R Visualization

– R Graph 및 패키지 소개 중심으로 -

넥스알 – 데이터분석팀 유충현

목차

- **Visualization의 필요성**
- **R Graphics**
- **유용한 R Graphics**
  - **lattice, SNS, wordCloud, maps**
- **Interactive R Graphics**
  - **rgl, iplots, rggobi**
- **Interface External Applications**
  - **GoogleMaps, GooleVis**
- **Another R Graphics**

Visualization의 필요성

Anscombe – regression

원천 데이터

Anscombe, Francis J. - American Statistician - **"Graphs in statistical analysis"** - 1973

관측수	I		II		III		IV	
	x1	y1	x2	y2	x3	y3	x4	y4
1	10	8.04	10	9.14	10	7.46	8	6.58
2	8	6.95	8	8.14	8	6.77	8	5.76
3	13	7.58	13	8.74	13	12.74	8	7.71
4	9	8.81	9	8.77	9	7.11	8	8.84
5	11	8.33	11	9.26	11	7.81	8	8.47
6	14	9.96	14	8.1	14	8.84	8	7.04
7	6	7.24	6	6.13	6	6.08	8	5.25
8	4	4.26	4	3.1	4	5.39	19	12.5
9	12	10.84	12	9.13	12	8.15	8	5.56
10	7	4.82	7	7.26	7	6.42	8	7.91
11	5	5.68	5	4.74	5	5.73	8	6.89

RUGK **R User Conference 2011** -3-

Visualization의 필요성

통계량 및 단순회귀분석

통계량

지표	I		II		III		IV	
	x1	y1	x2	y2	x3	y3	x4	y4
평균	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
분산	11	4.1273	11	4.1276	11	4.1226	11	4.1232
상관계수	0.8164205		0.8162365		0.8162867		0.8165214	

단순회귀분석

지표	I	II	III	IV
Coefficient Intercept	3.0	3.0	3.0	3.0
Coefficient x	0.5	0.5	0.5	0.5
Regression sum of squares	27.51	27.50	27.47	27.49
Residuals sum of squares	13.76	13.78	13.76	13.74
Estimated standard error of b1	0.12	0.12	0.12	0.12
Multiple R-square	0.67	0.67	0.67	0.67

RUGK   **R User Conference 2011**

Visualization의 필요성

Visualization

y ~ x | quartet
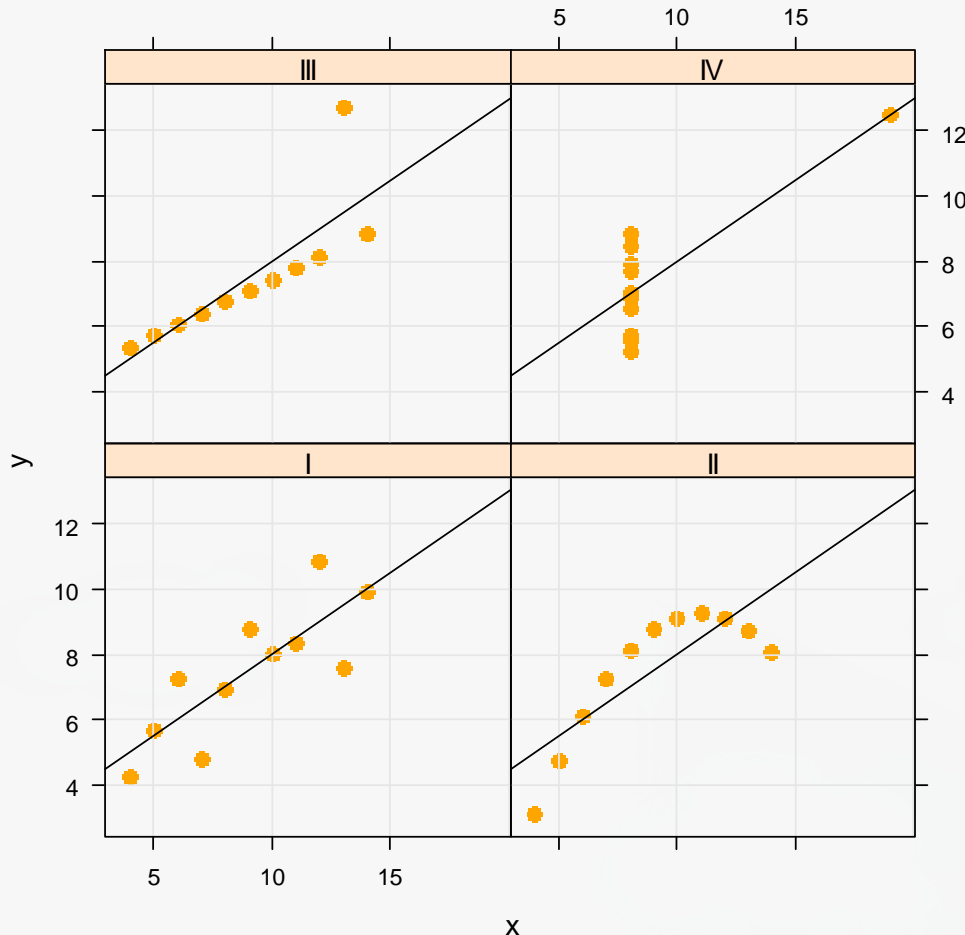
통계량 및 회귀계수 등의 수치는 동일

$$\hat{y} = 3.0 + 0.5x$$

산점도 상의 네 데이터 분포는 상이

Multivariate Visualization

R. A. Fisher's "The design of experiments"   **1930s ~ 1990s 인용된 자료**

Barley Yield (bushels/acre)

Visualization의 필요성
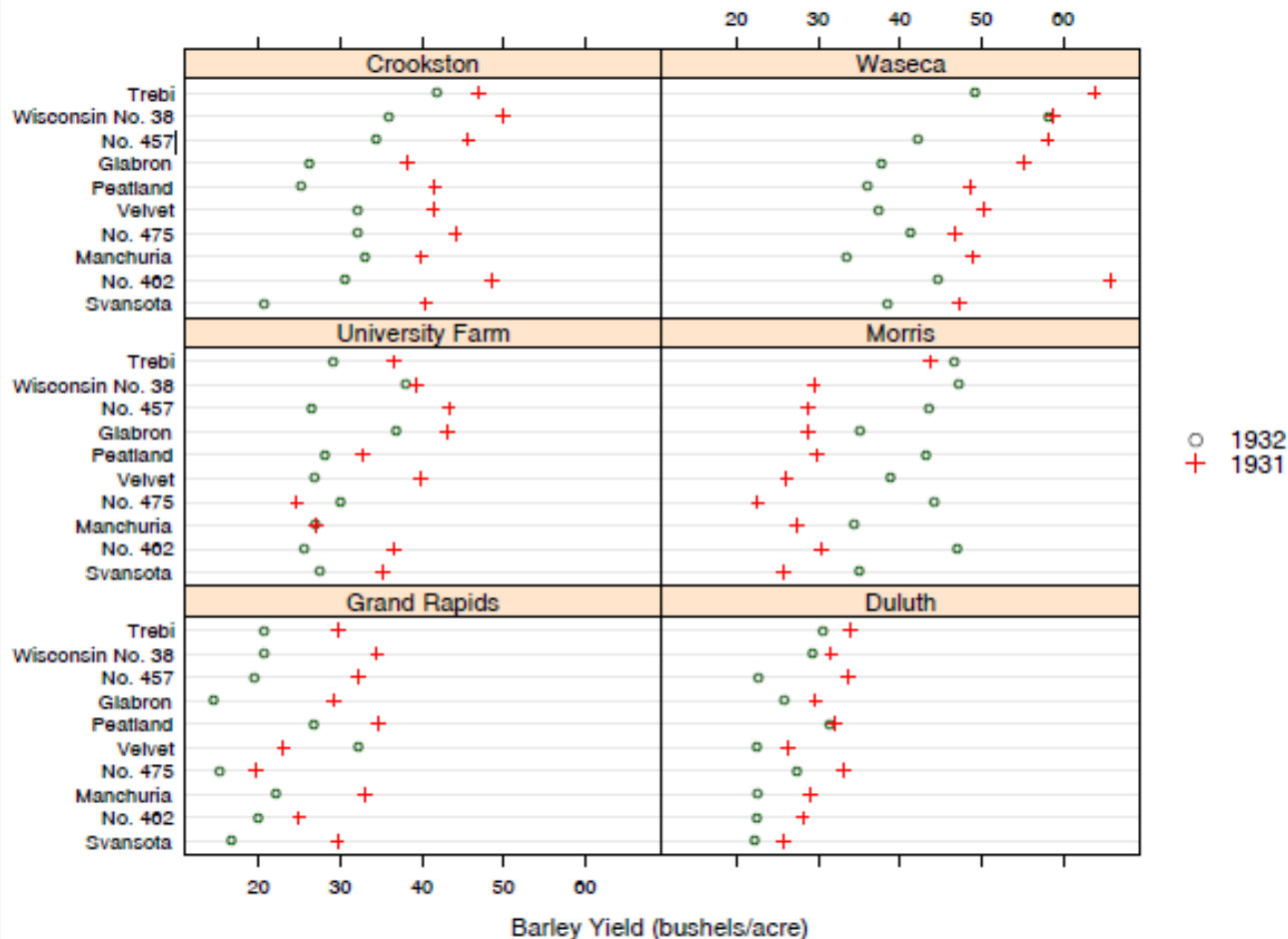
성별 합격율

1973년도 버클리 대학원의 6개 단과대학별 성별 합격여부 데이터

```
> apply(UCBAdmissions, c(1, 2), sum)
          Gender
Admit      Male   Female
Admitted   1198     557
Rejected   1493    1278


> prop.table(apply(UCBAdmissions, c
          Gender
Admit      Male        Female
Admitted   0.4451877   0.3035422
Rejected   0.5548123   0.6964578
```
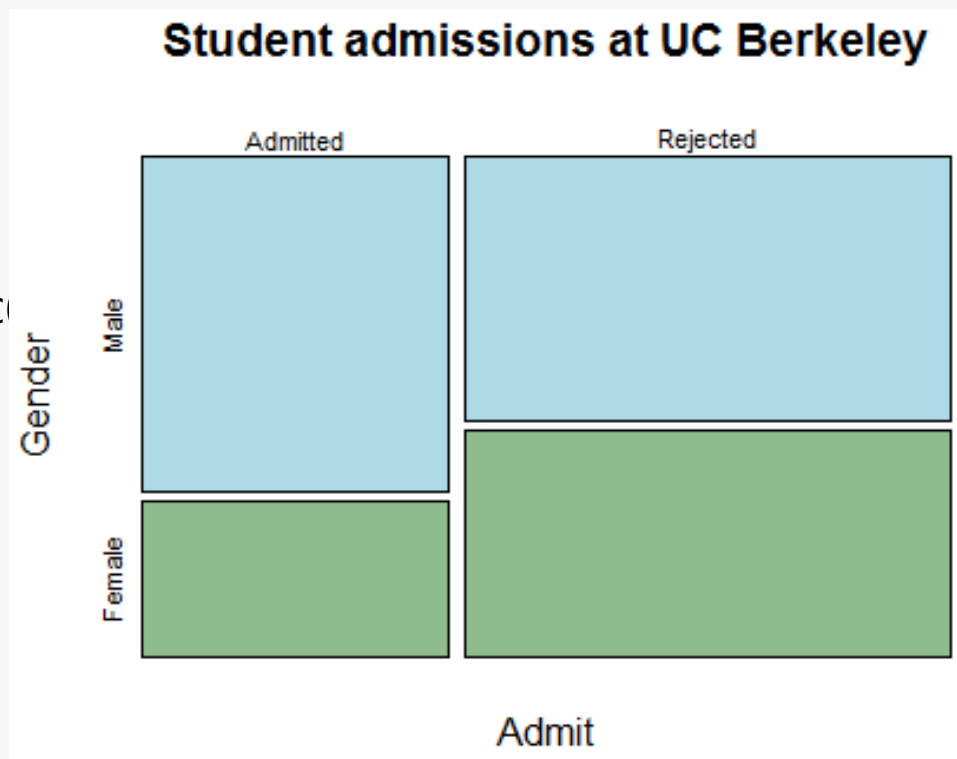

Student admissions at UC Berkeley

Visualization의 필요성

성별 단과대학별 합격율

Simpson's Paradox

R Graphics

R graphics architectures
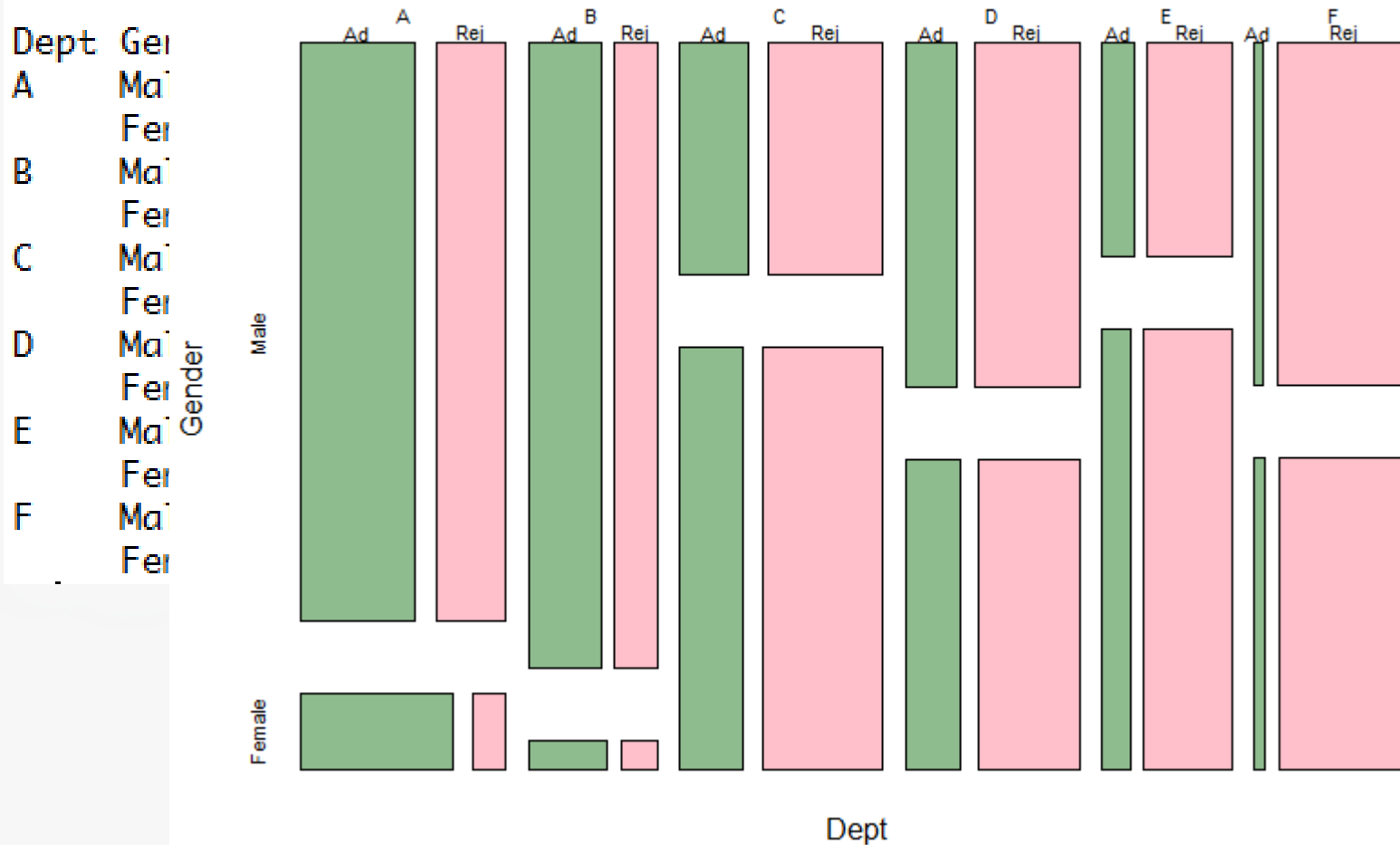
Hierarchical architecture

R Base

High Level Plots

plot, barplot, boxplot, pie, ggplot, lattice, rgl, sna, wordcloud, ...

Low Level Plots

points, lines, box, rect, polygon, text, title, mtext, legend, axis, grid

Graphics Devices

bmp, jpeg, png, tiff
pdf, postscript

External Application interface

GoogleMap, Archview, GoogleVis,

External Graphics Devices

Cairo, SVG, OpenGL, ...

Variety Graph Devices
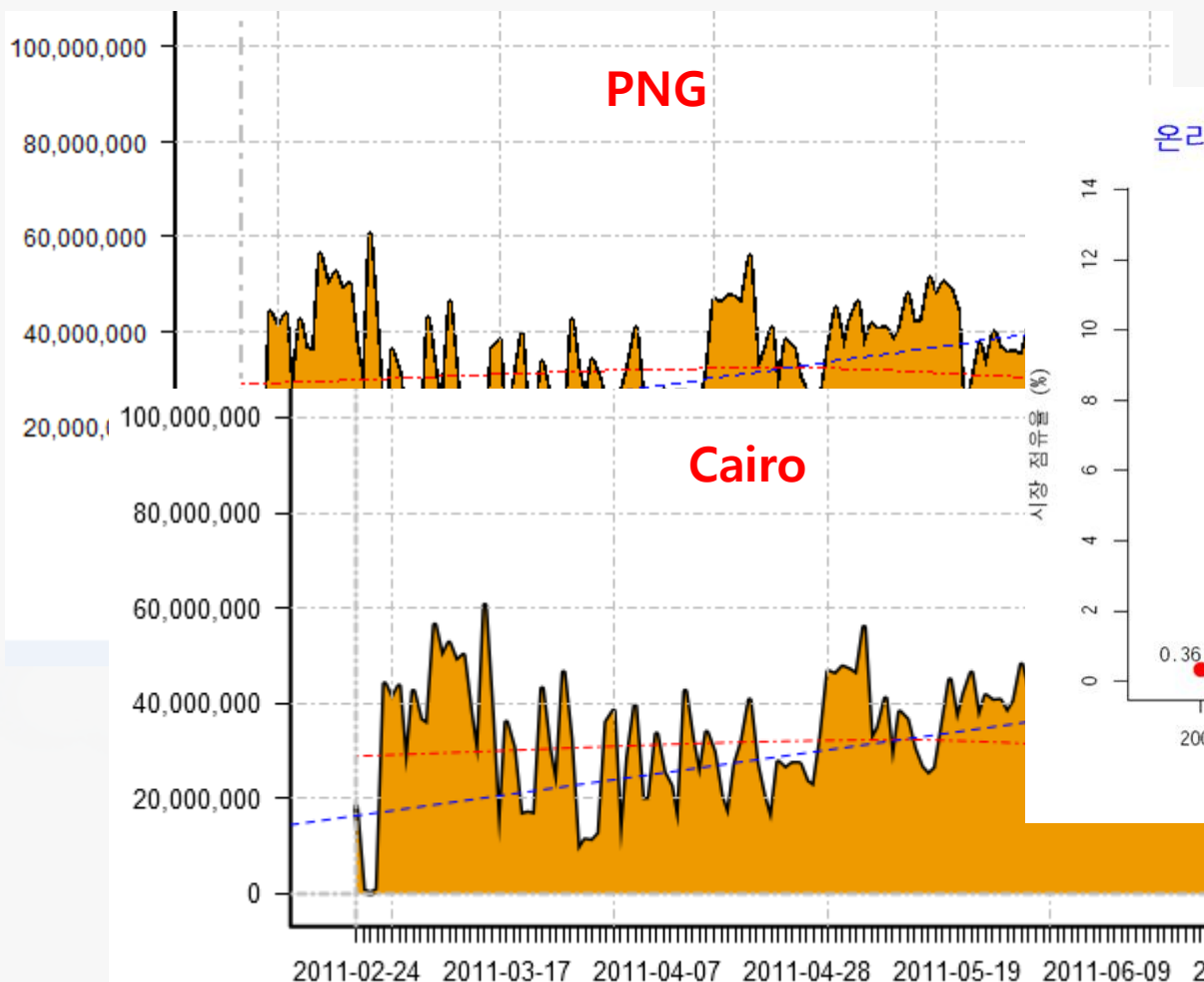
PNG

PDF

Cairo

온라인 자동차보험 년도별 시장점유율
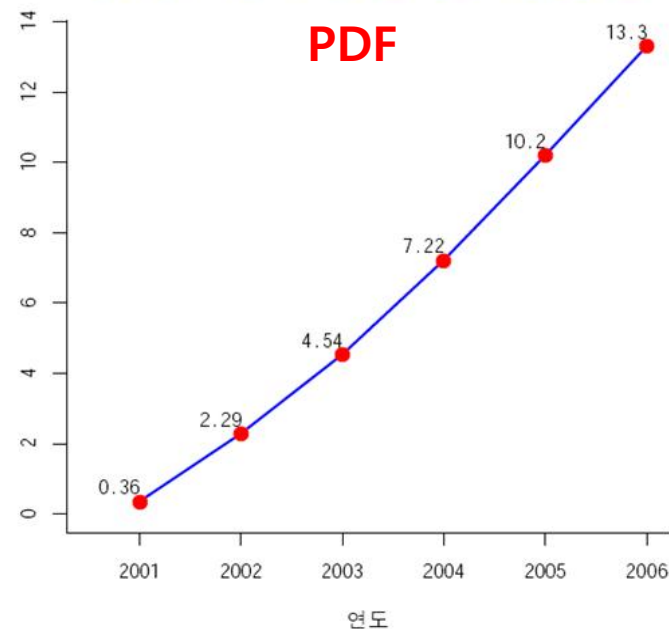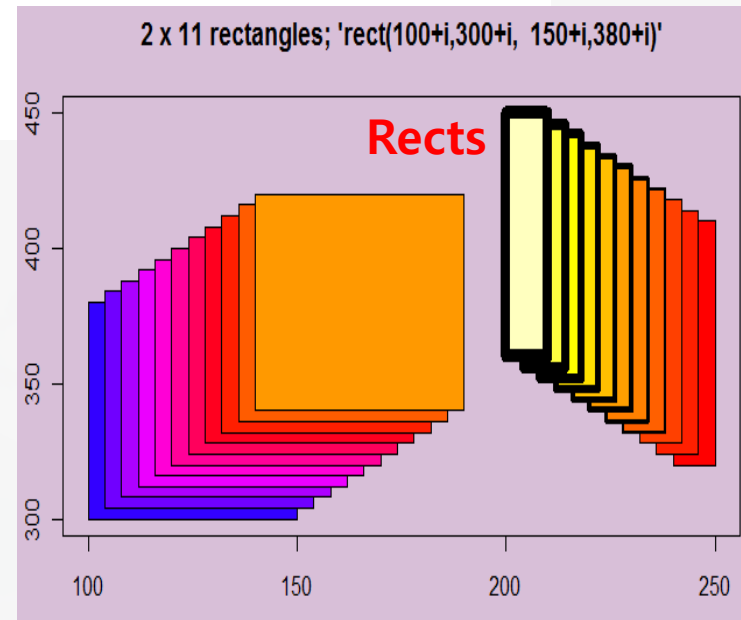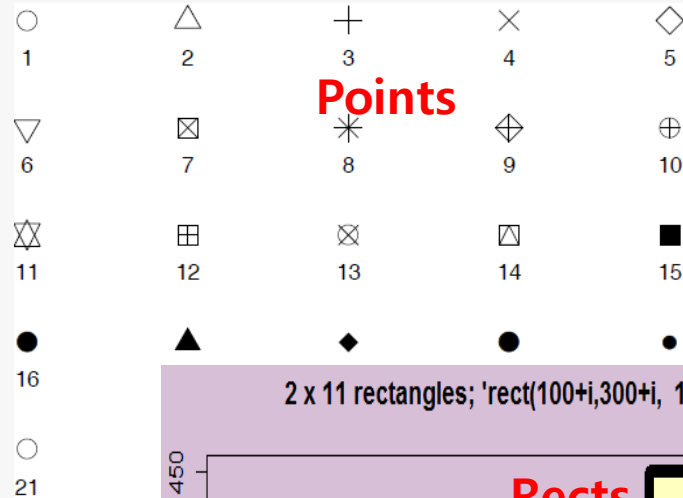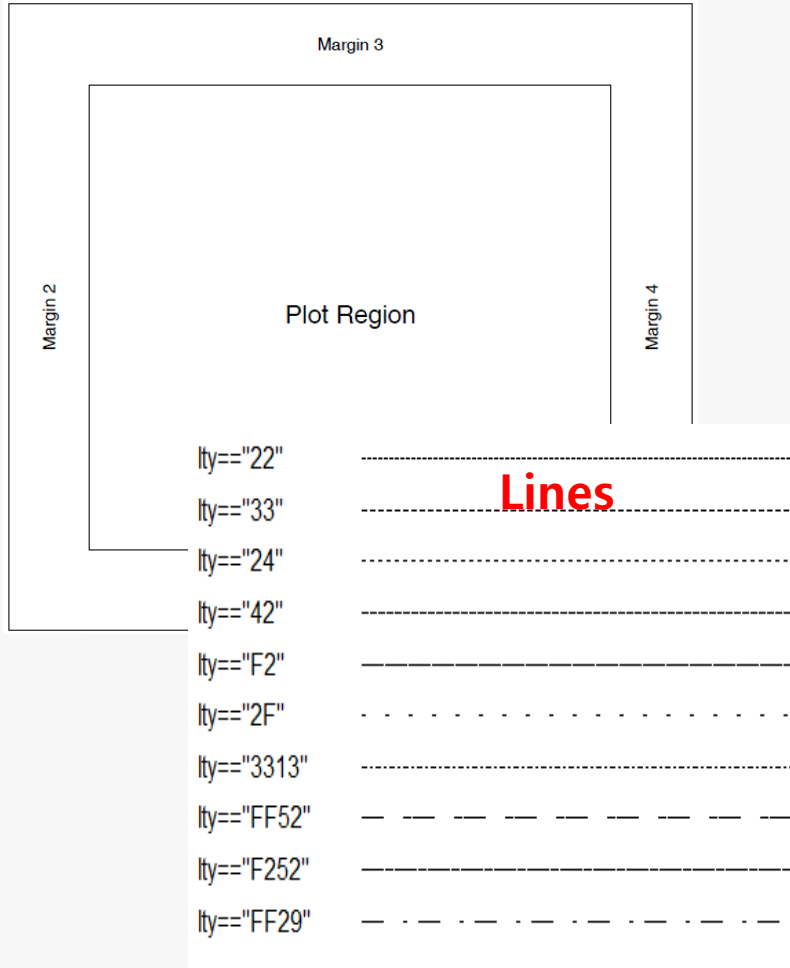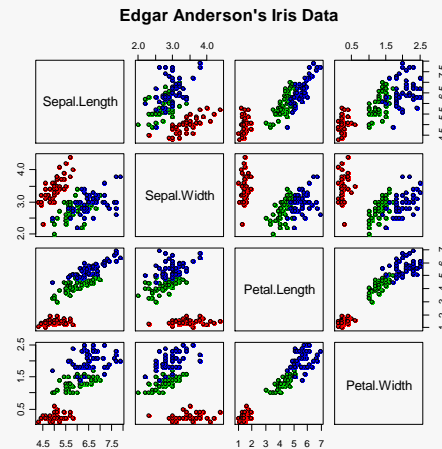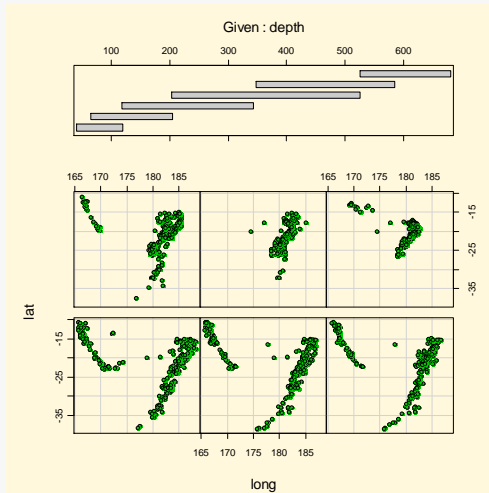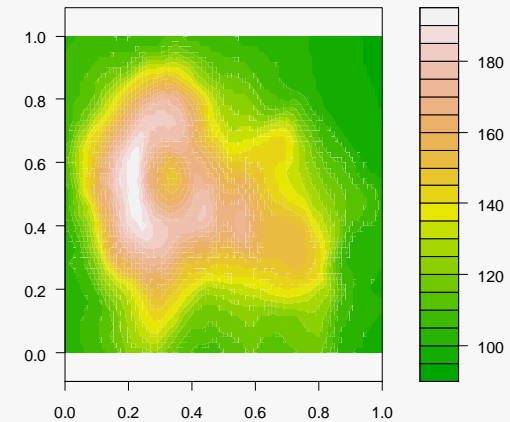
R Graphics

Atomic Plots

Margin 3

Margin 2

Plot Region

Margin 4

Points

Lines

lty=="22"
lty=="33"
lty=="24"
lty=="42"
lty=="F2"
lty=="2F"
lty=="3313"
lty=="FF52"
lty=="F252"
lty=="FF29"

Rects

2 x 11 rectangles; 'rect(100+i,300+i, 150+i,380+i)'

R Graphics

Variety Graph

유용한 R Graphics

lattice

RUGK · R User Conference 2011 · -13-
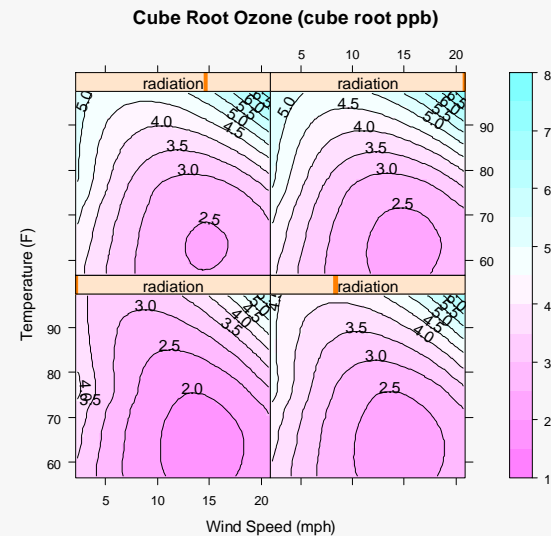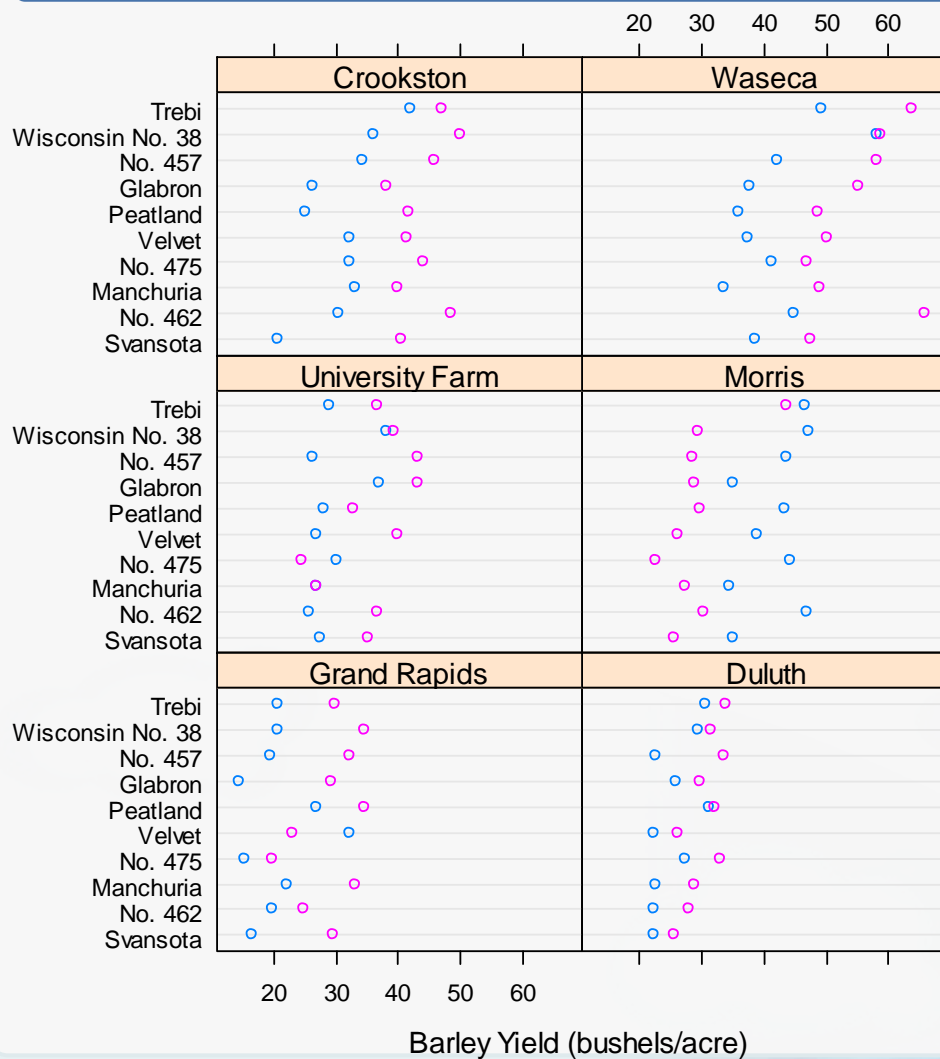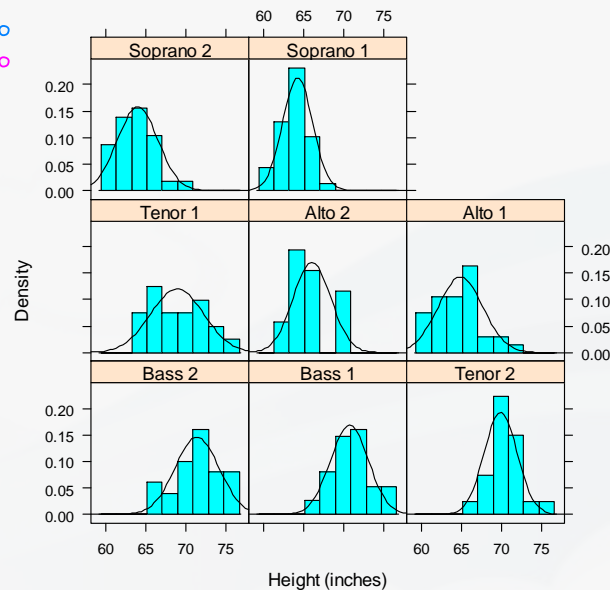
유용한 R Graphics

Social Network Analysis

R Journal 2011-1의 "Content-Based Social Network Analysis of Mailing Lists" 인용
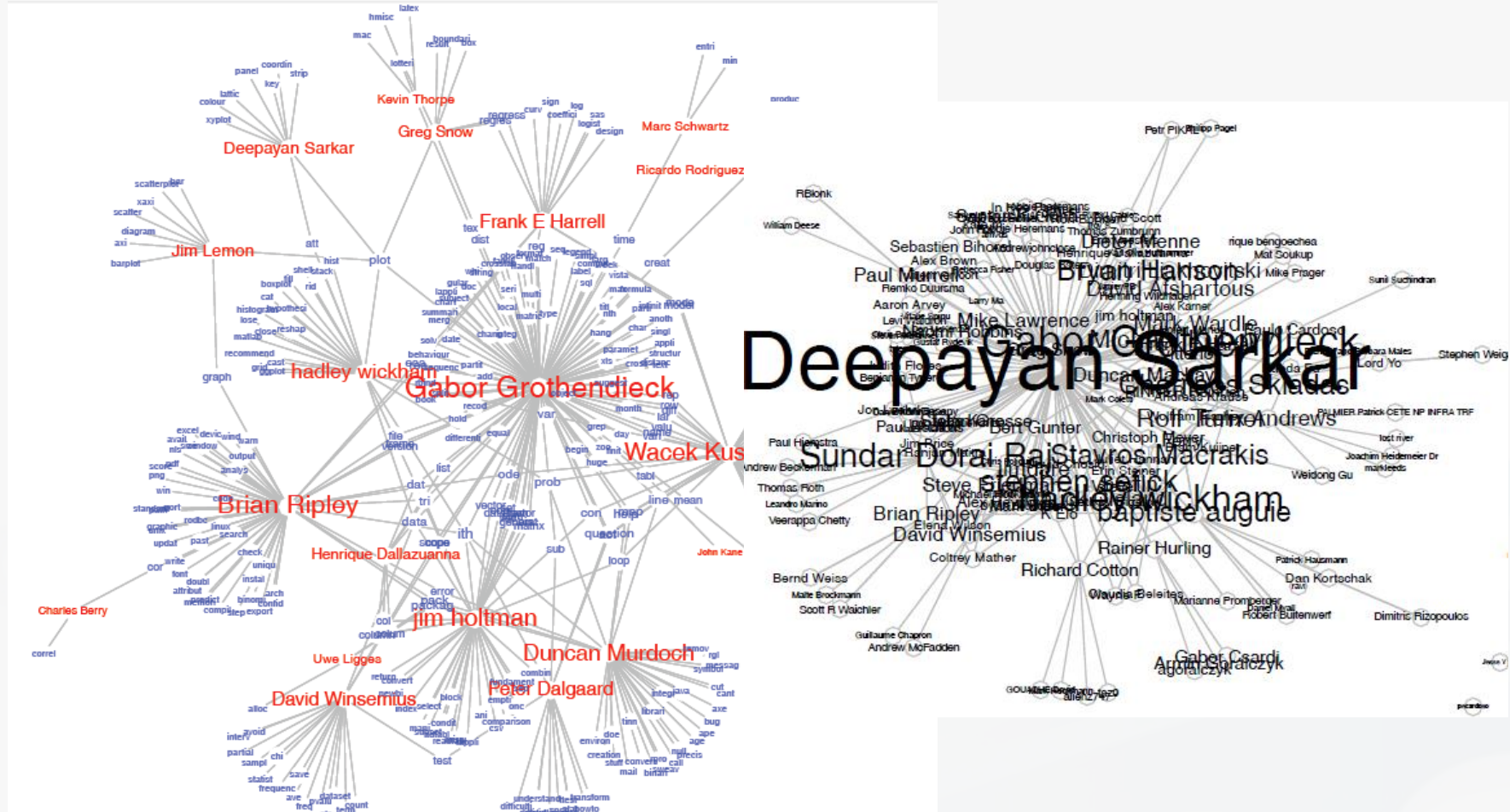
유용한 R Graphics

word cloud chart
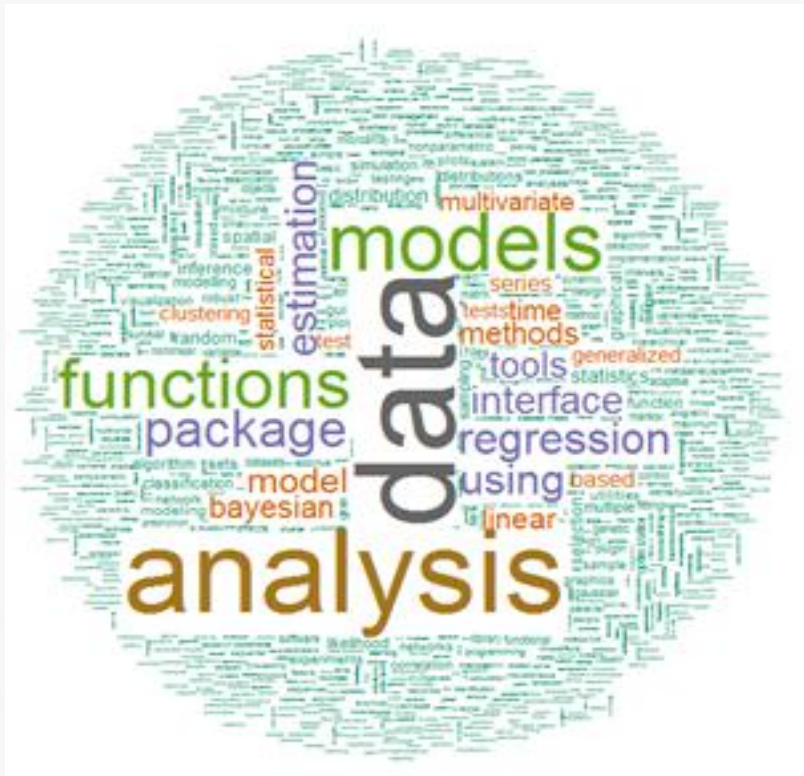
R Mailing List

R User Conference Survey

유용한 R Graphics

주제도

rworldmap

Happy Planet Index

HPI colour
- 2 good, 1 middle
- 1 good, 2 middle
- 3 middle
- 1 poor
- 2 poor or footprint v.poor
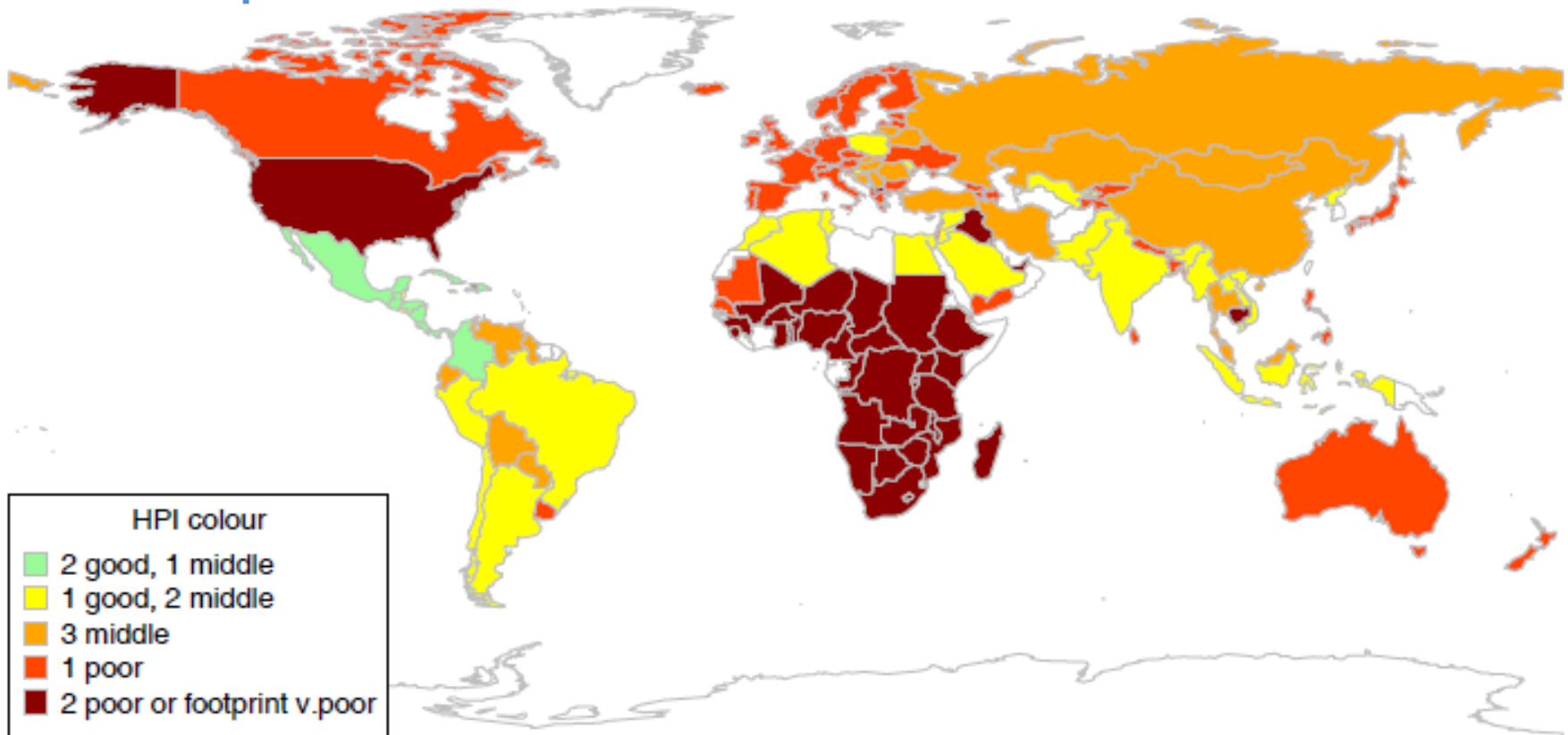
Interactive R Graphics

OpenGL interface
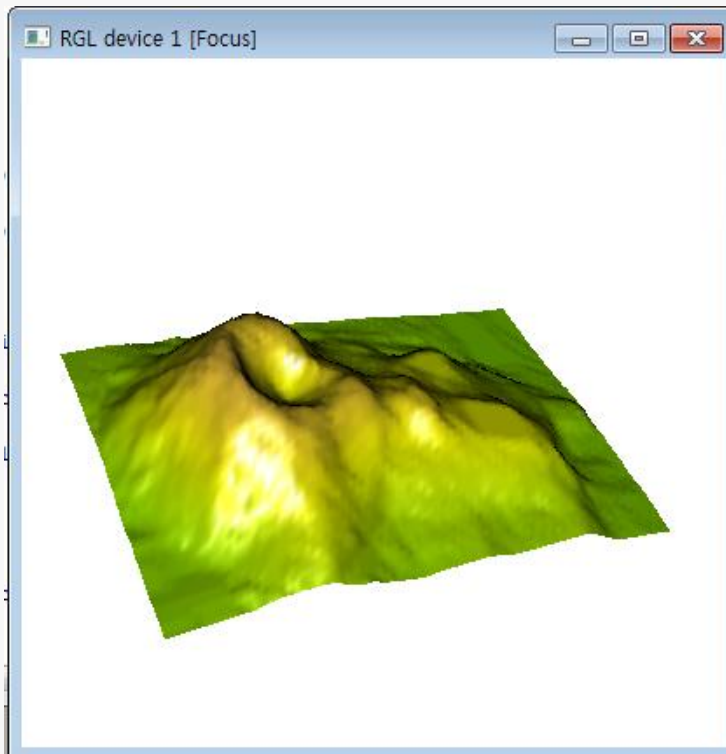
rgl

scatter3d
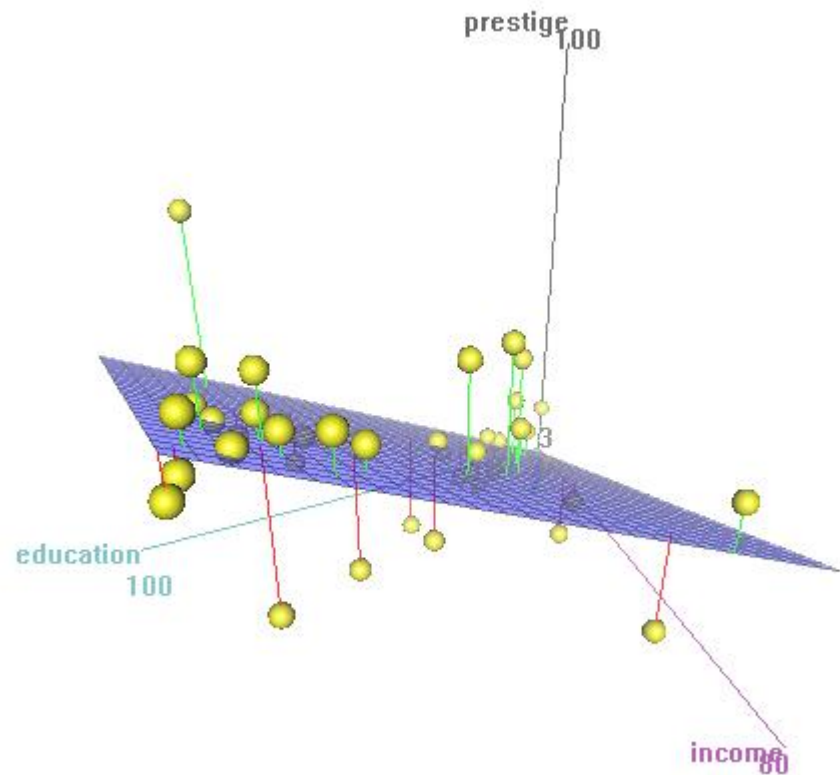
Interactive R Graphics

interactive plotting

Dynamic/Interactive Graph
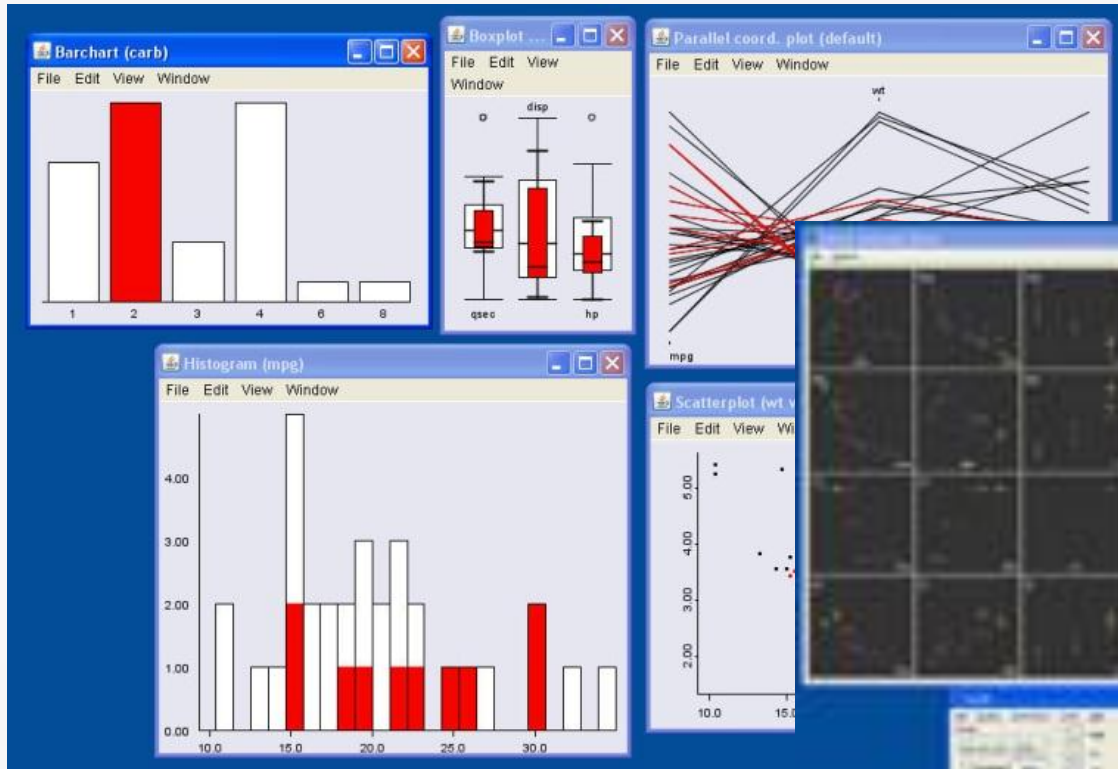
iplots
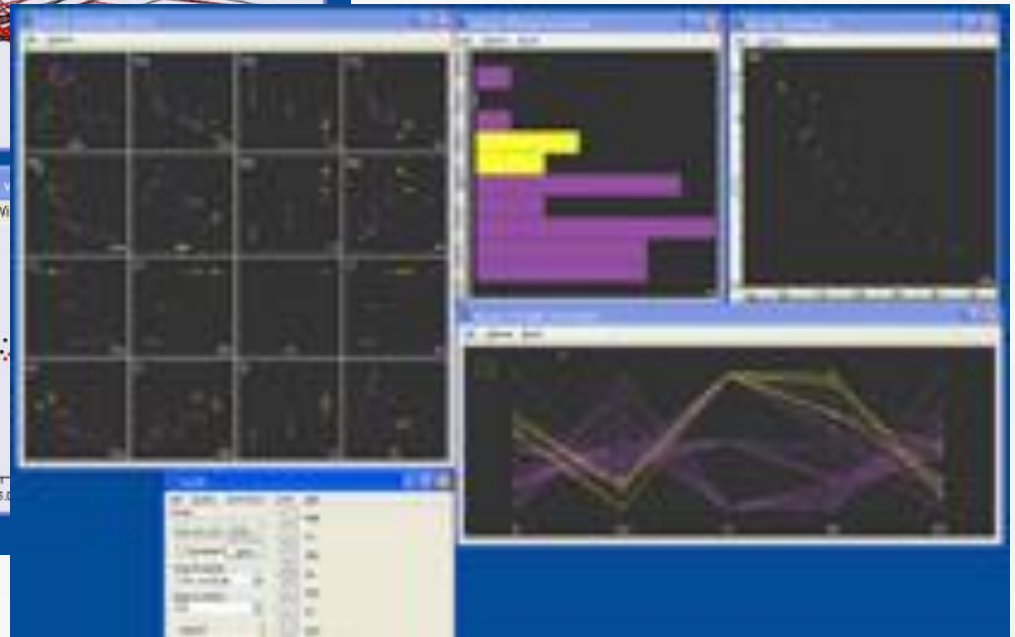
rggobi

Interface External Applications
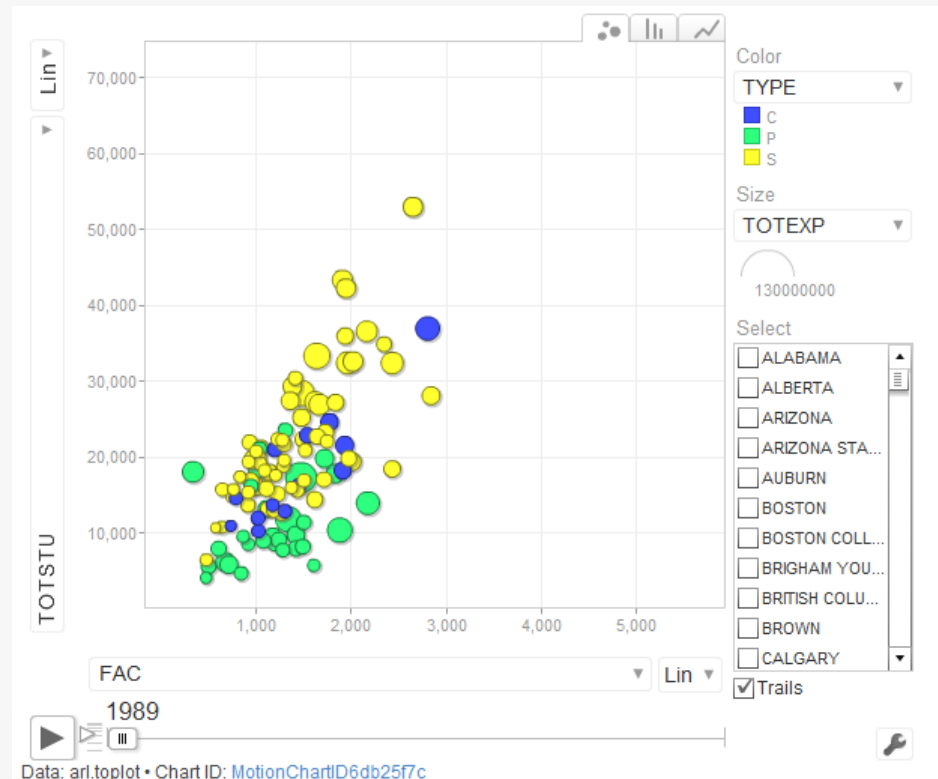
Google Maps

Google Maps & Google Earth

rcom

RGoogleMaps

RKML
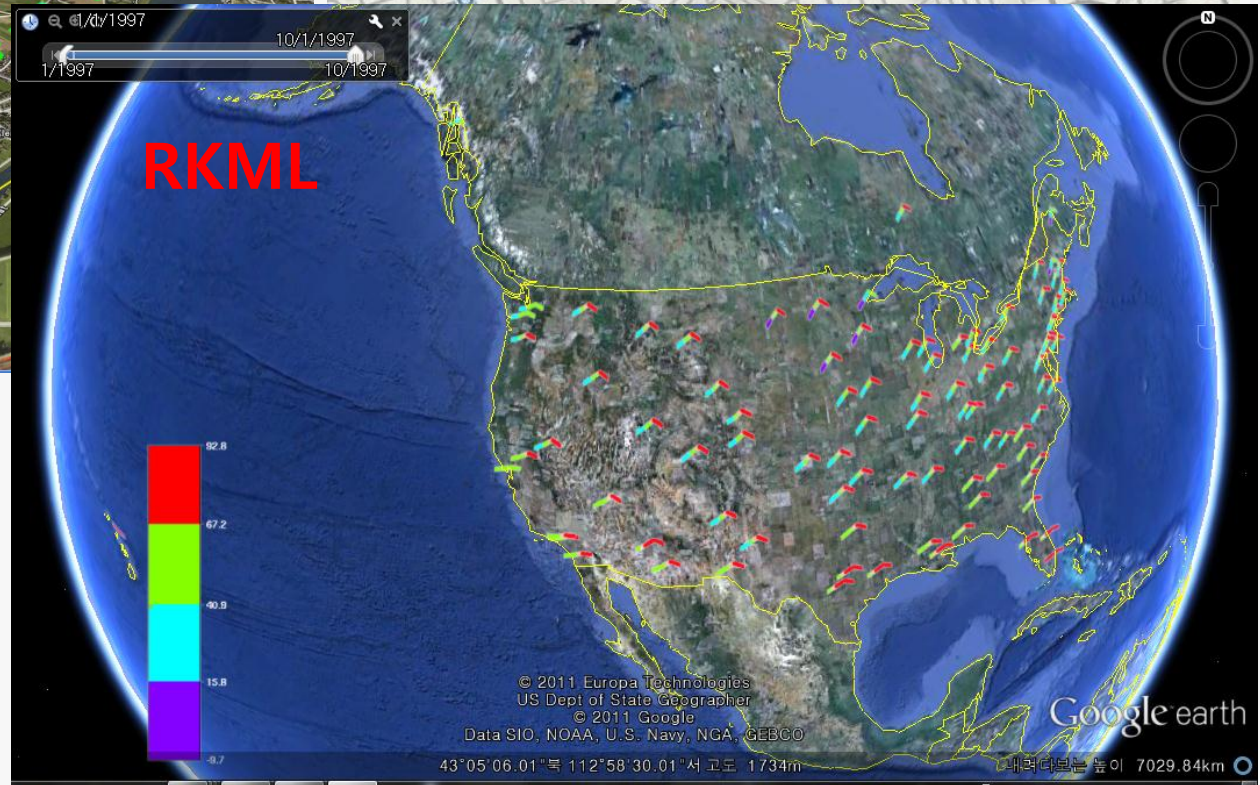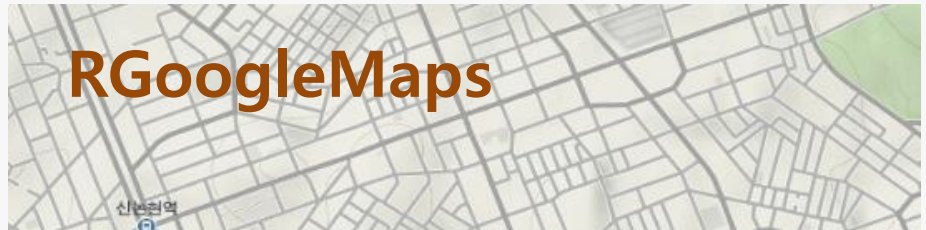
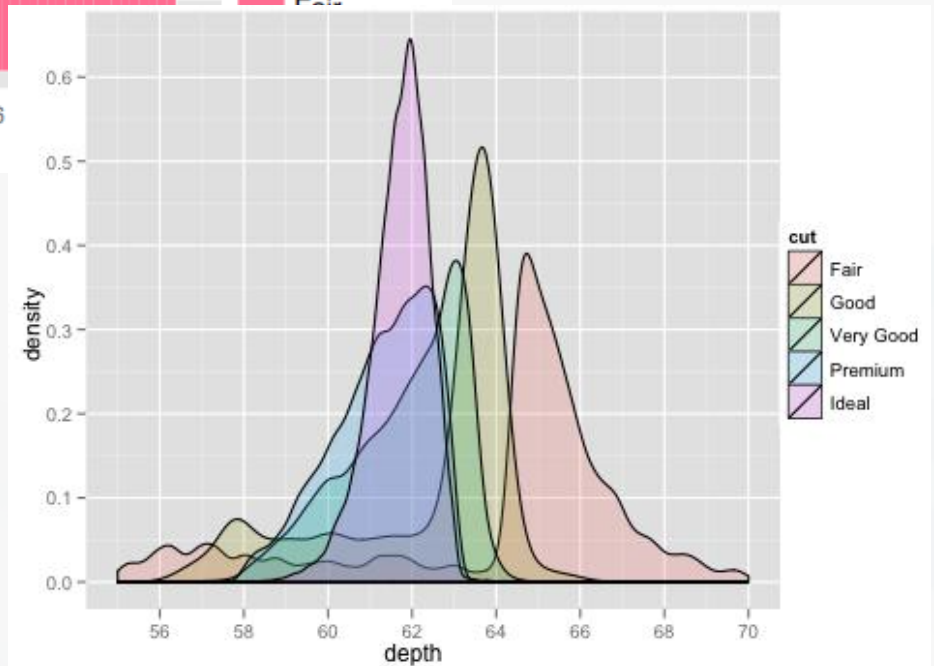# Interface External Applications
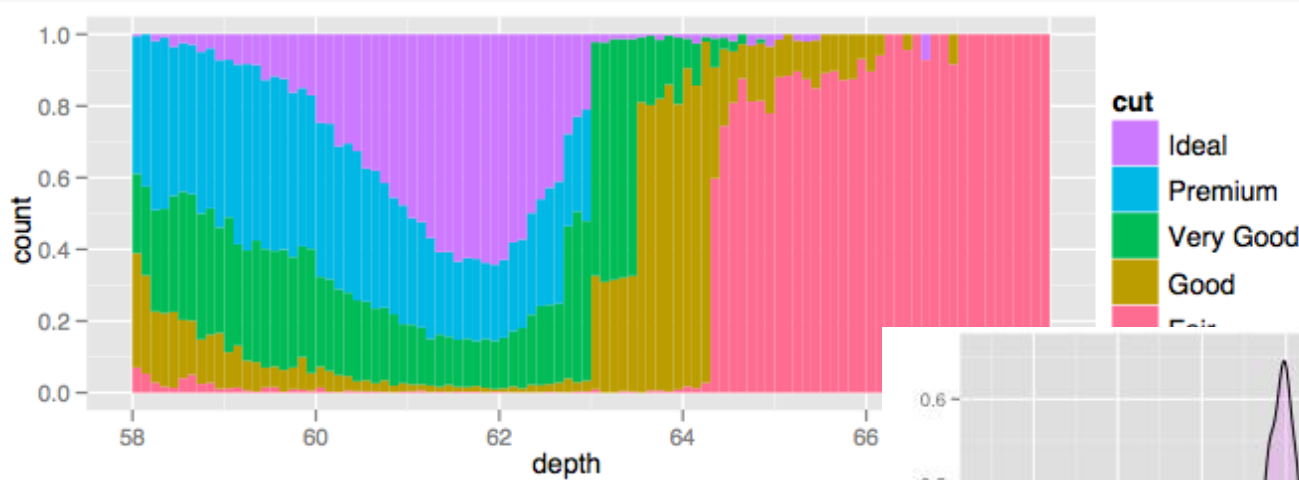
Google Visualization Interface

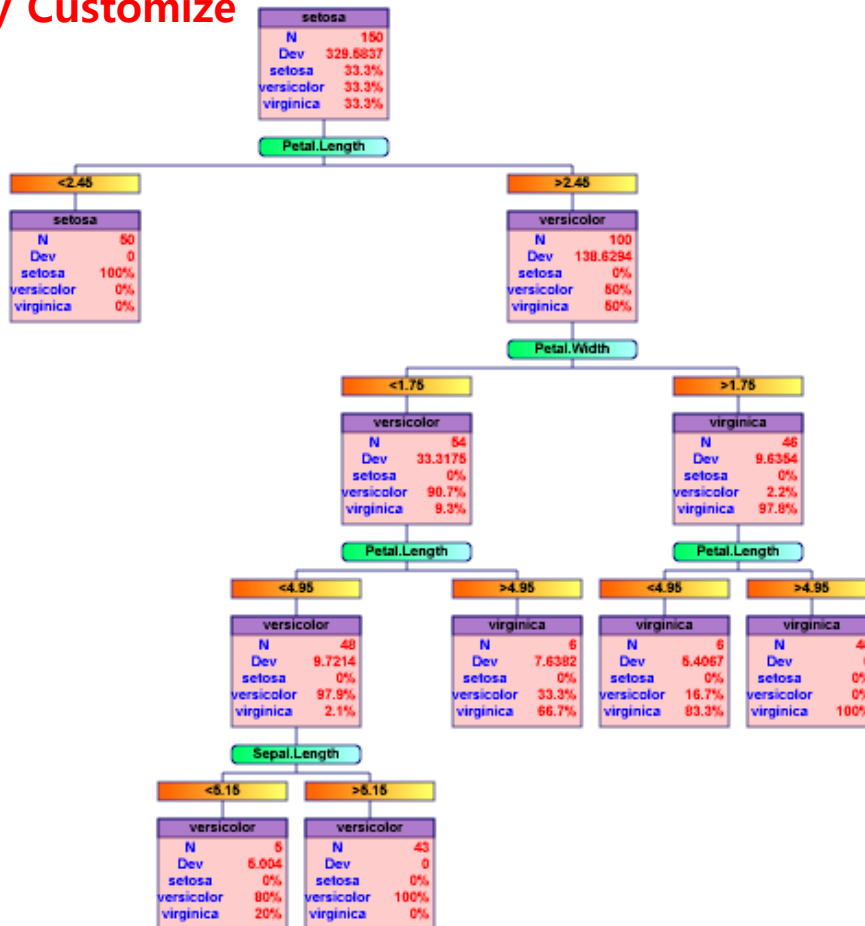
gvisGeoMap

gvisMotionChart

Another R Graphics

ggplot2

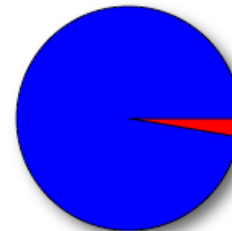RUGK   R User Conference 2011   -21-

Another R Graphics
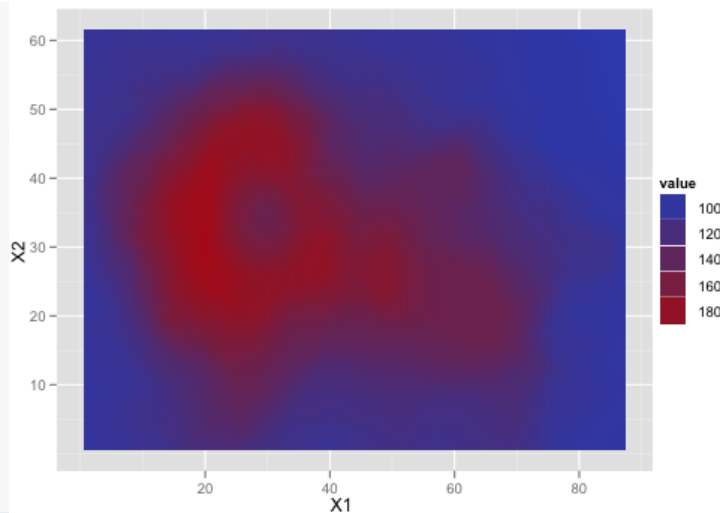
My Customizing Chart

My Customize

감사합니다

antony.ryu@nexr.com