Supplement code and figures

Xingmin Aaron Zhang

Factorize some variables

```
attach(mtcars)
mtcars$cyl = as.factor(mtcars$cyl)
mtcars$vs = as.factor(mtcars$vs)
mtcars$am = as.factor(mtcars$am)
mtcars$gear = as.factor(mtcars$gear)
mtcars$carb = as.factor(mtcars$carb)
```

We create pairwise plot to look at the relationship between each variables.

```
# plot inspired by http://www.sthda.com/english/wiki/scatter-plot-matrices-r-base-graphs
panel.cor <- function(x, y){
    usr <- par("usr"); on.exit(par(usr))
    par(usr = c(0, 1, 0, 1))
    r <- round(cor(as.numeric(x), as.numeric(y)), digits=2)
    txt <- paste0("R = ", r)
    cex.cor <- 0.8/strwidth(txt)
    text(0.5, 0.5, txt, cex = cex.cor * r)
}

upper.panel<-function(x, y){
  points(x,y, pch = 19, cex = 0.5)
}
pairs(mtcars[,-1], gap = 0.5, lower.panel = panel.cor, upper.panel = upper.panel,
      main ="mtcars: regressor correlation")
```

VIF of all variables

```
fit <- lm(mpg ~ ., data = mtcars)
library(car)
```

```
## Loading required package: carData
```

```
sqrt(vif(fit))
```

```
##           GVIF       Df GVIF^(1/(2*Df))
## cyl  11.319053 1.414214        1.834225
## disp  7.769536 1.000000        2.787389
## hp    5.312210 1.000000        2.304823
## drat  2.609533 1.000000        1.615405
## wt    4.881683 1.000000        2.209453
## qsec  3.284842 1.000000        1.812413
## vs    2.843970 1.000000        1.686407
## am    3.151269 1.000000        1.775181
## gear  7.131081 1.414214        1.634138
## carb 22.432384 2.236068        1.364858
```

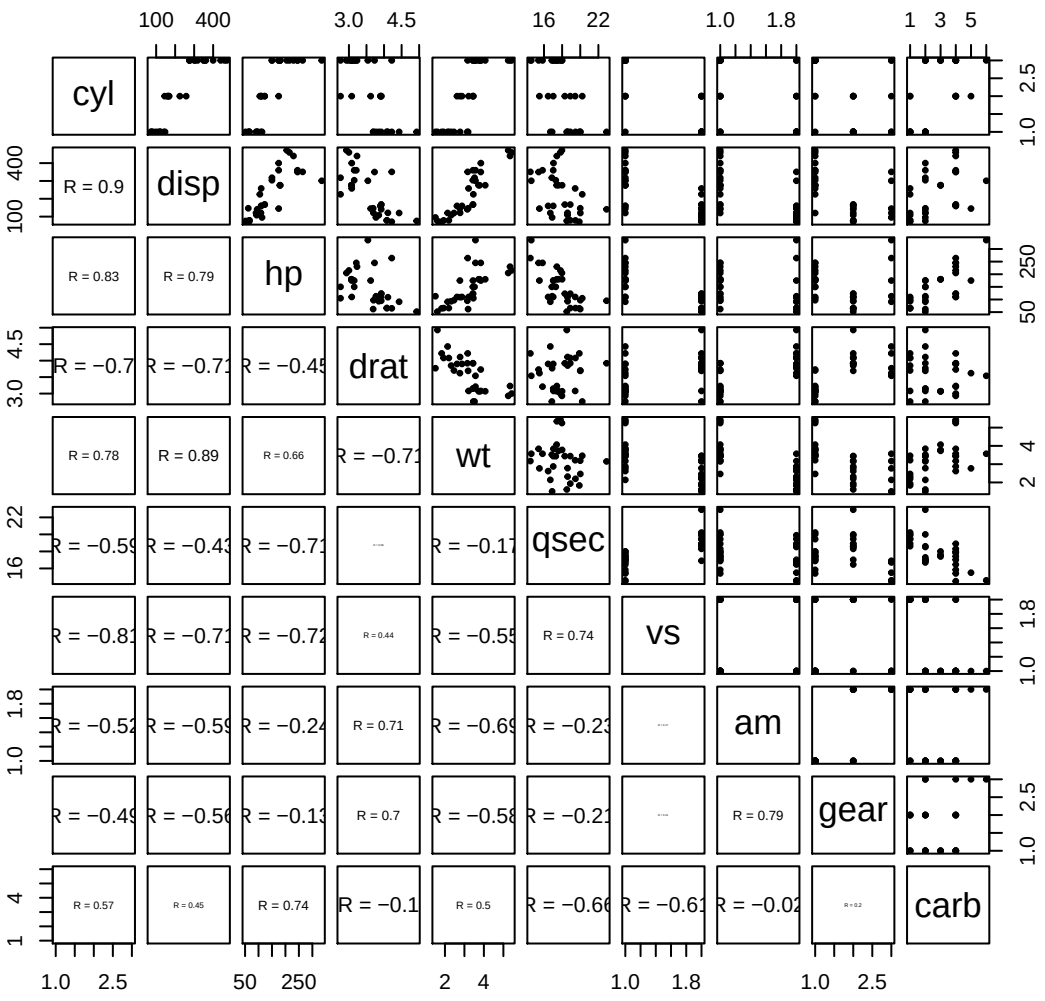

Figure 1: correlation matrix

VIF after removing hp

```
variables <- colnames(mtcars)
fit2 <- lm(mpg ~ ., data = mtcars[,variables != "hp"])
sqrt(vif(fit2))
```

```
##           GVIF       Df GVIF^(1/(2*Df))
## cyl   9.187241 1.414214        1.740990
## disp  7.563500 1.000000        2.750182
## drat  2.608198 1.000000        1.614992
## wt    4.845050 1.000000        2.201147
## qsec  3.269540 1.000000        1.808187
## vs    2.627953 1.000000        1.621096
## am    3.150134 1.000000        1.774862
## gear  6.314643 1.414214        1.585211
## carb 13.323707 2.236068        1.295575
```

VIF after removing both hp and disp

```
fit3 <- lm(mpg ~ ., data = mtcars[,variables != "hp" & variables != "disp"])
sqrt(vif(fit3))
```

```
##          GVIF       Df GVIF^(1/(2*Df))
## cyl  5.664030 1.414214        1.542700
## drat 2.607671 1.000000        1.614829
## wt   3.154976 1.000000        1.776225
## qsec 3.247714 1.000000        1.802142
## vs   2.611426 1.000000        1.615991
## am   2.961625 1.000000        1.720937
## gear 5.057345 1.414214        1.499618
## carb 7.120020 2.236068        1.216881
```

Build linear models and check the significance of residue reduction by adding one variable a time.

```
mtcars_subset <- mtcars[,variables != "hp" & variables != "disp"]
model1 <- lm(mpg ~ am, data = mtcars_subset)
model2 <- update(model1, mpg ~ am + cyl)
model3 <- update(model2, mpg ~ am + cyl + drat)
model4 <- update(model3, mpg ~ am + cyl + drat + wt)
model5 <- update(model4, mpg ~ am + cyl + drat + wt + qsec)
model6 <- update(model5, mpg ~ am + cyl + drat + wt + qsec + vs)
model7 <- update(model6, mpg ~ am + cyl + drat + wt + qsec + vs + gear)
model8 <- update(model7, mpg ~ am + cyl + drat + wt + qsec + vs + gear + carb)
anova(model1, model2, model3, model4, model5, model6, model7, model8)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl
## Model 3: mpg ~ am + cyl + drat
## Model 4: mpg ~ am + cyl + drat + wt
## Model 5: mpg ~ am + cyl + drat + wt + qsec
```

```
## Model 6: mpg ~ am + cyl + drat + wt + qsec + vs
## Model 7: mpg ~ am + cyl + drat + wt + qsec + vs + gear
## Model 8: mpg ~ am + cyl + drat + wt + qsec + vs + gear + carb
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     28 264.50  2    456.40 25.8134 7.057e-06 ***
## 3     27 264.32  1      0.17  0.0195  0.890559
## 4     26 182.75  1     81.57  9.2274  0.007433 **
## 5     25 159.14  1     23.61  2.6709  0.120580
## 6     24 159.14  1      0.00  0.0000  0.995586
## 7     22 158.86  2      0.27  0.0155  0.984625
## 8     17 150.29  5      8.58  0.1940  0.960629
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparison of final linear model with and without interation terms.

```
model <- lm(mpg ~ am + cyl + wt, data = mtcars_subset)
model_int <- update(model, mpg ~ am * cyl * wt )
anova(model, model_int)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am + cyl + wt
## Model 2: mpg ~ am + cyl + wt + am:cyl + am:wt + cyl:wt + am:cyl:wt
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     27 182.97
## 2     20 116.91  7    66.059 1.6144 0.1886
```