

## R Cheat sheet

```
x=1:10
x=seq(3, 20, by=4)
x=c(1, 99, 33)
x=c("Monday", "Tuesday", "Friday")
x=rep(2, 10)

y=length(x); y=y+1

x=function(p1, p2=2) { return(p1+p2) }
for (i in 1:10) print(c(i, i^2))

now=as.Date("22-06-2024", format="%d-%m-%Y") # ?strftime lists format options

x = data.frame(num=c(1, 2, 3, 4), name=c("b", "g", "c", "d"))
x$val=runif(nrow(x))

nrow(x); ncol(x) # number of rows and number of columns

str(x) # display first few rows of x
head(x$val, n=2); tail(x$val, n=3)

sort(x$name); order(x$name) # actually sort, and list sort order

plot(x$num, x$val)
plot( ~ num+val+name, data=x)

x_mod=glm(val ~ num, data=x) # build regression model
x_mod=glm(val ~ num+I(num^2), data=x) # build quadratic regression model
summary(x_mod) # summarise model
predict(x_mod) # predict values based on existing measurements
predict(x_mod, newdata=data.frame(num=-5:10)) # Specify point to predict at

# include standard error in prediction
x_p=predict(x_mod, se.fit=TRUE)
lines(x$num, x_p$fit, col="red") # show fitted points

# show two confidence bounds at 2.5% above and below the fit
lines(x$num, x_p$fit+qnorm(0.975)*x_p$se.fit, col="green")
lines(x$num, x_p$fit+qnorm(0.025)*x_p$se.fit, col="green")

y=subset(x, num > 2) # Extract subset of dataframe

install.packages("boot") # install a package
library("boot") # load a package

mean(x$val); median(x$val); sd(x$val)
sum(x$val); min(x$val); max(x$val); range(x$val)

data=read.csv("measurements.csv.xz") # default is comma separator
data=read.csv("measurements.csv", sep="|") # use | as column separator

plot(xvals, yvals)
plot(yvals) # equivalent to plot(1:length(yvals), yvals)
plot(xvals, yvals, log="y")
plot(xvals, yvals, xlim=c(1, 10))
plot(xvals, yvals, xlab="This is X", ylab="Y here")

plot(xvals, yvals, pch="+", col="pink")
points(more_x, more_y, col="green")
lines(x_points, y_points, col="blue")

loess_mod=loess(y_axis ~ x_axis, span=0.3)
loess_pred=predict(loess_mod)
lines(x_axis, loess_pred, col="blue")
```