

```

1   ###第6回基本的なグラフ作成###
2   #ヒストグラム,箱ひげ図,散布図はRの作図が便利#
3   #棒グラフ,帯グラフ,円グラフはEXCELの方が便利#
4   ###層別グラフではlattice関数を使う:引数scales="free"を使うと目盛りが
5   ###不揃いになることがありグラフの見え方が変わるので要注意
6   install.packages("vcd",dependencies = T)#dependencies=Tで関係Packageもinstall
7   install.packages("lattice",dependencies = T)
8   dat <- read.csv("data/cars93.csv" , #cars93.csvの読込
9                 header = T ,
10                row.names = 1 )
11  head(dat , 3)
12  ##量的データ(間隔尺度,比例尺度)
13  library(psych)
14  describe(dat[,c(5,7,12:14,17:22)]) #基本統計量の算出
15  ##層別統計量
16  by(dat[,c(5,7,12:14,17:22)] , #by()関数
17     dat$Origin , #層別変数
18     describe) #適用関数
19  ##数値表記と余白設定##
20  options(scipen=50) #指数表記を実数表記に
21  options(scipen=0) #指数表記する
22  par(mai=c(1.2, 1.2, 1, 1)) #余白設定inc(下,左,上,右)
23  ##ヒストグラム##
24  hist(dat[,c("Price")]) #ヒストグラム:Base Graph
25  hist(dat$"Price" , #グラフィック引数
26       col = "red" , #塗りつぶしの色
27       breaks="Scott", #breaks="Sturges""Scott" or 階級数
28       freq = T, #y軸:度数
29       xlab = "Price",ylab = "度数", #x,y軸ラベル
30       main = "Price Histogram", #タイトル
31       cex.main = 2.0 , cex.lab=1.6) #タイトル,軸ラベル文字サイズ
32  #層別ヒストグラム
33  library(lattice)
34  library(psych)
35  #scales="free":目盛り要注意(出力結果の比較)
36  histogram( ~ Price | Origin , #histogram()関数:~変数|層別変数
37            data=dat,
38            #scales = "free", #scales="free"パネル分割
39            layout=c(2,1), #パネル配列
40            xlab=list(label="Price", #xラベル
41                    cex=1.5,col="red"), #文字サイズ,色
42            ylab=list(label="度数",cex=1.5), #yラベル,文字サイズ
43            main = "Origin別")
44  by(dat$Price , #層別統計量,by()関数:
45     dat$Origin ,
46     describe)
47  #層別(2層)ヒストグラム
48  histogram( ~ Price | Origin + #histogram()関数:~変数|層別変数
49            DriveTrain , # |層別変数+層別変数
50            data=dat,
51            layout=c(2,3), #パネル配列
52            xlab=list(label="Price", #xラベル
53                    cex=1.5,col="red"), #文字サイズ,色
54            ylab=list(label="度数",cex=1.5), #yラベル,文字サイズ
55            #scales = "free", #scales="free"でパネル分割
56            main = "Origin.駆動形式別")
57
58  ##箱ひげ図##
59  boxplot(dat[,c("Price")]) #箱ひげ図:Base Graph
60  grid(lty = 3) #グリッド線(点線)

```

```

61  boxplot(dat$"Price",                                #グラフィック引数
62         xlab = "Price", ylab = "度数",
63         main = "Price Boxplot",
64         cex.main = 2.0 , cex.lab=1.6 , #タイトル, 軸ラベル文字サイズ
65         notch = T)                                #ノッチ: 中央値の95%信頼区間
66  grid(lty = 3)                                    #グリッド線(点線)
67  boxplot.stats(dat$"Price",                         #boxplot.stats() 関数: 箱ひげ図統計量
68             coef=1.5 ,                             #ひげ: 四分位範囲の1.5倍(0:最小, 最大値)
69             do.conf=T ,                             #中央値の95%信頼区間
70             do.out=T)                              #ひげの端点値を超えるデータの値
71  #stats: 下位ひげの最端値, 第1四分位数, 中央値, 第3四分位数, 上位ひげの最端値
72  #層別箱ひげ図
73  bp <- boxplot(Price~Origin,
74              data = dat,
75              main = "Origin別Price",
76              notch = T)                            #ノッチoverlapしていない=群間に有意差
77  grid(lty = 3)
78  colnames(bp$stats) <- bp$names                    #stats: 層別変数名
79  colnames(bp$conf) <- bp$names                    #conf: 層別変数名
80  rownames(bp$conf) <- c("Lower", "Upper")         #conf: 95%信頼区間下端上端値
81  bp                                                #統計量: 42行目からを参照
82
83  #層別箱ひげ図(パネル分割)
84  library(lattice)
85  bwplot(~Price | Origin,
86         data = dat,
87         #scales = "free",                        #scales="free"
88         layout = c(1,2),                        #1列2行
89         main = "Origin別Price",
90         xlab = "Price",
91         cex.lab=1.6 ,
92         notch = T)
93
94  ###質的データ(名義尺度, 順序尺度)###
95  #棒グラフ, 帯グラフ, 円グラフは表計算ソフトでの作図が便利#
96  t.f <- table(dat$Type)                          #table() 関数: 単純集計
97  barplot(t.f)                                    #棒グラフ: Base Graph
98  barplot(t.f,                                    #グラフィック引数
99         col = "red",                             #塗りつぶし色
100        ylim=c(0,30),                             #Y軸目盛り
101        space = 0.2,                               #棒の間隔
102        xlab = "Type",                             #X軸ラベル
103        ylab = "度数",                             #Y軸ラベル
104        main = "Type別棒グラフ",                  #タイトル
105        cex.main=2, cex.lab=1.5,                  #タイトル, ラベル文字サイズ
106        cex.names=1.5, cex.axis=1.5)              #項目, Y軸目盛りサイズ
107  box()                                           #枠線
108
109  ###2変数の関連: 量的データ###
110  round(cor(dat[,c("Price",
111                 "MPG.city",
112                 "EngineSize",
113                 "Fuel.tank.capacity",
114                 "Passengers")],
115         method="p"), 3)                          #method="pearson", kendall, spearman
116  library(psych)
117  cor.plot(cor(dat[,c("Price",
118                    "MPG.city",
119                    "EngineSize",
120                    "Fuel.tank.capacity",

```

```

121                                     "Passengers"]]))
122 ##散布図(回帰直線)##
123 par(mai=c(1.2, 1.2, 1, 1))           #余白指定inc(下,左,上,右)
124 plot(dat$EngineSize,dat$"Price")     #散布図:Base Graph
125 plot(dat$EngineSize,dat$"Price",     #グラフィック引数+回帰直線
126       pch=19,col="red",              #プロット点の形,色
127       xlab = "EngineSize",ylab="Price", #X,Y軸ラベル
128       main="散布図",                  #タイトル
129       cex.main=2,cex.lab=1.7)        #タイトル,ラベル文字サイズ
130 grid(lty=3)                           #パネルグリッド,線のタイプ
131 abline(lm(Price~EngineSize,          #abline()関数:lm()関数の結果を追加
132        data = dat ), lwd = 2)
133 #層別散布図
134 library(lattice)
135 #層ごとに色を変える
136 xyplot(data = dat,Price ~ EngineSize, #層別変数
137         group = Origin,                #層別変数
138         type=c("r","g","p"),          #プロット点の形
139         pch=19,                         #プロット点の形
140         xlab=list(label="EngineSize",  #X軸ラベル
141                   cex=1.5),           #文字サイズ
142         ylab=list(label="Price",       #Y軸ラベル
143                   cex=1.5),           #文字サイズ
144         main=list(label="層別散布図",  #タイトル
145                   cex=2),            #文字サイズ
146         auto.key =                      #凡例
147           list(title="Origin",        #凡例タイトル
148                 space="right",       #凡例位置
149                 cex=1.0)             #凡例文字の大きさ
150 #層別にパネル分割:scales="free"だと回帰直線の傾きが変わって見える
151 xyplot(data=dat,Price~EngineSize|Origin, #Y~X|層別変数
152         type=c("r","g","p"),          #c(r:Reg.line,g:Panel.Grid,p:Dot)
153         group = Origin,
154         xlab=list(label="EngineSize",  #X軸ラベル
155                   cex=2.0),           #文字サイズ
156         ylab=list(label="Price",       #Y軸ラベル
157                   cex=2.0),           #文字サイズ
158         main=list(label="層別散布図",  #タイトル
159                   cex=2),            #文字サイズ
160         #scales = "free",              #scales="free"
161         layout = c(2,1),               #パネルレイアウト
162         pch=19,cex=1.2,                #プロット点の形と大きさ
163         lwd=2)                          #回帰直線の幅
164 #層別(2層)にパネル分割
165 xyplot(data=dat,Price~EngineSize|Origin+ #Y~X|層別変数+層別変数
166         DriveTrain,                    #c(r:Reg.line,g:Panel.Grid,p:Dot)
167         type=c("r","g","p"),
168         group = Origin,
169         xlab=list(label="EngineSize",  #X軸ラベル
170                   cex=2.0),           #文字サイズ
171         ylab=list(label="Price",       #Y軸ラベル
172                   cex=2.0),           #文字サイズ
173         main=list(label="層別散布図",  #タイトル
174                   cex=2),            #文字サイズ
175         #scales = "free",              #scales="free":スケール不揃い
176         layout = c(2, 3),             #パネルレイアウト
177         pch=19,cex=1.2,                #プロット点の形と大きさ
178         lwd=2)                          #回帰直線の幅
179
180 ###順序尺度(評定尺度:間隔尺度と見なす場合がある)の2次元表現###

```

```

181  ##library(psych)に格納されているbfiデータを使う
182  #因子に対応した質問項目A1-A5, C1-C5, E1-E5, N1-N5, O1-O5, 性, 教育歴, 年齢
183  library(psych)
184  dat <- bfi
185  head(dat)
186  #dataは6段階の評定尺度(通常はクロス集計表で分析)#
187  #Jitter(点プロットグラフ)を使うと視覚的にクロス表を表現できる
188  table.s <- table(dat$N1,      #table()関数,引数(表側,表頭)
189                  dat$N2)
190  table.s      #plot()関数では度数を反映した散布図にならない
191  cor(dat$N1,dat$N2,      #相関係数:N1;怒りっぽい,N2;イライラしやすい
192       use = "complete") #NAがあるため
193  plot(jitter(bfi$N1),    #base Graph(x,y)#Jitterで表現すると
194       jitter(bfi$N2))
195  par(mai=c(1.2, 1.2, 1, 1))
196  plot(jitter(bfi$N1,2),  #jitter:引数2はデータの散布程度
197       jitter(bfi$N2,2), #formula:(x,y),(y~x)
198       xlab = "Get angry easily",
199       ylab = "Get imitated easily",
200       main = "N1/N2 Jitter Graf",
201       cex.main=2,cex.lab=1.5)
202
203  ####2変数の関連:質的データ###
204  dat <- read.csv("data/cars93.csv" , #cars93.csvの読み込
205                header = T ,
206                row.names = 1 )
207  mosaicplot(~ AirBags + Origin,      #mosaicplot()関数:~表頭+表側
208            data = dat,                #data名
209            cex.axis = 1.1,            #Category文字サイズ
210            xlab="Air Bag",
211            ylab="Origin",
212            main = "Mosaic Plot",      #タイトル
213            shad=T)                    #shad=T:χ2値の標準化残差
214  #残差2以下白,2~4:負の値赤,正の値ブルー
215  library(vcd)
216  mosaic(~ DriveTrain + Origin +      #mosaicplot()関数:3元クロス表の場合
217         Man.trans.avail,
218         data = dat,
219         main = "Mosaic Plot",
220         shade = T)
221  table.3 <- xtabs(~ DriveTrain + Origin+ #多元クロス表の作成
222                Man.trans.avail,        #次元数をあげる場合+で追加
223                data = dat)
224  pairs(table.3,                      #pairst()関数:モザイクPマトリックス
225        main = "Mosaic Plot",
226        shad=T, labels)                #labels:下三角に
227
228  assoc(table.3,shade=T) #assoc()関数:associationグラフ,矩形の高さは残差,
229  pairs(table.3,          #幅は期待度数の平方根,面積は観測度数と期待度数の差
230        upper_panel = pairs_assoc, #上側三角:associationグラフ
231        shade=T)
232

```