R 与统计图形

刘思喆、谢益辉

第三届中国 R 语言会议 @ 上海财经大学

2010 年 11 月 14 日

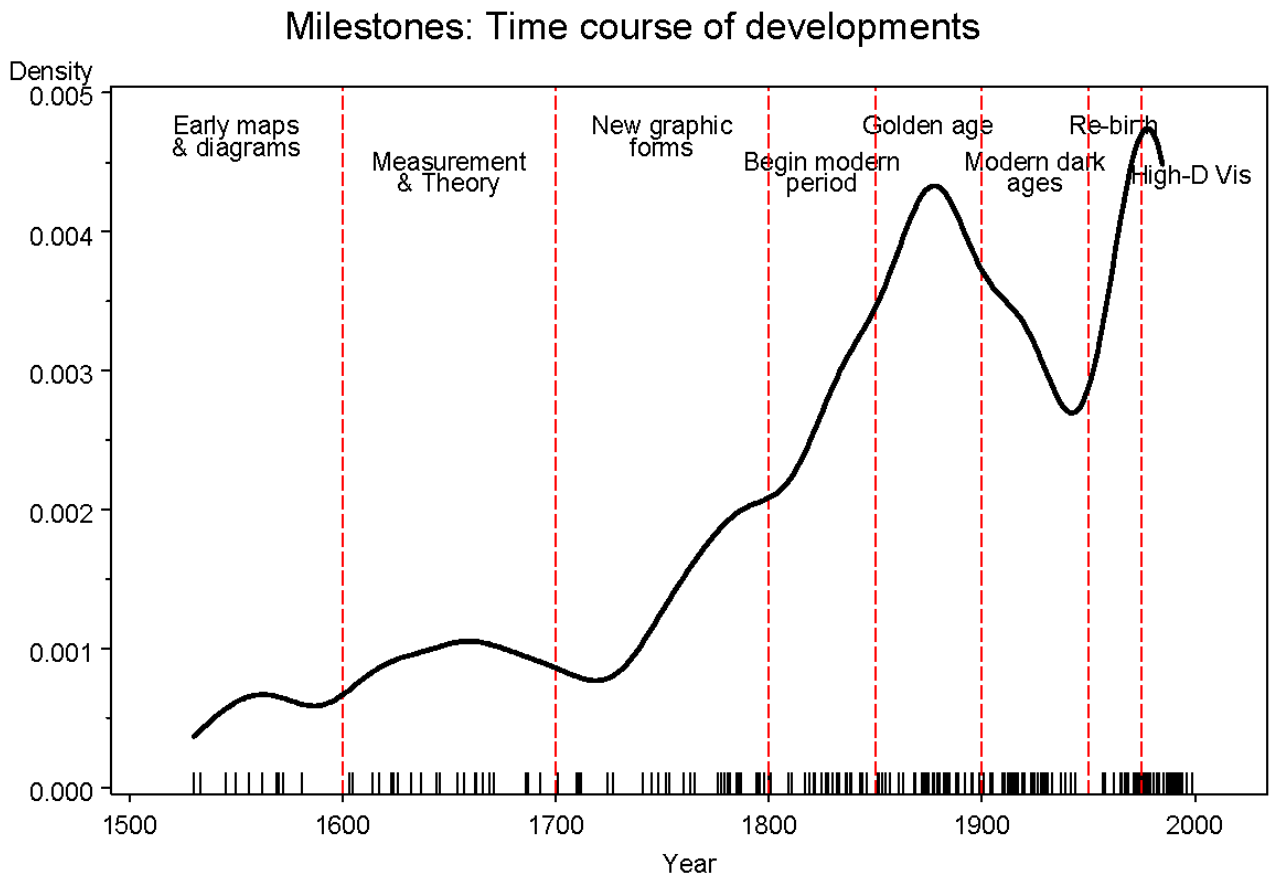
目录

1 历史

图形的各个历史时期：

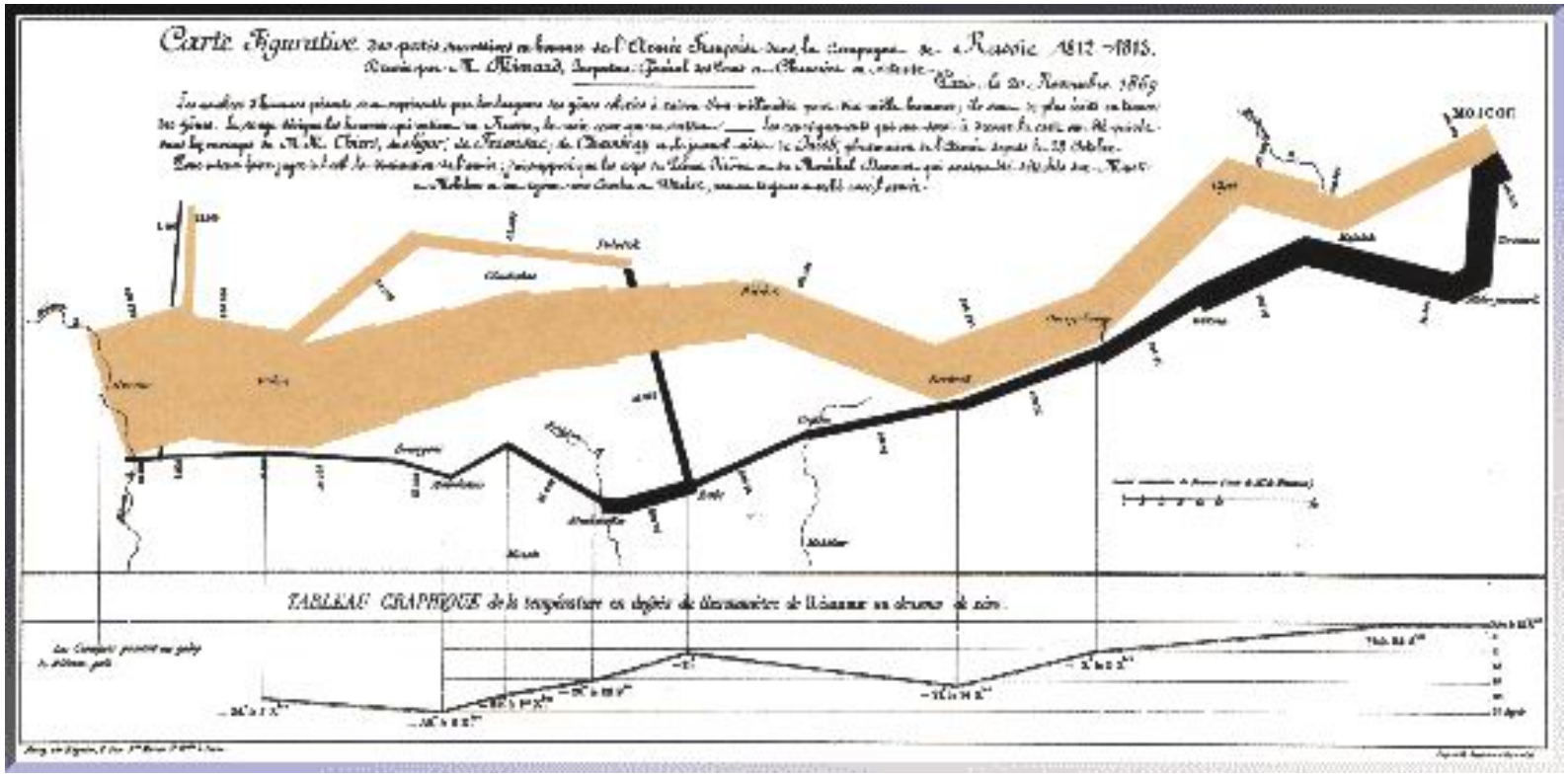
- Pre - 17th Century: Early maps and diagrams
- 1600 - 1699: Measurement and theory
- 1700 - 1799: New graphic forms
- 1800 - 1850: Beginnings of modern graphics
- 1850 - 1900: The Golden Age of statistical graphics
- 1900 - 1950: The modern dark ages
- 1950 - 1975: Re-birth of data visualization
- 1975 - present: High-D, interactive and dynamic data visualization

19 世纪后期是数据可视化的黄金时期；20 世纪中叶以后，随着计算机技术的发展，可视化技术又一次重生



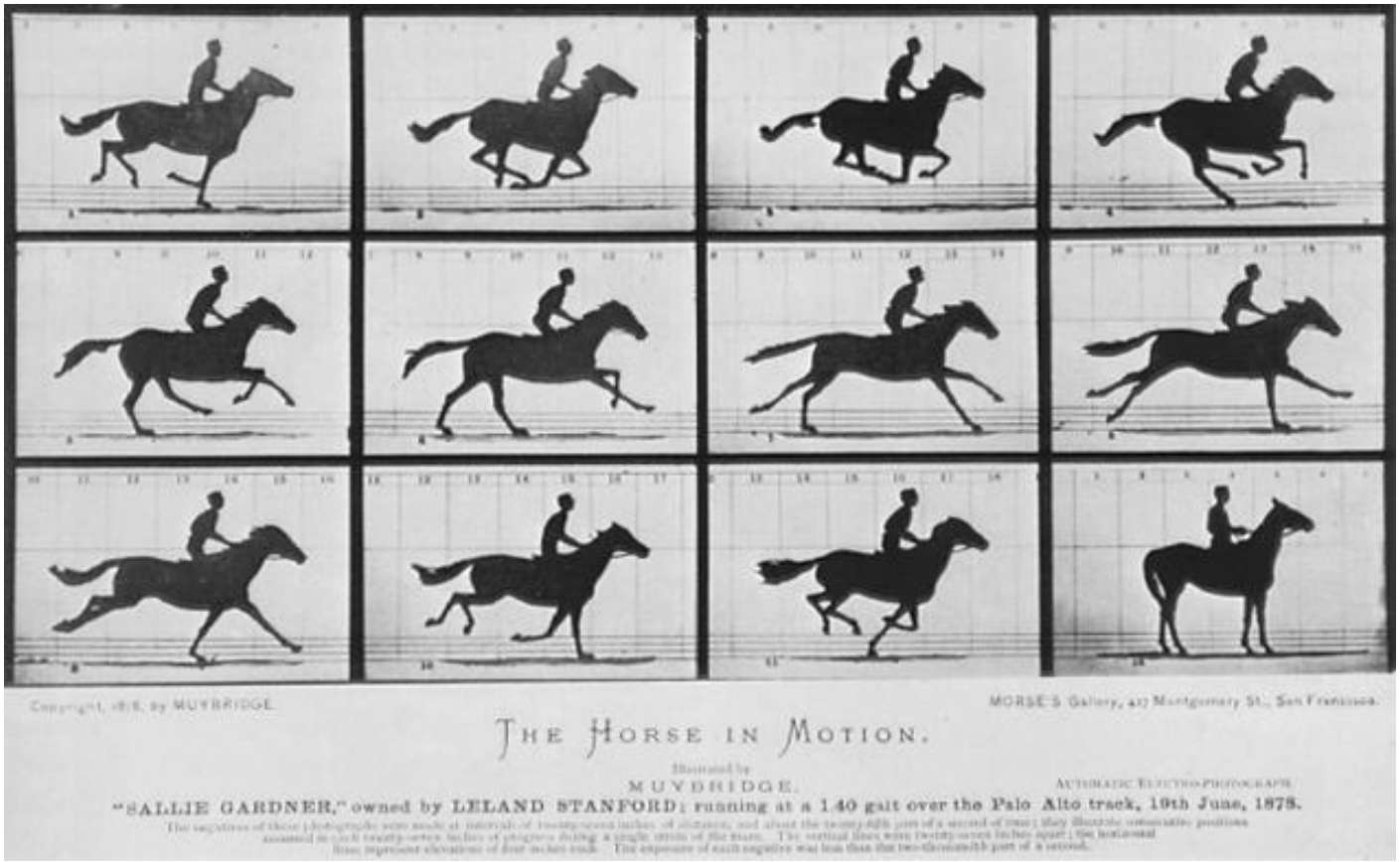
数据可视化历史中，各个时期里程碑事件的分布，由密度估计和坐标须展示。
来源：Handbook of Computational Statistics: Data Visualization(2006)

绘图历史上最好的图形 (1869)：拿破仑远征俄国



The French engineer, Charles Minardi (1781-1870), illustrated graphically the disastrous campaign of Napoleon against Russia in 1812. The width of the course is proportional to the number of surviving soldiers in the war campaign. In beige for the way in and in black for the return.

Edward Muybridge 的赛马动画 (1877)



来源：http://didyouknow.org/how-a-horse-kicked-off-the-movie-industry/

2 输出和设备

- 输出
- 设备

图形输出

- 当用户发出绘图命令后，R 会自动弹出图形窗口
- 图形窗口上，可选择 save 的文件类型
- savePlot

R 基础图形设备

R 图形设备

	名称	描述
屏幕显示	x11	X 窗口
	windows	Windows 窗口
文件设备	postscript	ps 格式文件
	pdf	pdf 格式文件
	pictex	供 LaTeX使用的文件
	png	png 格式文件
	jpeg	jpeg 格式文件
	bmp	bmp 格式文件
	xfig	供 XFIG 使用的图形格式
	win.metafile [a]	emf 格式的文件

[a]仅在 Windows 下有效

R 同时支持 Cairo 图形设备，这种图形设备可以支持高质量的矢量绘图 (PDF,PostScript and SVG)，高质量的点阵绘图 (PNG,JPEG,TIFF) 或屏幕输出。

3 一般性统计绘图

- 高级
- 低级
- 参数和颜色

plot 类函数

语法：plot(x, y, ...)

```
 [1] plot.acf*       plot.data.frame*  plot.decomposed.ts*
 [4] plot.default    plot.dendrogram*  plot.density
 [7] plot.ecdf       plot.factor*      plot.formula*
[10] plot.hclust*    plot.histogram*   plot.HoltWinters*
[13] plot.isoreg*    plot.lm           plot.medpolish*
[16] plot.mlm        plot.ppr*         plot.prcomp*
[19] plot.princomp*  plot.profile.nls* plot.spec
[22] plot.spec.phase plot.stepfun      plot.TukeyHSD
[25] plot.stl*       plot.table*       plot.ts
```

其他统计绘图

统计绘图函数

boxplot, bxp, cdplot,

contours, filled.contour,

fourfoldplot, hist, image,

matplot, mosaicplot, persp,

spineplot, symbols, stripchart, ···

参数

add = TRUE

axes = TRUE

元素

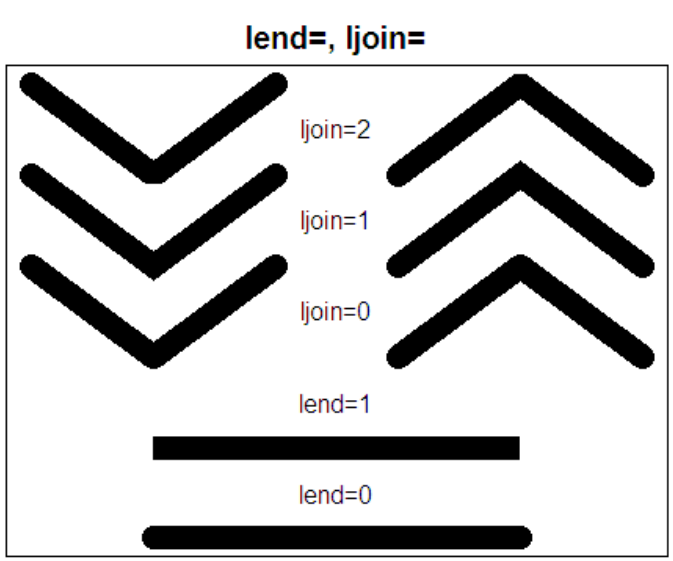
- 点 (points)
- 线 (lines, abline, segments, arrows)
- 多边形 (rect, polygon, box)

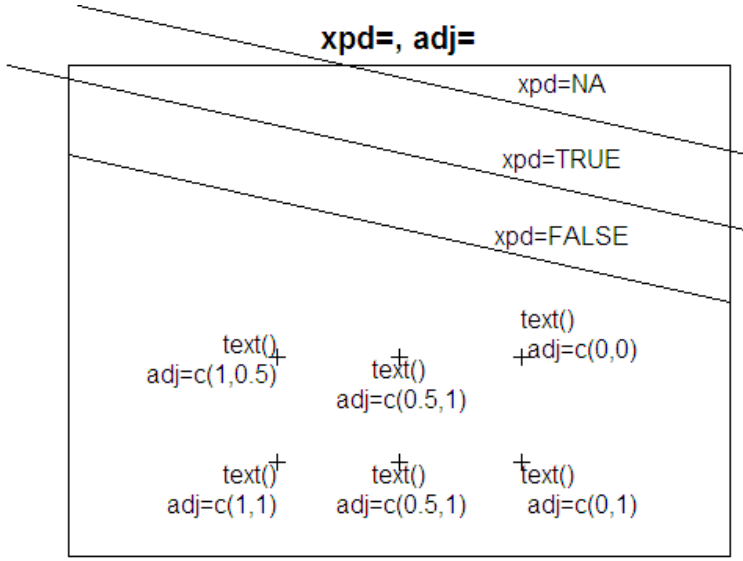
辅助

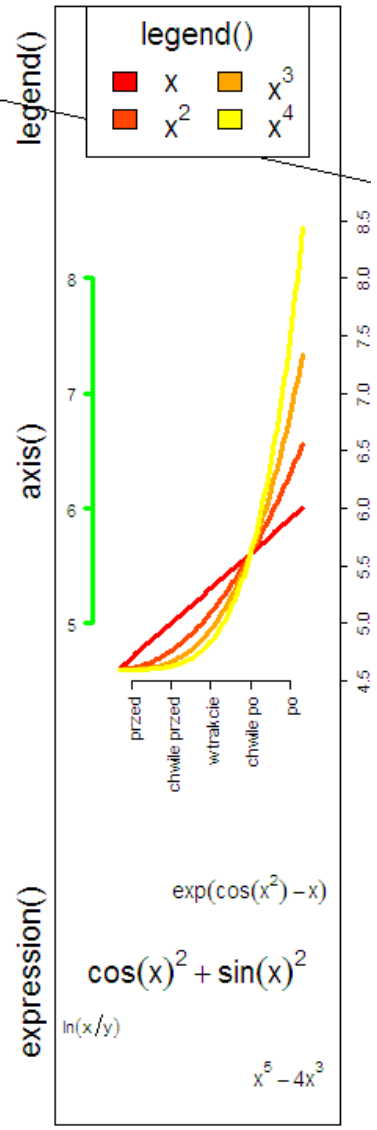
- 坐标轴 (axis)
- 文本 (title, legend, text, mtext)
- 公式及符号 (expression)

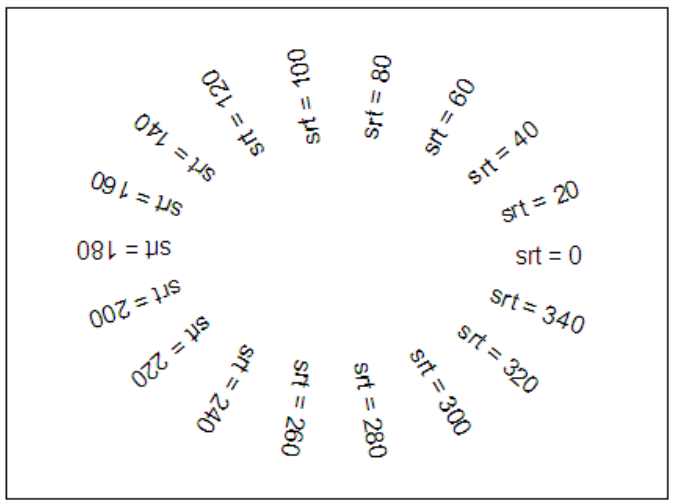
par

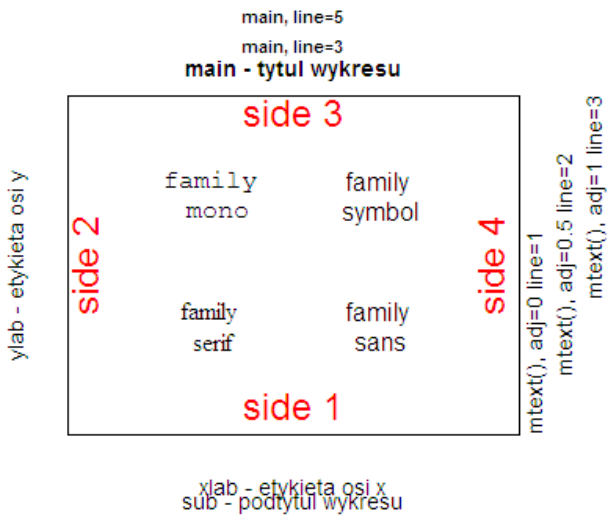
- par 是魔鬼 (72 个参数，常用的也达到近 20 个)
- 是绘图的必要条件 (除非你想忍受 R 的默认绘图状态)


lend=, ljoin=
ljoin=2
ljoin=1
ljoin=0
lend=1
lend=0


xpd=, adj=
xpd=NA
xpd=TRUE
xpd=FALSE
text()
adj=c(1,0.5)
text()
adj=c(0.5,1)
text()
adj=c(0,0)
text()
adj=c(1,1)
text()
adj=c(0.5,1)
text()
adj=c(0,1)


legend()
x
x^2
x^3
x^4
legend()
axis()
expression()
przed
chwile przed
w trakcie
chwile po
po
exp(cos(x^2) − x)
cos(x)^2 + sin(x)^2
ln(x/y)
x^5 − 4x^3

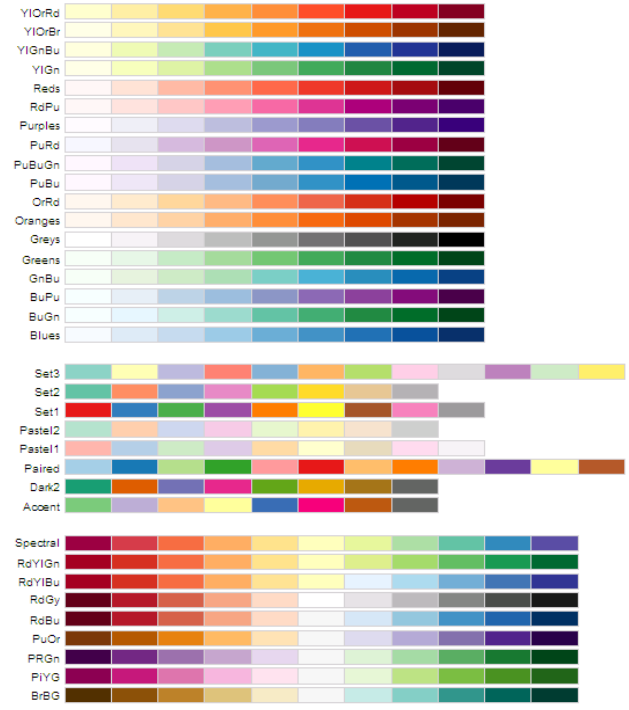

srt=
srt = 0
srt = 20
srt = 40
srt = 60
srt = 80
srt = 100
srt = 120
srt = 140
srt = 160
srt = 180
srt = 200
srt = 220
srt = 240
srt = 260
srt = 280
srt = 300
srt = 320
srt = 340


main, line=5
main, line=3
main - tytul wykresu
side 3
family
mono
family
symbol
side 2
side 4
family
serif
family
sans
side 1
ylab - etykieta osi y
mtext(), adj=0 line=1
mtext(), adj=0.5 line=2
mtext(), adj=1 line=3
xlab - etykieta osi x
sub - podtytuł wykresu

颜色

- *colors()*
- 颜色参数的传递
  - 颜色名称，如 red
  - 整数，对应当前调色板 *palette()*
  - 16 进制的三原色，如 #FF0000
- 特定主题调色板，如 *heat.colors()*
- 颜色的扩展包，如 `RColorBrewer`、`colorRamps`

Color names for colors from the RColorBrewer package



Color names for grDevices and colorRamps packages

Color names for colors(grDevices)

coral3	deeppink4	gray27	gray87	grey39	grey99	lightpink1	mistyrose1	pink4	slategray1	
coral2	deeppink3	gray26	gray86	grey38	grey98	lightpink	mistyrose	pink3	slategray	
coral1	deeppink2	gray25	gray85	grey37	grey97	lightgrey	mintcream	pink2	slateblue4	
coral	deeppink1	gray24	gray84	grey36	grey96	lightgreen	midnightblue	pink1	slateblue3	yellowgreen
chocolate4	deeppink	gray23	gray83	grey35	grey95	lightgray	mediumvioletred	pink	slateblue2	yellow4
chocolate3	darkviolet	gray22	gray82	grey34	grey94	lightgoldenrodyello	mediumturquoise	peru	slateblue1	yellow3
chocolate2	darkturquoise	gray21	gray81	grey33	grey93	lightgoldenrod4	mediumspringgree	peachpuff4	slateblue	yellow2
chocolate1	darkslategrey	gray20	gray80	grey32	grey92	lightgoldenrod3	mediumslateblue	peachpuff3	skyblue4	yellow1
chocolate	darkslategray4	gray19	gray79	grey31	grey91	lightgoldenrod2	mediumseagreen	peachpuff2	skyblue3	yellow
chartreuse4	darkslategray3	gray18	gray78	grey30	grey90	lightgoldenrod1	mediumpurple4	peachpuff1	skyblue2	whitesmoke
chartreuse3	darkslategray2	gray17	gray77	grey29	grey89	lightgoldenrod	mediumpurple3	peachpuff	skyblue1	wheat4
chartreuse2	darkslategray1	gray16	gray76	grey28	grey88	lightcyan4	mediumpurple2	papayawhip	skyblue	wheat3
chartreuse1	darkslategray	gray15	gray75	grey27	grey87	lightcyan3	mediumpurple1	palevioletred4	sienna4	wheat2
chartreuse	darkslateblue	gray14	gray74	grey26	grey86	lightcyan2	mediumpurple	palevioletred3	sienna3	wheat1
cadetblue4	darkseagreen4	gray13	gray73	grey25	grey85	lightcyan1	mediumorchid4	palevioletred2	sienna2	wheat
cadetblue3	darkseagreen3	gray12	gray72	grey24	grey84	lightcyan	mediumorchid3	palevioletred1	sienna1	violetred4
cadetblue2	darkseagreen2	gray11	gray71	grey23	grey83	lightcoral	mediumorchid2	palevioletred	sienna	violetred3
cadetblue1	darkseagreen1	gray10	gray70	grey22	grey82	lightblue4	mediumorchid1	paleturquoise4	seashell4	violetred2
cadetblue	darkseagreen	gray9	gray69	grey21	grey81	lightblue3	mediumorchid	paleturquoise3	seashell3	violetred1
burlywood4	darksalmon	gray8	gray68	grey20	grey80	lightblue2	mediumblue	paleturquoise2	seashell2	violetred
burlywood3	darkred	gray7	gray67	grey19	grey79	lightblue1	mediumaquamarin	paleturquoise1	seashell1	violet
burlywood2	darkorchid4	gray6	gray66	grey18	grey78	lightblue	maroon4	paleturquoise	seashell	turquoise4
burlywood1	darkorchid3	gray5	gray65	grey17	grey77	lemonchiffon4	maroon3	palegreen4	seagreen4	turquoise3
burlywood	darkorchid2	gray4	gray64	grey16	grey76	lemonchiffon3	maroon2	palegreen3	seagreen3	turquoise2
brown4	darkorchid1	gray3	gray63	grey15	grey75	lemonchiffon2	maroon1	palegreen2	seagreen2	turquoise1
brown3	darkorchid	gray2	gray62	grey14	grey74	lemonchiffon1	maroon	palegreen1	seagreen1	turquoise
brown2	darkorange4	gray1	gray61	grey13	grey73	lemonchiffon	magenta4	palegreen	seagreen	tomato4
brown1	darkorange3	gray0	gray60	grey12	grey72	lawngreen	magenta3	palegoldenrod	sandybrown	tomato3
brown	darkorange2	gray	gray59	grey11	grey71	lavenderblush4	magenta2	orchid4	salmon4	tomato2
blueviolet	darkorange1	goldenrod4	gray58	grey10	grey70	lavenderblush3	magenta1	orchid3	salmon3	tomato1
blue4	darkorange	goldenrod3	gray57	grey9	grey69	lavenderblush2	magenta	orchid2	salmon2	tomato
blue3	darkolivegreen4	goldenrod2	gray56	grey8	grey68	lavenderblush1	linen	orchid1	salmon1	thistle4
blue2	darkolivegreen3	goldenrod1	gray55	grey7	grey67	lavenderblush	limegreen	orchid	salmon	thistle3
blue1	darkolivegreen2	goldenrod	gray54	grey6	grey66	lavender	lightyellow4	orangered4	saddlebrown	thistle2
blue	darkolivegreen1	gold4	gray53	grey5	grey65	khaki4	lightyellow3	orangered3	royalblue4	thistle1
blanchedalmond	darkolivegreen	gold3	gray52	grey4	grey64	khaki3	lightyellow2	orangered2	royalblue3	thistle
black	darkmagenta	gold2	gray51	grey3	grey63	khaki2	lightyellow1	orangered1	royalblue2	tan4
bisque4	darkkhaki	gold1	gray50	grey2	grey62	khaki1	lightyellow	orangered	royalblue1	tan3
bisque3	darkgrey	gold	gray49	grey1	grey61	khaki	lightsteelblue4	orange4	royalblue	tan2
bisque2	darkgreen	ghostwhite	gray48	grey0	grey60	ivory4	lightsteelblue3	orange3	rosybrown4	tan1
bisque1	darkgray	gainsboro	gray47	grey	grey59	ivory3	lightsteelblue2	orange2	rosybrown3	tan
bisque	darkgoldenrod4	forestgreen	gray46	greenyellow	grey58	ivory2	lightsteelblue1	orange1	rosybrown2	steelblue4
beige	darkgoldenrod3	floralwhite	gray45	green4	grey57	ivory1	lightsteelblue	orange	rosybrown1	steelblue3
azure4	darkgoldenrod2	firebrick4	gray44	green3	grey56	ivory	lightslategrey	olivedrab4	rosybrown	steelblue2
azure3	darkgoldenrod1	firebrick3	gray43	green2	grey55	indianred4	lightslategray	olivedrab3	red4	steelblue1
azure2	darkgoldenrod	firebrick2	gray42	green1	grey54	indianred3	lightslateblue	olivedrab2	red3	steelblue
azure1	darkcyan	firebrick1	gray41	green	grey53	indianred2	lightskyblue4	olivedrab1	red2	springgreen4
azure	darkblue	firebrick	gray40	gray100	grey52	indianred1	lightskyblue3	olivedrab	red1	springgreen3
aquamarine4	cyan4	dodgerblue4	gray39	gray99	grey51	indianred	lightskyblue2	oldlace	red	springgreen2
aquamarine3	cyan3	dodgerblue3	gray38	gray98	grey50	hotpink4	lightskyblue1	navyblue	purple4	springgreen1
aquamarine2	cyan2	dodgerblue2	gray37	gray97	grey49	hotpink3	lightskyblue	navy	purple3	springgreen
aquamarine1	cyan1	dodgerblue1	gray36	gray96	grey48	hotpink2	lightseagreen	navajowhite4	purple2	snow4
aquamarine	cyan	dodgerblue	gray35	gray95	grey47	hotpink1	lightsalmon4	navajowhite3	purple1	snow3
antiquewhite4	cornsilk4	dimgrey	gray34	gray94	grey46	hotpink	lightsalmon3	navajowhite2	purple	snow2
antiquewhite3	cornsilk3	dimgray	gray33	gray93	grey45	honeydew4	lightsalmon2	navajowhite1	powderblue	snow1
antiquewhite2	cornsilk2	deepskyblue4	gray32	gray92	grey44	honeydew3	lightsalmon1	navajowhite	plum4	snow
antiquewhite1	cornsilk1	deepskyblue3	gray31	gray91	grey43	honeydew2	lightsalmon	moccasin	plum3	slategrey
antiquewhite	cornsilk	deepskyblue2	gray30	gray90	grey42	honeydew1	lightpink4	mistyrose4	plum2	slategray4
aliceblue	cornflowerblue	deepskyblue1	gray29	gray89	grey41	honeydew	lightpink3	mistyrose3	plum1	slategray3
white	coral4	deepskyblue	gray28	gray88	grey40	grey100	lightpink2	mistyrose2	plum	slategray2

绘图的例子

plot_confidence_interval.r

4 扩展
- 通用
- 特殊
- 交互
- 动画
- 图库

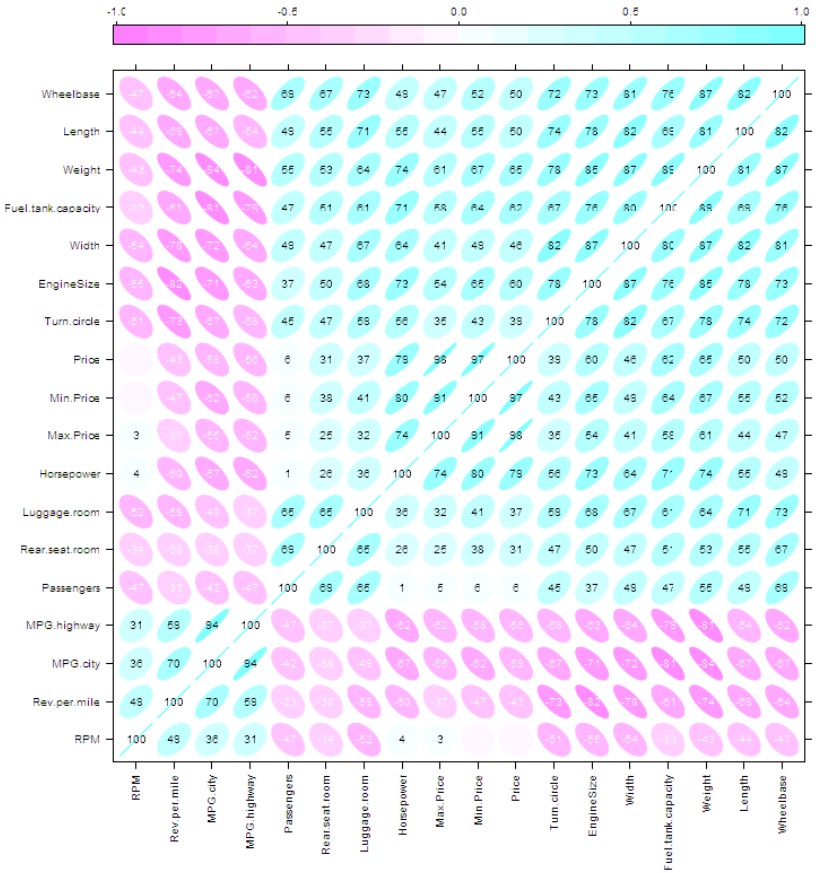
通用类

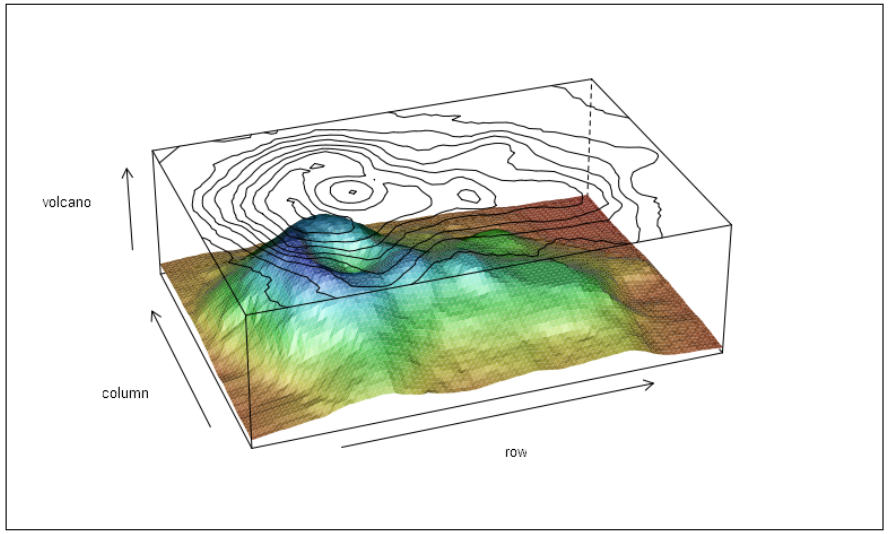
- lattice
- ggplot2

lattice

- The Trellis graphics of S(+) and The Elements of Graphing Data (Cleveland, 1985).
- The lattice package is an independent implementation of Trellis graphics
- lattice uses Paul Murrell's grid package, which provides more flexible low-level tools

lattice 绘图示例

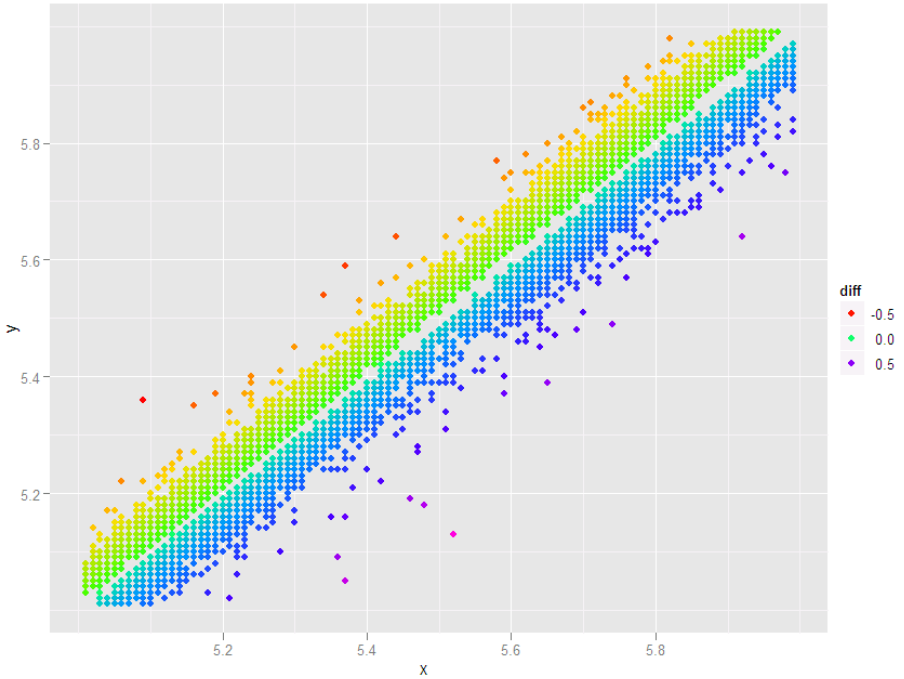


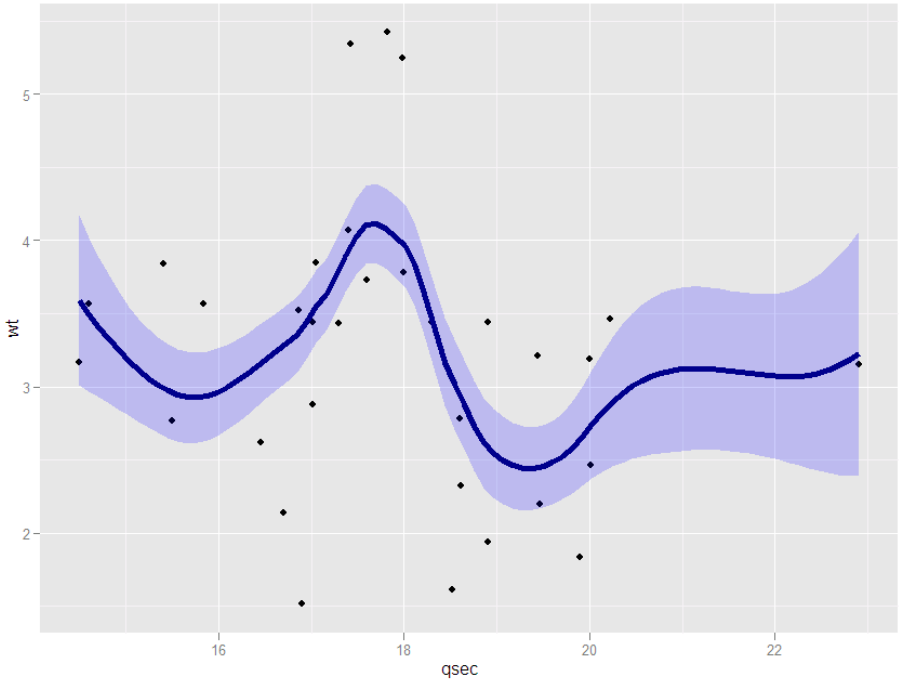


ggplot2

- The Grammar of Graphics[Wilkinson,2005] 在 R 中的实现
- 图层叠加的概念，如同魔方
  - 几何单位 (Geom, 点? 线? 光滑?)+ 统计变换 (Stat, 直方图? QQ 图?)+ 尺度表示 (Scale, 颜色渐变? 元素大小?)+ 坐标系 (Co-ord, 笛卡尔? 极坐标?)+ 面板分类 (Facet, 根据分类变量分别画图)+ 元素位置调整 (Position, 条形图并列或堆积? 散点随机微小打乱?)
  - 扩展了泛型函数：+ (使用非常形象)
- 细节设置自动化，例如图例

ggplot2 绘图示例

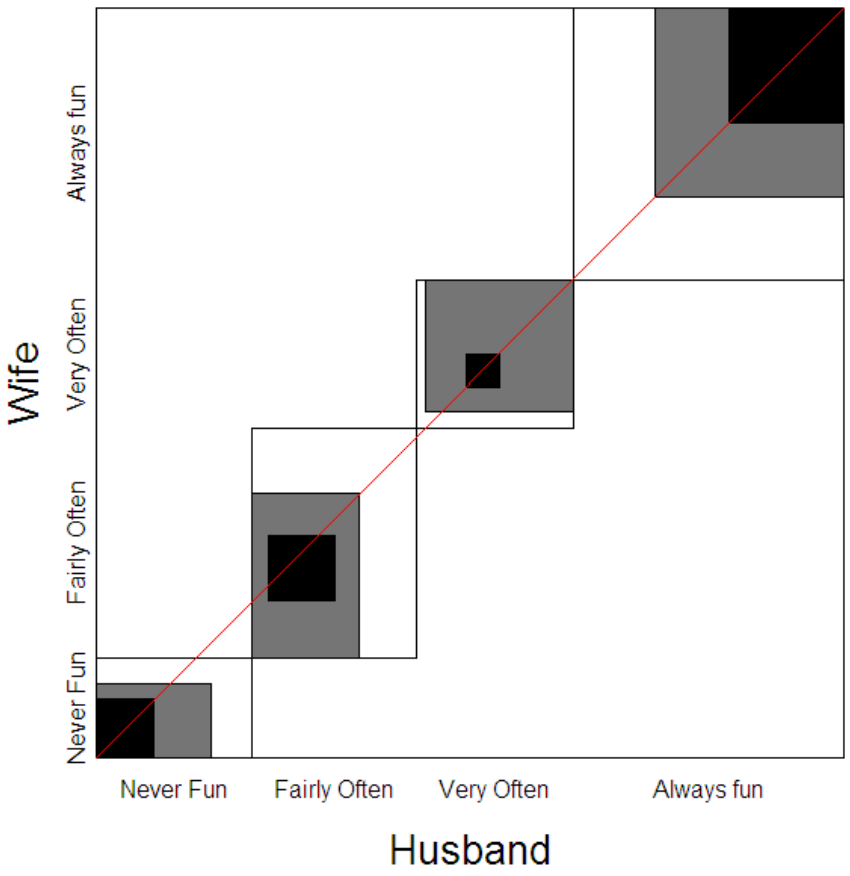




vcd

- Visualizing Categorical Data
- The package was inspired by the book "Visualizing Categorical Data" by Michael Friendly.

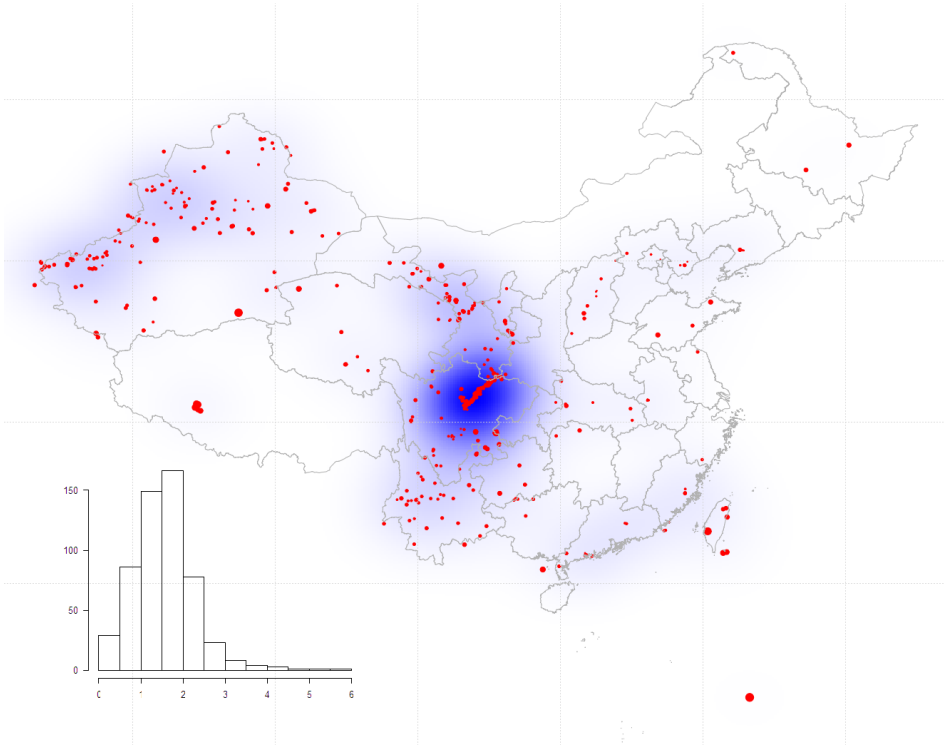
Bangdiwala's Observer Agreement Chart

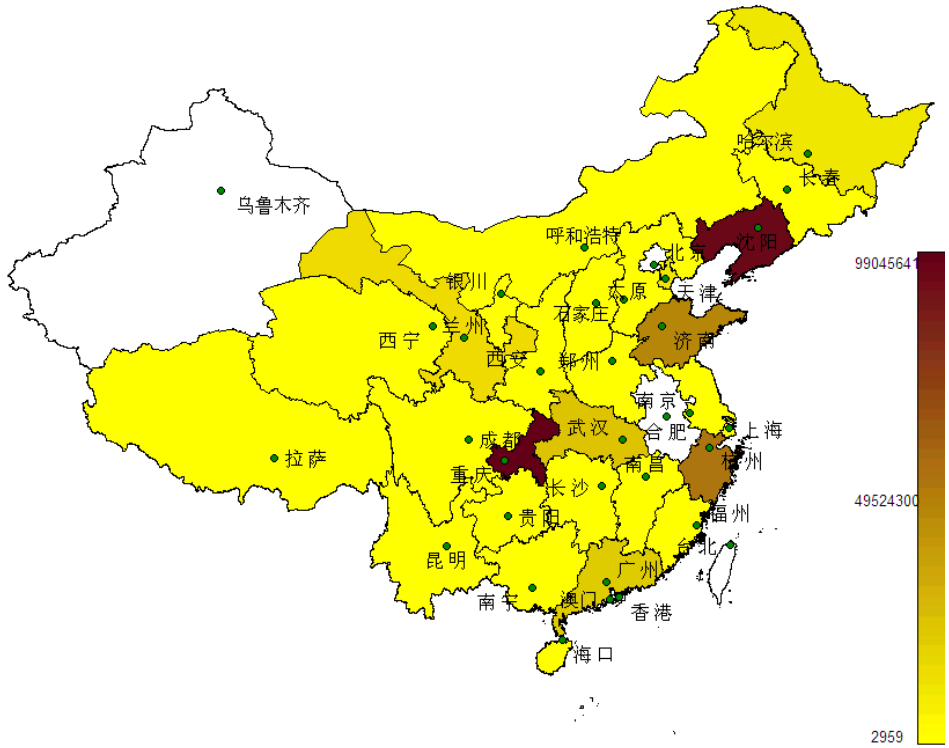


	Never Fun	Fairly Often	Very Often	Always fun
Never Fun	7	7	2	3
Fairly Often	2	8	3	7
Very Often	1	5	4	9
Always fun	2	8	9	14

Data from Hout et al. (1987) given by Agresti (1990) summarizing the responses of married couples to the questionnaire item: Sex is fun for me and my partner: (a) never or occasionally, (b) fairly often, (c) very often, (d) almost always.

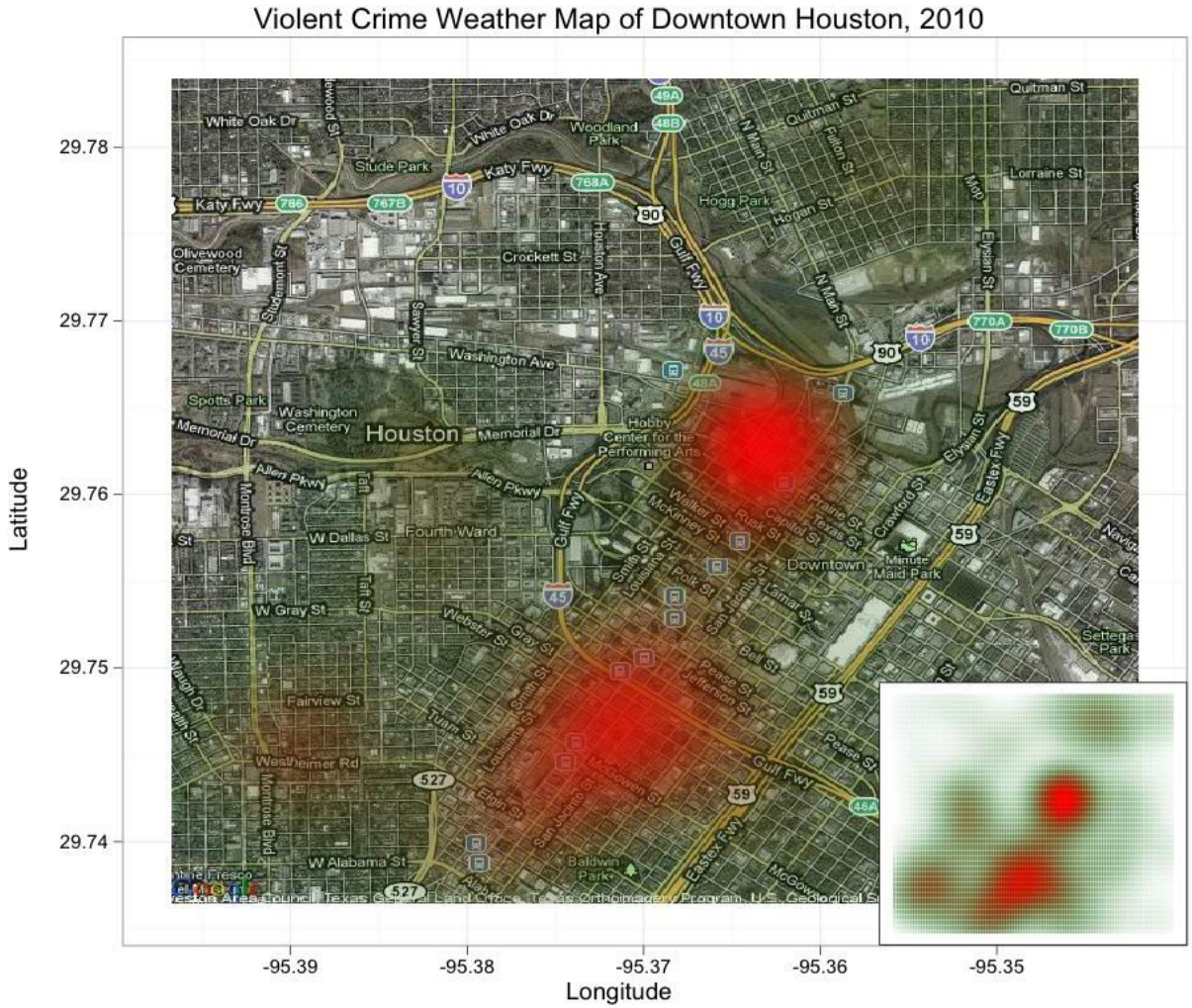
maps 系列





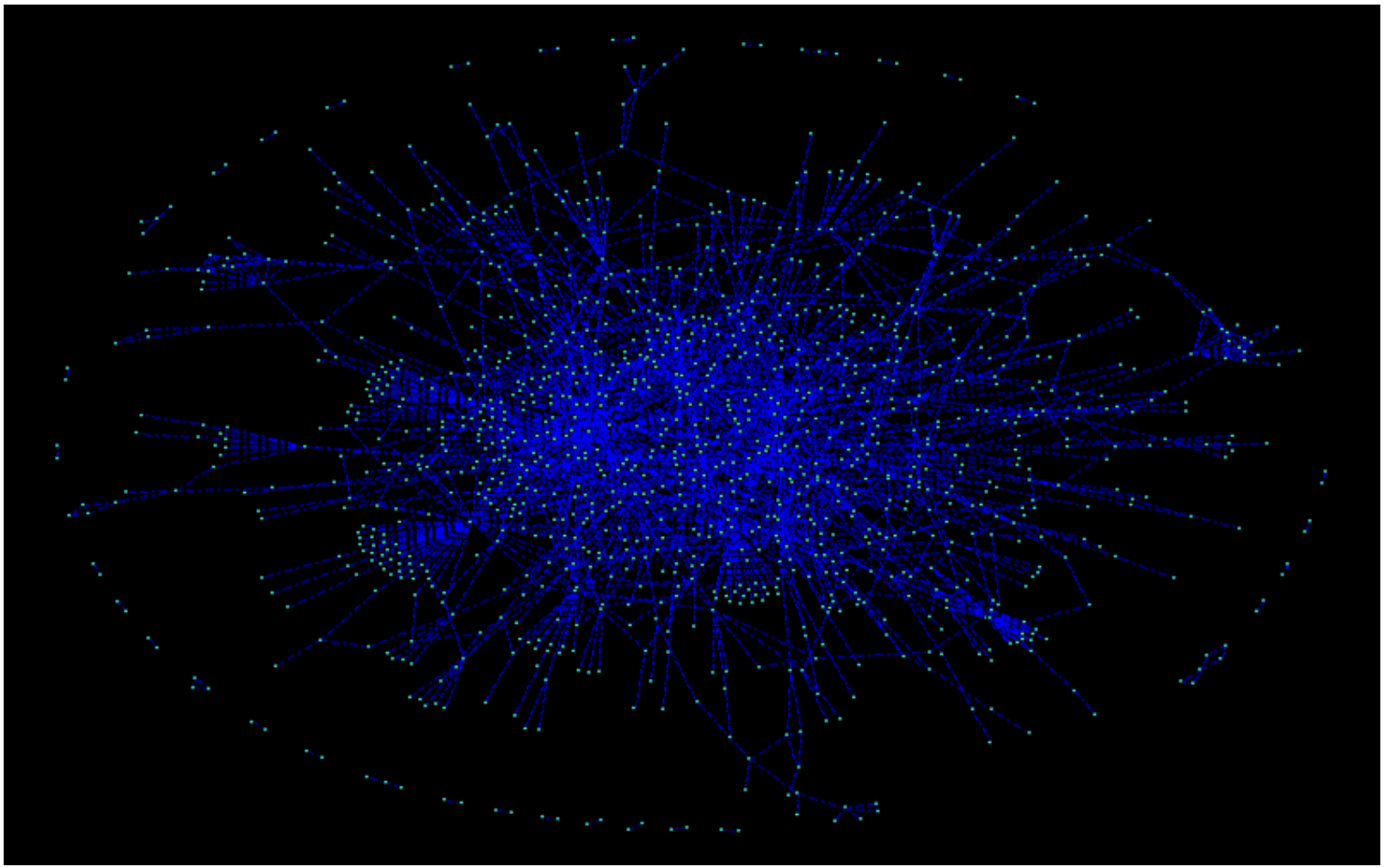
2010 年 11 月 5 日至 10 日国内地震情况以及某企业全国市场份额可视化。其他请参考：
http://www.bjt.name/2010/01/chinese-earthquake-visualization/

Google Map



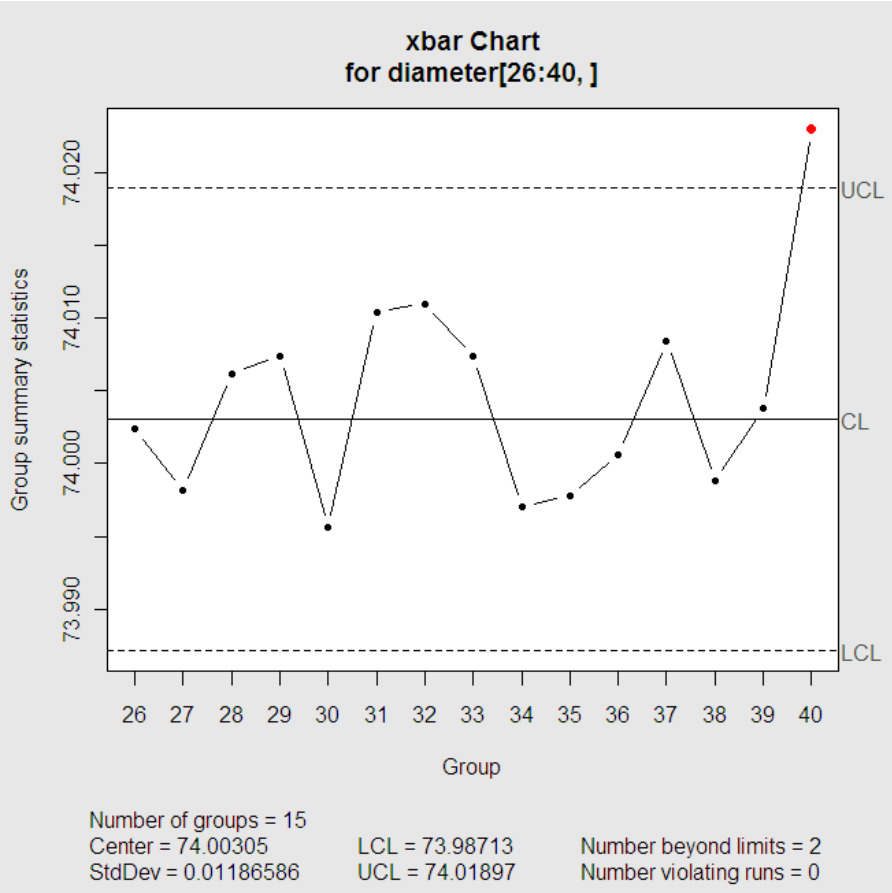
来源：https://github.com/hadley/ggplot2/wiki/

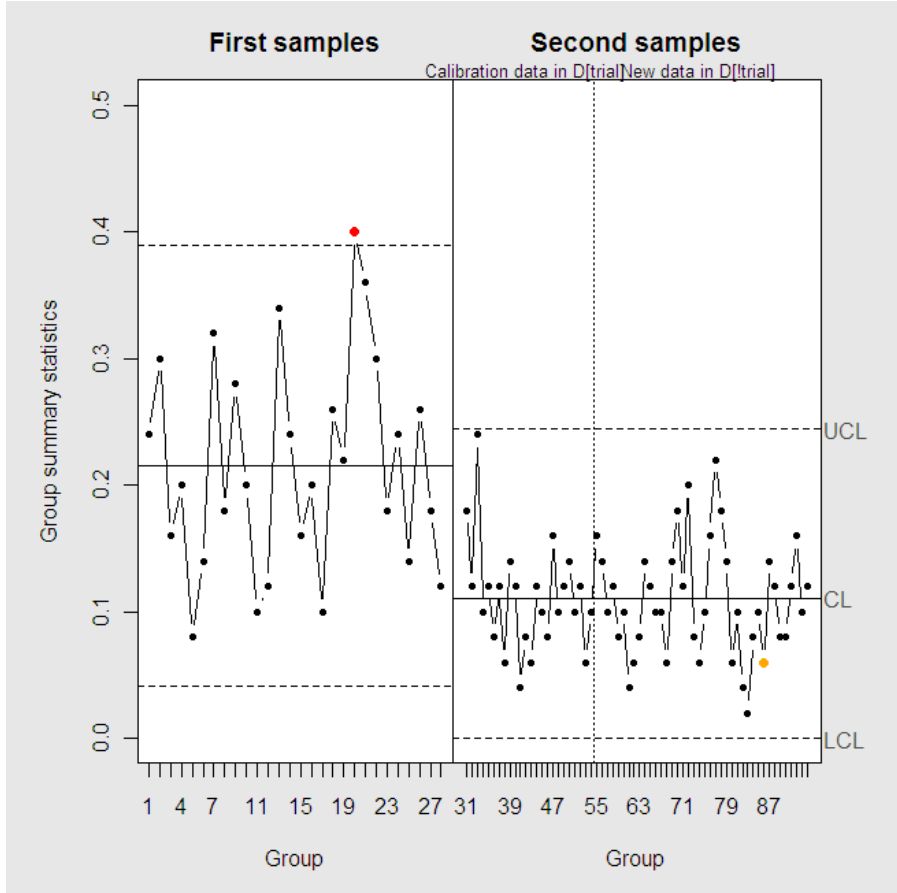
igraph



CRAN 上 R 包的网络关系：`http://www.bjt.name/2009/09/package-networks/`

Quality Control Charts





其他

高维数据展示相关：

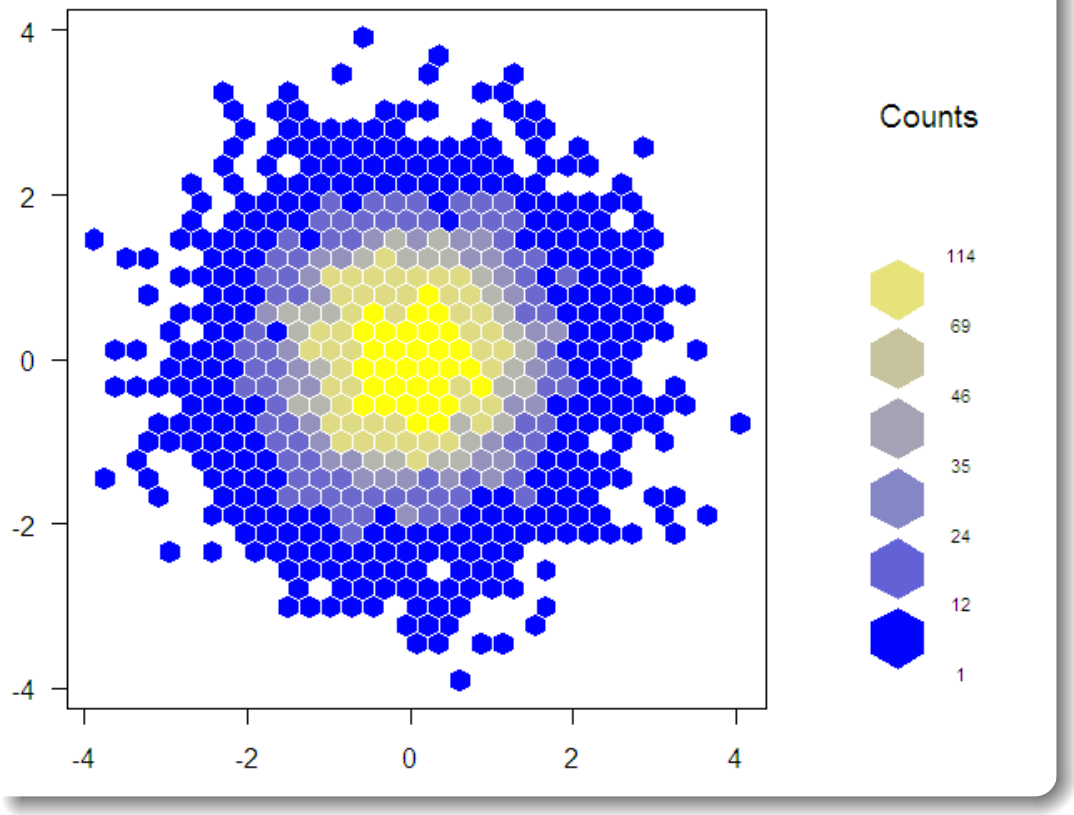
几何投影 (Geometric projection)

轮廓图,Andrews 调和曲线,···

基于图标 (Icon-based)

脸谱图, 星图, 雷达图, 花瓣图,···

大数据集



交互式绘图

- playwith
- rgl
- rggobi
- iplots
- 例子

动画相关的包

如果你有足够的想象力，可以自己来设计适合自己的动画！

或者借助一些包：

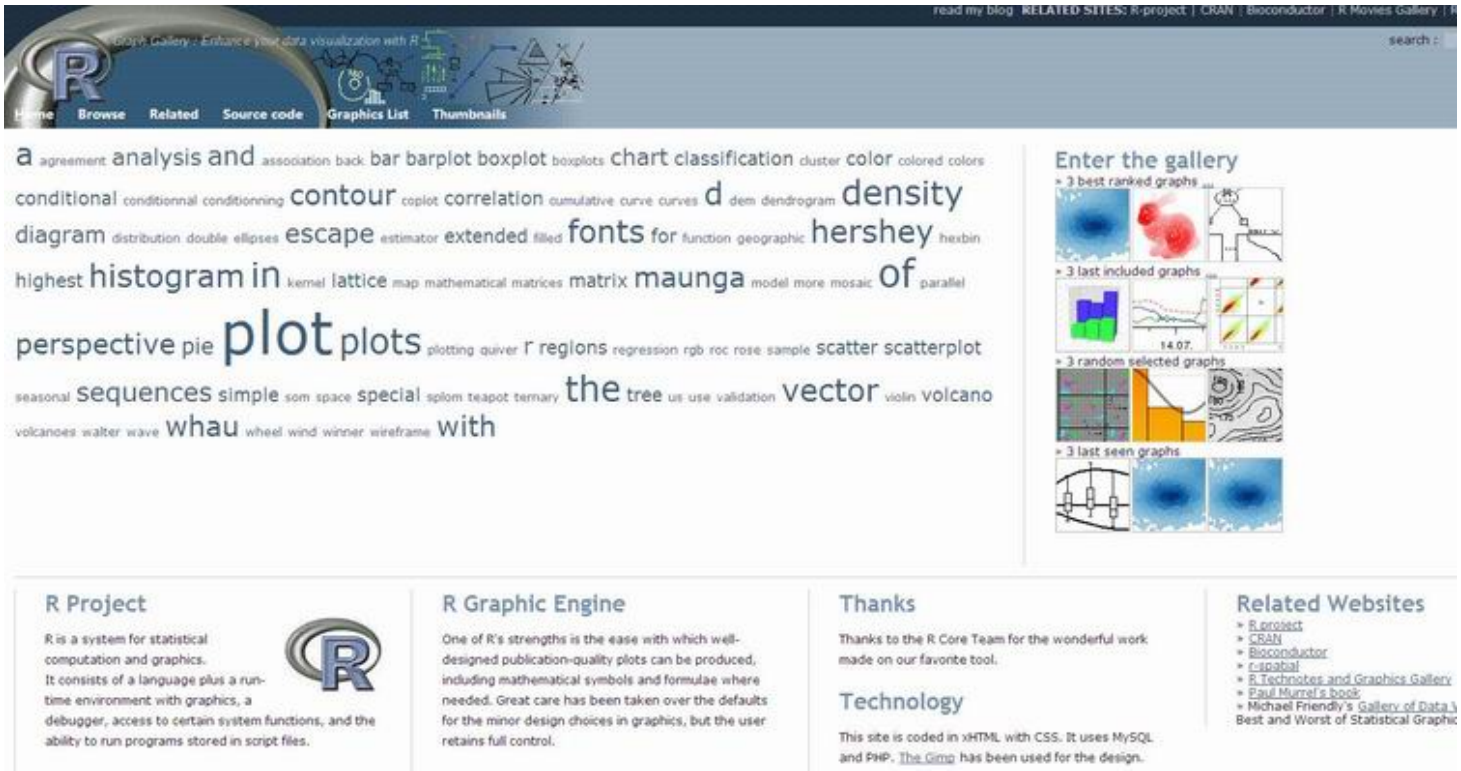
- `animation`(推荐)
- *write.gif* in package `caTools`
- `EBImage` pacakge in BioConductor

各类基础图形

- 一维数据：条形图、饼图、Cleveland 点图、坐标轴须、带状图
- 散点图：散点图、向日葵散点图
- 曲线：函数曲线
- 密度和分布：直方图、茎叶图、QQ 图
- 汇总：箱线图、因素效应图
- 分类数据关联：关联图、马赛克图
- 分类对连续：条件密度图、棘状图
- 分类画图：协同图
- 三维图形：颜色图、等高线图、三维透视图、平滑散点图
- 高维散点图：散点图矩阵、符号图

R 的图库

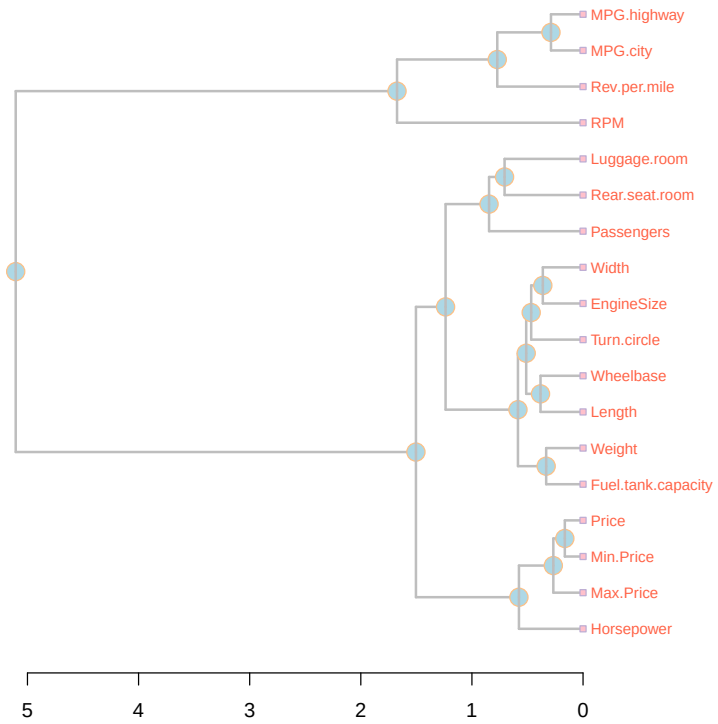
http://addictedtor.free.fr/graphiques/



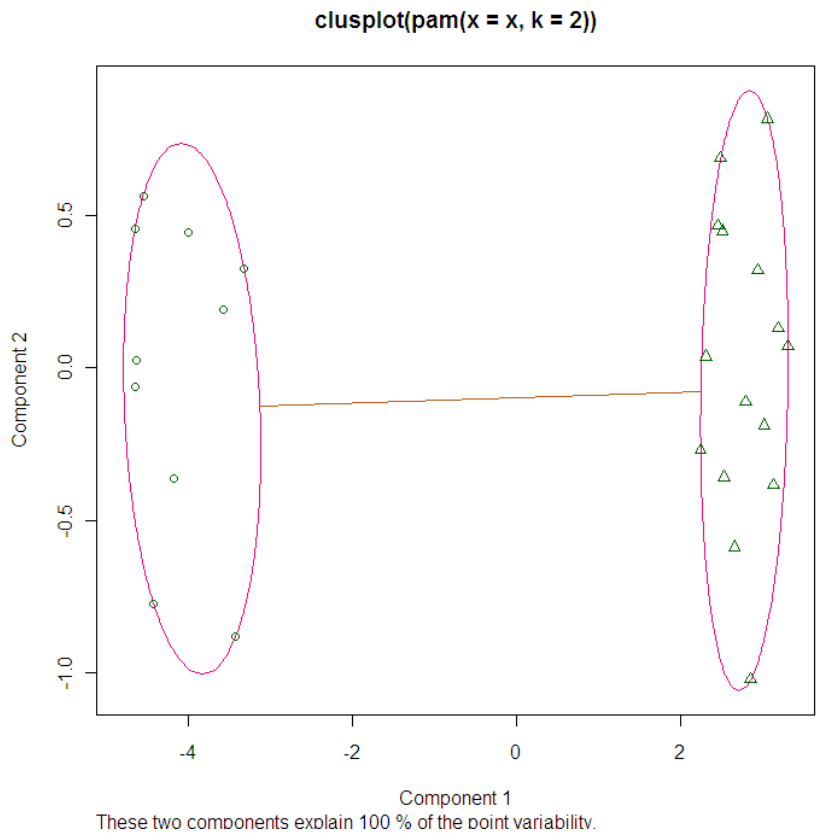
5 模型

聚类

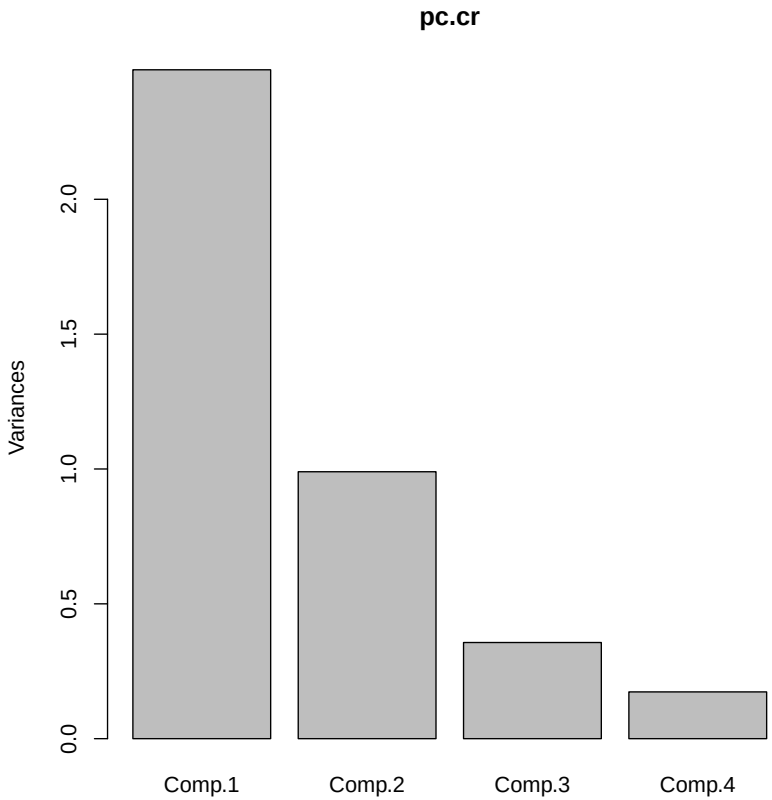
Hierarchical Clustering

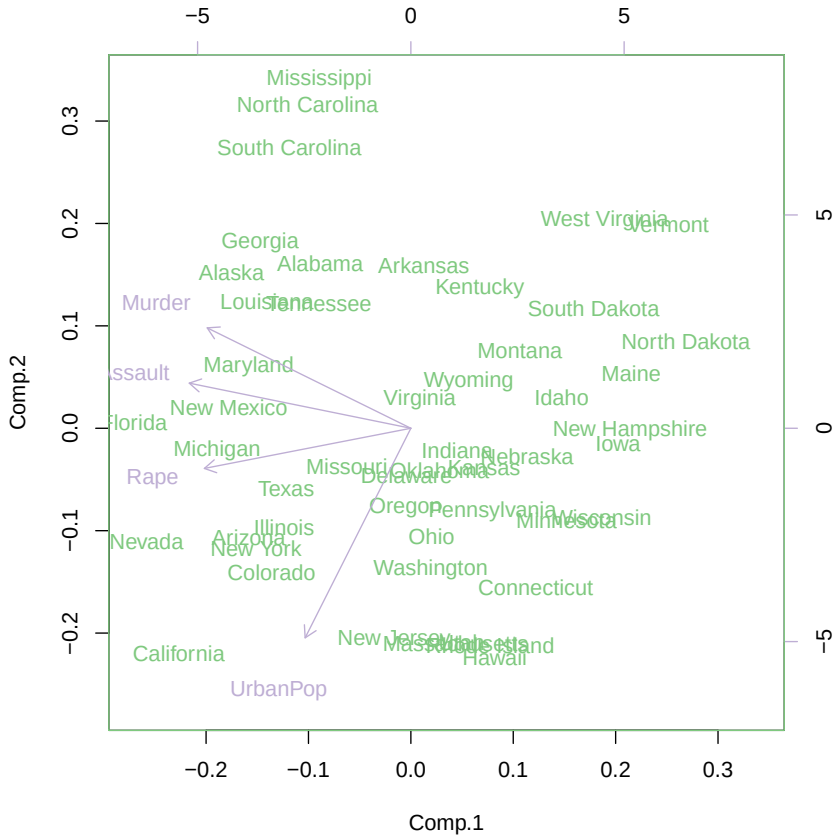


Partitioning Around Medoids

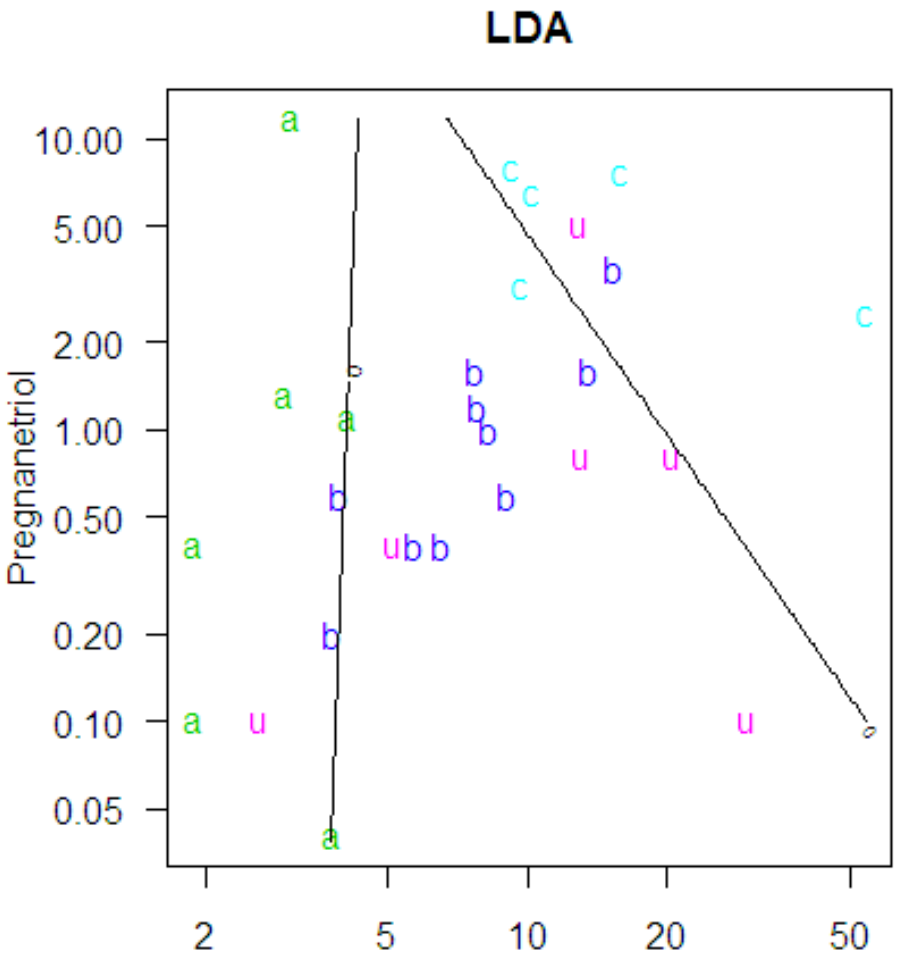


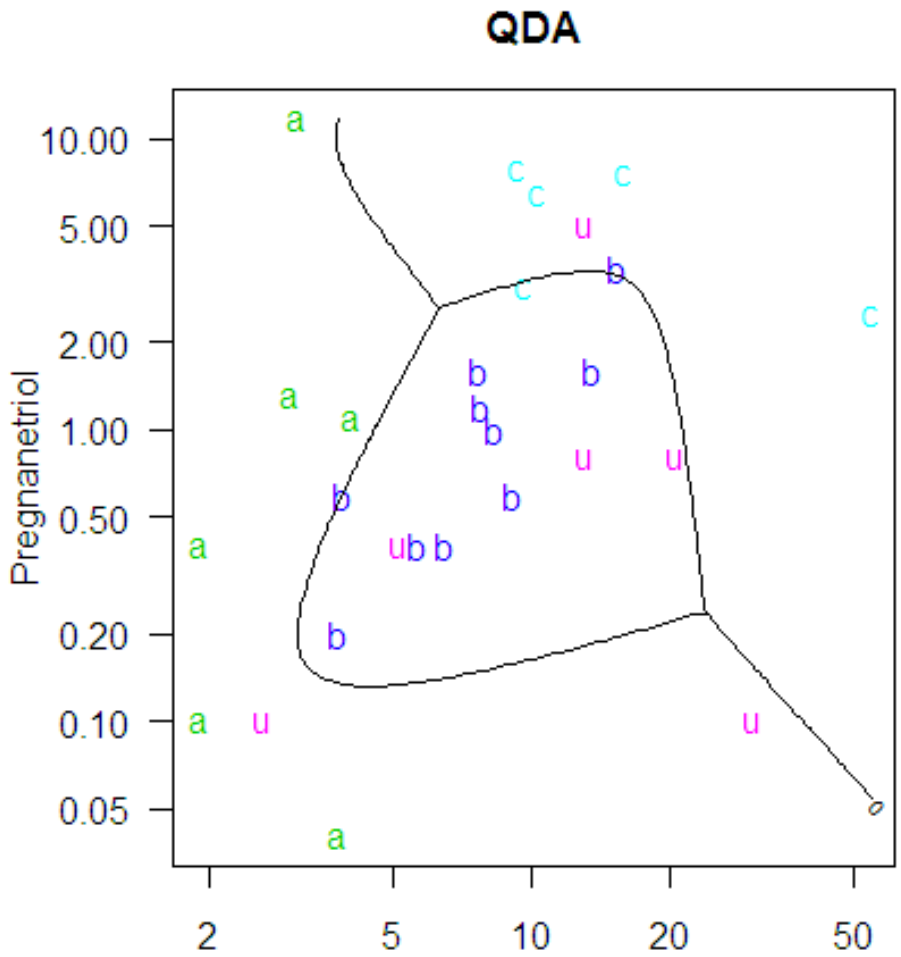
主成份和因子分析


pc.cr
Variances
0.0
0.5
1.0
1.5
2.0
Comp.1
Comp.2
Comp.3
Comp.4

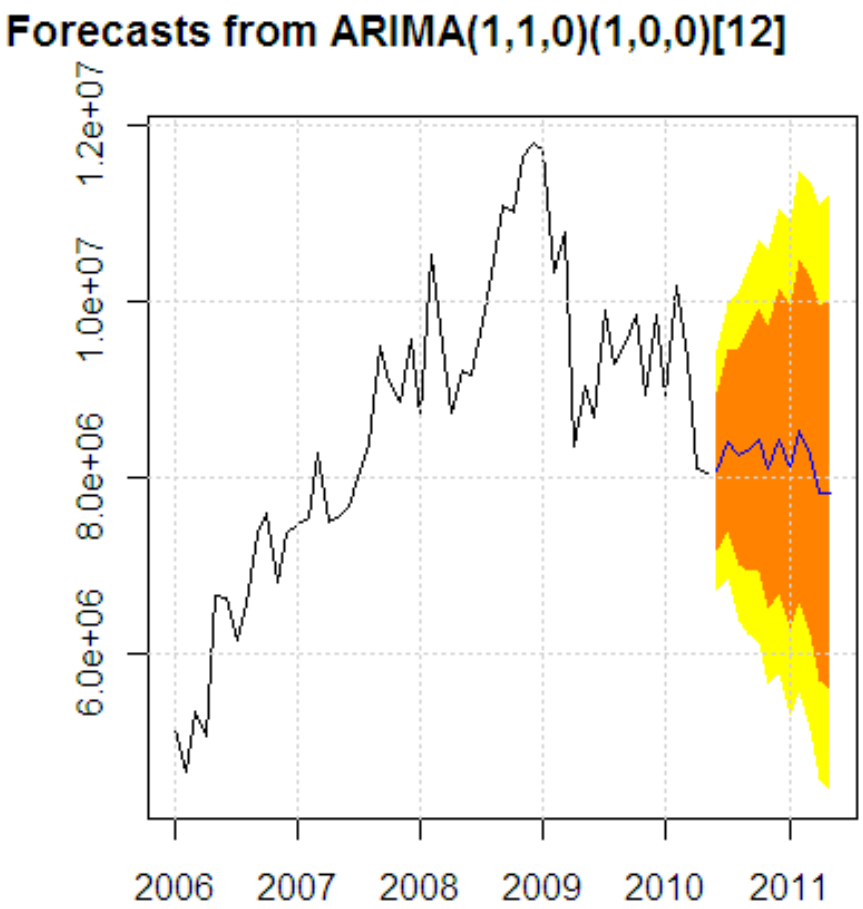

−5
0
5
Mississippi
North Carolina
South Carolina
West Virginia
Vermont
Georgia
Alabama
Arkansas
Alaska
Kentucky
Murder
Louisiana
Tennessee
South Dakota
North Dakota
Montana
Maryland
Maine
Wyoming
Virginia
Idaho
New Mexico
New Hampshire
Iowa
Michigan
Indiana
Nebraska
Missouri
Oklahoma
Kansas
Rape
Delaware
Texas
Oregon
Pennsylvania
Wisconsin
Minnesota
Illinois
Nevada
Arizona
Ohio
New York
Colorado
Washington
Connecticut
California
Hawaii
UrbanPop
Comp.2
Comp.1
−0.2
−0.1
0.0
0.1
0.2
0.3

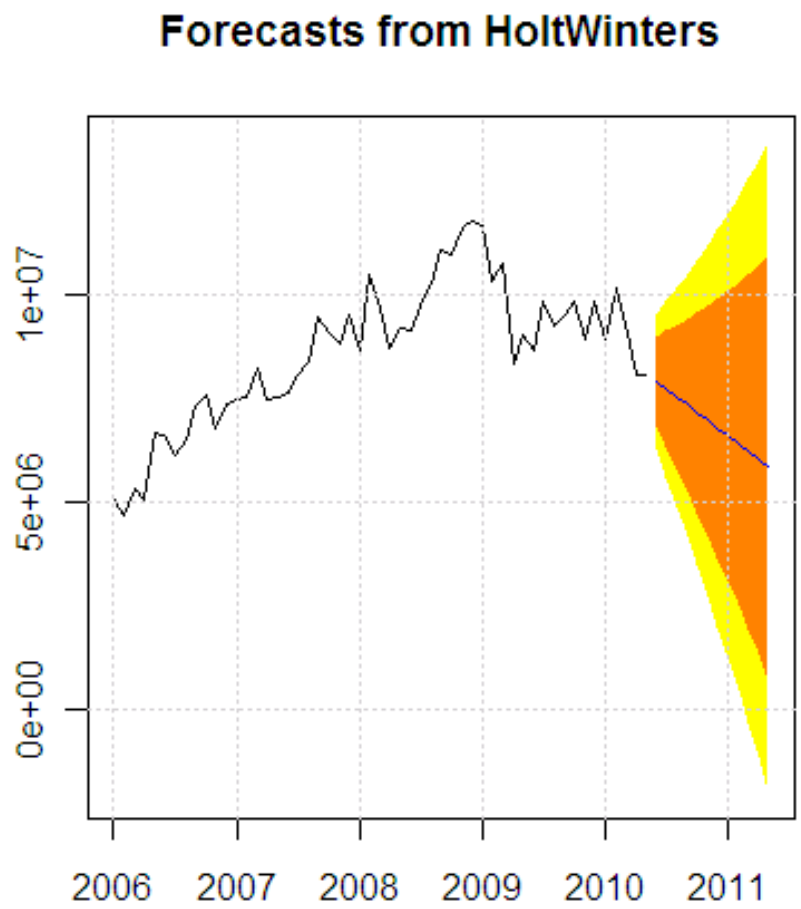
判别分析


LDA
Pregnanetriol
10.00
5.00
2.00
1.00
0.50
0.20
0.10
0.05
2
5
10
20
50


QDA
Pregnanetriol
10.00
5.00
2.00
1.00
0.50
0.20
0.10
0.05
2
5
10
20
50

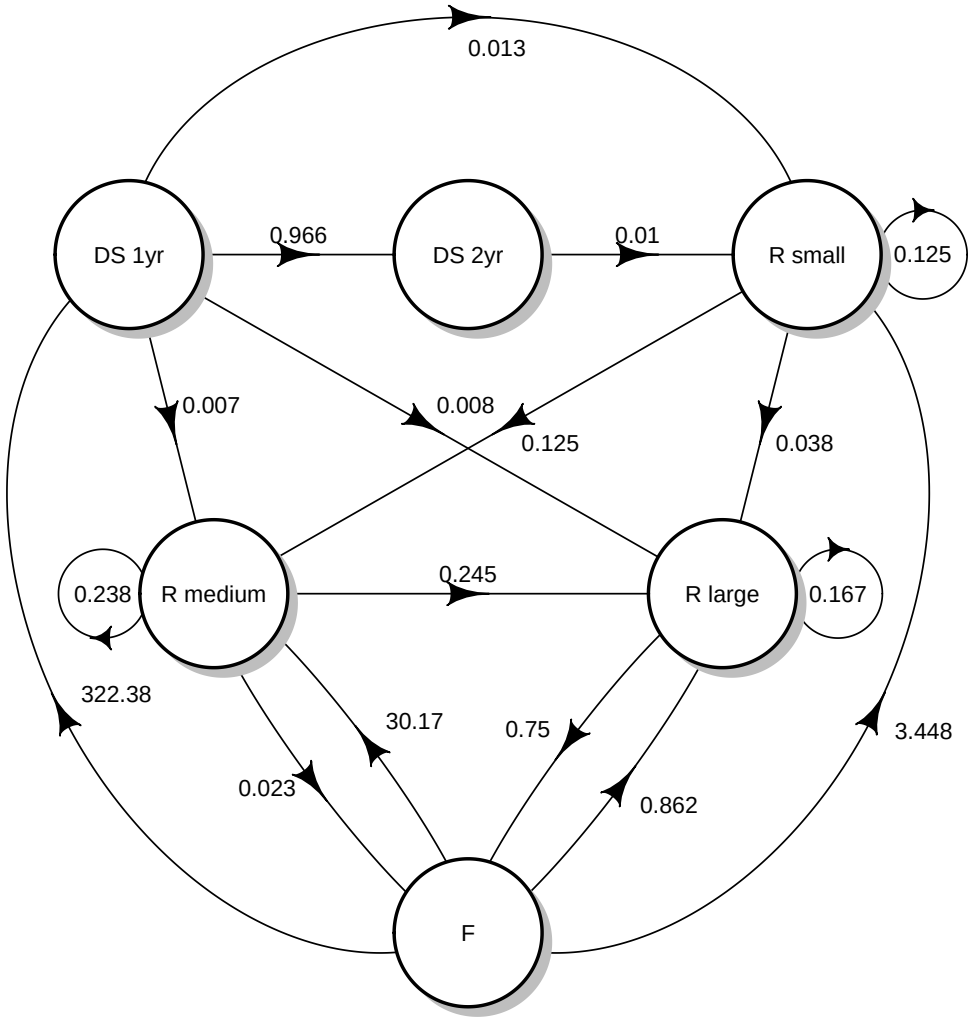
时间序列


Forecasts from ARIMA(1,1,0)(1,0,0)[12]
6.0e+06
8.0e+06
1.0e+07
1.2e+07
2006
2007
2008
2009
2010
2011


Forecasts from HoltWinters
0e+00
5e+06
1e+07
2006
2007
2008
2009
2010
2011

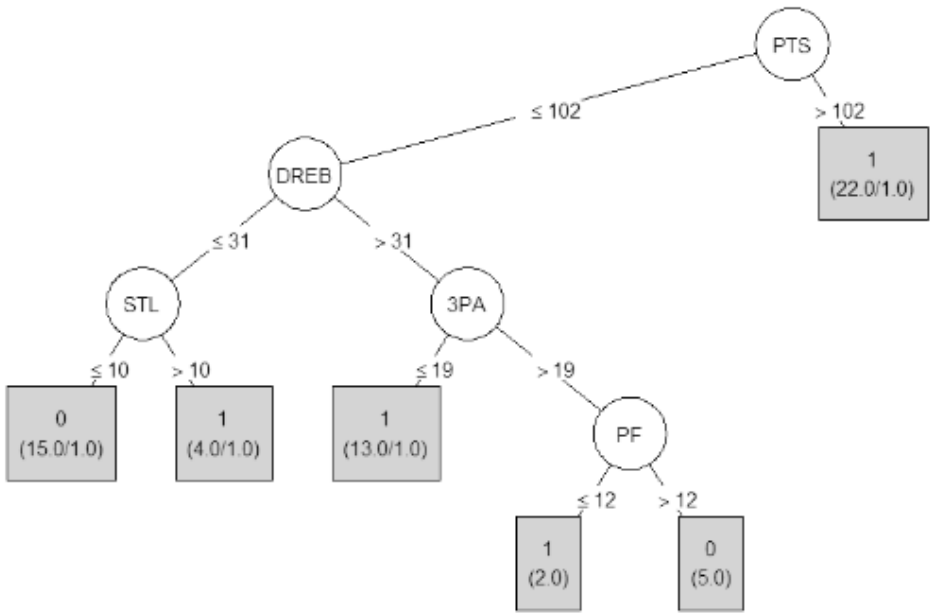
图模型

Life cycle of teasel



树型模型

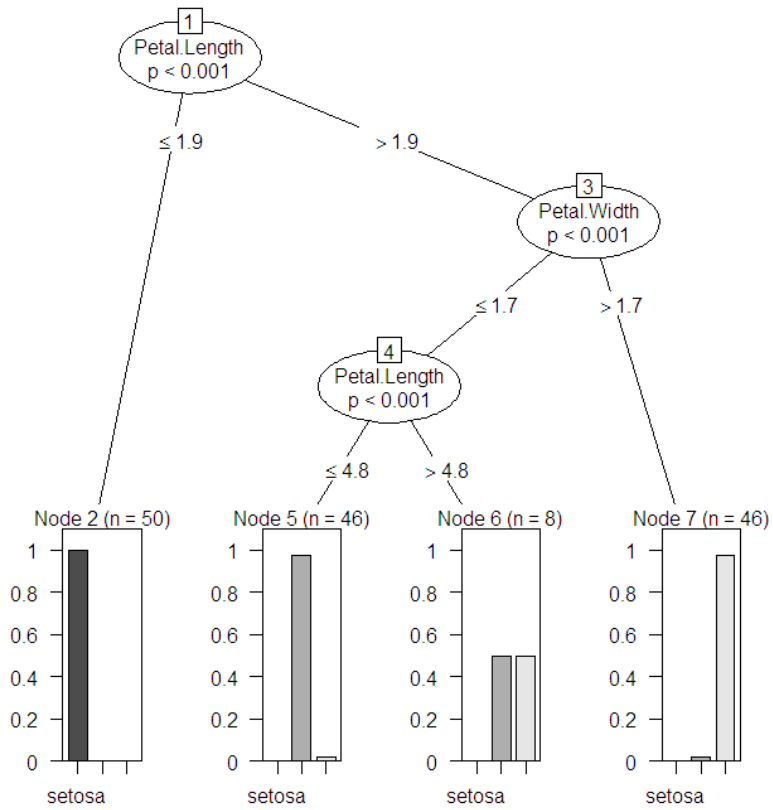
J48



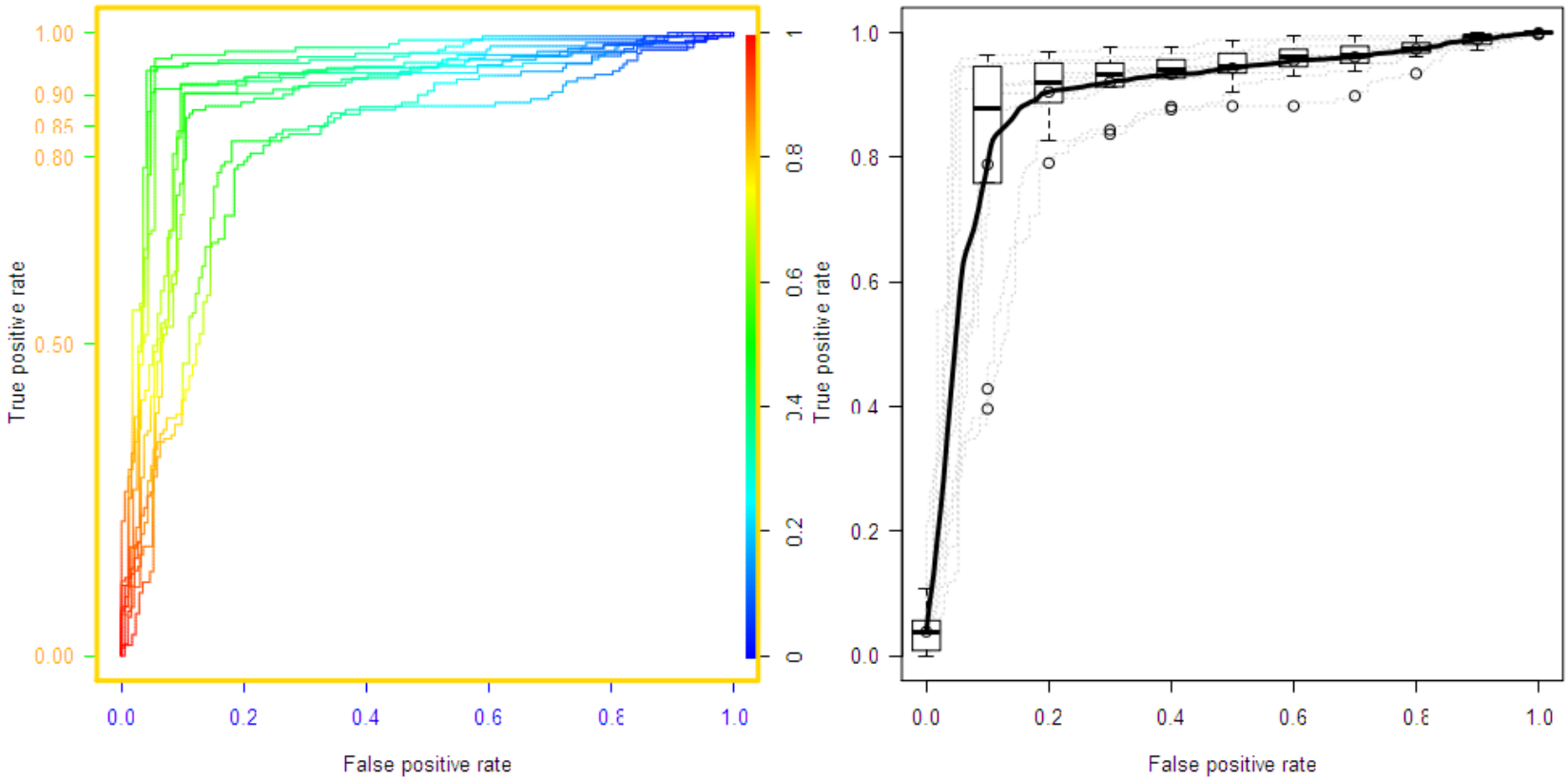
h

```
//bjt.name/2009/03/data-2008-2009-rockets
```

Conditional Inference Trees



ROC 曲线


True positive rate
False positive rate
1.00
0.90
0.85
0.80
0.50
0.00
0.0
0.2
0.4
0.6
0.8
1.0
0
0.2
0.4
0.6
0.8
1

参考文献


W. N. Venables and B. D. Ripley
Modern Applied Statistics with S 2002


John M. Chambers
Software for Data Analysis: Programming with R 2007


M. Friendly
A Brief History of Data Visualization 2005


Deepayan Sarkar
Lattice Multivariate Data Visualization with R 2007


Leland Wilkinson
The Grammar of Graphics, 2nd ed 2005


谢益辉
现代统计图形 2010

关于

刘思喆

- Email：sunbjt<at>gmail.com
- Blog：http://www.bjt.name

谢益辉

- Email：xie<at>yihui.name
- Blog：http://www.yihui.name

问题请关注：

- COS 论坛 R 版：http://cos.name/cn/forum/15
- 谢谢各位！

Jump to first slide