SubsetSelectionForward-Backward.R

vijaykalmath

2022-01-04

```r
# Subset Selection using Forward and Backware Stepwise Selection.

library(ISLR)

names(Hitters)
```

```
##  [1] "AtBat"     "Hits"      "HmRun"     "Runs"      "RBI"       "Walks"
##  [7] "Years"     "CAtBat"    "CHits"     "CHmRun"    "CRuns"     "CRBI"
## [13] "CWalks"    "League"    "Division"  "PutOuts"   "Assists"   "Errors"
## [19] "Salary"    "NewLeague"
```

```r
dim(Hitters)
```

```
## [1] 322  20
```

```r
# Removing NA from Hitters

Hitters = na.omit(Hitters)

# Dimensions after removing NA.
dim(Hitters)
```

```
## [1] 263  20
```

```r
# the Regsubsets function is part of the leaps library

library(leaps)

# calling regsubsets  with forward method

regfit.fwd = regsubsets(Salary~.,Hitters,nvmax=19,method="forward")


regfit.fwd.summary = summary(regfit.fwd)

regfit.fwd.summary$which
```
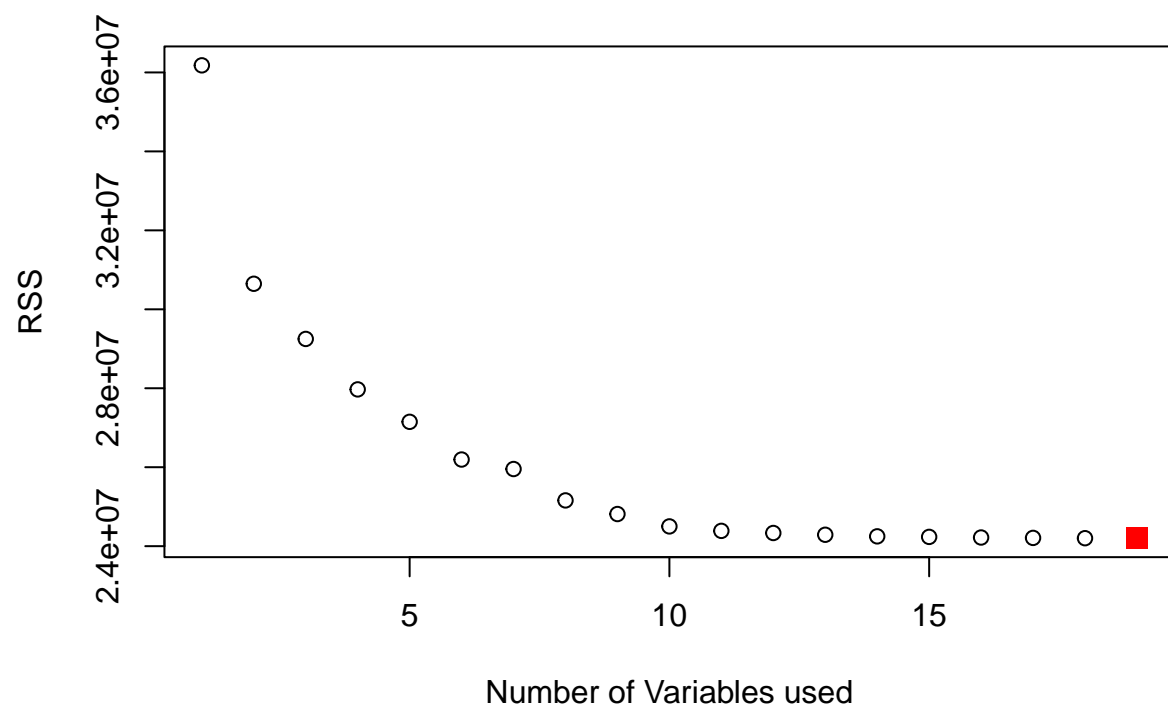
```
##   (Intercept) AtBat  Hits HmRun  Runs   RBI Walks Years CAtBat CHits CHmRun
## 1        TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
```

```
## 2           TRUE FALSE   TRUE FALSE FALSE FALSE FALSE FALSE   FALSE FALSE   FALSE
## 3           TRUE FALSE   TRUE FALSE FALSE FALSE FALSE FALSE   FALSE FALSE   FALSE
## 4           TRUE FALSE   TRUE FALSE FALSE FALSE FALSE FALSE   FALSE FALSE   FALSE
## 5           TRUE  TRUE   TRUE FALSE FALSE FALSE FALSE FALSE   FALSE FALSE   FALSE
## 6           TRUE  TRUE   TRUE FALSE FALSE FALSE  TRUE FALSE   FALSE FALSE   FALSE
## 7           TRUE  TRUE   TRUE FALSE FALSE FALSE  TRUE FALSE   FALSE FALSE   FALSE
## 8           TRUE  TRUE   TRUE FALSE FALSE FALSE  TRUE FALSE   FALSE FALSE   FALSE
## 9           TRUE  TRUE   TRUE FALSE FALSE FALSE  TRUE FALSE    TRUE FALSE   FALSE
## 10          TRUE  TRUE   TRUE FALSE FALSE FALSE  TRUE FALSE    TRUE FALSE   FALSE
## 11          TRUE  TRUE   TRUE FALSE FALSE FALSE  TRUE FALSE    TRUE FALSE   FALSE
## 12          TRUE  TRUE   TRUE FALSE  TRUE FALSE  TRUE FALSE    TRUE FALSE   FALSE
## 13          TRUE  TRUE   TRUE FALSE  TRUE FALSE  TRUE FALSE    TRUE FALSE   FALSE
## 14          TRUE  TRUE   TRUE  TRUE  TRUE FALSE  TRUE FALSE    TRUE FALSE   FALSE
## 15          TRUE  TRUE   TRUE  TRUE  TRUE FALSE  TRUE FALSE    TRUE  TRUE   FALSE
## 16          TRUE  TRUE   TRUE  TRUE  TRUE  TRUE  TRUE FALSE    TRUE  TRUE   FALSE
## 17          TRUE  TRUE   TRUE  TRUE  TRUE  TRUE  TRUE FALSE    TRUE  TRUE   FALSE
## 18          TRUE  TRUE   TRUE  TRUE  TRUE  TRUE  TRUE  TRUE    TRUE  TRUE   FALSE
## 19          TRUE  TRUE   TRUE  TRUE  TRUE  TRUE  TRUE  TRUE    TRUE  TRUE    TRUE
##      CRuns CRBI CWalks LeagueN DivisionW PutOuts Assists Errors NewLeagueN
## 1  FALSE TRUE  FALSE   FALSE     FALSE   FALSE   FALSE  FALSE      FALSE
## 2  FALSE TRUE  FALSE   FALSE     FALSE   FALSE   FALSE  FALSE      FALSE
## 3  FALSE TRUE  FALSE   FALSE     FALSE    TRUE   FALSE  FALSE      FALSE
## 4  FALSE TRUE  FALSE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 5  FALSE TRUE  FALSE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 6  FALSE TRUE  FALSE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 7  FALSE TRUE   TRUE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 8   TRUE TRUE   TRUE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 9   TRUE TRUE   TRUE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 10  TRUE TRUE   TRUE   FALSE      TRUE    TRUE    TRUE  FALSE      FALSE
## 11  TRUE TRUE   TRUE    TRUE      TRUE    TRUE    TRUE  FALSE      FALSE
## 12  TRUE TRUE   TRUE    TRUE      TRUE    TRUE    TRUE  FALSE      FALSE
## 13  TRUE TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE      FALSE
## 14  TRUE TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE      FALSE
## 15  TRUE TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE      FALSE
## 16  TRUE TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE      FALSE
## 17  TRUE TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE       TRUE
## 18  TRUE TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE       TRUE
## 19  TRUE TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE       TRUE
```
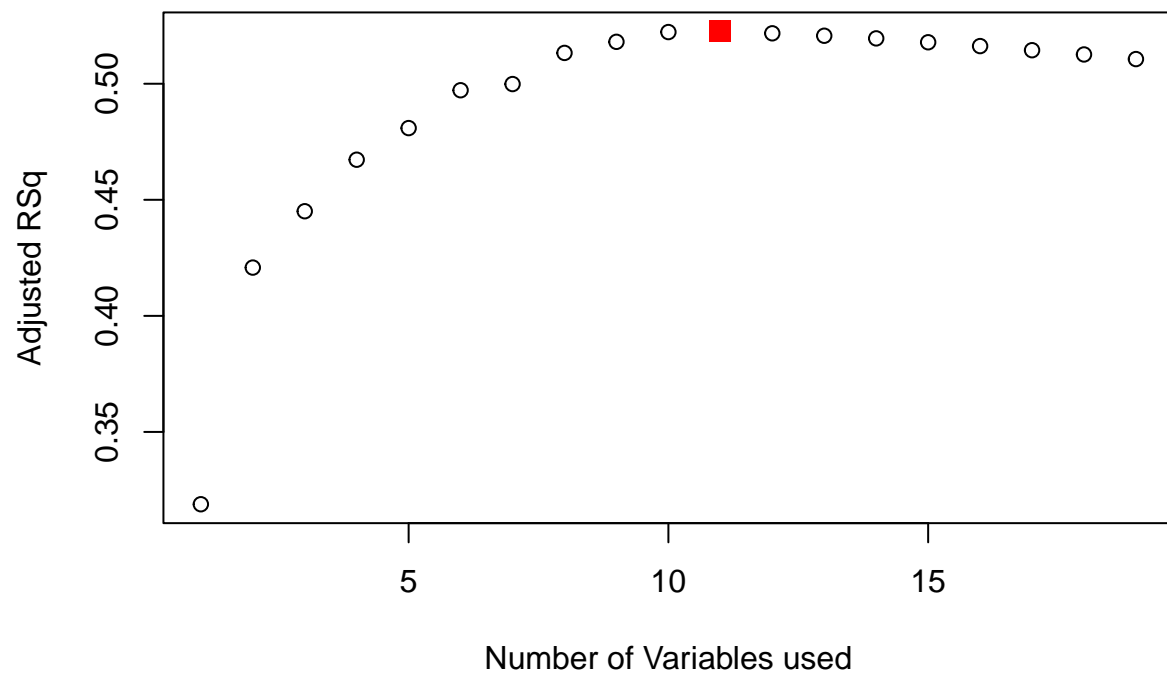
```
regfit.fwd.summary$rss
```

```
##  [1] 36179679 30646560 29249297 27970852 27149899 26194904 25954217 25159234
##  [9] 24814051 24500402 24387345 24333232 24289148 24248660 24235177 24219377
## [17] 24209447 24201837 24200700
```

```
plot(regfit.fwd.summary$rss, xlab = "Number of Variables used",ylab="RSS") +
points(which.min(regfit.fwd.summary$rss),min(regfit.fwd.summary$rss),col="red",cex=1.5,pch=15)
```
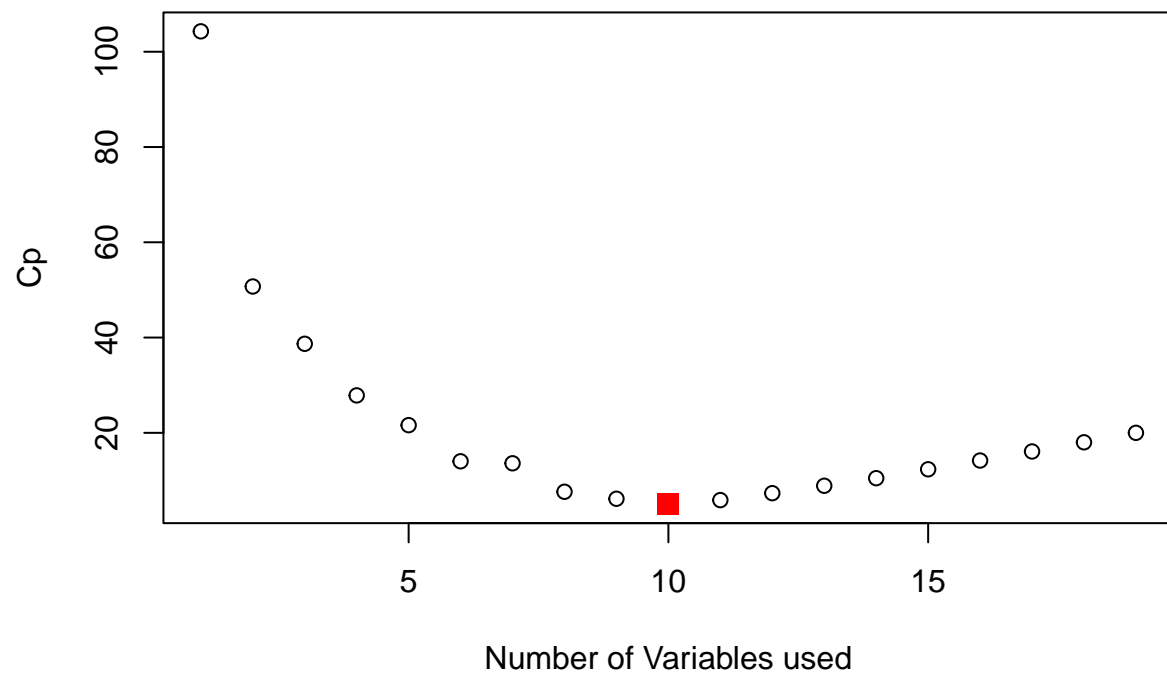
```
## integer(0)
```

```
plot(regfit.fwd.summary$adjr2, xlab = "Number of Variables used",ylab="Adjusted RSq") +
points(which.max(regfit.fwd.summary$adjr2),max(regfit.fwd.summary$adjr2),col="red",cex=1.5,pch=15)
```
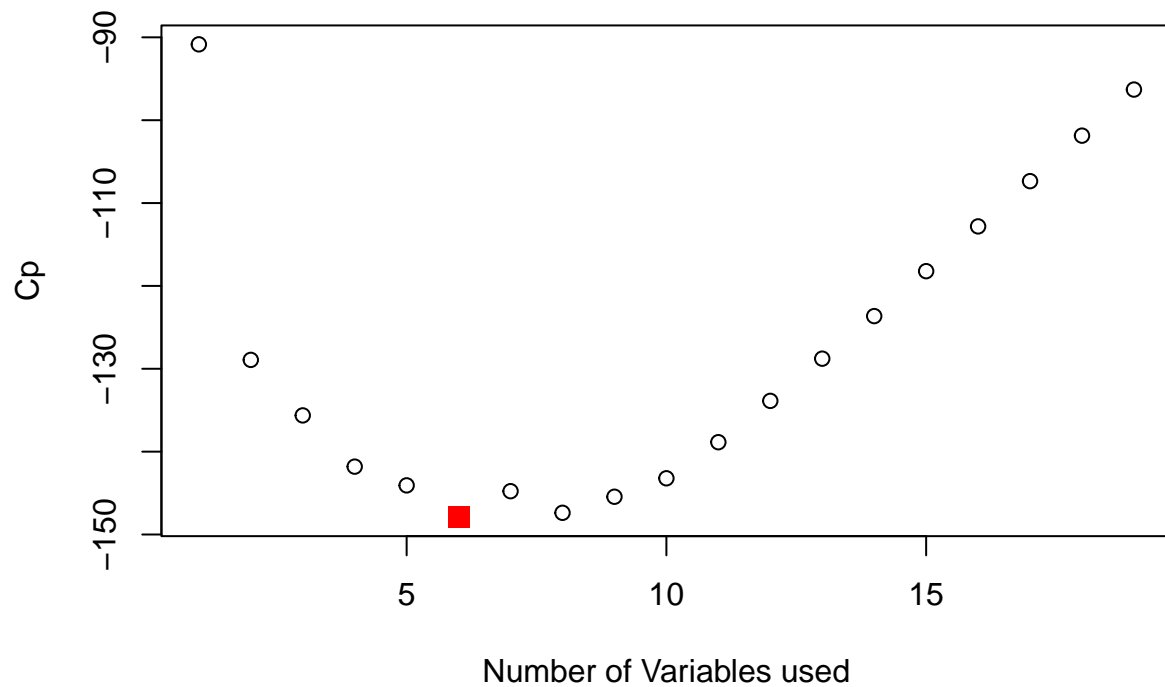
```
## integer(0)
```

```
plot(regfit.fwd.summary$cp, xlab = "Number of Variables used",ylab="Cp") +
points(which.min(regfit.fwd.summary$cp),min(regfit.fwd.summary$cp),col="red",cex=1.5,pch=15)
```

```
## integer(0)
```

```
plot(regfit.fwd.summary$bic, xlab = "Number of Variables used",ylab="Cp") + points(which.min(regfit.fwd
```

```
## integer(0)
```

```
# Backward Selection
```

```
regfit.bckwd = regsubsets(Salary~.,Hitters,nvmax=19,method="backward")
```

```
regfit.bckwd.summary = summary(regfit.bckwd)
```

```
regfit.bckwd.summary$which
```
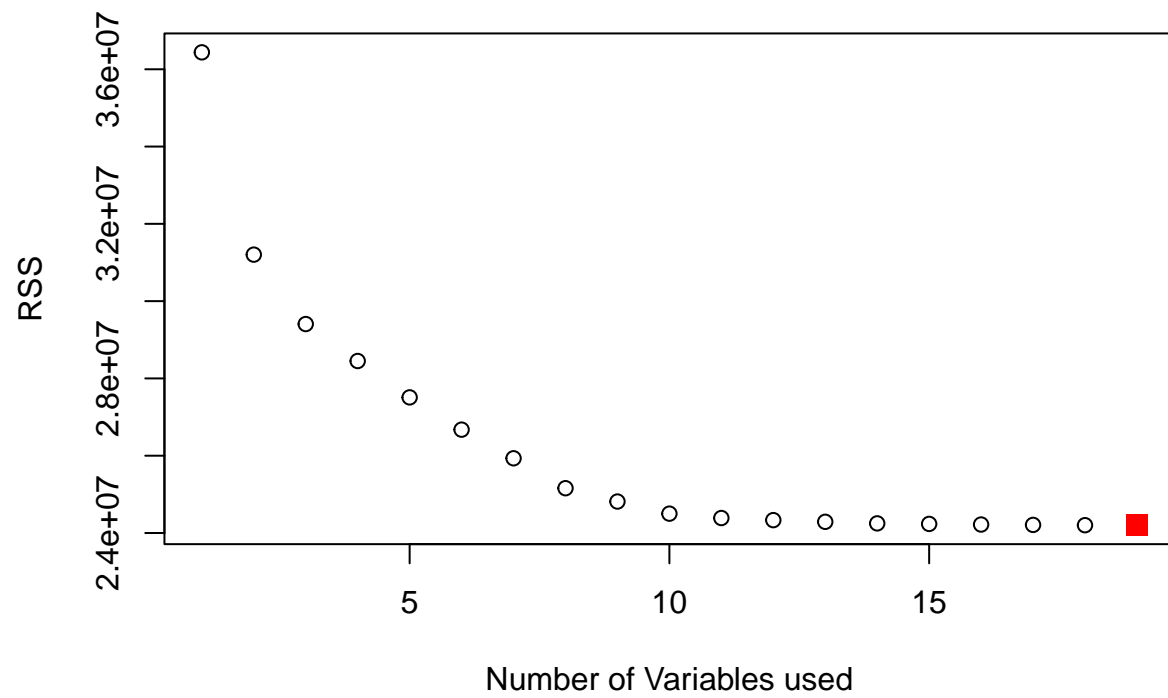
```
##    (Intercept) AtBat  Hits HmRun  Runs   RBI Walks Years CAtBat CHits CHmRun
## 1         TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 2         TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 3         TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 4         TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 5         TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE  FALSE FALSE  FALSE
## 6         TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE  FALSE FALSE  FALSE
## 7         TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE  FALSE FALSE  FALSE
## 8         TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE  FALSE FALSE  FALSE
## 9         TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE   TRUE FALSE  FALSE
## 10        TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE   TRUE FALSE  FALSE
## 11        TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE   TRUE FALSE  FALSE
```

```
## 12          TRUE   TRUE   TRUE FALSE   TRUE FALSE   TRUE FALSE    TRUE FALSE   FALSE
## 13          TRUE   TRUE   TRUE FALSE   TRUE FALSE   TRUE FALSE    TRUE FALSE   FALSE
## 14          TRUE   TRUE   TRUE  TRUE   TRUE FALSE   TRUE FALSE    TRUE FALSE   FALSE
## 15          TRUE   TRUE   TRUE  TRUE   TRUE FALSE   TRUE FALSE    TRUE  TRUE   FALSE
## 16          TRUE   TRUE   TRUE  TRUE   TRUE  TRUE   TRUE FALSE    TRUE  TRUE   FALSE
## 17          TRUE   TRUE   TRUE  TRUE   TRUE  TRUE   TRUE FALSE    TRUE  TRUE   FALSE
## 18          TRUE   TRUE   TRUE  TRUE   TRUE  TRUE   TRUE  TRUE    TRUE  TRUE   FALSE
## 19          TRUE   TRUE   TRUE  TRUE   TRUE  TRUE   TRUE  TRUE    TRUE  TRUE    TRUE
##      CRuns  CRBI CWalks LeagueN DivisionW PutOuts Assists Errors NewLeagueN
## 1    TRUE FALSE  FALSE   FALSE     FALSE   FALSE   FALSE  FALSE      FALSE
## 2    TRUE FALSE  FALSE   FALSE     FALSE   FALSE   FALSE  FALSE      FALSE
## 3    TRUE FALSE  FALSE   FALSE     FALSE    TRUE   FALSE  FALSE      FALSE
## 4    TRUE FALSE  FALSE   FALSE     FALSE    TRUE   FALSE  FALSE      FALSE
## 5    TRUE FALSE  FALSE   FALSE     FALSE    TRUE   FALSE  FALSE      FALSE
## 6    TRUE FALSE  FALSE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 7    TRUE FALSE   TRUE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 8    TRUE  TRUE   TRUE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 9    TRUE  TRUE   TRUE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 10   TRUE  TRUE   TRUE   FALSE      TRUE    TRUE    TRUE  FALSE      FALSE
## 11   TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE  FALSE      FALSE
## 12   TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE  FALSE      FALSE
## 13   TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE      FALSE
## 14   TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE      FALSE
## 15   TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE      FALSE
## 16   TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE      FALSE
## 17   TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE       TRUE
## 18   TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE       TRUE
## 19   TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE       TRUE
```
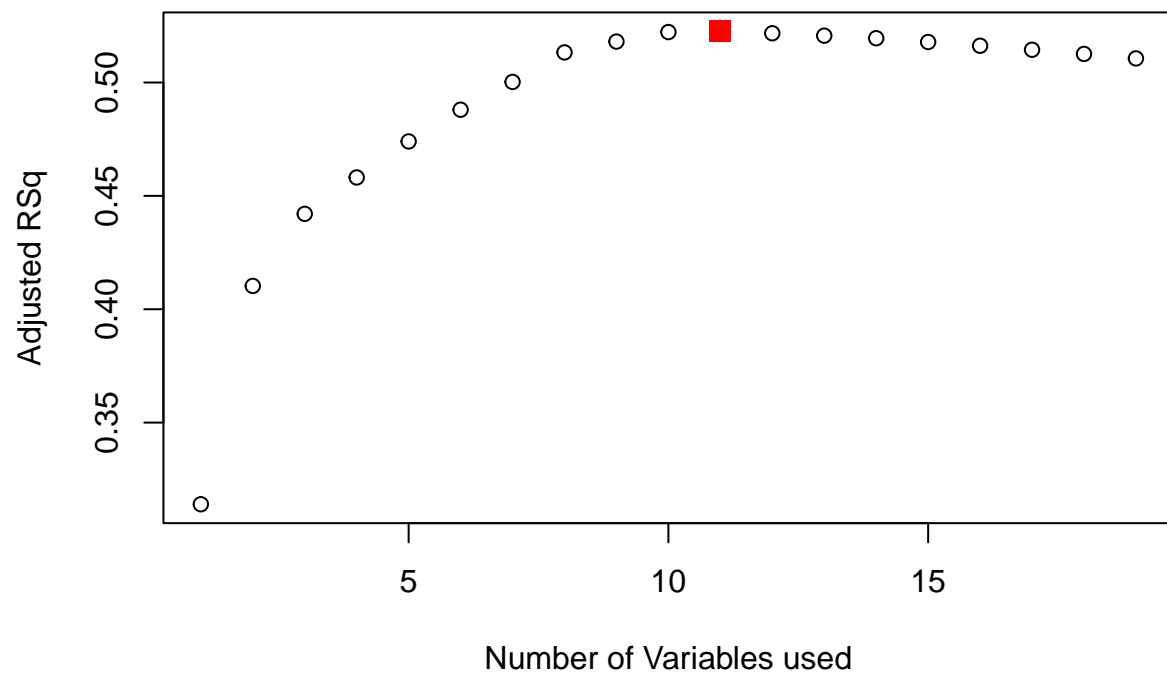
```
regfit.bckwd.summary$rss
```

```
##  [1] 36437951 31203460 29407297 28450807 27509524 26674092 25933487 25159234
##  [9] 24814051 24500402 24387345 24333232 24289148 24248660 24235177 24219377
## [17] 24209447 24201837 24200700
```

```r
plot(regfit.bckwd.summary$rss, xlab = "Number of Variables used",ylab="RSS") +
  points(which.min(regfit.bckwd.summary$rss),min(regfit.bckwd.summary$rss),col="red",cex=1.5,pch=15)
```
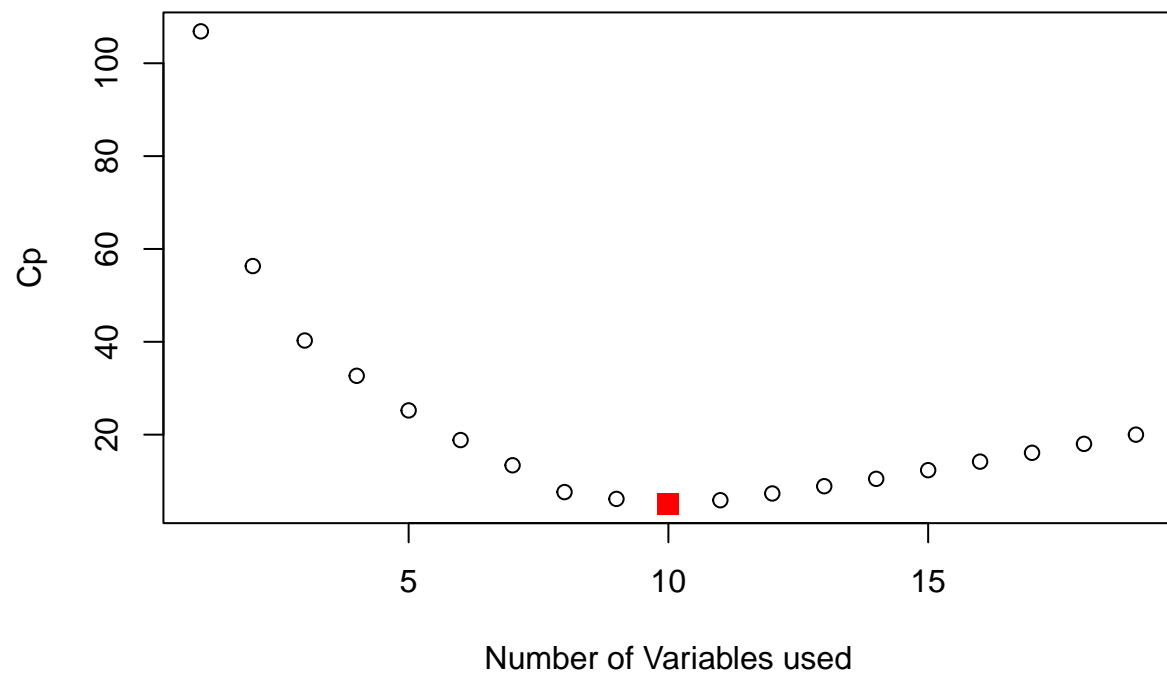
```
## integer(0)
```

```
plot(regfit.bckwd.summary$adjr2, xlab = "Number of Variables used",ylab="Adjusted RSq") +
  points(which.max(regfit.bckwd.summary$adjr2),max(regfit.bckwd.summary$adjr2),col="red",cex=1.5,pch=15)
```
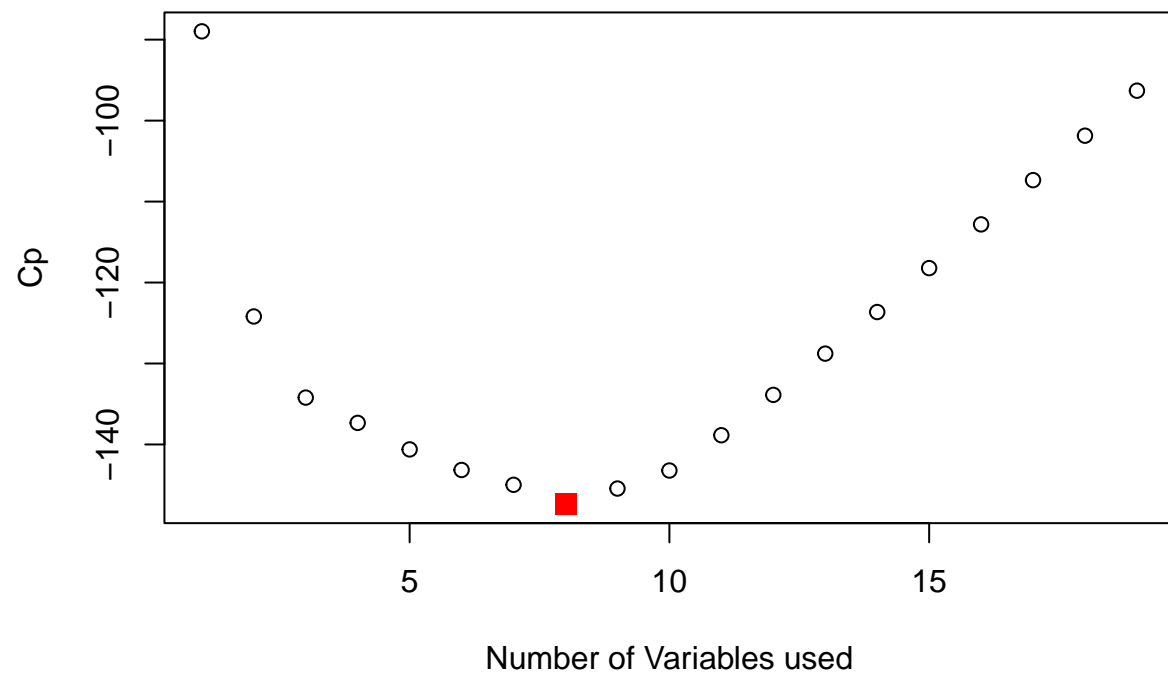
```
## integer(0)
```

```
plot(regfit.bckwd.summary$cp, xlab = "Number of Variables used",ylab="Cp") +
  points(which.min(regfit.bckwd.summary$cp),min(regfit.bckwd.summary$cp),col="red",cex=1.5,pch=15)
```

```
## integer(0)
```

```
plot(regfit.bckwd.summary$bic, xlab = "Number of Variables used",ylab="Cp") +
points(which.min(regfit.bckwd.summary$bic),min(regfit.bckwd.summary$bic),col="red",cex=1.5,pch=15)
```

```
## integer(0)
```