# Building Multi-Variate Models on London BikeSharing Data

Zhengyuan Shen

```
library(alr4)
```

```
## Loading required package: car
```

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

```
## Loading required package: effects
```

```
## Registered S3 methods overwritten by 'lme4':
##   method                          from
##   cooks.distance.influence.merMod car
##   influence.merMod                car
##   dfbeta.influence.merMod         car
##   dfbetas.influence.merMod        car
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
library(leaps)
library(MASS)
library(pls)
```

```
##
## Attaching package: 'pls'
```

```
## The following object is masked from 'package:stats':
##
##     loadings
```

```
library(trafo)
```

```
## Registered S3 method overwritten by 'pryr':
##   method      from
##   print.bytes Rcpp
```

```
##
## Attaching package: 'trafo'
```

```
## The following object is masked from 'package:MASS':
##
##     boxcox
```

```
library(VGAM)
```

```
## Loading required package: stats4
```

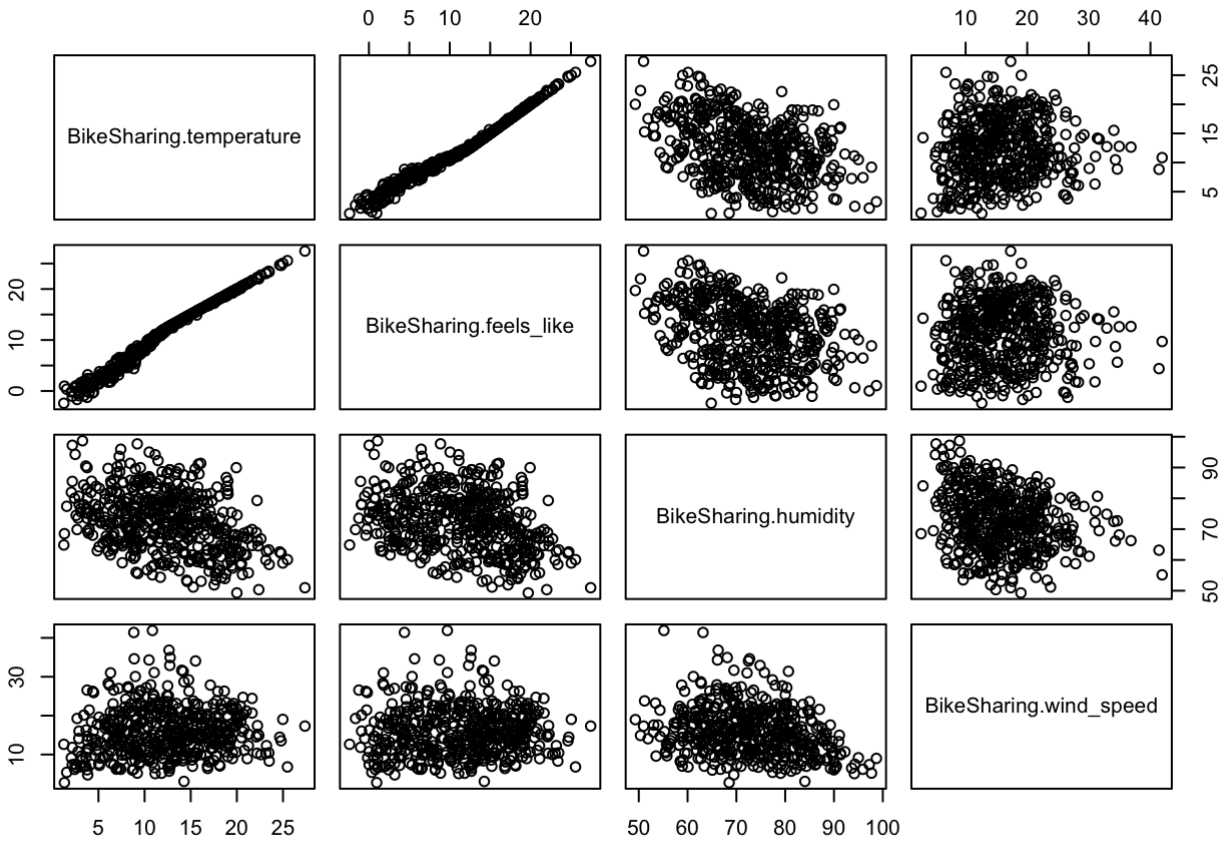
```
## Loading required package: splines
```

```
##
## Attaching package: 'VGAM'
```

```
## The following object is masked from 'package:trafo':
##
##     reciprocal
```

```
## The following object is masked from 'package:car':
##
##     logit
```

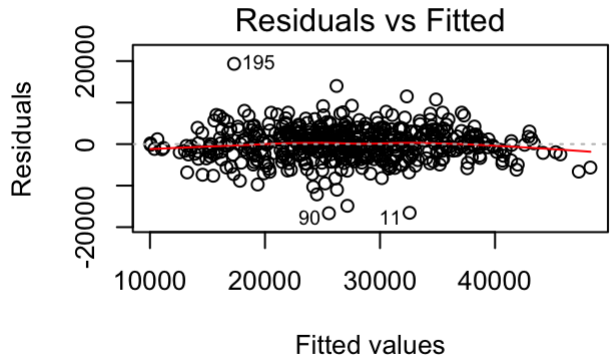
```
BikeSharing <- read.delim("bikesharing18.txt",sep = "")
pairs(data.frame(BikeSharing$temperature,BikeSharing$feels_like,BikeSharing$humidity,Bik
eSharing$wind_speed))
```

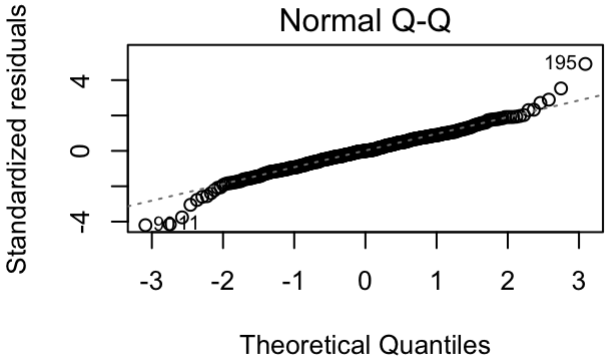


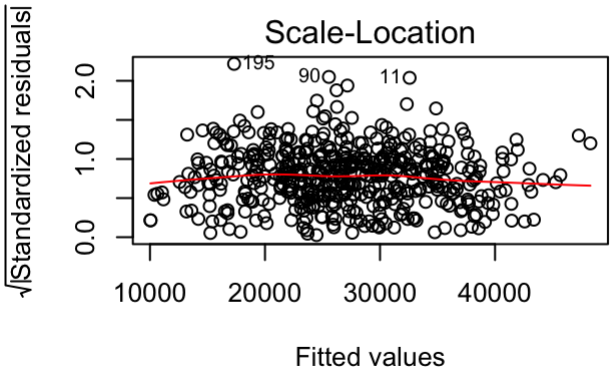

```
m1  = lm(N_bikes~temperature+feels_like+humidity+wind_speed+holiday+weekend+season,data
= BikeSharing)
summary(m1)
```

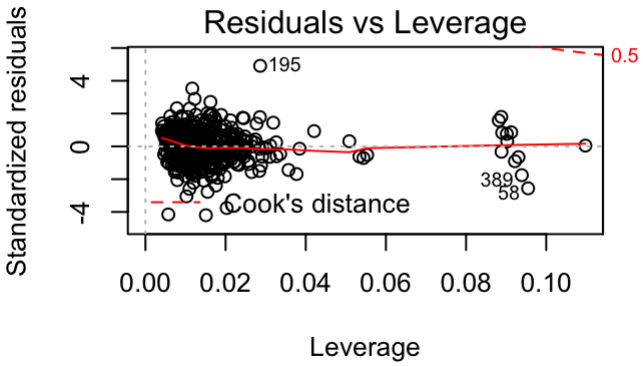
```
##
## Call:
## lm(formula = N_bikes ~ temperature + feels_like + humidity +
##     wind_speed + holiday + weekend + season, data = BikeSharing)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16648.7  -2430.3      7.2   2610.3  19338.3
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 51893.71    2175.84  23.850  < 2e-16 ***
## temperature   226.95     321.35   0.706   0.4804
## feels_like    545.91     265.94   2.053   0.0406 *
## humidity     -363.12      22.68 -16.008  < 2e-16 ***
## wind_speed   -357.92      32.07 -11.162  < 2e-16 ***
## holiday     -8153.36    1178.70  -6.917 1.44e-11 ***
## weekend     -5237.73     400.70 -13.072  < 2e-16 ***
## season        -87.24     183.91  -0.474   0.6354
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3994 on 492 degrees of freedom
## Multiple R-squared:  0.7722, Adjusted R-squared:  0.7689
## F-statistic: 238.2 on 7 and 492 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,2))
plot(m1)
```




```
m_all = lm(N_bikes~feels_like+humidity+wind_speed+holiday+weekend+feels_like*holiday+fee
ls_like*weekend+wind_speed*holiday+humidity*holiday+humidity*weekend+wind_speed*weekend,
data = BikeSharing)
Anova(m_all,type='II')
```

| | Sum Sq | Df | F value | Pr(>F) |
|---|---|---|---|---|
| | <dbl> | <dbl> | <dbl> | <dbl> |
| feels_like | 8984527787 | 1 | 590.69052053 | 4.316673e-86 |
| humidity | 5224559626 | 1 | 343.49026663 | 1.894593e-58 |
| wind_speed | 2292817818 | 1 | 150.74200696 | 2.204192e-30 |
| holiday | 822036833 | 1 | 54.04506240 | 8.323053e-13 |
| weekend | 2713945757 | 1 | 178.42919184 | 6.602646e-35 |
| feels_like:holiday | 2559828 | 1 | 0.16829668 | 6.818101e-01 |
| feels_like:weekend | 135348499 | 1 | 8.89852839 | 2.996902e-03 |
| wind_speed:holiday | 1065859 | 1 | 0.07007519 | 7.913391e-01 |
| humidity:holiday | 54368322 | 1 | 3.57446194 | 5.926661e-02 |
| humidity:weekend | 76597248 | 1 | 5.03590944 | 2.527455e-02 |

1-10 of 12 rows Previous **1** 2 Next

# Best Subset Selection

```
m.select =regsubsets(N_bikes~temperature+feels_like+humidity+wind_speed+holiday+weekend+
season+
                       temperature*holiday+temperature*weekend+temperature*season+
                       feels_like*holiday+feels_like*weekend+feels_like*season+
                       wind_speed*holiday+wind_speed*weekend+wind_speed*season+
                       humidity*holiday+humidity*weekend+humidity*season,data=BikeSharin
g,nvmax=20)
reg.summary = summary(m.select)
reg.summary
```

```
## Subset selection object
## Call: regsubsets.formula(N_bikes ~ temperature + feels_like + humidity +
##     wind_speed + holiday + weekend + season + temperature * holiday +
##     temperature * weekend + temperature * season + feels_like *
##     holiday + feels_like * weekend + feels_like * season + wind_speed *
##     holiday + wind_speed * weekend + wind_speed * season + humidity *
##     holiday + humidity * weekend + humidity * season, data = BikeSharing,
##     nvmax = 20)
## 19 Variables  (and intercept)
##                      Forced in Forced out
## temperature              FALSE      FALSE
## feels_like               FALSE      FALSE
## humidity                 FALSE      FALSE
## wind_speed               FALSE      FALSE
## holiday                  FALSE      FALSE
## weekend                  FALSE      FALSE
## season                   FALSE      FALSE
## temperature:holiday      FALSE      FALSE
## temperature:weekend      FALSE      FALSE
## temperature:season       FALSE      FALSE
## feels_like:holiday       FALSE      FALSE
## feels_like:weekend       FALSE      FALSE
## feels_like:season        FALSE      FALSE
## wind_speed:holiday       FALSE      FALSE
## wind_speed:weekend       FALSE      FALSE
## wind_speed:season        FALSE      FALSE
## humidity:holiday         FALSE      FALSE
## humidity:weekend         FALSE      FALSE
## humidity:season          FALSE      FALSE
## 1 subsets of each size up to 19
## Selection Algorithm: exhaustive
##           temperature feels_like humidity wind_speed holiday weekend season
## 1  ( 1 )  " "         "*"        " "      " "        " "     " "     " "
## 2  ( 1 )  " "         "*"        "*"      " "        " "     " "     " "
## 3  ( 1 )  " "         "*"        "*"      " "        " "     " "     " "
## 4  ( 1 )  " "         "*"        "*"      "*"        " "     " "     " "
## 5  ( 1 )  " "         "*"        "*"      "*"        " "     " "     " "
## 6  ( 1 )  " "         "*"        "*"      "*"        " "     " "     " "
## 7  ( 1 )  " "         "*"        "*"      "*"        " "     " "     " "
## 8  ( 1 )  " "         "*"        "*"      "*"        " "     " "     " "
## 9  ( 1 )  " "         "*"        "*"      "*"        " "     " "     " "
## 10  ( 1 ) " "         "*"        "*"      "*"        " "     " "     "*"
## 11  ( 1 ) " "         "*"        "*"      "*"        "*"     " "     "*"
## 12  ( 1 ) " "         "*"        "*"      "*"        "*"     " "     " "
## 13  ( 1 ) " "         "*"        "*"      "*"        "*"     "*"     " "
## 14  ( 1 ) " "         "*"        "*"      "*"        "*"     "*"     " "
## 15  ( 1 ) "*"         "*"        "*"      "*"        "*"     "*"     "*"
## 16  ( 1 ) " "         "*"        "*"      "*"        "*"     "*"     " "
## 17  ( 1 ) "*"         "*"        "*"      "*"        "*"     "*"     " "
## 18  ( 1 ) "*"         "*"        "*"      "*"        "*"     "*"     "*"
## 19  ( 1 ) "*"         "*"        "*"      "*"        "*"     "*"     "*"
##           temperature:holiday temperature:weekend temperature:season
## 1  ( 1 )  " "                 " "                 " "
```

```
## 2  ( 1 )  " "                  " "                   " "
## 3  ( 1 )  " "                  " "                   " "
## 4  ( 1 )  " "                  " "                   " "
## 5  ( 1 )  " "                  " "                   " "
## 6  ( 1 )  " "                  " "                   " "
## 7  ( 1 )  " "                  " "                   " "
## 8  ( 1 )  " "                  " "                   " "
## 9  ( 1 )  " "                  " "                   " "
## 10  ( 1 ) " "                  " "                   " "
## 11  ( 1 ) " "                  " "                   " "
## 12  ( 1 ) " "                  " "                   "*"
## 13  ( 1 ) " "                  " "                   "*"
## 14  ( 1 ) " "                  "*"                   "*"
## 15  ( 1 ) " "                  "*"                   " "
## 16  ( 1 ) "*"                  "*"                   "*"
## 17  ( 1 ) "*"                  "*"                   "*"
## 18  ( 1 ) "*"                  "*"                   "*"
## 19  ( 1 ) "*"                  "*"                   "*"
##           feels_like:holiday feels_like:weekend feels_like:season
## 1  ( 1 )  " "                  " "                   " "
## 2  ( 1 )  " "                  " "                   " "
## 3  ( 1 )  " "                  " "                   " "
## 4  ( 1 )  " "                  " "                   " "
## 5  ( 1 )  " "                  " "                   " "
## 6  ( 1 )  " "                  "*"                   " "
## 7  ( 1 )  " "                  "*"                   " "
## 8  ( 1 )  " "                  "*"                   "*"
## 9  ( 1 )  " "                  "*"                   "*"
## 10  ( 1 ) " "                  "*"                   "*"
## 11  ( 1 ) " "                  "*"                   "*"
## 12  ( 1 ) " "                  "*"                   "*"
## 13  ( 1 ) " "                  "*"                   "*"
## 14  ( 1 ) " "                  "*"                   "*"
## 15  ( 1 ) " "                  "*"                   "*"
## 16  ( 1 ) "*"                  "*"                   "*"
## 17  ( 1 ) "*"                  "*"                   "*"
## 18  ( 1 ) "*"                  "*"                   "*"
## 19  ( 1 ) "*"                  "*"                   "*"
##           wind_speed:holiday wind_speed:weekend wind_speed:season
## 1  ( 1 )  " "                  " "                   " "
## 2  ( 1 )  " "                  " "                   " "
## 3  ( 1 )  " "                  "*"                   " "
## 4  ( 1 )  " "                  " "                   " "
## 5  ( 1 )  " "                  " "                   " "
## 6  ( 1 )  " "                  " "                   " "
## 7  ( 1 )  " "                  "*"                   " "
## 8  ( 1 )  " "                  "*"                   " "
## 9  ( 1 )  " "                  "*"                   " "
## 10  ( 1 ) " "                  "*"                   "*"
## 11  ( 1 ) " "                  "*"                   "*"
## 12  ( 1 ) " "                  "*"                   "*"
## 13  ( 1 ) " "                  "*"                   "*"
## 14  ( 1 ) " "                  "*"                   "*"
## 15  ( 1 ) "*"                  "*"                   "*"
```

```
## 16  ( 1 ) " "                 "*"                 "*"
## 17  ( 1 ) " "                 "*"                 "*"
## 18  ( 1 ) " "                 "*"                 "*"
## 19  ( 1 ) "*"                 "*"                 "*"
##           humidity:holiday humidity:weekend humidity:season
## 1  ( 1 )  " "               " "               " "
## 2  ( 1 )  " "               " "               " "
## 3  ( 1 )  " "               " "               " "
## 4  ( 1 )  " "               "*"               " "
## 5  ( 1 )  "*"               "*"               " "
## 6  ( 1 )  "*"               "*"               " "
## 7  ( 1 )  "*"               "*"               " "
## 8  ( 1 )  "*"               "*"               " "
## 9  ( 1 )  "*"               "*"               "*"
## 10  ( 1 ) "*"               "*"               " "
## 11  ( 1 ) "*"               "*"               " "
## 12  ( 1 ) "*"               "*"               "*"
## 13  ( 1 ) "*"               "*"               "*"
## 14  ( 1 ) "*"               "*"               "*"
## 15  ( 1 ) "*"               "*"               " "
## 16  ( 1 ) "*"               "*"               "*"
## 17  ( 1 ) "*"               "*"               "*"
## 18  ( 1 ) "*"               "*"               "*"
## 19  ( 1 ) "*"               "*"               "*"
```

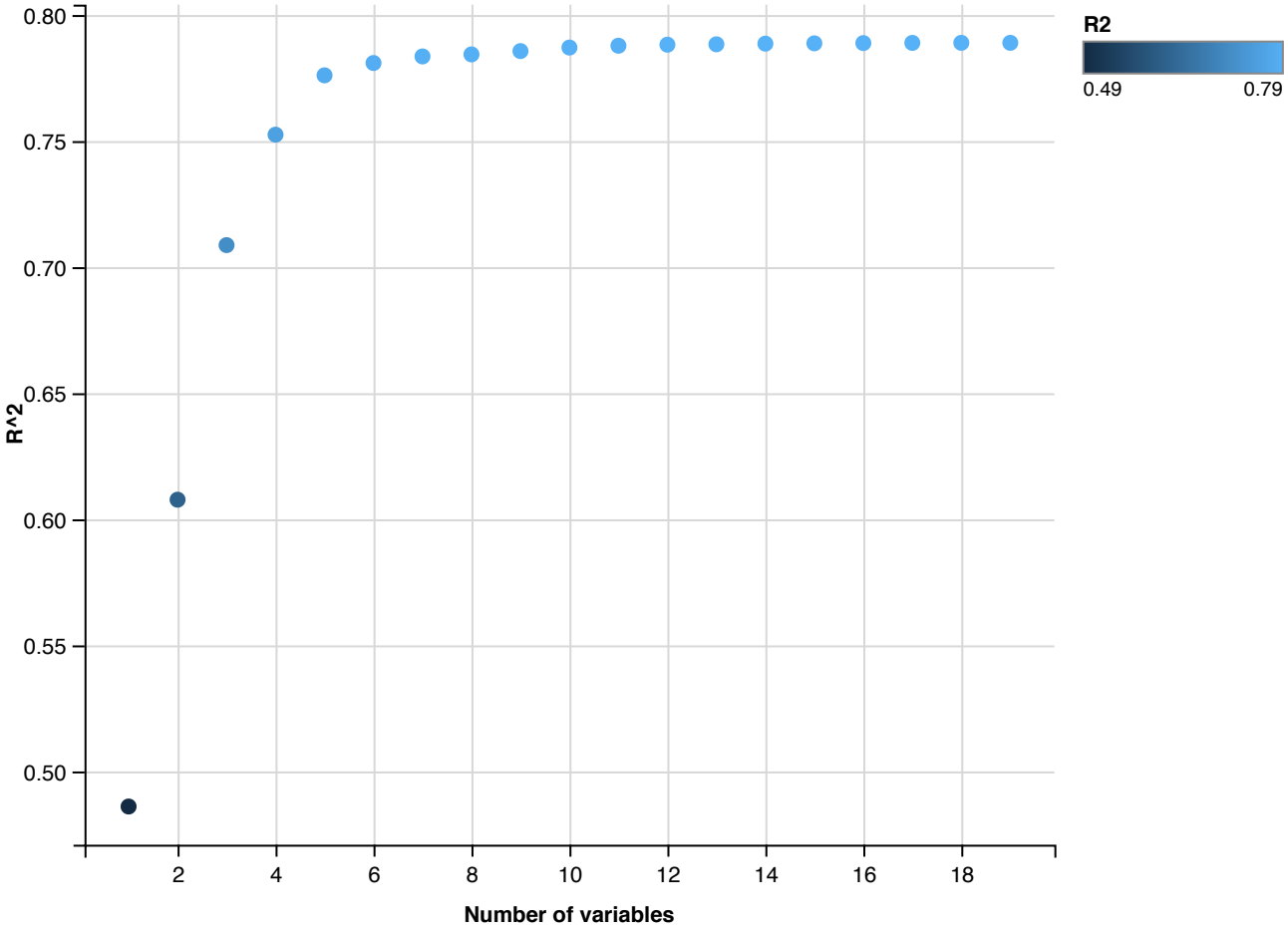
```
reg.summary$adjr2
```

```
##  [1] 0.4850612 0.6062035 0.7069845 0.7505448 0.7738142 0.7783012 0.7804956
##  [8] 0.7808679 0.7817611 0.7827414 0.7830404 0.7830140 0.7827299 0.7825421
## [15] 0.7822240 0.7818762 0.7814979 0.7810676 0.7806115
```

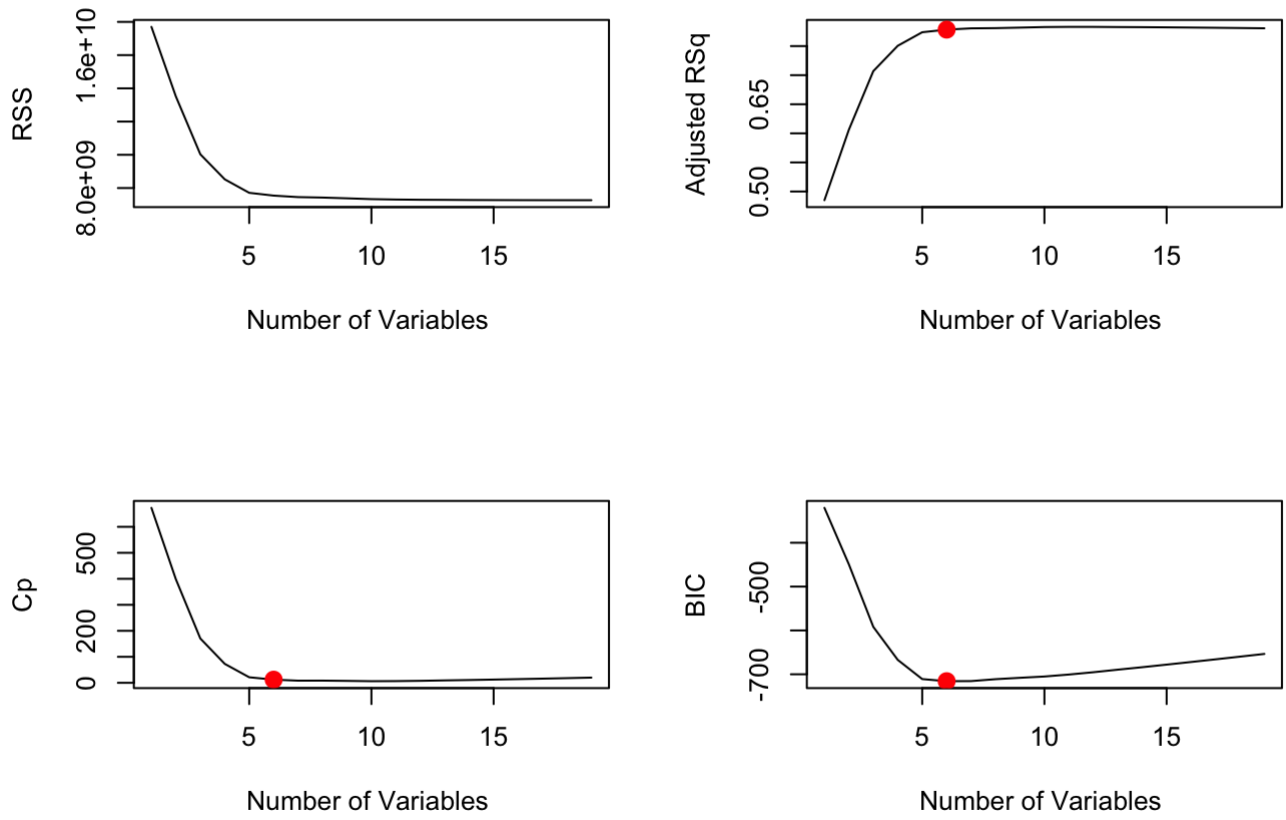
```
c(which.max(summary(m.select)$adjr2),max(summary(m.select)$adjr2))
```

```
## [1] 11.0000000  0.7830404
```

```
library(ggvis)
rsq <- as.data.frame(reg.summary$rsq)
names(rsq) <- "R2"
rsq %>%
        ggvis(x=~ c(1:nrow(rsq)), y=~R2 ) %>%
        layer_points(fill = ~ R2 ) %>%
        add_axis("y", title = "R^2") %>%
        add_axis("x", title = "Number of variables")
```




```
par(mfrow=c(2,2))
plot(reg.summary$rss ,xlab="Number of Variables ",ylab="RSS",type="l")
plot(reg.summary$adjr2 ,xlab="Number of Variables ", ylab="Adjusted RSq",type="l")
# which.max(reg.summary$adjr2)
points(6,reg.summary$adjr2[6], col="red",cex=2,pch=20)
plot(reg.summary$cp ,xlab="Number of Variables ",ylab="Cp", type='l')
# which.min(reg.summary$cp )
points(6,reg.summary$cp [6],col="red",cex=2,pch=20)
plot(reg.summary$bic ,xlab="Number of Variables ",ylab="BIC",type='l')
# which.min(reg.summary$bic )
points(6,reg.summary$bic [6],col="red",cex=2,pch=20)
```




```
#The model fitted from best subset selection
m_sub = lm(N_bikes~feels_like + humidity + wind_speed +feels_like:weekend + humidity:hol
iday + humidity:weekend,
           data = BikeSharing)
summary(m_sub)
```

```
##
## Call:
## lm(formula = N_bikes ~ feels_like + humidity + wind_speed + feels_like:weekend +
##     humidity:holiday + humidity:weekend, data = BikeSharing)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16829.0  -2460.1    191.7   2501.9  19307.3
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         51185.810   1763.072  29.032  < 2e-16 ***
## feels_like            678.121     34.271  19.787  < 2e-16 ***
## humidity             -338.182     20.192 -16.749  < 2e-16 ***
## wind_speed           -354.322     29.012 -12.213  < 2e-16 ***
## feels_like:weekend    186.135     56.126   3.316 0.000979 ***
## humidity:holiday     -112.115     15.087  -7.431 4.79e-13 ***
## humidity:weekend     -100.522      9.931 -10.123  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3912 on 493 degrees of freedom
## Multiple R-squared:  0.781,  Adjusted R-squared:  0.7783
## F-statistic:   293 on 6 and 493 DF,  p-value: < 2.2e-16
```

```
m_max = lm(N_bikes~temperature+feels_like+humidity+wind_speed+holiday+weekend+season+
                      temperature*holiday+temperature*weekend+temperature*season+
                      feels_like*holiday+feels_like*weekend+feels_like*season+
                      wind_speed*holiday+wind_speed*weekend+wind_speed*season+
                      humidity*holiday+humidity*weekend+humidity*season,data=BikeSharin
g)
Anova(m_max,type='II')
```

| | Sum Sq | Df | F value | Pr(>F) |
|---|---|---|---|---|
| | <dbl> | <dbl> | <dbl> | <dbl> |
| temperature | 1.092693e+07 | 1 | 7.215994e-01 | 3.960432e-01 |
| feels_like | 5.065073e+07 | 1 | 3.344906e+00 | 6.803362e-02 |
| humidity | 3.718521e+09 | 1 | 2.455661e+02 | 5.370569e-45 |
| wind_speed | 1.729032e+09 | 1 | 1.141829e+02 | 4.705424e-24 |
| holiday | 8.642165e+08 | 1 | 5.707169e+01 | 2.141800e-13 |
| weekend | 2.685205e+09 | 1 | 1.773273e+02 | 1.184349e-34 |
| season | 7.082474e+05 | 1 | 4.677170e-02 | 8.288710e-01 |
| temperature:holiday | 2.643925e+06 | 1 | 1.746012e-01 | 6.762414e-01 |
| temperature:weekend | 1.167517e+07 | 1 | 7.710124e-01 | 3.803434e-01 |

| | Sum Sq | Df | F value | Pr(>F) |
|---|---|---|---|---|
| | <dbl> | <dbl> | <dbl> | <dbl> |
| temperature:season | 1.307281e+06 | 1 | 8.633107e-02 | 7.690213e-01 |

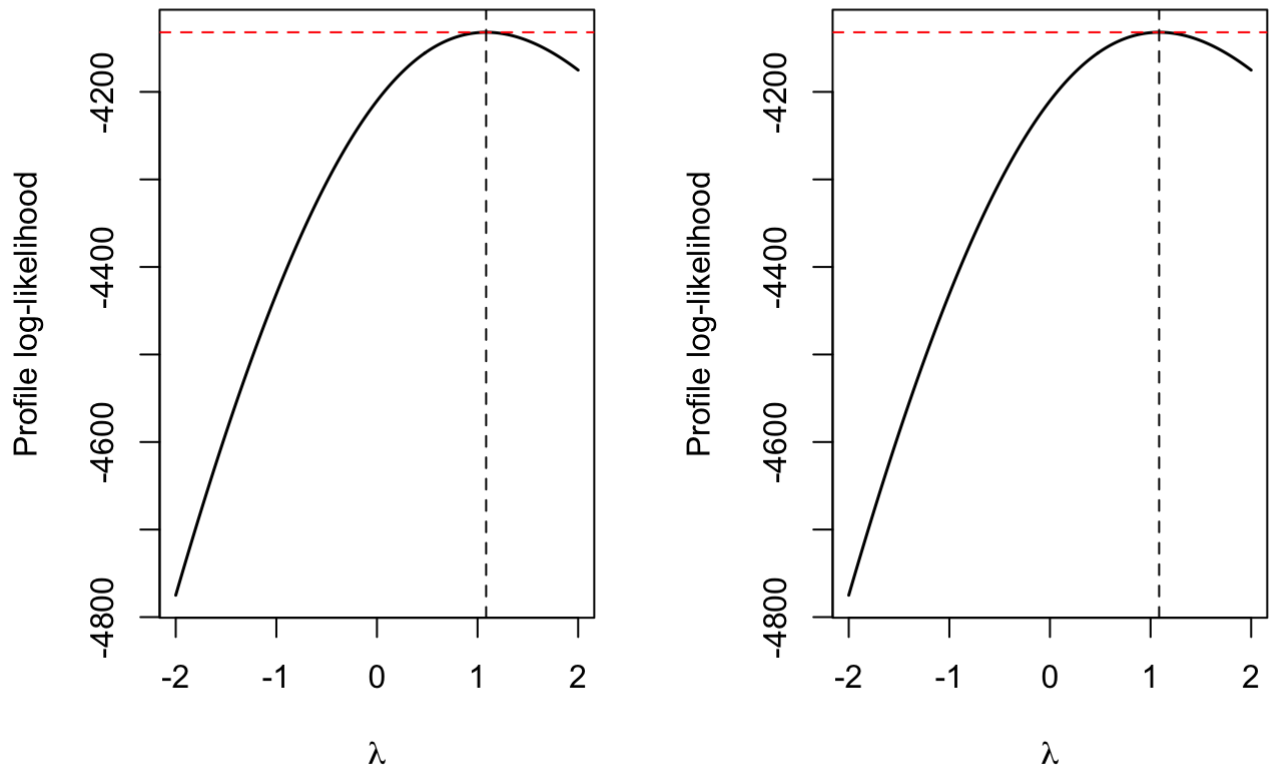
1-10 of 20 rows Previous **1** 2 Next

# Box-Cox and Yeo-Jhonson

```
par(mfrow=c(1,2))
boxcox(m_sub,plotit=TRUE)
```

```
## Box-Cox Transformation
##
## Estimation method:  ml
## Optimal parameter:  1.085186
## Loglike:  4131.914
##
## Summary of transformed variables
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12486   47730   58991   59748   73933  105844
```

```
best.lambda_bc = boxcox(m_sub)$lambdahat
```



```
best.lambda_bc
```

```
## [1] 1.085186
```

```
m_boxcox = lm(((N_bikes^best.lambda_bc-1)/best.lambda_bc)~feels_like + humidity + wind_s
peed +feels_like:weekend + humidity:holiday + humidity:weekend,data = BikeSharing)
summary(m_boxcox)
```

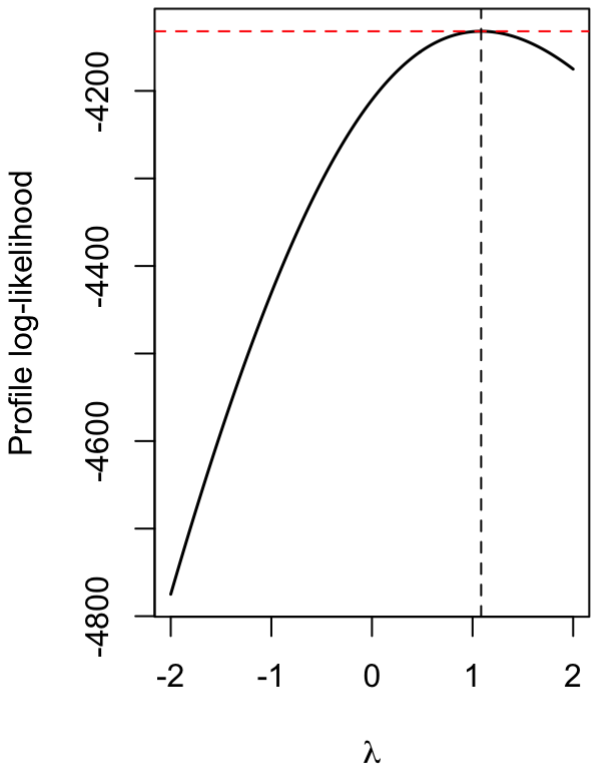
```
##
## Call:
## lm(formula = ((N_bikes^best.lambda_bc - 1)/best.lambda_bc) ~
##     feels_like + humidity + wind_speed + feels_like:weekend +
##         humidity:holiday + humidity:weekend, data = BikeSharing)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -38738  -5801    411   5839  45856
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        117124.56    4182.15  28.006  < 2e-16 ***
## feels_like           1631.63      81.29  20.071  < 2e-16 ***
## humidity             -808.68      47.90 -16.884  < 2e-16 ***
## wind_speed           -840.19      68.82 -12.209  < 2e-16 ***
## feels_like:weekend    401.90     133.14   3.019  0.00267 **
## humidity:holiday     -259.41      35.79  -7.248 1.64e-12 ***
## humidity:weekend     -229.83      23.56  -9.757  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9279 on 493 degrees of freedom
## Multiple R-squared:  0.7814, Adjusted R-squared:  0.7788
## F-statistic: 293.7 on 6 and 493 DF,  p-value: < 2.2e-16
```

```
best.lambda_yeo = yeojohnson(m_sub)$lambdahat
best.lambda_yeo
```

```
## [1] 1.085192
```

```
m_yeo =lm(yeo.johnson(N_bikes,best.lambda_yeo)~feels_like + humidity + wind_speed +feels
_like:weekend + humidity:holiday + humidity:weekend,data=BikeSharing)
summary(m_yeo)
```

```
##
## Call:
## lm(formula = yeo.johnson(N_bikes, best.lambda_yeo) ~ feels_like +
##     humidity + wind_speed + feels_like:weekend + humidity:holiday +
##     humidity:weekend, data = BikeSharing)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -38741  -5801    411   5839  45859
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)        117134.75    4182.45  28.006  < 2e-16 ***
## feels_like           1631.75      81.30  20.071  < 2e-16 ***
## humidity             -808.74      47.90 -16.884  < 2e-16 ***
## wind_speed           -840.25      68.82 -12.209  < 2e-16 ***
## feels_like:weekend    401.93     133.15   3.019  0.00267 **
## humidity:holiday     -259.43      35.79  -7.248 1.64e-12 ***
## humidity:weekend     -229.84      23.56  -9.757  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9280 on 493 degrees of freedom
## Multiple R-squared:  0.7814, Adjusted R-squared:  0.7788
## F-statistic: 293.7 on 6 and 493 DF,  p-value: < 2.2e-16
```



# Principle Component Regression

```
# #Without standardizing the variables
# m.pcr1 = pcr(N_bikes~.,data= BikeSharing,ncomp=7,validation='CV')
# prc1 = princomp(BikeSharing[,colnames(BikeSharing)!='N_bikes'])
# summary(m.pcr1)
# screeplot(prc1)
# m.pcr1$loadings

## With standardizing the variables
# BikeSharing[,colnames(BikeSharing)!='N_bikes'] = data.frame(scale(BikeSharing[,colname
s(BikeSharing)!='N_bikes']))
#
# m.pcr2 = pcr(N_bikes~.,data= BikeSharing,ncomp=7,validation='CV')
# prc2 = princomp(BikeSharing[,colnames(BikeSharing)!='N_bikes'])
# summary(m.pcr2)
# screeplot(prc2)
# m.pcr2$loadings
#
#
# #Pick the non-standardized variable model, with ncomp = 3
# BikeSharing <- read.delim("bikesharing18.txt",sep = "")
# m.pcr = pcr(N_bikes~.,data= BikeSharing,ncomp=3,validation='CV')
# MSEP(m.pcr)$val
# summary(m.pcr)


library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:ggvis':
##
##     resolution
```

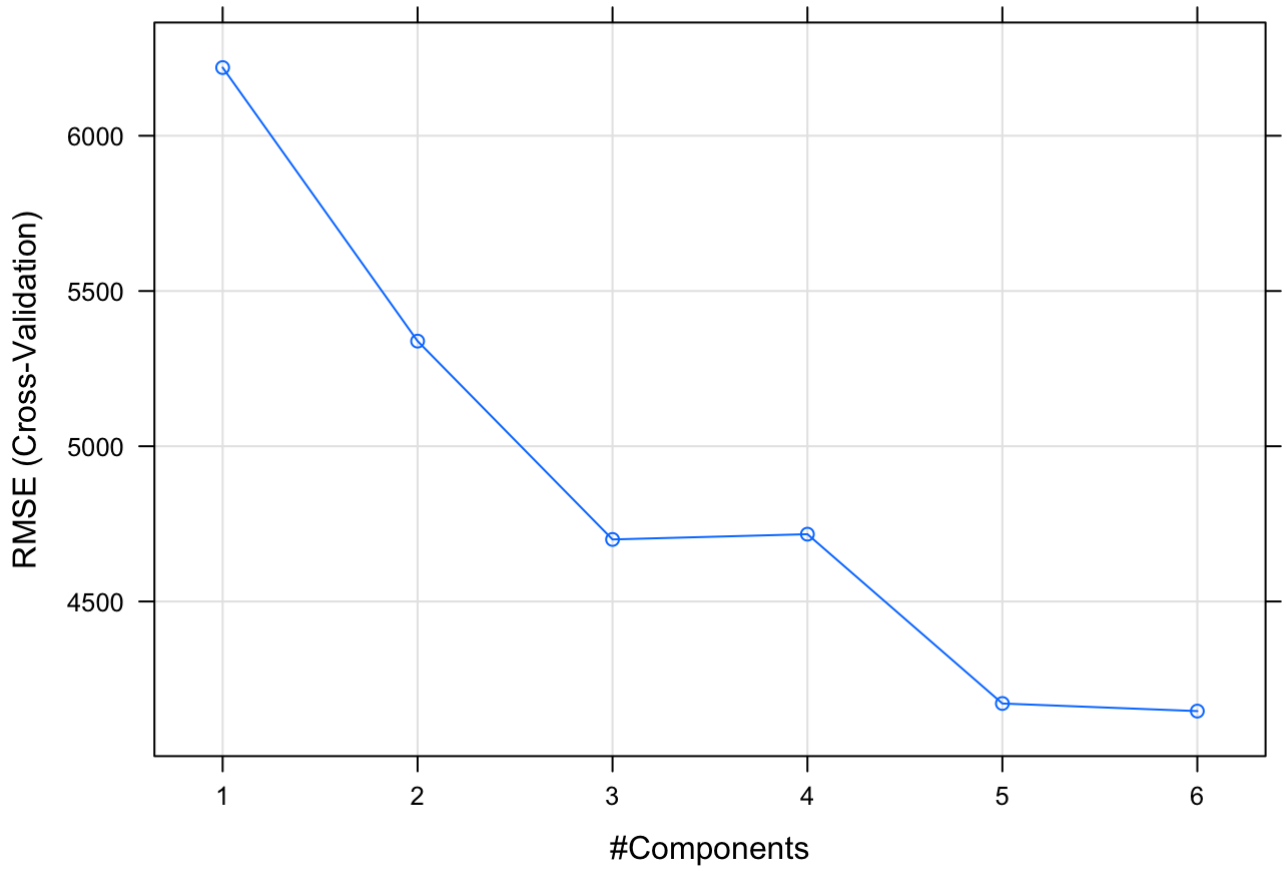
```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:VGAM':
##
##     predictors
```

```
## The following object is masked from 'package:pls':
##
##     R2
```

```
set.seed(123)
training.samples <- BikeSharing$N_bikes %>%
  createDataPartition(p = 0.66, list = FALSE)
train.data  <- BikeSharing[training.samples, ]
test.data <- BikeSharing[-training.samples, ]

set.seed(123)
model <- train(
  N_bikes~., data = train.data, method = "pcr",
  scale = FALSE,
  trControl = trainControl("cv", number = 7),
  tuneLength = 7
  )
# Plot model RMSE vs different values of components
plot(model)
```




```
# Print the best tuning parameter ncomp that
# minimize the cross-validation error, RMSE
model$bestTune
```

| | ncomp |
|---|---|
| | <dbl> |
| 6 | 6 |

1 row

```
summary(model$finalModel)
```

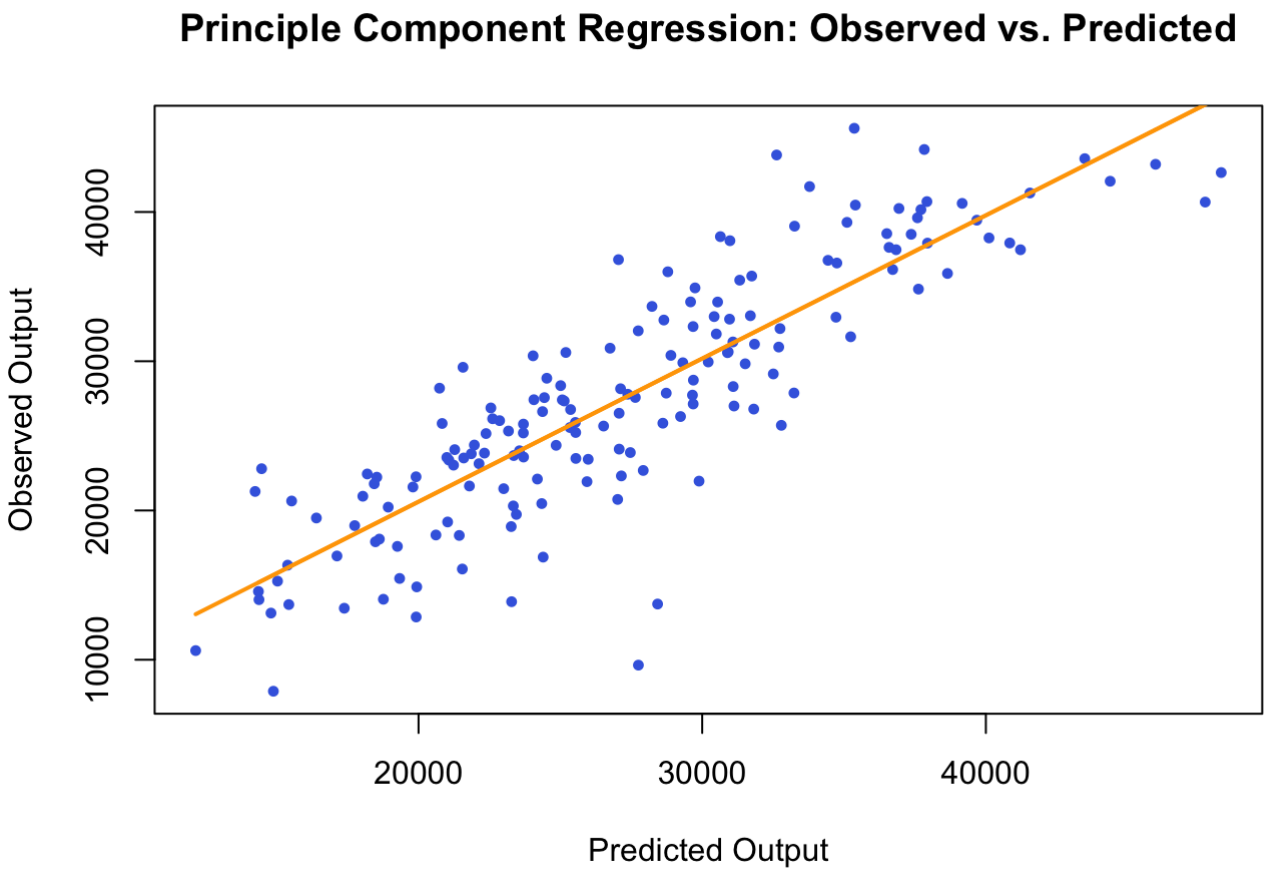
```
## Data:    X dimension: 332 7
##  Y dimension: 332 1
## Fit method: svdpc
## Number of components considered: 6
## TRAINING: % variance explained
##           1 comps  2 comps  3 comps  4 comps  5 comps  6 comps
## X           58.91    82.26    99.32    99.80    99.91    99.99
## .outcome    43.40    58.77    68.06    68.07    74.96    75.19
```

```
# Make predictions
predictions <- model %>% predict(test.data)
# Model performance metrics
data.frame(
  RMSE = caret::RMSE(predictions, test.data$N_bikes),
  Rsquare = caret::R2(predictions, test.data$N_bikes)
)
```

| RMSE | Rsquare |
|---|---|
| <dbl> | <dbl> |
| 4248.386 | 0.7394426 |

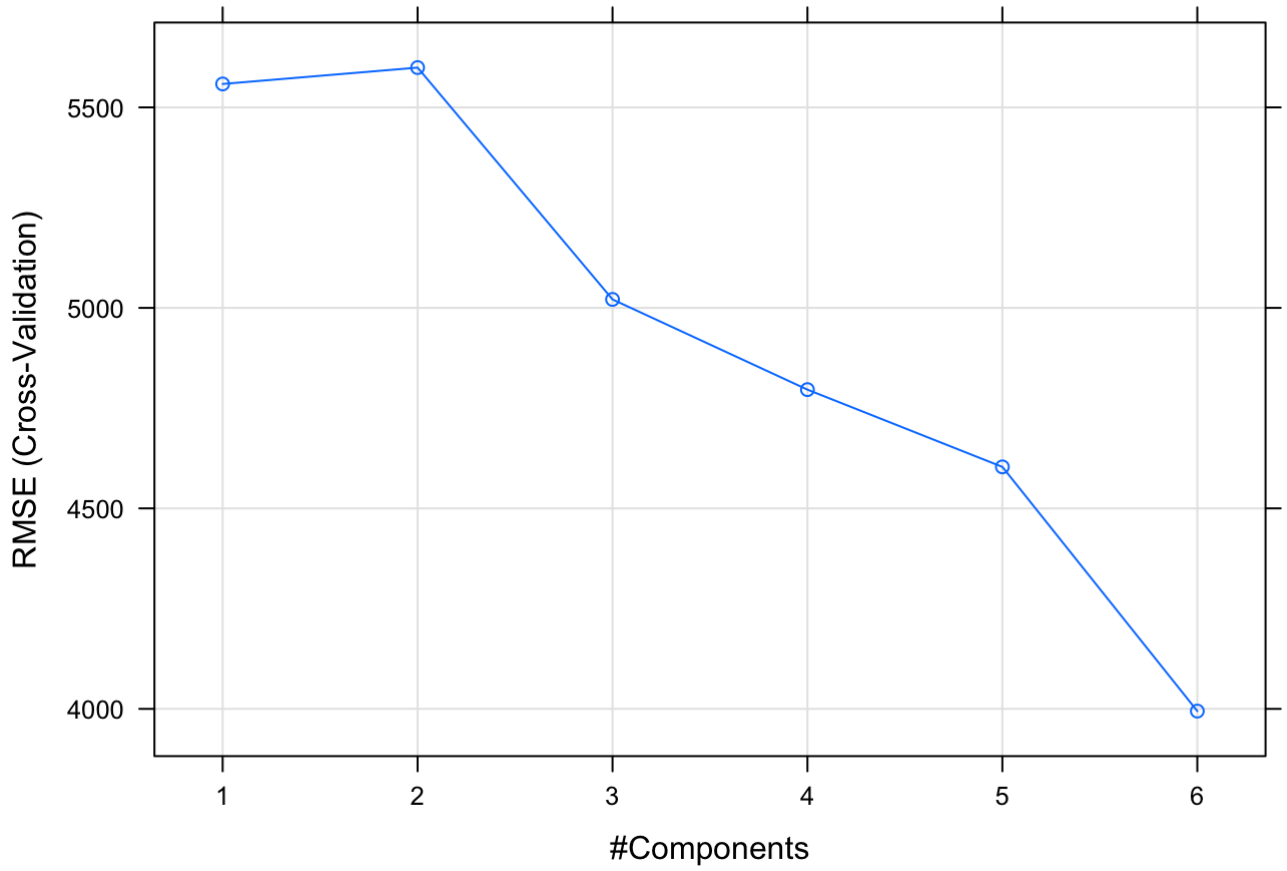
1 row

```
plot(predictions, test.data$N_bikes, pch=16, col="royalblue", cex=0.75,
xlab="Predicted Output",
ylab="Observed Output",
main="Principle Component Regression: Observed vs. Predicted")
lines(predictions, lm(a~b, data=data.frame(a=test.data$N_bikes, b=predictions))$fitted,
 lwd=2, col="orange")
```




```
#Normalizing data
BikeSharing[,colnames(BikeSharing)!='N_bikes'] = data.frame(scale(BikeSharing[,colnames
(BikeSharing)!='N_bikes']))
set.seed(123)
training.samples <- BikeSharing$N_bikes %>%
  createDataPartition(p = 0.66, list = FALSE)
train.data  <- BikeSharing[training.samples, ]
test.data <- BikeSharing[-training.samples, ]

set.seed(123)
model <- train(
  N_bikes~., data = train.data, method = "pcr",
  scale = FALSE,
  trControl = trainControl("cv", number = 7),
  tuneLength = 7
  )
# Plot model RMSE vs different values of components
plot(model)
```




```
# Print the best tuning parameter ncomp that
# minimize the cross-validation error, RMSE
model$bestTune
```

| | ncomp |
|---|---|
| | <dbl> |
| 6 | 6 |

1 row

```
summary(model$finalModel)
```

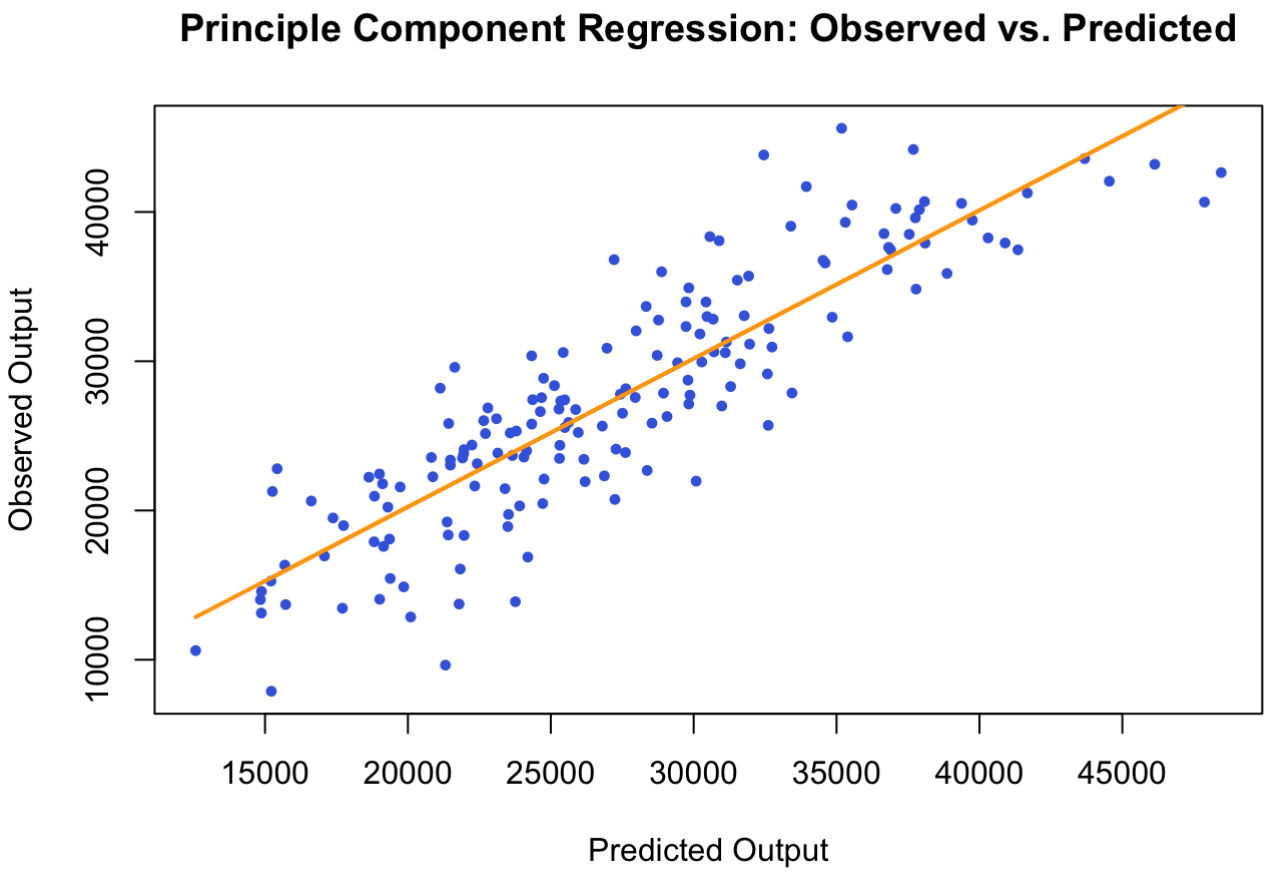
```
## Data:    X dimension: 332 7
##  Y dimension: 332 1
## Fit method: svdpc
## Number of components considered: 6
## TRAINING: % variance explained
##           1 comps  2 comps  3 comps  4 comps  5 comps  6 comps
## X           34.51    51.75    66.67    80.07    92.80    99.93
## .outcome    54.88    55.04    64.41    67.46    69.92    76.83
```

```
# Make predictions
predictions <- model %>% predict(test.data)
# Model performance metrics
data.frame(
  RMSE = caret::RMSE(predictions, test.data$N_bikes),
  Rsquare = caret::R2(predictions, test.data$N_bikes)
)
```

| RMSE <dbl> | Rsquare <dbl> |
|---|---|
| 3947.463 | 0.7733957 |

1 row

```
plot(predictions, test.data$N_bikes, pch=16, col="royalblue", cex=0.75,
xlab="Predicted Output",
ylab="Observed Output",
main="Principle Component Regression: Observed vs. Predicted")
lines(predictions, lm(a~b, data=data.frame(a=test.data$N_bikes, b=predictions))$fitted,
 lwd=2, col="orange")
```




```
library(xgboost)
library(Metrics)
```

```
##
## Attaching package: 'Metrics'
```

```
## The following objects are masked from 'package:caret':
##
##     precision, recall
```

```
training.samples <- BikeSharing$N_bikes %>%
  createDataPartition(p = 0.66, list = FALSE)
train.data  <- BikeSharing[training.samples, ]
test.data <- BikeSharing[-training.samples, ]
y_train = train.data[,1]
X_train = train.data[,-1]
X_train <- as.matrix(X_train)

y_test = test.data[,1]
X_test = test.data[,-1]
X_test <- as.matrix(X_test)

fit_xgb <- xgboost(X_train, y_train
                   , max_depth = 10
                   , eta = 0.02
                   , nthread = 4
                   , nrounds = 800
                   , subsample = .7
                   , colsample_bytree = .7
                   , booster = "gbtree"
                   , eval_metric = "rmse"
                   , objective="reg:linear")
```

```
## [1]  train-rmse:27743.531250
## [2]  train-rmse:27221.800781
## [3]  train-rmse:26712.132812
## [4]  train-rmse:26208.347656
## [5]  train-rmse:25720.515625
## [6]  train-rmse:25240.978516
## [7]  train-rmse:24762.300781
## [8]  train-rmse:24305.730469
## [9]  train-rmse:23853.441406
## [10] train-rmse:23408.259766
## [11] train-rmse:22971.455078
## [12] train-rmse:22538.347656
## [13] train-rmse:22119.998047
## [14] train-rmse:21704.425781
## [15] train-rmse:21292.203125
## [16] train-rmse:20901.816406
## [17] train-rmse:20518.195312
## [18] train-rmse:20134.470703
## [19] train-rmse:19762.724609
## [20] train-rmse:19399.187500
## [21] train-rmse:19047.445312
## [22] train-rmse:18700.113281
## [23] train-rmse:18371.136719
## [24] train-rmse:18032.853516
## [25] train-rmse:17703.134766
## [26] train-rmse:17384.726562
## [27] train-rmse:17066.320312
## [28] train-rmse:16754.599609
## [29] train-rmse:16447.916016
## [30] train-rmse:16150.884766
## [31] train-rmse:15861.286133
## [32] train-rmse:15570.099609
## [33] train-rmse:15294.421875
## [34] train-rmse:15028.643555
## [35] train-rmse:14757.707031
## [36] train-rmse:14486.216797
## [37] train-rmse:14222.744141
## [38] train-rmse:13968.128906
## [39] train-rmse:13726.092773
## [40] train-rmse:13484.792969
## [41] train-rmse:13244.374023
## [42] train-rmse:13008.515625
## [43] train-rmse:12783.118164
## [44] train-rmse:12559.764648
## [45] train-rmse:12336.700195
## [46] train-rmse:12115.701172
## [47] train-rmse:11901.664062
## [48] train-rmse:11695.286133
## [49] train-rmse:11497.737305
## [50] train-rmse:11304.179688
## [51] train-rmse:11108.521484
## [52] train-rmse:10922.160156
## [53] train-rmse:10731.785156
```

```
## [54] train-rmse:10539.979492
## [55] train-rmse:10361.614258
## [56] train-rmse:10184.431641
## [57] train-rmse:10009.558594
## [58] train-rmse:9836.306641
## [59] train-rmse:9669.949219
## [60] train-rmse:9503.525391
## [61] train-rmse:9345.850586
## [62] train-rmse:9185.360352
## [63] train-rmse:9037.725586
## [64] train-rmse:8886.983398
## [65] train-rmse:8734.384766
## [66] train-rmse:8589.279297
## [67] train-rmse:8453.056641
## [68] train-rmse:8313.802734
## [69] train-rmse:8185.420410
## [70] train-rmse:8059.378418
## [71] train-rmse:7923.378418
## [72] train-rmse:7796.852539
## [73] train-rmse:7671.351562
## [74] train-rmse:7553.328613
## [75] train-rmse:7436.678711
## [76] train-rmse:7320.436035
## [77] train-rmse:7200.841797
## [78] train-rmse:7089.483887
## [79] train-rmse:6976.525879
## [80] train-rmse:6873.501953
## [81] train-rmse:6767.017090
## [82] train-rmse:6666.192871
## [83] train-rmse:6567.474121
## [84] train-rmse:6462.243652
## [85] train-rmse:6358.239258
## [86] train-rmse:6259.083496
## [87] train-rmse:6161.566895
## [88] train-rmse:6064.799805
## [89] train-rmse:5974.864258
## [90] train-rmse:5890.081543
## [91] train-rmse:5805.908691
## [92] train-rmse:5719.312500
## [93] train-rmse:5635.772461
## [94] train-rmse:5552.648438
## [95] train-rmse:5469.339844
## [96] train-rmse:5400.510254
## [97] train-rmse:5325.775879
## [98] train-rmse:5250.012207
## [99] train-rmse:5180.275879
## [100]    train-rmse:5116.121582
## [101]    train-rmse:5045.099121
## [102]    train-rmse:4975.862305
## [103]    train-rmse:4906.013184
## [104]    train-rmse:4835.144043
## [105]    train-rmse:4766.580078
## [106]    train-rmse:4698.137695
## [107]    train-rmse:4627.566895
```

```
## [108]    train-rmse:4564.226074
## [109]    train-rmse:4504.202637
## [110]    train-rmse:4450.030273
## [111]    train-rmse:4389.895020
## [112]    train-rmse:4332.625488
## [113]    train-rmse:4276.886719
## [114]    train-rmse:4218.433105
## [115]    train-rmse:4161.882324
## [116]    train-rmse:4111.832031
## [117]    train-rmse:4061.996826
## [118]    train-rmse:4010.291504
## [119]    train-rmse:3953.902832
## [120]    train-rmse:3909.545166
## [121]    train-rmse:3853.952637
## [122]    train-rmse:3807.298828
## [123]    train-rmse:3754.910156
## [124]    train-rmse:3703.069580
## [125]    train-rmse:3658.929443
## [126]    train-rmse:3610.255371
## [127]    train-rmse:3565.666260
## [128]    train-rmse:3517.803467
## [129]    train-rmse:3471.909668
## [130]    train-rmse:3429.885498
## [131]    train-rmse:3393.559814
## [132]    train-rmse:3353.207520
## [133]    train-rmse:3316.645020
## [134]    train-rmse:3276.900635
## [135]    train-rmse:3232.897217
## [136]    train-rmse:3199.600342
## [137]    train-rmse:3164.021973
## [138]    train-rmse:3124.155029
## [139]    train-rmse:3089.548340
## [140]    train-rmse:3051.000732
## [141]    train-rmse:3014.766113
## [142]    train-rmse:2984.127930
## [143]    train-rmse:2952.810303
## [144]    train-rmse:2917.207764
## [145]    train-rmse:2886.307861
## [146]    train-rmse:2856.758057
## [147]    train-rmse:2829.162598
## [148]    train-rmse:2800.322754
## [149]    train-rmse:2770.081055
## [150]    train-rmse:2740.758301
## [151]    train-rmse:2710.939453
## [152]    train-rmse:2685.288818
## [153]    train-rmse:2654.124023
## [154]    train-rmse:2627.987061
## [155]    train-rmse:2603.485107
## [156]    train-rmse:2574.392822
## [157]    train-rmse:2548.642334
## [158]    train-rmse:2524.929443
## [159]    train-rmse:2500.286865
## [160]    train-rmse:2472.302979
## [161]    train-rmse:2449.703369
```

```
## [162]    train-rmse:2421.910156
## [163]    train-rmse:2400.241455
## [164]    train-rmse:2374.399170
## [165]    train-rmse:2351.449219
## [166]    train-rmse:2327.297607
## [167]    train-rmse:2301.128418
## [168]    train-rmse:2279.024414
## [169]    train-rmse:2259.474121
## [170]    train-rmse:2242.081543
## [171]    train-rmse:2222.509277
## [172]    train-rmse:2202.005127
## [173]    train-rmse:2178.126953
## [174]    train-rmse:2159.520996
## [175]    train-rmse:2139.286621
## [176]    train-rmse:2116.710693
## [177]    train-rmse:2094.339355
## [178]    train-rmse:2077.511963
## [179]    train-rmse:2059.480713
## [180]    train-rmse:2041.257812
## [181]    train-rmse:2025.269897
## [182]    train-rmse:2008.536987
## [183]    train-rmse:1990.047363
## [184]    train-rmse:1970.418457
## [185]    train-rmse:1952.654175
## [186]    train-rmse:1938.532104
## [187]    train-rmse:1920.705078
## [188]    train-rmse:1902.149780
## [189]    train-rmse:1888.717529
## [190]    train-rmse:1873.456055
## [191]    train-rmse:1854.398193
## [192]    train-rmse:1841.091309
## [193]    train-rmse:1824.729126
## [194]    train-rmse:1810.932739
## [195]    train-rmse:1791.933960
## [196]    train-rmse:1777.081665
## [197]    train-rmse:1761.270386
## [198]    train-rmse:1750.082520
## [199]    train-rmse:1736.101562
## [200]    train-rmse:1723.855835
## [201]    train-rmse:1708.864868
## [202]    train-rmse:1695.851074
## [203]    train-rmse:1679.488037
## [204]    train-rmse:1663.069092
## [205]    train-rmse:1652.534058
## [206]    train-rmse:1639.915894
## [207]    train-rmse:1625.977783
## [208]    train-rmse:1611.292847
## [209]    train-rmse:1599.584961
## [210]    train-rmse:1587.684692
## [211]    train-rmse:1577.029541
## [212]    train-rmse:1563.318481
## [213]    train-rmse:1549.054443
## [214]    train-rmse:1538.256104
## [215]    train-rmse:1525.852051
```

```
## [216]    train-rmse:1512.236206
## [217]    train-rmse:1499.499390
## [218]    train-rmse:1485.424927
## [219]    train-rmse:1472.894287
## [220]    train-rmse:1459.263184
## [221]    train-rmse:1446.096191
## [222]    train-rmse:1436.820557
## [223]    train-rmse:1424.882812
## [224]    train-rmse:1413.234497
## [225]    train-rmse:1402.587891
## [226]    train-rmse:1392.371460
## [227]    train-rmse:1383.109131
## [228]    train-rmse:1376.226685
## [229]    train-rmse:1364.655029
## [230]    train-rmse:1353.670776
## [231]    train-rmse:1342.640625
## [232]    train-rmse:1335.910278
## [233]    train-rmse:1326.055664
## [234]    train-rmse:1315.624878
## [235]    train-rmse:1305.260010
## [236]    train-rmse:1297.424683
## [237]    train-rmse:1286.511108
## [238]    train-rmse:1276.915527
## [239]    train-rmse:1267.100830
## [240]    train-rmse:1258.540283
## [241]    train-rmse:1248.128052
## [242]    train-rmse:1239.351562
## [243]    train-rmse:1230.786743
## [244]    train-rmse:1223.972534
## [245]    train-rmse:1217.058228
## [246]    train-rmse:1207.585327
## [247]    train-rmse:1198.423828
## [248]    train-rmse:1189.002686
## [249]    train-rmse:1179.613647
## [250]    train-rmse:1170.701294
## [251]    train-rmse:1162.062500
## [252]    train-rmse:1152.641113
## [253]    train-rmse:1142.710815
## [254]    train-rmse:1132.348511
## [255]    train-rmse:1126.051025
## [256]    train-rmse:1117.499023
## [257]    train-rmse:1111.070923
## [258]    train-rmse:1102.713745
## [259]    train-rmse:1096.026978
## [260]    train-rmse:1086.269287
## [261]    train-rmse:1077.637573
## [262]    train-rmse:1068.922363
## [263]    train-rmse:1064.730957
## [264]    train-rmse:1055.606323
## [265]    train-rmse:1047.673950
## [266]    train-rmse:1042.374634
## [267]    train-rmse:1036.812988
## [268]    train-rmse:1030.233154
## [269]    train-rmse:1023.082153
```

```
## [270]    train-rmse:1014.900818
## [271]    train-rmse:1007.253052
## [272]    train-rmse:1002.228271
## [273]    train-rmse:996.697571
## [274]    train-rmse:990.601501
## [275]    train-rmse:985.431580
## [276]    train-rmse:981.181641
## [277]    train-rmse:974.408691
## [278]    train-rmse:968.434204
## [279]    train-rmse:965.172485
## [280]    train-rmse:960.383789
## [281]    train-rmse:953.143127
## [282]    train-rmse:946.859741
## [283]    train-rmse:939.425415
## [284]    train-rmse:933.087646
## [285]    train-rmse:926.878540
## [286]    train-rmse:921.696716
## [287]    train-rmse:915.811401
## [288]    train-rmse:911.903809
## [289]    train-rmse:907.006226
## [290]    train-rmse:900.148315
## [291]    train-rmse:894.573669
## [292]    train-rmse:888.859680
## [293]    train-rmse:883.806213
## [294]    train-rmse:878.357483
## [295]    train-rmse:872.805725
## [296]    train-rmse:867.807678
## [297]    train-rmse:863.119751
## [298]    train-rmse:859.031128
## [299]    train-rmse:854.806152
## [300]    train-rmse:850.860168
## [301]    train-rmse:844.665344
## [302]    train-rmse:839.771240
## [303]    train-rmse:834.503296
## [304]    train-rmse:828.773132
## [305]    train-rmse:824.137634
## [306]    train-rmse:819.328735
## [307]    train-rmse:813.786255
## [308]    train-rmse:808.183899
## [309]    train-rmse:805.742493
## [310]    train-rmse:801.410950
## [311]    train-rmse:796.990051
## [312]    train-rmse:791.335754
## [313]    train-rmse:785.102661
## [314]    train-rmse:781.866272
## [315]    train-rmse:777.517334
## [316]    train-rmse:771.622131
## [317]    train-rmse:767.515991
## [318]    train-rmse:763.874451
## [319]    train-rmse:758.611389
## [320]    train-rmse:755.924255
## [321]    train-rmse:751.644958
## [322]    train-rmse:748.244202
## [323]    train-rmse:743.013611
```

```
## [324]    train-rmse:738.568787
## [325]    train-rmse:734.080688
## [326]    train-rmse:731.502319
## [327]    train-rmse:727.385437
## [328]    train-rmse:723.855652
## [329]    train-rmse:719.661560
## [330]    train-rmse:716.766846
## [331]    train-rmse:710.849426
## [332]    train-rmse:706.254211
## [333]    train-rmse:702.621765
## [334]    train-rmse:698.704041
## [335]    train-rmse:694.002991
## [336]    train-rmse:691.618225
## [337]    train-rmse:687.534302
## [338]    train-rmse:684.415222
## [339]    train-rmse:680.462708
## [340]    train-rmse:678.083862
## [341]    train-rmse:675.511230
## [342]    train-rmse:671.565247
## [343]    train-rmse:668.237915
## [344]    train-rmse:664.607788
## [345]    train-rmse:661.414246
## [346]    train-rmse:657.721375
## [347]    train-rmse:655.312256
## [348]    train-rmse:652.114136
## [349]    train-rmse:649.788452
## [350]    train-rmse:648.471680
## [351]    train-rmse:644.893311
## [352]    train-rmse:641.398621
## [353]    train-rmse:638.242065
## [354]    train-rmse:633.905823
## [355]    train-rmse:629.340149
## [356]    train-rmse:625.422974
## [357]    train-rmse:621.204346
## [358]    train-rmse:618.642700
## [359]    train-rmse:615.588623
## [360]    train-rmse:612.944397
## [361]    train-rmse:609.276184
## [362]    train-rmse:605.911926
## [363]    train-rmse:602.339050
## [364]    train-rmse:599.114441
## [365]    train-rmse:595.777161
## [366]    train-rmse:593.307800
## [367]    train-rmse:591.145264
## [368]    train-rmse:587.473572
## [369]    train-rmse:585.214966
## [370]    train-rmse:581.388062
## [371]    train-rmse:578.013489
## [372]    train-rmse:575.280151
## [373]    train-rmse:572.318665
## [374]    train-rmse:569.902832
## [375]    train-rmse:566.112854
## [376]    train-rmse:563.149475
## [377]    train-rmse:559.802490
```

```
## [378]    train-rmse:555.963379
## [379]    train-rmse:552.476868
## [380]    train-rmse:549.815125
## [381]    train-rmse:547.662537
## [382]    train-rmse:544.405151
## [383]    train-rmse:540.998535
## [384]    train-rmse:538.729370
## [385]    train-rmse:535.969910
## [386]    train-rmse:532.629944
## [387]    train-rmse:529.273804
## [388]    train-rmse:526.678406
## [389]    train-rmse:523.575134
## [390]    train-rmse:520.773376
## [391]    train-rmse:517.155090
## [392]    train-rmse:514.952148
## [393]    train-rmse:512.185669
## [394]    train-rmse:509.212372
## [395]    train-rmse:506.594818
## [396]    train-rmse:504.169952
## [397]    train-rmse:501.785522
## [398]    train-rmse:499.511932
## [399]    train-rmse:497.579651
## [400]    train-rmse:495.031128
## [401]    train-rmse:492.345184
## [402]    train-rmse:489.247009
## [403]    train-rmse:485.940521
## [404]    train-rmse:484.392395
## [405]    train-rmse:481.971222
## [406]    train-rmse:479.873962
## [407]    train-rmse:477.572052
## [408]    train-rmse:475.127655
## [409]    train-rmse:472.462433
## [410]    train-rmse:469.907623
## [411]    train-rmse:466.891602
## [412]    train-rmse:463.779175
## [413]    train-rmse:461.822632
## [414]    train-rmse:458.889923
## [415]    train-rmse:456.743195
## [416]    train-rmse:454.775604
## [417]    train-rmse:452.818237
## [418]    train-rmse:450.774048
## [419]    train-rmse:449.121460
## [420]    train-rmse:446.722900
## [421]    train-rmse:444.802979
## [422]    train-rmse:442.521271
## [423]    train-rmse:440.116425
## [424]    train-rmse:438.484528
## [425]    train-rmse:435.657135
## [426]    train-rmse:433.645203
## [427]    train-rmse:431.584198
## [428]    train-rmse:430.051575
## [429]    train-rmse:427.709900
## [430]    train-rmse:425.758606
## [431]    train-rmse:423.292664
```

```
## [432]    train-rmse:420.742554
## [433]    train-rmse:418.792114
## [434]    train-rmse:416.485474
## [435]    train-rmse:414.640106
## [436]    train-rmse:412.599213
## [437]    train-rmse:411.040558
## [438]    train-rmse:409.344055
## [439]    train-rmse:406.820740
## [440]    train-rmse:405.159332
## [441]    train-rmse:402.380890
## [442]    train-rmse:399.696106
## [443]    train-rmse:398.152069
## [444]    train-rmse:395.682648
## [445]    train-rmse:393.607971
## [446]    train-rmse:391.936890
## [447]    train-rmse:390.195038
## [448]    train-rmse:388.014984
## [449]    train-rmse:385.390198
## [450]    train-rmse:383.389954
## [451]    train-rmse:380.977661
## [452]    train-rmse:379.116669
## [453]    train-rmse:377.458160
## [454]    train-rmse:375.176636
## [455]    train-rmse:373.257996
## [456]    train-rmse:371.568878
## [457]    train-rmse:369.385315
## [458]    train-rmse:367.434814
## [459]    train-rmse:365.245392
## [460]    train-rmse:363.297455
## [461]    train-rmse:361.900665
## [462]    train-rmse:360.487183
## [463]    train-rmse:358.737488
## [464]    train-rmse:356.706299
## [465]    train-rmse:354.824585
## [466]    train-rmse:353.242279
## [467]    train-rmse:351.461914
## [468]    train-rmse:349.232666
## [469]    train-rmse:347.747162
## [470]    train-rmse:346.313507
## [471]    train-rmse:344.578705
## [472]    train-rmse:342.629425
## [473]    train-rmse:341.263123
## [474]    train-rmse:339.652740
## [475]    train-rmse:337.507019
## [476]    train-rmse:335.373413
## [477]    train-rmse:334.079987
## [478]    train-rmse:332.687408
## [479]    train-rmse:330.280701
## [480]    train-rmse:328.609009
## [481]    train-rmse:326.404938
## [482]    train-rmse:325.296967
## [483]    train-rmse:322.848053
## [484]    train-rmse:321.482635
## [485]    train-rmse:319.663605
```

```
## [486]    train-rmse:317.932281
## [487]    train-rmse:317.455475
## [488]    train-rmse:316.222076
## [489]    train-rmse:314.276855
## [490]    train-rmse:313.357849
## [491]    train-rmse:311.820618
## [492]    train-rmse:310.290894
## [493]    train-rmse:308.242584
## [494]    train-rmse:306.281860
## [495]    train-rmse:304.650635
## [496]    train-rmse:302.929016
## [497]    train-rmse:301.541534
## [498]    train-rmse:300.763092
## [499]    train-rmse:299.586639
## [500]    train-rmse:298.899902
## [501]    train-rmse:296.458008
## [502]    train-rmse:295.089630
## [503]    train-rmse:293.241638
## [504]    train-rmse:292.530273
## [505]    train-rmse:290.880493
## [506]    train-rmse:289.976349
## [507]    train-rmse:288.907440
## [508]    train-rmse:287.923920
## [509]    train-rmse:286.056976
## [510]    train-rmse:284.297119
## [511]    train-rmse:282.717072
## [512]    train-rmse:281.360504
## [513]    train-rmse:279.912048
## [514]    train-rmse:278.953766
## [515]    train-rmse:277.705902
## [516]    train-rmse:276.166290
## [517]    train-rmse:275.491455
## [518]    train-rmse:274.096436
## [519]    train-rmse:272.746887
## [520]    train-rmse:271.495880
## [521]    train-rmse:269.815399
## [522]    train-rmse:268.431885
## [523]    train-rmse:267.034393
## [524]    train-rmse:265.853973
## [525]    train-rmse:264.402588
## [526]    train-rmse:262.742401
## [527]    train-rmse:261.451752
## [528]    train-rmse:260.061340
## [529]    train-rmse:259.197723
## [530]    train-rmse:257.854553
## [531]    train-rmse:257.315399
## [532]    train-rmse:256.700775
## [533]    train-rmse:255.514450
## [534]    train-rmse:254.171158
## [535]    train-rmse:252.885391
## [536]    train-rmse:251.567886
## [537]    train-rmse:250.611359
## [538]    train-rmse:249.070328
## [539]    train-rmse:248.075974
```

```
## [540]    train-rmse:246.526535
## [541]    train-rmse:245.348816
## [542]    train-rmse:243.980255
## [543]    train-rmse:242.314972
## [544]    train-rmse:241.254623
## [545]    train-rmse:240.178391
## [546]    train-rmse:238.940048
## [547]    train-rmse:237.856293
## [548]    train-rmse:237.181717
## [549]    train-rmse:236.001083
## [550]    train-rmse:235.288101
## [551]    train-rmse:233.660355
## [552]    train-rmse:232.841339
## [553]    train-rmse:231.515884
## [554]    train-rmse:230.221512
## [555]    train-rmse:228.763916
## [556]    train-rmse:227.423050
## [557]    train-rmse:226.250839
## [558]    train-rmse:225.256744
## [559]    train-rmse:224.187912
## [560]    train-rmse:223.326126
## [561]    train-rmse:222.909424
## [562]    train-rmse:221.159576
## [563]    train-rmse:219.815094
## [564]    train-rmse:218.893280
## [565]    train-rmse:217.961288
## [566]    train-rmse:216.671539
## [567]    train-rmse:215.771576
## [568]    train-rmse:214.660782
## [569]    train-rmse:213.236328
## [570]    train-rmse:212.813660
## [571]    train-rmse:211.953918
## [572]    train-rmse:210.935120
## [573]    train-rmse:209.842163
## [574]    train-rmse:209.428116
## [575]    train-rmse:207.941498
## [576]    train-rmse:207.045197
## [577]    train-rmse:206.059464
## [578]    train-rmse:204.967316
## [579]    train-rmse:204.098587
## [580]    train-rmse:203.004410
## [581]    train-rmse:201.271561
## [582]    train-rmse:200.635025
## [583]    train-rmse:199.770996
## [584]    train-rmse:198.371414
## [585]    train-rmse:197.836578
## [586]    train-rmse:196.522400
## [587]    train-rmse:196.032074
## [588]    train-rmse:195.337311
## [589]    train-rmse:194.519424
## [590]    train-rmse:193.567139
## [591]    train-rmse:192.522705
## [592]    train-rmse:191.735840
## [593]    train-rmse:190.693344
```

```
## [594]    train-rmse:189.823227
## [595]    train-rmse:188.501404
## [596]    train-rmse:187.533691
## [597]    train-rmse:186.734528
## [598]    train-rmse:185.574188
## [599]    train-rmse:184.566620
## [600]    train-rmse:183.717682
## [601]    train-rmse:182.796768
## [602]    train-rmse:181.843719
## [603]    train-rmse:181.010239
## [604]    train-rmse:180.202789
## [605]    train-rmse:179.596161
## [606]    train-rmse:178.701782
## [607]    train-rmse:177.838425
## [608]    train-rmse:177.148163
## [609]    train-rmse:176.363068
## [610]    train-rmse:175.494446
## [611]    train-rmse:174.550095
## [612]    train-rmse:173.572586
## [613]    train-rmse:173.165588
## [614]    train-rmse:172.573456
## [615]    train-rmse:171.515854
## [616]    train-rmse:170.875748
## [617]    train-rmse:170.067337
## [618]    train-rmse:169.352280
## [619]    train-rmse:168.014404
## [620]    train-rmse:167.271545
## [621]    train-rmse:166.387909
## [622]    train-rmse:165.743088
## [623]    train-rmse:165.052078
## [624]    train-rmse:164.145294
## [625]    train-rmse:162.887955
## [626]    train-rmse:162.303223
## [627]    train-rmse:161.752426
## [628]    train-rmse:161.170807
## [629]    train-rmse:160.462128
## [630]    train-rmse:159.450241
## [631]    train-rmse:158.528534
## [632]    train-rmse:157.660507
## [633]    train-rmse:156.847534
## [634]    train-rmse:156.157562
## [635]    train-rmse:155.384583
## [636]    train-rmse:154.787216
## [637]    train-rmse:154.147598
## [638]    train-rmse:153.196106
## [639]    train-rmse:152.762024
## [640]    train-rmse:151.850677
## [641]    train-rmse:151.291504
## [642]    train-rmse:150.467682
## [643]    train-rmse:149.739517
## [644]    train-rmse:149.180145
## [645]    train-rmse:148.575623
## [646]    train-rmse:147.565247
## [647]    train-rmse:146.727570
```

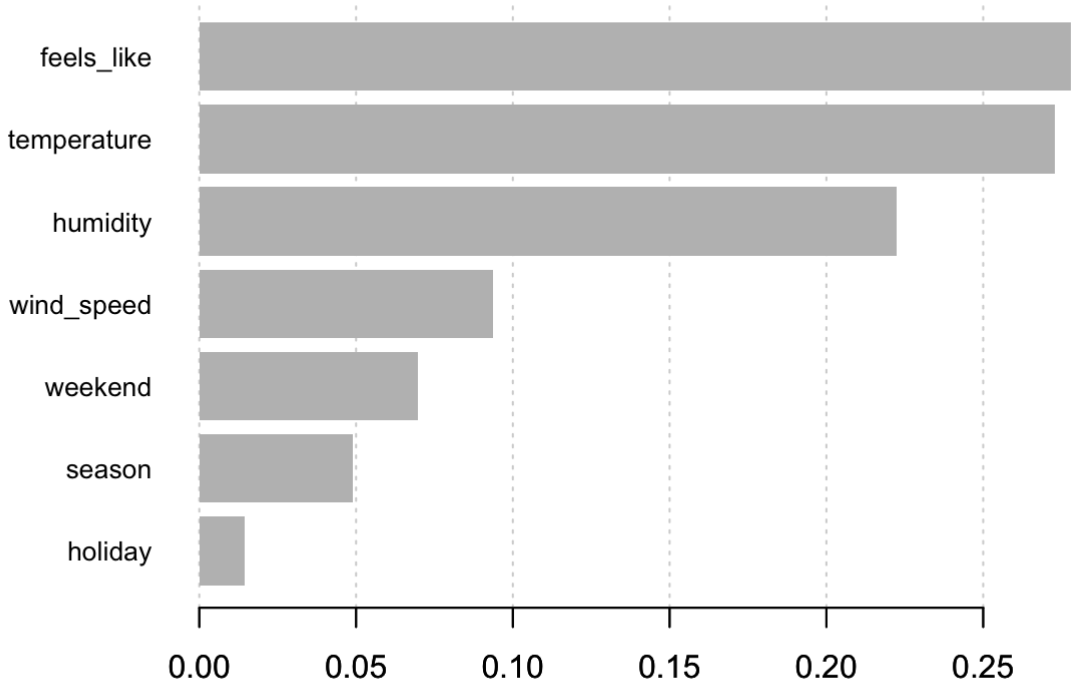
```
## [648]    train-rmse:146.176102
## [649]    train-rmse:145.628250
## [650]    train-rmse:144.762070
## [651]    train-rmse:143.992554
## [652]    train-rmse:143.109070
## [653]    train-rmse:142.137939
## [654]    train-rmse:141.322845
## [655]    train-rmse:140.714874
## [656]    train-rmse:139.941666
## [657]    train-rmse:139.155319
## [658]    train-rmse:138.493759
## [659]    train-rmse:137.981522
## [660]    train-rmse:137.232880
## [661]    train-rmse:136.719299
## [662]    train-rmse:136.251404
## [663]    train-rmse:135.779251
## [664]    train-rmse:134.710831
## [665]    train-rmse:133.989883
## [666]    train-rmse:133.280685
## [667]    train-rmse:132.962280
## [668]    train-rmse:132.053757
## [669]    train-rmse:131.423233
## [670]    train-rmse:130.841919
## [671]    train-rmse:130.063156
## [672]    train-rmse:129.426559
## [673]    train-rmse:128.861954
## [674]    train-rmse:128.343002
## [675]    train-rmse:127.633026
## [676]    train-rmse:127.110550
## [677]    train-rmse:126.472954
## [678]    train-rmse:125.981728
## [679]    train-rmse:125.313057
## [680]    train-rmse:124.726379
## [681]    train-rmse:123.866768
## [682]    train-rmse:123.581703
## [683]    train-rmse:122.875076
## [684]    train-rmse:122.648842
## [685]    train-rmse:121.877899
## [686]    train-rmse:121.240128
## [687]    train-rmse:120.599937
## [688]    train-rmse:120.078949
## [689]    train-rmse:119.291779
## [690]    train-rmse:118.557640
## [691]    train-rmse:117.846733
## [692]    train-rmse:117.327347
## [693]    train-rmse:116.803528
## [694]    train-rmse:116.115593
## [695]    train-rmse:115.814880
## [696]    train-rmse:115.062546
## [697]    train-rmse:114.427513
## [698]    train-rmse:114.004425
## [699]    train-rmse:113.694557
## [700]    train-rmse:112.977196
## [701]    train-rmse:112.406425
```

```
## [702]    train-rmse:111.665642
## [703]    train-rmse:111.099846
## [704]    train-rmse:110.734924
## [705]    train-rmse:110.260368
## [706]    train-rmse:109.612473
## [707]    train-rmse:109.021713
## [708]    train-rmse:108.495285
## [709]    train-rmse:108.145699
## [710]    train-rmse:107.671150
## [711]    train-rmse:107.163239
## [712]    train-rmse:106.538887
## [713]    train-rmse:105.908508
## [714]    train-rmse:105.284164
## [715]    train-rmse:104.700920
## [716]    train-rmse:104.357895
## [717]    train-rmse:103.684113
## [718]    train-rmse:103.041183
## [719]    train-rmse:102.525322
## [720]    train-rmse:101.945755
## [721]    train-rmse:101.587616
## [722]    train-rmse:101.047745
## [723]    train-rmse:100.382217
## [724]    train-rmse:99.958008
## [725]    train-rmse:99.524231
## [726]    train-rmse:98.997765
## [727]    train-rmse:98.609261
## [728]    train-rmse:98.155312
## [729]    train-rmse:97.570976
## [730]    train-rmse:97.052063
## [731]    train-rmse:96.467766
## [732]    train-rmse:95.988510
## [733]    train-rmse:95.514183
## [734]    train-rmse:95.040367
## [735]    train-rmse:94.673027
## [736]    train-rmse:94.154640
## [737]    train-rmse:93.775993
## [738]    train-rmse:93.304619
## [739]    train-rmse:92.900444
## [740]    train-rmse:92.481018
## [741]    train-rmse:92.045090
## [742]    train-rmse:91.505585
## [743]    train-rmse:91.022530
## [744]    train-rmse:90.690674
## [745]    train-rmse:90.264374
## [746]    train-rmse:89.858391
## [747]    train-rmse:89.360527
## [748]    train-rmse:89.094467
## [749]    train-rmse:88.586891
## [750]    train-rmse:88.078224
## [751]    train-rmse:87.817978
## [752]    train-rmse:87.413513
## [753]    train-rmse:86.967369
## [754]    train-rmse:86.469780
## [755]    train-rmse:85.983482
```

```
## [756]    train-rmse:85.780403
## [757]    train-rmse:85.482544
## [758]    train-rmse:84.959801
## [759]    train-rmse:84.531670
## [760]    train-rmse:84.016388
## [761]    train-rmse:83.743118
## [762]    train-rmse:83.389290
## [763]    train-rmse:83.058952
## [764]    train-rmse:82.428413
## [765]    train-rmse:82.115379
## [766]    train-rmse:81.763657
## [767]    train-rmse:81.392952
## [768]    train-rmse:81.119263
## [769]    train-rmse:80.767845
## [770]    train-rmse:80.490318
## [771]    train-rmse:80.140213
## [772]    train-rmse:79.609993
## [773]    train-rmse:79.186119
## [774]    train-rmse:78.749573
## [775]    train-rmse:78.454376
## [776]    train-rmse:78.061440
## [777]    train-rmse:77.504303
## [778]    train-rmse:77.097023
## [779]    train-rmse:76.693268
## [780]    train-rmse:76.358383
## [781]    train-rmse:75.952156
## [782]    train-rmse:75.563896
## [783]    train-rmse:75.121452
## [784]    train-rmse:74.620422
## [785]    train-rmse:74.233864
## [786]    train-rmse:73.947838
## [787]    train-rmse:73.814430
## [788]    train-rmse:73.506378
## [789]    train-rmse:73.128410
## [790]    train-rmse:72.813698
## [791]    train-rmse:72.576759
## [792]    train-rmse:72.245232
## [793]    train-rmse:71.708481
## [794]    train-rmse:71.299179
## [795]    train-rmse:70.891647
## [796]    train-rmse:70.655846
## [797]    train-rmse:70.326317
## [798]    train-rmse:69.887207
## [799]    train-rmse:69.684853
## [800]    train-rmse:69.356483
```

```
y_hat_xgb <- predict(fit_xgb, X_test)

## Plot the feature importance
importance_matrix <- xgb.importance(colnames(X_train), model = fit_xgb)
xgb.plot.importance(importance_matrix = importance_matrix[1:7])
```

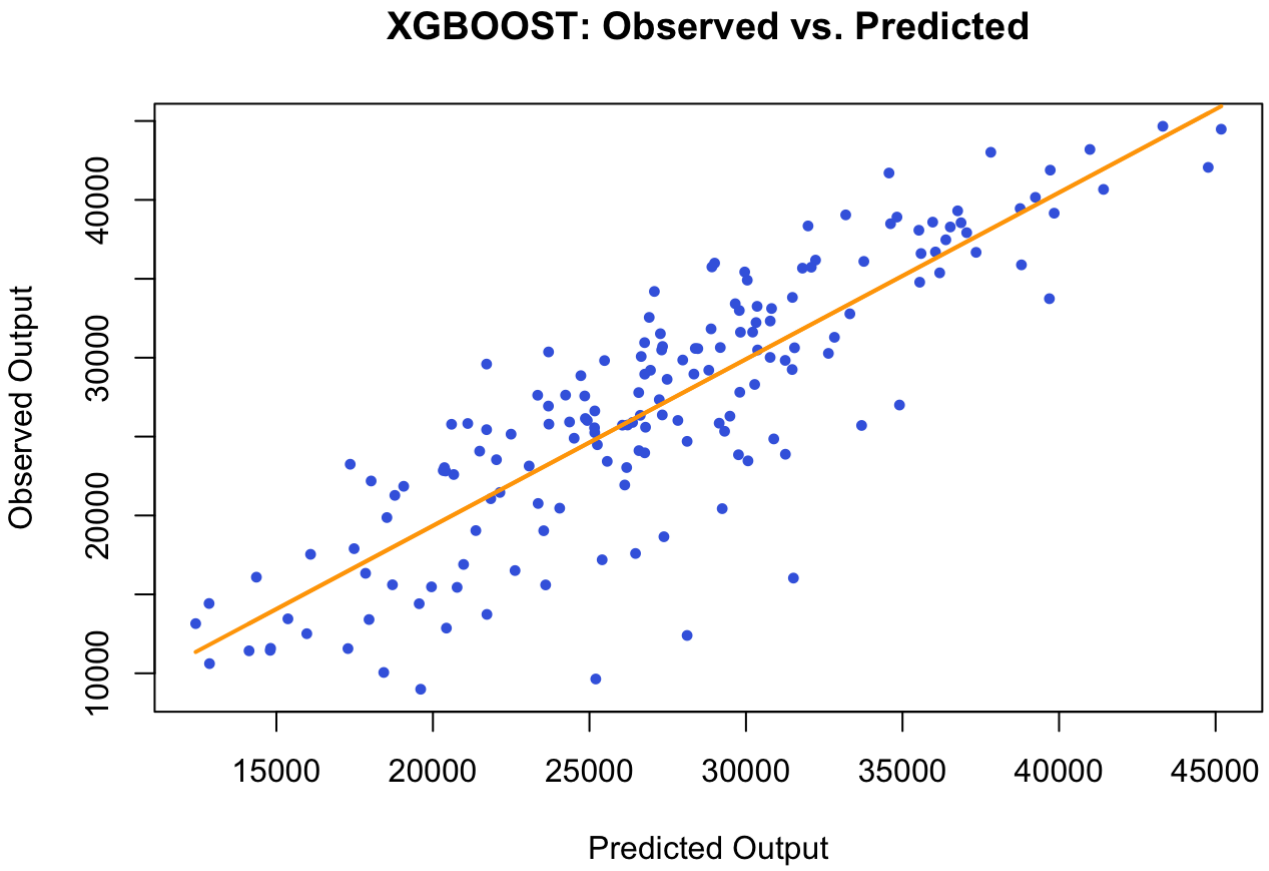



```
data.frame(
  RMSE = RMSE(y_hat_xgb,y_test),
  Rsquare = R2(y_hat_xgb,y_test)
)
```

| RMSE | Rsquare |
|---|---|
| <dbl> | <dbl> |
| 4372.736 | 0.7307053 |

1 row

```
plot(y_hat_xgb, y_test, pch=16, col="royalblue", cex=0.75,
xlab="Predicted Output",
ylab="Observed Output",
main="XGBOOST: Observed vs. Predicted")
lines(y_hat_xgb, lm(a~b, data=data.frame(a=y_test, b=y_hat_xgb))$fitted, lwd=2, col="ora
nge")
```


XGBOOST: Observed vs. Predicted
Observed Output
10000
20000
30000
40000
15000
20000
25000
30000
35000
40000
45000
Predicted Output