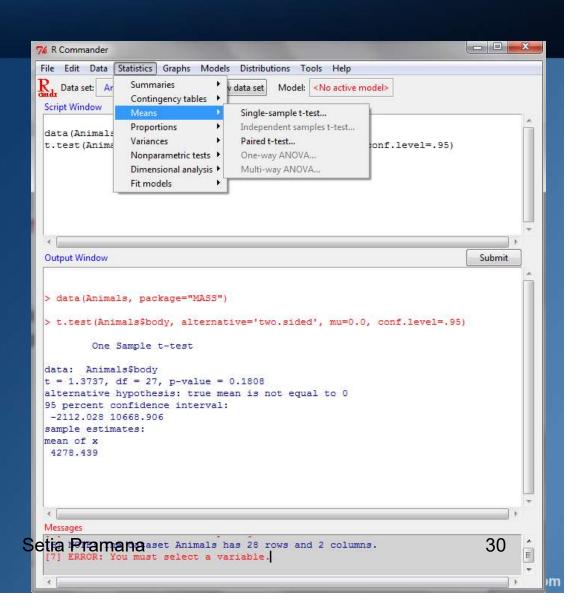


RGUI Developed using tcltk



RGUI: RCommander

- Rcommander.com
- Helpful for R beginner
- Install inside R



RGUI: Web Based App

Setia Pramana

31

WebBUGS

- Conducting Bayesian Statistical Analysis
 Online
- CombinesOpenBUGS and R

www.webbugs.psychstat.org

```
Welcome JSS » New I Current I Re-run I Copy I Email I Share I All Analyses I Modules I FAQ
  Model
 Name of analysis Meta-analysis of correlation
                                                                                                                                                                                                                                            Email notification
  1. Model: Type or select a model Meta-analysis
                                                                                                                                                                                                # Help
           for (i in 1:m){
                 z[i] < -.5*log((l+r[i])/(l-r[i]))
                  pre.phi[i] \leftarrow (n[i]-3)*a[i]
                  z[i] ~ dnorm(zeta[i], pre.phi[i])
                 zeta[i] ~ dnorm(beta, pre.tau)
          beta ~ dnorm(0, 1.0E-6)
           rho <- (exp(2*beta)-1)/((exp(2*beta)+1))
           pre.tau ~ dgamma(.001,.001)
           tau <- 1/pre.tau
 2. Data: Input or select data No + Convert data Help
   \texttt{list}(\texttt{m} = \texttt{11}, \texttt{r} = \texttt{c}(0.210, 0.252, 0.123, 0.330, 0.400, 0.340, 0.110, 0.147, 0.110, 0.360, 0.138), \texttt{n} = \texttt{n}(0.210, 0.252, 0.123, 0.330, 0.400, 0.340, 0.110, 0.147, 0.110, 0.360, 0.138), \texttt{n} = \texttt{n}(0.210, 0.252, 0.123, 0.330, 0.400, 0.340, 0.110, 0.147, 0.110, 0.360, 0.138), \texttt{n} = \texttt{n}(0.210, 0.252, 0.123, 0.330, 0.400, 0.340, 0.110, 0.147, 0.110, 0.360, 0.138), \texttt{n} = \texttt{n}(0.210, 0.252, 0.123, 0.330, 0.400, 0.340, 0.110, 0.147, 0.110, 0.360, 0.138), \texttt{n} = \texttt{n}(0.210, 0.252, 0.123, 0.330, 0.400, 0.340, 0.110, 0.147, 0.110, 0.360, 0.138), \texttt{n} = \texttt{n}(0.210, 0.252, 0.123, 0.123, 0.330, 0.400, 0.340, 0.110, 0.147, 0.110, 0.360, 0.138), \texttt{n} = \texttt{n}(0.210, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0.123, 0
   c(215,132,309,117,307,1212,175,380,86,74,361), a =
   c(1.0,1.0,1.0,1.0,0.8,1.0,1.0,1.0,1.0,1.0,1.0))
 3. Initial values: Multiple sets of initial values are allowed. Convert data Help
   list(beta = 0, pre.tau = 1 )
   list(beta = 1, pre.tau = 1 )
   list(beta = -1, pre.tau = 1)
     Next
```

Setia Pramana 32

WebBUGS Admin » Login I Logout I Profile I Forgot password

RGUI: Shiny

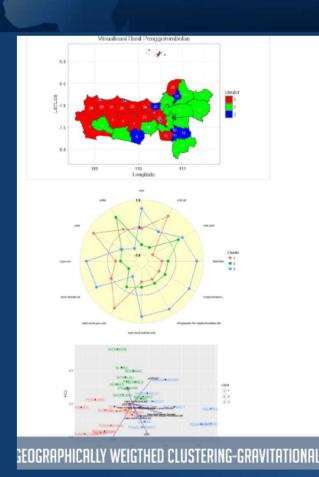
- A new package from Rstudio to build interactive web applications with R.
- Really Easy!
- Build useful web applications with only a few lines of code—no JavaScript required.
- Self learning: http://shiny.rstudio.com/
- http://www.showmeshiny.com/

Our Recent R Packages

Name	Title	Brief Description	Author	Repository
spatialClust	Spatial Clustering	Clustering analysis with pay attentation on	Imam Habib Pamungkas, Setia	CRAN
11,339 No. 12	165.7	membership via spatial effects	Pramana	
advclust	S4 Object Oriented for	Advance on clustering with fuzzy clustering for	Achmad Fauzi Bagus F, Setia	CRAN
	Advanced Clustering(overlapping cluster and objects on gray area.	Pramana	
	Fuzzy Clustering and	Cluster Ensemble performs combining several		
	Cluster Ensemble)	result as one robust and stable result.		
RcmdrPlugin.Fuzzy	R commander plugin	Graphical User interface via Rcmdr Plugin for	Achmad Fauzi Bagus F, Setia	CRAN
Clust	for fuzzy clustering	fuzzy clustering analysis	pramana	
MetaheuristicFPA	Metaheuristic with	Optimization of function objectives to get	Amanda Pratama Putra, Margaretha	CRAN
	Flower Pollinantion	global optimum of parameter by using Flower	Ari Anggorowati	1.5000000000000000000000000000000000000
	Algorithm	Pollination Algorithm		
Multiplier	Social Accounting	Graphical User Interface for performing SAM	Tiara Ratna Dewi, Aisyah Fitri	R-Forge
	Matrix and Finansial	(Social Accounting Matrix) and FSAM (Financial	Yuniarshi	88.0
	Social Accounting	Social Accounting Matrix)	AND ELECTRICAL SECTION (IN)	
	Matrix			
RcmdrPlugin.PCAR	Robust PCA plugin for	Graphical User Interface for Robust Principal	Monalisa Sipahutar, Setia Pramana	CRAN
obust	Rcmdr	Component Analysis (PCA) with Hubert	N	
2000	2015-00000000000000000000000000000000000	Algorithm for Dimension Reduction		

34

Our Recent R Packages



spatialClust

An R Package for Cluster Spatial Data

Available on CRAN:

September 3rd 2016

"spatialClust" - an R package

by Imam Habib Pamungkas, S.S.T and Setia Pramana, Ph.D

Setia Pramana 35

Our Recent R Packages

- Kalingga
- Muria
- C++

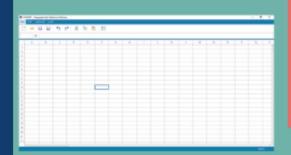




ASGARD is a statistics software used to perform geographically weighted regression (GWR). This software was made in 2016 and currently contains some basic GWR functions like GWR, Geographically Weighted Poisson Regression (GWPR), Geographically Weighted Logistic Regression (GWLR), Geographically Weighted Negative-Binomial Regression (GWNBR) and some Assumption Test related to GWR. In addition, ASGARD is also integrated with the map that make it easier for users to performs analysis.

MAIN FEATURES

Spreadsheet



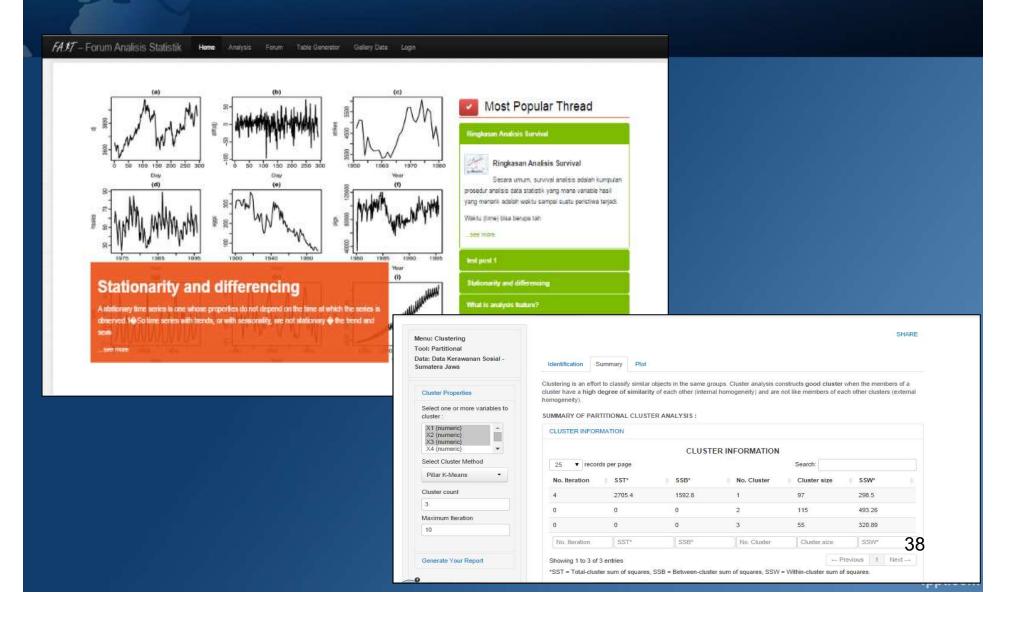
Fairly complete functions

- · GWR
- · GWPR
- · GWLR
- · GWNBR
- · Variance Inflation Factor
- · Breusch-Pagan Test



Map Visualization
Map Visualization can help
users to understand the
circumstances of the
observation area.

RGUI using Shiny: FAST



RGUI using C#: Wires

 For Spatial Data Analysis



RGUI using C#: Wires

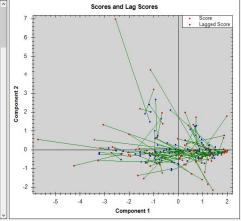
File Thematic Analysis Tools

Features



Data View Variable View Ma	p View Result View
Exploratory Spatial Data Analysis	Provide calculation of spatial autocorrelation based on Moran's I, Gearcy's C, Local Indicators of Spatial Association (LISA)
Spatial Weight Matrix	Spatial interactions among observations
Spatial Clustering	Clustering observation with spatial attributes
Spatial Regression	Regression analysis with spatial depedency
Regional Inequality	Inequality analysis especially on poverty subjects
Spatial Shift Share	Comparing growing rate of several sector based on spatial
Kriging	Imputation on missing data with spatial attributes

	Score Component 1	Lag Score Component 1	Score Component 2	Lag Score Component 2	Inequality
SELAPARANG	-2.5570	-1.1572	6.9862	1.4920	5.6697
SUMBAWA	-1.1325	0.8284	4.2860	-0.0731	4.7798
DOMPU	-3.0534	0.1945	1.3361	-0.1566	3.5745
WOJA	-3.4202	-0.1159	0.5342	0.0115	3.3454
PRAYA	-5.7190	-2.5356	0.5513	-0.0462	3.2390
AIKMEL	-3.2690	-0.2523	-0.5785	-0.6618	3.0178
MOYO UTARA	1.6801	-0.3151	-0.1405	2.1151	3.0114
RHEE	1.9859	0.8254	1.7949	-0.3738	2.4597
SUKAMULIA	0.9716	-0.9446	-1.5866	-0.1341	2.4045
SELONG	-2.0031	-0.1675	0.8245	-0.6772	2.3716
SEMBALUN	1.3885	-0.8893	-0.1614	-0.7276	2.3471
TAMBORA	1.4137	0.4499	-2.1472	-0.0816	2.2794
KURIPAN	0.6826	-1.5507	-0.0123	-0.2446	2.2453
SOROMANDI	1.2874	-0.8441	-0.3371	0.2568	2.2127
PUJUT	-4.2471	-2.3220	-0.8469	-0.2324	2.0208
RABA	0.2869	0.8763	1.9864	0.0585	2.0160
MATARAM	-0.8592	-0.9215	0.6684	2.6770	2.0096
ALAS	0.4113	1.3838	1.3021	-0.4249	1.9820
GERUNG	-2.0509	-0.0880	-0.1390	-0.3074	1.9701
GUNUNG SARI	-1.7310	-0.6951	-0.0028	1.6604	1.9594
KAYANGAN	0.9773	-0.9436	-0.2941	-0.3153	1.9210
TARANO	1.2161	-0.5865	-0.2837	-0.0498	1.8177
TALIWANG	-0.0186	1.6467	0.5946	-0.0904	1.8007



b	after					
e		0		1		
f	▶ 0	90		19		
r	1	3		15		
е	CCR: 82.6	5772 9	-			
K	ABKOTA_N	10	Before	After		
01			0	0		
03	03		0	0		
06	06		0	0		
07	07		0	0		
08	08		0	0		
10)		0	0		
12	12		0	0		
14	14		0	0		
74			0	0		
79)		0	0		
01			0	0		
02	2		0	0		
03	8		0	0		
04	1		0	0		
05	i		0	0		
06	5		0	0		
07	r:		0	0		
08	3		0	0		

40

Dynamic Report Generation

Produce documents automatically: pdf, doc, html Packages:

- Sweave
- knitr
- Markdown

Big Data and R

Data

Cleaning Filtering Aggregation

Collect Subset



Big Data

SparkR

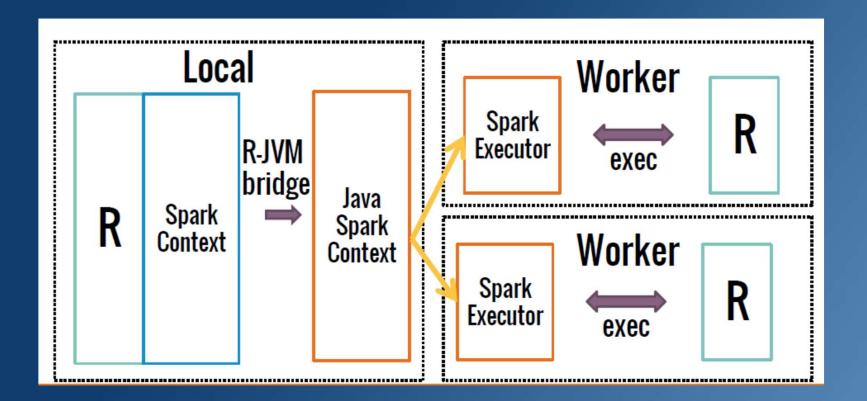


SparkR is a language binding that seamlessly integrates R with Spark, and enables native R programs to scale in a distributed setting

Setia Pramana

43

SparkR Architecture



A Start

Sehaga laiknya bidang innu yang berindipt kagalda mutuntahka, statistika tak dapat dipelaja sesara penyadutin dengat mendiana, ten mendiana kasa. Seletian merabahan salah koosep pedu praktisi dan pengalaman langsung dalam mendiangan hipotesia, menali sangking merakod dan mengalay data modelika melasupaan lasi statistik untuk menga-

Baktiw Hissan, Ph.D. Senior Ecocatisticom European Organization for Research and Treatment of Cancer, Belgium

Aflapun yang dibahsa dalam tuku ini, sebagai be - Pangenalah RGU itan RStudio - Statolik Deslerjah dan Visualsiani Dala - Visualinasi dengan ggolot2 - Statolik jeferenasi dan Uji Hipotesis - Anahais Keragaman (Anova)

- Analies Regres den Korelani Persograman dengen Riden abdyr Progenian Rickermander Analisa Data dengan Richmander Regres Logaria





DASAR-DASAR STATISTIKA Dengan Software R Konsep dan Aptikasi

Setia Pramana, Ph.D Ricky Yordani, M.Stat Robert Kurniawan, M.Si Budi Yuniarto, M.Si

DASAR-DASAR

STATISTIKA Dengan Software R
Konsep dan Aplikasi

EDISI KEDUA

IN MEDIA









Konsep Serta Implementasi

Kebutuhan akan eksplorasi dan analisis data semakin meningkat beberapa tahun terakhir. Metode eskplorasi dan analisis data juga mulai bergeser ke arah penggunaan data mining dan wetode esspionasi dari aminasi dara juga malari bergaeri ker arin proggamari dara iniming dari beberapa algoritma machine learning. Hal ini mendorong perubahan kurikulum dan materi yang harus disampaikan dan dikuasai mahasiswa khusuanya mahasiswa jurusan statistik. Buku ini sangat saya rekomendasikan baik kepada mahasiswa maupun para pengajar karena buku ini tidak hanya memberikan teori namun juga mengajarkan bagaimana mengaplikasikan teori tersebut dalam contoh-contoh praktis. Buku ini juga memberikan keberagaman aplikasi dari data mining dengan tipe data yang berbeda-beda yang dapat diaplikasikan dengan software R.

Dr. Erni Tri Astuti, M.Math - Direktur Politeknik Statistika STIS

R merupakan salah satu alat pengolahan data yang sangat ampuh. Dengan bahasa yang lugas dan "to-the-point", penulis berhasil menyajikan data mining dengan pendekatan praktis menggunakan R. Buku ini merupakan batu pijakan yang sangat berguna buat para aspiring data scientist yang ingin menggeluti bidang data science

Syafri Bahar S.Si., M.Sc., FRM - Vice President of Data Science GOJEK.

Bahasan buku ini mencakup:

- Pengantar Data Mining
 Eksplorasi dan Visualisasi Data
- Regresi Linear dan Logistik Analisis Komponen Utama
- Multivariate Anova
- Supervised Learning (KNN, Decision Tree, Random Forest, dll)
 Unsupervised Learning (Cluster Analysis)
- Text Mining
- 9. Analisis Sentimen
- 10. Data Mining dalam Bioinformatika



DATA

MINING dengan R

Harga P. Jawa Rp.

Setia Pramana Setia Pramana, **Budi Yuniarto** Siti Mariyah Ibnu Santoso Rani Nooraeni DATA dengan Konsep Serta Implementasi 9

Book Chapters



Dan Lin Ziv Shkedy Daniel Yekutieli Dhammika Amaratunga Luc Bijnens *Editors*

Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R

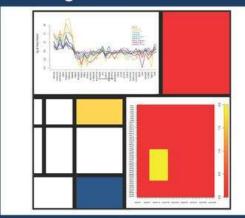
Order-Restricted Analysis of Microarray Data





Chapman & Hall/CRC Biostatistics Series

Applied Biclustering
Methods for Big and
High-Dimensional Data
Using R



edited by Adetayo Kasim Ziv Shkedy • Sebastian Kaiser Sepp Hochreiter • Willem Talloen



Conclusion

If statistics programs/languages were cars...











Thank you for your attention!

