CheatSheet - Model Development

Command	Syntax	Description	Example
install package	`install.packages("packagename")`	`install.packages` is used to install the packages from the R library.	`install.packages("tidyverse")`
load package	`library(packagename)`	`library()` Load the package from R library.	`library(tidyverse)`
download.file	`download.file(url, destfile, method, quiet = FALSE, mode = "w",cacheOK = TRUE,headers = NULL, …)`	`download.file()` to download the file locally using the download.file() function.	`download.file(url, destfile = "lax_to_jfk.tar.gz")`
untar	`untar()`	`untar()` is used to extract files from a tar archive is done with untar function from the utils package.	`untar("lax_to_jfk.tar.gz")`
Simple Linear Regression			
Linear Model Function	`lm(formula, data, subset, weights, na.action, method = "qr", model = TRUE, x = FALSE, y = FALSE, qr = TRUE, singular.ok = TRUE, contrasts = NULL, offset, …)`	`lm()` is used to fit linear models. It can be used to carry out regression, single stratum analysis of variance and analysis of covariance (although aov may provide a more convenient interface for these). formula an object of class "formula" a symbolic description of the model to be fitted. na.action a function which indicates what should happen when the data contain NAs. method the method to be used; for fitting, currently only method = "qr" is supported; method = "model.frame" returns the model frame (the same as with model = TRUE, see below). model, x, y, qr logicals. If TRUE the corresponding components of the fit (the model frame, the model matrix, the response, the QR decomposition) are returned. singular If FALSE (the default in S but not in R) a singular fit is an error.	`lm(arrdelayminutes ~ depdelayminutes, data = aa_delays)`

filter	`filter()`	`filter()` function screens out observations based on values.	`filter(carrierDelay != "na", reporting_airline == "aa")`
head	`head(x)`	`head(x)` function returns the first part of a vector, matrix, table, data frame or function.	`head(aa_delays)`
summary	`summary(model)`	`summary()` function is a generic function used to produce result summaries of the results of various model fitting functions.	`summary(linear_model)`
data.frame	`data.frame(object)`	The function data.frame() creates data frames, tightly coupled collections of variables which share many of the properties of matrices and of lists.	`data.frame(depdelayminutes = c(12, 19, 24))`
predict	`predict(object..)`	The predict() function in R is used to predict the values based on the input data.	`predict(linear_model, newdata = new_depdelay, interval = "confidence")`
Multiple Linear Regression			
MLR model Function	`lm(y ~ x1+x2+x3...,data)`	In multiple regression we build a model having more than one predictor variable and one response variable.	`lm(arrdelayminutes ~ depdelayminutes + lateaircraftdelay, data = aa_delays)`
$ (dollar symbol)	`df$object`	The `$` operator is used to extract or subset a specific part of a data object.	`mlr$coefficients`
Assessing Models Visually			
ggplot	`ggplot(df, aes(x, y, other aesthetics))`	ggplot is a plotting package that makes it simple to create complex plots from data in a data frame. data Default dataset to use for plot. If not already a data.frame, will be converted to one by fortify(). If not specified, must be supplied in each layer added to the plot. mapping Default list of aesthetic mappings to use for plot. If not specified, must be supplied in each layer added to the plot. ... Other arguments passed on to methods. Not currently used. environment DEPRECATED. Used prior to tidy evaluation.	`ggplot(aa_delays, aes(x = depdelayminutes, y = arrdelayminutes))`

geom_point	`geom_point()`	The function geom_point() adds a layer of points to your plot, which creates a scatterplot.	`ggplot(data=null,aes(x , noisy.y)) + geom_point() + geom_smooth(method = "lm")`
geom_smooth	`geom_smooth(objects...)`	geom_smooth() for adding smoothed conditional means / regression line.	`ggplot(data=null,aes(x , noisy.y)) + geom_point() + geom_smooth(method = "lm")`
geom_segment	`geom_segment( mapping = NULL, data = NULL,...)`	geom_segment() draws a straight line between points (x, y) and (xend, yend).	`geom_segment(aes(xend = depdelayminutes, yend = predicted), alpha = .2)`
theme_bw	`theme_bw(base_size = 12, base_family = "")`	A theme with white background and black gridlines.	`ggplot(data=null,aes(x , noisy.y)) + geom_point() + geom_smooth(method = "lm") + theme_bw()`
cor	`cor(object)`	cor() computes the correlation coefficient.	`cor(aa_delays$depdelay minutes, aa_delays$arrdelayminu tes)`
Polynomial Regression			
Polynomial regression function	`lm(y ~ poly(x, degree, raw = true))`	Polynomial Regression is a form of linear regression in which the relationship between the independent variable x and dependent variable y is modeled as an nth degree polynomial.	`lm(temp ~ poly(time, 4, raw = true))`
Assessing the Model			
R-squared	`r.squared(object, model = NULL, type = c("cor", "rss", "ess"), dfcor = FALSE)`	`r.squared()` computes R squared or adjusted R squared for plm objects. It allows to define on which transformation of the data the (adjusted) R squared is to be computed and which method for calculation is used. object an object of class plm, model on which transformation of the data the R-squared is to be computed. I type indicates method which is used to compute R squared. dfcor if TRUE, the adjusted R squared is computed.	`summary(linear_model)$ r.squared`
Mean Squared Error (MSE)	`mean(x, …)`	`mean()` compute the mean squared error regression loss.	`mean(linear_model$resi duals^2)`

Author(s)

D.M. Naidu

Changelog

Date	Version	Changed by	Change Description
2020-08-11	1.0	D.M. Naidu	Initial Version