

Web Deployment with RWas

KRUG 7차 Meet up

(2012-08-16)

유충현

bdboy@r-project.kr

antony.ryu@nexr.com

Agenda

● 개요

- 시스템 개념도
- 시스템 구성도
- 시스템 구현 방법
- 주요 내용

● 기본 기능

- 통계분석 기능
- 테이블 출력 기능
- 그래프 출력 기능
- Trellis 그래프 출력 기능
- 마이닝 그래프 출력 기능
- Excel 출력 기능

● Exception

- Error Recovery

- Function Exit Code

● Debugging

- Trace Back

- Debug

- Browser Mode

- Browse

- Recover

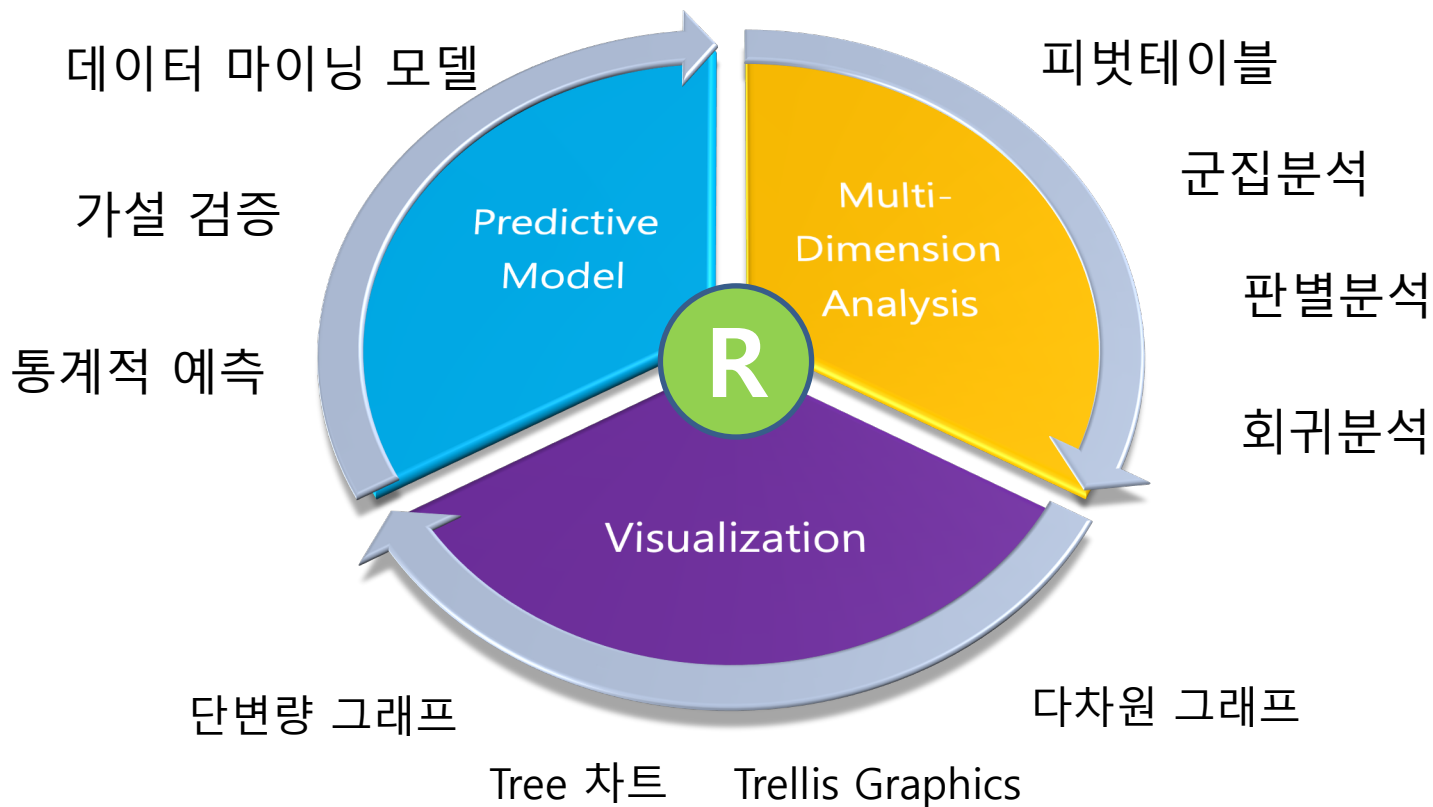
● Q&A

개요

시스템 개념도

개요

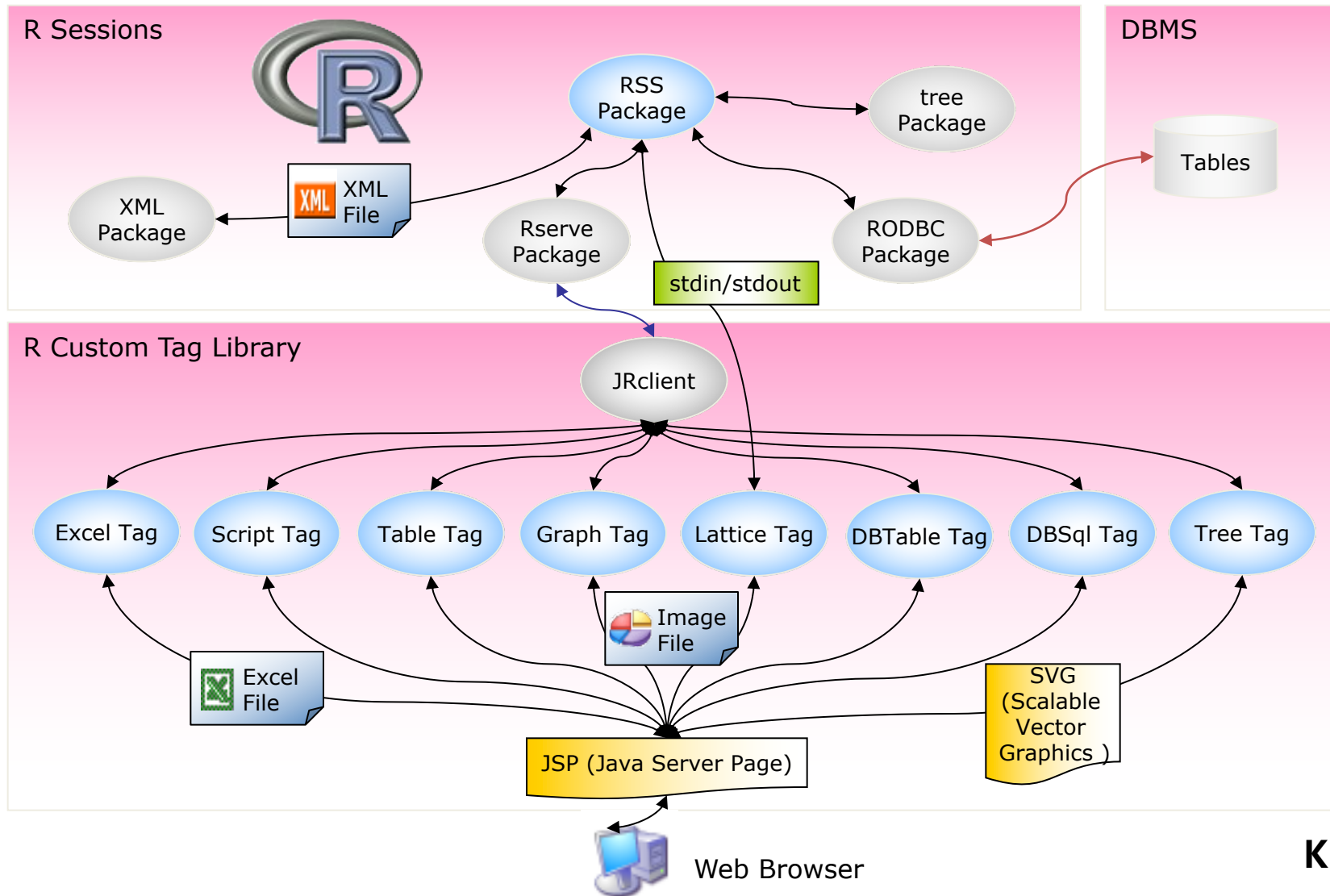
RWas(R Web Application Server)는 통계 프로그램 언어인 **R**을 이용한 Predictive Model, Multi-Dimension Analysis, Visualization 등의 통계적 자료분석의 방법론을 웹 어플리케이션에 구현할 수 있는 통계분석 서버



시스템 구성도

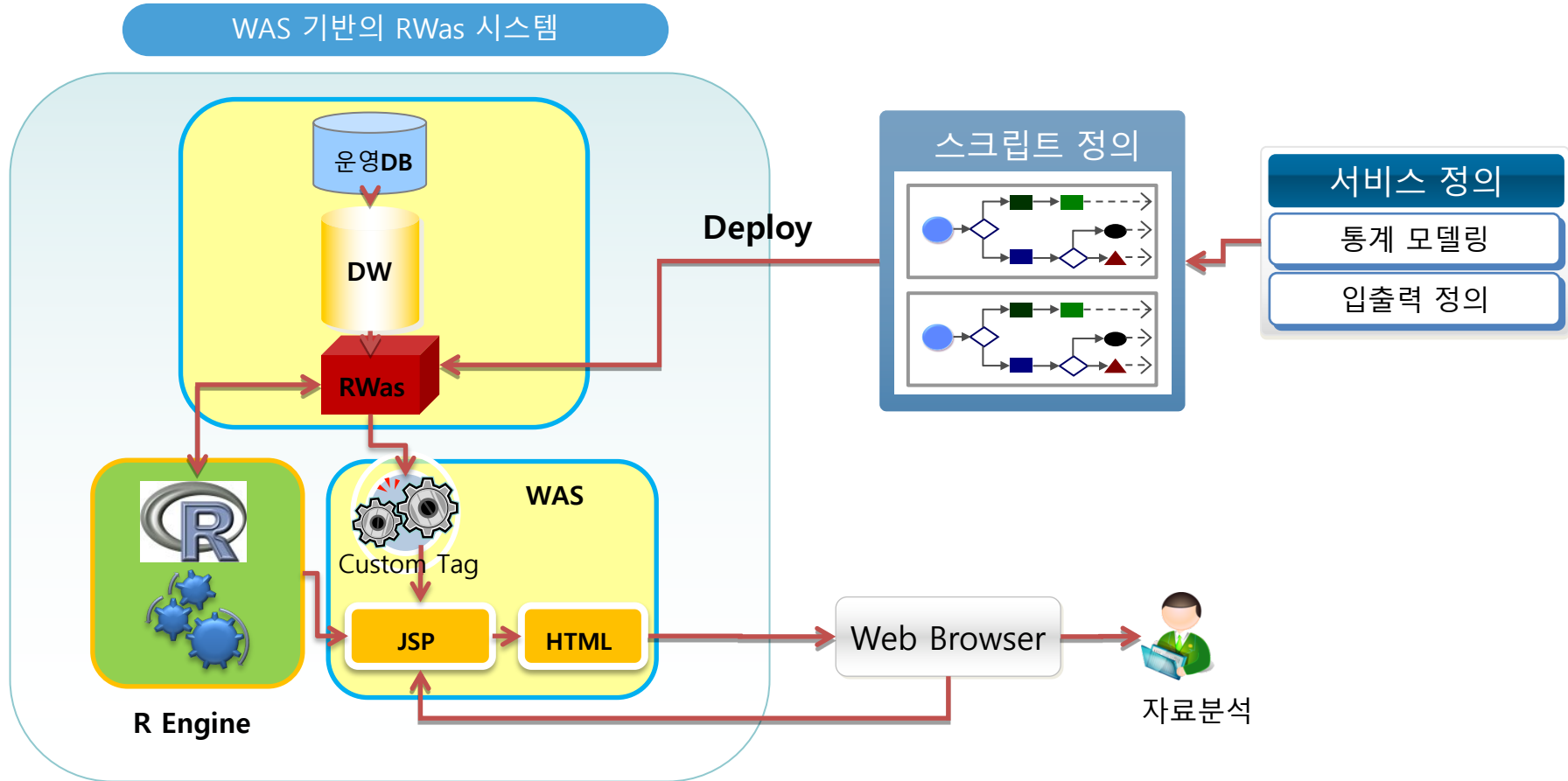
개요

RWas는 통계분석 엔진인 **R Sessions**와 웹 어플리케이션 인터페이스인 **R Custom Tag Library**로 구성되었으며 **DBMS와 상호 연동**하는 인터페이스 라이브러리를 제공



시스템 구현 방법

개요



주요내용

개요

통계분석 기반의 기능

통계분석

통계학 기반의 통계분석 수행 및 결과 출력
마이닝 분석 지원 및 결과 출력
통계분석 Script의 수행 및 결과 출력

Visualization

다양한 통계 그래프의 출력
다차원 분석 도구인 Trellis 그래프의 출력

마이닝 Chart

Classification Tree Model
Regression Tree Model
Tree Model 그래프
SVG 기반의 Tree 그래프 및 검증 그래프
각종 마이닝 Chart

데이터 입출력

SAS, SPSS 데이터 Import
GIF, PNG, JPG, PDF, PS 파일 Export
ODBC를 통한 DBMS 입출력 기능
- Oracle, Informix, DB2, MYSQL 연동
- Access, EXCEL 연동

BI 기반의 기능

DBMS 기반 집계

TABLE 데이터 출력 기능
- 테이블 집계 통계
SQL 기반의 결과 출력 기능
- Cross Table 출력
- Tabulate 출력

데이터 집계 기능

집계테이블 출력
피벗 테이블 출력
테이블 헤더 색상 및 밴딩 기능
집계 결과의 EXCEL Export 기능
- Excel View
- Excel Download

기본 기능

통계분석 기능

기본 기능

통계 분석 언어인 R Script를 수행하고 결과를 출력

(결과)

```
> abc = rnorm(5)
> abc = abc*100
> abc
[1] 111.95711 -25.80146  57.02010 116.00430 -69.56722
> x=1:10
> y=x+rnorm(10)
> z=lm(y~x)
> summary(z)
```

Call:

```
lm(formula = y ~ x)
```

Residuals:

Min	1Q	Median	3Q	Max
-9.7221	-3.9742	0.1469	3.4722	9.7878

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.2388	4.4820	-0.500	0.631
x	0.4421	0.7223	0.612	0.558

Residual standard error: 6.561 on 8 degrees of freedom

Multiple R-squared: 0.04473, Adjusted R-squared: -0.07468

F-statistic: 0.3746 on 1 and 8 DF, p-value: 0.5575

테이블 출력 기능

기본 기능

자료를 집계하고 집계 테이블을 Tabulate Table 과 Crosstab Table로 출력

RowName	age.group	center	Freq
1	+19	일반장기1센터	0
2	20~24	일반장기1센터	3513
3	25~29	일반장기1센터	4963
4	30~34	일반장기1센터	3666
5	35~39	일반장기1센터	4161
6	40~44	일반장기1센터	4053
7	45~49	일반장기1센터	3377
8	50+	일반장기1센터	3503
9	+19	일반장기2센터	1
10	20~24	일반장기2센터	3465
11	25~29	일반장기2센터	4886
12	30~34	일반장기2센터	3598
13	35~39	일반장기2센터	4029
14	40~44	일반장기2센터	3904
15	45~49	일반장기2센터	3276
16	50+	일반장기2센터	3432

Tabulate Table

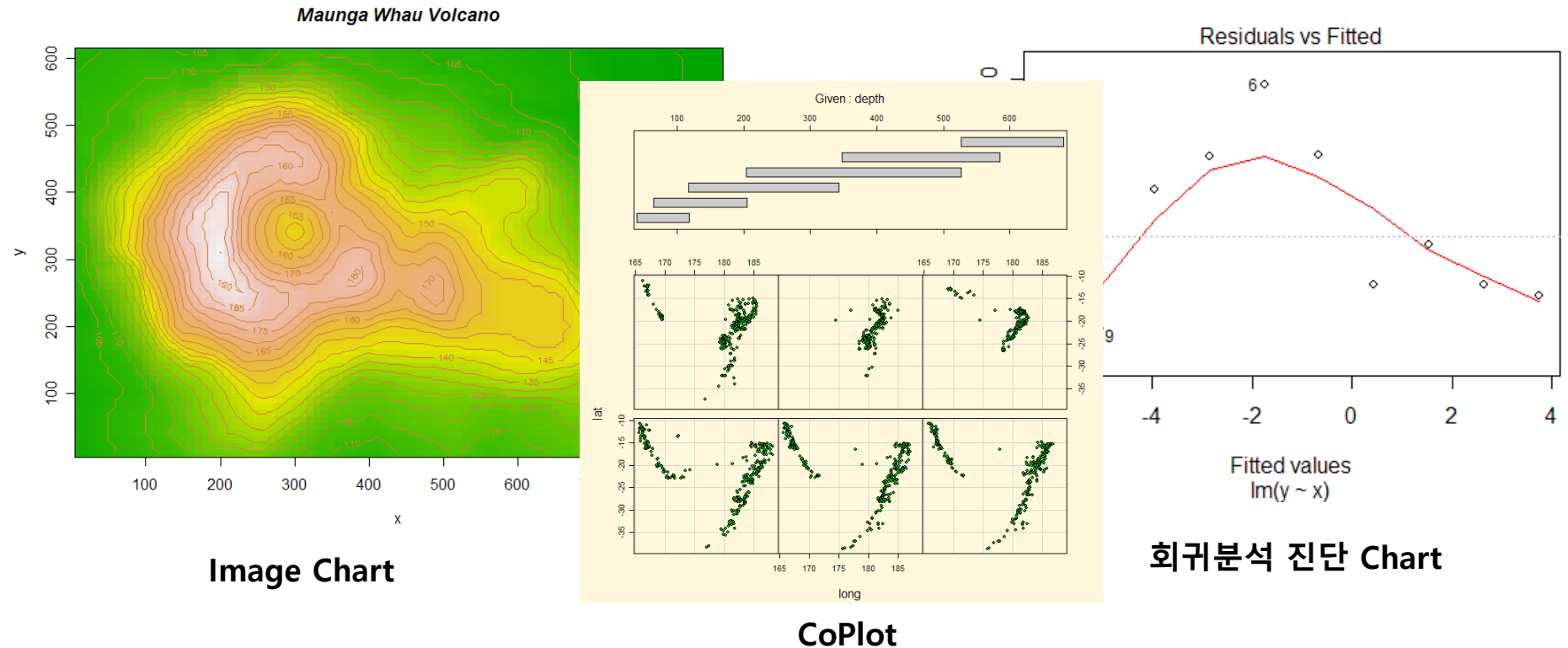
age.group	center	
	일반장기1센터	일반장기2센터
+19	0	1
20~24	3513	3465
25~29	4963	4886
30~34	3666	3598
35~39	4161	4029
40~44	4053	3904
45~49	3377	3276
50+	3503	3432

Crosstab Table

그래프 출력 기능

기본 기능

통계분석 및 통계분석의 결과로 도출되는 다양한 통계 그래프를 웹 어플리케이션 화면에 출력



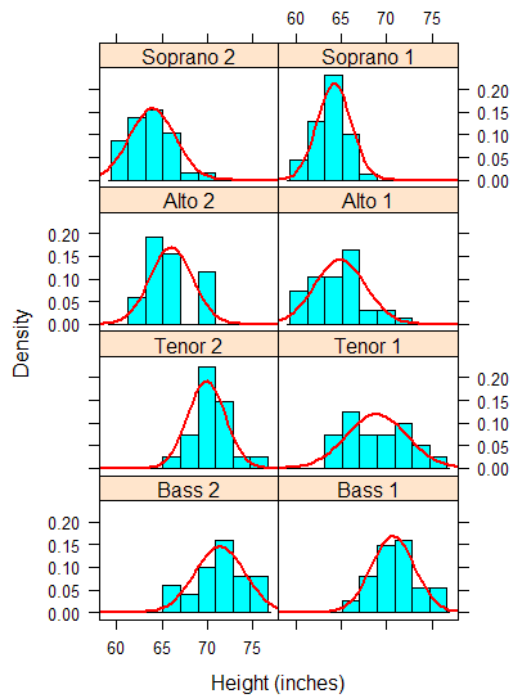
회귀분석 진단 Chart

Trellis 그래프 출력 기능

기본 기능

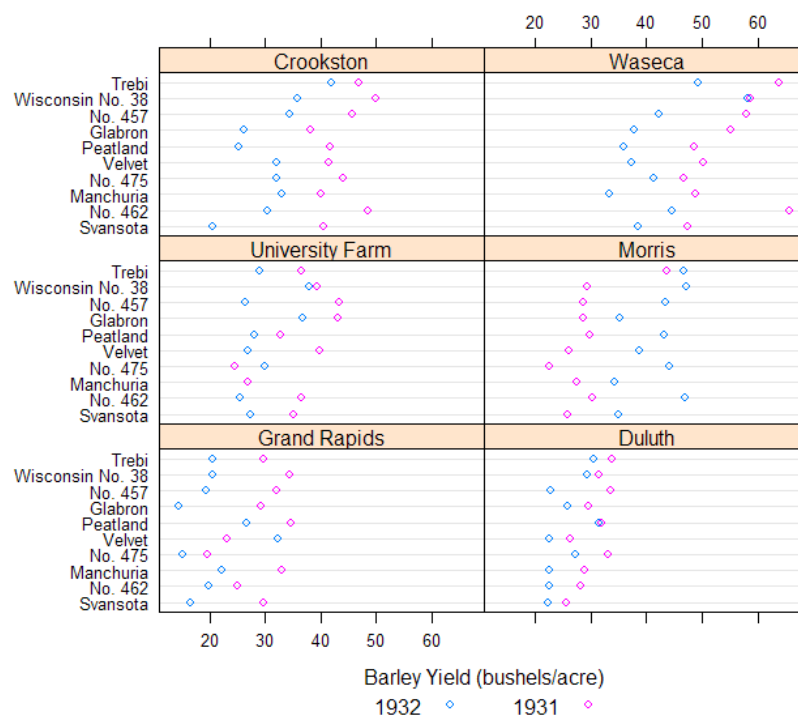
다차원 데이터의 데이터의 상호 관계 및 분포를 파악할 수 있는 여러 종류의 Trellis 그래프를 출력

Heights of New York Choral Society singers



Histogram

Yield data from a Minnesota barley trial

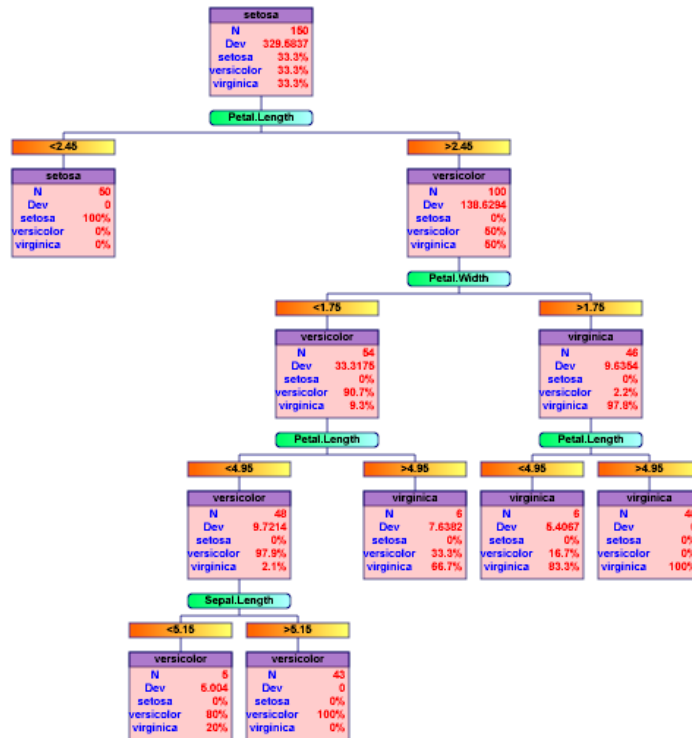


Dot Plot

마이닝 그래프 출력 기능

기본 기능

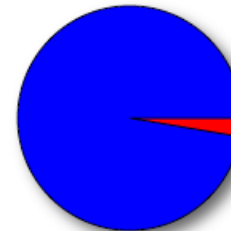
Classification Tree 및 Regression Tree 마이닝 Model을 적합하고 결과를 그래프로 출력



Classification Tree Chart

Tree Type
 Call
 Used variables
 Terminal nodes Cnt
 Resid mean deviance
 Misclass error rate

Classification tree
 tree(tree formula= Species ~ . data= iris)
 Petal.Length / Petal.Width / Sepal.Length
 6
 0.1253 = 18.0489 / 144
 0.027 = 4 / 150



Miss classification - 2.67 %
 Right classification - 97.33 %

Model 검증 Chart

Excel 출력 기능

기본 기능

집계한 통계자료는 Excel Export 기능을 이용하여 Excel 파일로 출력

	A	B	C	D	E	F	G	H
1	age_group	sex	Freq					
2	00~04세	남자	1237301					
3	05~09세	남자	1654228					
4	10~14세	남자	1816318					
5	15~19세	남자	1626378					
6	20~24세	남자	1915902					
7	25~29세	남자	1858332					
8	30~34세	남자	2059913					
9	35~39세	남자	2065668					
10	40~44세	남자	2082427					
11	45~49세	남자	1961859					
12	50~54세	남자	1426597					
13	55~59세	남자	1126997					
14	60~64세	남자	897384					
15	65~69세	남자	755949					
16	70~74세	남자	514241					
17	75~79세	남자	270632					
18	80~84세	남자	136705					
19	85세이상	남자	58819					
20	00~04세	여자	1145049					
21	05~09세	여자	1514659					
22	10~14세	여자	1618573					
23	15~19세	여자	1474145					
24	20~24세	여자	1746221					
25	25~29세	여자	1813515					
26	30~34세	여자	2036369					
27	35~39세	여자	2047117					

Appendix

● Chart Display 방법

- image 파일을 특정 디렉터리(Context root의 하위 디렉터리)에 떨군 후 Html 페이지에서 해당 파일을 img tag로 링크를 걸
 - 일반적으로 *.png 파일이 웹에서 용량 대비 시인성이 높음
- 정기적으로 이미지 파일을 삭제하는 Script를 수행하는 것이 좋음

● lattice package의 사용

- Rserve의 eval 메소드는 lattice 패키지를 지원하지 않음
 - 출력한 그래프를 보면 하얀 공백으로 출력됨
- Rscript나 R 실행파일을 이용한 리다이렉션을 이용해서 그래픽 파일을 생성함
 - 예시


```
"C:/R/R-2.12.2/bin/i386/R" --slave --no-save --no-restore
< "C:/Developments/Tomcat 5.5/webapps/RWas/temp/2012-08-19/18824923A0D662C06DA83A9BC84A03B4/2PIYA1R7.r" >
"C:/Developments/Tomcat 5.5/webapps/RWas/temp/2012-08-19/18824923A0D662C06DA83A9BC84A03B4/2PIYA1R7.out"
```

● Multi-User

- OS별로 차이가 있는 것 같음
- Meetup 시 Q&A에서 인지함
 - Windows에서는 Multi-User를 위해서 포트를 바꿔어서 Rserve를 몇 개 더 띄워야 하는 것 같음

● R에서 Web Deployment를 위해서는 Rserve 패키지를 이용하는 것이 가장 효과적이라는 중론임

R 웹 어플리케이션 서버의 개발 및 응용에 대한 연구_발표자료

Demo

Q & A