454GroupProject

2024-11-13

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

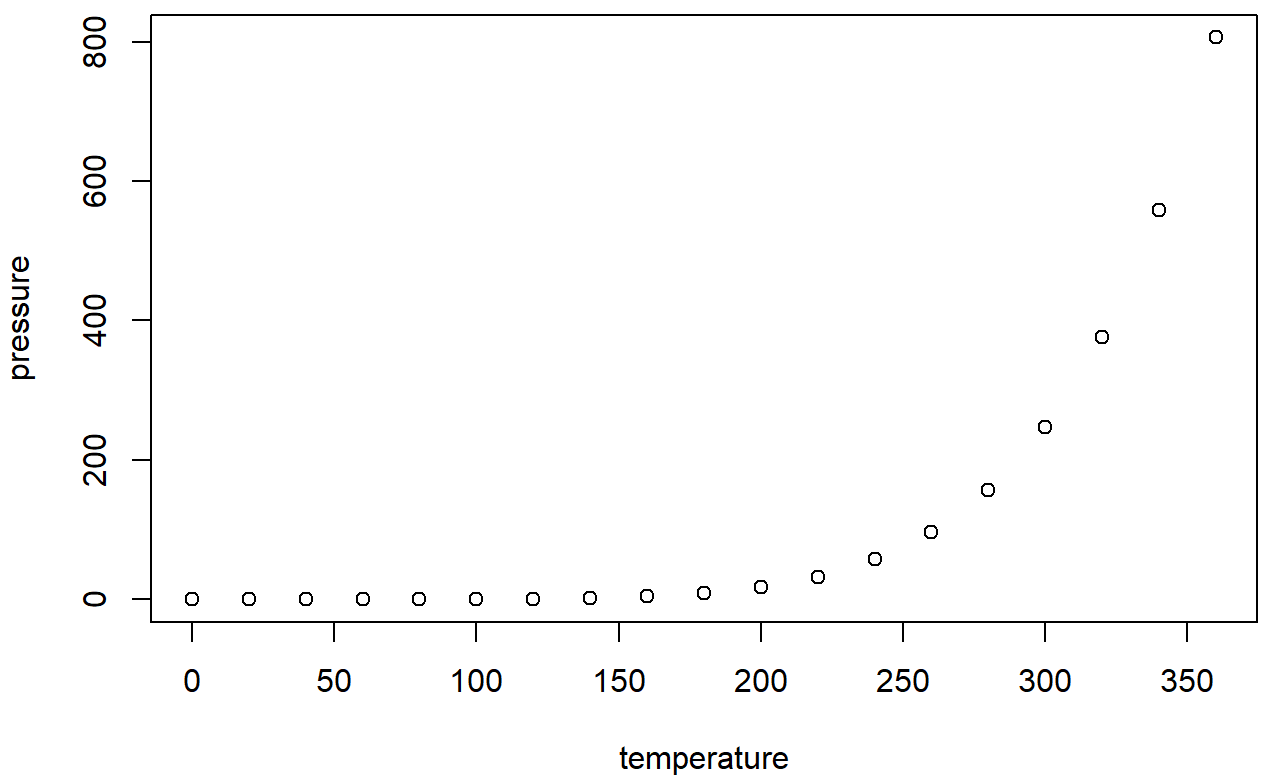
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
library(caret)
```

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

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

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ranger)
library(ggplot2)
```

```
worldbank <- read.csv("C:/Users/conno/Downloads/WorldBankGR (1).csv")
```

```
head(worldbank)
```

```
##   Is.DRC Is.China Is.Russia Is.USA East.Asia...Pacific Europe.and.Central.Asia
## 1      0        0         0      0                   0                       1
## 2      0        0         0      0                   0                       1
## 3      0        0         0      0                   0                       1
## 4      0        0         0      0                   0                       1
## 5      0        0         0      0                   0                       1
## 6      0        0         0      0                   0                       1
##   Latin.America...Carrebian Middle.East...North.Africa North.America South.Asia
## 1                         0                           0             0          0
## 2                         0                           0             0          0
## 3                         0                           0             0          0
## 4                         0                           0             0          0
## 5                         0                           0             0          0
## 6                         0                           0             0          0
##   Sub.Saharan.Africa IncomeGroupRanking Year Birth.rate Death.rate
## 1                  0                 3 2014      12.26       7.22
## 2                  0                 3 2013      12.26       7.10
## 3                  0                 3 2012      12.20       7.00
## 4                  0                 3 2011      12.10       6.92
## 5                  0                 3 2010      12.00       6.84
## 6                  0                 3 2009      11.95       6.76
##   Electric.power.consumption         GDP GDP.per.capita
## 1                    2309.37 13228200000        4578.67
## 2                    2533.25 12776300000        4413.08
## 3                    2118.33 12319800000        4247.61
## 4                    2205.70 12890900000        4437.18
## 5                    1943.34 11927000000        4094.36
## 6                    1835.68 12044200000        4114.13
##   Individuals.using.the.Internet Infant.mortality.rate Life.expectancy
## 1                          60.10                   8.9           77.81
## 2                          57.20                   9.5           77.55
## 3                          54.66                  10.2           77.25
## 4                          49.00                  11.0           76.91
## 5                          45.00                  11.9           76.56
## 6                          41.20                  12.9           76.22
##   Population.density Unemployment..
## 1             105.44          17.49
## 2             105.66          15.87
## 3             105.85          13.38
## 4             106.03          13.48
## 5             106.32          14.09
## 6             106.84          13.67
```

```
sample <- sample.int(n = nrow(worldbank), size = nrow(worldbank)*0.8, replace = F)
worldbank_train <- worldbank[sample, ] ##Yields training dataset that is the training percentage %
worldbank_validation <- worldbank[-sample, ] ##Yields validation dataset
```

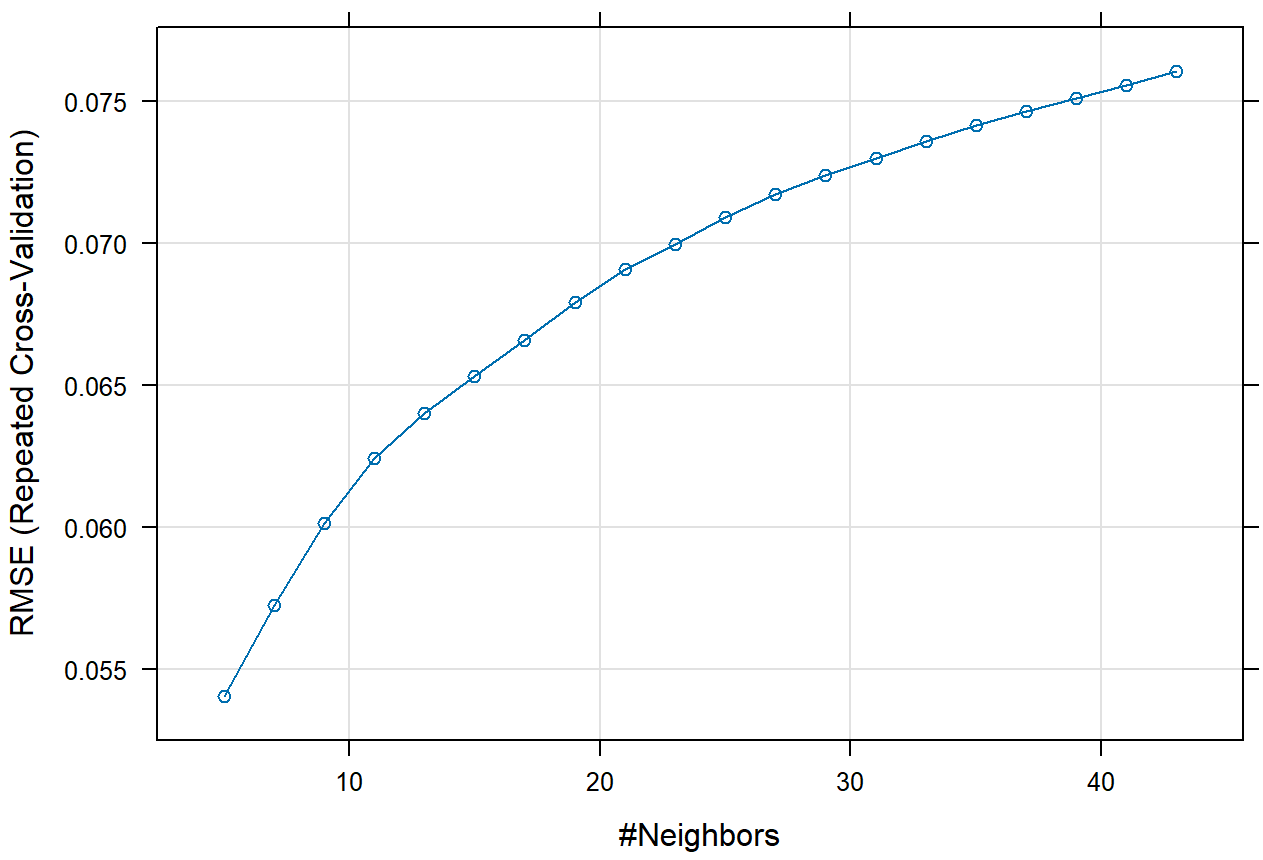
```
preProcValues <- preProcess(worldbank, method = c("range")) ##Uses column minimums & maximums to normalize values around 0 u
sing original data
```

```
worldbank_train_norm <- predict(preProcValues, worldbank_train) #Using the normalizing object, normalize the rows in the dat
aframe and save it new to a new one
worldbank_validation_norm <- predict(preProcValues, worldbank_validation) #Using the normalizing object, normalize the rows
in the dataframe and save it new to a new one
```

```
##Train kNN model on normalized data & select optimal k
ctrl <- trainControl(method="repeatedcv",repeats = 3)  #Set training parameters
knnFit <- train(Life.expectancy ~ Is.DRC + Is.China + Is.Russia + Is.USA + East.Asia...Pacific + Europe.and.Central.Asia + L
atin.America...Carrebian + Middle.East...North.Africa + North.America + South.Asia + Sub.Saharan.Africa + IncomeGroupRanking
+ Year + Electric.power.consumption + GDP + GDP.per.capita + Individuals.using.the.Internet + Population.density + Unemploym
ent.., data = worldbank_train_norm, method = "knn", trControl = ctrl, tuneLength = 20) #Test various values of k on normaliz
ed training data.
knnFit ##Displays the relative performance of different values of k
```

```
## k-Nearest Neighbors
##
## 2220 samples
##   19 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 1998, 1997, 1997, 2000, 1998, 1998, ...
## Resampling results across tuning parameters:
##
##   k   RMSE        Rsquared   MAE
##    5  0.05402226  0.9344809  0.03755880
##    7  0.05722940  0.9265174  0.04101281
##    9  0.06013179  0.9187805  0.04384209
##   11  0.06242278  0.9125093  0.04591509
##   13  0.06401335  0.9079555  0.04737137
##   15  0.06529957  0.9042436  0.04861998
##   17  0.06656687  0.9005074  0.04962620
##   19  0.06792919  0.8963129  0.05053068
##   21  0.06909347  0.8927283  0.05129087
##   23  0.06996653  0.8900062  0.05182073
##   25  0.07092157  0.8869901  0.05242335
##   27  0.07174006  0.8843246  0.05292356
##   29  0.07240480  0.8821644  0.05330156
##   31  0.07299791  0.8802301  0.05370725
##   33  0.07360561  0.8782289  0.05408687
##   35  0.07414744  0.8764243  0.05441994
##   37  0.07466819  0.8746489  0.05472912
##   39  0.07512569  0.8730609  0.05501467
##   41  0.07557929  0.8714838  0.05533306
##   43  0.07607792  0.8697688  0.05568916
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was k = 5.
```

```
plot(knnFit) #Plot the accuracy of various k values
```




```
knn_validation_predictions <- predict(knnFit,newdata = worldbank_validation_norm) #Generate validation data predictions
```

```
head(worldbank_validation_norm)
```

```
##    Is.DRC Is.China Is.Russia Is.USA East.Asia...Pacific Europe.and.Central.Asia
## 3       0        0         0      0                   0                       1
## 5       0        0         0      0                   0                       1
## 7       0        0         0      0                   0                       1
## 9       0        0         0      0                   0                       1
## 12      0        0         0      0                   0                       1
## 13      0        0         0      0                   0                       1
##    Latin.America...Carrebian Middle.East...North.Africa North.America
## 3                          0                          0             0
## 5                          0                          0             0
## 7                          0                          0             0
## 9                          0                          0             0
## 12                         0                          0             0
## 13                         0                          0             0
##    South.Asia Sub.Saharan.Africa IncomeGroupRanking      Year Birth.rate
## 3           0                  0          0.6666667 0.9130435 0.10013061
## 5           0                  0          0.6666667 0.8260870 0.09577710
## 7           0                  0          0.6666667 0.7391304 0.09512407
## 9           0                  0          0.6666667 0.6521739 0.10448411
## 12          0                  0          0.6666667 0.5217391 0.14040052
## 13          0                  0          0.6666667 0.4782609 0.15650849
##    Death.rate Electric.power.consumption          GDP GDP.per.capita
## 3   0.3257492                 0.03825678 0.0006772711     0.03491376
## 5   0.3168701                 0.03506215 0.0006548248     0.03362292
## 7   0.3063263                 0.02911799 0.0007093634     0.03594921
## 9   0.2907880                 0.02182690 0.0004816241     0.02417542
## 12  0.2674806                 0.02640734 0.0002939288     0.01468581
## 13  0.2641509                 0.02839560 0.0002217315     0.01113969
##    Individuals.using.the.Internet Infant.mortality.rate Life.expectancy
## 3                     0.556845966            0.06135866       0.8435341
## 5                     0.458435208            0.07377648       0.8265054
## 7                     0.243072535            0.08838568       0.8104640
## 9                     0.097901385            0.10372535       0.7983712
## 12                    0.009881826            0.12856099       0.7838105
## 13                    0.003973105            0.13732652       0.7776407
##    Population.density Unemployment..
## 3          0.01353131      0.3559093
## 5          0.01359225      0.3750675
## 7          0.01375431      0.3472747
## 9          0.01396823      0.4371290
## 12         0.01419122      0.4670804
## 13         0.01424438      0.4673502
```

```
postResample(pred=knn_validation_predictions, worldbank_validation_norm$Life.expectancy)
```

```
##       RMSE   Rsquared        MAE
## 0.04990680 0.94362431 0.03650703
```