CAPITAL OF STATISTICS

PROFESSION, HUMANITY & INTEGRITY

R 与统计作图及其实例

统计之都五周年系列活动 — 第四届 R 会议上海会场

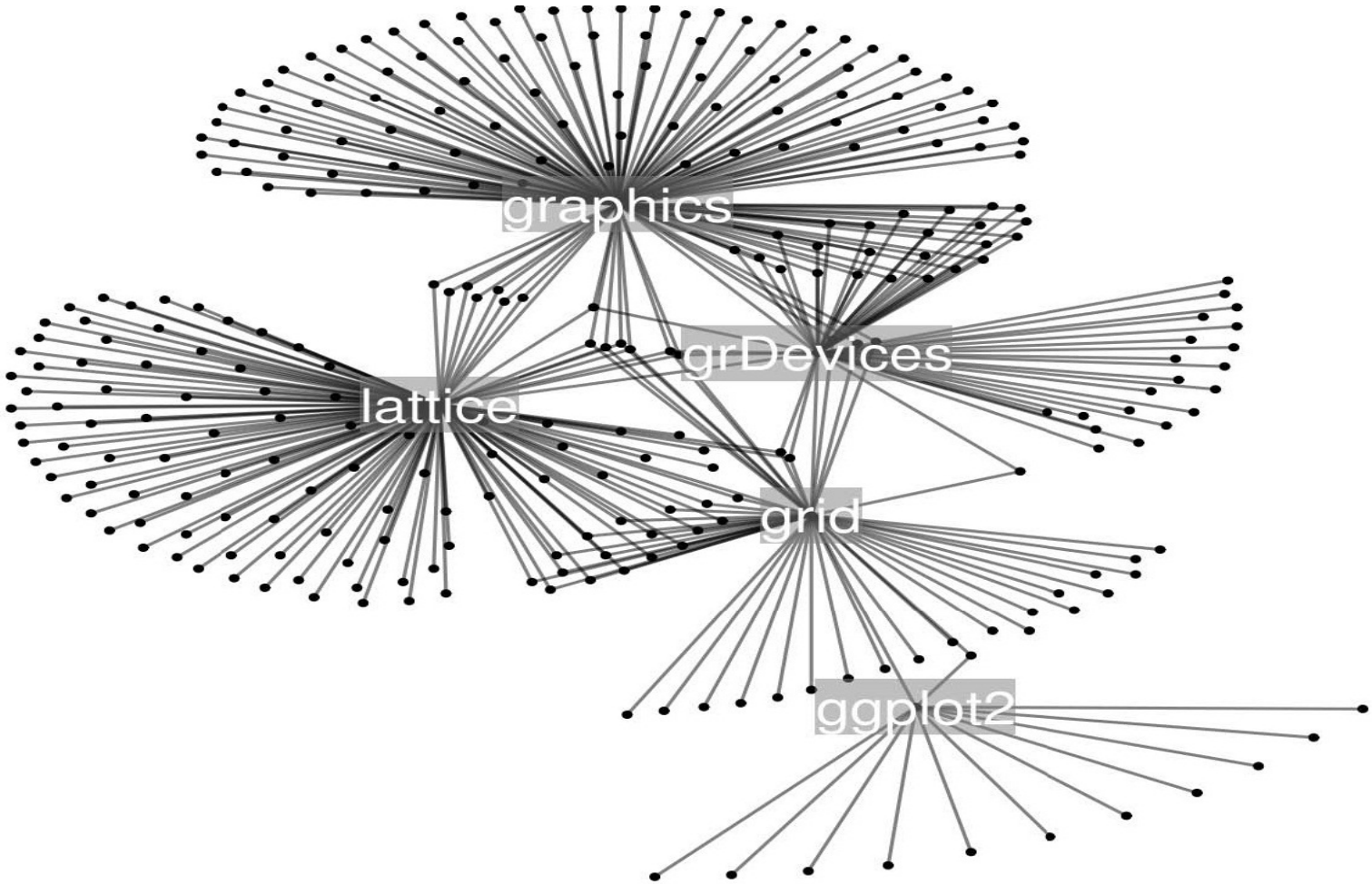
熊熹　邱怡轩　高涛　魏太云

中国人民大学统计学院

2011 年 11 月

提纲

R 中的作图系统



图形系统的特点

graphics

grid

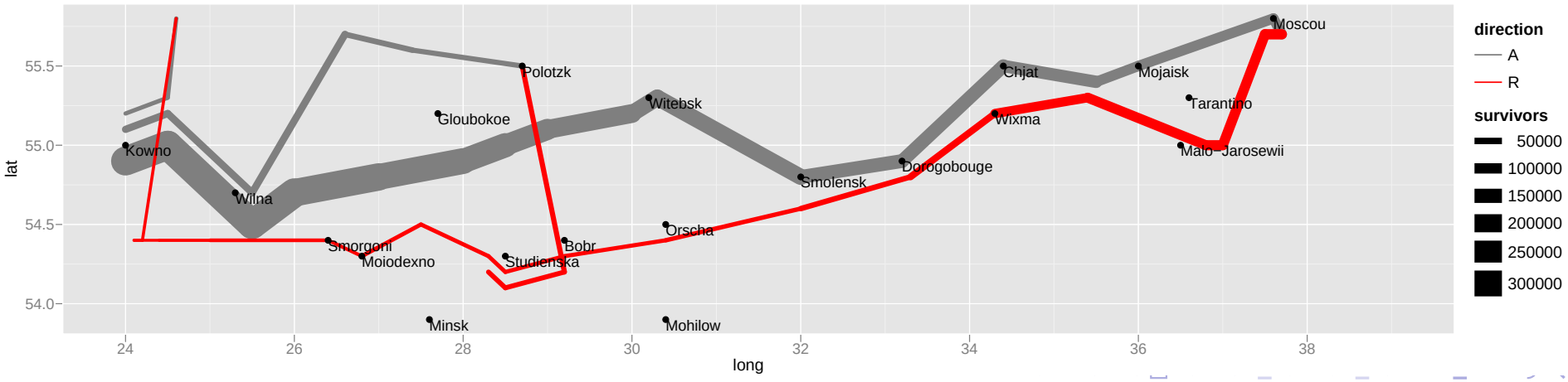
lattice

ggplot

提纲

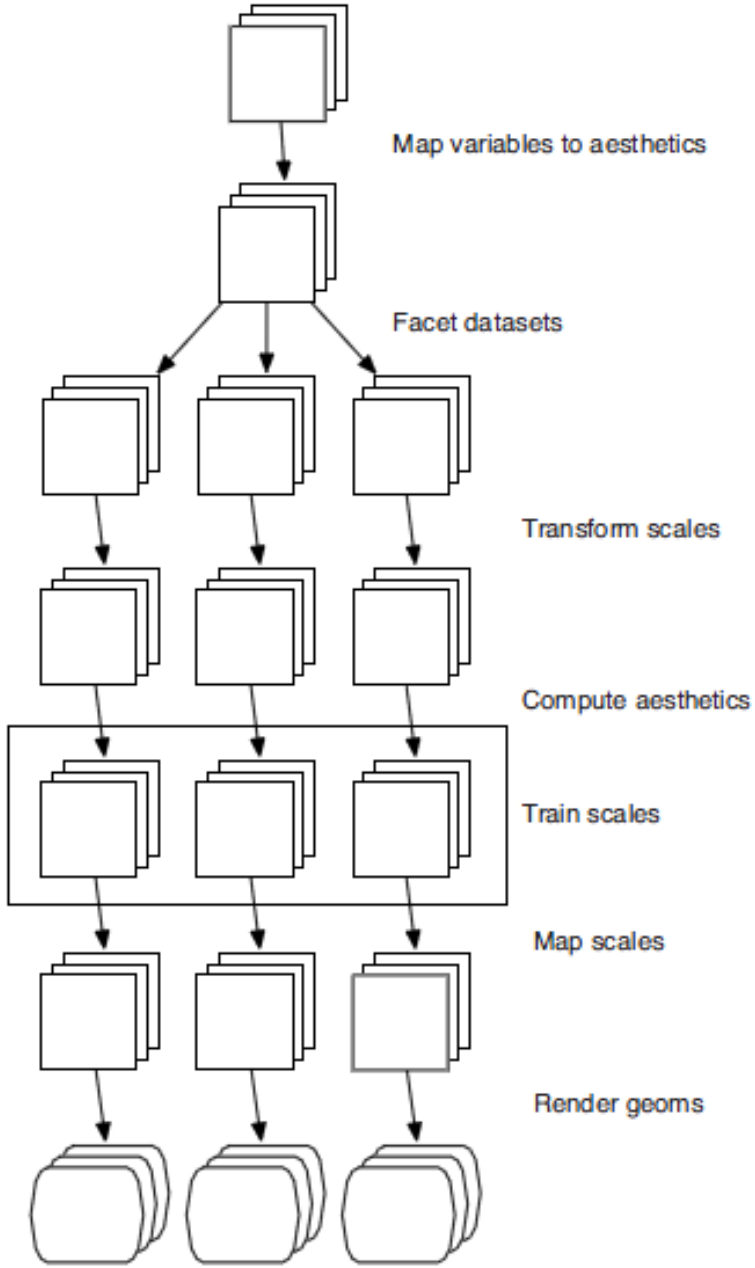
回眸一瞥

```
> ggplot(cities, aes(x = long, y = lat)) +
+ geom_path(aes(size = survivors, colour = direction, group
= group), data=troops) +
+ geom_point() +
+ geom_text(aes(label = city), hjust=0, vjust=1, size=4) +
+ scale_size(to = c(1, 10)) +
+ scale_colour_manual(values = c("grey50","red")) +
+ scale_x_continuous(limits = c(24, 39))
```



ggplot2 的理念

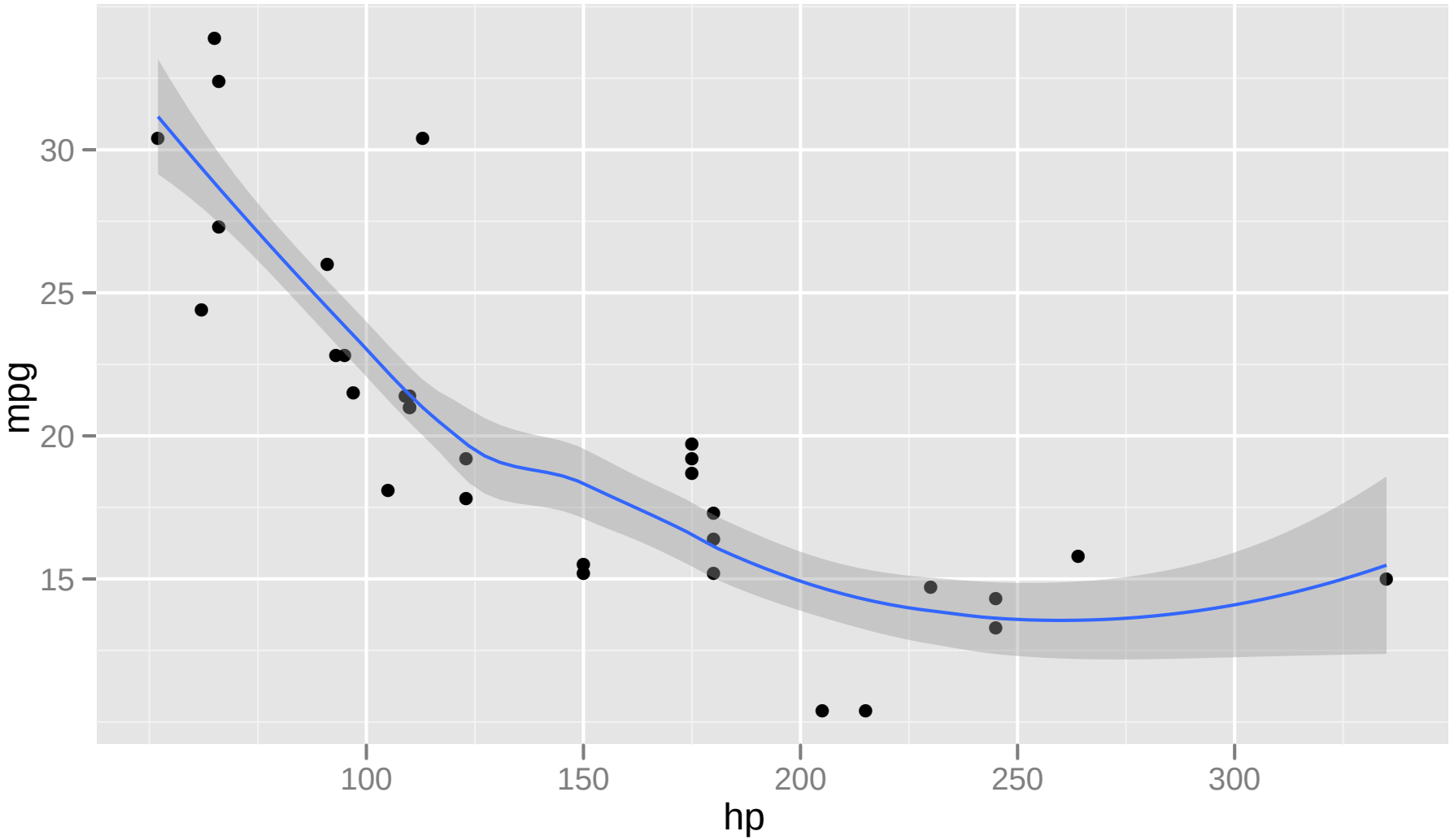
1. 层 (layer) 的叠加
   - 数据集
   - 变量与“美感”间的映射 (aesthetic mapping)
   - 统计转化
   - 几何形状
   - 位置调整
2. 专门的一套语法，比如"+" 的灵活应用


Map variables to aesthetics
Facet datasets
Transform scales
Compute aesthetics
Train scales
Map scales
Render geoms

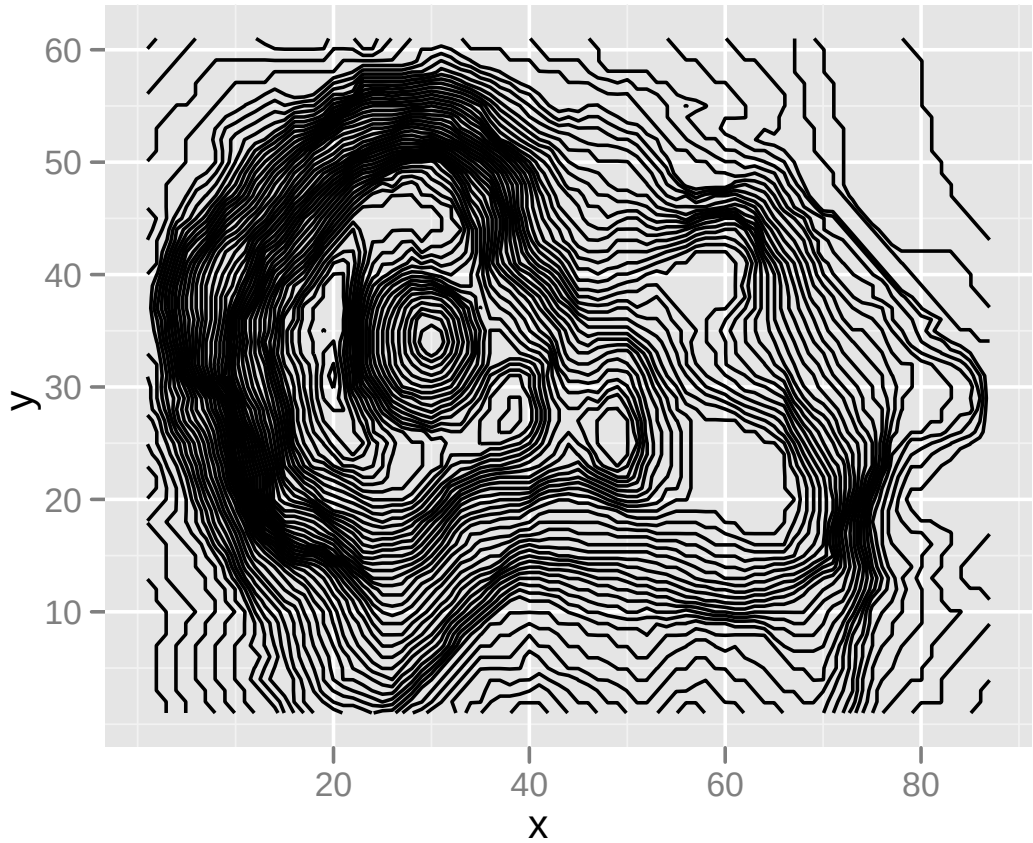
ggplot2 的系统架构

1. 几何形状 (geom)
2. 统计量 (statistic)
3. 标度 (scale)
4. 坐标系 (coordinate system)
5. 切片 (facet)
6. 位置调整 (Position adjustments)
7. 主题 (theme)

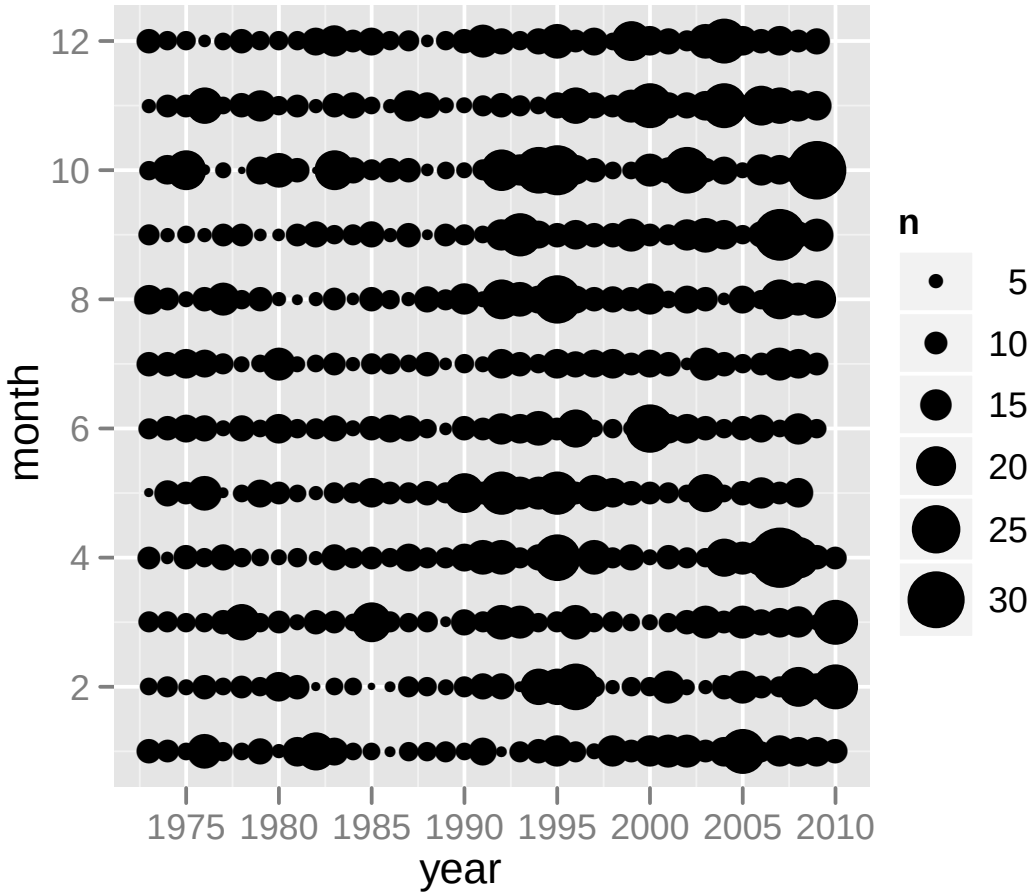
```
> print(qplot(hp, mpg, data=mtcars) + geom_smooth())
```



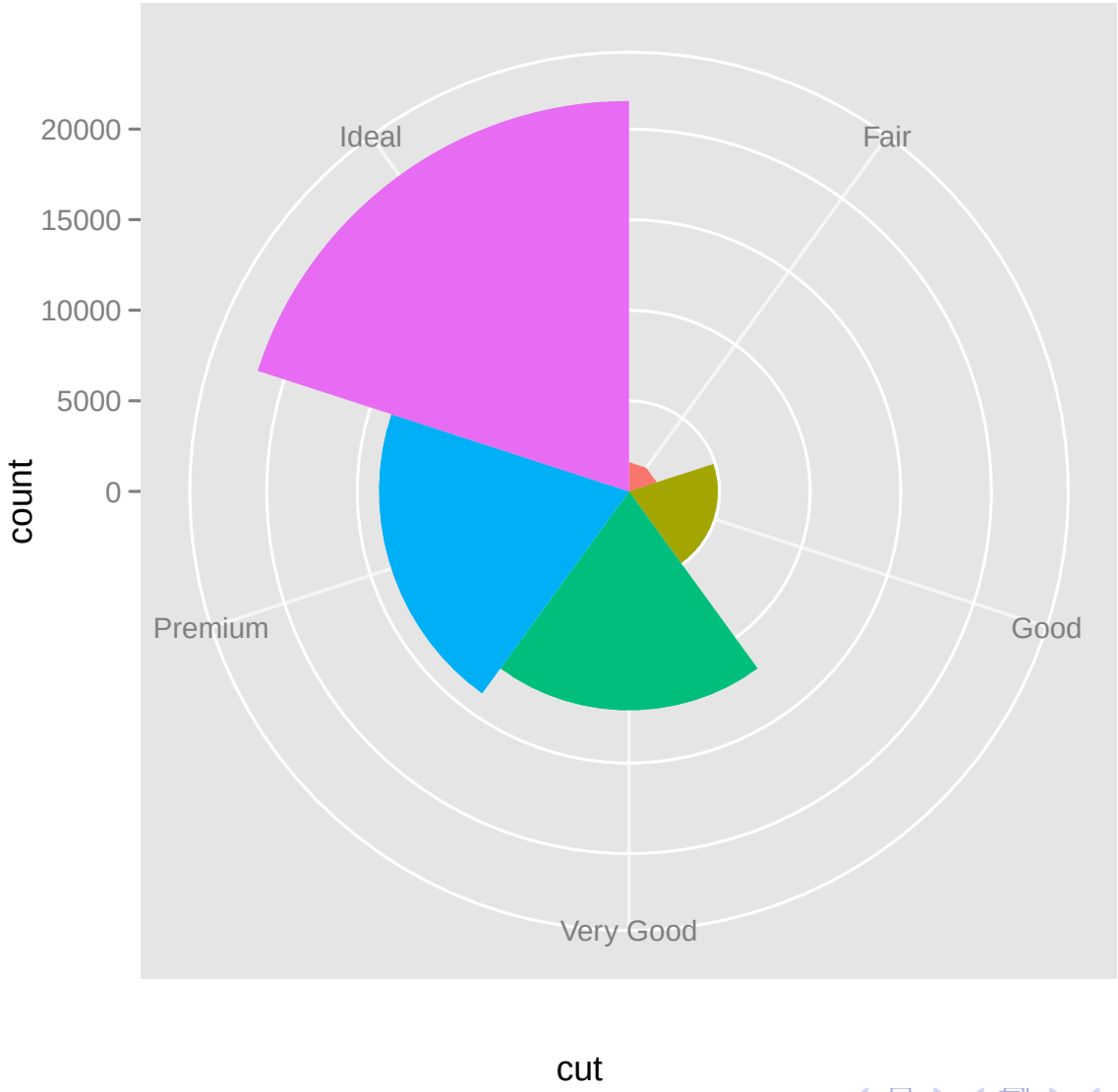
```
> volcano3d <- melt(volcano)
> names(volcano3d) <- c("x", "y", "z")
> v <- ggplot(volcano3d, aes(x, y, z = z))
> v + stat_contour(binwidth = 2)
```



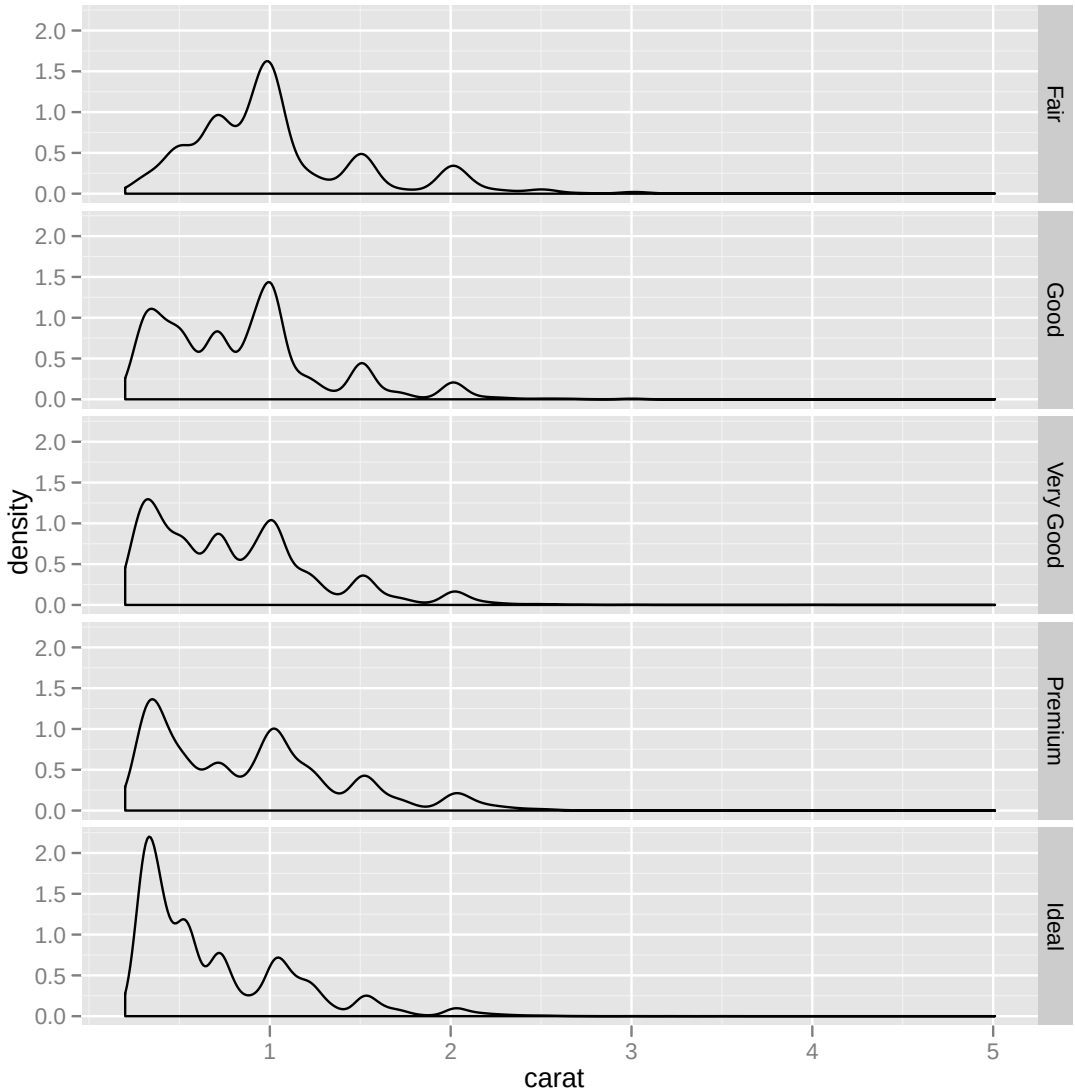
```
> data(quake6, package = "MSG")
> p = ggplot(quake6, aes(x = year, y = month))
> print(p + stat_sum(aes(size = ..n..)) + scale_size(to =
c(1, 8)))
```



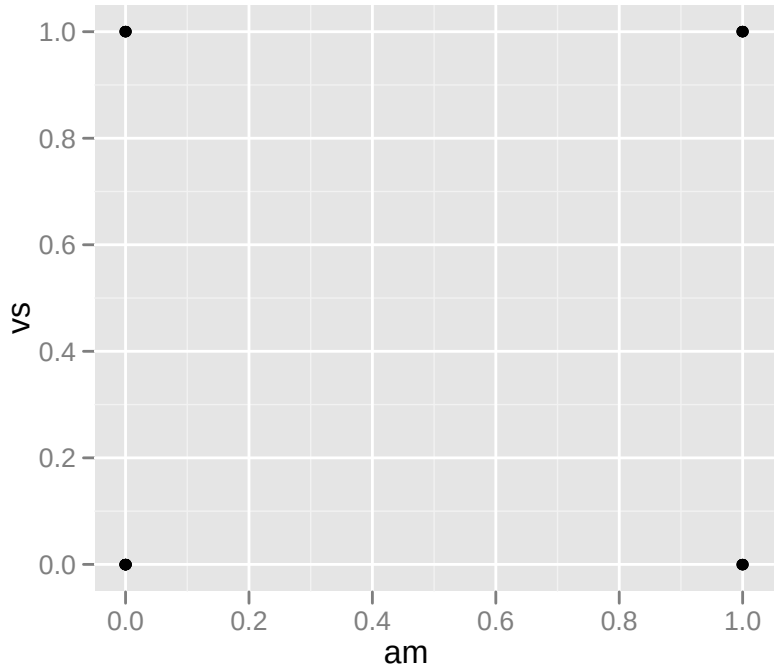
```
> p = qplot(cut, data=diamonds, fill=cut) + coord_polar()
> print(p+opts(legend.position="none")+geom_bar(width=1))
```

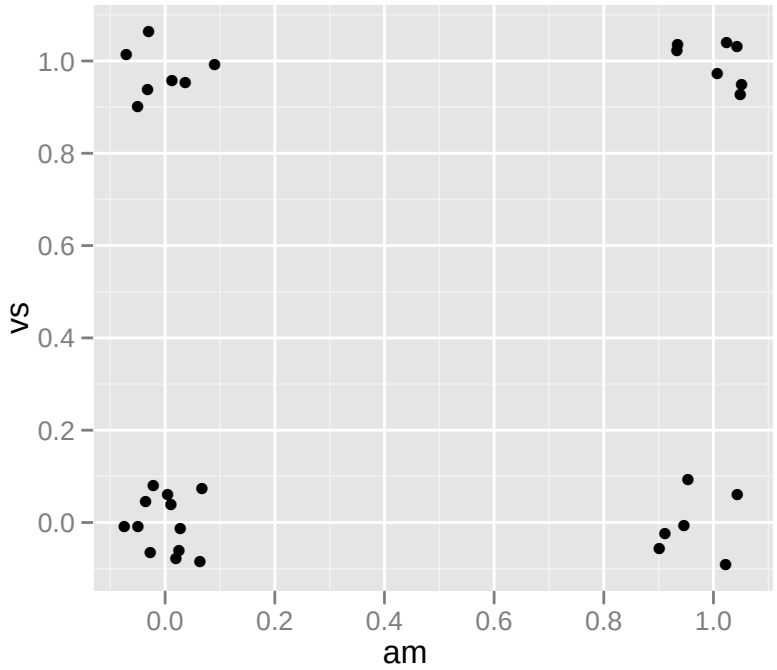


```
> print(qplot(carat, data = diamonds, geom = "density",
facets = cut ~ .))
```

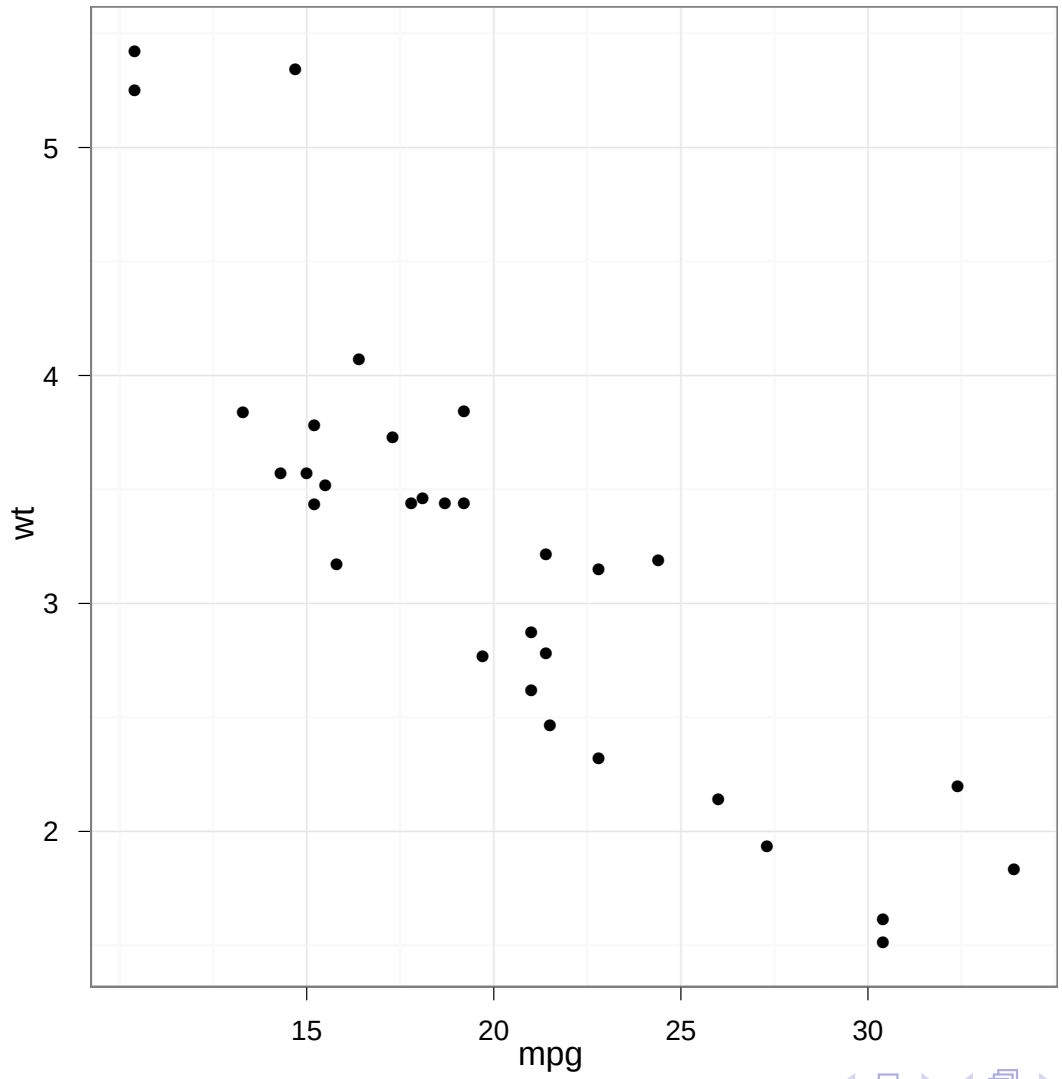


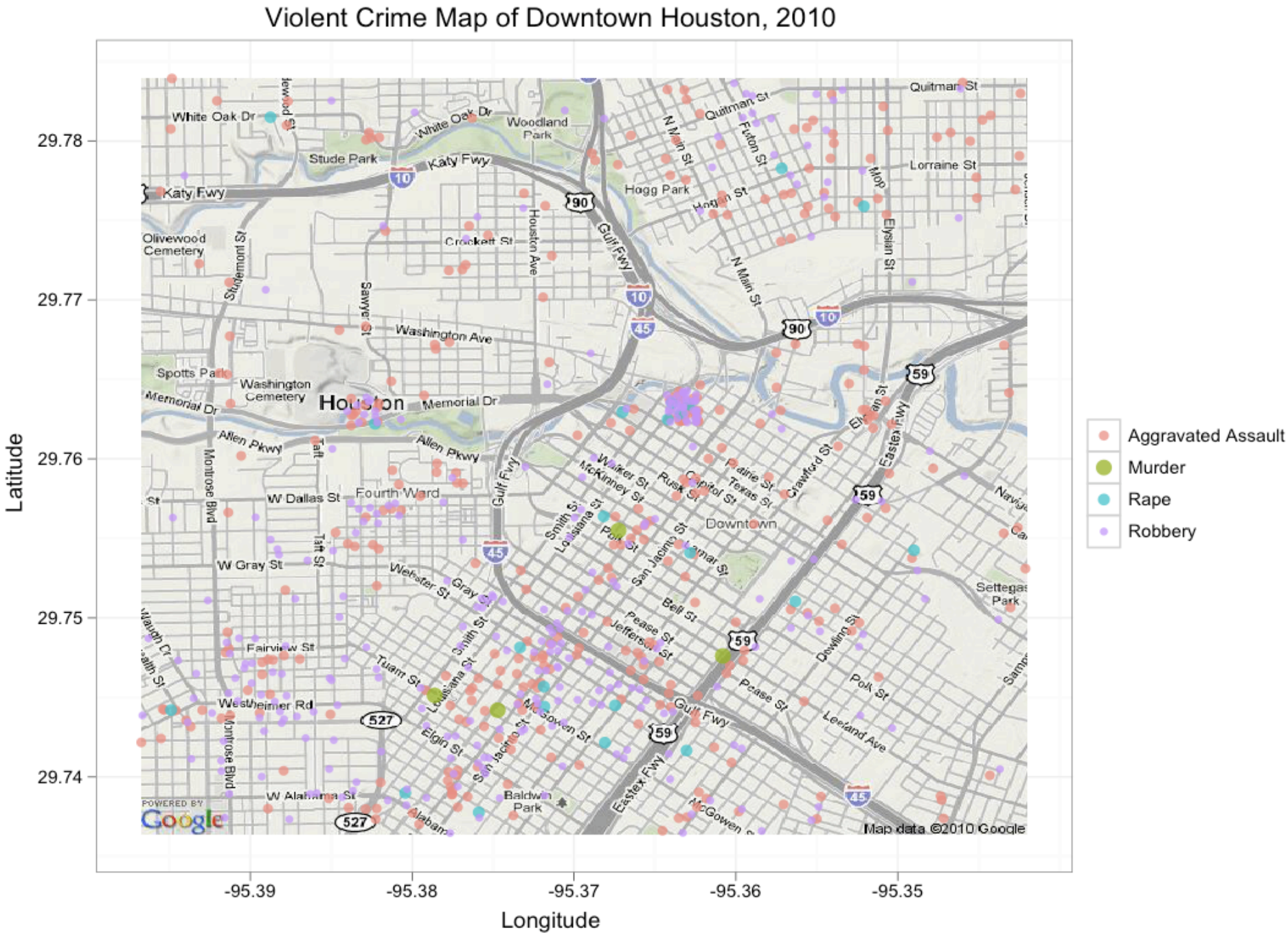
```
> qplot(am, vs, data=mtcars)
> qplot(am, vs, data=mtcars, position=position_jitter(w=0.1,
h=0.1))
```

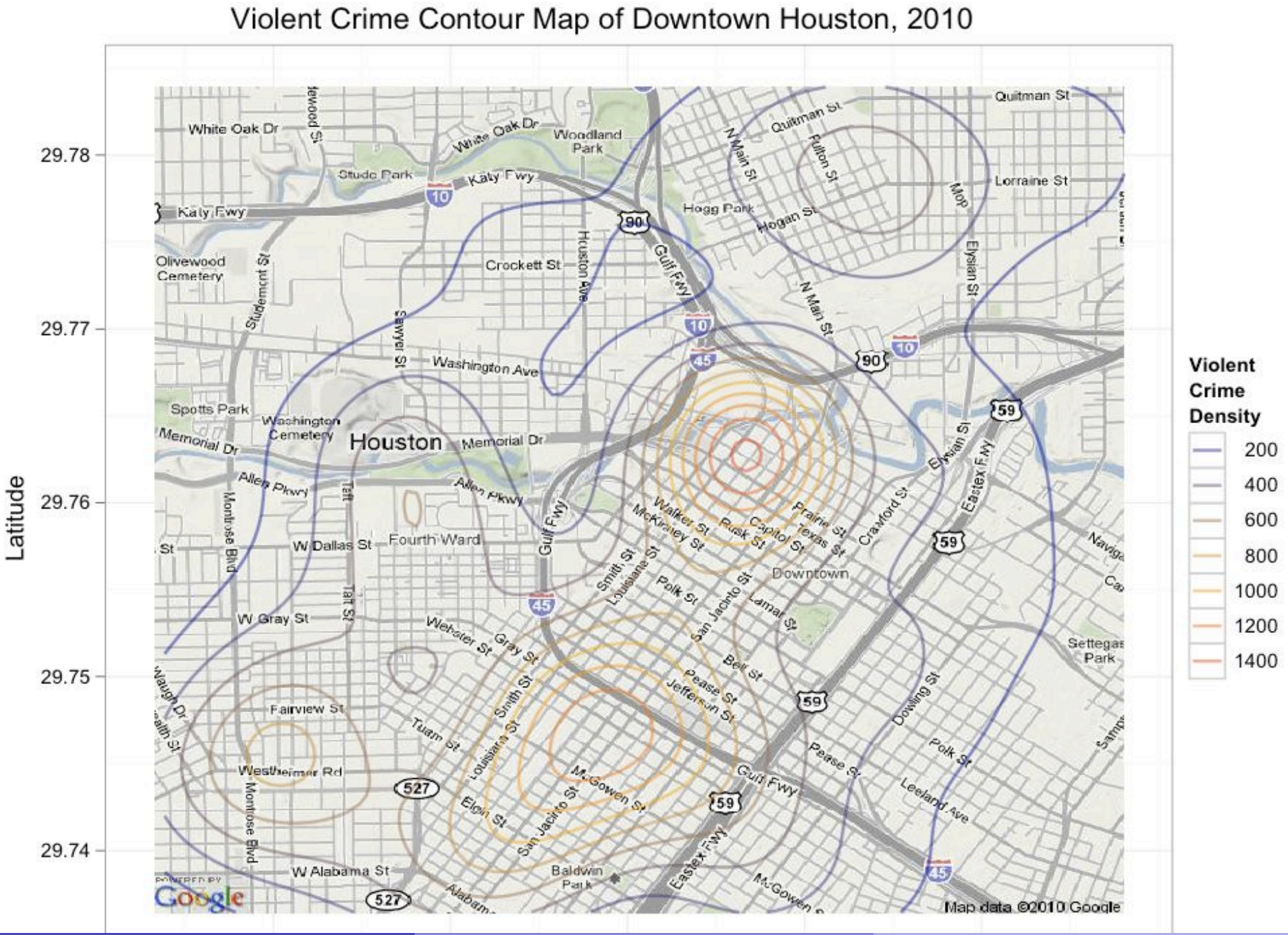




```
> old = theme_set(theme_bw()) # 设置黑白主题
> qplot(mpg, wt, data = mtcars)
```




Violent Crime Map of Downtown Houston, 2010
Latitude
29.78
29.77
29.76
29.75
29.74
-95.39
-95.38
-95.37
-95.36
-95.35
Longitude
Aggravated Assault
Murder
Rape
Robbery
Houston
Downtown
Fourth Ward
Washington Cemetery
Spotts Park
Woodland Park
Hogg Park
Stude Park
Olivewood Cemetery
Baldwin Park
Settegast Park
White Oak Dr
Katy Fwy
Gulf Fwy
Memorial Dr
Allen Pkwy
Washington Ave
W Dallas St
W Gray St
Westheimer Rd
W Alabama St
Crockett St
Quitman St
Lorraine St
Pease St
Polk St
Leeland Ave
Bell St
Webster St
Fairview St
Elgin St
Map data ©2010 Google
POWERED BY Google


Violent Crime Contour Map of Downtown Houston, 2010
Latitude
29.78
29.77
29.76
29.75
29.74
Houston
Downtown
Fourth Ward
Spotts Park
Washington Cemetery
Olivewood Cemetery
Woodland Park
Hogg Park
Studc Park
Baldwin Park
Settegast Park
Katy Fwy
Crockett St
Washington Ave
Memorial Dr
Allen Pkwy
W Dallas St
W Gray St
Fairview St
Westheimer Rd
W Alabama St
Quitman St
Lorraine St
Gulf Fwy
Eastex Fwy
Pease St
Polk St
Bell St
Lamar St
Leeland Ave
Map data ©2010 Google
Violent Crime Density
200
400
600
800
1000
1200
1400

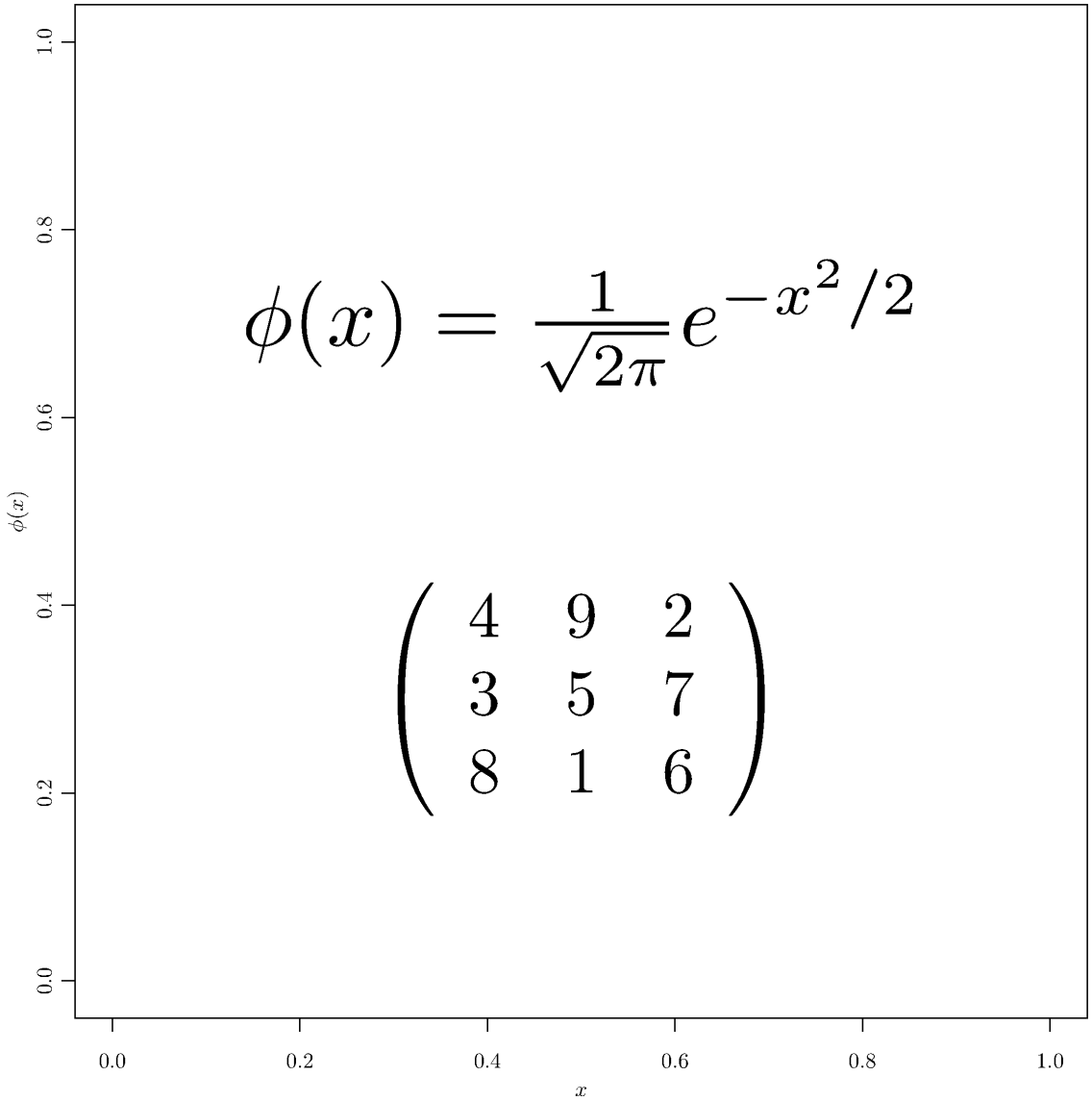
提纲

tikzDevice

- R + tikz + TEX

```
> library(tikzDevice)
> tikz('form.tex',standAlone=TRUE)
> par(mar = c(3, 3, 0.1, 0.1), mgp = c(2, 0.9, 0))
> plot(0:1,0:1, type="n", xlab="$x$", ylab="$\\phi(x)$")
> fo1 <- "$\\phi(x)=\\frac{1}{\\sqrt{2\\pi}}e^{{-x^2}/{2}}$"

> fo2 <- "$\\left(\\begin{array}{ccc}
        4 & 9 & 2 \\\\
        3 & 5 & 7 \\\\
        8 & 1 & 6 \\\\
 \\end{array}\\right)$"
> text(0.5, 0.7, fo1, cex = 4)
> x1 <- grconvertX(0.5,,'device')
> y1 <- grconvertY(0.3,,'device')
> tikzAnnotate(paste('\\node[scale= 3] at
        (',x1,',',y1,') {',fo2,'};'))
> dev.off()
> tools::texi2dvi("form.tex", pdf = TRUE)
> system(paste(getOption("pdfviewer"), "form.pdf"))
```

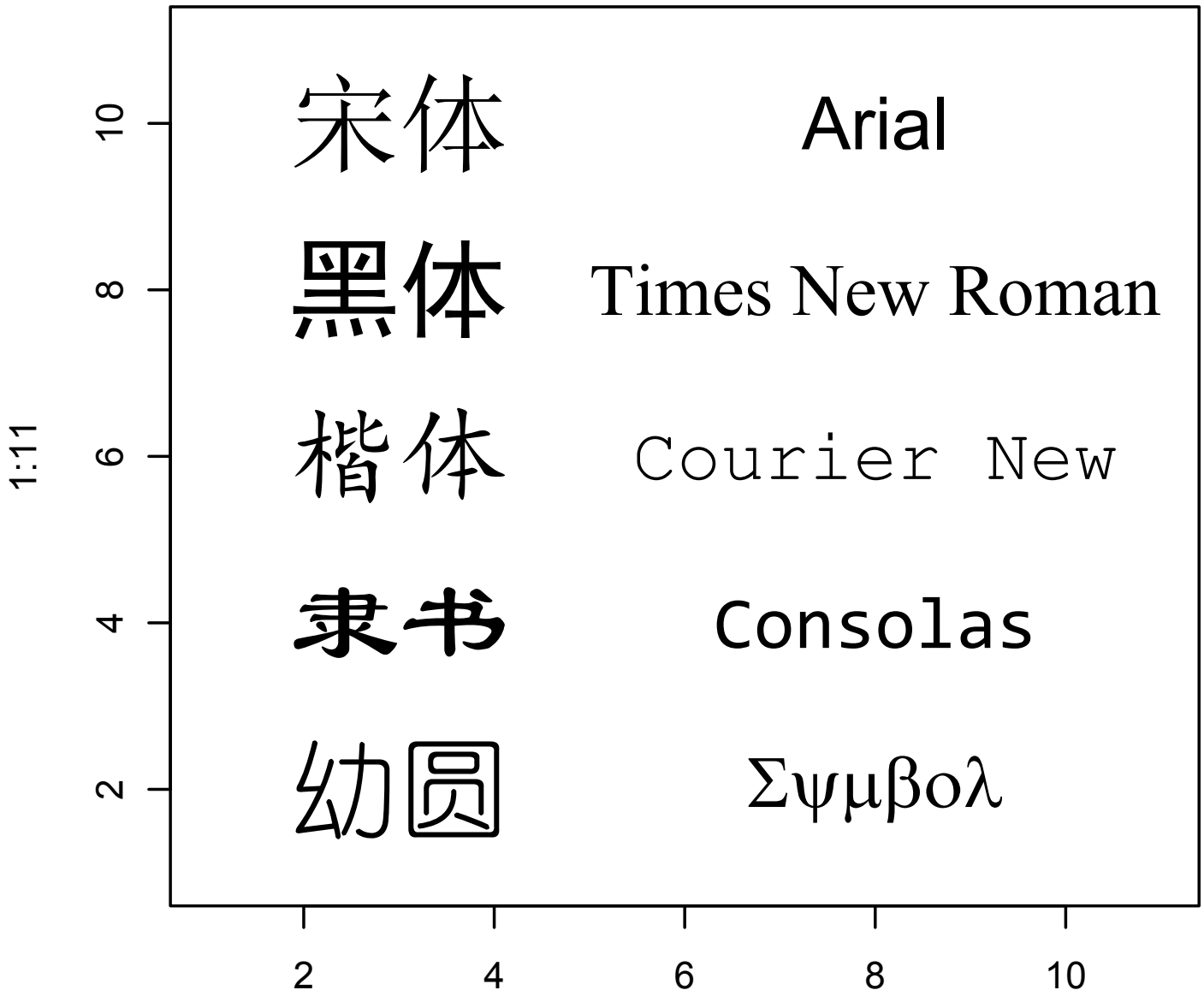

$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
$\begin{pmatrix} 4 & 9 & 2 \\ 3 & 5 & 7 \\ 8 & 1 & 6 \end{pmatrix}$
$\phi(x)$
x
0.0
0.2
0.4
0.6
0.8
1.0

Cairo 系统

- 支持多种输出格式 (SVG、PDF、PNG、PS 等)
- 支持抗锯齿，图形质量高
- 字体可以任意调用

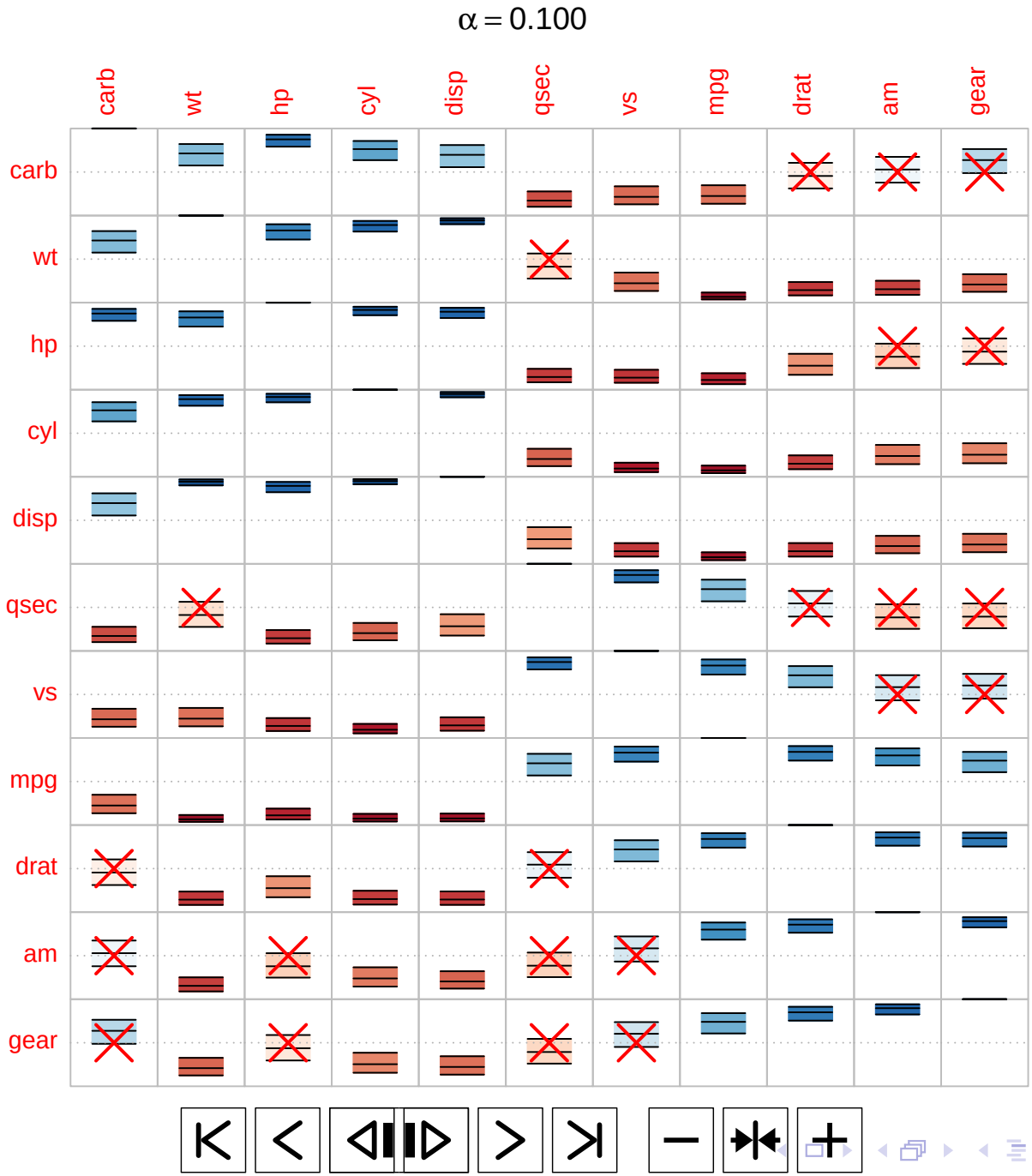
```
> require(Cairo)
> CairoPDF("font.pdf")
> plot(1:11,1:11,type="n")
> text(3,10," 宋体",family="SimSun",cex=3)
> text(3,8," 黑体",family="SimHei",cex=3)
> text(3,6," 楷体",family="KaiTi_GB2312",cex=3)
> text(3,4," 隶书",family="LiSu",cex=3)
> text(3,2," 幼圆",family="YouYuan",cex=3)
> text(8,10,"Arial",family="Arial",cex=2)
> text(8,8,"Times New Roman",family="Times New Roman",cex=2)

> text(8,6,"Courier New",family="Courier New",cex=2)
> text(8,4,"Consolas",family="Consolas",cex=2)
> text(8,2,"Symbol",family="Symbol",cex=2)
> dev.off()
```


宋体
Arial
黑体
Times New Roman
楷体
Courier New
隶书
Consolas
幼圆
Σψμβολ
1:11
2
4
6
8
10
2
4
6
8
10

动画系统

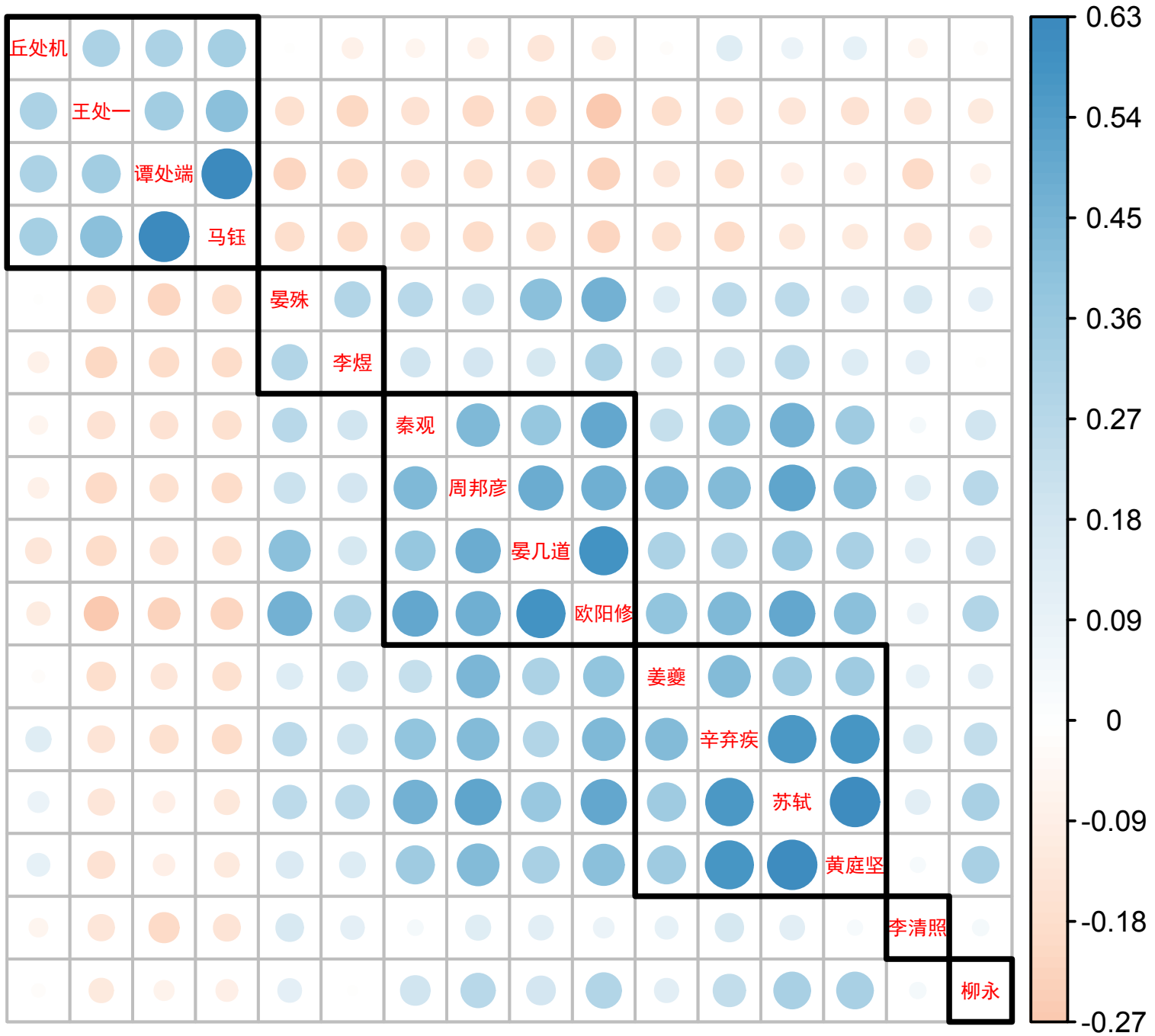
- animation 包
- R2SWF 包


α = 0.100
carb
wt
hp
cyl
disp
qsec
vs
mpg
drat
am
gear

提纲

统计词话

```
> library(MSG)
> library(corrplot)
> ## or download SongWords from http://goo.gl/GzOyT
> SongCorr = cor(SongWords)
> ## 作者词风关系矩阵图
> corrplot(SongCorr, order="hc", diag=FALSE, cl.range="m",
+   addrect=6, addtextlabel="d", tl.cex=0.8)
> ## 关键字网络图
> demo("SongWordsNet", package = "MSG")
```

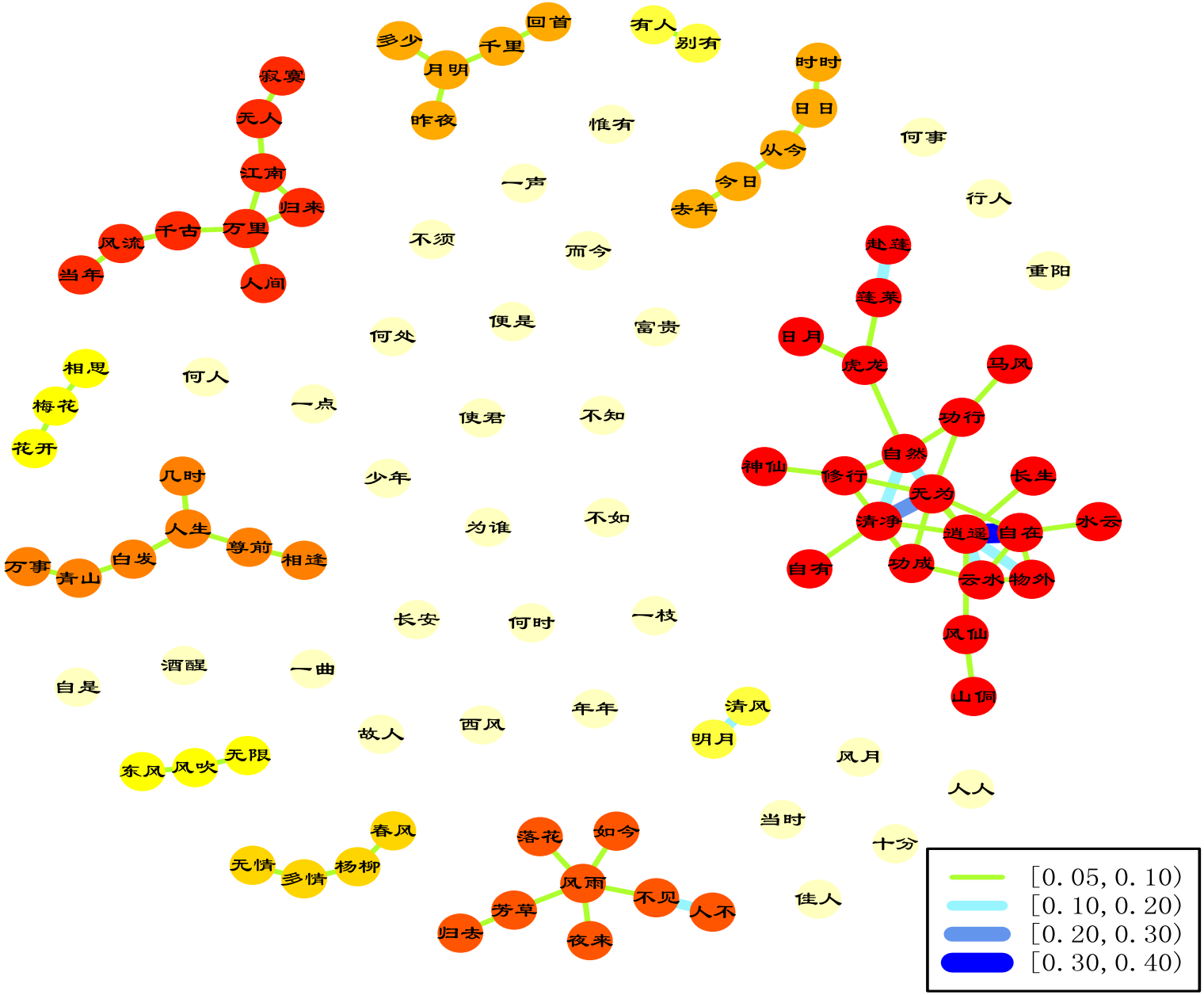

丘处机
王处一
谭处端
马钰
晏殊
李煜
秦观
周邦彦
晏几道
欧阳修
姜夔
辛弃疾
苏轼
黄庭坚
李清照
柳永
0.63
0.54
0.45
0.36
0.27
0.18
0.09
0
-0.09
-0.18
-0.27

词人分类

第一类：马钰、丘处机、谭处端、王处一 (全真七子之四)
第二类：晏殊、李煜
第三类：秦观、周邦彦、欧阳修、晏几道
第四类：姜夔、辛弃疾、黄庭坚、苏轼
第五类：李清照
第六类：柳永

插播：李清照《词论》节选

李煜语虽甚奇，所谓“亡国之音哀以思”也。柳永词虽协音律，而词语尘下。张子野、宋子京兄弟、沈唐、元绛、晁次之辈，虽时时有妙语，而破碎何足名家！晏殊、欧阳修、苏轼 学际天人，然皆句读不茸之诗尔，且常不协音律。王安石、曾巩，文章似西汉，若作一小歌词，则人必绝倒，不可读也。词别是一家，至晏几道、贺铸、秦观、黄庭坚出，始能知之。然晏苦无铺叙；贺苦少重典；秦即专主情致，而少故实；黄即尚故实而多疵病。


回首
多少
千里
月明
昨夜
有人
别有
惟有
时时
日日
从今
今日
去年
何事
一声
行人
寂寞
无人
江南
归来
千古
风流
万里
当年
人间
不须
而今
重阳
赴蓬
蓬莱
便是
何处
富贵
日月
虎龙
马风
相思
梅花
何人
一点
使君
不知
功行
花开
自然
神仙
修行
长生
几时
少年
无为
人生
为谁
不如
清净
逍遥
自在
水云
尊前
相逢
万事
白发
青山
自有
功成
云水
物外
长安
何时
一枝
风仙
酒醒
一曲
山侗
自是
年年
清风
西风
故人
明月
无限
东风
风吹
风月
人人
当时
春风
落花
如今
十分
无情
杨柳
多情
风雨
不见
佳人
芳草
人不
归去
夜来
[0.05, 0.10)
[0.10, 0.20)
[0.20, 0.30)
[0.30, 0.40)

第一类 (21 个节点)：自然、逍遥、物外、无为、蓬莱、修行、清净、山侗、长生、功成、云水、自在、马风、神仙、水云、风仙、自有、日月、赴蓬、功行、虎龙

第二类 (9 个节点)：人间、风流、无人、归来、江南、万里、千古、当年、寂寞

第三类 (8 个节点)：归去、落花、风雨、如今、芳草、不见、人不、夜来

第四类 (7 个节点)：尊前、万事、白发、相逢、人生、青山、几时

第五类 (5 个节点)：千里、多少、回首、月明、昨夜

第六类 (5 个节点)：今日、去年、时时、日日、从今

第七类 (4 个节点)：春风、多情、无情、杨柳

第八类 (3 个节点)：东风、风吹、无限

第九类 (3 个节点)：相思、梅花、花开

第十类 (2 个节点)：明月、清风

参考文献


谢益辉, 现代统计图形 (to appear), 电子工业出版社, 2012.


Hadley Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer, 2009.

谢谢大家！