Online Supplementary Material

Calculating the relative importance of multiple regression predictor variables using dominance analysis and random forests

Contents

Note

The figure labels in the pdf version of this supplemental materials are not displaying correctly when viewed through OSF's online reader. To see the figure labels, please download the pdf file and open it on your computer.

Environment

```
sessionInfo()
```

```
## R version 4.2.0 (2022-04-22)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur/Monterey 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRlapack.dylib
##
## locale:
## [1] ja_JP.UTF-8/ja_JP.UTF-8/ja_JP.UTF-8/C/ja_JP.UTF-8/ja_JP.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] compiler_4.2.0  magrittr_2.0.3  fastmap_1.1.0   cli_3.3.0
##  [5] tools_4.2.0     htmltools_0.5.2 rstudioapi_0.13 yaml_2.3.5
##  [9] stringi_1.7.6   rmarkdown_2.14  knitr_1.39      stringr_1.4.0
## [13] xfun_0.30       digest_0.6.29   rlang_1.0.2     evaluate_0.15
```

Note: Run all code to reproduce the results

Install Required Packages

```
list.of.packages <- c("remote","MASS","ppcor","lavaan","semPlot","glmnet",
                      "dominanceanalysis","relaimpo","yhat","boot","ggplot2",
                      "randomForest","Boruta","car","kernlab","xgboost","caret")
new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages, repos = "http://cran.us.r-project.org")
```

Install rpsych from archive

This is necessary because these packages are not updated for the current version of R.

```
library(remotes)
install_version("rpsychi", "0.8", repos = "http://cran.us.r-project.org")
```

```
## Downloading package from url: http://cran.us.r-project.org/src/contrib/Archive/rpsychi/rpsychi_0.8.ta
```

```
install_version("dominanceanalysis", "2.0.0", repos = "http://cran.us.r-project.org")
```

```
## Downloading package from url: http://cran.us.r-project.org/src/contrib/Archive/dominanceanalysis/dom
```

Table 1

```
correl <- matrix(c(
  1, 0.62, 0.43, 0.56, 0.61,
  0.62, 1, 0.67, 0.55, 0.72,
  0.43, 0.67, 1, 0.67, 0.65,
  0.56, 0.55, 0.67, 1, 0.79,
  0.61, 0.72, 0.65, 0.79, 1),
  nrow=5,
  dimnames=list(c("Speaking","Vocabulary","Grammar","Writing","Reading"),
                c("Speaking","Vocabulary","Grammar","Writing","Reading")))

# Multiple Regression Analysis
library(rpsychi)
multreg.second(Speaking ~ Vocabulary+Grammar+Writing+Reading, corr=correl, n=100)
```

```
## $corr.partial.corr
##            Speaking Vocabulary Grammar Writing Reading
## Speaking       1.00      0.375  -0.167   0.235   0.106
## Vocabulary     0.62      1.000   0.449  -0.267   0.429
## Grammar        0.43      0.670   1.000   0.406   0.010
## Writing        0.56      0.550   0.670   1.000   0.585
## Reading        0.61      0.720   0.650   0.790   1.000
##
## $corr.confidence
##            Speaking Vocabulary Grammar Writing Reading
## Speaking      1.000      0.728   0.578   0.681   0.720
## Vocabulary    0.482      1.000   0.766   0.674   0.803
## Grammar       0.255      0.545   1.000   0.766   0.751
## Writing       0.409      0.396   0.545   1.000   0.854
## Reading       0.470      0.610   0.520   0.703   1.000
##
## $omnibus.es
##   Rsq lower upper
```

```
## 0.474 0.309 0.568
##
## $standardized.estimates
##            estimates  lower upper   std
## Vocabulary     0.467  0.235 0.699 0.118
## Grammar       -0.189 -0.414 0.035 0.115
## Writing        0.310  0.052 0.568 0.131
## Reading        0.152 -0.134 0.438 0.146
##
## $power
##  small medium  large
##  0.165  0.874  0.999
```

```
# Creating Simulated Data
library(MASS)
set.seed(88)
mu <- rep(c(0), times = length(colnames(correl)))
simdat <- mvrnorm(n=100, mu=mu, Sigma=correl, empirical=TRUE)
colnames(simdat) <- c("Speaking","Vocabulary","Grammar","Writing","Reading")
cor(simdat)
```

```
##            Speaking Vocabulary Grammar Writing Reading
## Speaking       1.00       0.62    0.43    0.56    0.61
## Vocabulary     0.62       1.00    0.67    0.55    0.72
## Grammar        0.43       0.67    1.00    0.67    0.65
## Writing        0.56       0.55    0.67    1.00    0.79
## Reading        0.61       0.72    0.65    0.79    1.00
```

```
# Calculating p-values
z <- scale(as.data.frame(simdat))
z <- data.frame(z)
summary(lm(Speaking~ ., z))
```

```
##
## Call:
## lm(formula = Speaking ~ ., data = z)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.66226 -0.40085 -0.04744  0.49712  1.64736
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.329e-16  7.401e-02   0.000 1.000000
## Vocabulary   4.669e-01  1.184e-01   3.944 0.000153 ***
## Grammar     -1.893e-01  1.147e-01  -1.651 0.102087
## Writing      3.099e-01  1.314e-01   2.358 0.020411 *
## Reading      1.520e-01  1.462e-01   1.040 0.300906
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7401 on 95 degrees of freedom
## Multiple R-squared:  0.4744, Adjusted R-squared:  0.4523
## F-statistic: 21.43 on 4 and 95 DF,  p-value: 1.269e-12
```

```r
# Partial and Semipartial Correlation (Not in the Paper)
library(ppcor)
pcor(simdat)$estimate # Partial Correlation
```

```
##              Speaking Vocabulary      Grammar    Writing     Reading
## Speaking    1.0000000  0.3750810 -0.166987688  0.2351664 0.106115497
## Vocabulary  0.3750810  1.0000000  0.449087023 -0.2671930 0.429121732
## Grammar    -0.1669877  0.4490870  1.000000000  0.4056505 0.009673428
## Writing     0.2351664 -0.2671930  0.405650540  1.0000000 0.585148738
## Reading     0.1061155  0.4291217  0.009673428  0.5851487 1.000000000
```

```r
spcor(simdat)$estimate # Semipartial Correlation
```

```
##               Speaking Vocabulary      Grammar    Writing     Reading
## Speaking    1.00000000  0.2933487 -0.122789353  0.1754140 0.077370068
## Vocabulary  0.23565308  1.0000000  0.292729085 -0.1614853 0.276693048
## Grammar    -0.10831644  0.3214479  1.000000000  0.2838319 0.006186849
## Writing     0.13309847 -0.1525292  0.244138597  1.0000000 0.396943833
## Reading     0.05400663  0.2404271  0.004895646  0.3651693 1.000000000
```

```r
# Path Tracing (Just for Reference)
# Speaking and Vocabulary
0.47+(-0.19*0.67)+(0.31*0.55)+(0.15*0.72)
```

```
## [1] 0.6212
```

```r
# Speaking and Grammer
-0.19+(0.47*0.67)+(0.31*0.67)+(0.15*0.65)
```

```
## [1] 0.4301
```

```r
# Speaking and Writing
0.31+(0.47*0.55)+(-0.19*0.67)+(0.15*0.79)
```

```
## [1] 0.5597
```

```r
# Speaking and Writing
0.15+(0.47*0.72)+(-0.19*0.65)+(0.31*0.79)
```

```
## [1] 0.6098
```

Figure 1 (Quite Literally)

Oswald, F. L. (2021, June 19). Regression illustrated.jpg. *Fred Oswald's Quick Files.* https://osf.io/adnj2

```r
correl <- matrix(c(
  1, 0.5, 0.4, 0.4,
  0.5, 1, 0.2, 0.5,
  0.4, 0.2, 1, 0.3,
  0.4, 0.5, 0.3, 1),
  nrow=4,
  dimnames=list(c("X1","X2","X3","Y"),
                c("X1","X2","X3","Y")))

# Multiple Regression Analysis
library(rpsychi)
multreg.second(Y~ X1+X2+X3, corr=correl, n=100)
```

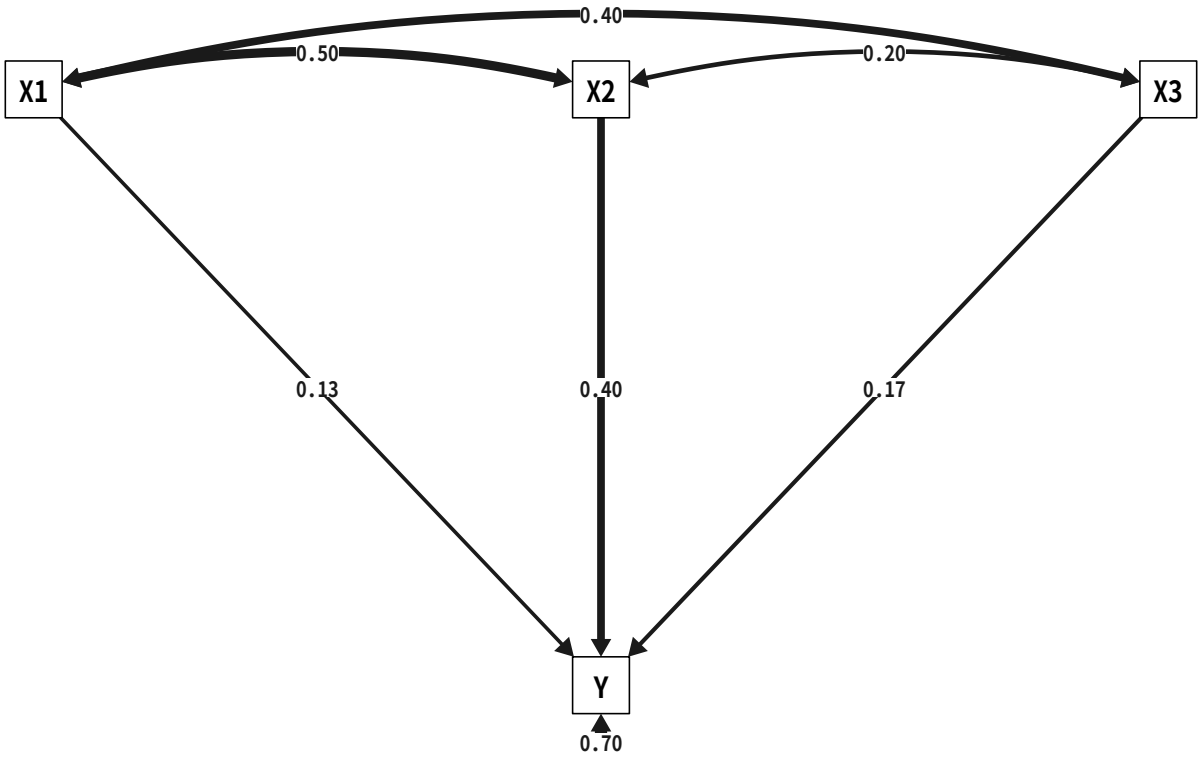
```
## $corr.partial.corr
##      Y    X1    X2     X3
## Y  1.0 0.128 0.383  0.180
## X1 0.4 1.000 0.379  0.322
## X2 0.5 0.500 1.000 -0.069
## X3 0.3 0.400 0.200  1.000
##
## $corr.confidence
##        Y    X1    X2    X3
## Y  1.000 0.553 0.634 0.469
## X1 0.221 1.000 0.634 0.553
## X2 0.337 0.337 1.000 0.381
## X3 0.110 0.221 0.004 1.000
##
## $omnibus.es
##   Rsq lower upper
## 0.303 0.143 0.418
##
## $standardized.estimates
##    estimates  lower upper   std
## X1     0.133 -0.073 0.339 0.105
## X2     0.400  0.207 0.593 0.098
## X3     0.167 -0.016 0.349 0.093
##
## $power
##  small medium  large
##  0.186  0.905  0.999
```

```r
# SEM
library(lavaan)
regression.model <-'
Y ~ a*X1 + b*X2 + c*X3
# residual variance of Y
Y ~~ z*Y
'
regression.fit <- sem(regression.model, sample.cov=correl, sample.nobs=100)
summary(regression.fit, rsquare=TRUE)
```

```
## lavaan 0.6-11 ended normally after 1 iterations
```

```
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                         4
##
##   Number of observations                           100
##
## Model Test User Model:
##
##   Test statistic                                 0.000
##   Degrees of freedom                                 0
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   Y ~
##     X1         (a)    0.133    0.103    1.294    0.196
##     X2         (b)    0.400    0.096    4.150    0.000
##     X3         (c)    0.167    0.091    1.830    0.067
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .Y          (z)    0.690    0.098    7.071    0.000
##
## R-Square:
##                    Estimate
##     Y                 0.303
```

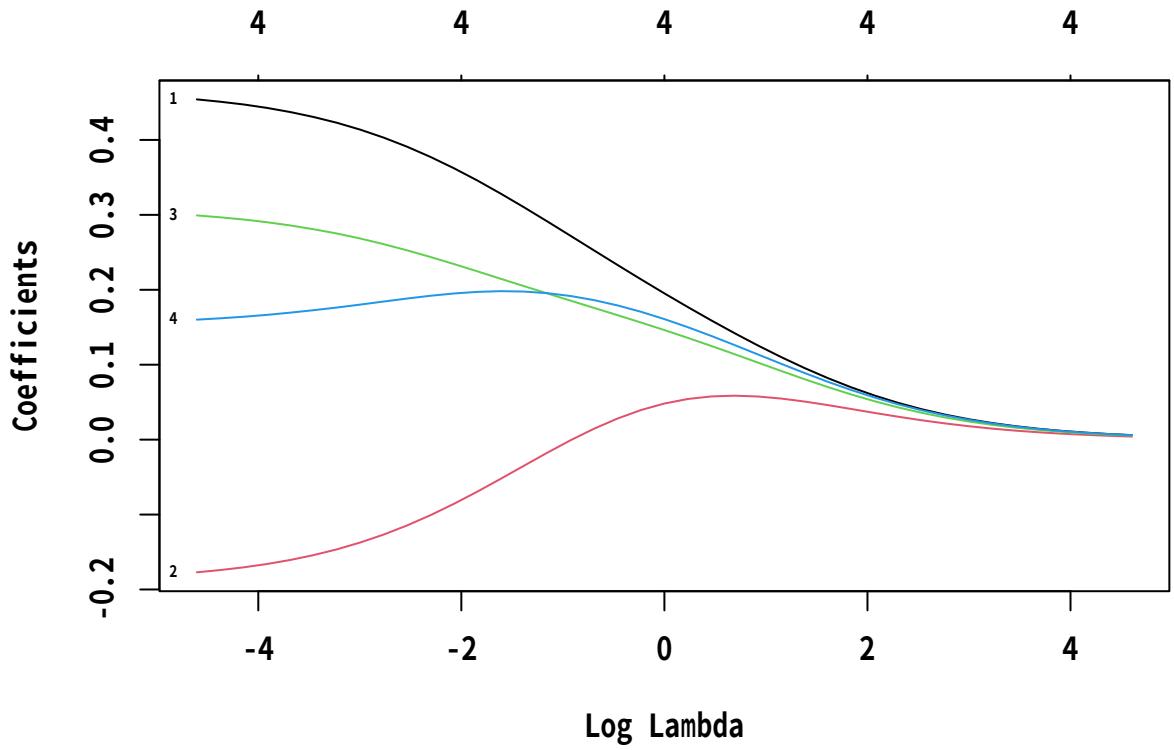
```
library(semPlot)
semPaths(regression.fit, "std", style="lisrel",
         mar=c(6,1,3,1), edge.label.cex=.8, fade=F, theme = 'gray')
```



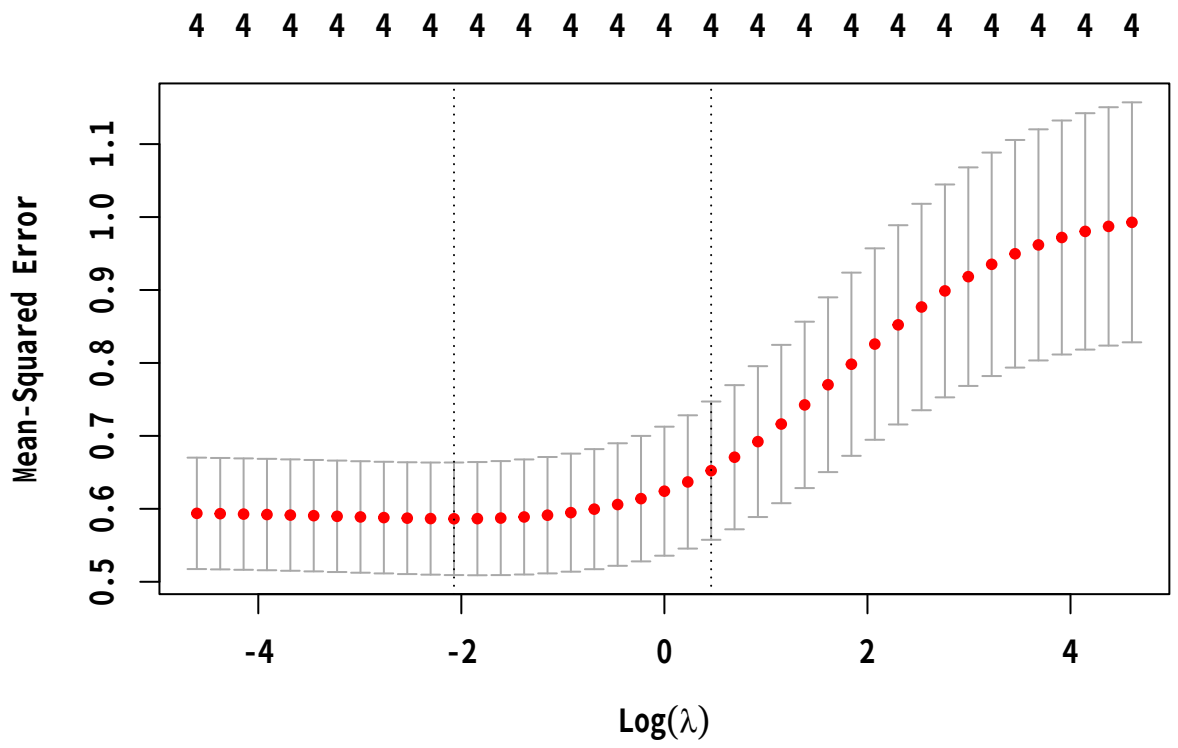
Regularization (Ridge Regression)

```
correl <- matrix(c(
  1, 0.62, 0.43, 0.56, 0.61,
  0.62, 1, 0.67, 0.55, 0.72,
  0.43, 0.67, 1, 0.67, 0.65,
  0.56, 0.55, 0.67, 1, 0.79,
  0.61, 0.72, 0.65, 0.79, 1),
  nrow=5,
  dimnames=list(c("Speaking","Vocabulary","Grammar","Writing","Reading"),
                c("Speaking","Vocabulary","Grammar","Writing","Reading")))

# Creating Simulated Data
library(MASS)
set.seed(88)
mu <- rep(c(0), times = length(colnames(correl)))
simdat <- mvrnorm(n=100, mu=mu, Sigma=correl, empirical=TRUE)
colnames(simdat) <- c("Speaking","Vocabulary","Grammar","Writing","Reading")
sim.df <- as.data.frame(simdat)

# Apply the Ridge Regression
```

```
library(glmnet)
predictors <- as.matrix(sim.df[-1])
response_variable <- as.matrix(sim.df[1])
lambdas <- 10^seq(2, -2, by = -.1)
fit <- glmnet(predictors, response_variable, alpha = 0, lambda = lambdas)
# alpha: ridge = 0, lasso = 1, elastic net = between 0 and 1
plot(fit, xvar = "lambda", label = TRUE)
```




```
# Cross-validation
set.seed(88) # to get the same result
lambda_calc <- cv.glmnet(predictors, response_variable, alpha = 0,
                         lambda = lambdas, grouped = FALSE)
# alpha: ridge = 0, lasso = 1, elastic net = between 0 and 1
plot(lambda_calc)
```




```
# 1000 cross-validations and take the average of
# lambda with minimum average error (lambda.min)
LambdaCalcs <- NULL
for (i in 1:1000){
  lambda_calc <- cv.glmnet(predictors, response_variable, alpha = 0,
                           lambda = lambdas, grouped = FALSE)
  LambdaCalcs <- cbind(LambdaCalcs, lambda_calc$lambda.min)
}
optlambda <- mean(LambdaCalcs)
# Get the standardized beta coefficients
options(scipen=999) # to turn off scientific notation (e notation)
coef(fit, s = optlambda)
```

```
## 5 x 1 sparse Matrix of class "dgCMatrix"
##                                       s1
## (Intercept)  0.000000000000000008082579
## Vocabulary   0.371642530167936802065753
## Grammar     -0.095130569070466325376501
## Writing      0.240225082709787679524638
## Reading      0.193172356492457786281136
```

Table 3 (Dominance Analysis)

```
# Dominance Analysis with Correlation Matrix (using dominanceanalysis package)
library(dominanceanalysis)
lm.cov <- lmWithCov(Reading~ Vocabulary+Grammar+Writing+Speaking, correl)
da <- dominanceAnalysis(lm.cov)
print(da)
```

```
##
## Dominance analysis
## Predictors: Vocabulary, Grammar, Writing, Speaking
## Fit-indices: r2
##
## * Fit index:  r2
##                  complete    conditional        general
## Vocabulary      Grmm,Spkn      Grmm,Spkn      Grmm,Spkn
## Grammar                                           Spkn
## Writing    Vcbl,Grmm,Spkn Vcbl,Grmm,Spkn Vcbl,Grmm,Spkn
## Speaking
##
## Average contribution:
##    Writing Vocabulary    Grammar   Speaking
##      0.292      0.200      0.133      0.119
```

```
# Dominance Analysis with the Simulated Data
lm.out <- lm(Speaking ~., sim.df)

# Relative Importance Analysis (=Dominance Analysis in Larson-Hall, 2016)
library(relaimpo)
calc.relimp(lm.out)
```

```
## Response variable: Speaking
## Total response variance: 1
## Analysis based on 100 observations
##
## 4 Regressors:
## Vocabulary Grammar Writing Reading
## Proportion of variance explained by model: 47.44%
## Metrics are not normalized (rela=FALSE).
##
## Relative importance metrics:
##
##                   lmg
## Vocabulary 0.17571475
## Grammar    0.05221536
## Writing    0.11412262
## Reading    0.13232929
##
## Average coefficients for different model sizes:
##
##              1X        2Xs        3Xs        4Xs
```

```
## Vocabulary 0.62 0.47499239  0.4417621  0.4669131
## Grammar    0.43 0.06131287 -0.0914224 -0.1893001
## Writing    0.56 0.33837447  0.2767278  0.3099320
## Reading    0.61 0.45261954  0.3229486  0.1520214
```

```
# CI and Statistical Test (Nimon & Oswald, 2013)
library(yhat)
regrOut <- calc.yhat(lm.out)
# Bootstrapping
library(boot)
set.seed(88)
boot.out <- boot(sim.df, boot.yhat, 1000, lmOut=lm.out, regrout0=regrOut)
# Summary Statistics of the Bootstrap Data
result <- booteval.yhat(regrOut, bty= "perc", boot.out)
# See the Results
regrOut #Compare a range of relative importance indices
```

```
## $PredictorMetrics
##                 b   Beta    r    rs   rs2 Unique Common  CD:0  CD:1  CD:2  CD:3
## Vocabulary  0.467  0.467 0.62 0.900 0.810  0.086  0.298 0.384 0.136 0.097 0.086
## Grammar    -0.189 -0.189 0.43 0.624 0.390  0.015  0.170 0.185 0.003 0.006 0.015
## Writing     0.310  0.310 0.56 0.813 0.661  0.031  0.283 0.314 0.073 0.039 0.031
## Reading     0.152  0.152 0.61 0.886 0.784  0.006  0.366 0.372 0.106 0.045 0.006
## Total          NA     NA   NA    NA 2.645  0.138  1.117 1.255 0.318 0.187 0.138
##            GenDom  Pratt   RLW
## Vocabulary  0.176  0.289 0.180
## Grammar     0.052 -0.081 0.048
## Writing     0.114  0.174 0.119
## Reading     0.132  0.093 0.128
## Total       0.474  0.475 0.475
##
## $OrderedPredictorMetrics
##            b Beta r rs rs2 Unique Common CD:0 CD:1 CD:2 CD:3 GenDom Pratt RLW
## Vocabulary 1    1 1  1   1      1      2    1    1    1    1      1     1   1
## Grammar    3    3 4  4   4      3      4    4    4    4    3      4     4   4
## Writing    2    2 3  3   3      2      3    3    3    3    2      3     2   3
## Reading    4    4 2  2   2      4      1    2    2    2    4      2     3   2
##
## $PairedDominanceMetrics
##                     Comp Cond Gen
## Vocabulary>Grammar  1.0  1.0   1
## Vocabulary>Writing  1.0  1.0   1
## Vocabulary>Reading  1.0  1.0   1
## Grammar>Writing     0.0  0.0   0
## Grammar>Reading     0.5  0.5   0
## Writing>Reading     0.5  0.5   0
##
## $APSRelatedMetrics
##                                   Commonality  % Total    R2 Vocabulary.Inc
## Vocabulary                              0.086    0.181 0.384             NA
## Grammar                                 0.015    0.032 0.185          0.200
## Writing                                 0.031    0.065 0.314          0.140
## Reading                                 0.006    0.013 0.372          0.068
## Vocabulary,Grammar                     -0.015   -0.032 0.385             NA
```

```
## Vocabulary,Writing                      -0.016   -0.035 0.453            NA
## Grammar,Writing                         -0.011   -0.024 0.319         0.149
## Vocabulary,Reading                       0.063    0.133 0.440            NA
## Grammar,Reading                          0.000    0.000 0.374         0.070
## Writing,Reading                          0.053    0.111 0.388         0.071
## Vocabulary,Grammar,Writing               0.013    0.028 0.468            NA
## Vocabulary,Grammar,Reading               0.005    0.011 0.444            NA
## Vocabulary,Writing,Reading               0.067    0.141 0.459            NA
## Grammar,Writing,Reading                 -0.003   -0.007 0.388         0.086
## Vocabulary,Grammar,Writing,Reading       0.181    0.381 0.474            NA
## Total                                    0.474    1.000    NA            NA
##                                    Grammar.Inc Writing.Inc Reading.Inc
## Vocabulary                               0.000       0.069       0.056
## Grammar                                     NA       0.134       0.189
## Writing                                  0.005          NA       0.075
## Reading                                  0.002       0.016          NA
## Vocabulary,Grammar                          NA       0.084       0.059
## Vocabulary,Writing                       0.015          NA       0.006
## Grammar,Writing                             NA          NA       0.069
## Vocabulary,Reading                       0.004       0.019          NA
## Grammar,Reading                             NA       0.014          NA
## Writing,Reading                          0.000          NA          NA
## Vocabulary,Grammar,Writing                  NA          NA       0.006
## Vocabulary,Grammar,Reading                  NA       0.031          NA
## Vocabulary,Writing,Reading               0.015          NA          NA
## Grammar,Writing,Reading                     NA          NA          NA
## Vocabulary,Grammar,Writing,Reading          NA          NA          NA
## Total                                       NA          NA          NA
```

```
regrOut$PredictorMetrics[,12] #GenDom Weight
```

```
## Vocabulary    Grammar    Writing    Reading      Total
##      0.176      0.052      0.114      0.132      0.474
```

```
regrOut$OrderedPredictorMetrics[,12] #GenDom Order
```

```
## Vocabulary    Grammar    Writing    Reading
##          1          4          3          2
```

```
regrOut$PairedDominanceMetrics #GenDom Comparisons
```

```
##                    Comp Cond Gen
## Vocabulary>Grammar  1.0  1.0   1
## Vocabulary>Writing  1.0  1.0   1
## Vocabulary>Reading  1.0  1.0   1
## Grammar>Writing     0.0  0.0   0
## Grammar>Reading     0.5  0.5   0
## Writing>Reading     0.5  0.5   0
```

```
# (0:Xi dominates Xj / 1:Xj dominates Xi / 0.5:Dominance not established)
result$combCIpm[,12, drop=FALSE] #DA weights with upper/lower CI
```

```
##                          GenDom
## Vocabulary 0.176(0.090,0.267)
## Grammar    0.052(0.029,0.104)
## Writing    0.114(0.056,0.184)
## Reading    0.132(0.063,0.214)
```

```
result$combCIpmDiff[,"GenDom", drop=FALSE] #Comparisons of PVs
```

```
##                                    GenDom
## Vocabulary-Grammar    0.124(0.040,0.207)*
## Vocabulary-Writing    0.062(-0.065,0.177)
## Vocabulary-Reading    0.044(-0.078,0.150)
## Grammar-Writing      -0.062(-0.128,0.001)
## Grammar-Reading     -0.080(-0.153,-0.007)*
## Writing-Reading      -0.018(-0.100,0.070)
```

Figure 2

```
library(ggplot2)
dat <- data.frame(
  dw = regrOut$PredictorMetrics[,12][1:(length(regrOut$PredictorMetrics[,12])-1)],
  low = result$lowerCIpm[,12],
  up = result$upperCIpm[,12])
valnames <- rownames(dat)

ggplot(dat) +
  theme(axis.text.x = element_text(size = 10),
        axis.text.y = element_text(size = 10)) +
  geom_bar(aes(x=reorder(valnames,dw), y=dw),
           stat="identity", fill="skyblue", alpha=0.6) +
  geom_errorbar(aes(x=rownames(dat), ymin=low, ymax=up),
                width=0.3, colour="red", alpha=0.9, size=0.8) +
  geom_label(aes(x=valnames, y=dw, label = dw, vjust=0.45),
             position = position_dodge(width=0.9)) +
  labs(x = "Predictor Variables",
       y = "Dominance Weights",
       caption = paste0("R-squared = ",
                        regrOut$PredictorMetrics[,12]
                        [length(regrOut$PredictorMetrics[,12])],
                        " / ",
                        "N = ", nrow(simdat)
                        )) +
  ylim(c(-0.001, 0.33)) +
  geom_text(x = 2.5, y = 0.31, label = "*", size = 5) +
  geom_segment(x = 1, xend = 1, y = 0.3, yend = 0.29, size = 0.3) +
  geom_segment(x = 1, xend = 4, y = 0.3, yend = 0.3, size = 0.3) +
  geom_segment(x = 4, xend = 4, y = 0.3, yend = 0.29, size = 0.3) +
  geom_text(x = 2.0, y = 0.27, label = "*", size = 5) +
```

```
  geom_segment(x = 1, xend = 1, y = 0.26, yend = 0.25, size = 0.3) +
  geom_segment(x = 1, xend = 3, y = 0.26, yend = 0.26, size = 0.3) +
  geom_segment(x = 3, xend = 3, y = 0.26, yend = 0.25, size = 0.3) +
  coord_flip()
```

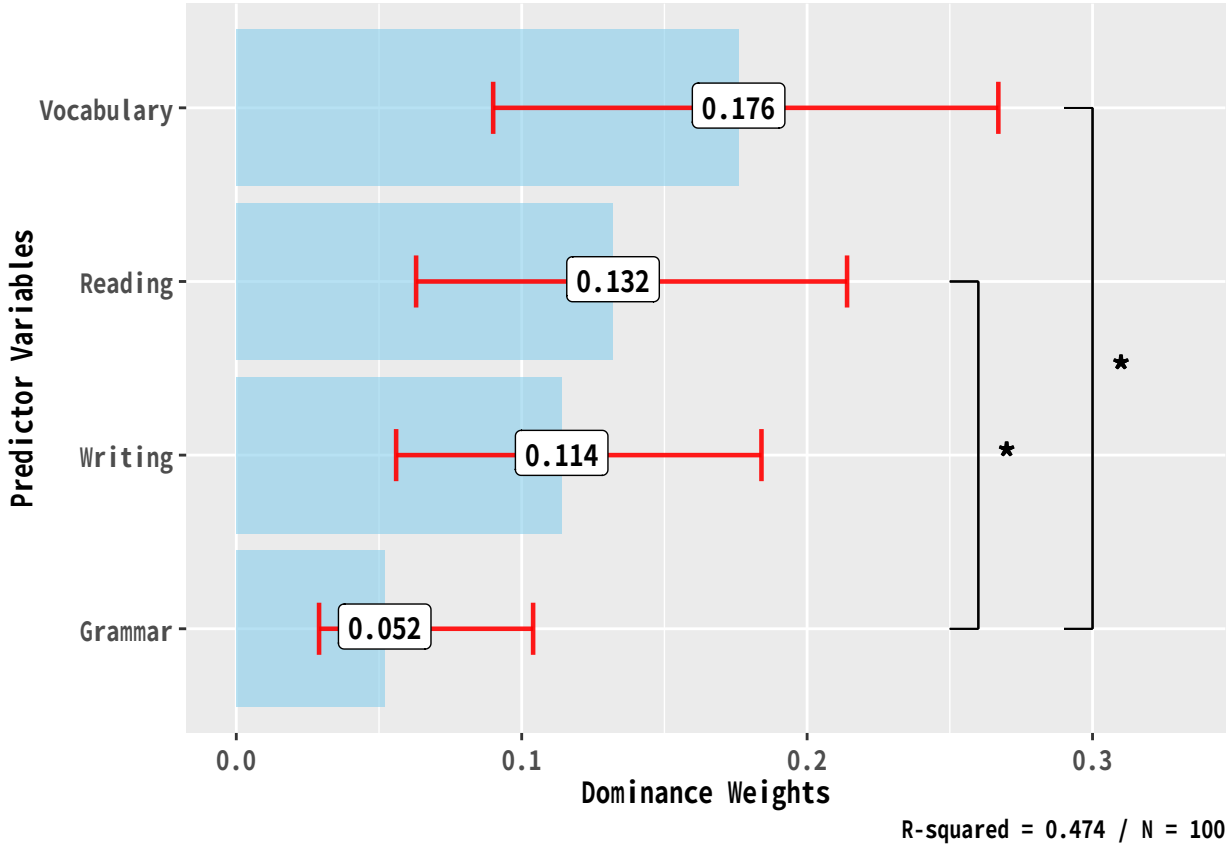
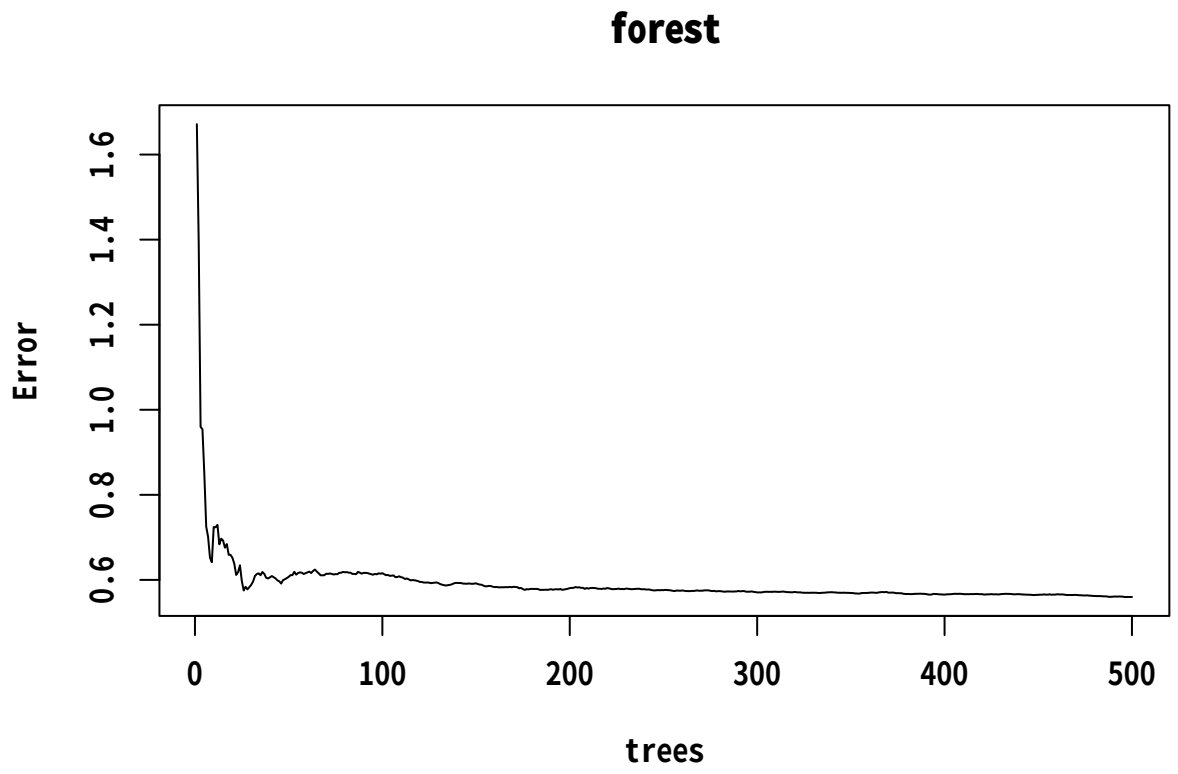


Figure 4

```
# Random Forest
library(randomForest)
forest <- randomForest(Speaking~., data=sim.df)
print(forest)
```

```
##
## Call:
##  randomForest(formula = Speaking ~ ., data = sim.df)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 1
##
##           Mean of squared residuals: 0.5598912
##                     % Var explained: 43.45
```

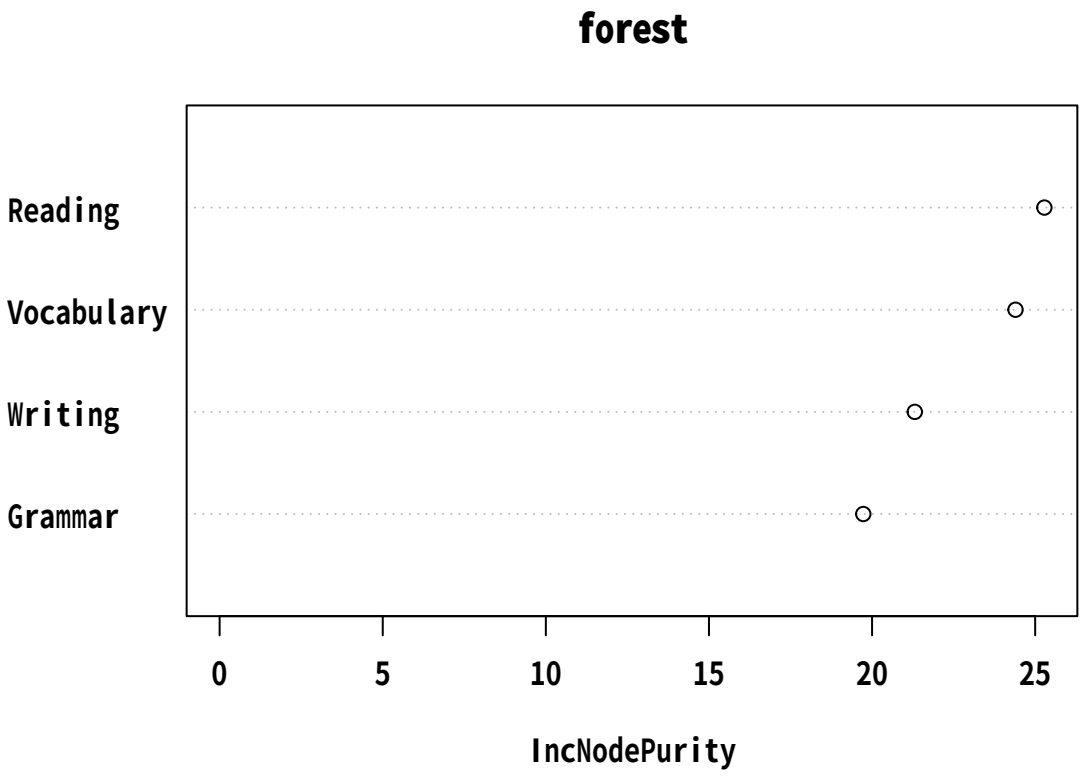
```
plot(forest)
```




```
forest$importance
```

```
##             IncNodePurity
## Vocabulary      24.39715
## Grammar         19.73061
## Writing         21.30974
## Reading         25.28152
```

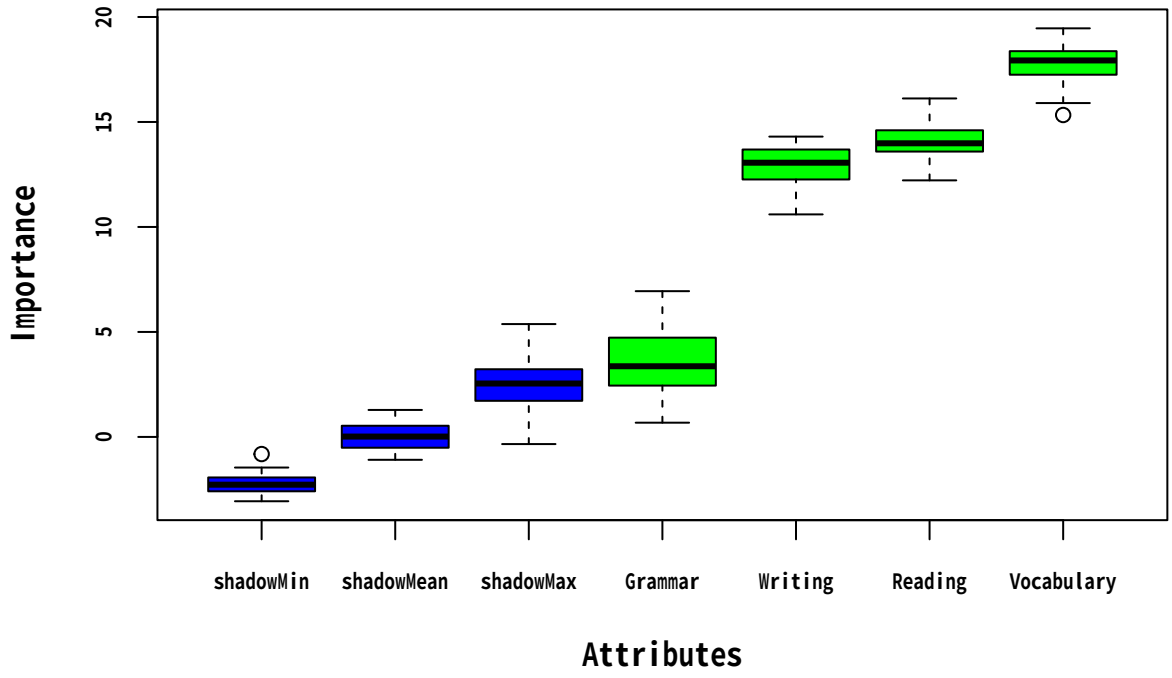
```
varImpPlot(forest)
```




```
# Boruta
library(Boruta)
set.seed(88)
boruta <- Boruta(Speaking~., maxRuns = 200, data = sim.df, doTrace = 2)
print(boruta)
```

```
## Boruta performed 28 iterations in 0.3937151 secs.
##  4 attributes confirmed important: Grammar, Reading, Vocabulary,
## Writing;
##  No attributes deemed unimportant.
```

```
plot(boruta, cex.axis=0.7)
```




```
attStats(boruta)
```

```
##               meanImp medianImp     minImp    maxImp  normHits  decision
## Vocabulary 17.823628 17.929024 15.3315548 19.458267 1.0000000 Confirmed
## Grammar     3.569753  3.365247  0.6776103  6.942289 0.7857143 Confirmed
## Writing    12.872331 13.058745 10.5974782 14.302943 1.0000000 Confirmed
## Reading    14.076095 13.983513 12.2187482 16.120454 1.0000000 Confirmed
```

Reproduction (Goh et al., 2020)

```
# Correlation Matrix
correl <- matrix(c(1, 0.67, 0.41, 0.02, -0.35, 0.35, 0.45,
                   0.67, 1, 0.18, 0.05, -0.28, 0.24, 0.31,
                   0.41, 0.18, 1, 0.21, -0.32, 0.34, 0.43,
                   0.02, 0.05, 0.21, 1, -0.03, 0.22, 0.47,
                   -0.35, -0.28, -0.32, -0.03, 1, -0.13, -0.22,
                   0.35, 0.24, 0.34, 0.22, -0.13, 1, 0.33,
                   0.45, 0.31, 0.43, 0.47, -0.22, 0.33, 1),
                 nrow=7,
                 dimnames=list(c("Score", "Wordcount", "CLI", "Commas", "Stopwords",
```

```
                                "Linking", "WordsSentence"),
                                c("Score", "Wordcount", "CLI", "Commas", "Stopwords",
                                "Linking", "WordsSentence")))

# Multiple Regression Analysis
library(rpsychi)
multreg.second(Score~ Wordcount+CLI+Commas+Stopwords+Linking+WordsSentence,
               corr=correl, n=200)
```

```
## $corr.partial.corr
##                Score Wordcount    CLI Commas Stopwords Linking WordsSentence
## Score           1.00      0.599  0.244 -0.247    -0.118   0.158         0.281
## Wordcount       0.67      1.000 -0.174  0.062    -0.100   0.029         0.016
## CLI             0.41      0.180  1.000  0.070    -0.213   0.177         0.198
## Commas          0.02      0.050  0.210  1.000     0.033   0.129         0.468
## Stopwords      -0.35     -0.280 -0.320 -0.030     1.000   0.042        -0.028
## Linking         0.35      0.240  0.340  0.220    -0.130   1.000         0.068
## WordsSentence   0.45      0.310  0.430  0.470    -0.220   0.330         1.000
##
## $corr.confidence
##                 Score Wordcount    CLI Commas Stopwords Linking WordsSentence
## Score           1.000     0.740  0.519  0.158    -0.222   0.466         0.554
## Wordcount       0.586     1.000  0.311  0.187    -0.147   0.367         0.430
## CLI             0.288     0.042  1.000  0.339    -0.190   0.457         0.537
## Commas         -0.119    -0.089  0.073  1.000     0.109   0.348         0.571
## Stopwords      -0.466    -0.403 -0.439 -0.168     1.000   0.009        -0.084
## Linking         0.222     0.105  0.211  0.084    -0.264   1.000         0.448
## WordsSentence   0.332     0.179  0.310  0.354    -0.348   0.200         1.000
##
## $omnibus.es
##   Rsq lower upper
## 0.598 0.502 0.653
##
## $standardized.estimates
##                estimates  lower  upper   std
## Wordcount          0.521  0.423  0.620 0.050
## CLI                0.188  0.082  0.293 0.054
## Commas            -0.185 -0.288 -0.083 0.052
## Stopwords         -0.082 -0.180  0.015 0.050
## Linking            0.112  0.013  0.211 0.050
## WordsSentence      0.240  0.124  0.355 0.059
##
## $power
##  small medium  large
##  0.261  0.992  1.000
```

```
# Dominance Analysis (using dominanceanalysis package)
library(dominanceanalysis)
lm.cov <- lmWithCov(Score ~ Wordcount+CLI+Commas+Stopwords+Linking+WordsSentence, correl)
da <- dominanceAnalysis(lm.cov)
print(da)
```

```
##
```

```
## Dominance analysis
## Predictors: Wordcount, CLI, Commas, Stopwords, Linking, WordsSentence
## Fit-indices: r2
##
## * Fit index:  r2
##                               complete             conditional
## Wordcount     CLI,Cmms,Stpw,Lnkn,WrdS CLI,Cmms,Stpw,Lnkn,WrdS
## CLI                                                   Stpw,Lnkn
## Commas
## Stopwords
## Linking
## WordsSentence          Cmms,Stpw,Lnkn      CLI,Cmms,Stpw,Lnkn
##                                general
## Wordcount     CLI,Cmms,Stpw,Lnkn,WrdS
## CLI                    Cmms,Stpw,Lnkn
## Commas
## Stopwords                         Cmms
## Linking                      Cmms,Stpw
## WordsSentence      CLI,Cmms,Stpw,Lnkn
##
## Average contribution:
##     Wordcount WordsSentence           CLI       Linking     Stopwords
##         0.315         0.097         0.073         0.048         0.047
##        Commas
##         0.018
```

```
# Simulated Dataset Using Means, SDs, and Correlations
mu <- c(25.95, 228.03, 7.94, 0.76, 0.15, 0.007, 15.24)
stddev <- c(12.16, 97.47, 2.01, 0.53, 0.03, 0.007, 5.91)
corMat <- matrix(c(1, 0.67, 0.41, 0.02, -0.35, 0.35, 0.45,
                   0.67, 1, 0.18, 0.05, -0.28, 0.24, 0.31,
                   0.41, 0.18, 1, 0.21, -0.32, 0.34, 0.43,
                   0.02, 0.05, 0.21, 1, -0.03, 0.22, 0.47,
                   -0.35, -0.28, -0.32, -0.03, 1, -0.13, -0.22,
                   0.35, 0.24, 0.34, 0.22, -0.13, 1, 0.33,
                   0.45, 0.31, 0.43, 0.47, -0.22, 0.33, 1),
                 ncol = 7)
covMat <- stddev %*% t(stddev) * corMat
library(MASS)
set.seed(88)
dat1 <- mvrnorm(n = 200, mu = mu, Sigma = covMat, empirical = TRUE)
colnames(dat1) <- c("Score", "Wordcount", "CLI", "Commas", "Stopwords",
                    "Linking", "WordsSentence")
dat1 <- as.data.frame(dat1)
colMeans(dat1) # Means
```

```
##         Score     Wordcount           CLI        Commas     Stopwords
##        25.950       228.030         7.940         0.760         0.150
##       Linking WordsSentence
##         0.007        15.240
```

```
apply(dat1, 2, sd) # SDs
```

```
##         Score     Wordcount           CLI        Commas     Stopwords
```

```
##        12.160        97.470         2.010         0.530         0.030
##       Linking WordsSentence
##         0.007         5.910
```

```
cor(dat1) # Correlation Matrix
```

```
##                Score Wordcount   CLI Commas Stopwords Linking WordsSentence
## Score           1.00      0.67  0.41   0.02     -0.35    0.35          0.45
## Wordcount       0.67      1.00  0.18   0.05     -0.28    0.24          0.31
## CLI             0.41      0.18  1.00   0.21     -0.32    0.34          0.43
## Commas          0.02      0.05  0.21   1.00     -0.03    0.22          0.47
## Stopwords      -0.35     -0.28 -0.32  -0.03      1.00   -0.13         -0.22
## Linking         0.35      0.24  0.34   0.22     -0.13    1.00          0.33
## WordsSentence   0.45      0.31  0.43   0.47     -0.22    0.33          1.00
```

```
# Multiple Regression Analysis
library(rpsychi)
multreg(Score~ Wordcount+CLI+Commas+Stopwords+Linking+WordsSentence, data=dat1)
```

```
## $samp.stat
##                       m     sd
## Score            25.950 12.160
## Wordcount       228.030 97.470
## CLI               7.940  2.010
## Commas            0.760  0.530
## Stopwords         0.150  0.030
## Linking           0.007  0.007
## WordsSentence    15.240  5.910
##
## $corr.partial.corr
##                Score Wordcount    CLI Commas Stopwords Linking WordsSentence
## Score           1.00     0.599  0.244 -0.247    -0.118   0.158         0.281
## Wordcount       0.67     1.000 -0.174  0.062    -0.100   0.029         0.016
## CLI             0.41     0.180  1.000  0.070    -0.213   0.177         0.198
## Commas          0.02     0.050  0.210  1.000     0.033   0.129         0.468
## Stopwords      -0.35    -0.280 -0.320 -0.030     1.000   0.042        -0.028
## Linking         0.35     0.240  0.340  0.220    -0.130   1.000         0.068
## WordsSentence   0.45     0.310  0.430  0.470    -0.220   0.330         1.000
##
## $corr.confidence
##                 Score Wordcount    CLI Commas Stopwords Linking WordsSentence
## Score           1.000     0.740  0.519  0.158    -0.222   0.466         0.554
## Wordcount       0.586     1.000  0.311  0.187    -0.147   0.367         0.430
## CLI             0.288     0.042  1.000  0.339    -0.190   0.457         0.537
## Commas         -0.119    -0.089  0.073  1.000     0.109   0.348         0.571
## Stopwords      -0.466    -0.403 -0.439 -0.168     1.000   0.009        -0.084
## Linking         0.222     0.105  0.211  0.084    -0.264   1.000         0.448
## WordsSentence   0.332     0.179  0.310  0.354    -0.348   0.200         1.000
##
## $omnibus.es
##   Rsq lower upper
## 0.598 0.502 0.653
##
```

```
## $raw.estimates
##                estimates   lower   upper   std
## Intercept          1.465      NA      NA    NA
## Wordcount          0.065   0.053   0.077 0.050
## CLI                1.135   0.494   1.775 0.054
## Commas            -4.250  -6.621  -1.880 0.052
## Stopwords        -33.350 -73.118   6.418 0.050
## Linking          194.692  21.785 367.599 0.050
## WordsSentence      0.493   0.254   0.732 0.059
##
## $standardized.estimates
##                estimates  lower  upper   std
## Wordcount          0.521  0.423  0.620 0.050
## CLI                0.188  0.082  0.293 0.054
## Commas            -0.185 -0.288 -0.083 0.052
## Stopwords         -0.082 -0.180  0.015 0.050
## Linking            0.112  0.013  0.211 0.050
## WordsSentence      0.240  0.124  0.355 0.059
##
## $power
##  small medium  large
##  0.261  0.992  1.000
```

```
lm.out <- lm(Score ~., dat1)

# Calculating p-values
z <- scale(dat1)
z <- data.frame(z)
summary(lm(Score~ ., z))
```

```
##
## Call:
## lm(formula = Score ~ ., data = z)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.84374 -0.42567 -0.03523  0.39587  1.53001
##
## Coefficients:
##                              Estimate               Std. Error t value
## (Intercept)    0.00000000000000009349  0.045505498100877263268   0.000
## Wordcount      0.52124469843165954774  0.050142466955968169329  10.395
## CLI            0.18755223548123939325  0.053679575719487448993   3.494
## Commas        -0.18525621742477174636  0.052381917597320623581  -3.537
## Stopwords     -0.08227735472442142261  0.049744017767831619214  -1.654
## Linking        0.11207599094020087582  0.050465773820258172377   2.221
## WordsSentence  0.23975100936925616834  0.058904186105827087416   4.070
##                           Pr(>|t|)
## (Intercept)               1.000000
## Wordcount     < 0.0000000000000002 ***
## CLI                       0.000590 ***
## Commas                    0.000507 ***
## Stopwords                 0.099750 .
## Linking                   0.027525 *
```

```
## WordsSentence            0.0000685 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6435 on 193 degrees of freedom
## Multiple R-squared:  0.5983, Adjusted R-squared:  0.5858
## F-statistic: 47.92 on 6 and 193 DF,  p-value: < 0.00000000000000022
```

```
# Check the VIF (variance inflation factor)
library(car)
vif(lm.out) # VIF>10 shows multicollinearity. VIF<2 recommended.
```

```
##     Wordcount           CLI        Commas     Stopwords       Linking
##      1.208111      1.384565      1.318433      1.188987      1.223740
## WordsSentence
##      1.667200
```

```
# Relative importance analysis (=Dominance Analysis in Larson-Hall, 2016)
library(relaimpo)
calc.relimp(lm.out)
```

```
## Response variable: Score
## Total response variance: 147.8656
## Analysis based on 200 observations
##
## 6 Regressors:
## Wordcount CLI Commas Stopwords Linking WordsSentence
## Proportion of variance explained by model: 59.83%
## Metrics are not normalized (rela=FALSE).
##
## Relative importance metrics:
##
##                      lmg
## Wordcount     0.31473012
## CLI           0.07293881
## Commas        0.01783422
## Stopwords     0.04723990
## Linking       0.04822695
## WordsSentence 0.09736686
##
## Average coefficients for different model sizes:
##
##                          1X           2Xs          3Xs          4Xs
## Wordcount        0.08358674    0.07775359   0.07348302   0.07018525
## CLI              2.48039801    1.99672836   1.64886236   1.40661747
## Commas           0.45886792   -1.73851437  -2.96867066  -3.68106412
## Stopwords     -141.86666667 -108.88809517 -84.26239595 -64.42175753
## Linking        608.00000000  464.83472084 365.50772584 293.85228863
## WordsSentence    0.92588832    0.79697263   0.69668147   0.61664023
##                         5Xs          6Xs
## Wordcount        0.06746194   0.06502858
## CLI              1.24293203   1.13464437
## Commas          -4.07417738  -4.25040680
```

```
## Stopwords     -47.67116820 -33.34975445
## Linking       238.97508546 194.69200712
## WordsSentence   0.55028876   0.49329480
```

```
# Dominance Analysis (Same Result)
library(dominanceanalysis)
da.result <- dominanceAnalysis(lm.out)
print(da.result)
```

```
##
## Dominance analysis
## Predictors: Wordcount, CLI, Commas, Stopwords, Linking, WordsSentence
## Fit-indices: r2
##
## * Fit index:  r2
##                               complete             conditional
## Wordcount     CLI,Cmms,Stpw,Lnkn,WrdS CLI,Cmms,Stpw,Lnkn,WrdS
## CLI                                                  Stpw,Lnkn
## Commas
## Stopwords
## Linking
## WordsSentence          Cmms,Stpw,Lnkn      CLI,Cmms,Stpw,Lnkn
##                               general
## Wordcount     CLI,Cmms,Stpw,Lnkn,WrdS
## CLI                    Cmms,Stpw,Lnkn
## Commas
## Stopwords                      Cmms
## Linking                   Cmms,Stpw
## WordsSentence      CLI,Cmms,Stpw,Lnkn
##
## Average contribution:
##     Wordcount WordsSentence           CLI       Linking     Stopwords
##         0.315         0.097         0.073         0.048         0.047
##        Commas
##         0.018
```

```
# CI and Statistical Test (Nimon & Oswald, 2013)
library(yhat)
regrOut <- calc.yhat(lm.out)
# Bootstrapping
library(boot)
set.seed(88)
boot.out <- boot(dat1, boot.yhat, 1000, lmOut=lm.out, regrout0=regrOut)
# Summary Statistics of the Bootstrap Data
result <- booteval.yhat(regrOut, bty= "perc", boot.out)
# See the Results
regrOut #Compare a range of relative importance indices
```

```
## $PredictorMetrics
##                      b   Beta     r     rs   rs2 Unique Common  CD:0  CD:1  CD:2
## Wordcount        0.065  0.521  0.67  0.866 0.750  0.225  0.224 0.449 0.369 0.316
## CLI              1.135  0.188  0.41  0.530 0.281  0.025  0.143 0.168 0.102 0.065
## Commas          -4.250 -0.185  0.02  0.026 0.001  0.026 -0.026 0.000 0.011 0.019
```

```
## Stopwords     -33.350 -0.082 -0.35 -0.452 0.205  0.006  0.117 0.123 0.073 0.043
## Linking       194.692  0.112  0.35  0.452 0.205  0.010  0.112 0.122 0.071 0.042
## WordsSentence   0.493  0.240  0.45  0.582 0.338  0.034  0.168 0.203 0.136 0.094
## Total              NA     NA    NA     NA 1.780  0.326  0.738 1.065 0.762 0.579
##                 CD:3  CD:4  CD:5 GenDom  Pratt   RLW
## Wordcount     0.279 0.250 0.225  0.315  0.349 0.317
## CLI           0.044 0.033 0.025  0.073  0.077 0.074
## Commas        0.024 0.026 0.026  0.018 -0.004 0.015
## Stopwords     0.025 0.014 0.006  0.047  0.029 0.047
## Linking       0.027 0.017 0.010  0.048  0.039 0.049
## WordsSentence 0.068 0.049 0.034  0.097  0.108 0.096
## Total         0.467 0.389 0.326  0.598  0.598 0.598
##
## $OrderedPredictorMetrics
##               b Beta r rs rs2 Unique Common CD:0 CD:1 CD:2 CD:3 CD:4 CD:5
## Wordcount     6    1 1  1   1      1      1    1    1    1    1    1    1
## CLI           4    3 3  3   3      4      3    3    3    3    3    3    4
## Commas        3    4 6  6   6      3      6    6    6    6    6    4    3
## Stopwords     2    6 4  4   4      6      4    4    4    4    5    6    6
## Linking       1    5 5  5   5      5      5    5    5    5    4    5    5
## WordsSentence 5    2 2  2   2      2      2    2    2    2    2    2    2
##               GenDom Pratt RLW
## Wordcount          1     1   1
## CLI                3     3   3
## Commas             6     6   6
## Stopwords          5     5   5
## Linking            4     4   4
## WordsSentence      2     2   2
##
## $PairedDominanceMetrics
##                          Comp Cond Gen
## Wordcount>CLI             1.0  1.0   1
## Wordcount>Commas          1.0  1.0   1
## Wordcount>Stopwords       1.0  1.0   1
## Wordcount>Linking         1.0  1.0   1
## Wordcount>WordsSentence   1.0  1.0   1
## CLI>Commas                0.5  0.5   1
## CLI>Stopwords             0.5  1.0   1
## CLI>Linking               0.5  1.0   1
## CLI>WordsSentence         0.5  0.0   0
## Commas>Stopwords          0.5  0.5   0
## Commas>Linking            0.5  0.5   0
## Commas>WordsSentence      0.0  0.0   0
## Stopwords>Linking         0.5  0.5   0
## Stopwords>WordsSentence   0.0  0.0   0
## Linking>WordsSentence     0.0  0.0   0
##
## $APSRelatedMetrics
##                                                    Commonality  % Total    R2
## Wordcount                                                0.225    0.376 0.449
## CLI                                                      0.025    0.042 0.168
## Commas                                                   0.026    0.044 0.000
## Stopwords                                                0.006    0.010 0.123
## Linking                                                  0.010    0.017 0.122
```

```
## WordsSentence                                        0.034    0.058 0.203
## Wordcount,CLI                                       -0.005   -0.009 0.535
## Wordcount,Commas                                     0.020    0.034 0.449
## CLI,Commas                                          -0.001   -0.001 0.173
## Wordcount,Stopwords                                  0.027    0.046 0.478
## CLI,Stopwords                                        0.009    0.014 0.221
## Commas,Stopwords                                     0.002    0.003 0.123
## Wordcount,Linking                                    0.021    0.036 0.487
## CLI,Linking                                          0.010    0.016 0.218
## Commas,Linking                                      -0.003   -0.005 0.126
## Stopwords,Linking                                    0.000   -0.001 0.217
## Wordcount,WordsSentence                              0.061    0.101 0.514
## CLI,WordsSentence                                    0.024    0.040 0.260
## Commas,WordsSentence                                -0.018   -0.030 0.250
## Stopwords,WordsSentence                              0.002    0.003 0.269
## Linking,WordsSentence                                0.005    0.009 0.248
## Wordcount,CLI,Commas                                 0.000   -0.001 0.541
## Wordcount,CLI,Stopwords                              0.009    0.015 0.543
## Wordcount,Commas,Stopwords                           0.006    0.010 0.478
## CLI,Commas,Stopwords                                 0.001    0.002 0.225
## Wordcount,CLI,Linking                                0.005    0.008 0.547
## Wordcount,Commas,Linking                            -0.003   -0.005 0.490
## CLI,Commas,Linking                                  -0.001   -0.002 0.229
## Wordcount,Stopwords,Linking                          0.001    0.002 0.511
## CLI,Stopwords,Linking                                0.002    0.003 0.269
## Commas,Stopwords,Linking                             0.000    0.000 0.220
## Wordcount,CLI,WordsSentence                          0.014    0.023 0.558
## Wordcount,Commas,WordsSentence                      -0.019   -0.032 0.538
## CLI,Commas,WordsSentence                            -0.004   -0.007 0.308
## Wordcount,Stopwords,WordsSentence                    0.014    0.023 0.532
## CLI,Stopwords,WordsSentence                          0.008    0.014 0.301
## Commas,Stopwords,WordsSentence                      -0.001   -0.002 0.307
## Wordcount,Linking,WordsSentence                      0.017    0.028 0.531
## CLI,Linking,WordsSentence                            0.012    0.020 0.286
## Commas,Linking,WordsSentence                         0.000   -0.001 0.303
## Stopwords,Linking,WordsSentence                      0.000    0.000 0.308
## Wordcount,CLI,Commas,Stopwords                       0.001    0.001 0.548
## Wordcount,CLI,Commas,Linking                         0.000    0.000 0.556
## Wordcount,CLI,Stopwords,Linking                      0.005    0.008 0.556
## Wordcount,Commas,Stopwords,Linking                   0.000   -0.001 0.515
## CLI,Commas,Stopwords,Linking                         0.000    0.000 0.278
## Wordcount,CLI,Commas,WordsSentence                   0.000   -0.001 0.583
## Wordcount,CLI,Stopwords,WordsSentence                0.020    0.033 0.565
## Wordcount,Commas,Stopwords,WordsSentence            -0.005   -0.008 0.553
## CLI,Commas,Stopwords,WordsSentence                  -0.002   -0.003 0.342
## Wordcount,CLI,Linking,WordsSentence                  0.018    0.030 0.565
## Wordcount,Commas,Linking,WordsSentence               0.000    0.000 0.559
## CLI,Commas,Linking,WordsSentence                     0.001    0.002 0.340
## Wordcount,Stopwords,Linking,WordsSentence            0.003    0.004 0.547
## CLI,Stopwords,Linking,WordsSentence                  0.003    0.005 0.327
## Commas,Stopwords,Linking,WordsSentence               0.000    0.000 0.353
## Wordcount,CLI,Commas,Stopwords,Linking               0.000    0.000 0.564
## Wordcount,CLI,Commas,Stopwords,WordsSentence        -0.002   -0.003 0.588
## Wordcount,CLI,Commas,Linking,WordsSentence           0.002    0.004 0.593
```

```
## Wordcount,CLI,Stopwords,Linking,WordsSentence              0.015    0.025 0.572
## Wordcount,Commas,Stopwords,Linking,WordsSentence           0.000   -0.001 0.573
## CLI,Commas,Stopwords,Linking,WordsSentence                 0.000    0.000 0.373
## Wordcount,CLI,Commas,Stopwords,Linking,WordsSentence       0.001    0.002 0.598
## Total                                                      0.598    1.000    NA
##                                                  Wordcount.Inc CLI.Inc
## Wordcount                                                   NA   0.087
## CLI                                                      0.367      NA
## Commas                                                   0.449   0.172
## Stopwords                                                0.355   0.099
## Linking                                                  0.364   0.096
## WordsSentence                                            0.311   0.058
## Wordcount,CLI                                               NA      NA
## Wordcount,Commas                                            NA   0.092
## CLI,Commas                                               0.368      NA
## Wordcount,Stopwords                                         NA   0.066
## CLI,Stopwords                                            0.322      NA
## Commas,Stopwords                                         0.355   0.102
## Wordcount,Linking                                           NA   0.060
## CLI,Linking                                              0.329      NA
## Commas,Linking                                           0.364   0.103
## Stopwords,Linking                                        0.295   0.052
## Wordcount,WordsSentence                                     NA   0.044
## CLI,WordsSentence                                        0.298      NA
## Commas,WordsSentence                                     0.288   0.059
## Stopwords,WordsSentence                                  0.263   0.032
## Linking,WordsSentence                                    0.283   0.038
## Wordcount,CLI,Commas                                        NA      NA
## Wordcount,CLI,Stopwords                                     NA      NA
## Wordcount,Commas,Stopwords                                  NA   0.071
## CLI,Commas,Stopwords                                     0.324      NA
## Wordcount,CLI,Linking                                       NA      NA
## Wordcount,Commas,Linking                                    NA   0.066
## CLI,Commas,Linking                                       0.327      NA
## Wordcount,Stopwords,Linking                                 NA   0.044
## CLI,Stopwords,Linking                                    0.286      NA
## Commas,Stopwords,Linking                                 0.294   0.058
## Wordcount,CLI,WordsSentence                                 NA      NA
## Wordcount,Commas,WordsSentence                              NA   0.045
## CLI,Commas,WordsSentence                                 0.275      NA
## Wordcount,Stopwords,WordsSentence                           NA   0.033
## CLI,Stopwords,WordsSentence                              0.264      NA
## Commas,Stopwords,WordsSentence                           0.246   0.035
## Wordcount,Linking,WordsSentence                             NA   0.034
## CLI,Linking,WordsSentence                                0.279      NA
## Commas,Linking,WordsSentence                             0.256   0.037
## Stopwords,Linking,WordsSentence                          0.240   0.019
## Wordcount,CLI,Commas,Stopwords                              NA      NA
## Wordcount,CLI,Commas,Linking                                NA      NA
## Wordcount,CLI,Stopwords,Linking                             NA      NA
## Wordcount,Commas,Stopwords,Linking                          NA   0.049
## CLI,Commas,Stopwords,Linking                             0.286      NA
## Wordcount,CLI,Commas,WordsSentence                          NA      NA
## Wordcount,CLI,Stopwords,WordsSentence                       NA      NA
```

```
## Wordcount,Commas,Stopwords,WordsSentence                        NA   0.035
## CLI,Commas,Stopwords,WordsSentence                            0.246     NA
## Wordcount,CLI,Linking,WordsSentence                              NA     NA
## Wordcount,Commas,Linking,WordsSentence                           NA   0.034
## CLI,Commas,Linking,WordsSentence                              0.252     NA
## Wordcount,Stopwords,Linking,WordsSentence                        NA   0.025
## CLI,Stopwords,Linking,WordsSentence                           0.245     NA
## Commas,Stopwords,Linking,WordsSentence                        0.220   0.020
## Wordcount,CLI,Commas,Stopwords,Linking                           NA     NA
## Wordcount,CLI,Commas,Stopwords,WordsSentence                     NA     NA
## Wordcount,CLI,Commas,Linking,WordsSentence                       NA     NA
## Wordcount,CLI,Stopwords,Linking,WordsSentence                    NA     NA
## Wordcount,Commas,Stopwords,Linking,WordsSentence                 NA   0.025
## CLI,Commas,Stopwords,Linking,WordsSentence                    0.225     NA
## Wordcount,CLI,Commas,Stopwords,Linking,WordsSentence             NA     NA
## Total                                                            NA     NA
##                                                        Commas.Inc Stopwords.Inc
## Wordcount                                                   0.000         0.029
## CLI                                                         0.005         0.053
## Commas                                                         NA         0.122
## Stopwords                                                   0.000            NA
## Linking                                                     0.003         0.094
## WordsSentence                                               0.047         0.066
## Wordcount,CLI                                               0.006         0.008
## Wordcount,Commas                                               NA         0.029
## CLI,Commas                                                     NA         0.052
## Wordcount,Stopwords                                         0.000            NA
## CLI,Stopwords                                               0.003            NA
## Commas,Stopwords                                               NA            NA
## Wordcount,Linking                                           0.003         0.024
## CLI,Linking                                                 0.011         0.051
## Commas,Linking                                                 NA         0.094
## Stopwords,Linking                                           0.003            NA
## Wordcount,WordsSentence                                     0.024         0.018
## CLI,WordsSentence                                           0.048         0.041
## Commas,WordsSentence                                           NA         0.057
## Stopwords,WordsSentence                                     0.038            NA
## Linking,WordsSentence                                       0.055         0.060
## Wordcount,CLI,Commas                                           NA         0.007
## Wordcount,CLI,Stopwords                                     0.005            NA
## Wordcount,Commas,Stopwords                                     NA            NA
## CLI,Commas,Stopwords                                           NA            NA
## Wordcount,CLI,Linking                                       0.009         0.008
## Wordcount,Commas,Linking                                       NA         0.024
## CLI,Commas,Linking                                             NA         0.049
## Wordcount,Stopwords,Linking                                 0.003            NA
## CLI,Stopwords,Linking                                       0.009            NA
## Commas,Stopwords,Linking                                       NA            NA
## Wordcount,CLI,WordsSentence                                 0.025         0.007
## Wordcount,Commas,WordsSentence                                 NA         0.015
## CLI,Commas,WordsSentence                                       NA         0.034
## Wordcount,Stopwords,WordsSentence                           0.021            NA
## CLI,Stopwords,WordsSentence                                 0.041            NA
## Commas,Stopwords,WordsSentence                                 NA            NA
```

```
## Wordcount,Linking,WordsSentence                                0.028         0.017
## CLI,Linking,WordsSentence                                      0.054         0.041
## Commas,Linking,WordsSentence                                      NA         0.050
## Stopwords,Linking,WordsSentence                                0.045            NA
## Wordcount,CLI,Commas,Stopwords                                    NA            NA
## Wordcount,CLI,Commas,Linking                                      NA         0.008
## Wordcount,CLI,Stopwords,Linking                                0.008            NA
## Wordcount,Commas,Stopwords,Linking                                NA            NA
## CLI,Commas,Stopwords,Linking                                      NA            NA
## Wordcount,CLI,Commas,WordsSentence                                NA         0.005
## Wordcount,CLI,Stopwords,WordsSentence                          0.023            NA
## Wordcount,Commas,Stopwords,WordsSentence                          NA            NA
## CLI,Commas,Stopwords,WordsSentence                                NA            NA
## Wordcount,CLI,Linking,WordsSentence                            0.028         0.007
## Wordcount,Commas,Linking,WordsSentence                            NA         0.014
## CLI,Commas,Linking,WordsSentence                                  NA         0.033
## Wordcount,Stopwords,Linking,WordsSentence                      0.025            NA
## CLI,Stopwords,Linking,WordsSentence                            0.046            NA
## Commas,Stopwords,Linking,WordsSentence                            NA            NA
## Wordcount,CLI,Commas,Stopwords,Linking                            NA            NA
## Wordcount,CLI,Commas,Stopwords,WordsSentence                      NA            NA
## Wordcount,CLI,Commas,Linking,WordsSentence                        NA         0.006
## Wordcount,CLI,Stopwords,Linking,WordsSentence                  0.026            NA
## Wordcount,Commas,Stopwords,Linking,WordsSentence                  NA            NA
## CLI,Commas,Stopwords,Linking,WordsSentence                        NA            NA
## Wordcount,CLI,Commas,Stopwords,Linking,WordsSentence              NA            NA
## Total                                                             NA            NA
##                                                          Linking.Inc
## Wordcount                                                      0.038
## CLI                                                            0.050
## Commas                                                         0.126
## Stopwords                                                      0.094
## Linking                                                           NA
## WordsSentence                                                  0.046
## Wordcount,CLI                                                  0.012
## Wordcount,Commas                                               0.041
## CLI,Commas                                                     0.057
## Wordcount,Stopwords                                            0.034
## CLI,Stopwords                                                  0.048
## Commas,Stopwords                                               0.098
## Wordcount,Linking                                                 NA
## CLI,Linking                                                       NA
## Commas,Linking                                                    NA
## Stopwords,Linking                                                 NA
## Wordcount,WordsSentence                                        0.017
## CLI,WordsSentence                                              0.026
## Commas,WordsSentence                                           0.053
## Stopwords,WordsSentence                                        0.039
## Linking,WordsSentence                                             NA
## Wordcount,CLI,Commas                                           0.015
## Wordcount,CLI,Stopwords                                        0.012
## Wordcount,Commas,Stopwords                                     0.037
## CLI,Commas,Stopwords                                           0.053
## Wordcount,CLI,Linking                                             NA
```

```
## Wordcount,Commas,Linking                                        NA
## CLI,Commas,Linking                                              NA
## Wordcount,Stopwords,Linking                                     NA
## CLI,Stopwords,Linking                                           NA
## Commas,Stopwords,Linking                                        NA
## Wordcount,CLI,WordsSentence                                  0.007
## Wordcount,Commas,WordsSentence                               0.021
## CLI,Commas,WordsSentence                                     0.032
## Wordcount,Stopwords,WordsSentence                            0.016
## CLI,Stopwords,WordsSentence                                  0.026
## Commas,Stopwords,WordsSentence                               0.046
## Wordcount,Linking,WordsSentence                                 NA
## CLI,Linking,WordsSentence                                       NA
## Commas,Linking,WordsSentence                                    NA
## Stopwords,Linking,WordsSentence                                 NA
## Wordcount,CLI,Commas,Stopwords                               0.015
## Wordcount,CLI,Commas,Linking                                    NA
## Wordcount,CLI,Stopwords,Linking                                 NA
## Wordcount,Commas,Stopwords,Linking                              NA
## CLI,Commas,Stopwords,Linking                                    NA
## Wordcount,CLI,Commas,WordsSentence                           0.010
## Wordcount,CLI,Stopwords,WordsSentence                        0.007
## Wordcount,Commas,Stopwords,WordsSentence                     0.020
## CLI,Commas,Stopwords,WordsSentence                           0.032
## Wordcount,CLI,Linking,WordsSentence                             NA
## Wordcount,Commas,Linking,WordsSentence                          NA
## CLI,Commas,Linking,WordsSentence                                NA
## Wordcount,Stopwords,Linking,WordsSentence                       NA
## CLI,Stopwords,Linking,WordsSentence                             NA
## Commas,Stopwords,Linking,WordsSentence                          NA
## Wordcount,CLI,Commas,Stopwords,Linking                          NA
## Wordcount,CLI,Commas,Stopwords,WordsSentence                 0.010
## Wordcount,CLI,Commas,Linking,WordsSentence                      NA
## Wordcount,CLI,Stopwords,Linking,WordsSentence                   NA
## Wordcount,Commas,Stopwords,Linking,WordsSentence                NA
## CLI,Commas,Stopwords,Linking,WordsSentence                      NA
## Wordcount,CLI,Commas,Stopwords,Linking,WordsSentence            NA
## Total                                                           NA
##                                                  WordsSentence.Inc
## Wordcount                                                    0.065
## CLI                                                          0.092
## Commas                                                       0.249
## Stopwords                                                    0.146
## Linking                                                      0.126
## WordsSentence                                                   NA
## Wordcount,CLI                                                0.022
## Wordcount,Commas                                             0.089
## CLI,Commas                                                   0.135
## Wordcount,Stopwords                                          0.054
## CLI,Stopwords                                                0.080
## Commas,Stopwords                                             0.184
## Wordcount,Linking                                            0.044
## CLI,Linking                                                  0.068
## Commas,Linking                                               0.177
```

```
## Stopwords,Linking                                          0.091
## Wordcount,WordsSentence                                       NA
## CLI,WordsSentence                                             NA
## Commas,WordsSentence                                          NA
## Stopwords,WordsSentence                                       NA
## Linking,WordsSentence                                         NA
## Wordcount,CLI,Commas                                       0.042
## Wordcount,CLI,Stopwords                                    0.021
## Wordcount,Commas,Stopwords                                 0.075
## CLI,Commas,Stopwords                                       0.117
## Wordcount,CLI,Linking                                      0.018
## Wordcount,Commas,Linking                                   0.069
## CLI,Commas,Linking                                         0.111
## Wordcount,Stopwords,Linking                                0.036
## CLI,Stopwords,Linking                                      0.058
## Commas,Stopwords,Linking                                   0.133
## Wordcount,CLI,WordsSentence                                   NA
## Wordcount,Commas,WordsSentence                                NA
## CLI,Commas,WordsSentence                                      NA
## Wordcount,Stopwords,WordsSentence                             NA
## CLI,Stopwords,WordsSentence                                   NA
## Commas,Stopwords,WordsSentence                                NA
## Wordcount,Linking,WordsSentence                               NA
## CLI,Linking,WordsSentence                                     NA
## Commas,Linking,WordsSentence                                  NA
## Stopwords,Linking,WordsSentence                               NA
## Wordcount,CLI,Commas,Stopwords                             0.040
## Wordcount,CLI,Commas,Linking                               0.036
## Wordcount,CLI,Stopwords,Linking                            0.017
## Wordcount,Commas,Stopwords,Linking                         0.058
## CLI,Commas,Stopwords,Linking                               0.095
## Wordcount,CLI,Commas,WordsSentence                            NA
## Wordcount,CLI,Stopwords,WordsSentence                         NA
## Wordcount,Commas,Stopwords,WordsSentence                      NA
## CLI,Commas,Stopwords,WordsSentence                            NA
## Wordcount,CLI,Linking,WordsSentence                           NA
## Wordcount,Commas,Linking,WordsSentence                        NA
## CLI,Commas,Linking,WordsSentence                              NA
## Wordcount,Stopwords,Linking,WordsSentence                     NA
## CLI,Stopwords,Linking,WordsSentence                           NA
## Commas,Stopwords,Linking,WordsSentence                        NA
## Wordcount,CLI,Commas,Stopwords,Linking                     0.034
## Wordcount,CLI,Commas,Stopwords,WordsSentence                  NA
## Wordcount,CLI,Commas,Linking,WordsSentence                    NA
## Wordcount,CLI,Stopwords,Linking,WordsSentence                 NA
## Wordcount,Commas,Stopwords,Linking,WordsSentence              NA
## CLI,Commas,Stopwords,Linking,WordsSentence                    NA
## Wordcount,CLI,Commas,Stopwords,Linking,WordsSentence          NA
## Total                                                         NA
```

```
regrOut$PredictorMetrics[,14] #GenDom Weight
```

```
##     Wordcount           CLI        Commas     Stopwords       Linking
##         0.315         0.073         0.018         0.047         0.048
```

```
## WordsSentence         Total
##         0.097         0.598
```

```
regrOut$OrderedPredictorMetrics[,14] #GenDom Order
```

```
##     Wordcount           CLI        Commas     Stopwords       Linking
##             1             3             6             5             4
## WordsSentence
##             2
```

```
regrOut$PairedDominanceMetrics #GenDom Comparisons
```

```
##                          Comp Cond Gen
## Wordcount>CLI             1.0  1.0   1
## Wordcount>Commas          1.0  1.0   1
## Wordcount>Stopwords       1.0  1.0   1
## Wordcount>Linking         1.0  1.0   1
## Wordcount>WordsSentence   1.0  1.0   1
## CLI>Commas                0.5  0.5   1
## CLI>Stopwords             0.5  1.0   1
## CLI>Linking               0.5  1.0   1
## CLI>WordsSentence         0.5  0.0   0
## Commas>Stopwords          0.5  0.5   0
## Commas>Linking            0.5  0.5   0
## Commas>WordsSentence      0.0  0.0   0
## Stopwords>Linking         0.5  0.5   0
## Stopwords>WordsSentence   0.0  0.0   0
## Linking>WordsSentence     0.0  0.0   0
```

```
# (0:Xi dominates Xj / 1:Xj dominates Xi / 0.5:Dominance not established)
result$combCIpm[,14, drop=FALSE] #DA weights with upper/lower CI
```

```
##                            GenDom
## Wordcount     0.315(0.231,0.392)
## CLI           0.073(0.034,0.127)
## Commas        0.018(0.007,0.049)
## Stopwords     0.047(0.017,0.090)
## Linking       0.048(0.015,0.103)
## WordsSentence 0.097(0.050,0.158)
```

```
result$combCIpmDiff[,"GenDom", drop=FALSE] #Comparisons of PVs
```

```
##                                      GenDom
## Wordcount-CLI              0.242(0.127,0.337)*
## Wordcount-Commas           0.297(0.202,0.378)*
## Wordcount-Stopwords        0.268(0.169,0.351)*
## Wordcount-Linking          0.267(0.150,0.355)*
## Wordcount-WordsSentence    0.218(0.107,0.313)*
## CLI-Commas                 0.055(0.001,0.111)*
## CLI-Stopwords             0.026(-0.035,0.088)
## CLI-Linking               0.025(-0.036,0.086)
```

```
## CLI-WordsSentence         -0.024(-0.099,0.053)
## Commas-Stopwords          -0.029(-0.073,0.015)
## Commas-Linking            -0.030(-0.084,0.015)
## Commas-WordsSentence     -0.079(-0.142,-0.018)*
## Stopwords-Linking         -0.001(-0.069,0.056)
## Stopwords-WordsSentence   -0.050(-0.121,0.022)
## Linking-WordsSentence     -0.049(-0.119,0.029)
```

Figure 5

```
library(ggplot2)
dat <- data.frame(
  dw = regrOut$PredictorMetrics[,14][1:(length(regrOut$PredictorMetrics[,14])-1)],
  low = result$lowerCIpm[,14],
  up = result$upperCIpm[,14])
valnames <- rownames(dat)

ggplot(dat) +
  theme(axis.text.x = element_text(size = 10),
        axis.text.y = element_text(size = 10)) +
  geom_bar(aes(x=reorder(valnames,dw), y=dw),
           stat="identity", fill="skyblue", alpha=0.6) +
  geom_errorbar(aes(x=rownames(dat), ymin=low, ymax=up),
                width=0.3, colour="red", alpha=0.9, size=0.8) +
  geom_label(aes(x=valnames, y=dw, label = dw, vjust=0.45),
             position = position_dodge(width=0.9)) +
  labs(x = "Predictor Variables",
       y = "Dominance Weights",
       caption = paste0("R-squared = ",
                        regrOut$PredictorMetrics[,14]
                        [length(regrOut$PredictorMetrics[,14])],
                        " / ",
                        "N = ", nrow(dat1)
                        )) +
  ylim(c(-0.001, 0.55)) +
  geom_text(x = 5.5, y = 0.43, label = "*", size = 5) +
  geom_segment(x = 6, xend = 6, y = 0.42, yend = 0.40, size = 0.3) +
  geom_segment(x = 6, xend = 5, y = 0.42, yend = 0.42, size = 0.3) +
  geom_segment(x = 5, xend = 5, y = 0.42, yend = 0.40, size = 0.3) +
  geom_text(x = 5, y = 0.46, label = "*", size = 5) +
  geom_segment(x = 6, xend = 6, y = 0.45, yend = 0.43, size = 0.3) +
  geom_segment(x = 6, xend = 4, y = 0.45, yend = 0.45, size = 0.3) +
  geom_segment(x = 4, xend = 4, y = 0.45, yend = 0.43, size = 0.3) +
  geom_text(x = 4.5, y = 0.49, label = "*", size = 5) +
  geom_segment(x = 6, xend = 6, y = 0.48, yend = 0.46, size = 0.3) +
  geom_segment(x = 6, xend = 3, y = 0.48, yend = 0.48, size = 0.3) +
  geom_segment(x = 3, xend = 3, y = 0.48, yend = 0.46, size = 0.3) +
  geom_text(x = 4, y = 0.52, label = "*", size = 5) +
  geom_segment(x = 6, xend = 6, y = 0.51, yend = 0.49, size = 0.3) +
```

```
  geom_segment(x = 6, xend = 2, y = 0.51, yend = 0.51, size = 0.3) +
  geom_segment(x = 2, xend = 2, y = 0.51, yend = 0.49, size = 0.3) +
  geom_text(x = 3.5, y = 0.55, label = "*", size = 5) +
  geom_segment(x = 6, xend = 6, y = 0.54, yend = 0.52, size = 0.3) +
  geom_segment(x = 6, xend = 1, y = 0.54, yend = 0.54, size = 0.3) +
  geom_segment(x = 1, xend = 1, y = 0.54, yend = 0.52, size = 0.3) +
  geom_text(x = 3, y = 0.23, label = "*", size = 5) +
  geom_segment(x = 5, xend = 5, y = 0.22, yend = 0.20, size = 0.3) +
  geom_segment(x = 5, xend = 1, y = 0.22, yend = 0.22, size = 0.3) +
  geom_segment(x = 1, xend = 1, y = 0.22, yend = 0.20, size = 0.3) +
  geom_text(x = 2.5, y = 0.18, label = "*", size = 5) +
  geom_segment(x = 4, xend = 4, y = 0.17, yend = 0.15, size = 0.3) +
  geom_segment(x = 4, xend = 1, y = 0.17, yend = 0.17, size = 0.3) +
  geom_segment(x = 1, xend = 1, y = 0.17, yend = 0.15, size = 0.3) +
  coord_flip()
```

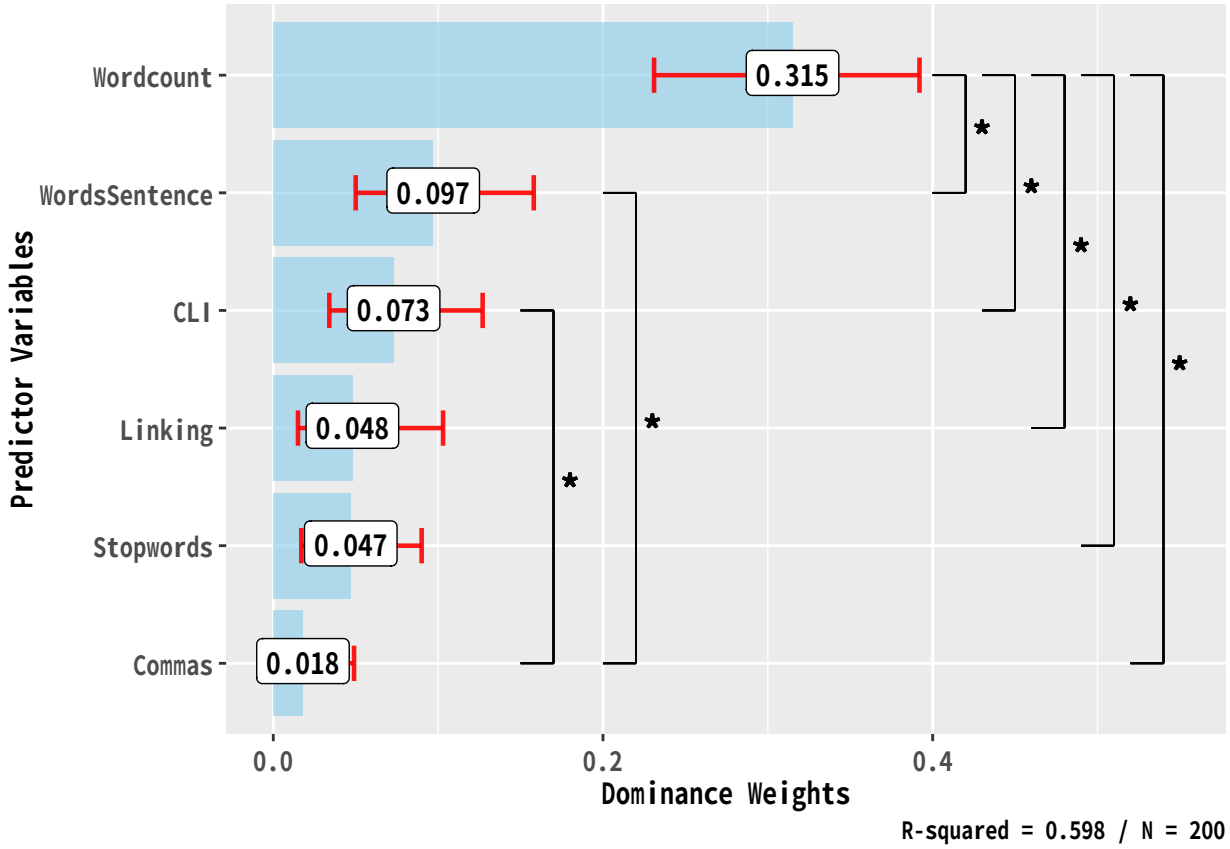


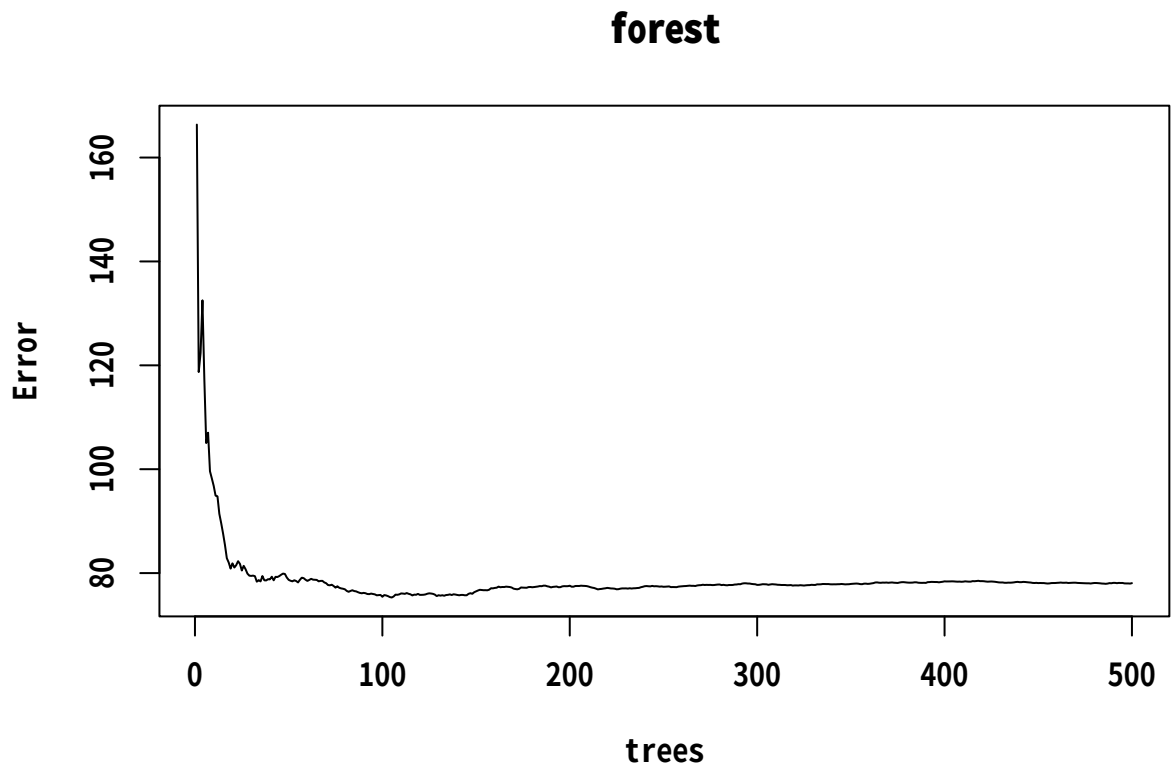
Figure 6

```
# Random Forest
library(randomForest)
```

```
forest <- randomForest(Score~., data=dat1)
print(forest)
```

```
##
## Call:
##  randomForest(formula = Score ~ ., data = dat1)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 2
##
##           Mean of squared residuals: 78.06284
##                     % Var explained: 46.94
```

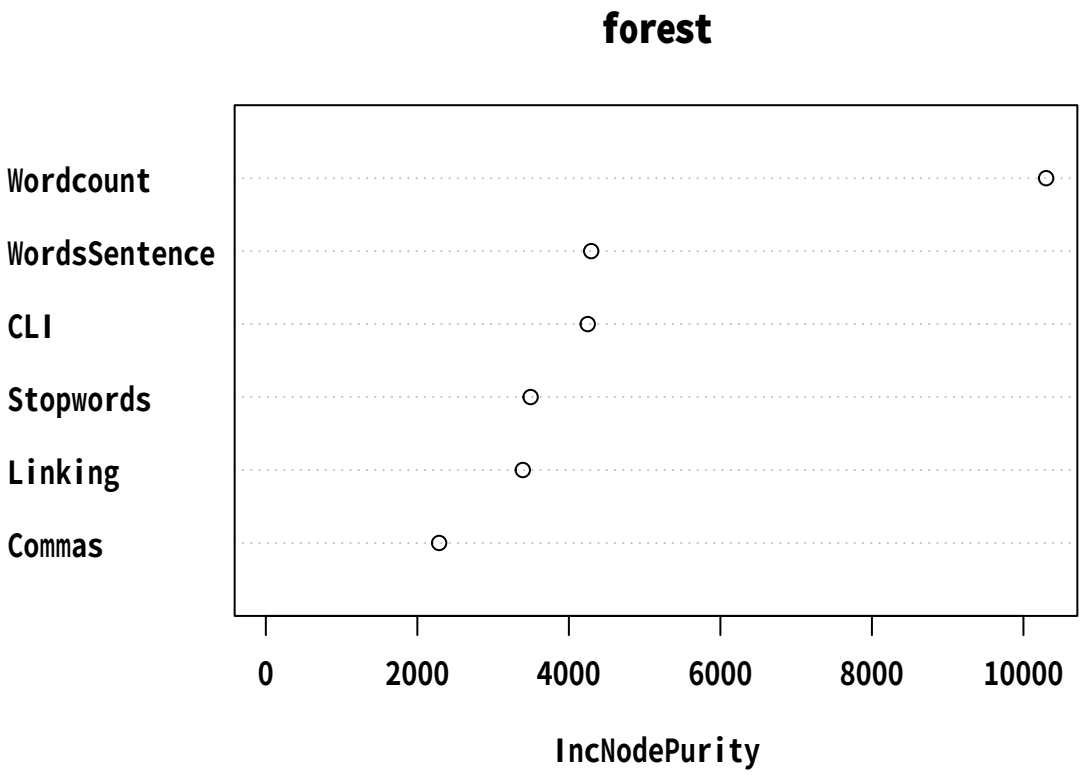
```
plot(forest)
```




```
forest$importance
```

```
##               IncNodePurity
## Wordcount         10299.564
## CLI                4248.687
## Commas             2287.207
## Stopwords          3494.619
## Linking            3391.637
## WordsSentence      4294.565
```

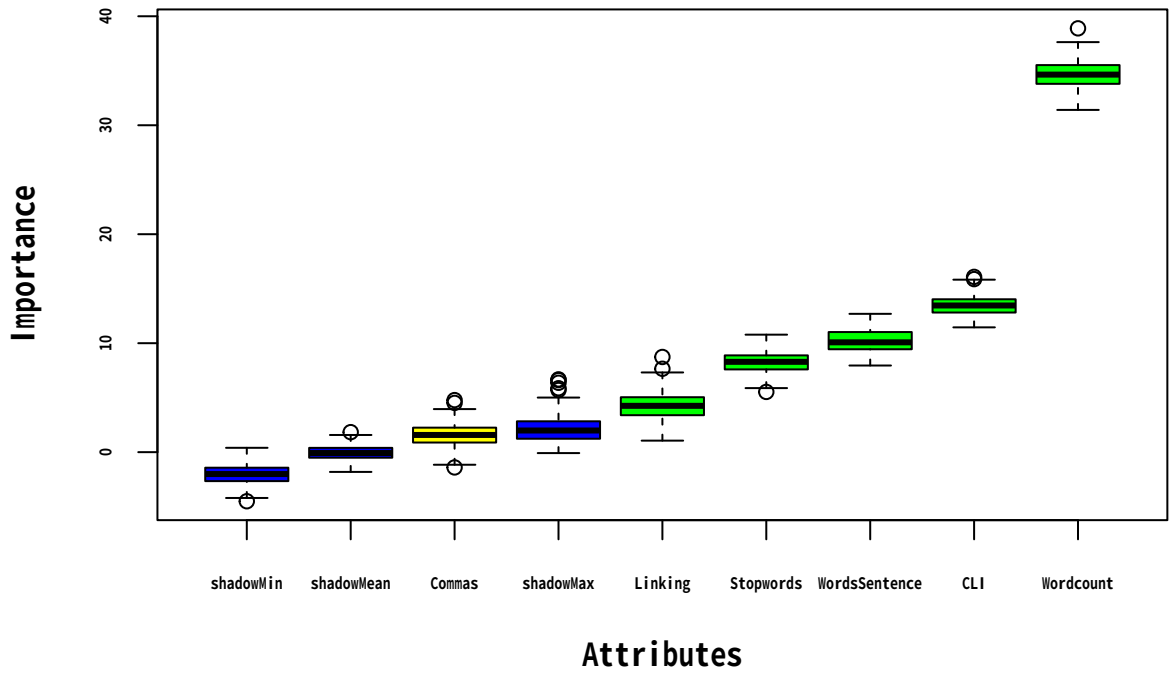
```
varImpPlot(forest)
```




```
# Boruta
library(Boruta)
set.seed(88)
boruta <- Boruta(Score~., maxRuns = 200, data = dat1, doTrace = 2)
print(boruta)
```

```
## Boruta performed 199 iterations in 4.142513 secs.
##  5 attributes confirmed important: CLI, Linking, Stopwords, Wordcount,
## WordsSentence;
##  No attributes deemed unimportant.
##  1 tentative attributes left: Commas;
```

```
plot(boruta, cex.axis = 0.5)
```




```
attStats(boruta)
```

```
##                 meanImp medianImp    minImp    maxImp  normHits  decision
## Wordcount     34.689979 34.647507 31.406242 38.891381 1.0000000 Confirmed
## CLI           13.486228 13.460720 11.452826 16.084404 1.0000000 Confirmed
## Commas         1.562773  1.580389 -1.400805  4.777331 0.4120603 Tentative
## Stopwords      8.269685  8.289149  5.528578 10.782055 1.0000000 Confirmed
## Linking        4.278329  4.251459  1.058463  8.735209 0.9145729 Confirmed
## WordsSentence 10.214595 10.077946  7.951354 12.693180 1.0000000 Confirmed
```

Cross-validation of Algorithms

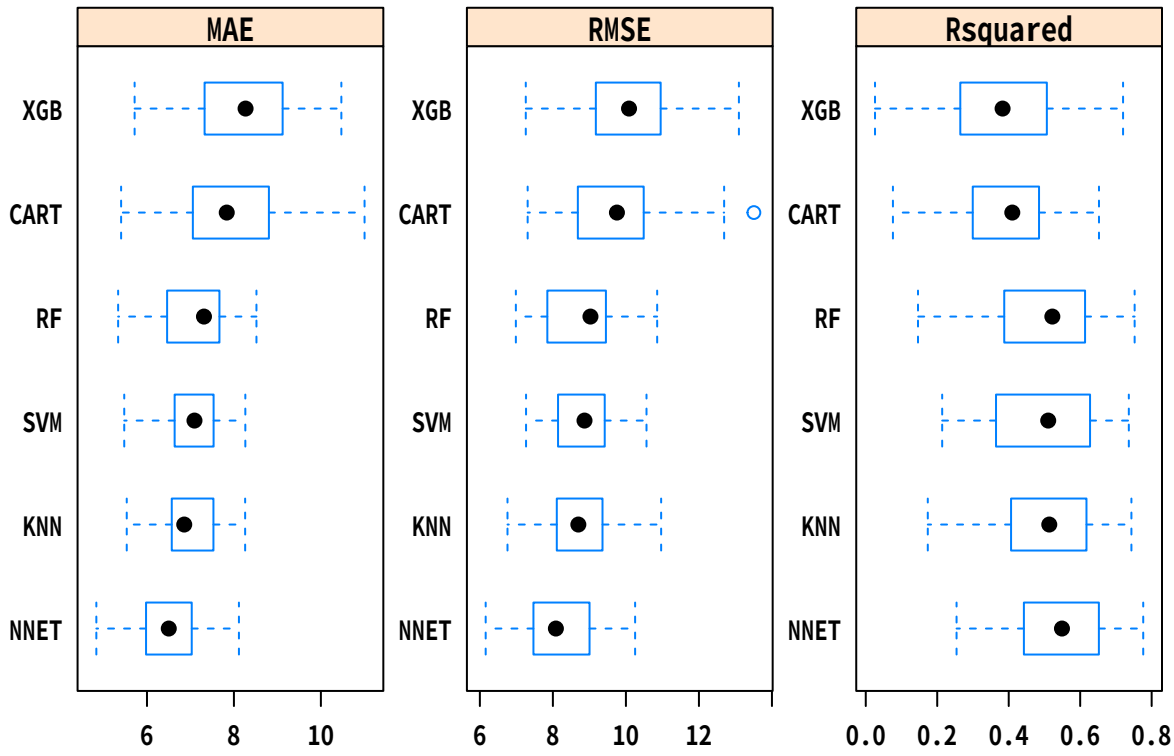
```
library(caret)

# prepare training scheme
control <- trainControl(method="repeatedcv", number=10, repeats=3)

# CART (Classification and Regression Trees)
set.seed(88)
fit.cart <- train(Score~., data=dat1, method="rpart", trControl=control,
```

```
                 tuneGrid = data.frame(cp = c(0.01, 0.05, 0.1)),
                 preProcess = c('center', 'scale'))

# SVM (Support Vector Machine with Radial Basis Function)
set.seed(88)
fit.svm <- train(Score~., data=dat1, method="svmRadial",
                 trControl=control, preProcess = c('center', 'scale'))

# kNN (k-Nearest Neighbors)
set.seed(88)
fit.knn <- train(Score~., data=dat1, method="knn", trControl=control,
                 preProcess = c('center', 'scale'))

# Random Forest
set.seed(88)
fit.rf <- train(Score~., data=dat1, method="rf", trControl=control,
                preProcess = c('center', 'scale'))

# Neural Network
set.seed(88)
fit.nnet <- train(Score~., data = dat1, method="nnet", trControl = control,
                  preProcess = c('center', 'scale'), linout = TRUE, trace=FALSE)

# XgboostLinear (eXtreme Gradient Boosting)
set.seed(88)
fit.xgbLinear <- train(Score~., data=dat1, method="xgbLinear",
                       trControl=control, preProcess = c('center', 'scale'))

# Collect Resamples
results <- resamples(list(CART=fit.cart, SVM=fit.svm, KNN=fit.knn,
                          RF=fit.rf, NNET=fit.nnet, XGB=fit.xgbLinear))

# Summarize the Result
summary(results)
```

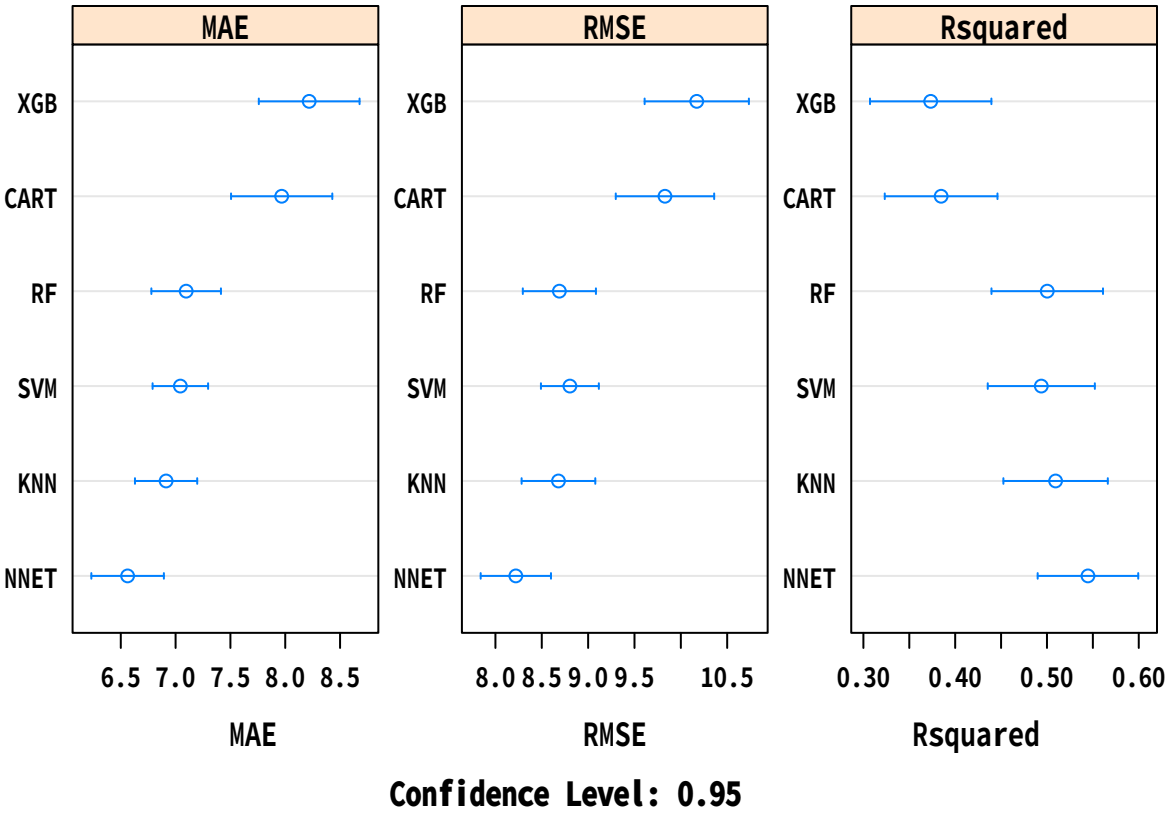
```
##
## Call:
## summary.resamples(object = results)
##
## Models: CART, SVM, KNN, RF, NNET, XGB
## Number of resamples: 30
##
## MAE
##          Min.  1st Qu.   Median     Mean  3rd Qu.      Max. NA's
## CART 5.405289 7.082223 7.834562 7.967500 8.772692 11.004552    0
## SVM  5.477013 6.653310 7.092682 7.043372 7.511497  8.262595    0
## KNN  5.534898 6.585466 6.857098 6.913275 7.528687  8.257881    0
## RF   5.336864 6.492356 7.311149 7.096340 7.639370  8.518174    0
## NNET 4.835460 6.007423 6.501087 6.562389 7.021524  8.114789    0
## XGB  5.715725 7.341998 8.267864 8.218345 9.030610 10.470718    0
##
## RMSE
##          Min.  1st Qu.    Median      Mean   3rd Qu.     Max. NA's
```

```
## CART 7.308737 8.685144  9.753906  9.828851 10.457049 13.50553    0
## SVM  7.264931 8.143864  8.866044  8.803258  9.394488 10.56388    0
## KNN  6.756455 8.110122  8.697439  8.679857  9.298676 10.96518    0
## RF   6.986130 7.849983  9.028978  8.689779  9.440174 10.85638    0
## NNET 6.159081 7.478173  8.082193  8.220080  8.997473 10.25245    0
## XGB  7.257378 9.226271 10.085816 10.171664 10.918776 13.09532    0
##
## Rsquared
##            Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## CART 0.07599797 0.3007584 0.4099998 0.3846782 0.4739510 0.6525623    0
## SVM  0.21383764 0.3809937 0.5105411 0.4938134 0.6269018 0.7360675    0
## KNN  0.17358565 0.4159621 0.5134891 0.5094005 0.6172054 0.7433024    0
## RF   0.14616005 0.3887521 0.5221258 0.5002814 0.6121381 0.7523641    0
## NNET 0.25403572 0.4497454 0.5490355 0.5446481 0.6509964 0.7763809    0
## XGB  0.02587729 0.2682327 0.3828831 0.3733386 0.5036111 0.7199376    0
```

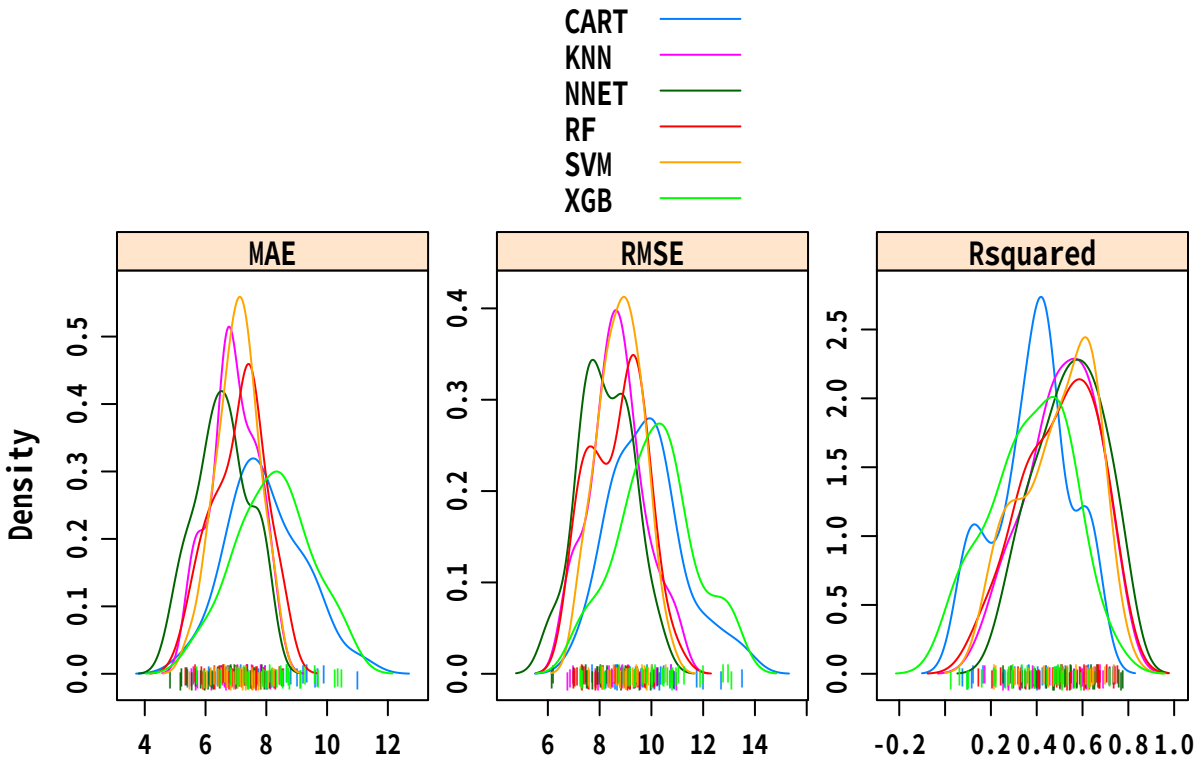
```r
# Box and Whisker Plots to Compare Models
scales <- list(x=list(relation="free"), y=list(relation="free"))
bwplot(results, scales=scales)
```



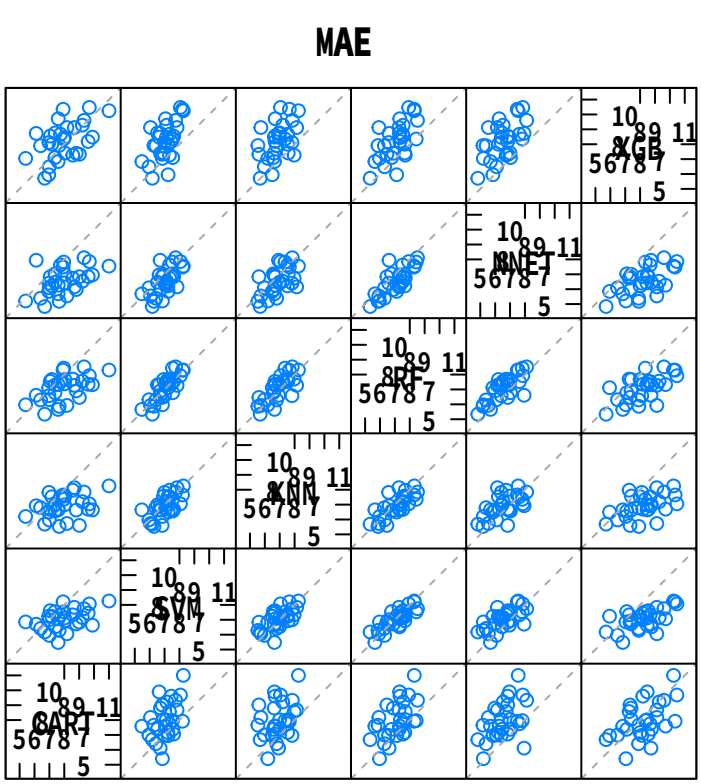

```r
# Dot plots of Accuracy with 95% Confidence Intervals
scales <- list(x=list(relation="free"), y=list(relation="free"))
dotplot(results, scales=scales)
```




```
# Density Plots of Accuracy
# to evaluate the overlap in the estimated behavior of algorithms
scales <- list(x=list(relation="free"), y=list(relation="free"))
densityplot(results, scales=scales, pch = "|", auto.key=T)
```




```
# Pair-wise Scatterplots of Predictions
splom(results)
```




```
# Difference in Model Predictions
options(scipen=999) # to turn off scientific notation (e notation)
diffs <- diff(results)
# Pair-wise Comparisons
summary(diffs)
```

```
##
## Call:
## summary.diff.resamples(object = diffs)
##
## p-value adjustment: bonferroni
## Upper diagonal: estimates of the difference
## Lower diagonal: p-value for H0: difference = 0
##
## MAE
##      CART          SVM           KNN           RF            NNET
## CART                0.92413       1.05422       0.87116       1.40511
## SVM  0.0014572                    0.13010      -0.05297       0.48098
## KNN  0.0007683     1.0000000                   -0.18306       0.35089
## RF   0.0015145     1.0000000     1.0000000                    0.53395
## NNET 0.00000320413 0.0071499     0.2713067     0.0002615
## XGB  1.0000000     0.00000720097 0.00000424694 0.00000878952 0.00000001222
##      XGB
## CART -0.25084
## SVM  -1.17497
## KNN  -1.30507
```

```
## RF   -1.12201
## NNET -1.65596
## XGB
##
## RMSE
##      CART          SVM       KNN           RF            NNET
## CART                1.025592  1.148993      1.139071      1.608771
## SVM  0.0004926               0.123401      0.113479      0.583178
## KNN  0.0004799     1.0000000               -0.009922     0.459777
## RF   0.0001225     1.0000000 1.0000000                   0.469699
## NNET 0.00000288416 0.0172183 0.3064224     0.0376564
## XGB  1.0000000     0.0001076 0.00009399864 0.00000494955 0.00000007451
##      XGB
## CART -0.342814
## SVM  -1.368406
## KNN  -1.491807
## RF   -1.481885
## NNET -1.951584
## XGB
##
## Rsquared
##      CART          SVM       KNN           RF            NNET          XGB
## CART               -0.109135 -0.124722     -0.115603     -0.159970      0.011340
## SVM  0.0017174               -0.015587     -0.006468     -0.050835      0.120475
## KNN  0.0017921     1.0000000                0.009119     -0.035248      0.136062
## RF   0.0029558     1.0000000 1.0000000                   -0.044367      0.126943
## NNET 0.0000463215  0.1419321 1.0000000     0.1281312                    0.171310
## XGB  1.0000000     0.0001136 0.0000060545  0.0000002974  0.0000008489
```

The results of cross-validation of machine learning algorithms above suggest that the prediction of the random forest model is better than or almost equal to that of other equally powerful machine learning models.

End of Document