DC Bike Sharing Analysis

Stephanie Spitzer, Anamika Kakhanal, and Susan Khan

```
library (dplyr)
```

```
## Registered S3 method overwritten by 'dplyr':
##   method               from
##   as.data.frame.tbl_df tibble

##
## 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 (readr)
library (rpart)
library (rpart.plot)
data<-read_csv("day.csv")
```

```
## Parsed with column specification:
## cols(
##   instant = col_integer(),
##   dteday = col_date(format = ""),
##   season = col_integer(),
##   yr = col_integer(),
##   mnth = col_integer(),
##   holiday = col_integer(),
##   weekday = col_integer(),
##   workingday = col_integer(),
##   weathersit = col_integer(),
##   temp = col_double(),
##   atemp = col_double(),
##   hum = col_double(),
##   windspeed = col_double(),
##   casual = col_integer(),
##   registered = col_integer(),
##   cnt = col_integer()
## )
```

Checking for missing data

```
sum(is.na(data$atemp))
```

```
## [1] 0
```

```
sum(is.na(data$weekday))
```

```
## [1] 0
```

```r
sum(is.na(data$weathersit))
```

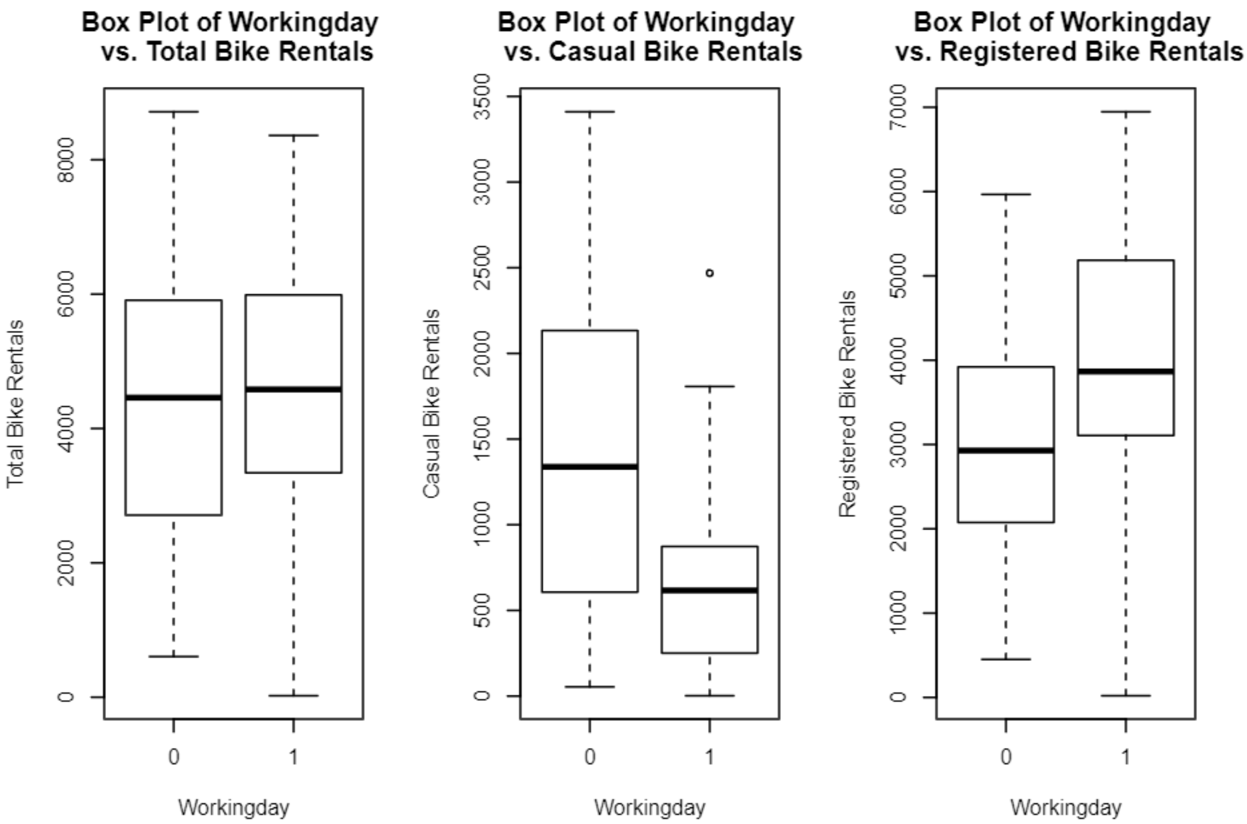
```
## [1] 0
```

From the three lines of codes above, we see that for each variable we are interested in, there is no missing data.

Exploratory Analysis

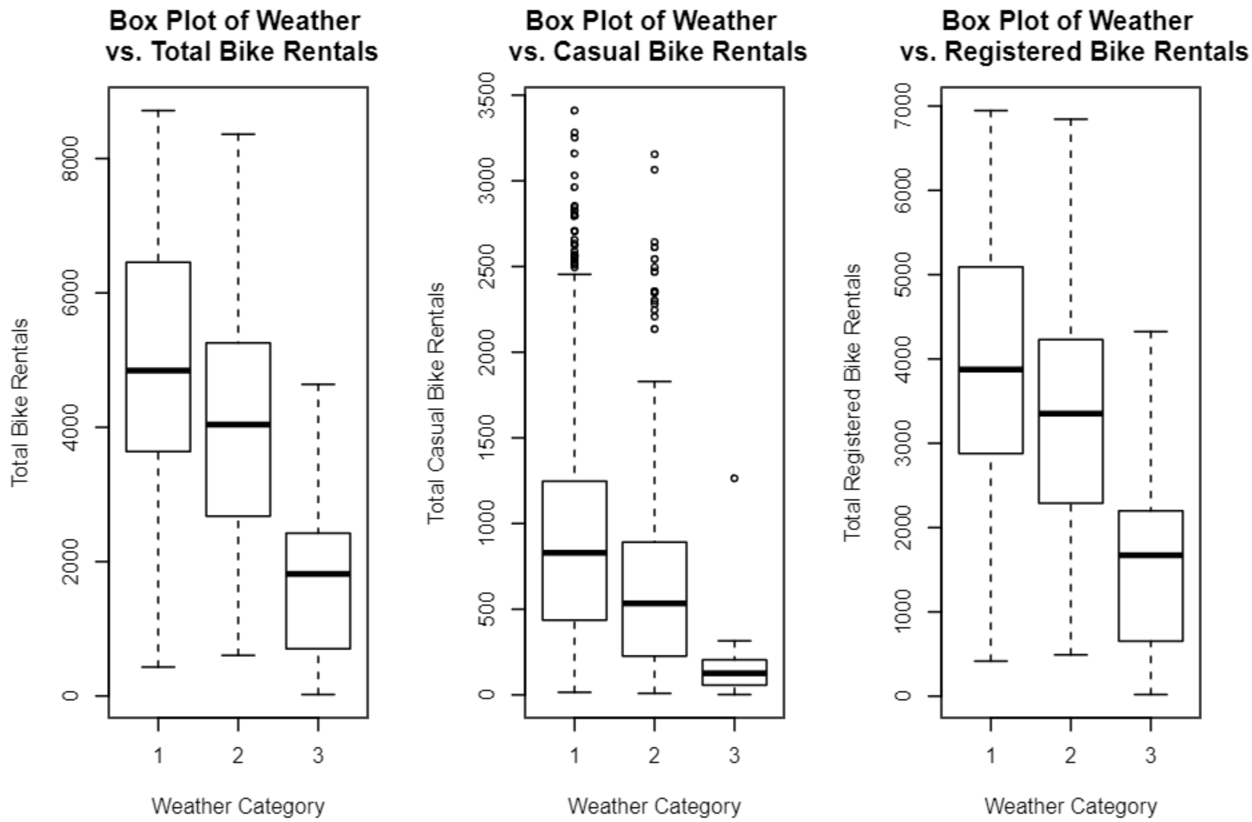
Workingday Box Plots

```r
par(mfrow = c(1,3))
boxplot(cnt ~ workingday, data = data, xlab = "Workingday", ylab = "Total Bike Rentals", main = "Box Pl
boxplot(casual ~ workingday, data = data, xlab = "Workingday", ylab = "Casual Bike Rentals", main = "Bo
boxplot(registered ~ workingday, data = data, xlab = "Workingday", ylab = "Registered Bike Rentals", ma
```



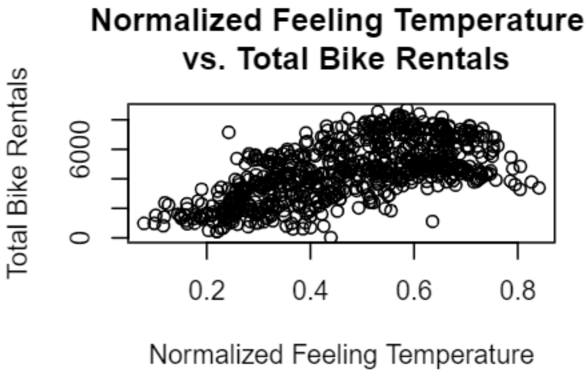
Weather Box Plots

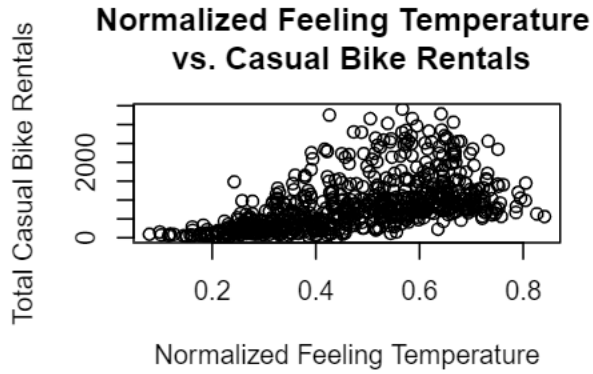
```r
par(mfrow = c(1,3))
boxplot(cnt ~ weathersit, data = data, xlab = "Weather Category", ylab = "Total Bike Rentals", main = "
boxplot(casual ~ weathersit, data = data, xlab = "Weather Category", ylab = "Total Casual Bike Rentals"
boxplot(registered ~ weathersit, data = data, xlab = "Weather Category", ylab = "Total Registered Bike 
```

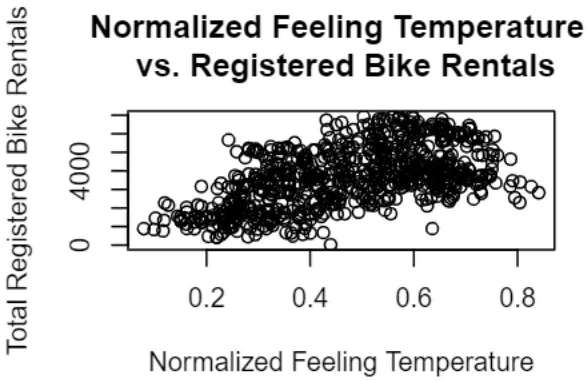


Normalized Feeling Temp Scatter Plots

```
par(mfrow = c(2, 2))
plot(data$atemp, data$cnt, xlab = "Normalized Feeling Temperature", ylab = "Total Bike Rentals", main =
plot(data$atemp, data$casual, xlab = "Normalized Feeling Temperature", ylab = "Total Casual Bike Rentals
plot(data$atemp, data$registered, xlab = "Normalized Feeling Temperature", ylab = "Total Registered Bike
```





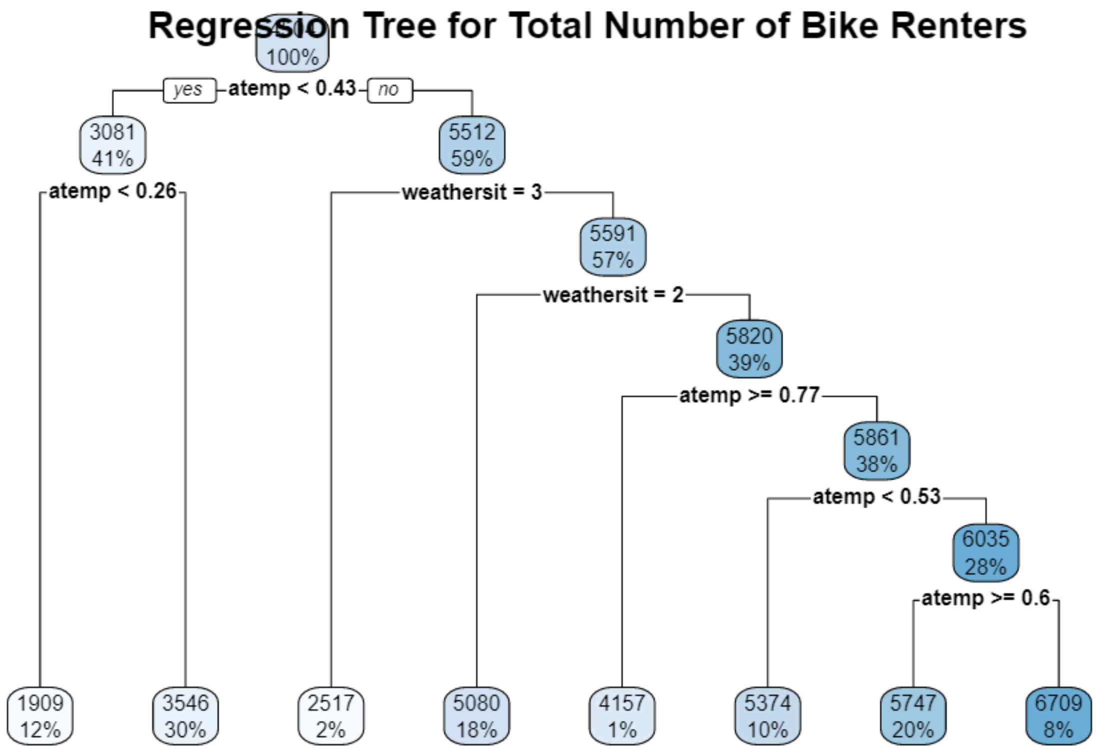


For the weather and the working day plots, we see that the plots are non-linear because these variables are categorical. For the atemp plot, we see a positive linear relationship, which means as the temperature increases, more people will go out biking. Based on the results from the plots, we see that a tree-based regression using count as the dependent variable and weather, working day, and atemp as our independent variables is appropriate because we can easily split the data. Since atemp is relatively linear, we will use the Gini coefficient to help split the data into separate regions recursively. Additionally, there are four categories for the weather variable, but from the weather plot, we see that there are no days with category four weather (i.e., heavy rain, ice pallets, thunderstorms, mist, fog, or snow). Therefore, when conducting our analysis, we will only be using the other three weather categories.

Relatively Advanced Method From Class (Tree-Based Regression)

Tree Based Regression for Total

```
#Recursive Binary Splitting
options(repr.plot.width=20,repr.plot.height=20)

data$workingday <- as.factor(data$workingday) #changes from integer to categorical
data$weathersit <- as.factor(data$weathersit) #changes from integer to categorical

count.tree <- rpart(cnt~atemp + workingday + weathersit, method="anova", data=data)

rpart.plot(count.tree, uniform = TRUE)
title("Regression Tree for Total Number of Bike Renters", cex = 0.5)
```




```r
summary(count.tree)
```

```
## Call:
## rpart(formula = cnt ~ atemp + workingday + weathersit, data = data,
##     method = "anova")
##   n= 731
##
##           CP nsplit rel error    xerror       xstd
## 1 0.38256827      0 1.0000000 1.0044145 0.04046663
## 2 0.06020466      1 0.6174317 0.6507717 0.02994502
## 3 0.03696960      2 0.5572271 0.5954162 0.02987557
## 4 0.01777407      3 0.5202575 0.5568799 0.02592968
## 5 0.01019724      4 0.5024834 0.5360394 0.02460772
## 6 0.01000000      7 0.4718917 0.5259857 0.02502552
##
## Variable importance
##      atemp weathersit
##         90         10
##
## Node number 1: 731 observations,    complexity param=0.3825683
##   mean=4504.349, MSE=3747654
##   left son=2 (303 obs) right son=3 (428 obs)
##   Primary splits:
##       atemp      < 0.4308565 to the left,  improve=0.382568300, (0 missing)
##       weathersit splits as  RLL, improve=0.063943010, (0 missing)
##       workingday splits as  LR, improve=0.003740064, (0 missing)
##
## Node number 2: 303 observations,    complexity param=0.06020466
##   mean=3081.251, MSE=2346720
##   left son=4 (86 obs) right son=5 (217 obs)
```

```
##   Primary splits:
##       atemp      < 0.2607295 to the left,  improve=0.23195470, (0 missing)
##       weathersit splits as  RRL, improve=0.06188205, (0 missing)
##       workingday splits as  LR, improve=0.01159276, (0 missing)
##
## Node number 3: 428 observations,    complexity param=0.0369696
##   mean=5511.822, MSE=2290701
##   left son=6 (11 obs) right son=7 (417 obs)
##   Primary splits:
##       weathersit splits as  RRL, improve=0.103302200, (0 missing)
##       atemp      < 0.5296815 to the left,  improve=0.044495530, (0 missing)
##       workingday splits as  RL, improve=0.002113189, (0 missing)
##
## Node number 4: 86 observations
##   mean=1909.291, MSE=1003270
##
## Node number 5: 217 observations
##   mean=3545.714, MSE=2119088
##
## Node number 6: 11 observations
##   mean=2516.727, MSE=1439988
##
## Node number 7: 417 observations,    complexity param=0.01777407
##   mean=5590.83, MSE=2070265
##   left son=14 (129 obs) right son=15 (288 obs)
##   Primary splits:
##       weathersit splits as  RL-, improve=0.0564029300, (0 missing)
##       atemp      < 0.501571  to the left,  improve=0.0316158400, (0 missing)
##       workingday splits as  RL, improve=0.0003208998, (0 missing)
##
## Node number 14: 129 observations
##   mean=5080.248, MSE=2087606
##
## Node number 15: 288 observations,    complexity param=0.01019724
##   mean=5819.528, MSE=1893426
##   left son=30 (7 obs) right son=31 (281 obs)
##   Primary splits:
##       atemp      < 0.773667  to the right, improve=0.0363647900, (0 missing)
##       workingday splits as  RL, improve=0.0003670386, (0 missing)
##
## Node number 30: 7 observations
##   mean=4157, MSE=598708
##
## Node number 31: 281 observations,    complexity param=0.01019724
##   mean=5860.943, MSE=1855110
##   left son=62 (74 obs) right son=63 (207 obs)
##   Primary splits:
##       atemp      < 0.529669  to the left,  improve=0.045654800, (0 missing)
##       workingday splits as  RL, improve=0.001572414, (0 missing)
##
## Node number 62: 74 observations
##   mean=5374.203, MSE=1920185
##
## Node number 63: 207 observations,    complexity param=0.01019724
```

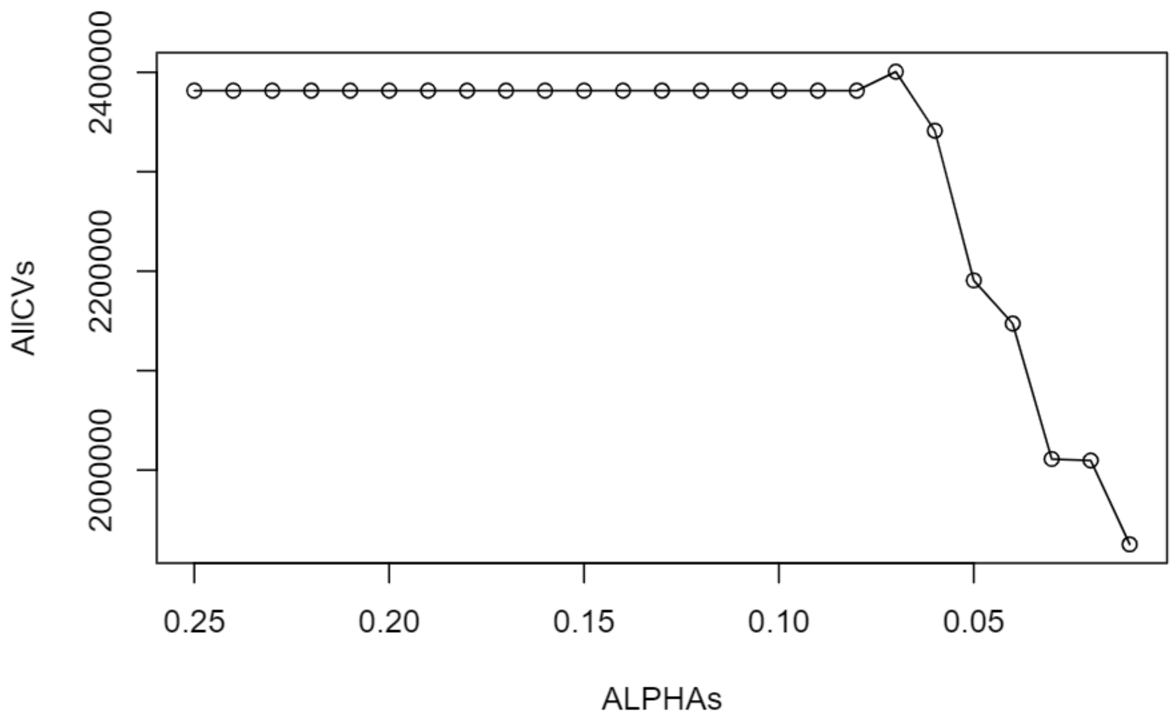
```
##    mean=6034.947, MSE=1716874
##    left son=126 (145 obs) right son=127 (62 obs)
##    Primary splits:
##        atemp      < 0.6000685 to the right, improve=0.113052100, (0 missing)
##        workingday splits as  RL, improve=0.001489815, (0 missing)
##
## Node number 126: 145 observations
##    mean=5746.862, MSE=1431058
##
## Node number 127: 62 observations
##    mean=6708.694, MSE=1737284
```

Pruned Tree for Total Bike Rentals

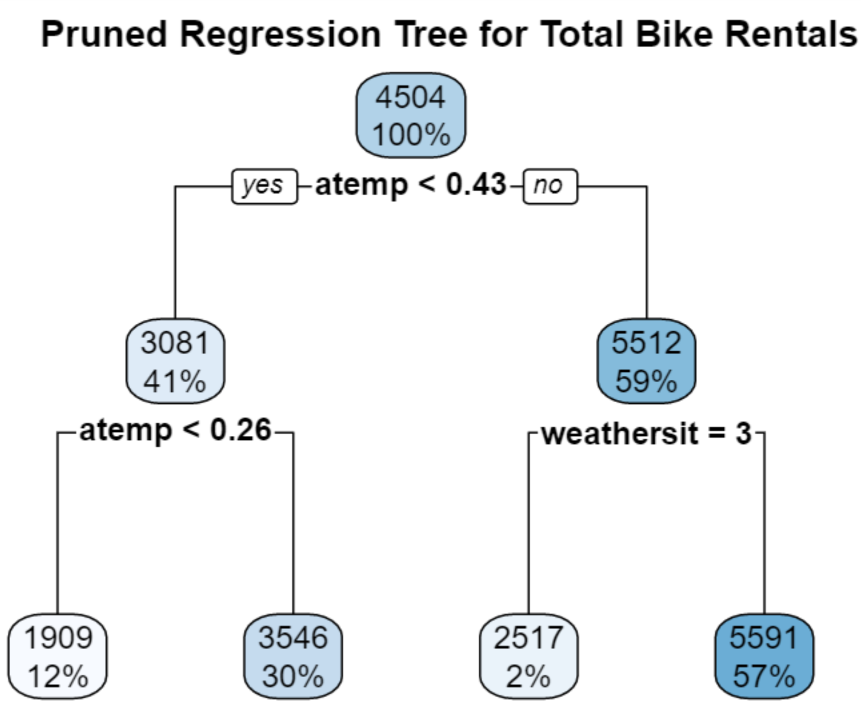
```r
#Complexity Pruning
N = length(data)
K = 10
ALPHAs= seq(0.25,0.01,-0.01)
# split into K folds
Kfolds = split(data,1:K)
```

```
## Warning in split.default(x = seq_len(nrow(x)), f = f, drop = drop, ...):
## data length is not a multiple of split variable
```

```r
AllCVs = rep(0,length(ALPHAs))

i=1
for (ALPHA in ALPHAs){
    MSEs = rep(0,K)
    for(k in 1:K){
        trainingIndices = setdiff(1:K,k)
        trainingData = do.call(rbind,Kfolds[trainingIndices])
        testingData  = Kfolds[[k]]

        BigTree <- rpart(cnt~atemp + workingday + weathersit, method="anova", data=trainingData)
        smallerTree = prune(BigTree, cp = ALPHA)

        predictions = predict(smallerTree, testingData)
        MSEs[k] = t(testingData$cnt - predictions)%*%(testingData$cnt - predictions)/nrow(testingData)
    }
    AllCVs[i] = mean(MSEs)
    i=i+1
}

plot(ALPHAs,AllCVs, xlim=c(0.25,0.01) )
lines(ALPHAs,AllCVs)
```

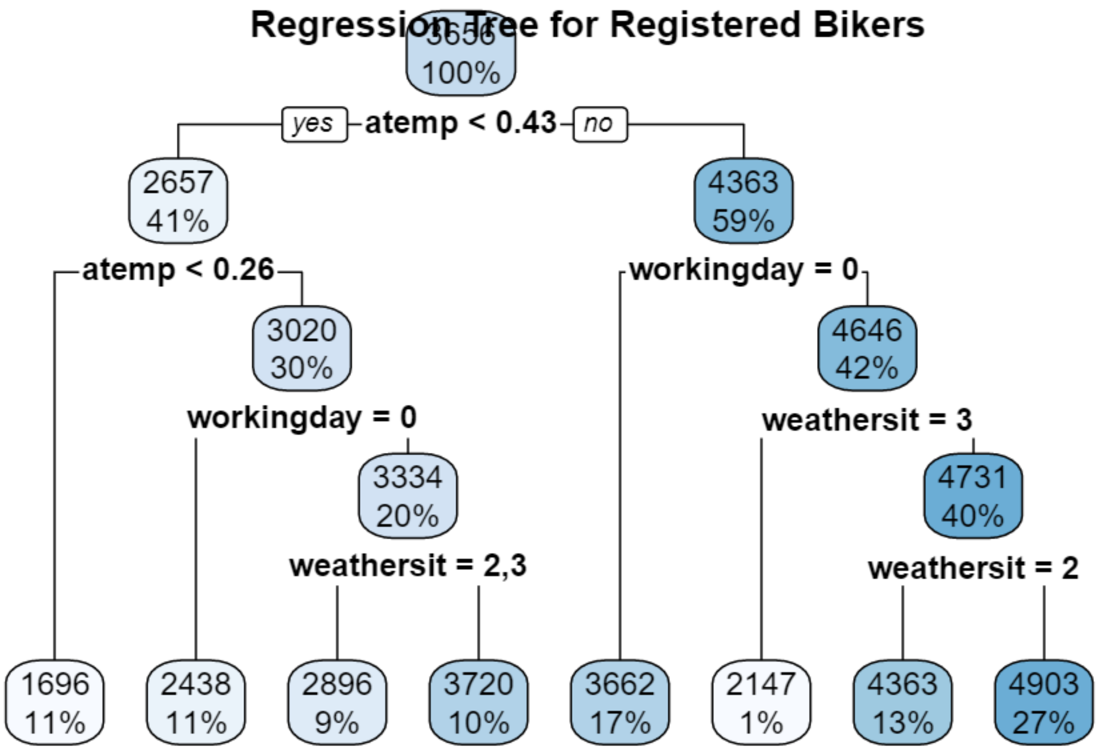



```r
#Pruned Tree
prune.total<- prune(count.tree, cp=0.03) # from cptable

# plot the pruned tree
rpart.plot(prune.total, uniform=TRUE)
title("Pruned Regression Tree for Total Bike Rentals")
```



Tree-Based Regression for Registered Bikers

```r
#Recursive Binary Splitting
registered.tree <- rpart(registered~atemp + workingday + weathersit, method="anova", data=data)

rpart.plot(registered.tree, uniform =TRUE)
title("Regression Tree for Registered Bikers")
```




```r
summary(registered.tree)
```

```
## Call:
## rpart(formula = registered ~ atemp + workingday + weathersit,
##     data = data, method = "anova")
##   n= 731
##
##           CP nsplit rel error    xerror       xstd
## 1 0.29051946      0 1.0000000 1.0021211 0.04195643
## 2 0.05945857      1 0.7094805 0.7185669 0.02950334
## 3 0.04781571      2 0.6500220 0.6634903 0.03004455
## 4 0.03633657      3 0.6022063 0.6139530 0.02878739
## 5 0.02262316      4 0.5658697 0.5798279 0.02534010
## 6 0.01361291      5 0.5432465 0.5582973 0.02433754
## 7 0.01050642      6 0.5296336 0.5604168 0.02479796
## 8 0.01000000      7 0.5191272 0.5500890 0.02448670
##
## Variable importance
##      atemp workingday weathersit
##         73         15         12
##
## Node number 1: 731 observations,    complexity param=0.2905195
```

```
##   mean=3656.172, MSE=2431070
##   left son=2 (303 obs) right son=3 (428 obs)
##   Primary splits:
##       atemp      < 0.4308565 to the left,  improve=0.29051950, (0 missing)
##       workingday splits as  LR, improve=0.09235954, (0 missing)
##       weathersit splits as  RRL, improve=0.05055062, (0 missing)
##
## Node number 2: 303 observations,    complexity param=0.05945857
##   mean=2657.353, MSE=1771702
##   left son=4 (83 obs) right son=5 (220 obs)
##   Primary splits:
##       atemp      < 0.2578935 to the left,  improve=0.19683210, (0 missing)
##       workingday splits as  LR, improve=0.07768392, (0 missing)
##       weathersit splits as  RRL, improve=0.05742509, (0 missing)
##
## Node number 3: 428 observations,    complexity param=0.04781571
##   mean=4363.28, MSE=1691591
##   left son=6 (123 obs) right son=7 (305 obs)
##   Primary splits:
##       workingday splits as  LR, improve=0.11736710, (0 missing)
##       weathersit splits as  RRL, improve=0.07014692, (0 missing)
##       atemp      < 0.5296815 to the left,  improve=0.02452893, (0 missing)
##   Surrogate splits:
##       atemp < 0.7885045 to the right, agree=0.717, adj=0.016, (0 split)
##
## Node number 4: 83 observations
##   mean=1695.928, MSE=654841.7
##
## Node number 5: 220 observations,    complexity param=0.02262316
##   mean=3020.073, MSE=1712769
##   left son=10 (77 obs) right son=11 (143 obs)
##   Primary splits:
##       workingday splits as  LR, improve=0.10669560, (0 missing)
##       weathersit splits as  RLL, improve=0.03531297, (0 missing)
##       atemp      < 0.321956  to the left,  improve=0.01990192, (0 missing)
##   Surrogate splits:
##       atemp < 0.4125545 to the right, agree=0.664, adj=0.039, (0 split)
##
## Node number 6: 123 observations
##   mean=3661.634, MSE=975018.3
##
## Node number 7: 305 observations,    complexity param=0.03633657
##   mean=4646.239, MSE=1701966
##   left son=14 (10 obs) right son=15 (295 obs)
##   Primary splits:
##       weathersit splits as  RRL, improve=0.12439650, (0 missing)
##       atemp      < 0.5296815 to the left,  improve=0.03854313, (0 missing)
##
## Node number 10: 77 observations
##   mean=2437.506, MSE=1361205
##
## Node number 11: 143 observations,    complexity param=0.01361291
##   mean=3333.762, MSE=1620927
##   left son=22 (67 obs) right son=23 (76 obs)
```

```
##   Primary splits:
##       weathersit splits as  RLL, improve=0.10436780, (0 missing)
##       atemp      < 0.332973  to the left,  improve=0.02592348, (0 missing)
##   Surrogate splits:
##       atemp < 0.3588895 to the right, agree=0.58, adj=0.104, (0 split)
##
## Node number 14: 10 observations
##   mean=2147.1, MSE=1016477
##
## Node number 15: 295 observations,    complexity param=0.01050642
##   mean=4730.956, MSE=1506307
##   left son=30 (94 obs) right son=31 (201 obs)
##   Primary splits:
##       weathersit splits as  RL-, improve=0.04201787, (0 missing)
##       atemp      < 0.5296815 to the left,  improve=0.02039954, (0 missing)
##
## Node number 22: 67 observations
##   mean=2895.701, MSE=1674909
##
## Node number 23: 76 observations
##   mean=3719.947, MSE=1255026
##
## Node number 30: 94 observations
##   mean=4363.074, MSE=1472287
##
## Node number 31: 201 observations
##   mean=4903, MSE=1429326
```

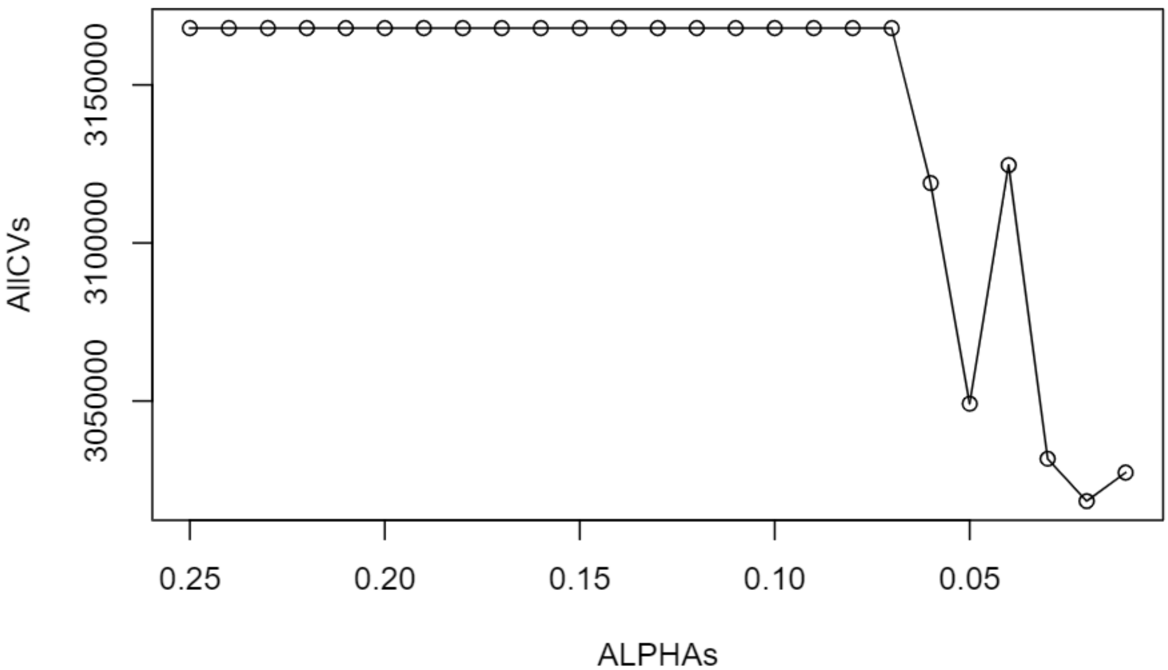
Pruned Tree for Registered Bikers

```r
#Complexity Pruning
N = length(data)
K = 10
ALPHAs= seq(0.25,0.01,-0.01)
# split into K folds
Kfolds = split(data,1:K)
```

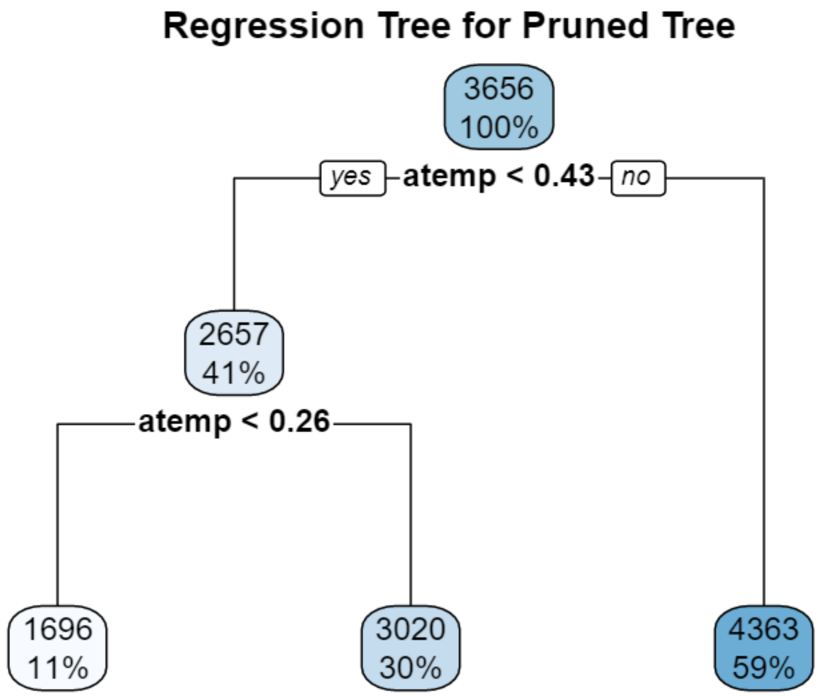
```
## Warning in split.default(x = seq_len(nrow(x)), f = f, drop = drop, ...):
## data length is not a multiple of split variable
```

```r
AllCVs = rep(0,length(ALPHAs))

i=1
for (ALPHA in ALPHAs){
    MSEs = rep(0,K)
    for(k in 1:K){
        trainingIndices = setdiff(1:K,k)
        trainingData = do.call(rbind,Kfolds[trainingIndices])
        testingData  = Kfolds[[k]]

        BigTree <- rpart(registered~atemp + workingday + weathersit, method="anova", data=trainingData)
        smallerTree = prune(BigTree, cp = ALPHA)
```

```
        predictions = predict(smallerTree, testingData)
        MSEs[k] = t(testingData$cnt - predictions)%*%(testingData$cnt - predictions)/nrow(testingData)
    }
    AllCVs[i] = mean(MSEs)
    i=i+1
}

plot(ALPHAs,AllCVs, xlim=c(0.25,0.01) )
lines(ALPHAs,AllCVs)
```

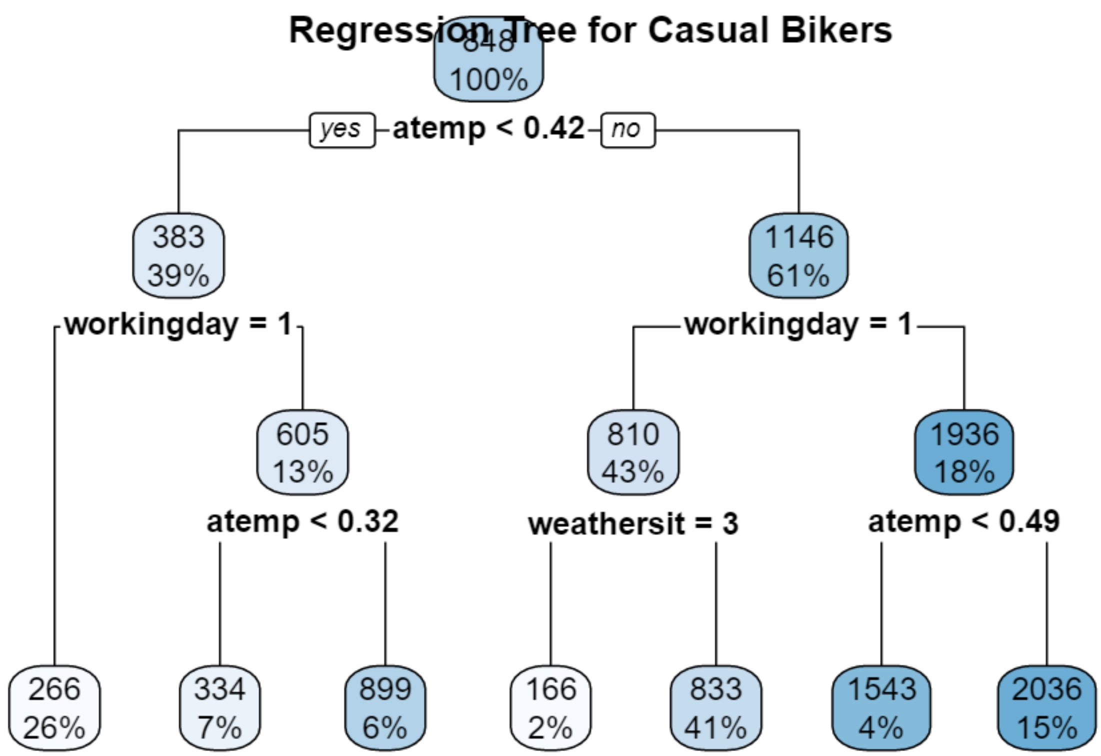



```
#Pruned Tree
prune.registered <- prune(registered.tree, cp=0.05)# from cp graph

# plot the pruned tree
rpart.plot(prune.registered, uniform=TRUE)
title("Regression Tree for Pruned Tree")
```



#Tree-Based Regression for Casual Bikers

```r
#Recursive Binary Splitting
casual.tree <- rpart(casual~atemp + workingday + weathersit, method="anova", data=data)

rpart.plot(casual.tree, uniform=TRUE)
title("Regression Tree for Casual Bikers")
```



```r
summary(casual.tree)
```

```
## Call:
## rpart(formula = casual ~ atemp + workingday + weathersit, data = data,
##     method = "anova")
##   n= 731
##
##           CP nsplit rel error    xerror       xstd
## 1 0.31894036      0 1.0000000 1.0032222 0.06755632
## 2 0.02203466      2 0.3621193 0.3766321 0.03103055
## 3 0.01519603      4 0.3180500 0.3342008 0.02834977
## 4 0.01374834      5 0.3028539 0.3370210 0.02856322
## 5 0.01000000      6 0.2891056 0.3384770 0.02816899
##
## Variable importance
## workingday      atemp weathersit
##         50         48          2
##
## Node number 1: 731 observations,    complexity param=0.3189404
##   mean=848.1765, MSE=470805.5
##   left son=2 (285 obs) right son=3 (446 obs)
##   Primary splits:
##       atemp      < 0.417285  to the left,  improve=0.29416540, (0 missing)
##       workingday splits as  RL, improve=0.26836980, (0 missing)
##       weathersit splits as  RLL, improve=0.04925213, (0 missing)
##
## Node number 2: 285 observations,    complexity param=0.02203466
##   mean=382.6316, MSE=131083.1
##   left son=4 (187 obs) right son=5 (98 obs)
##   Primary splits:
##       workingday splits as  RL, improve=0.19752650, (0 missing)
##       atemp      < 0.3102565 to the left,  improve=0.17618810, (0 missing)
##       weathersit splits as  RLL, improve=0.05008473, (0 missing)
##   Surrogate splits:
##       atemp < 0.1170525 to the right, agree=0.67, adj=0.041, (0 split)
##
## Node number 3: 446 observations,    complexity param=0.3189404
##   mean=1145.666, MSE=460898.1
##   left son=6 (313 obs) right son=7 (133 obs)
##   Primary splits:
##       workingday splits as  RL, improve=0.57546390, (0 missing)
##       weathersit splits as  RRL, improve=0.05263446, (0 missing)
##       atemp      < 0.5659775 to the left,  improve=0.03986356, (0 missing)
##   Surrogate splits:
##       atemp < 0.427125  to the right, agree=0.715, adj=0.045, (0 split)
##
## Node number 4: 187 observations
##   mean=266.1444, MSE=57142.27
##
## Node number 5: 98 observations,    complexity param=0.02203466
##   mean=604.9082, MSE=196874.9
##   left son=10 (51 obs) right son=11 (47 obs)
##   Primary splits:
##       atemp      < 0.315586  to the left,  improve=0.4036293, (0 missing)
```

```
##       weathersit splits as  RLL, improve=0.0477719, (0 missing)
##   Surrogate splits:
##       weathersit splits as  LRL, agree=0.551, adj=0.064, (0 split)
##
## Node number 6: 313 observations,    complexity param=0.01374834
##   mean=809.9553, MSE=99606.77
##   left son=12 (11 obs) right son=13 (302 obs)
##   Primary splits:
##       weathersit splits as  RRL, improve=0.1517666, (0 missing)
##       atemp      < 0.565142  to the left,  improve=0.1436688, (0 missing)
##
## Node number 7: 133 observations,    complexity param=0.01519603
##   mean=1935.722, MSE=421736.5
##   left son=14 (27 obs) right son=15 (106 obs)
##   Primary splits:
##       atemp      < 0.4873645 to the left,  improve=0.09323870, (0 missing)
##       weathersit splits as  RLL, improve=0.06469288, (0 missing)
##
## Node number 10: 51 observations
##   mean=334.2941, MSE=46393.42
##
## Node number 11: 47 observations
##   mean=898.5532, MSE=194471.4
##
## Node number 12: 11 observations
##   mean=165.7273, MSE=7065.835
##
## Node number 13: 302 observations
##   mean=833.4205, MSE=87309.87
##
## Node number 14: 27 observations
##   mean=1542.815, MSE=504887.5
##
## Node number 15: 106 observations
##   mean=2035.802, MSE=351218.3
```

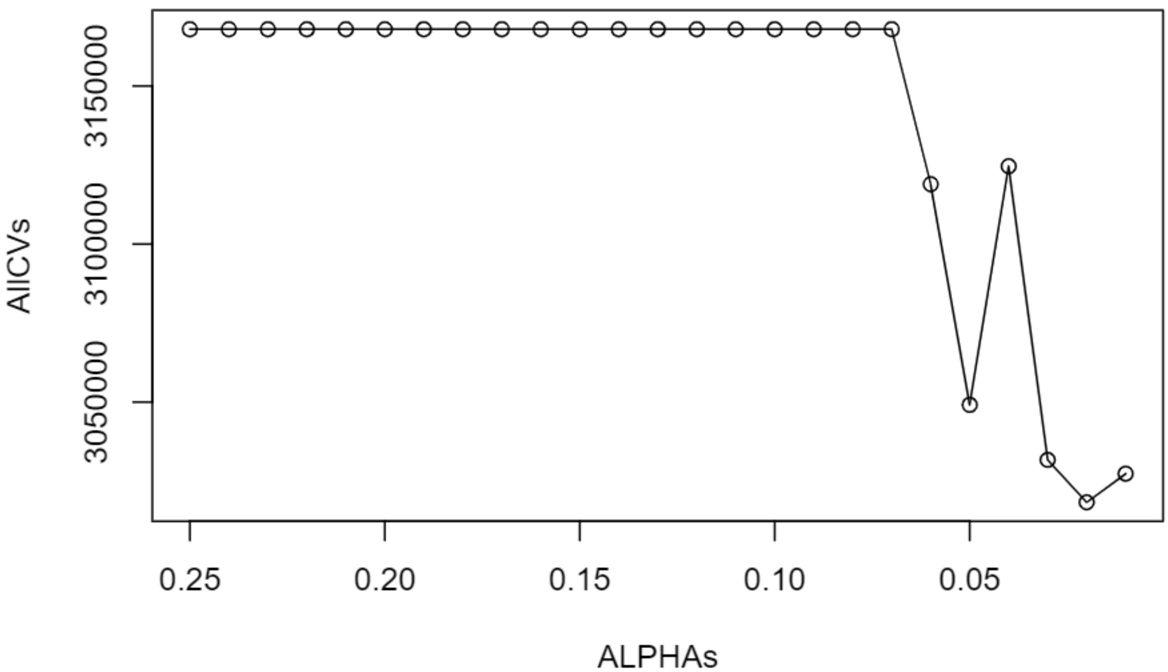
Pruned Tree for Casual Bikers

```r
#Complexity Pruning
N = length(data)
K = 10
ALPHAs= seq(0.25,0.01,-0.01)
# split into K folds
Kfolds = split(data,1:K)
```

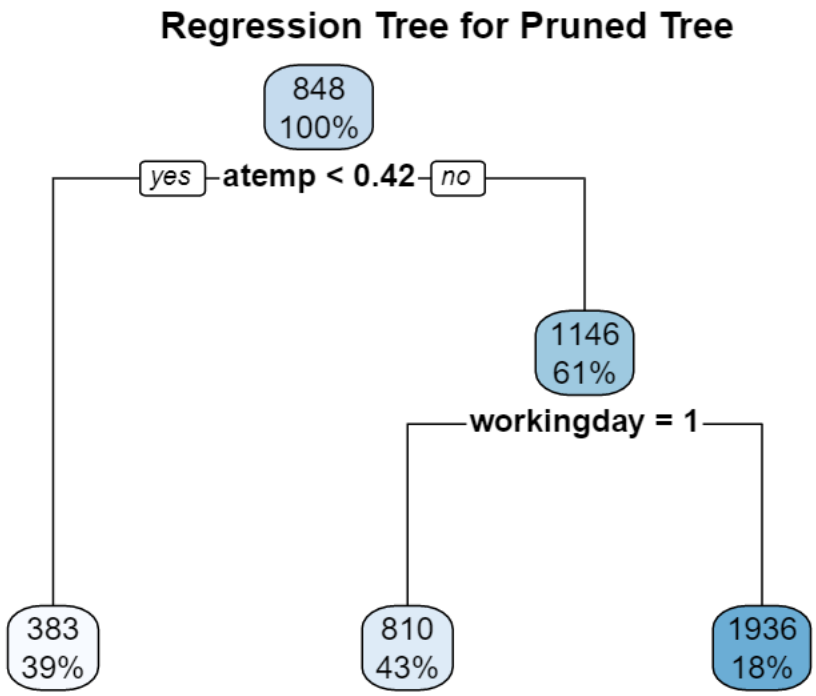
```
## Warning in split.default(x = seq_len(nrow(x)), f = f, drop = drop, ...):
## data length is not a multiple of split variable
```

```r
AllCVs = rep(0,length(ALPHAs))

i=1
for (ALPHA in ALPHAs){
    MSEs = rep(0,K)
```

```r
    for(k in 1:K){
        trainingIndices = setdiff(1:K,k)
        trainingData = do.call(rbind,Kfolds[trainingIndices])
        testingData  = Kfolds[[k]]

        BigTree <- rpart(registered~atemp + workingday + weathersit, method="anova", data=trainingData)
        smallerTree = prune(BigTree, cp = ALPHA)

        predictions = predict(smallerTree, testingData)
        MSEs[k] = t(testingData$cnt - predictions)%*%(testingData$cnt - predictions)/nrow(testingData)
    }
    AllCVs[i] = mean(MSEs)
    i=i+1
}

plot(ALPHAs,AllCVs, xlim=c(0.25,0.01) )
lines(ALPHAs,AllCVs)
```




```r
#Pruned Tree
prune.casual <- prune(casual.tree, cp=0.05)# from cp graph

# plot the pruned tree
rpart.plot(prune.casual, uniform=TRUE)
title("Regression Tree for Pruned Tree")
```



Out-of-Class Advanced Method

Poisson Regression for Total Bike Rentals

```r
count.pois <- glm(cnt ~ workingday + weathersit + atemp, family = poisson(link = "log"), data = data)
summary(count.pois)
```

```
##
## Call:
## glm(formula = cnt ~ workingday + weathersit + atemp, family = poisson(link = "log"),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -67.754  -18.890   -1.313   14.602   59.870
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  7.637644   0.002122 3599.58   <2e-16 ***
## workingday1  0.043052   0.001201   35.83   <2e-16 ***
## weathersit2 -0.125341   0.001215 -103.17   <2e-16 ***
## weathersit3 -0.858249   0.005195 -165.20   <2e-16 ***
## atemp        1.614471   0.003530  457.40   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 668801  on 730  degrees of freedom
```

```
## Residual deviance: 378122  on 726  degrees of freedom
## AIC: 385530
##
## Number of Fisher Scoring iterations: 4
```

```
exp(count.pois$coefficients)
```

```
##  (Intercept)  workingday1  weathersit2  weathersit3        atemp
## 2074.8491454    1.0439925    0.8821962    0.4239036    5.0252298
```

Calculation of Overdispersion for Total Bike Rentals

```
count.quasi <- glm(cnt ~ workingday + weathersit + atemp, family = quasipoisson (link = "log"), data = d
summary(count.quasi)
```

```
##
## Call:
## glm(formula = cnt ~ workingday + weathersit + atemp, family = quasipoisson(link = "log"),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -67.754  -18.890   -1.313   14.602   59.870
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.63764    0.04781 159.760  < 2e-16 ***
## workingday1  0.04305    0.02707   1.590    0.112
## weathersit2 -0.12534    0.02737  -4.579 5.50e-06 ***
## weathersit3 -0.85825    0.11705  -7.332 6.06e-13 ***
## atemp        1.61447    0.07953  20.301  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 507.6562)
##
##     Null deviance: 668801  on 730  degrees of freedom
## Residual deviance: 378122  on 726  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

Poisson Regression for Registered Bikers

```
registered.pois <- glm(registered ~ workingday + weathersit + atemp, family = poisson(link = "log"), dat
summary(registered.pois)
```

```
##
## Call:
## glm(formula = registered ~ workingday + weathersit + atemp, family = poisson(link = "log"),
##     data = data)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -57.898  -17.967   -0.972   13.970   44.650
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  7.393289   0.002369  3121.1   <2e-16 ***
## workingday1  0.285229   0.001406   202.9   <2e-16 ***
## weathersit2 -0.115268   0.001340   -86.0   <2e-16 ***
## weathersit3 -0.797439   0.005490  -145.3   <2e-16 ***
## atemp        1.337731   0.003894   343.5   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 526854  on 730  degrees of freedom
## Residual deviance: 309274  on 726  degrees of freedom
## AIC: 316535
##
## Number of Fisher Scoring iterations: 4
```

```r
exp(registered.pois$coefficients)
```

```
##  (Intercept)  workingday1  weathersit2  weathersit3        atemp
## 1625.0414109    1.3300666    0.8911271    0.4504812    3.8103868
```

Calculaiton of Overdispersion for Registered Bikers

```r
registered.quasi <- glm(registered ~ workingday + weathersit + atemp, family = quasipoisson (link = "lo
summary(registered.quasi)
```

```
##
## Call:
## glm(formula = registered ~ workingday + weathersit + atemp, family = quasipoisson(link = "log"),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -57.898  -17.967   -0.972   13.970   44.650
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.39329    0.04822 153.321  < 2e-16 ***
## workingday1  0.28523    0.02861   9.969  < 2e-16 ***
## weathersit2 -0.11527    0.02728  -4.225 2.70e-05 ***
## weathersit3 -0.79744    0.11175  -7.136 2.34e-12 ***
## atemp        1.33773    0.07928  16.874  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 414.3999)
```

```
##
##     Null deviance: 526854  on 730  degrees of freedom
## Residual deviance: 309274  on 726  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

Poisson Regression for Casual Bikers

```r
casual.poisson <- glm(casual ~ workingday + weathersit + atemp, family = poisson(link = "log"), data = d
summary(casual.poisson)
```

```
##
## Call:
## glm(formula = casual ~ workingday + weathersit + atemp, family = poisson(link = "log"),
##     data = data)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -49.30  -10.14   -2.82    6.72   54.73
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.863403   0.005045 1162.13   <2e-16 ***
## workingday1 -0.841787   0.002547 -330.54   <2e-16 ***
## weathersit2 -0.170872   0.002885  -59.22   <2e-16 ***
## weathersit3 -1.262686   0.016131  -78.28   <2e-16 ***
## atemp        2.831899   0.008427  336.04   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 378438  on 730  degrees of freedom
## Residual deviance: 130365  on 726  degrees of freedom
## AIC: 136359
##
## Number of Fisher Scoring iterations: 5
```

```r
exp(casual.poisson$coefficients)
```

```
## (Intercept) workingday1 weathersit2 weathersit3       atemp
## 351.9195468   0.4309397   0.8429296   0.2828931  16.9776627
```

Calculation of Overdispersion for Casual Bikers

```r
casual.quasi <- glm(casual ~ workingday + weathersit + atemp, family = quasipoisson (link = "log"), data
summary(casual.quasi)
```

```
##
## Call:
```

```
## glm(formula = casual ~ workingday + weathersit + atemp, family = quasipoisson(link = "log"),
##     data = data)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -49.30  -10.14   -2.82    6.72   54.73
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.86340    0.06911  84.839  < 2e-16 ***
## workingday1 -0.84179    0.03489 -24.130  < 2e-16 ***
## weathersit2 -0.17087    0.03952  -4.323 1.75e-05 ***
## weathersit3 -1.26269    0.22096  -5.715 1.61e-08 ***
## atemp        2.83190    0.11544  24.532  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 187.6388)
##
##     Null deviance: 378438  on 730  degrees of freedom
## Residual deviance: 130365  on 726  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```