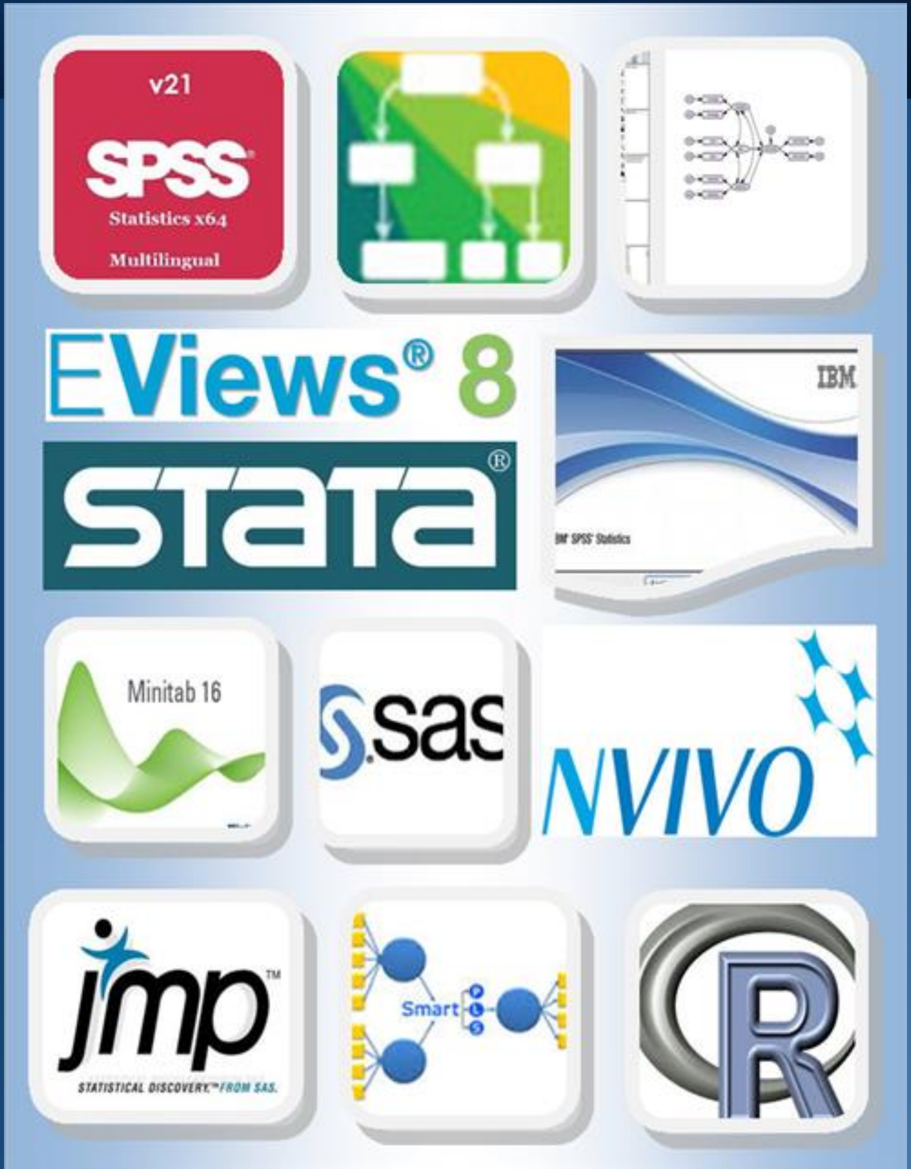
Introduction to R

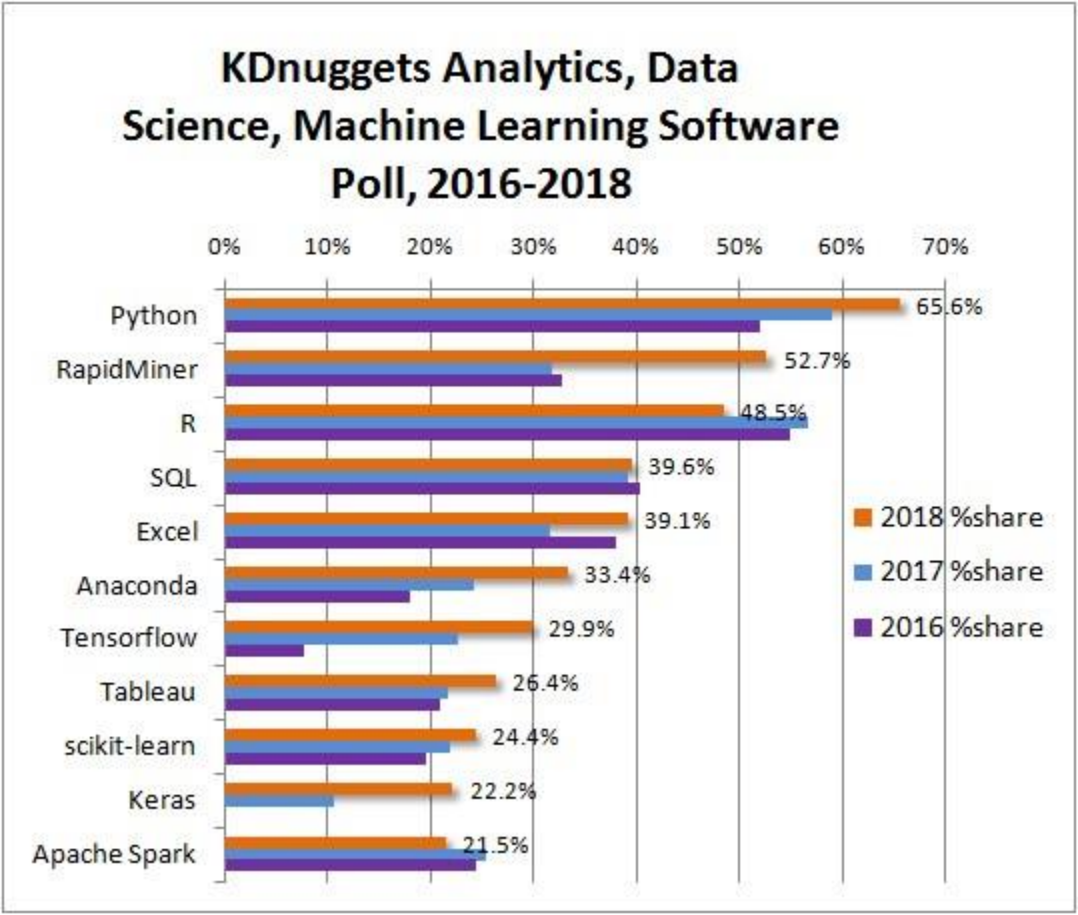
Setia Pramana

Politeknik Statistika STIS

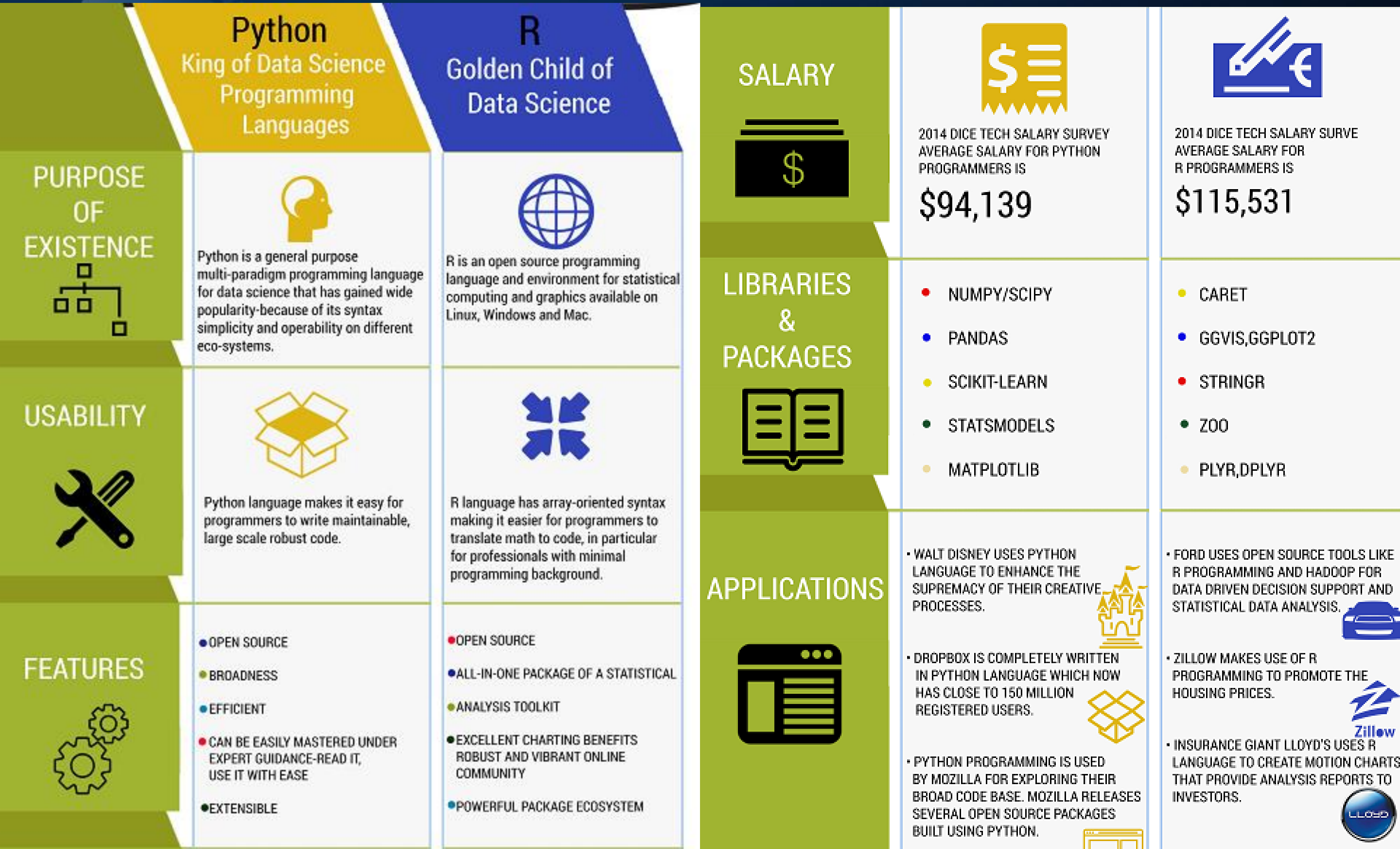
Available Statistical Packages



Statistical Software Used



Python or R?



	Python King of Data Science Programming Languages	R Golden Child of Data Science
PURPOSE OF EXISTENCE	Python is a general purpose multi-paradigm programming language for data science that has gained wide popularity-because of its syntax simplicity and operability on different eco-systems.	R is an open source programming language and environment for statistical computing and graphics available on Linux, Windows and Mac.
USABILITY	Python language makes it easy for programmers to write maintainable, large scale robust code.	R language has array-oriented syntax making it easier for programmers to translate math to code, in particular for professionals with minimal programming background.
FEATURES	● OPEN SOURCE ● BROADNESS ● EFFICIENT ● CAN BE EASILY MASTERED UNDER EXPERT GUIDANCE-READ IT, USE IT WITH EASE ● EXTENSIBLE	● OPEN SOURCE ● ALL-IN-ONE PACKAGE OF A STATISTICAL ● ANALYSIS TOOLKIT ● EXCELLENT CHARTING BENEFITS ROBUST AND VIBRANT ONLINE COMMUNITY ● POWERFUL PACKAGE ECOSYSTEM

SALARY	2014 DICE TECH SALARY SURVEY AVERAGE SALARY FOR PYTHON PROGRAMMERS IS $94,139	2014 DICE TECH SALARY SURVE AVERAGE SALARY FOR R PROGRAMMERS IS $115,531
LIBRARIES & PACKAGES	● NUMPY/SCIPY ● PANDAS ● SCIKIT-LEARN ● STATSMODELS ● MATPLOTLIB	● CARET ● GGVIS,GGPLOT2 ● STRINGR ● ZOO ● PLYR,DPLYR
APPLICATIONS	• WALT DISNEY USES PYTHON LANGUAGE TO ENHANCE THE SUPREMACY OF THEIR CREATIVE PROCESSES. • DROPBOX IS COMPLETELY WRITTEN IN PYTHON LANGUAGE WHICH NOW HAS CLOSE TO 150 MILLION REGISTERED USERS. • PYTHON PROGRAMMING IS USED BY MOZILLA FOR EXPLORING THEIR BROAD CODE BASE. MOZILLA RELEASES SEVERAL OPEN SOURCE PACKAGES BUILT USING PYTHON.	• FORD USES OPEN SOURCE TOOLS LIKE R PROGRAMMING AND HADOOP FOR DATA DRIVEN DECISION SUPPORT AND STATISTICAL DATA ANALYSIS. • ZILLOW MAKES USE OF R PROGRAMMING TO PROMOTE THE HOUSING PRICES. • INSURANCE GIANT LLOYD'S USES R LANGUAGE TO CREATE MOTION CHARTS THAT PROVIDE ANALYSIS REPORTS TO INVESTORS.

Python or R?

- https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis

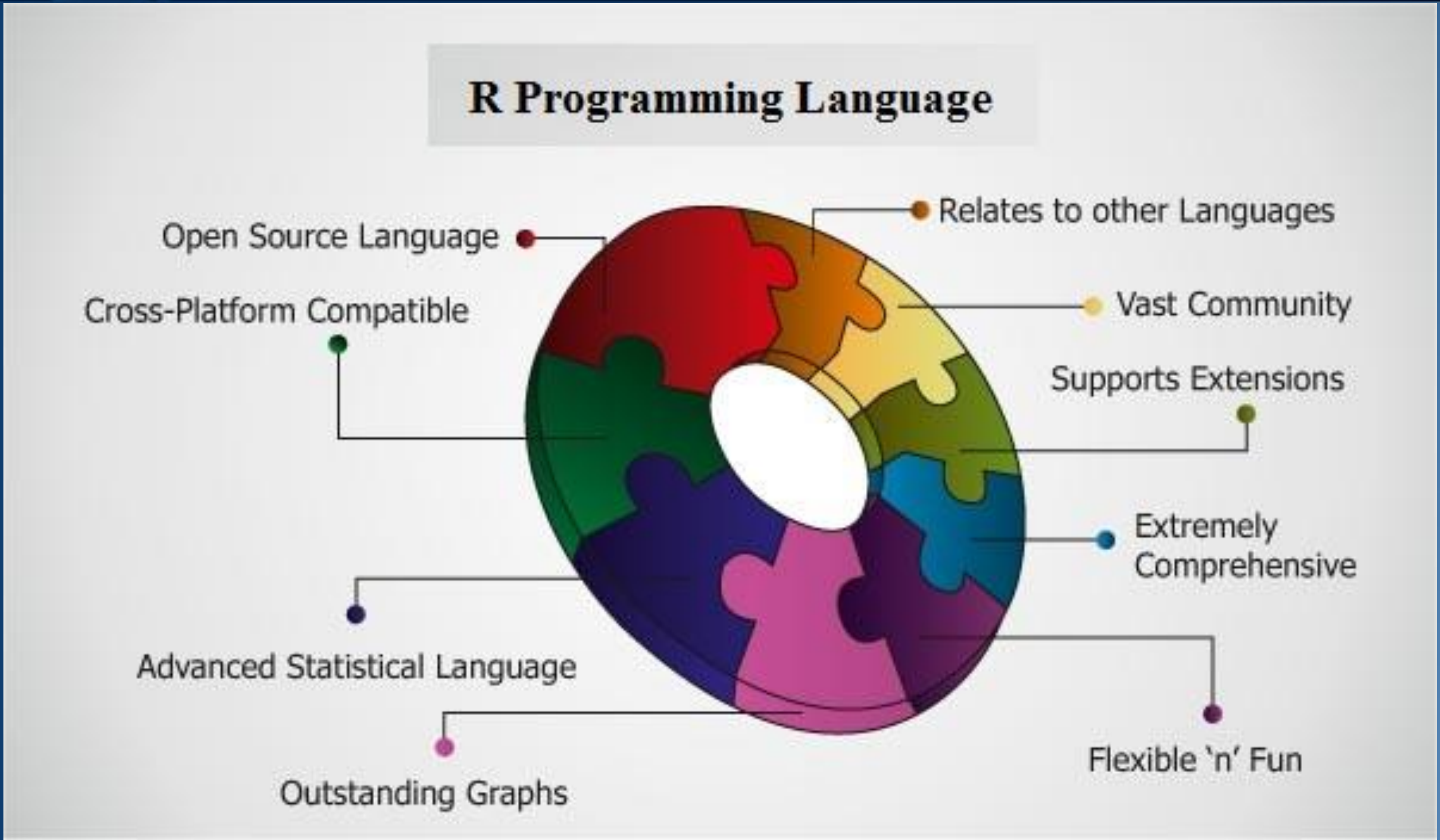
What is R?

- 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

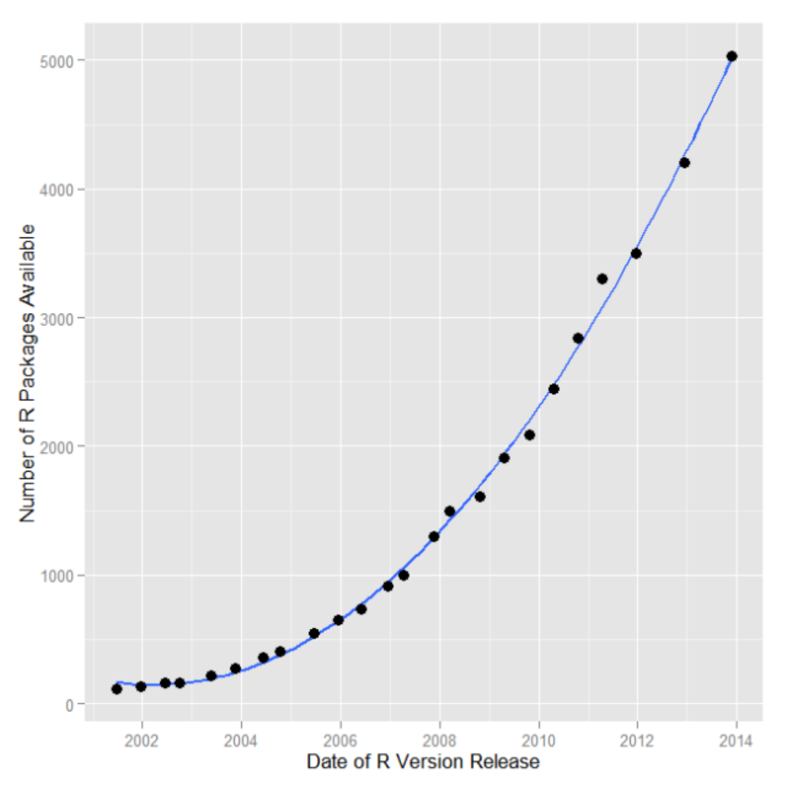
- Effective data handling and storage facility,
- A suite of operators for calculations on arrays, in particular matrices
- A large, coherent, integrated collection of intermediate tools for data analysis,
- Graphical facilities for data analysis and display either on-screen or on hardcopy
- Well-developed, 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
Open Source Language
Relates to other Languages
Cross-Platform Compatible
Vast Community
Supports Extensions
Extremely Comprehensive
Advanced Statistical Language
Flexible 'n' Fun
Outstanding Graphs

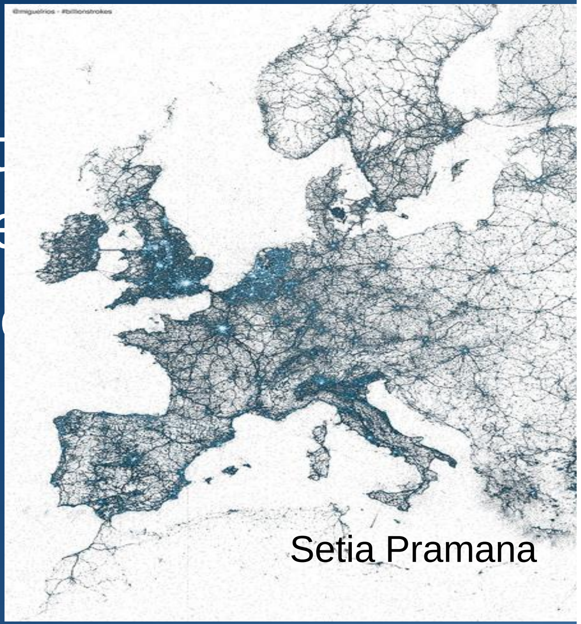
Why R?

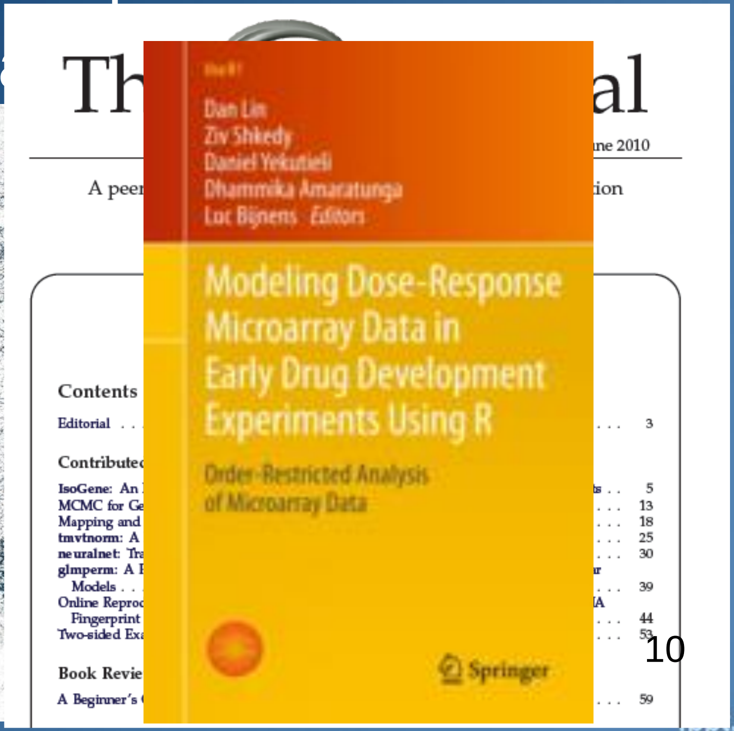
- It is not only statistical software but also a language
- 10.000 add-on packages → lots of pre-prepared packages (http://cran.r-project.org/web/packages/)
- With many applications http://cran.r-project.org/web/views/, http://www.revolutionanalytics.com/r-language-features-applications-and-extensions#thirdparty .
- 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 visualiz
- Can be used
- Well Support
  resources-we
- And many m





Th al

ne 2010

A peer ion

Contents

Editorial . . . 3

Contributed

IsoGene: An . . . 5
MCMC for Ge . . . 13
Mapping and . . . 18
tmvtnorm: A . . . 25
neuralnet: Tra . . . 30
glmperm: A
Models . . . 39
Online Repro
Fingerprint . . . 44
Two-sided Ex . . . 53

Book Revie

A Beginner's . . . 59

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

Springer

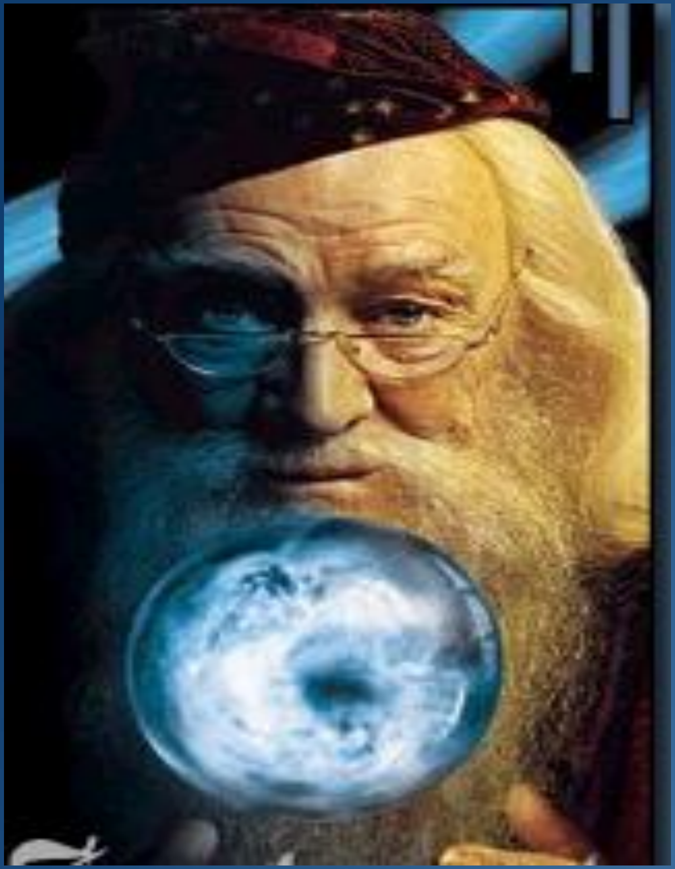
Why R?

- 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 → multiple cores, either on a single machine or across a network.

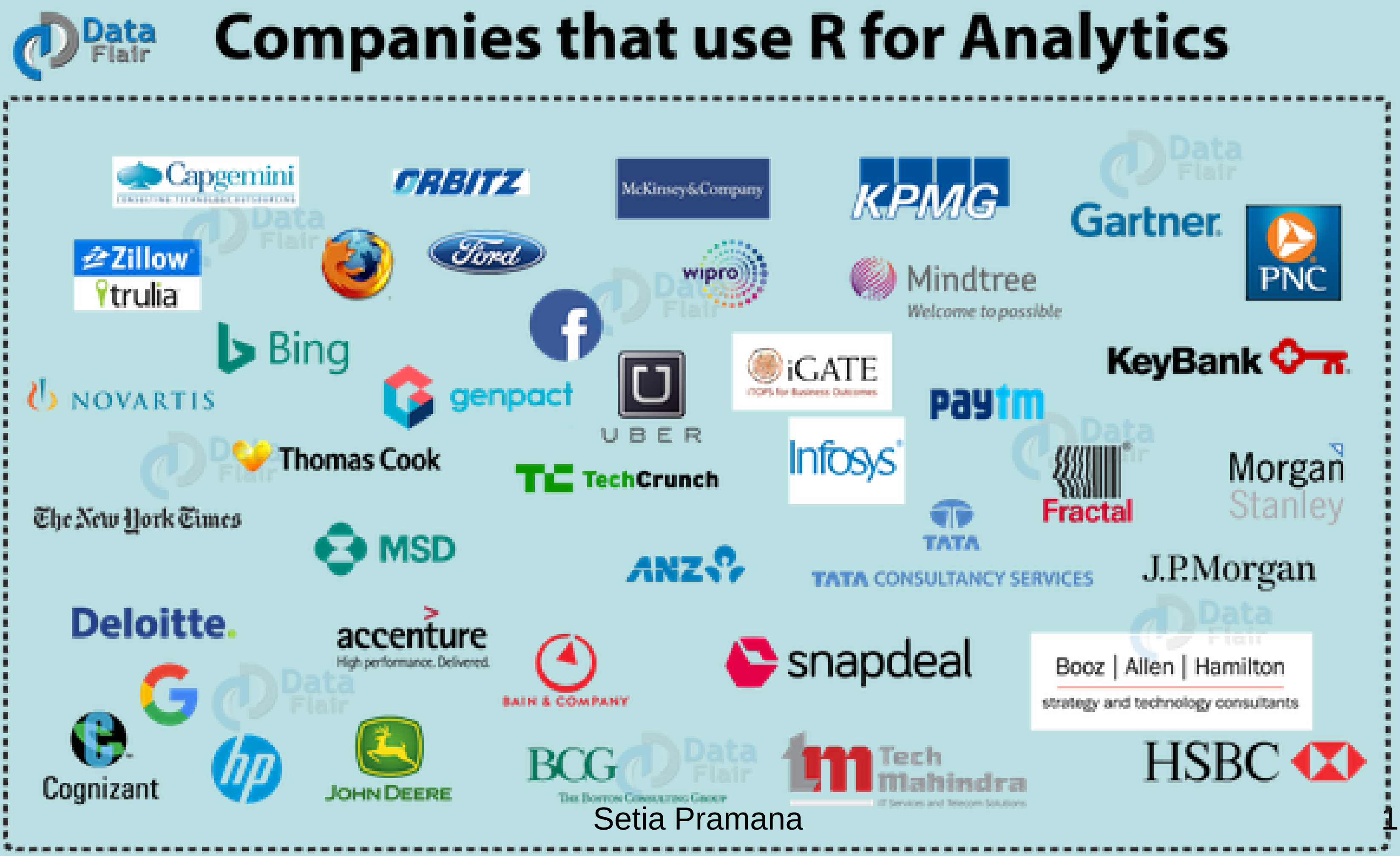
But…..

- 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

WizaRd

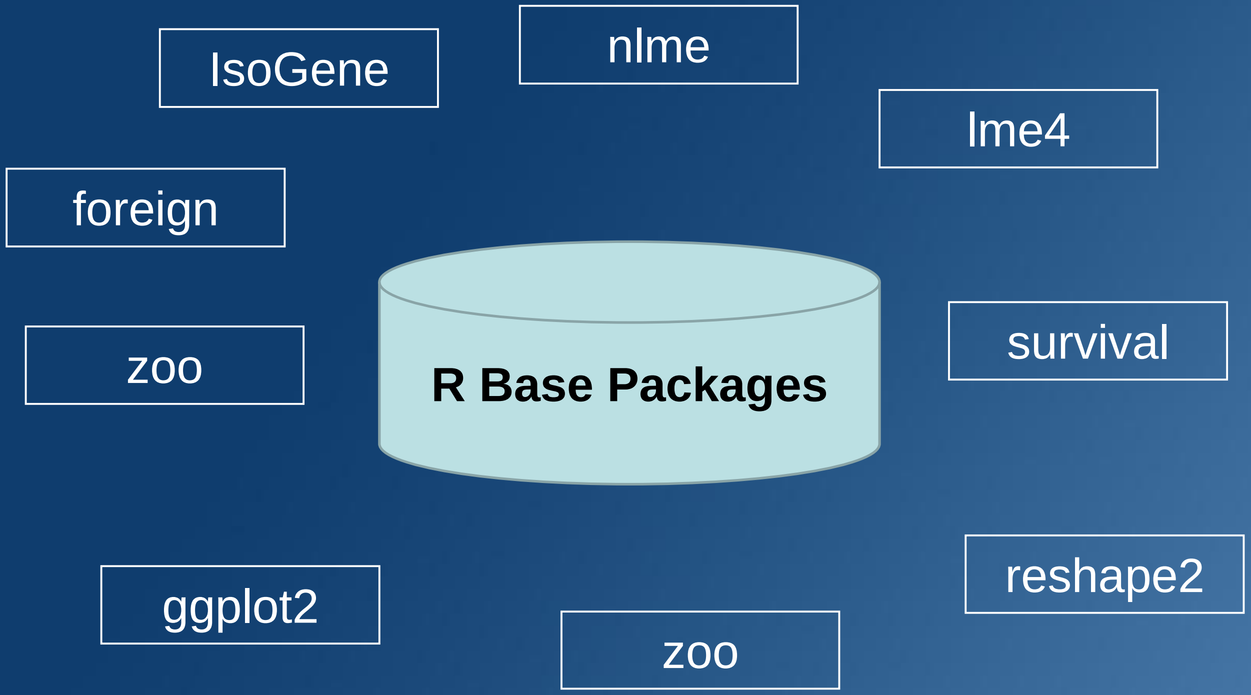


Learning R



Companies that use R for Analytics

R Library/packages



IsoGene

nlme

lme4

foreign

zoo

R Base Packages

survival

ggplot2

zoo

reshape2

My R Packages

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

Cutting Edge Technologies

Finance

- https://cran.r-project.org/web/views/Finance.html
- zoo - Provides the most popular format for saving time series objects in R.
- xts - Very flexible tools for manipulating time series data sets.
- quantmod - Tools for downloading financial data, plotting common charts, and doing technical analysis.
- The QuantTools package offers enhanced quantitative trading and modeling tools.
- The Risk package computes 26 financial risk measures for any continuous distribution.

Data Mining

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

Social Media

- Text mining: tm
- Topic modelling: topicmodels, 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

Big Data

- RHadoop - a collection of five R packages that allow users to manage and analyze data with Hadoop, developed by Revolution Analytics

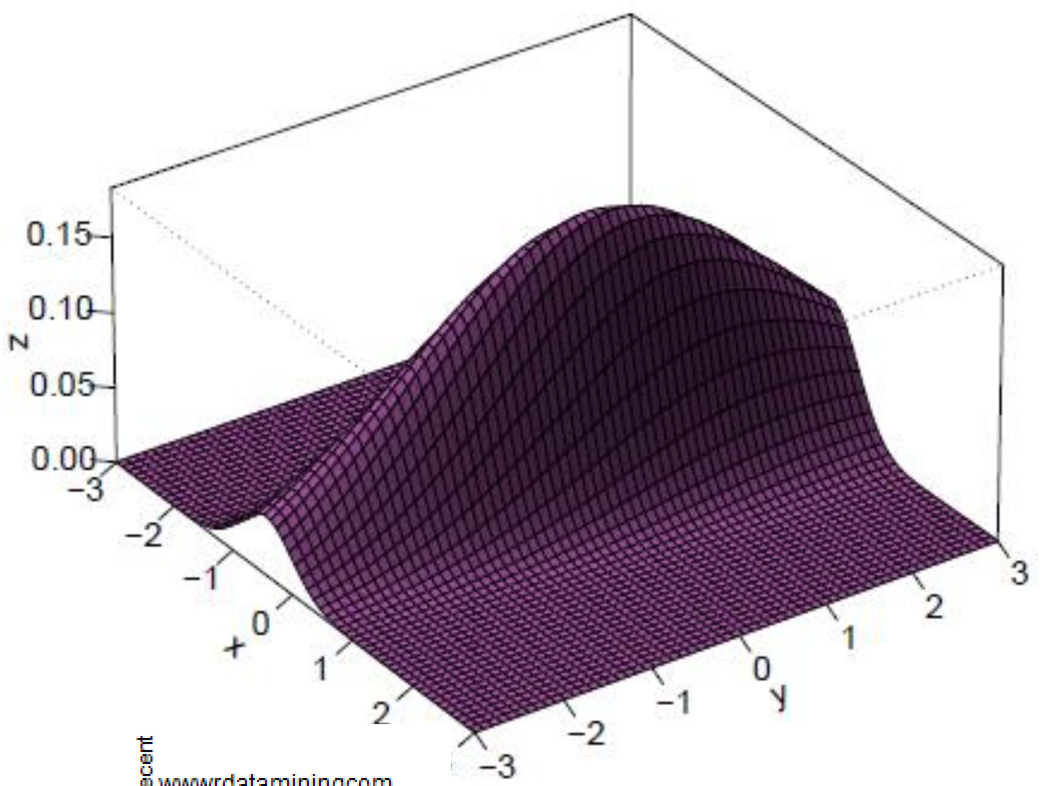
- RHIPE - an R and Hadoop Integrated Programming Environment

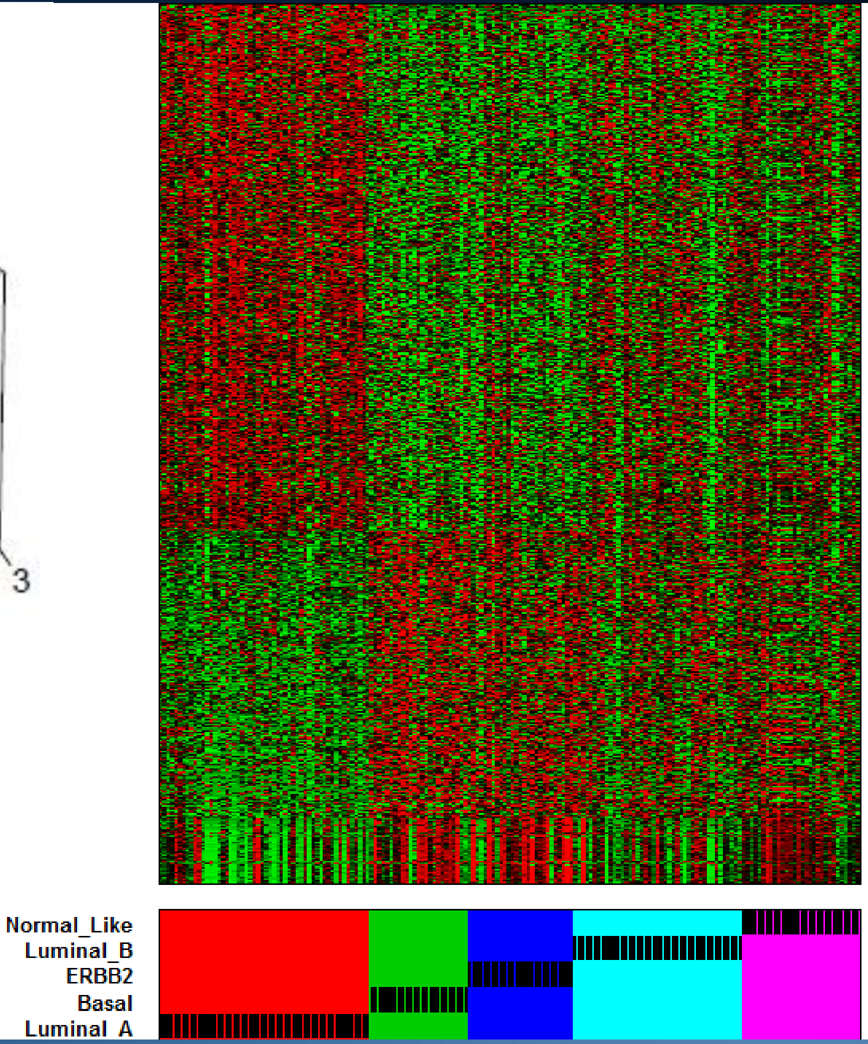
- More……

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/

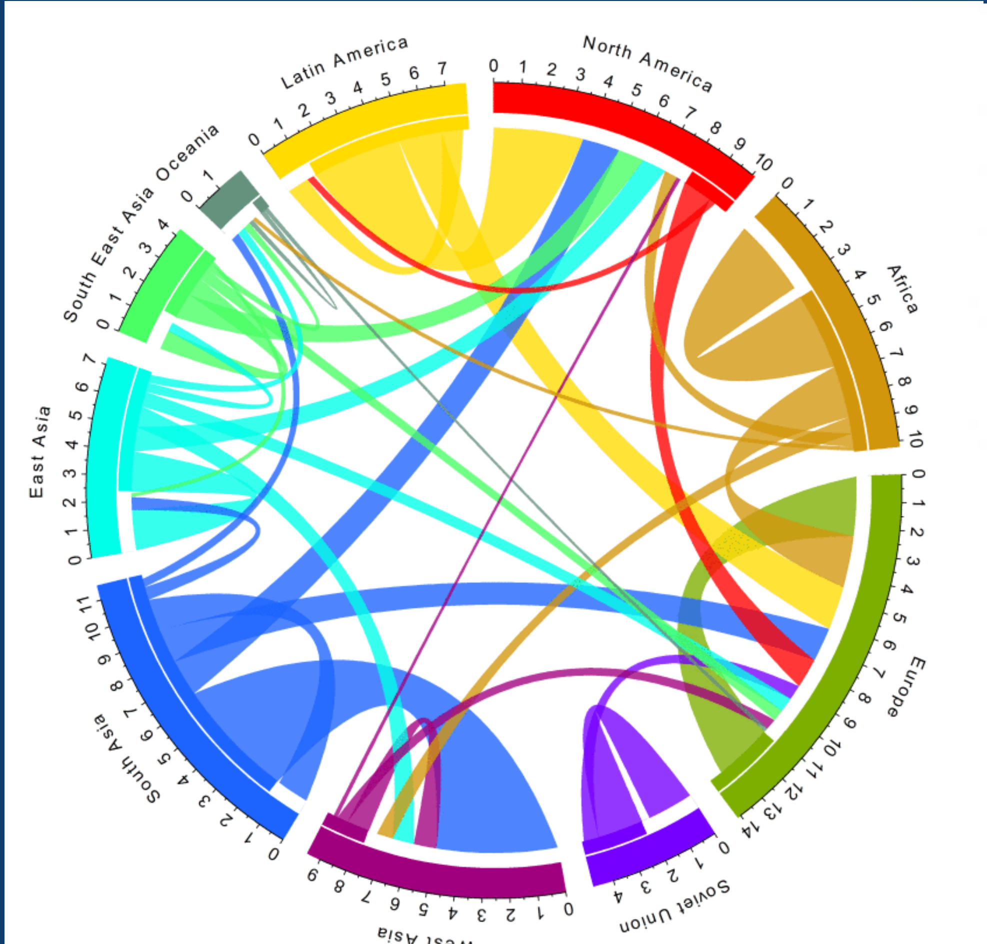
R Graphics


0.15
0.10
0.05
0.00
z
−3
−2
−1
0
1
2
x
−3
−2
−1
0
1
2
3
y

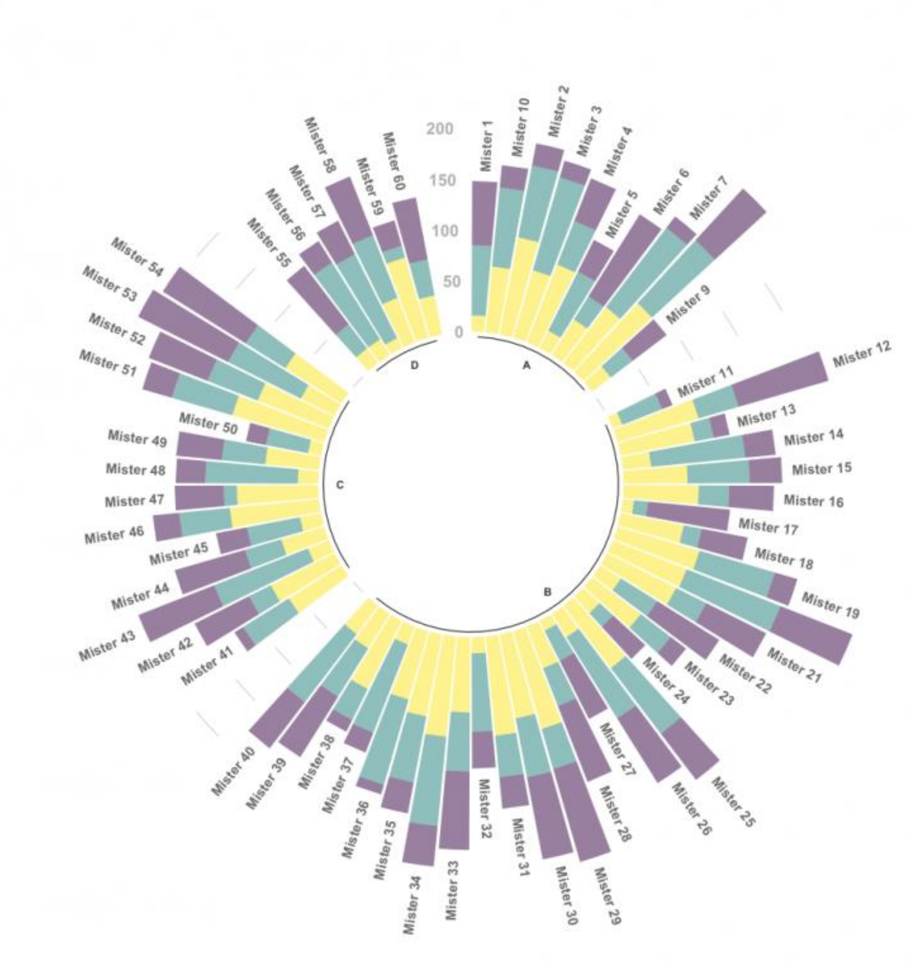

Normal_Like
Luminal_B
ERBB2
Basal
Luminal_A

recent
wwwrdataminingcom
twitter
slides
tutorials
using
try
data
time
introduction
tools
series
users
spatial
free
answers
rdatamining
code
packages
notes
pdf
analysis
text
examples
itemset
graphics
techniques
clustering
fast
frequent
package
lecture
r
talk
network
mining
learn
poll
charts
tutorial

R Graphics



Based on estimates from:
Abel and Sander (2014) *Science* Vol. 343 no. 6178 pp. 1520 – 1522



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

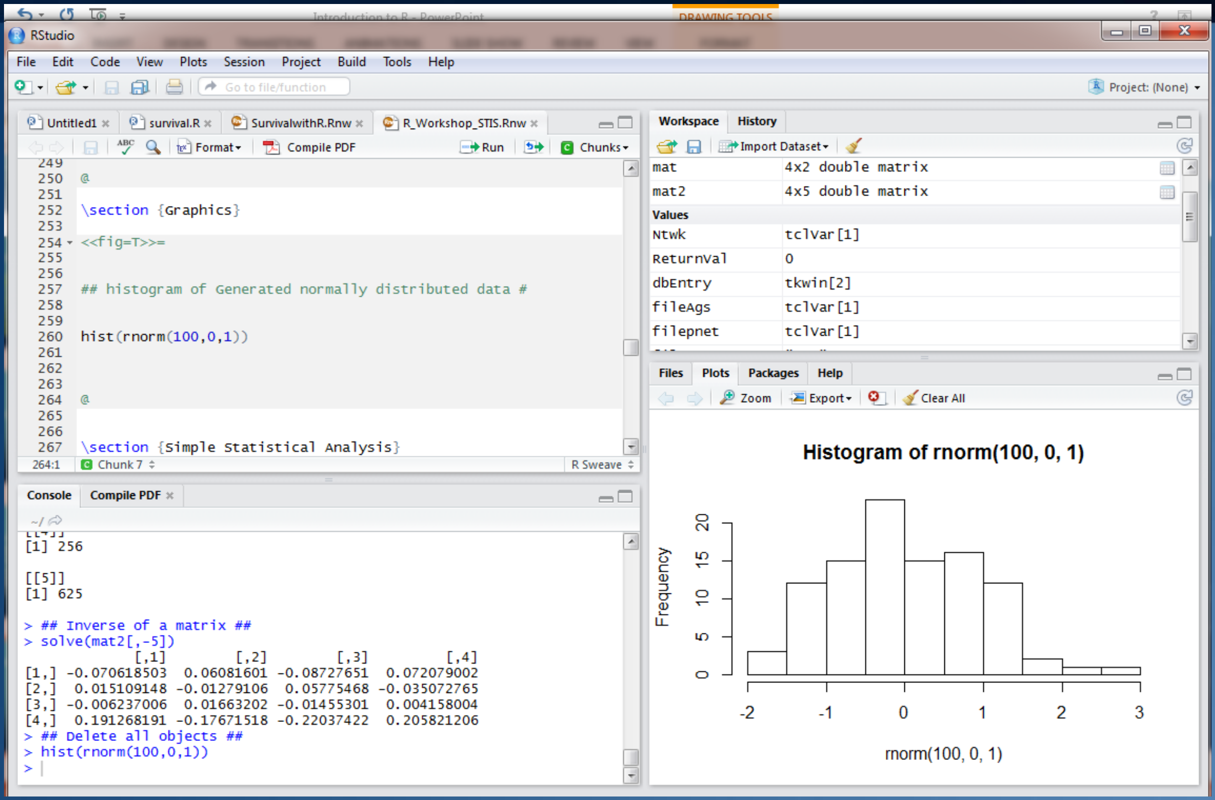
- Integrated development environment (IDE)/Script Editors aimed to provide feature-rich environments to edit R scripts and code: Rstudio (www.rstudio.com), and architect (www.Openanalytics.eu)
- Web based application: the Rweb (Banfield, 1999), R.Net (www.u.arizona.edu/~ryckman/Net.php), or gWidgetsWWW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



```
@

\section {Graphics}

<<fig=T>>=


## histogram of Generated normally distributed data #


hist(rnorm(100,0,1))



@


\section {Simple Statistical Analysis}
```

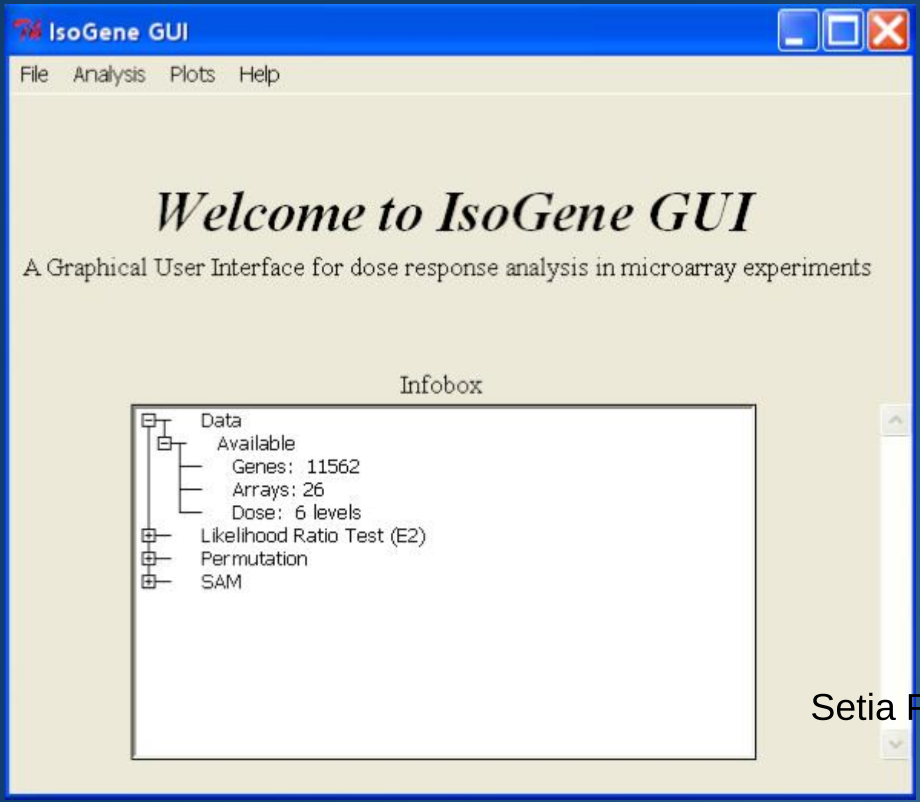
```
[1] 256

[[5]]
[1] 625

> ## Inverse of a matrix ##
> solve(mat2[,-5])
             [,1]        [,2]        [,3]         [,4]
[1,] -0.070618503  0.06081601 -0.08727651  0.072079002
[2,]  0.015109148 -0.01279106  0.05775468 -0.035072765
[3,] -0.006237006  0.01663202 -0.01455301  0.004158004
[4,]  0.191268191 -0.17671518 -0.22037422  0.205821206
> ## Delete all objects ##
> hist(rnorm(100,0,1))
> 
```

mat	4x2 double matrix
mat2	4x5 double matrix
Values	
Ntwk	tclvar[1]
ReturnVal	0
dbEntry	tkwin[2]
fileAgs	tclvar[1]
filepnet	tclvar[1]

RGUI Developed using tcltk



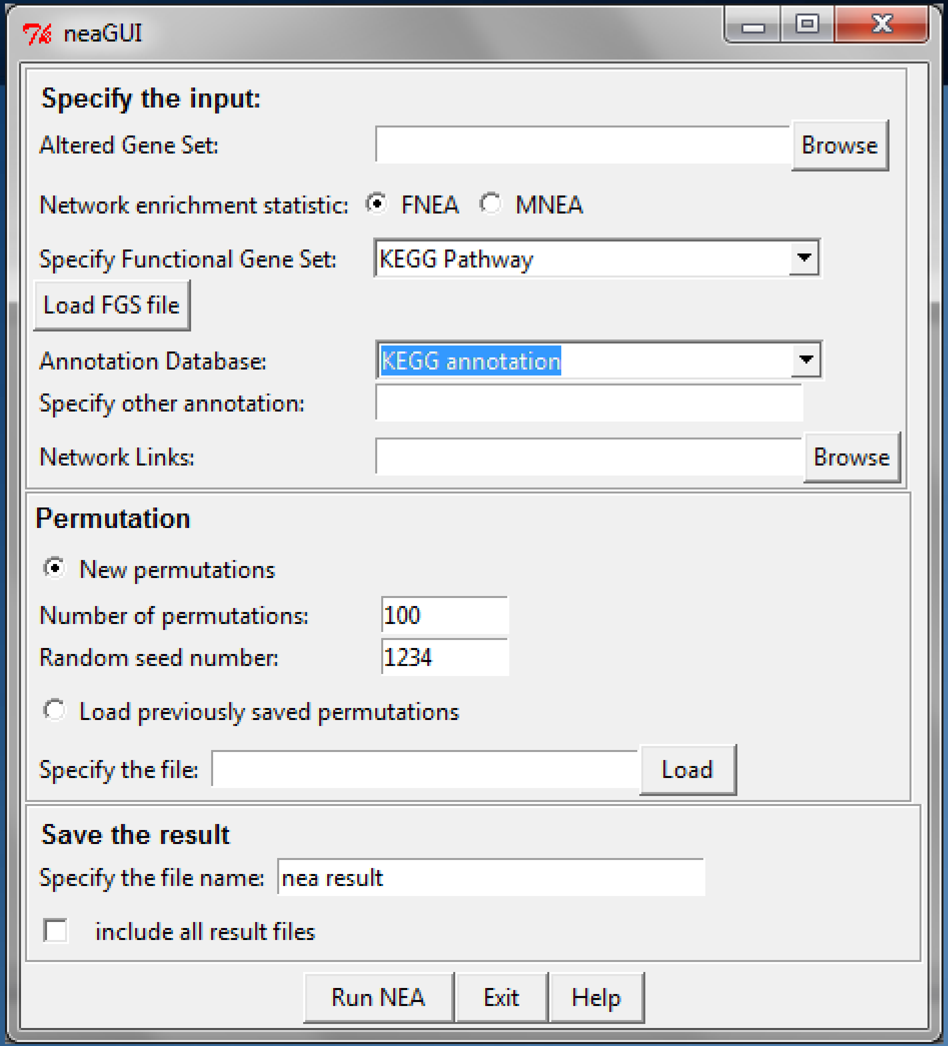
IsoGene GUI

Welcome to IsoGene GUI

A Graphical User Interface for dose response analysis in microarray experiments

Infobox

```
Data
  Available
    Genes:  11562
    Arrays: 26
    Dose:  6 levels
Likelihood Ratio Test (E2)
Permutation
SAM
```



neaGUI

Specify the input:

Altered Gene Set:

Network enrichment statistic: ● FNEA ○ MNEA

Specify Functional Gene Set: KEGG Pathway

Annotation Database: KEGG annotation

Specify other annotation:

Network Links:

Permutation

● New permutations

Number of permutations: 100

Random seed number: 1234

○ Load previously saved permutations

Specify the file:

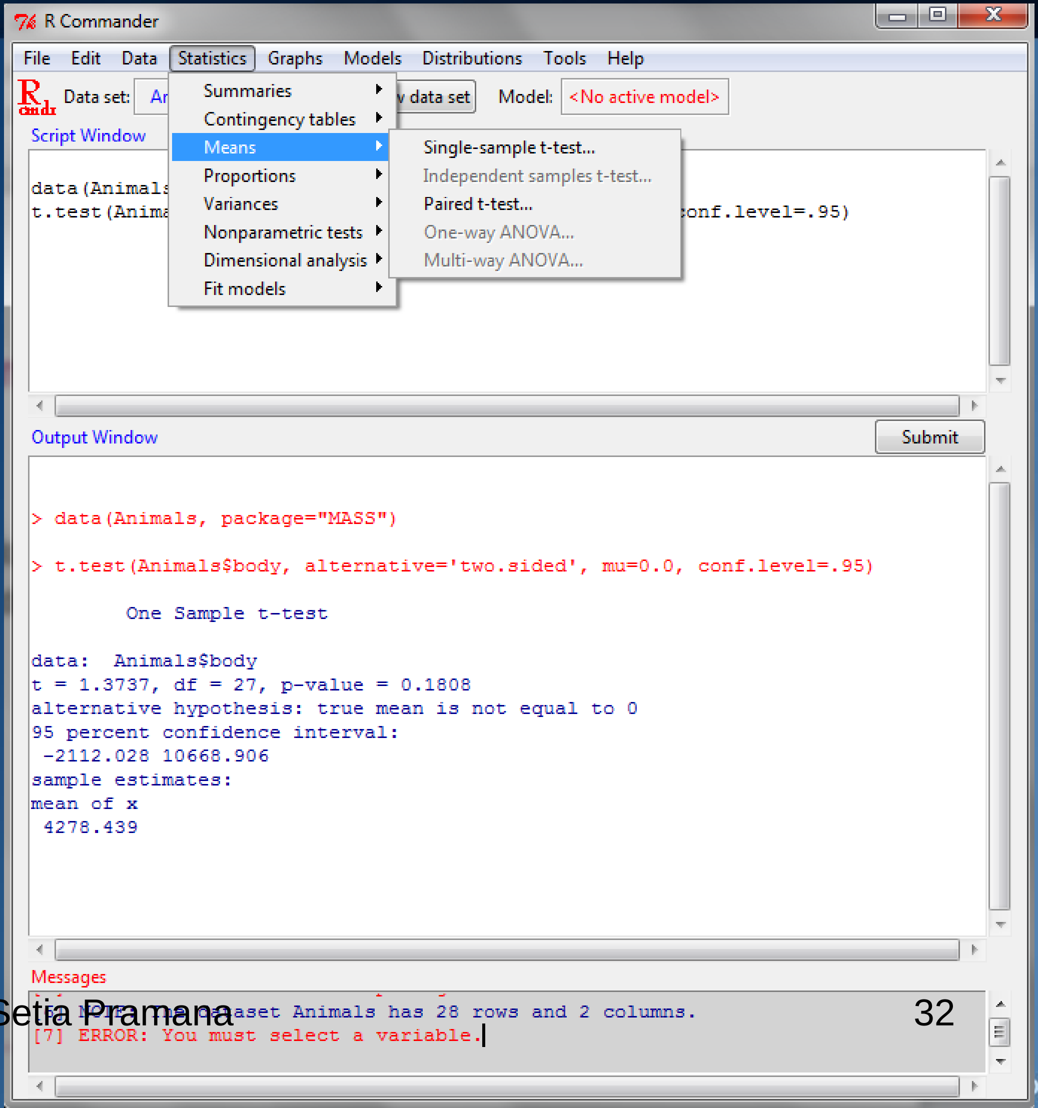
Save the result

Specify the file name: nea result

□ include all result files

RGUI: RCommander

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



```
data(Animals
t.test(Anima                                              conf.level=.95)
```

```
> data(Animals, package="MASS")

> t.test(Animals$body, alternative='two.sided', mu=0.0, conf.level=.95)

        One Sample t-test

data:  Animals$body
t = 1.3737, df = 27, p-value = 0.1808
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -2112.028 10668.906
sample estimates:
mean of x
 4278.439
```

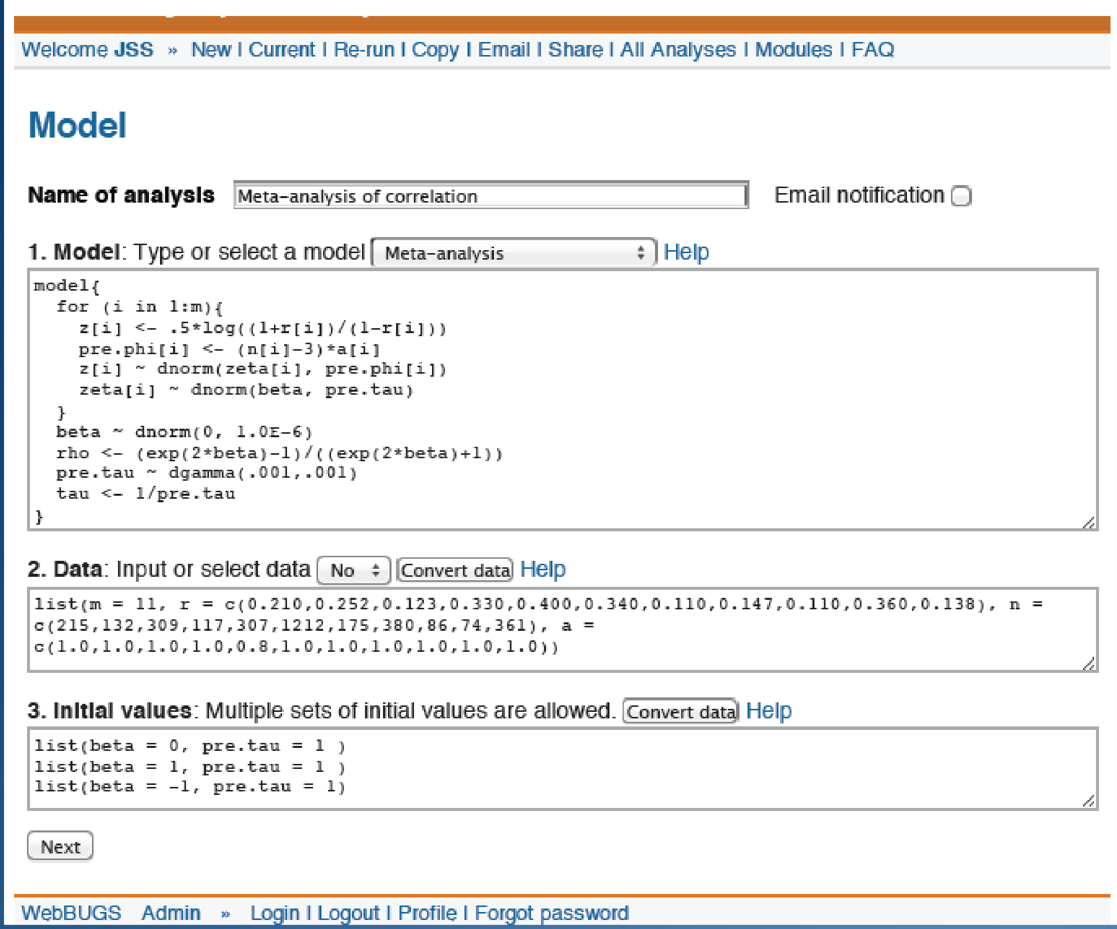
```
[6] NOTE: The dataset Animals has 28 rows and 2 columns.
[7] ERROR: You must select a variable.
```

RGUI: Web Based App

WebBUGS

- Conducting Bayesian Statistical Analysis Online
- Combines OpenBUGS 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

```
model{
  for (i in 1:m){
    z[i] <- .5*log((1+r[i])/(1-r[i]))
    pre.phi[i] <- (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

```
list(m = 11, r = c(0.210,0.252,0.123,0.330,0.400,0.340,0.110,0.147,0.110,0.360,0.138), n =
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))
```

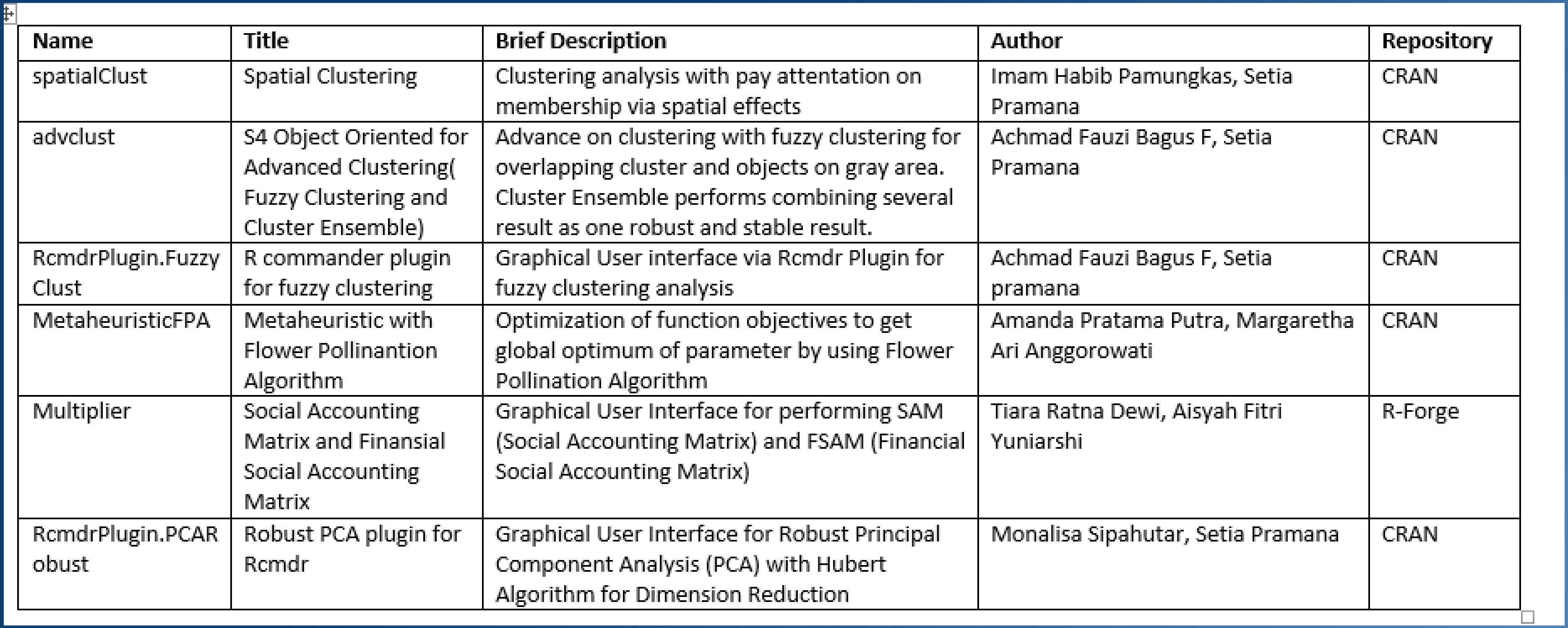
3. Initial values: Multiple sets of initial values are allowed.

```
list(beta = 0, pre.tau = 1 )
list(beta = 1, pre.tau = 1 )
list(beta = -1, pre.tau = 1)
```

RGUI: Shiny

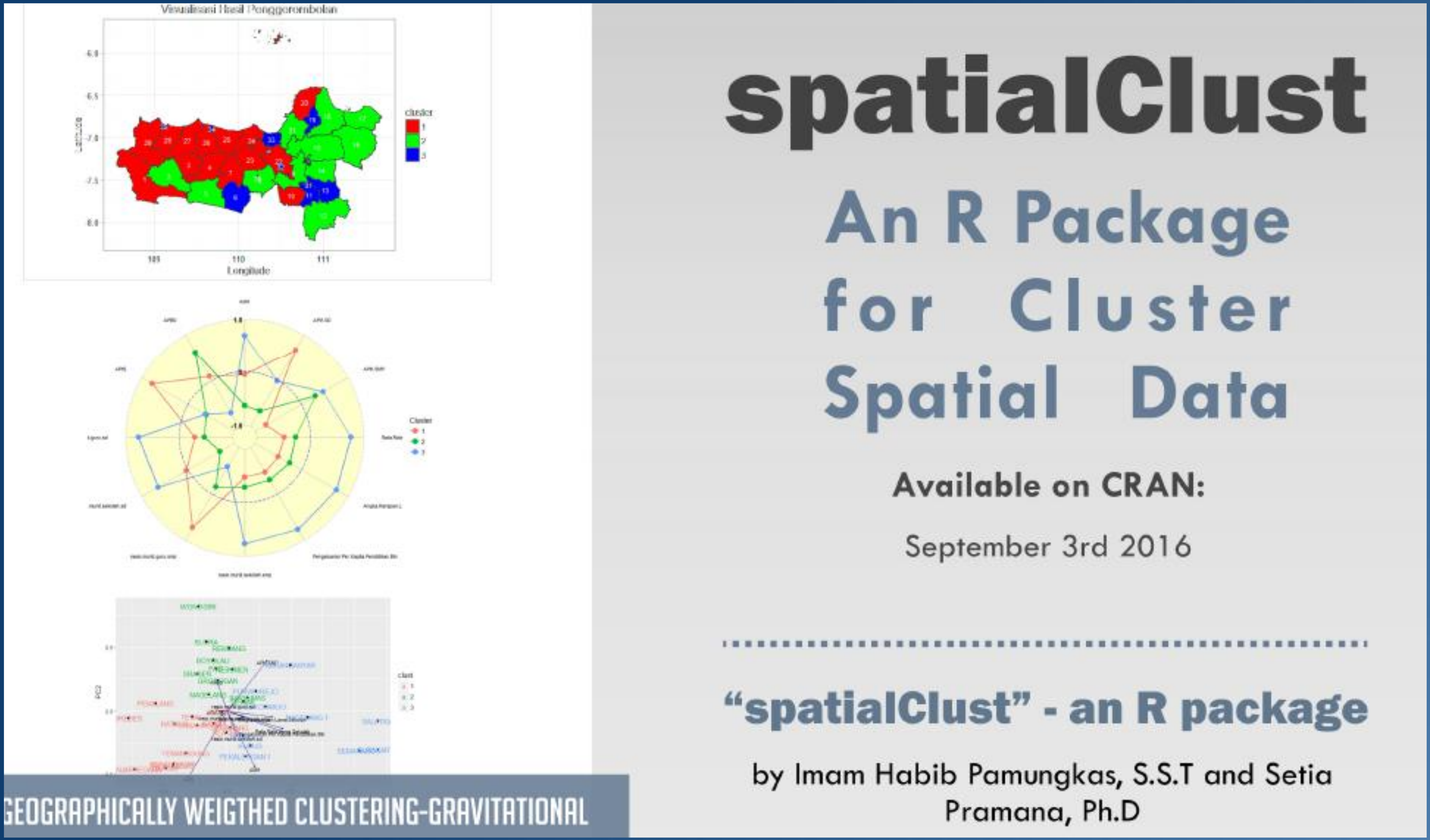
- A new package from Rstudio to build interactive web applications with R.
- Really Easy!
- Build useful web applications with only a few lines of code—no JavaScript required.
- Self learning: http://shiny.rstudio.com/
- http://www.showmeshiny.com/

Our Recent R Packages



Name	Title	Brief Description	Author	Repository
spatialClust	Spatial Clustering	Clustering analysis with pay attentation on membership via spatial effects	Imam Habib Pamungkas, Setia Pramana	CRAN
advclust	S4 Object Oriented for Advanced Clustering(Fuzzy Clustering and Cluster Ensemble)	Advance on clustering with fuzzy clustering for overlapping cluster and objects on gray area. Cluster Ensemble performs combining several result as one robust and stable result.	Achmad Fauzi Bagus F, Setia Pramana	CRAN
RcmdrPlugin.FuzzyClust	R commander plugin for fuzzy clustering	Graphical User interface via Rcmdr Plugin for fuzzy clustering analysis	Achmad Fauzi Bagus F, Setia pramana	CRAN
MetaheuristicFPA	Metaheuristic with Flower Pollinantion Algorithm	Optimization of function objectives to get global optimum of parameter by using Flower Pollination Algorithm	Amanda Pratama Putra, Margaretha Ari Anggorowati	CRAN
Multiplier	Social Accounting Matrix and Finansial Social Accounting Matrix	Graphical User Interface for performing SAM (Social Accounting Matrix) and FSAM (Financial Social Accounting Matrix)	Tiara Ratna Dewi, Aisyah Fitri Yuniarshi	R-Forge
RcmdrPlugin.PCARobust	Robust PCA plugin for Rcmdr	Graphical User Interface for Robust Principal Component Analysis (PCA) with Hubert Algorithm for Dimension Reduction	Monalisa Sipahutar, Setia Pramana	CRAN

Our Recent R Packages


Visualisasi Hasil Penggerombolan
Latitude
Longitude
109
110
111
cluster
1
2
3
Cluster

GEOGRAPHICALLY WEIGTHED CLUSTERING-GRAVITATIONAL

spatialClust

An R Package for Cluster Spatial Data

Available on CRAN:

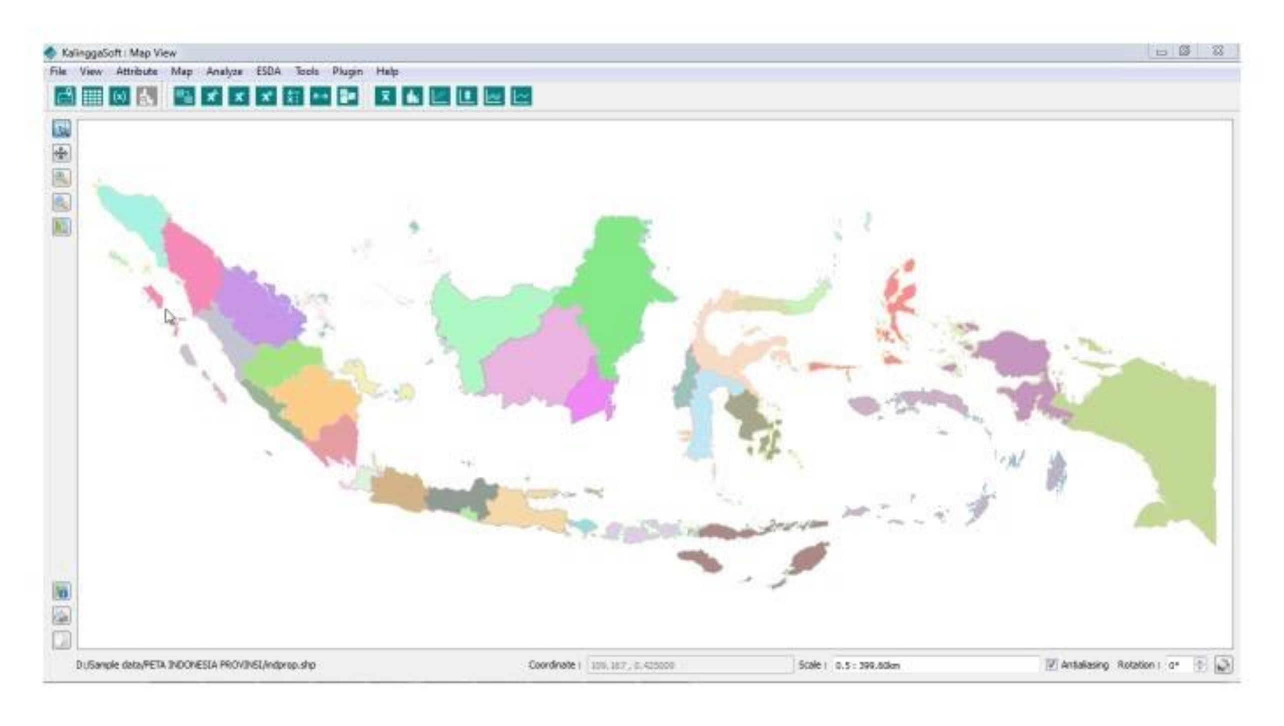
September 3rd 2016

"spatialClust" - an R package

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

Our Recent R Packages

- Kalingga
- Muria
- C++



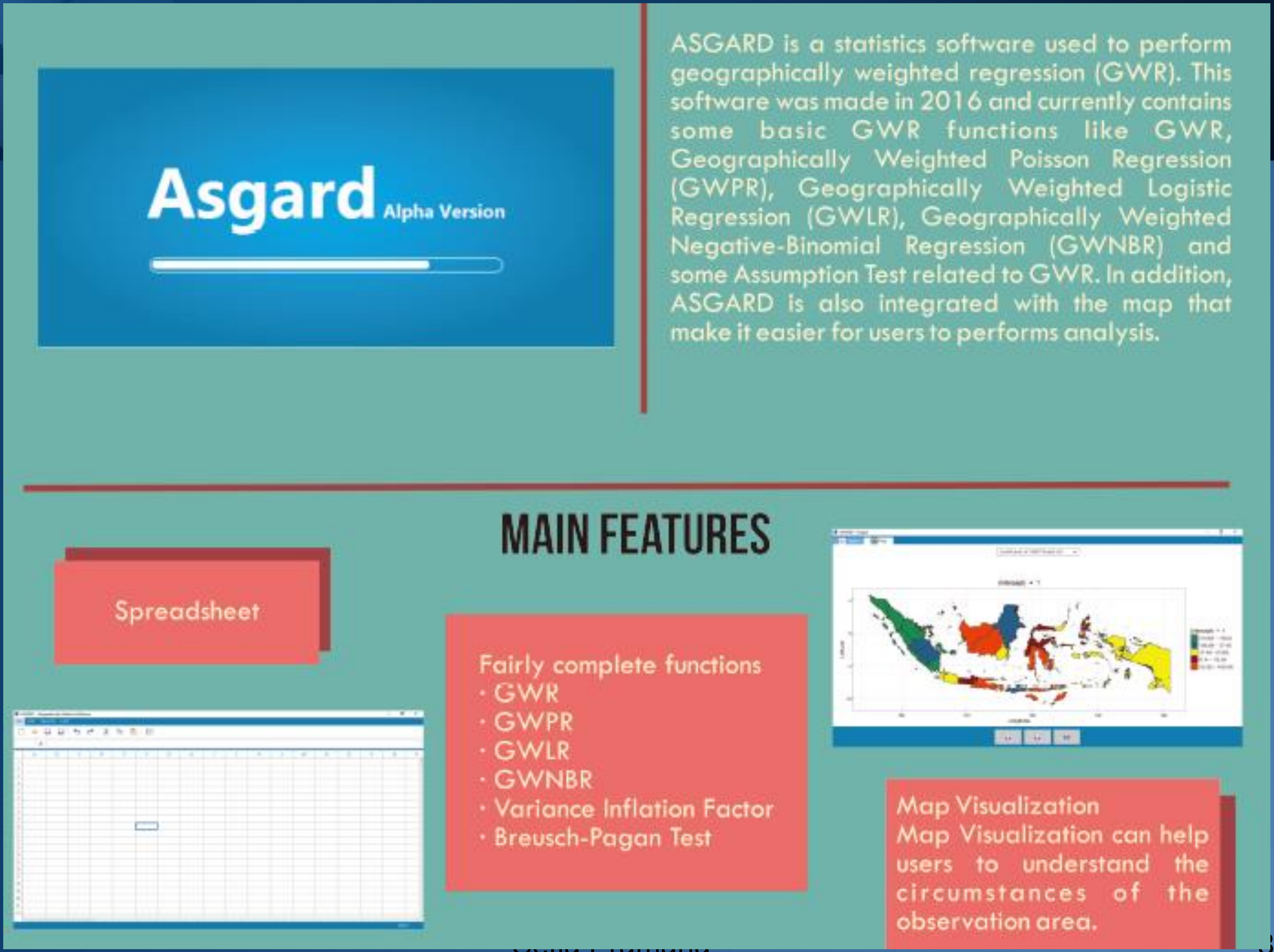
D:/Sample data/PETA INDONESIA PROVINSI/indprop.shp

Coordinate : 109.167 , 0.425000

Scale : 0.5 : 399.60km

☑ Antialiasing

Rotation : 0°



Asgard Alpha Version

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.

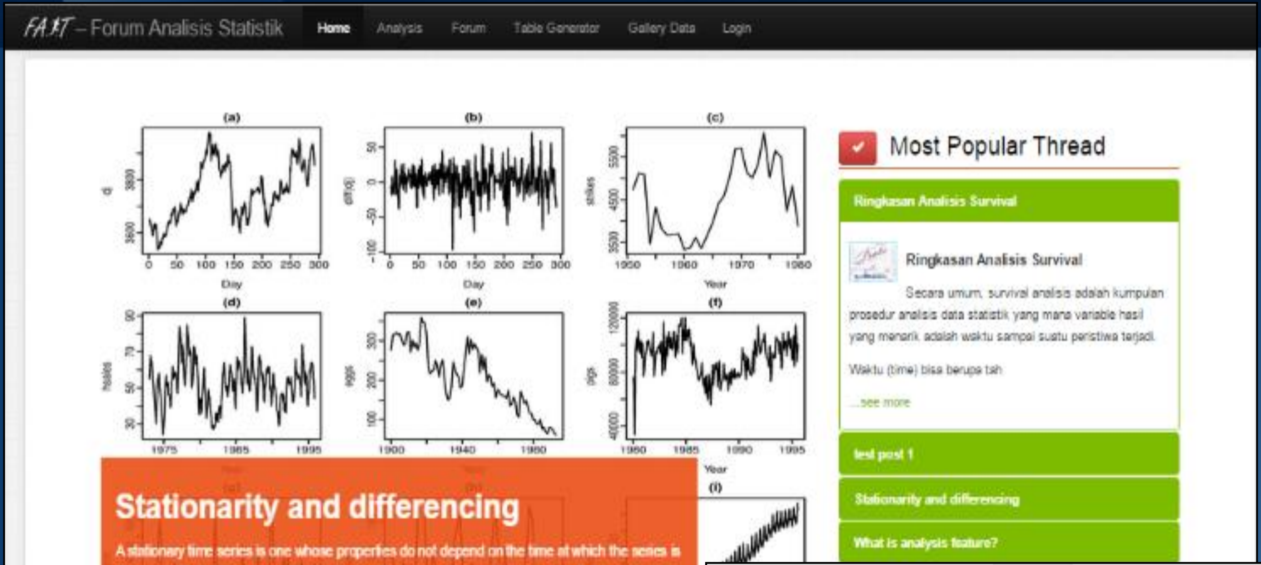
MAIN FEATURES

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



Most Popular Thread

Ringkasan Analisis Survival

Ringkasan Analisis Survival

Secara umum, survival analisis adalah kumpulan prosedur analisis data statistik yang mana variable hasil yang menarik adalah waktu sampai suatu peristiwa terjadi.

Waktu (time) bisa berupa tah

...see more

test post 1

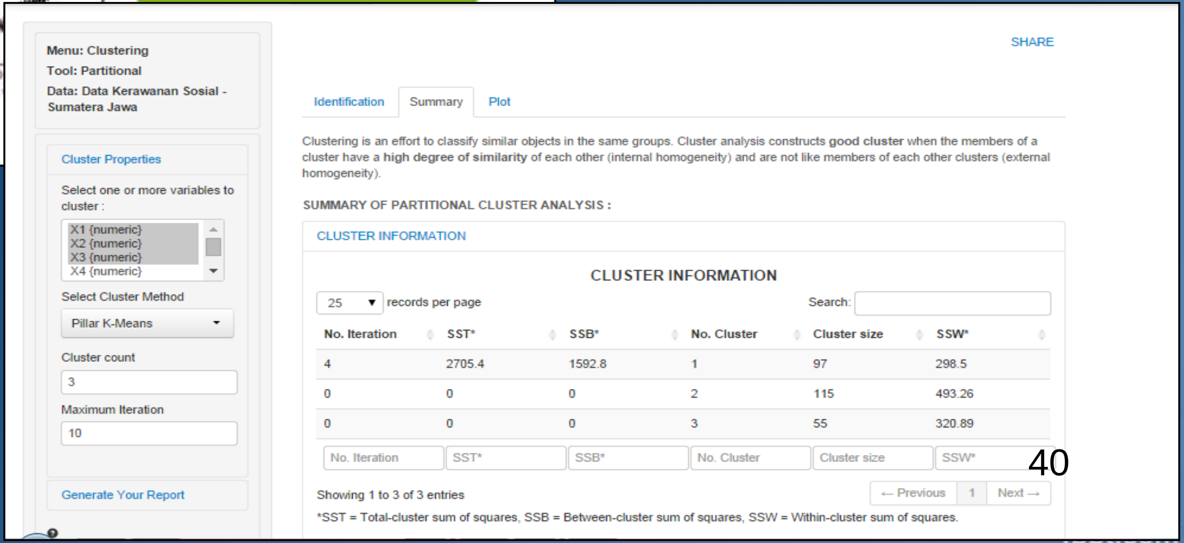
Stationarity and differencing

What is analysis feature?

Stationarity and differencing

A stationary time series is one whose properties do not depend on the time at which the series is observed 1�So time series with trends, or with seasonality, are not stationary � the trend and seas

...see more



Menu: Clustering
Tool: Partitional
Data: Data Kerawanan Sosial - Sumatera Jawa

Select one or more variables to cluster :

X1 {numeric}
X2 {numeric}
X3 {numeric}
X4 {numeric}

Select Cluster Method: Pillar K-Means

Cluster count: 3

Maximum Iteration: 10

Clustering is an effort to classify similar objects in the same groups. Cluster analysis constructs **good cluster** when the members of a cluster have a **high degree of similarity** of each other (internal homogeneity) and are not like members of each other clusters (external homogeneity).

SUMMARY OF PARTITIONAL CLUSTER ANALYSIS :

CLUSTER INFORMATION

CLUSTER INFORMATION

25 ▾ records per page   Search:

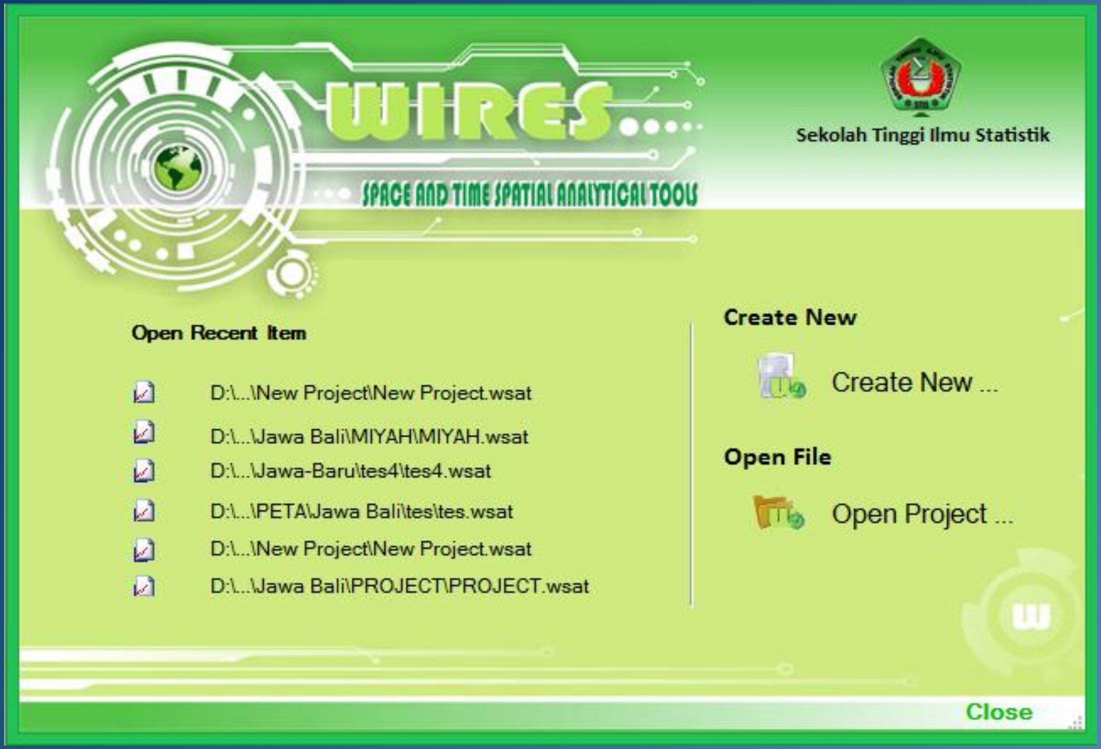
No. Iteration	SST*	SSB*	No. Cluster	Cluster size	SSW*
4	2705.4	1592.8	1	97	298.5
0	0	0	2	115	493.26
0	0	0	3	55	320.89

Showing 1 to 3 of 3 entries   ← Previous 1 Next →

*SST = Total-cluster sum of squares, SSB = Between-cluster sum of squares, SSW = Within-cluster sum of squares.

RGUI using C#: Wires

- For Spatial Data Analysis



WIRES....
SPACE AND TIME SPATIAL ANALYTICAL TOOLS
Sekolah Tinggi Ilmu Statistik

Open Recent Item

D:\...\New Project\New Project.wsat
D:\...\Jawa Bali\MIYAH\MIYAH.wsat
D:\...\Jawa-Baru\tes4\tes4.wsat
D:\...\PETA\Jawa Bali\tes\tes.wsat
D:\...\New Project\New Project.wsat
D:\...\Jawa Bali\PROJECT\PROJECT.wsat

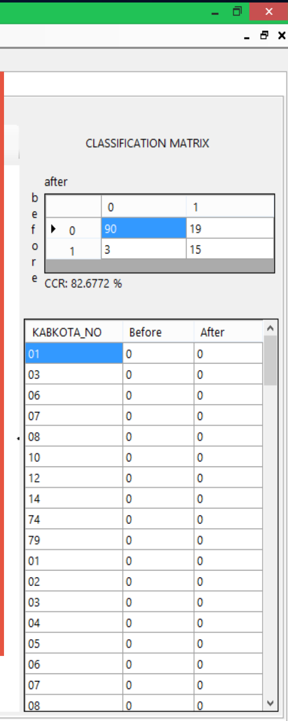
Create New
Create New ...
Open File
Open Project ...
Close

RGUI using C#: Wires

Features

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

WIRES
SPACE AND TIME SPATIAL ANALYTICAL TOOLS
Sekolah Tinggi Ilmu Statistik
Open Recent Item
D:\...\New Project\New Project.wsat
D:\...\Jawa Bali\MIYAH\MIYAH.wsat
D:\...\Jawa-Baru\test4\test4.wsat
D:\...\PETA\Jawa Bali\tes\tes.wsat
D:\...\New Project\New Project.wsat
D:\...\Jawa Bali\PROJECT\PROJECT.wsat
Create New
Create New ...
Open File
Open Project ...
Close



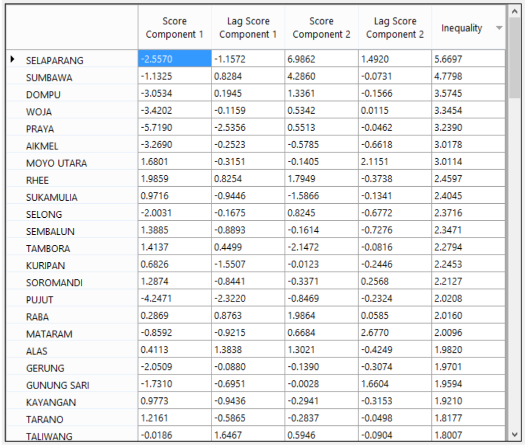
CLASSIFICATION MATRIX

after / before

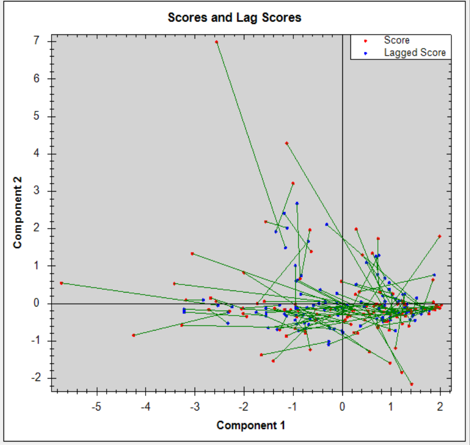
	0	1
0	90	19
1	3	15

CCR: 82.6772 %

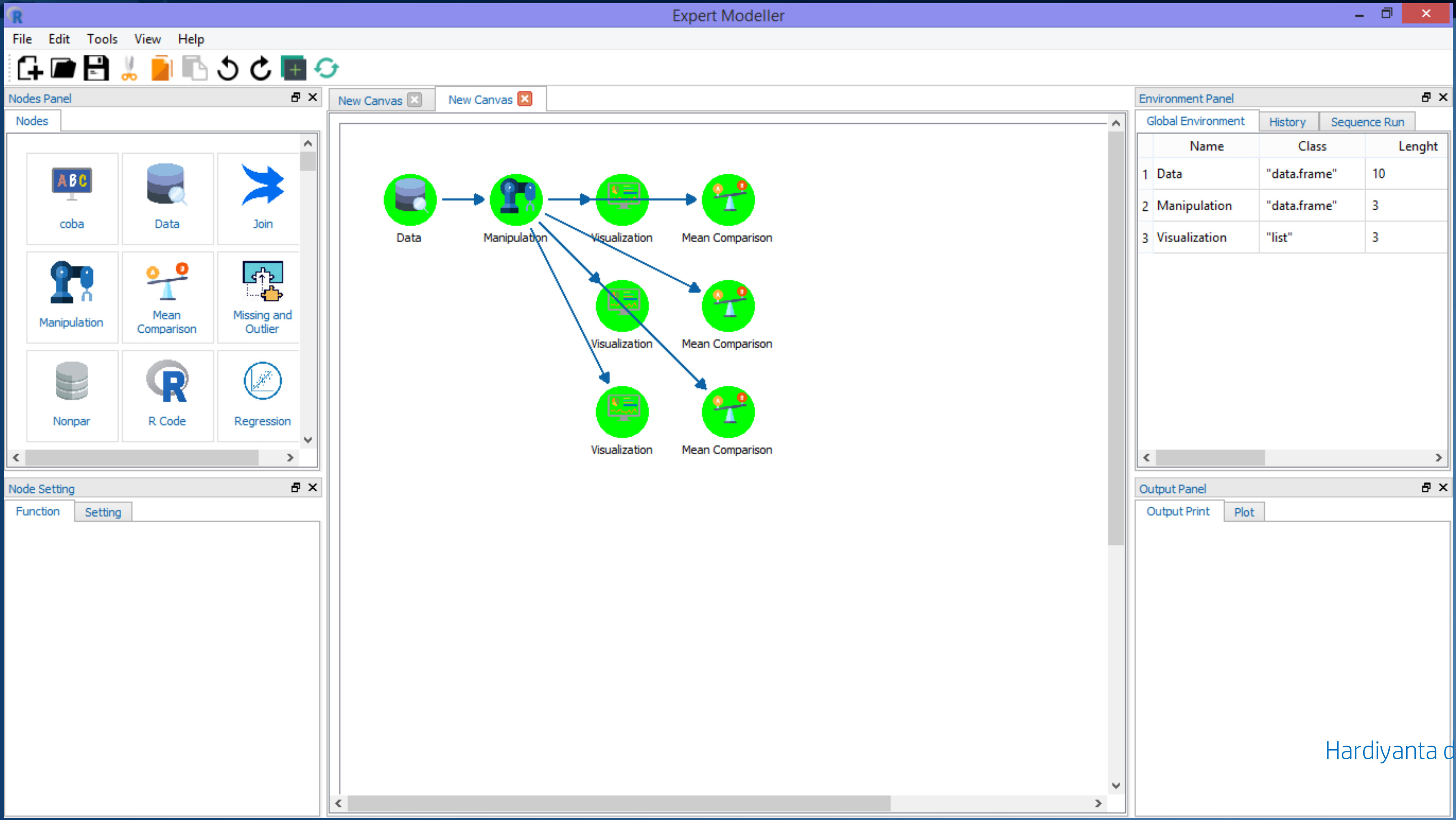
KABKOTA_NO	Before	After
01	0	0
03	0	0
06	0	0
07	0	0
08	0	0
10	0	0
12	0	0
14	0	0
74	0	0
79	0	0
01	0	0
02	0	0
03	0	0
04	0	0
05	0	0
06	0	0
07	0	0
08	0	0



	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



R Expert Modeler



	Name	Class	Lenght
1	Data	"data.frame"	10
2	Manipulation	"data.frame"	3
3	Visualization	"list"	3

Hardiyanta dan Pramana 2018

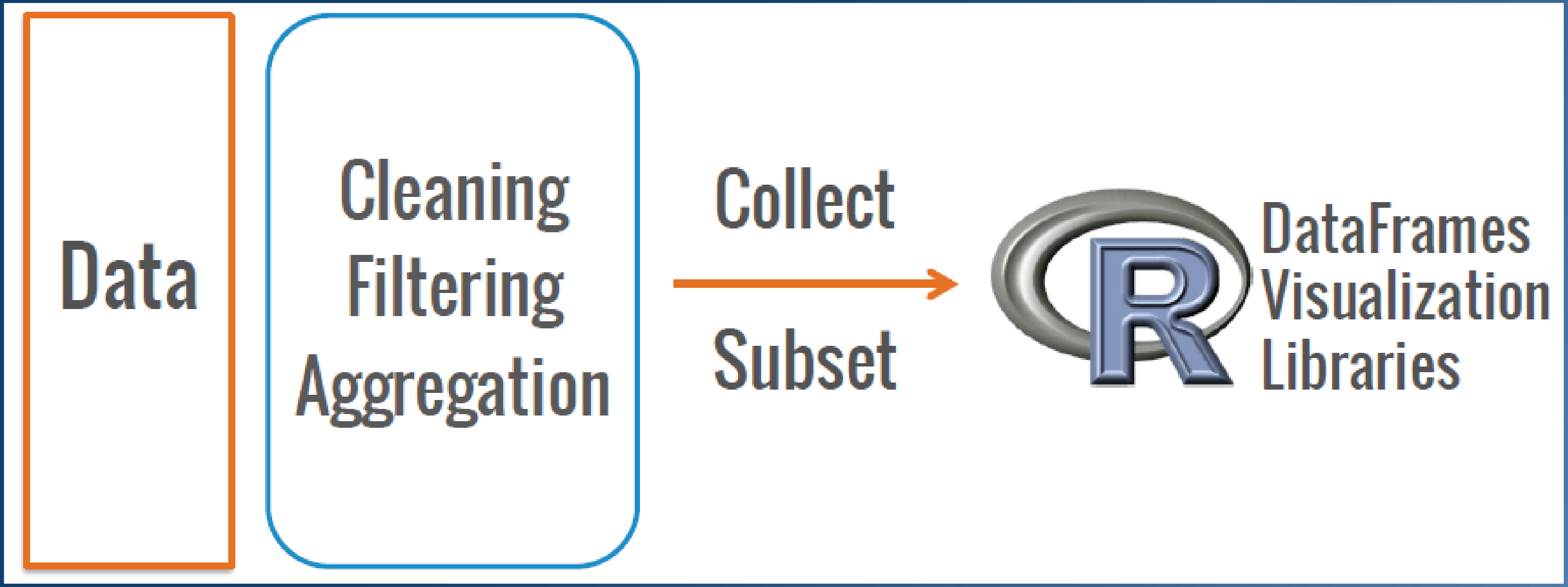
Dynamic Report Generation

Produce documents automatically: pdf, doc, html

Packages:

- Sweave
- knitr
- Markdown

Big Data and R



Data

Cleaning
Filtering
Aggregation

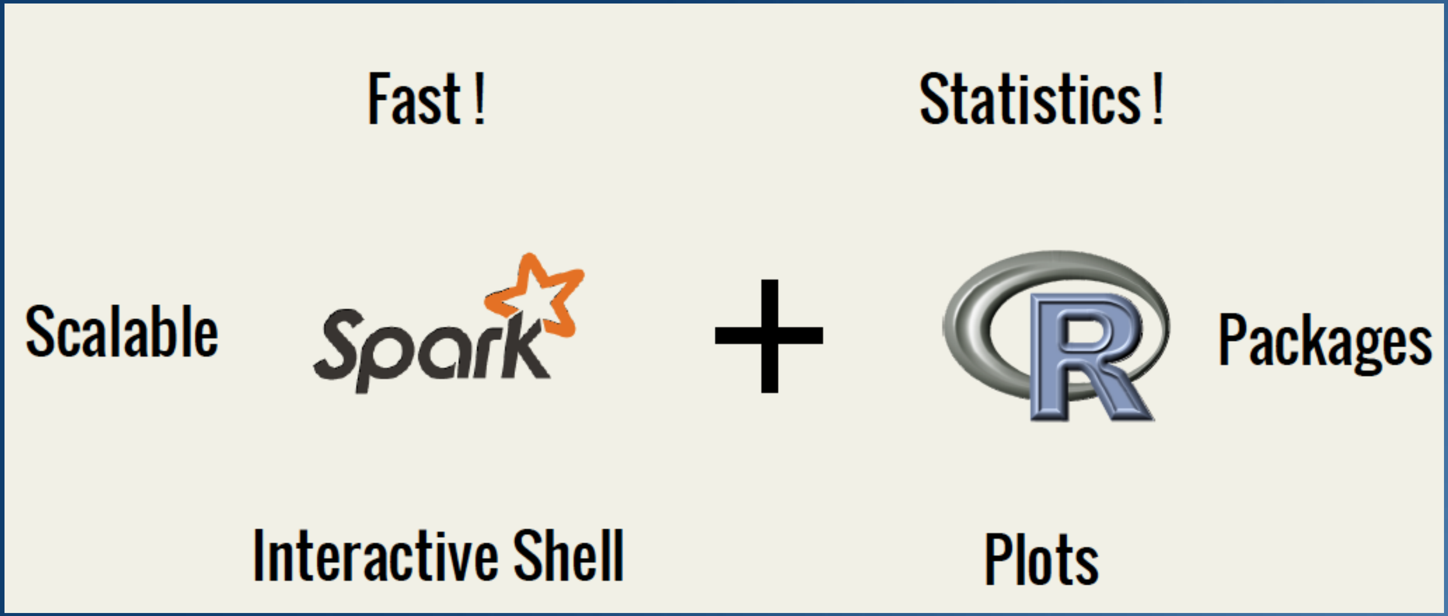
Collect

Subset

DataFrames
Visualization
Libraries

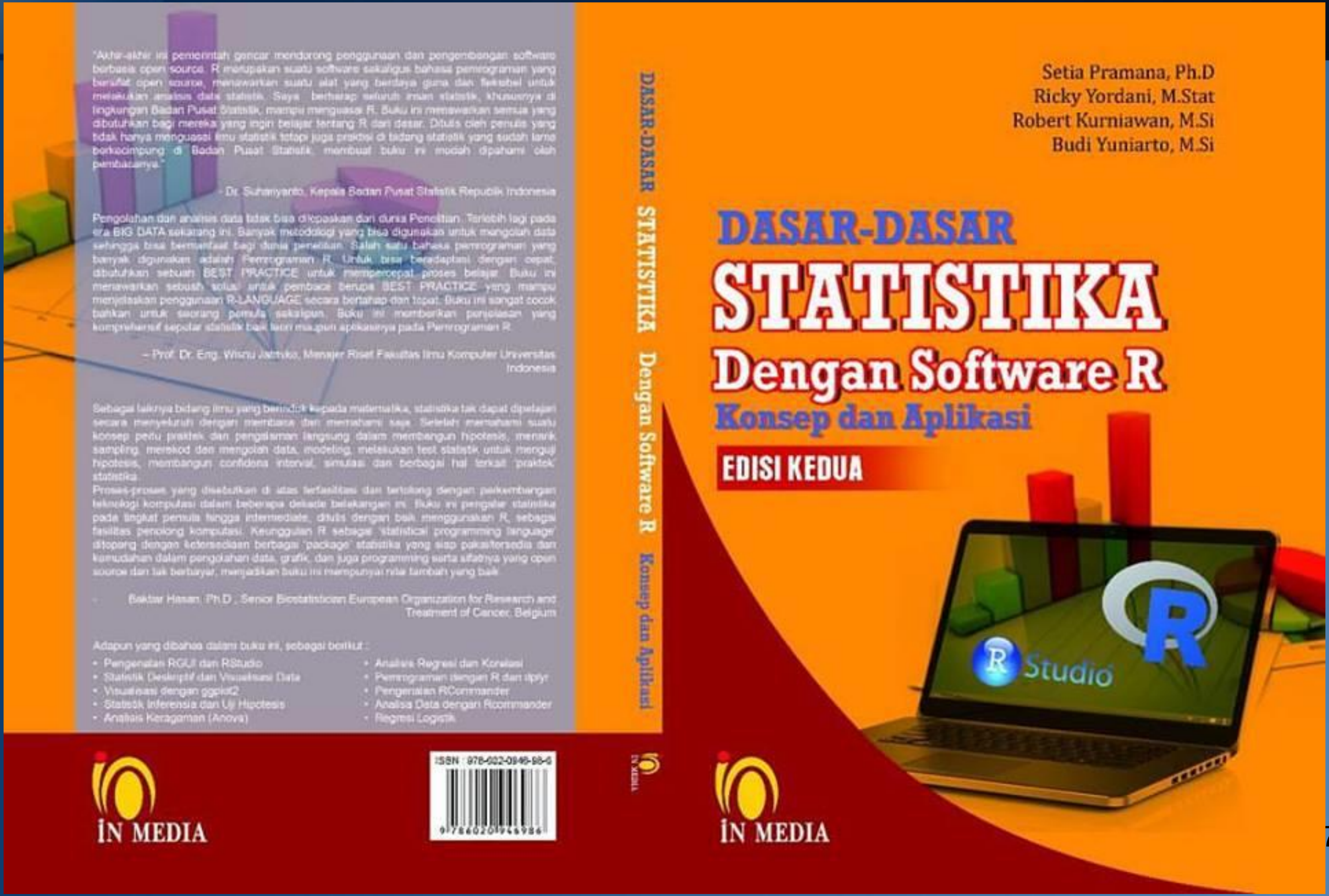
Big Data

- **SparkR**



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

A Start



"Akhir-akhir ini pemerintah gencar mendorong penggunaan dan pengembangan software berbasis open source. R merupakan suatu software sekaligus bahasa pemrograman yang bersifat open source, menawarkan suatu alat yang berdaya guna dan fleksibel untuk melakukan analisis data statistik. Saya berharap seluruh insan statistik, khususnya di lingkungan Badan Pusat Statistik, mampu menguasai R. Buku ini menawarkan semua yang dibutuhkan bagi mereka yang ingin belajar tentang R dari dasar. Ditulis oleh penulis yang tidak hanya menguasai ilmu statistik tetapi juga praktisi di bidang statistik yang sudah lama berkecimpung di Badan Pusat Statistik, membuat buku ini mudah dipahami oleh pembacanya."

- Dr. Suhariyanto, Kepala Badan Pusat Statistik Republik Indonesia

Pengolahan dan analisis data tidak bisa dilepaskan dari dunia Penelitian. Terlebih lagi pada era BIG DATA sekarang ini. Banyak metodologi yang bisa digunakan untuk mengolah data sehingga bisa bermanfaat bagi dunia penelitian. Salah satu bahasa pemrograman yang banyak digunakan adalah Pemrograman R. Untuk bisa beradaptasi dengan cepat, dibutuhkan sebuah BEST PRACTICE untuk mempercepat proses belajar. Buku ini menawarkan sebuah solusi untuk pembaca berupa BEST PRACTICE yang mampu menjelaskan penggunaan R-LANGUAGE secara bertahap dan tepat. Buku ini sangat cocok bahkan untuk seorang pemula sekalipun. Buku ini memberikan penjelasan yang komprehensif seputar statistik baik teori maupun aplikasinya pada Pemrograman R.

– Prof. Dr. Eng. Wisnu Jatmiko, Manajer Riset Fakultas Ilmu Komputer Universitas Indonesia

Sebagai lazimnya bidang ilmu yang bersandar kepada matematika, statistika tak dapat dipelajari secara menyeluruh dengan membaca dan memahami saja. Setelah memahami suatu konsep perlu praktek dan pengalaman langsung dalam membangun hipotesis, menarik sampling, merekod dan mengolah data, modeling, melakukan test statistik untuk menguji hipotesis, membangun confidens interval, simulasi dan berbagai hal terkait 'praktek' statistika.
Proses-proses yang disebutkan di atas terfasilitasi dan tertolong dengan perkembangan teknologi komputasi dalam beberapa dekade belakangan ini. Buku ini pengantar statistika pada tingkat pemula hingga intermediate, ditulis dengan baik menggunakan R, sebagai fasilitas penolong komputasi. Keunggulan R sebagai 'statistical programming language' ditopang dengan ketersediaan berbagai 'package' statistika yang siap pakai/tersedia dan kemudahan dalam pengolahan data, grafik, dan juga programming serta sifatnya yang open source dan tak berbayar, menjadikan buku ini mempunyai nilai tambah yang baik.

- Baktiar Hasan, Ph.D., Senior Biostatistician European Organization for Research and Treatment of Cancer, Belgium

Adapun yang dibahas dalam buku ini, sebagai berikut :

- Pengenalan RGUI dan RStudio
- Statistik Deskriptif dan Visualisasi Data
- Visualisasi dengan ggplot2
- Statistik Inferensia dan Uji Hipotesis
- Analisis Keragaman (Anova)
- Analisis Regresi dan Korelasi
- Pemrograman dengan R dan dplyr
- Pengenalan RCommander
- Analisa Data dengan Rcommander
- Regresi Logistik

IN MEDIA

ISBN : 978-602-0946-98-6
9786020946986

DASAR-DASAR STATISTIKA Dengan Software R Konsep dan Aplikasi

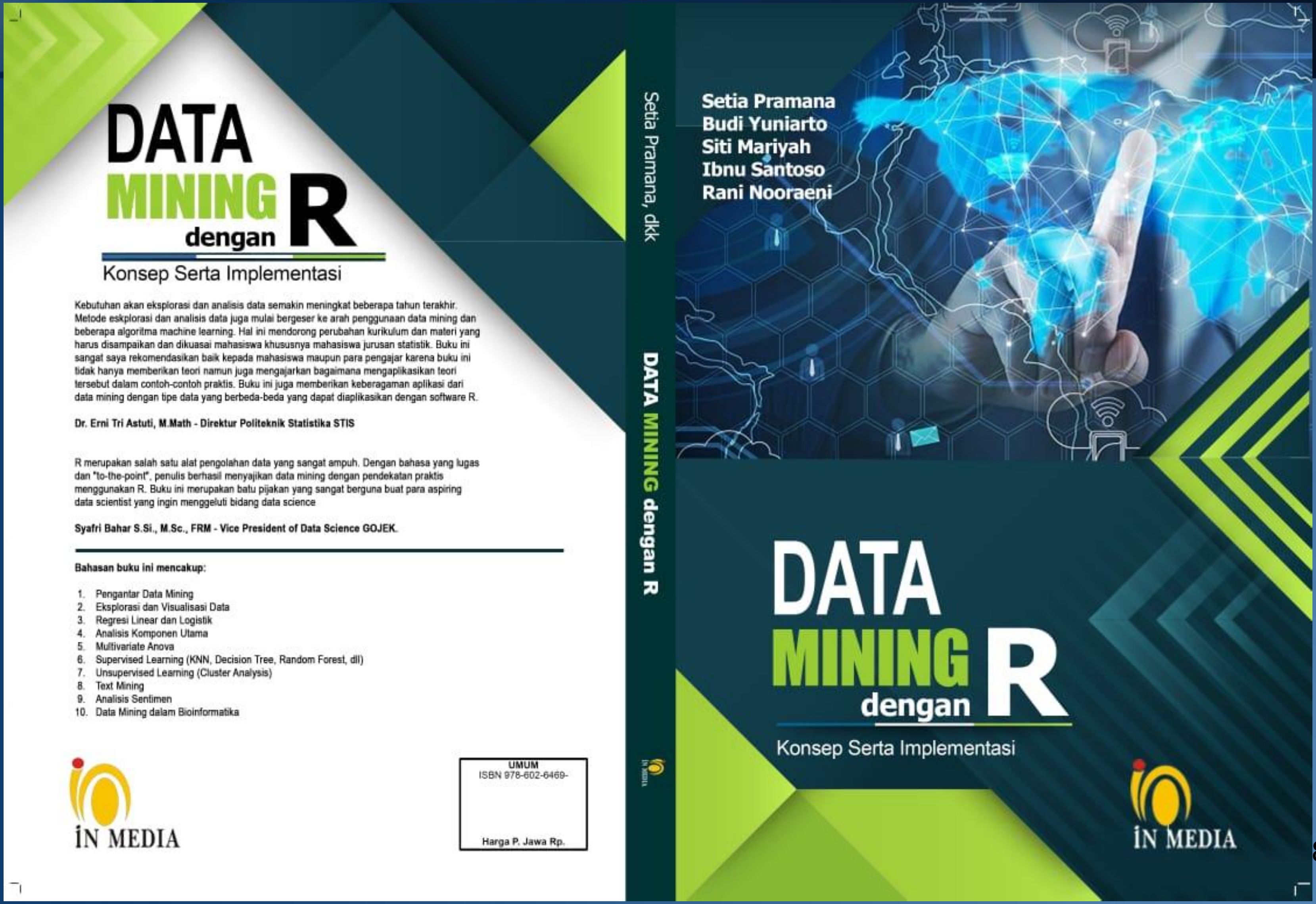
IN MEDIA

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



DATA MINING dengan R

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 beberapa algoritma machine learning. Hal ini mendorong perubahan kurikulum dan materi yang harus disampaikan dan dikuasai mahasiswa khususnya 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:

1. Pengantar Data Mining
2. Eksplorasi dan Visualisasi Data
3. Regresi Linear dan Logistik
4. Analisis Komponen Utama
5. Multivariate Anova
6. Supervised Learning (KNN, Decision Tree, Random Forest, dll)
7. Unsupervised Learning (Cluster Analysis)
8. Text Mining
9. Analisis Sentimen
10. Data Mining dalam Bioinformatika

IN MEDIA

UMUM
ISBN 978-602-6469-

Harga P. Jawa Rp.

Setia Pramana, dkk

DATA MINING dengan R

IN MEDIA

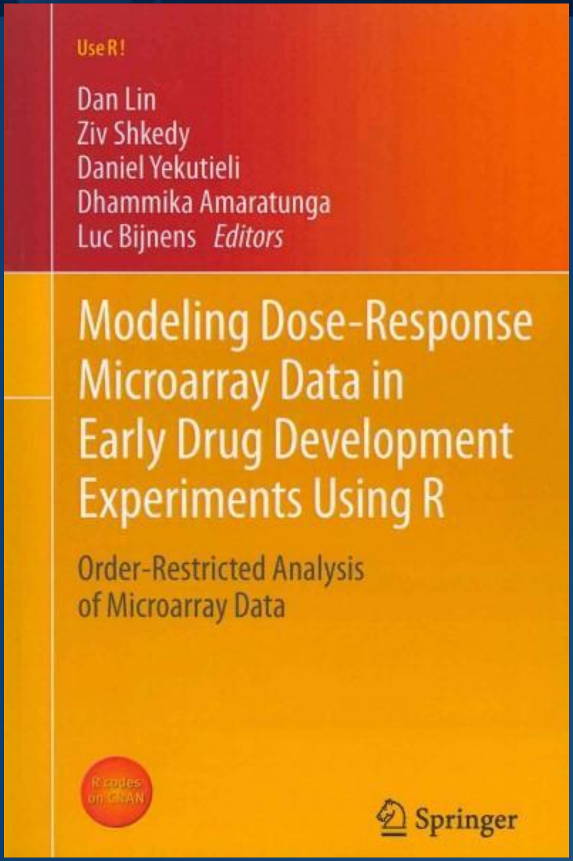
Setia Pramana
Budi Yuniarto
Siti Mariyah
Ibnu Santoso
Rani Nooraeni

DATA MINING dengan R

Konsep Serta Implementasi

IN MEDIA

Book Chapters



Use R!

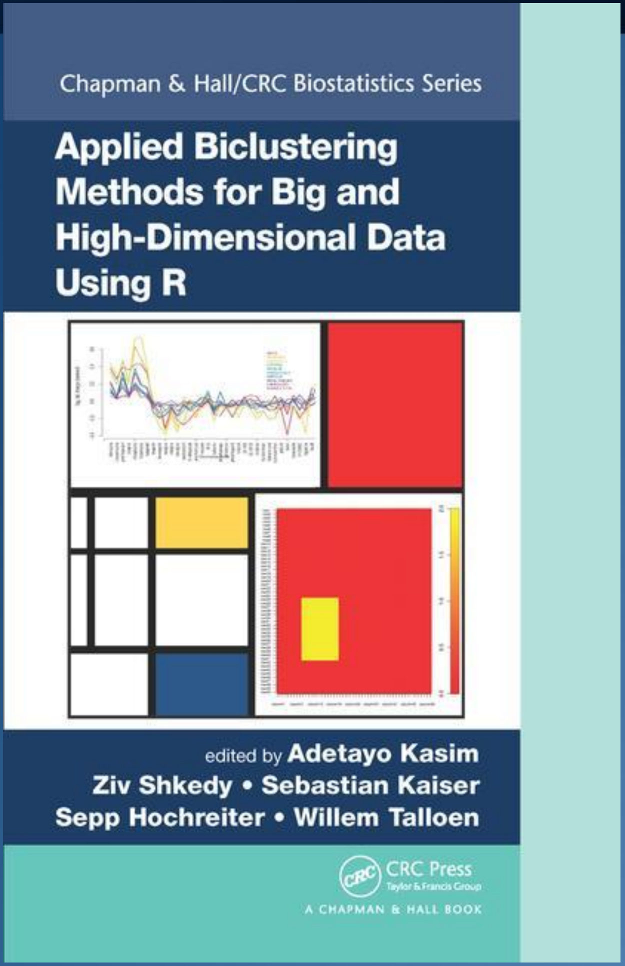
Dan Lin
Ziv Shkedy
Daniel Yekutieli
Dhammika Amaratunga
Luc Bijnens *Editors*

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

Order-Restricted Analysis of Microarray Data

R codes on CRAN

Springer



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

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Conclusion

If statistics programs/languages were cars...

Thank you for your attention!