Cheatography

Exploratory data analysis in R Cheat Sheet

by xeonkai via cheatography.com/31513/cs/9762/

Barplot

`barplo t(h eight)` where `height` is vector or matrix

Options: `horiz=TRUE`, `main`, `xlab`, `ylab`, `names.arg`

If `height` is matrix, stacked and grouped bar plot is produced

Options: `beside =TRUE`, `col`, `legend`

If `height` is factor or ordered factor, use `plot()` function

Spinograms

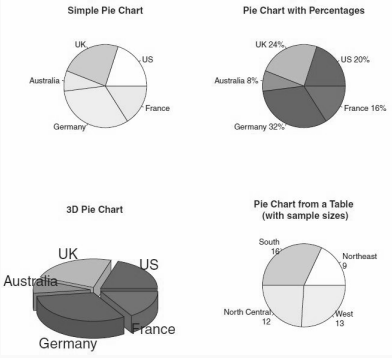
`spine()` function in `vcd` package

Pie charts

```
par(mfrow=c(2, 2))
slices <- c(10, 12,4, 16, 8)
lbls <- c("US", "UK", "Australia", "Germany", "France")

pie(slices, labels = lbls,
    main="Simple Pie Chart")

pct <- round(slices/sum(slices)*100)
lbls2 <- paste(lbls, " ", pct, "%", sep="")
pie(slices, labels=lbls2, col=rainbow(length(lbls2)),
    main="Pie Chart with Percentages")

library(plotrix)
pie3D(slices, labels=lbls,explode=0.1,
      main="3D Pie Chart ")

mytable <- table(state.region)
lbls3 <- paste(names(mytable), "\n", mytable, sep="")
pie(mytable, labels = lbls3,
    main="Pie Chart from a Table\n (with sample sizes)")
```

1. Combine four graphs into one
2. Add percentages to pie chart
3. Create chart from table

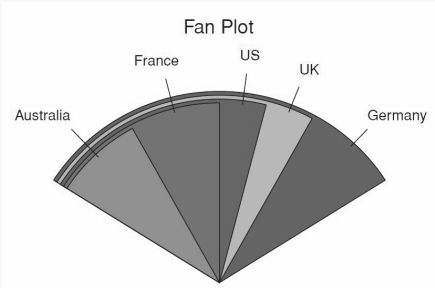
Pie charts



Fan plot

`fan.plot()` in `plotrix` package

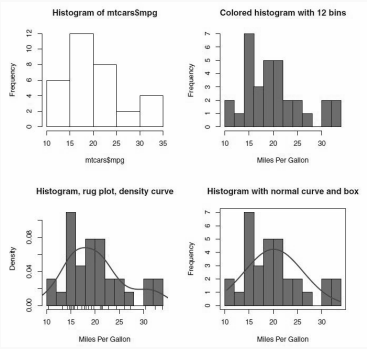
Fan plot



Histogram

```
par(mfrow=c(2,2))

hist(mtcars$mpg)                          <- 1 Simple histogram

hist(mtcars$mpg,                          <- 2 With specified bins and color
     breaks=12,
     col="red",
     xlab="Miles Per Gallon",
     main="Colored histogram with 12 bins")

hist(mtcars$mpg,                          <- 3 With rug plot and frame
     freq=FALSE,
     breaks=12,
     col="red",
     xlab="Miles Per Gallon",
     main="Histogram, rug plot, density curve")
rug(jitter(mtcars$mpg))
lines(density(mtcars$mpg), col="blue", lwd=2)

x <- mtcars$mpg                           <- 4 With normal curve
h<-hist(x,
        breaks=12,
        col="red",
        xlab="Miles Per Gallon",
        main="Histogram with normal curve and box")
xfit<-seq(min(x), max(x), length=40)
yfit<-dnorm(xfit, mean=mean(x), sd=sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue", lwd=2)
box()
```

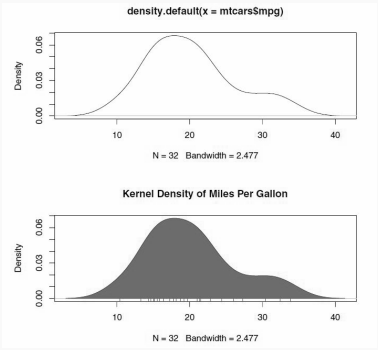
Histogram



Kernel density plot

```
par(mfrow=c(2,1))
d <- density(mtcars$mpg)

plot(d)

d <- density(mtcars$mpg)
plot(d, main="Kernel Density of Miles Per Gallon")
polygon(d, col="red", border="blue")
rug(mtcars$mpg, col="brown")
```

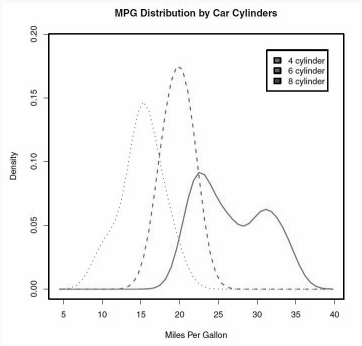
Kernel density plot



Kernel density plot

```
par(lwd=2)                                    <- 1 Double width of plotted lines
library(sm)
attach(mtcars)

cyl.f <- factor(cyl, levels= c(4,6,8),        <- 2 Create grouping factor
                labels = c("4 cylinder", "6 cylinder",
                           "8 cylinder"))

sm.density.compare(mpg, cyl, xlab="Miles Per Gallon")   <- 3 Plot densities
title(main="MPG Distribution by Car Cylinders")

colfill<-c(2:(1+length(levels(cyl.f))))       <- 4 Add legend via mouse click
legend(locator(1), levels(cyl.f), fill=colfill)

detach(mtcars)
```

Kernel density plot



Sponsored by **CrosswordCheats.com**
Learn to solve cryptic crosswords!
http://crosswordcheats.com

Cheatography

Exploratory data analysis in R Cheat Sheet

by xeonkai via cheatography.com/31513/cs/9762/

Box plots

Five-number summary

> Minimum, lower quartile (25th percentile), median (50th percentile), upper quartile (75th percentile), maximum

Outliers

> Values outside the range of +- 1.5*IQR

Example

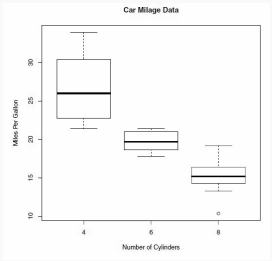
```
boxplo t(m tca rs$mpg,
main="Box plot", ylab="Miles
per Galleo ns")
```

```
boxplo t.s tat s(m tca rs$mpg)
```

Parallel box plots for comparison

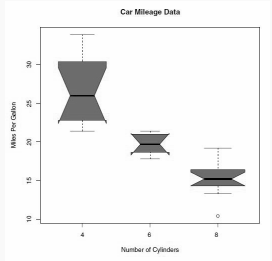
```
boxplo t(f ormula, data=d -
ata frame)
```

Box plots




```
boxplo t(mpg ~ cyl, data=m -
tcars, main="Car Mileage Data",
xlab="N umber of Cylind ers ",
ylab="Miles per Gallon ")
```

Box plots

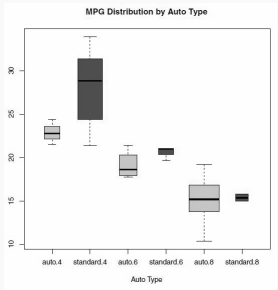



```
boxplo t(mpg ~ cyl, data=m -
tcars, notch= TRUE, varwid th= -
TRUE, col="re d", main="Car
Mileage Data", xlab="N umber of
Cylind ers ", ylab="Miles per
Gallon ")
```

Box plots (2 crossed factors)

```
mtcars$cyl.f <- factor(mtcars$cyl,
                       levels=c(4,6,8),
                       labels=c("4","6","8"))      <- Create factor for # of cylinders

mtcars$am.f <- factor(mtcars$am,
                      levels=c(0,1),
                      labels=c("auto", "standard"))   <- Create factor for transmission type

boxplot(mpg ~ am.f *cyl.f,                        <- Generate box plot
        data=mtcars,
        varwidth=TRUE,
        col=c("gold","darkgreen"),
        main="MPG Distribution by Auto Type",
        xlab="Auto Type")
```

Box plots (2 crossed factors)

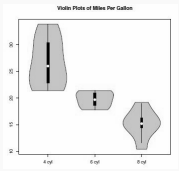


Violin plots

Violin plots are kernel density plots superimposed in a mirror image fashion over box plots

```
vioplo t(x1, x2, … , names=,
col=)
```

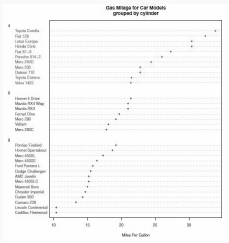
Violin plots



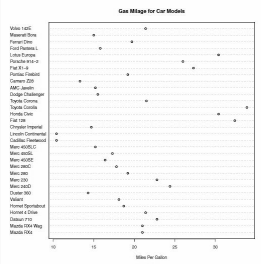

```
librar y(v ioplot)
x1 <- mtcars $mp g[m tca rs$ -
cyl==4]
x2 <- mtcars $mp g[m tca rs$ -
cyl==6]
x3 <- mtcars $mp g[m tca rs$ -
cyl==8]
vioplo t(x1, x2, x3, names= c("4
cyl", "6 cyl", "8 cyl"),
col="go ld")
title( " Violin Plots of Miles
Per Gallon ")
```

Dot plots

Dot plots - grouped, sorted and colored




```
x <- mtcars [or der (mt car -
s$m pg),]
x$cyl <- factor (x$cyl)
x$colo r[x $cy l==4] <- " red "
x$colo r[x $cy l==6] <- " blu e"
x$colo r[x $cy l==8] <- " dar -
kgr een "
dotcha rt( x$mpg, labels =
row.na mes(x), cex=.7, groups =
x$cyl, gcolor = " bla ck", color
= x$color, pch=19, main = "Gas
Mileage for Car Models \ng -
rouped by cylind er", xlab = " -
Miles Per Gallon ")
```




```
dotcha rt( mtc ars $mpg,
labels =ro w.n ame s(m tcars),
cex=.7, main="Gas Mileage for
Car Models ", xlab="Miles Per
Gallon ")
```


Sponsored by **CrosswordCheats.com**
Learn to solve cryptic crosswords!
http://crosswordcheats.com