# Capstone Data Logistic Regression - Predict Krummholz

*Tom Thorpe*

*July 25, 2018*

## Objective

Use Logistic regression to predict tree coverage.

```
# Include required libraries.

library(gsubfn)
```

```
## Loading required package: proto
```

```
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(ggplot2)
library(ggridges) # for easier viewing of sub-group distributions
library(ROCR)
```

```
## Loading required package: gplots

##
## Attaching package: 'gplots'

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

```
suppressMessages(library(latticeExtra, warn.conflicts = FALSE, quietly=TRUE))
#library(latticeExtra)

  curTime=Sys.time()
  print(paste("Forest Cover Logistic script started at",curTime))
```

```
## [1] "Forest Cover Logistic script started at 2018-08-12 18:37:05"
```

```
#Point to data. The forestcover_clean_full.csv is the cleaned data to be graphed.

calcROC <- 1
saveFileName="ForestCoverLogisticStats.csv"

infile="C:/Users/Tom/git/datasciencefoundation/ForestCoverage/forestcover_clean_full.csv"
#infile="C:/Users/Tom/git/datasciencefoundation/ForestCoverage/forestcover_clean.csv"
#infile="C:/Users/Tom/git/datasciencefoundation/ForestCoverage/forestcoversmall_clean_full.csv"
```

```r
#infile="C:/Users/Tom/git/datasciencefoundation/ForestCoverage/forestcoversmall_clean.csv"
out2file="C:/Users/Tom/git/datasciencefoundation/ForestCoverage/forestcover_graph.csv"
#out1file="C:/Users/Tom/git/datasciencefoundation/ForestCoverage/forestcoversmall_clean_full.csv"
#out2file="C:/Users/Tom/git/datasciencefoundation/ForestCoverage/forestcoversmall_clean.csv"

alphaVal<-0.05 # large data
#alphaVal<-0.1  # small data
```

```r
forestcover <- read.csv(infile,header=TRUE,sep=",") %>% tbl_df()
  curTime=Sys.time()
  print(paste("Forest Cover data load completed at",curTime))
```

```
## [1] "Forest Cover data load completed at 2018-08-12 18:37:45"
```

```r
forestcover$SoilType<-as.factor(forestcover$SoilType)
forestcover$ClimateZone<-as.factor(forestcover$ClimateZone)
forestcover$GeoZone<-as.factor(forestcover$GeoZone)

 #glimpse(forestcover)

 #knitr::knit_exit()

# table(forestcover$Sed_mix)
```

A table showing the number of occurrences for each tree type is shown below.

```r
covCount<-data.frame(table(forestcover$CovName))
totCount<-nrow(forestcover)
covCount <- mutate(covCount,Percent = as.integer(covCount$Freq*1000/totCount)/10)
LodgePct<-covCount$Percent[covCount$Var1=="Lodgepole"]
SpruceFirPct<-covCount$Percent[covCount$Var1=="Spruce&Fir"]
LodgeAndSpruceFir<-LodgePct+SpruceFirPct
#```
#```{r echo=TRUE}
covCount
```

```
##            Var1   Freq Percent
## 1         Aspen   9493     1.6
## 2 Cotton&Willow   2747     0.4
## 3    DouglasFir  17367     2.9
## 4     Krummholz  20510     3.5
## 5     Lodgepole 283301    48.7
## 6     Ponderosa  35754     6.1
## 7    Spruce&Fir 211840    36.4
```

Lodge pole Pine represents 48.7 percent of the sample. So always guessing “Lodge pole” would provide success rate of 48.7 percent and can be used as a baseline for comparing our predictions. Krummholz represent the next largest number of trees. The two together represent 85.1 percent.

## Logistic Model Accuracy Function

A function to help determine threshold for best accuracy and testing is shown below.

```r
source("logisticAccuracy.R") # for function calcLogisticModelAccuracy
#save("calcLogisticModelAccuracy", file="logisticAccuracy.Rdata")
```

```
bestThreshIndex=11
```

# Create Training and Testing Sets

Split data into training and testing data for logistic regression. The split is based on cover type so that the different coverage types will be split proportionately for all cover types in the training and test sets.

```
library(caTools)
set.seed(127)
split = sample.split(forestcover$CovType, 0.70) # we want 65% in the training set
forestTrain = subset(forestcover, split == TRUE)
forestTest  = subset(forestcover, split == FALSE)
```

Check training set coverage percentages and compare with test set to ensure there is a representative amount of data in each set for each coverage type.

### View Training Set Coverage Percentages

Check training set coverage percentages.

```
covCount<-data.frame(table(forestTrain$CovName))
totCount<-nrow(forestTrain)
covCount <- mutate(covCount,Percent = as.integer(covCount$Freq*1000/totCount)/10)
covCount
```

```
##             Var1   Freq Percent
## 1          Aspen   6645     1.6
## 2 Cotton&Willow   1923     0.4
## 3    DouglasFir  12157     2.9
## 4      Krummholz  14357     3.5
## 5      Lodgepole 198311    48.7
## 6      Ponderosa  25028     6.1
## 7    Spruce&Fir 148288    36.4
```

### View Test Set Coverage Percentages

Check test set coverage percentages.

```
covCount<-data.frame(table(forestTest$CovName))
totCount<-nrow(forestTest)
covCount <- mutate(covCount,Percent = as.integer(covCount$Freq*1000/totCount)/10)
covCount
```

```
##             Var1  Freq Percent
## 1          Aspen  2848     1.6
## 2 Cotton&Willow   824     0.4
## 3    DouglasFir  5210     2.9
## 4      Krummholz  6153     3.5
## 5      Lodgepole 84990    48.7
## 6      Ponderosa 10726     6.1
## 7    Spruce&Fir 63552    36.4
```

```
# knitr::knit_exit() # exit early

#glimpse(forestTrain)
```

```
#glimpse(forestTest)
#summary(forestTrain)
#summary(forestTest)
#table(forestTrain$Sed_mix)
#table(forestTrain$GeoName)
#table(forestTrain$Krummholz)
#table(forestTest$Krummholz)

# the above all work without error.

#table(forestTest$Rock_Land)
# Get the following error with above code:
#  Error in table(SpfFir_test$Rock_Land) : object 'SpfFir_test' not found
#    Calls: <Anonymous> ... withCallingHandlers -> withVisible -> eval -> eval -> table


#table(forestTrain$Rock_Land)
#table(forestTest$Rock_Land)
#table(forestTrain$Rubbly)
#table(forestTest$Rubbly)

#table(forestTrain$Sed_mix)
#table(forestTrain$Gateview)
#table(forestTrain$Rubbly)
#table(forestTest$Sed_mix)
#table(forestTest$Gateview)
#table(forestTest$Rubbly)

############# Start Start Start Start Start Start Start Start #################
```

# Krummholz Logistic Regression

Logistic regression models are created and compared for the Krummholz coverage type. The outcome is based on the binary 'Krummholz' variable.

## Krummholz Logistic Regression - All Variables

### Create Krummholz Logistic Model - All Vars

Create the Krummholz logistic model for the Aggregated Soil data using all independent variables.

### Krummholz All Aggregated Soil Types

The original project used aggregated Soil Types. Compute a logistic regression model using the aggregated soil types to see how the dis-aggregated / individuated variables compare.

```
  # You can remove the levels of the factor variables using the option exclude:
  #   lm(dependent ~ factor(independent1, exclude=c('b','d')) + independent2)
  #   This way the factors b, d will not be included in the regression.
```

```r
curTime=Sys.time()
print(paste("Krummholz aggregated Logistic Model Calculation started at",curTime))
```

```
## [1] "Krummholz aggregated Logistic Model Calculation started at 2018-08-12 18:37:47"
```

```r
Krumm_Agg_LogMod =
  glm(Krummholz ~
        Elev +      # Elevation in meters of data cell
        Aspect +    # Direction in degrees slope faces
        Slope +     # Slope / steepness of hill in degrees (0 to 90)
        H2OHD +     # Horizontal distance in meters to nearest water
        H2OVD +     # Vertical distance in meters to nearest water
        RoadHD +    # Horizontal distance in meters to nearest road
        FirePtHD + # Horizontal distance in meters to nearest fire point
        Shade9AM + Shade12PM + Shade3PM + # Amount of shade at 9am, 12pm and 3pm
        # Wilderness areas:
          RWwild + NEwild + CMwild + CPwild +
        # Aggregated Soil type:
          ST01 + ST02 + ST03 + ST04 + ST05 + ST06 + ST07 + ST08 + ST09 + ST10 +
          ST11 + ST12 + ST13 + ST14 + ST15 + ST16 + ST17 + ST18 + ST19 + ST20 +
          ST21 + ST22 + ST23 + ST24 + ST25 + ST26 + ST27 + ST28 + ST29 + ST30 +
          ST31 + ST32 + ST33 + ST34 + ST35 + ST36 + ST37 + ST38 + ST39 + ST40 ,
        data=forestTrain, family=binomial)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
# save model for later use
Krumm_Agg_All_LogMod = Krumm_Agg_LogMod
save("Krumm_Agg_All_LogMod", file="Krumm_Agg_All_LogMod.Rdata")

Krumm_Agg_All_aic<-as.integer(Krumm_Agg_LogMod$aic)
Krumm_Agg_All_aic
```

```
## [1] 46631
```

```r
curTime=Sys.time()
print(paste("Krummholz aggregated Logistic Model Calculation completed at",curTime))
```

```
## [1] "Krummholz aggregated Logistic Model Calculation completed at 2018-08-12 18:40:08"
```

Check the coefficients for the Krummholz model using all aggregated data.

```r
summary(Krumm_Agg_LogMod)
```

```
##
## Call:
## glm(formula = Krummholz ~ Elev + Aspect + Slope + H2OHD + H2OVD +
##     RoadHD + FirePtHD + Shade9AM + Shade12PM + Shade3PM + RWwild +
##     NEwild + CMwild + CPwild + ST01 + ST02 + ST03 + ST04 + ST05 +
##     ST06 + ST07 + ST08 + ST09 + ST10 + ST11 + ST12 + ST13 + ST14 +
##     ST15 + ST16 + ST17 + ST18 + ST19 + ST20 + ST21 + ST22 + ST23 +
##     ST24 + ST25 + ST26 + ST27 + ST28 + ST29 + ST30 + ST31 + ST32 +
##     ST33 + ST34 + ST35 + ST36 + ST37 + ST38 + ST39 + ST40, family = binomial,
##     data = forestTrain)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -2.4896  -0.0865  -0.0275  -0.0010   4.8812
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.036e+09  4.118e+11   0.003 0.997994
## Elev         1.559e-02  1.765e-04  88.317  < 2e-16 ***
## Aspect       6.539e-04  1.622e-04   4.032 5.53e-05 ***
## Slope       -4.283e-02  3.406e-03 -12.575  < 2e-16 ***
## H2OHD       -2.161e-03  7.510e-05 -28.781  < 2e-16 ***
## H2OVD       -1.577e-03  2.846e-04  -5.541 3.00e-08 ***
## RoadHD      -1.448e-04  1.225e-05 -11.821  < 2e-16 ***
## FirePtHD     3.587e-04  1.265e-05  28.355  < 2e-16 ***
## Shade9AM    -1.377e-02  2.774e-03  -4.964 6.89e-07 ***
## Shade12PM    1.079e-02  2.337e-03   4.619 3.86e-06 ***
## Shade3PM    -2.316e-02  2.203e-03 -10.512  < 2e-16 ***
## RWwild      -1.036e+09  4.118e+11  -0.003 0.997994
## NEwild      -1.036e+09  4.118e+11  -0.003 0.997994
## CMwild      -1.036e+09  4.118e+11  -0.003 0.997994
## CPwild      -1.036e+09  4.118e+11  -0.003 0.997994
## ST01         1.629e+00  5.200e+02   0.003 0.997500
## ST02        -1.090e+01  3.332e+02  -0.033 0.973907
## ST03        -9.607e+00  3.960e+02  -0.024 0.980646
## ST04         3.569e+00  1.740e-01  20.511  < 2e-16 ***
## ST05         1.090e+00  6.839e+02   0.002 0.998728
## ST06        -5.131e-01  4.193e+02  -0.001 0.999024
## ST07        -1.357e+01  3.440e+03  -0.004 0.996852
## ST08        -1.379e+01  2.514e+03  -0.005 0.995623
## ST09        -8.067e+00  1.015e+03  -0.008 0.993659
## ST10        -1.226e+01  1.554e+02  -0.079 0.937096
## ST11        -1.375e+01  2.819e+02  -0.049 0.961109
## ST12        -1.311e+01  1.801e+02  -0.073 0.941990
## ST13        -3.739e+00  4.538e-01  -8.240  < 2e-16 ***
## ST14        -6.211e+00  9.995e+02  -0.006 0.995041
## ST15         5.023e+00  1.687e+04   0.000 0.999762
## ST16        -1.477e+01  4.804e+02  -0.031 0.975474
## ST17        -1.375e+01  4.939e+02  -0.028 0.977792
## ST18        -1.307e+01  5.402e+02  -0.024 0.980691
## ST19        -3.548e+00  1.003e+00  -3.538 0.000403 ***
## ST20        -1.539e+01  2.948e+02  -0.052 0.958379
## ST21        -2.455e+00  3.082e-01  -7.966 1.63e-15 ***
## ST22        -2.689e+00  1.154e-01 -23.303  < 2e-16 ***
## ST23        -1.775e+00  7.511e-02 -23.630  < 2e-16 ***
## ST24        -2.096e+00  1.031e-01 -20.326  < 2e-16 ***
## ST25        -1.930e+01  1.542e+03  -0.013 0.990014
## ST26        -1.638e+01  6.043e+02  -0.027 0.978380
## ST27        -1.392e+00  2.260e-01  -6.158 7.37e-10 ***
## ST28        -1.271e+01  1.036e+03  -0.012 0.990213
## ST29         2.707e-01  6.966e-02   3.885 0.000102 ***
## ST30         6.333e-01  1.057e-01   5.993 2.06e-09 ***
## ST31        -2.723e+00  9.826e-02 -27.710  < 2e-16 ***
## ST32        -2.781e+00  6.475e-02 -42.957  < 2e-16 ***
## ST33        -2.383e+00  7.550e-02 -31.562  < 2e-16 ***
## ST34        -2.384e-01  1.861e-01  -1.281 0.200068
## ST35         3.633e-01  7.788e-02   4.665 3.09e-06 ***
```

```
## ST36         5.263e-01  2.306e-01   2.282 0.022486 *
## ST37         2.234e+01  1.934e+03   0.012 0.990782
## ST38         7.618e-01  4.828e-02  15.781  < 2e-16 ***
## ST39         1.042e+00  4.853e-02  21.477  < 2e-16 ***
## ST40                NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 124217  on 406708  degrees of freedom
## Residual deviance:  46523  on 406655  degrees of freedom
## AIC: 46631
##
## Number of Fisher Scoring iterations: 20
```

WOW! The intercept is huge and listed as not significant. Wilderness area and several soil types are not significant and can be removed in the next iteration.

### Krummholz All Individuated Soil Types

Create a logistic model using the Individuated variables that were derived from the Soil Types. The Soil Type was the intersection of climate zone, geology zone, soil families, and rock content. These variables are used instead of the Soil types.

```r
curTime=Sys.time()
print(paste("Krummholz Individual Logistic Model Calculation started at",curTime))
```

```
## [1] "Krummholz Individual Logistic Model Calculation started at 2018-08-12 18:40:09"
```

```r
Krumm_Ind_LogMod =
  glm(Krummholz ~
        Elev +      # Elevation in meters of cell
        Aspect +    # Direction in degrees slope faces
        Slope +     # Slope / steepness of hill in degrees (0 to 90)
        H2OHD +     # Horizontal distance in meters to nearest water
        H2OVD +     # Vertical distance in meters to nearest water
        RoadHD +    # Horizontal distance in meters to nearest road
        FirePtHD + # Horizontal distance in meters to nearest fire point
        Shade9AM + Shade12PM + Shade3PM + # Amount of shade at 9am, 12pm and 3pm
        # Wilderness areas:
          RWwild + NEwild + CMwild + CPwild +
        # Climate Zone:
        # ClimateName +
          Montane_low + Montane + SubAlpine + Alpine + Dry + Non_Dry +
        # Geology Zone:
        # GeoName +
          Alluvium + Glacial + Sed_mix + Ign_Meta +
        # Soil Family:
          Aquolis_cmplx + Argiborolis_Pachic + Borohemists_cmplx + Bross +
          Bullwark + Bullwark_Cmplx + Catamount + Catamount_cmplx +
          Cathedral + Como + Cryaquepts_cmplx + Cryaquepts_Typic + Cryaquolls +
          Cryaquolls_cmplx + Cryaquolls_Typic + Cryaquolls_Typic_cmplx +
          Cryoborolis_cmplx + Cryorthents + Cryorthents_cmplx + Cryumbrepts +
          Cryumbrepts_cmplx + Gateview + Gothic + Granile + Haploborolis +
```

```r
           Legault + Legault_cmplx + Leighcan + Leighcan_cmplx + Leighcan_warm +
           Moran + Ratake + Ratake_cmplx + Rogert + Supervisor_Limber_cmplx +
           Troutville + Unspecified + Vanet + Wetmore +
         # Soil Rock composition:
           Bouldery_ext + Rock_Land + Rock_Land_cmplx + Rock_Outcrop +
           Rock_Outcrop_cmplx + Rubbly + Stony + Stony_extreme + Stony_very +
           Till_Substratum ,
         data=forestTrain, family=binomial)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
# save model for later use
Krumm_Ind_All_LogMod = Krumm_Ind_LogMod
save("Krumm_Ind_All_LogMod", file="Krumm_Ind_All_LogMod.Rdata")

#table(forestTrain$GeoName)
#table(forestTrain$Sed_mix)
#table(forestTrain$Gateview)
# above: Error in table(SpfFir_test$Gateview) : object 'SpfFir_train' not found <-------

Krumm_Ind_All_aic<-as.integer(Krumm_Ind_LogMod$aic)
Krumm_Ind_All_aic
```

```
## [1] 46642
```

```r
summary(Krumm_Ind_LogMod)
```

```
##
## Call:
## glm(formula = Krummholz ~ Elev + Aspect + Slope + H2OHD + H2OVD +
##     RoadHD + FirePtHD + Shade9AM + Shade12PM + Shade3PM + RWwild +
##     NEwild + CMwild + CPwild + Montane_low + Montane + SubAlpine +
##     Alpine + Dry + Non_Dry + Alluvium + Glacial + Sed_mix + Ign_Meta +
##     Aquolis_cmplx + Argiborolis_Pachic + Borohemists_cmplx +
##     Bross + Bullwark + Bullwark_Cmplx + Catamount + Catamount_cmplx +
##     Cathedral + Como + Cryaquepts_cmplx + Cryaquepts_Typic +
##     Cryaquolls + Cryaquolls_cmplx + Cryaquolls_Typic + Cryaquolls_Typic_cmplx +
##     Cryoborolis_cmplx + Cryorthents + Cryorthents_cmplx + Cryumbrepts +
##     Cryumbrepts_cmplx + Gateview + Gothic + Granile + Haploborolis +
##     Legault + Legault_cmplx + Leighcan + Leighcan_cmplx + Leighcan_warm +
##     Moran + Ratake + Ratake_cmplx + Rogert + Supervisor_Limber_cmplx +
##     Troutville + Unspecified + Vanet + Wetmore + Bouldery_ext +
##     Rock_Land + Rock_Land_cmplx + Rock_Outcrop + Rock_Outcrop_cmplx +
##     Rubbly + Stony + Stony_extreme + Stony_very + Till_Substratum,
##     family = binomial, data = forestTrain)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4891  -0.0864  -0.0276  -0.0013   4.8825
##
## Coefficients: (15 not defined because of singularities)
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)             1.010e+11  2.876e+12   0.035 0.971980
## Elev                    1.560e-02  2.123e-04  73.475  < 2e-16 ***
## Aspect                  6.523e-04  1.683e-04   3.876 0.000106 ***
```

```
## Slope                    -4.285e-02  3.486e-03 -12.290  < 2e-16 ***
## H2OHD                    -2.160e-03  7.880e-05 -27.412  < 2e-16 ***
## H2OVD                    -1.576e-03  2.976e-04  -5.297 1.18e-07 ***
## RoadHD                   -1.448e-04  1.247e-05 -11.610  < 2e-16 ***
## FirePtHD                  3.586e-04  1.368e-05  26.208  < 2e-16 ***
## Shade9AM                 -1.375e-02  2.798e-03  -4.916 8.82e-07 ***
## Shade12PM                 1.077e-02  2.370e-03   4.546 5.47e-06 ***
## Shade3PM                 -2.314e-02  2.221e-03 -10.420  < 2e-16 ***
## RWwild                    1.467e+10  1.020e+12   0.014 0.988521
## NEwild                    1.467e+10  1.020e+12   0.014 0.988521
## CMwild                    1.467e+10  1.020e+12   0.014 0.988521
## CPwild                    1.467e+10  1.020e+12   0.014 0.988521
## Montane_low              -1.707e+11  8.264e+12  -0.021 0.983522
## Montane                  -1.296e+11  1.313e+13  -0.010 0.992128
## SubAlpine                -1.157e+11  2.365e+12  -0.049 0.960986
## Alpine                   -1.157e+11  2.365e+12  -0.049 0.960986
## Dry                       3.564e+11  1.602e+13   0.022 0.982247
## Non_Dry                  -4.674e+10  5.361e+12  -0.009 0.993043
## Alluvium                 -2.958e+11  6.621e+12  -0.045 0.964371
## Glacial                   5.112e+11  6.420e+12   0.080 0.936543
## Sed_mix                  -3.425e+11  8.674e+12  -0.039 0.968503
## Ign_Meta                         NA         NA      NA       NA
## Aquolis_cmplx            -1.218e+01  3.156e+04   0.000 0.999692
## Argiborolis_Pachic               NA         NA      NA       NA
## Borohemists_cmplx        -5.112e+11  6.420e+12  -0.080 0.936543
## Bross                     1.980e+00  5.314e+02   0.004 0.997027
## Bullwark                  6.065e+10  1.413e+13   0.004 0.996575
## Bullwark_Cmplx            6.065e+10  1.413e+13   0.004 0.996575
## Catamount                -3.527e+00  1.197e+00  -2.947 0.003205 **
## Catamount_cmplx          -6.242e-01  1.231e-01  -5.072 3.94e-07 ***
## Cathedral                 1.018e+11  1.144e+13   0.009 0.992903
## Como                      2.326e+00  5.314e+02   0.004 0.996508
## Cryaquepts_cmplx          2.180e+00  5.314e+02   0.004 0.996727
## Cryaquepts_Typic          2.958e+11  6.621e+12   0.045 0.964371
## Cryaquolls                4.721e+11  1.708e+13   0.028 0.977947
## Cryaquolls_cmplx          2.622e+00  2.076e-01  12.628  < 2e-16 ***
## Cryaquolls_Typic          8.069e+11  1.191e+13   0.068 0.945998
## Cryaquolls_Typic_cmplx    9.139e-01  1.168e-01   7.827 4.99e-15 ***
## Cryoborolis_cmplx                NA         NA      NA       NA
## Cryorthents               5.825e-02  1.211e+00   0.048 0.961638
## Cryorthents_cmplx         2.516e+01  3.231e+03   0.008 0.993787
## Cryumbrepts                      NA         NA      NA       NA
## Cryumbrepts_cmplx                NA         NA      NA       NA
## Gateview                  4.721e+11  1.708e+13   0.028 0.977947
## Gothic                    2.354e-01  8.629e+03   0.000 0.999978
## Granile                  -1.470e+01  1.129e+03  -0.013 0.989611
## Haploborolis              1.018e+11  1.144e+13   0.009 0.992903
## Legault                   6.065e+10  1.413e+13   0.004 0.996575
## Legault_cmplx                    NA         NA      NA       NA
## Leighcan                 -4.359e-02  5.314e+02   0.000 0.999935
## Leighcan_cmplx            2.801e+00  5.314e+02   0.005 0.995795
## Leighcan_warm            -2.597e+00  5.314e+02  -0.005 0.996101
## Moran                            NA         NA      NA       NA
## Ratake                    1.018e+11  1.144e+13   0.009 0.992903
```

```
## Ratake_cmplx             -9.750e+00  1.239e+03  -0.008 0.993723
## Rogert                    1.763e+11  1.507e+13   0.012 0.990662
## Supervisor_Limber_cmplx          NA         NA      NA       NA
## Troutville               -4.505e+11  1.490e+13  -0.030 0.975884
## Unspecified                      NA         NA      NA       NA
## Vanet                     1.018e+11  1.144e+13   0.009 0.992903
## Wetmore                  -1.604e+00  1.255e+03  -0.001 0.998980
## Bouldery_ext                     NA         NA      NA       NA
## Rock_Land                 3.628e-01  9.973e-02   3.638 0.000275 ***
## Rock_Land_cmplx           1.758e+00  5.314e+02   0.003 0.997360
## Rock_Outcrop                     NA         NA      NA       NA
## Rock_Outcrop_cmplx        3.243e+00  1.195e+00   2.714 0.006652 **
## Rubbly                           NA         NA      NA       NA
## Stony                            NA         NA      NA       NA
## Stony_extreme                    NA         NA      NA       NA
## Stony_very                       NA         NA      NA       NA
## Till_Substratum          -5.112e+11  6.420e+12  -0.080 0.936543
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 124217  on 406708  degrees of freedom
## Residual deviance:  46525  on 406650  degrees of freedom
## AIC: 46643
##
## Number of Fisher Scoring iterations: 21
```

```r
  curTime=Sys.time()
  print(paste("Krummholz Individual Logistic Model Calculation completed at",curTime))
```

```
## [1] "Krummholz Individual Logistic Model Calculation completed at 2018-08-12 18:45:12"
```

```r
  #table(forestTest$Rock_Land)
  # Get the following error with above code:
  #  Error in table(SpfFir_test$Rock_Land) : object 'SpfFir_test' not found
  #    Calls: <Anonymous> ... withCallingHandlers -> withVisible -> eval -> eval -> table
```

#### Predict Krummholz Logistic Model Probabilities - All Aggregated Vars

#### Krummholz Probabilities - All Aggregated Data

Predict the probability of Krummholz for aggregated Data - all variables.

```r
# Predict Krummholz Agg Data - all variables

  Krumm_Agg_Train_predict= predict(Krumm_Agg_LogMod, type="response")
  Krumm_Agg_Train_Logit= predict(Krumm_Agg_LogMod)
  summary(Krumm_Agg_Train_predict)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000000 0.0000087 0.0005160 0.0353008 0.0052917 1.0000000
```

```r
  str(Krumm_Agg_Train_predict)
```

```
##  Named num [1:406709] 1.28e-05 1.33e-05 2.85e-10 1.60e-05 9.33e-06 ...
```

```
##  - attr(*, "names")= chr [1:406709] "1" "2" "3" "4" ...
```

```
  #plot(table(Krumm_Agg_Train_predict))
  #plot(table(Krumm_Agg_Train_Logit))
  dens<-data.frame(table(Krumm_Agg_Train_predict))
# str(dens)

  Krumm_Agg_Test_predict= predict(Krumm_Agg_LogMod, type="response",newdata=forestTest)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type =
## ifelse(type == : prediction from a rank-deficient fit may be misleading
```

```
  summary(Krumm_Agg_Test_predict)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000000 0.0000091 0.0005237 0.0355914 0.0053854 1.0000000
```

```
   str(Krumm_Agg_Test_predict)
```

```
##  Named num [1:174303] 1.75e-04 1.54e-05 2.05e-11 5.55e-12 9.00e-13 ...
##  - attr(*, "names")= chr [1:174303] "1" "2" "3" "4" ...
```

#### Krummholz Probabilities - All Individuated Data

Predict the probability of Krummholz for Individual Data - all variables.

```
  Krumm_Ind_Train_predict= predict(Krumm_Ind_LogMod, type="response")
  summary(Krumm_Ind_Train_predict)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000000 0.0000095 0.0005190 0.0352936 0.0052854 1.0000000
```

```
  Krumm_Ind_Test_predict= predict(Krumm_Ind_LogMod, type="response",newdata=forestTest)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type =
## ifelse(type == : prediction from a rank-deficient fit may be misleading
```

```
  summary(Krumm_Ind_Test_predict)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000000 0.0000099 0.0005264 0.0355843 0.0053767 1.0000000
```

#### Krummholz Receiver Operating Characteristic (ROC) - All Vars

#### Krummholz Receiver ROC - All Aggregated Data

Next, look at the True Positive and False Positive rates based on threshold value for the aggregated data.

```
  if (calcROC) {
    curTime=Sys.time()
    print(paste("ROC graph 1 started at",curTime))

    ROCpred_Krumm_Agg = prediction(Krumm_Agg_Train_predict, forestTrain$Krummholz)
    summary(ROCpred_Krumm_Agg)
    ROCperf_Krumm_Agg = performance(ROCpred_Krumm_Agg, "tpr", "fpr")
    summary(ROCperf_Krumm_Agg)

    Krumm_Agg_All_ROC_AUC = as.numeric(performance(ROCpred_Krumm_Agg, "auc")@y.values)
```

```
    Krumm_Agg_All_ROC_AUC=as.integer(as.numeric(Krumm_Agg_All_ROC_AUC)*1000)/10
    print(paste("Krumm_Agg_All_ROC_AUC=",Krumm_Agg_All_ROC_AUC))

    jpeg(filename="Fig-ROCR_perf_Krumm_Agg.jpg")
    plot(ROCperf_Krumm_Agg, colorize=TRUE, print.cutoffs.at=seq(0,1,0.1), text.adj=c(-0.2,1.7))
    dev.off()
  } else {
    Krumm_Agg_All_ROC_AUC = 84.2
  }
```

```
## [1] "ROC graph 1 started at 2018-08-12 18:45:18"
## [1] "Krumm_Agg_All_ROC_AUC= 98"
```

```
## pdf
##   2
```

### Krummholz Receiver ROC - All Individuated Data

The Response Operating Curve for the individuated data is shown below.

```
  if (calcROC) {
    curTime=Sys.time()
    print(paste("ROCR graph 2 started at",curTime))

    ROCpred_Krumm_Ind = prediction(Krumm_Ind_Train_predict, forestTrain$Krummholz)
    summary(ROCpred_Krumm_Ind)
    ROCperf_Krumm_Ind = performance(ROCpred_Krumm_Ind, "tpr", "fpr")
    summary(ROCperf_Krumm_Ind)

    Krumm_Ind_All_ROC_AUC = as.numeric(performance(ROCpred_Krumm_Ind, "auc")@y.values)
    Krumm_Ind_All_ROC_AUC=as.integer(as.numeric(Krumm_Ind_All_ROC_AUC)*1000)/10
    print(paste("Krumm_Ind_All_ROC_AUC=",Krumm_Ind_All_ROC_AUC))

    jpeg(filename="Fig-ROCR_perf_Krumm_Ind.jpg")
    plot(ROCperf_Krumm_Ind, colorize=TRUE, print.cutoffs.at=seq(0,1,0.1), text.adj=c(-0.2,1.7))
    dev.off()
  } else {
    Krumm_Ind_All_ROC_AUC = 84.2
  }
```

```
## [1] "ROCR graph 2 started at 2018-08-12 18:48:03"
## [1] "Krumm_Ind_All_ROC_AUC= 98"
```

```
## pdf
##   2
```

The threshold graphs are essentially identical. This is making me think that there is not much difference between the two models. The AIC score for the Soil Type model is AIC: 351676 and for the individuated variables is: AIC: 351839. The Soil type model AIC score is 0.046% better than the individuated model.

```
  curTime=Sys.time()
  print(paste("ROCR graph 2 completed at",curTime))
```

```
## [1] "ROCR graph 2 completed at 2018-08-12 18:50:33"
```

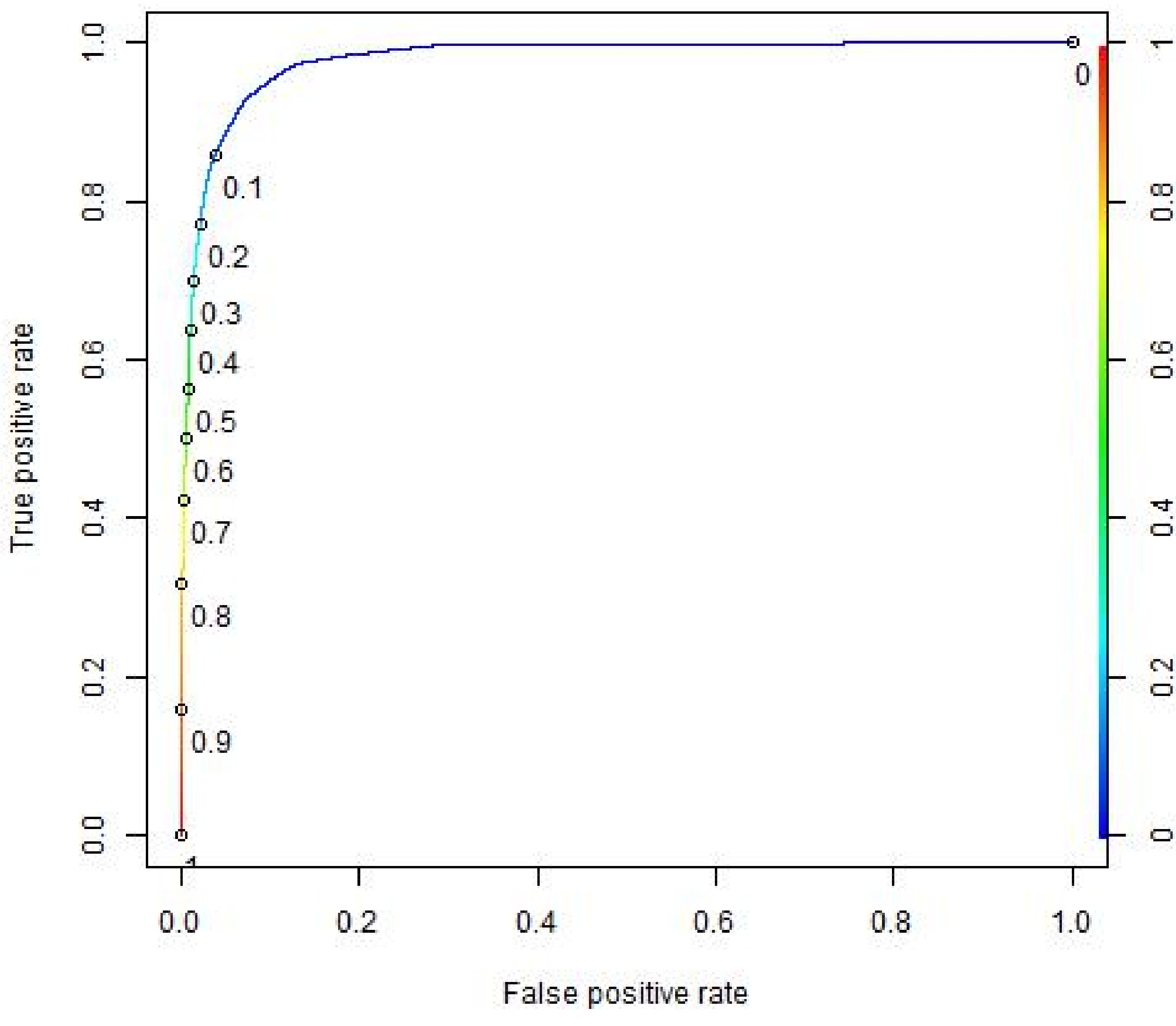


Figure 1: Krummholz ROC for All Aggregated Data

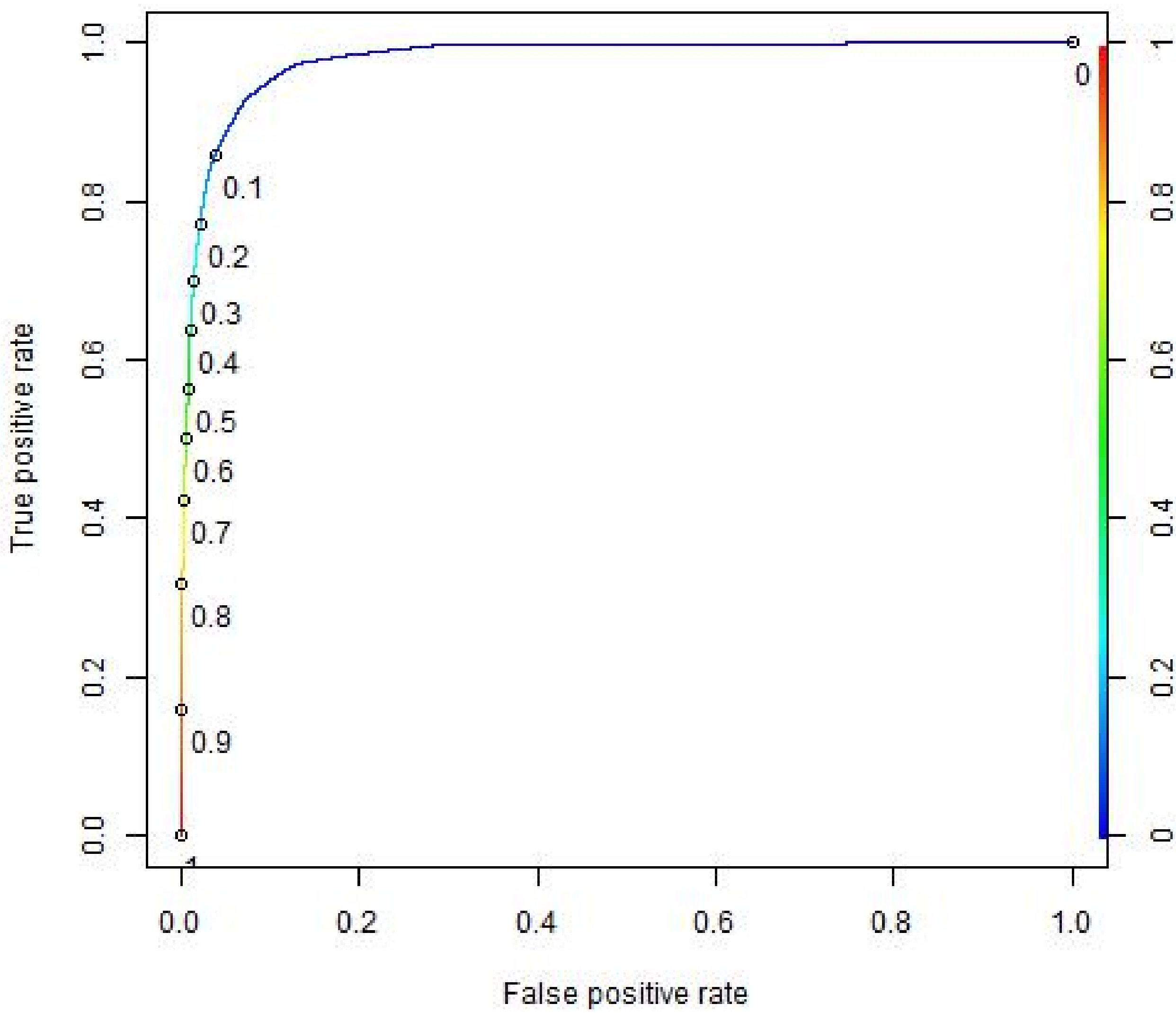


Figure 2: Krummholz ROC for All Individuated Data

### Calculate Accuracy of Krummholz Logisitic Models - All Vars

#### Calculate Krummholz Aggregated Data Logisitic Model Accuracy - All Vars

Find best threshold for Krummholz using all aggregated data.

```
result = calcLogisticModelAccuracy (forestTrain$Krummholz, Krumm_Agg_Train_predict,
                        0.0, 1, 10, "Krummholz", "Other", 1,1)
```

```
## [1] "Searching for threshold producing best Sensitivity_Specificity"
## [1] "start= 0 end= 1 inc= 0.1"
## [1] "Thresh=0, Accuracy=3.5%, BaseAcc(Other)=96.4%, Sens=100%, Spec=0%, Sens^2+Spec^2=-2"
## [1] "Thresh=0.1, Accuracy=95.8%, BaseAcc(Other)=96.4%, Sens=85.7%, Spec=96.2%, Sens^2+Spec^2=1.662"
## [1] "Thresh=0.2, Accuracy=97.1%, BaseAcc(Other)=96.4%, Sens=77.2%, Spec=97.8%, Sens^2+Spec^2=1.554"
## [1] "Thresh=0.3, Accuracy=97.6%, BaseAcc(Other)=96.4%, Sens=69.9%, Spec=98.6%, Sens^2+Spec^2=1.462"
## [1] "Thresh=0.4, Accuracy=97.7%, BaseAcc(Other)=96.4%, Sens=63.7%, Spec=99%, Sens^2+Spec^2=1.386"
## [1] "Thresh=0.5, Accuracy=97.7%, BaseAcc(Other)=96.4%, Sens=56.3%, Spec=99.2%, Sens^2+Spec^2=1.302"
## [1] "Thresh=0.6, Accuracy=97.7%, BaseAcc(Other)=96.4%, Sens=49.9%, Spec=99.4%, Sens^2+Spec^2=1.239"
## [1] "Thresh=0.7, Accuracy=97.6%, BaseAcc(Other)=96.4%, Sens=42.4%, Spec=99.7%, Sens^2+Spec^2=1.174"
## [1] "Thresh=0.8, Accuracy=97.4%, BaseAcc(Other)=96.4%, Sens=31.8%, Spec=99.8%, Sens^2+Spec^2=1.099"
## [1] "Thresh=0.9, Accuracy=96.9%, BaseAcc(Other)=96.4%, Sens=15.9%, Spec=99.9%, Sens^2+Spec^2=1.024"
## [1] "Thresh=1, Accuracy=96.4%, BaseAcc(Other)=96.4%, Sens=0%, Spec=100%, Sens^2+Spec^2=-2"
## [1] "Best Sensitivity_Specificity threshold= 0.1 inc= 0.1"
## [1] "========================================"
## [1] "start= 0 end= 0.2 inc= 0.01"
## [1] "Thresh=0, Accuracy=3.5%, BaseAcc(Other)=96.4%, Sens=100%, Spec=0%, Sens^2+Spec^2=-2"
## [1] "Thresh=0.01, Accuracy=83.8%, BaseAcc(Other)=96.4%, Sens=98%, Spec=83.3%, Sens^2+Spec^2=1.655"
## [1] "Thresh=0.02, Accuracy=88.9%, BaseAcc(Other)=96.4%, Sens=96.3%, Spec=88.6%, Sens^2+Spec^2=1.714"
## [1] "Thresh=0.03, Accuracy=91.2%, BaseAcc(Other)=96.4%, Sens=94.3%, Spec=91.1%, Sens^2+Spec^2=1.719"
## [1] "Thresh=0.04, Accuracy=92.6%, BaseAcc(Other)=96.4%, Sens=92.8%, Spec=92.6%, Sens^2+Spec^2=1.721"
## [1] "Thresh=0.05, Accuracy=93.6%, BaseAcc(Other)=96.4%, Sens=91%, Spec=93.7%, Sens^2+Spec^2=1.708"
## [1] "Thresh=0.06, Accuracy=94.3%, BaseAcc(Other)=96.4%, Sens=89.6%, Spec=94.5%, Sens^2+Spec^2=1.696"
## [1] "Thresh=0.07, Accuracy=94.9%, BaseAcc(Other)=96.4%, Sens=88.4%, Spec=95.1%, Sens^2+Spec^2=1.687"
## [1] "Thresh=0.08, Accuracy=95.3%, BaseAcc(Other)=96.4%, Sens=87.4%, Spec=95.6%, Sens^2+Spec^2=1.678"
## [1] "Thresh=0.09, Accuracy=95.6%, BaseAcc(Other)=96.4%, Sens=86.5%, Spec=95.9%, Sens^2+Spec^2=1.669"
## [1] "Thresh=0.1, Accuracy=95.8%, BaseAcc(Other)=96.4%, Sens=85.7%, Spec=96.2%, Sens^2+Spec^2=1.662"
## [1] "Thresh=0.11, Accuracy=96.1%, BaseAcc(Other)=96.4%, Sens=85.1%, Spec=96.5%, Sens^2+Spec^2=1.656"
## [1] "Thresh=0.12, Accuracy=96.2%, BaseAcc(Other)=96.4%, Sens=84.3%, Spec=96.7%, Sens^2+Spec^2=1.647"
## [1] "Thresh=0.13, Accuracy=96.4%, BaseAcc(Other)=96.4%, Sens=83.5%, Spec=96.9%, Sens^2+Spec^2=1.637"
## [1] "Thresh=0.14, Accuracy=96.5%, BaseAcc(Other)=96.4%, Sens=82.8%, Spec=97%, Sens^2+Spec^2=1.627"
## [1] "Thresh=0.15, Accuracy=96.6%, BaseAcc(Other)=96.4%, Sens=82%, Spec=97.2%, Sens^2+Spec^2=1.618"
## [1] "Thresh=0.16, Accuracy=96.7%, BaseAcc(Other)=96.4%, Sens=81.2%, Spec=97.3%, Sens^2+Spec^2=1.607"
## [1] "Thresh=0.17, Accuracy=96.9%, BaseAcc(Other)=96.4%, Sens=80.4%, Spec=97.5%, Sens^2+Spec^2=1.598"
## [1] "Thresh=0.18, Accuracy=96.9%, BaseAcc(Other)=96.4%, Sens=79.5%, Spec=97.6%, Sens^2+Spec^2=1.585"
## [1] "Thresh=0.19, Accuracy=97%, BaseAcc(Other)=96.4%, Sens=78.5%, Spec=97.7%, Sens^2+Spec^2=1.571"
## [1] "Best Sensitivity_Specificity threshold= 0.04 inc= 0.01"
## [1] "========================================"
## [1] "start= 0.03 end= 0.05 inc= 0.001"
## [1] "Thresh=0.03, Accuracy=91.2%, BaseAcc(Other)=96.4%, Sens=94.3%, Spec=91.1%, Sens^2+Spec^2=1.719"
## [1] "Thresh=0.031, Accuracy=91.4%, BaseAcc(Other)=96.4%, Sens=94.1%, Spec=91.3%, Sens^2+Spec^2=1.72"
## [1] "Thresh=0.032, Accuracy=91.5%, BaseAcc(Other)=96.4%, Sens=94%, Spec=91.5%, Sens^2+Spec^2=1.721"
## [1] "Thresh=0.033, Accuracy=91.7%, BaseAcc(Other)=96.4%, Sens=93.8%, Spec=91.6%, Sens^2+Spec^2=1.722
## [1] "Thresh=0.034, Accuracy=91.9%, BaseAcc(Other)=96.4%, Sens=93.7%, Spec=91.8%, Sens^2+Spec^2=1.722
## [1] "Thresh=0.035, Accuracy=92%, BaseAcc(Other)=96.4%, Sens=93.6%, Spec=92%, Sens^2+Spec^2=1.722"
## [1] "Thresh=0.036, Accuracy=92.2%, BaseAcc(Other)=96.4%, Sens=93.4%, Spec=92.1%, Sens^2+Spec^2=1.722
```

```
## [1] "Thresh=0.037, Accuracy=92.3%, BaseAcc(Other)=96.4%, Sens=93.3%, Spec=92.2%, Sens^2+Spec^2=1.722
## [1] "Thresh=0.038, Accuracy=92.4%, BaseAcc(Other)=96.4%, Sens=93.1%, Spec=92.4%, Sens^2+Spec^2=1.721
## [1] "Thresh=0.039, Accuracy=92.5%, BaseAcc(Other)=96.4%, Sens=92.9%, Spec=92.5%, Sens^2+Spec^2=1.721
## [1] "Thresh=0.04, Accuracy=92.6%, BaseAcc(Other)=96.4%, Sens=92.8%, Spec=92.6%, Sens^2+Spec^2=1.721"
## [1] "Thresh=0.041, Accuracy=92.8%, BaseAcc(Other)=96.4%, Sens=92.7%, Spec=92.8%, Sens^2+Spec^2=1.722
## [1] "Thresh=0.042, Accuracy=92.9%, BaseAcc(Other)=96.4%, Sens=92.5%, Spec=92.9%, Sens^2+Spec^2=1.72"
## [1] "Thresh=0.043, Accuracy=93%, BaseAcc(Other)=96.4%, Sens=92.4%, Spec=93%, Sens^2+Spec^2=1.72"
## [1] "Thresh=0.044, Accuracy=93.1%, BaseAcc(Other)=96.4%, Sens=92.2%, Spec=93.1%, Sens^2+Spec^2=1.719
## [1] "Thresh=0.045, Accuracy=93.2%, BaseAcc(Other)=96.4%, Sens=92%, Spec=93.2%, Sens^2+Spec^2=1.717"
## [1] "Thresh=0.046, Accuracy=93.3%, BaseAcc(Other)=96.4%, Sens=91.9%, Spec=93.3%, Sens^2+Spec^2=1.716
## [1] "Thresh=0.047, Accuracy=93.4%, BaseAcc(Other)=96.4%, Sens=91.7%, Spec=93.4%, Sens^2+Spec^2=1.715
## [1] "Thresh=0.048, Accuracy=93.4%, BaseAcc(Other)=96.4%, Sens=91.5%, Spec=93.5%, Sens^2+Spec^2=1.713
## [1] "Thresh=0.049, Accuracy=93.5%, BaseAcc(Other)=96.4%, Sens=91.3%, Spec=93.6%, Sens^2+Spec^2=1.711
## [1] "========================================"
## [1] "Best Threshold=0.035"
## [1] "Best Sensitivity_Specificity=1.72284651400005"
```

```r
curThresh = as.numeric(result[bestThreshIndex])
Krumm_Agg_All_threshold = curThresh
```

The accuracy for the best threshold on the training set for Krummholz using all aggregated data is shown below.

```r
result = calcLogisticModelAccuracy (forestTrain$Krummholz, Krumm_Agg_Train_predict,
                       curThresh, curThresh, 1, "Krummholz", "Other", 3)
```

```
## [1] "Model Performance for threshold= 0.035"
## [1] "predicted performance="
##                        Predicted
## Actual                  FALSE=Predict:Other TRUE=Predict:Krummholz
##   0=Actual:Other          361015 (TN)         31337 (FP)
##   1=Actual:Krummholz      918 (FN)            13439 (TP)
## [1] "Sensitivity= 0.936059065264331 (True positive rate of Krummholz = TP/(TP+FN) = 13439 /( 13439 +
## [1] "Specificity= 0.920130393116385 (True negative rate of Other = TN/(TN+FP) = 361015 /( 361015 + 3
## [1] "Sens^2+Spec^2=1.722"
## [1] "Baseline (Other) Accuracy=0.964699"
## [1] "Logistic Accuracy=0.920692"
```

The accuracy for the best threshold on the testing set for Krummholz using all aggregated data is shown below.

```r
result = calcLogisticModelAccuracy (forestTest$Krummholz, Krumm_Agg_Test_predict,
                       curThresh, curThresh, 1, "Krummholz", "Other", 3,
                       saveFile=saveFileName, desc="Krummholz All Aggregate Vars",
                       AIC=Krumm_Agg_All_aic, AUC=Krumm_Agg_All_ROC_AUC)
```

```
## [1] "Model Performance for threshold= 0.035"
## [1] "predicted performance="
##                        Predicted
## Actual                  FALSE=Predict:Other TRUE=Predict:Krummholz
##   0=Actual:Other          154424 (TN)         13726 (FP)
##   1=Actual:Krummholz      398 (FN)            5755 (TP)
## [1] "Sensitivity= 0.93531610596457 (True positive rate of Krummholz = TP/(TP+FN) = 5755 /( 5755 + 39
## [1] "Specificity= 0.918370502527505 (True negative rate of Other = TN/(TN+FP) = 154424 /( 154424 + 1
## [1] "Sens^2+Spec^2=1.718"
## [1] "Baseline (Other) Accuracy=0.964699"
## [1] "Logistic Accuracy=0.918968"
```

```
# retVal = c(modelPerformance, sensitivity,specificity) # TN, FN, FP, TP, sens, spec
# c(funcStat,accuracy,baseline,retVal)
list[RC, Krumm_Agg_All_model_acc, Krumm_Agg_All_baseline_acc,
    TN, FN, FP, TP, Krumm_Agg_All_sens, Krumm_Agg_All_spec] <- result
if (RC != "OK") {
  print(paste("Error - terminating:",RC))
  knitr:knit_exit()
}
Krumm_Agg_All_model_acc = as.integer(as.numeric(Krumm_Agg_All_model_acc)*1000)/10
Krumm_Agg_All_baseline_acc = as.integer(as.numeric(Krumm_Agg_All_baseline_acc)*1000)/10
Krumm_Agg_All_sens = as.integer(as.numeric(Krumm_Agg_All_sens)*1000)/10
Krumm_Agg_All_spec = as.integer(as.numeric(Krumm_Agg_All_spec)*1000)/10
```

#### Calculate Krummholz Individuated Data Logisitic Model Accuracy - All Vars

Find best threshold for Krummholz using all individuated data.

```
result = calcLogisticModelAccuracy (forestTrain$Krummholz, Krumm_Ind_Train_predict,
                        0.0, 1, 10, "Krummholz", "Other", 1,1)
```

```
## [1] "Searching for threshold producing best Sensitivity_Specificity"
## [1] "start= 0 end= 1 inc= 0.1"
## [1] "Thresh=0, Accuracy=3.5%, BaseAcc(Other)=96.4%, Sens=100%, Spec=0%, Sens^2+Spec^2=-2"
## [1] "Thresh=0.1, Accuracy=95.8%, BaseAcc(Other)=96.4%, Sens=85.7%, Spec=96.2%, Sens^2+Spec^2=1.662"
## [1] "Thresh=0.2, Accuracy=97.1%, BaseAcc(Other)=96.4%, Sens=77.2%, Spec=97.8%, Sens^2+Spec^2=1.554"
## [1] "Thresh=0.3, Accuracy=97.6%, BaseAcc(Other)=96.4%, Sens=69.9%, Spec=98.6%, Sens^2+Spec^2=1.462"
## [1] "Thresh=0.4, Accuracy=97.7%, BaseAcc(Other)=96.4%, Sens=63.7%, Spec=99%, Sens^2+Spec^2=1.386"
## [1] "Thresh=0.5, Accuracy=97.7%, BaseAcc(Other)=96.4%, Sens=56.3%, Spec=99.2%, Sens^2+Spec^2=1.302"
## [1] "Thresh=0.6, Accuracy=97.7%, BaseAcc(Other)=96.4%, Sens=49.9%, Spec=99.4%, Sens^2+Spec^2=1.239"
## [1] "Thresh=0.7, Accuracy=97.6%, BaseAcc(Other)=96.4%, Sens=42.4%, Spec=99.7%, Sens^2+Spec^2=1.174"
## [1] "Thresh=0.8, Accuracy=97.4%, BaseAcc(Other)=96.4%, Sens=31.8%, Spec=99.8%, Sens^2+Spec^2=1.099"
## [1] "Thresh=0.9, Accuracy=96.9%, BaseAcc(Other)=96.4%, Sens=15.9%, Spec=99.9%, Sens^2+Spec^2=1.024"
## [1] "Thresh=1, Accuracy=96.4%, BaseAcc(Other)=96.4%, Sens=0%, Spec=100%, Sens^2+Spec^2=-2"
## [1] "Best Sensitivity_Specificity threshold= 0.1 inc= 0.1"
## [1] "========================================"
## [1] "start= 0 end= 0.2 inc= 0.01"
## [1] "Thresh=0, Accuracy=3.5%, BaseAcc(Other)=96.4%, Sens=100%, Spec=0%, Sens^2+Spec^2=-2"
## [1] "Thresh=0.01, Accuracy=83.8%, BaseAcc(Other)=96.4%, Sens=98%, Spec=83.3%, Sens^2+Spec^2=1.656"
## [1] "Thresh=0.02, Accuracy=88.9%, BaseAcc(Other)=96.4%, Sens=96.3%, Spec=88.6%, Sens^2+Spec^2=1.713"
## [1] "Thresh=0.03, Accuracy=91.2%, BaseAcc(Other)=96.4%, Sens=94.3%, Spec=91.1%, Sens^2+Spec^2=1.719"
## [1] "Thresh=0.04, Accuracy=92.7%, BaseAcc(Other)=96.4%, Sens=92.8%, Spec=92.6%, Sens^2+Spec^2=1.721"
## [1] "Thresh=0.05, Accuracy=93.6%, BaseAcc(Other)=96.4%, Sens=91%, Spec=93.7%, Sens^2+Spec^2=1.708"
## [1] "Thresh=0.06, Accuracy=94.3%, BaseAcc(Other)=96.4%, Sens=89.5%, Spec=94.5%, Sens^2+Spec^2=1.696"
## [1] "Thresh=0.07, Accuracy=94.9%, BaseAcc(Other)=96.4%, Sens=88.4%, Spec=95.1%, Sens^2+Spec^2=1.687"
## [1] "Thresh=0.08, Accuracy=95.3%, BaseAcc(Other)=96.4%, Sens=87.4%, Spec=95.6%, Sens^2+Spec^2=1.678"
## [1] "Thresh=0.09, Accuracy=95.6%, BaseAcc(Other)=96.4%, Sens=86.5%, Spec=95.9%, Sens^2+Spec^2=1.669"
## [1] "Thresh=0.1, Accuracy=95.8%, BaseAcc(Other)=96.4%, Sens=85.7%, Spec=96.2%, Sens^2+Spec^2=1.662"
## [1] "Thresh=0.11, Accuracy=96.1%, BaseAcc(Other)=96.4%, Sens=85.1%, Spec=96.5%, Sens^2+Spec^2=1.656"
## [1] "Thresh=0.12, Accuracy=96.2%, BaseAcc(Other)=96.4%, Sens=84.3%, Spec=96.7%, Sens^2+Spec^2=1.647"
## [1] "Thresh=0.13, Accuracy=96.4%, BaseAcc(Other)=96.4%, Sens=83.5%, Spec=96.9%, Sens^2+Spec^2=1.637"
## [1] "Thresh=0.14, Accuracy=96.5%, BaseAcc(Other)=96.4%, Sens=82.7%, Spec=97%, Sens^2+Spec^2=1.627"
## [1] "Thresh=0.15, Accuracy=96.6%, BaseAcc(Other)=96.4%, Sens=82%, Spec=97.2%, Sens^2+Spec^2=1.618"
## [1] "Thresh=0.16, Accuracy=96.8%, BaseAcc(Other)=96.4%, Sens=81.2%, Spec=97.3%, Sens^2+Spec^2=1.607"
## [1] "Thresh=0.17, Accuracy=96.9%, BaseAcc(Other)=96.4%, Sens=80.4%, Spec=97.5%, Sens^2+Spec^2=1.597"
```

```
## [1] "Thresh=0.18, Accuracy=96.9%, BaseAcc(Other)=96.4%, Sens=79.5%, Spec=97.6%, Sens^2+Spec^2=1.585"
## [1] "Thresh=0.19, Accuracy=97%, BaseAcc(Other)=96.4%, Sens=78.5%, Spec=97.7%, Sens^2+Spec^2=1.572"
## [1] "Best Sensitivity_Specificity threshold= 0.04 inc= 0.01"
## [1] "========================================"
## [1] "start= 0.03 end= 0.05 inc= 0.001"
## [1] "Thresh=0.03, Accuracy=91.2%, BaseAcc(Other)=96.4%, Sens=94.3%, Spec=91.1%, Sens^2+Spec^2=1.719"
## [1] "Thresh=0.031, Accuracy=91.4%, BaseAcc(Other)=96.4%, Sens=94.1%, Spec=91.3%, Sens^2+Spec^2=1.72"
## [1] "Thresh=0.032, Accuracy=91.6%, BaseAcc(Other)=96.4%, Sens=94%, Spec=91.5%, Sens^2+Spec^2=1.721"
## [1] "Thresh=0.033, Accuracy=91.7%, BaseAcc(Other)=96.4%, Sens=93.8%, Spec=91.6%, Sens^2+Spec^2=1.722
## [1] "Thresh=0.034, Accuracy=91.9%, BaseAcc(Other)=96.4%, Sens=93.7%, Spec=91.8%, Sens^2+Spec^2=1.722
## [1] "Thresh=0.035, Accuracy=92%, BaseAcc(Other)=96.4%, Sens=93.5%, Spec=92%, Sens^2+Spec^2=1.722"
## [1] "Thresh=0.036, Accuracy=92.2%, BaseAcc(Other)=96.4%, Sens=93.4%, Spec=92.1%, Sens^2+Spec^2=1.722
## [1] "Thresh=0.037, Accuracy=92.3%, BaseAcc(Other)=96.4%, Sens=93.2%, Spec=92.2%, Sens^2+Spec^2=1.722
## [1] "Thresh=0.038, Accuracy=92.4%, BaseAcc(Other)=96.4%, Sens=93.1%, Spec=92.4%, Sens^2+Spec^2=1.721
## [1] "Thresh=0.039, Accuracy=92.5%, BaseAcc(Other)=96.4%, Sens=92.9%, Spec=92.5%, Sens^2+Spec^2=1.721
## [1] "Thresh=0.04, Accuracy=92.7%, BaseAcc(Other)=96.4%, Sens=92.8%, Spec=92.6%, Sens^2+Spec^2=1.721"
## [1] "Thresh=0.041, Accuracy=92.8%, BaseAcc(Other)=96.4%, Sens=92.7%, Spec=92.8%, Sens^2+Spec^2=1.722
## [1] "Thresh=0.042, Accuracy=92.9%, BaseAcc(Other)=96.4%, Sens=92.5%, Spec=92.9%, Sens^2+Spec^2=1.72"
## [1] "Thresh=0.043, Accuracy=93%, BaseAcc(Other)=96.4%, Sens=92.4%, Spec=93%, Sens^2+Spec^2=1.719"
## [1] "Thresh=0.044, Accuracy=93.1%, BaseAcc(Other)=96.4%, Sens=92.2%, Spec=93.1%, Sens^2+Spec^2=1.719
## [1] "Thresh=0.045, Accuracy=93.2%, BaseAcc(Other)=96.4%, Sens=92%, Spec=93.2%, Sens^2+Spec^2=1.717"
## [1] "Thresh=0.046, Accuracy=93.3%, BaseAcc(Other)=96.4%, Sens=91.9%, Spec=93.3%, Sens^2+Spec^2=1.716
## [1] "Thresh=0.047, Accuracy=93.4%, BaseAcc(Other)=96.4%, Sens=91.7%, Spec=93.4%, Sens^2+Spec^2=1.715
## [1] "Thresh=0.048, Accuracy=93.4%, BaseAcc(Other)=96.4%, Sens=91.5%, Spec=93.5%, Sens^2+Spec^2=1.713
## [1] "Thresh=0.049, Accuracy=93.5%, BaseAcc(Other)=96.4%, Sens=91.3%, Spec=93.6%, Sens^2+Spec^2=1.711
## [1] "========================================"
## [1] "Best Threshold=0.034"
## [1] "Best Sensitivity_Specificity=1.72269554146574"
```

```r
curThresh = as.numeric(result[bestThreshIndex])
Krumm_Ind_All_threshold = curThresh
```

The accuracy for the best threshold on the training set for Krummholz using all individuated data is shown below.

```r
result = calcLogisticModelAccuracy (forestTrain$Krummholz, Krumm_Ind_Train_predict,
                        curThresh, curThresh, 1, "Krummholz", "Other", 3)
```

```
## [1] "Model Performance for threshold= 0.034"
## [1] "predicted performance="
##                        Predicted
## Actual                  FALSE=Predict:Other TRUE=Predict:Krummholz
##   0=Actual:Other            360398 (TN)         31954 (FP)
##   1=Actual:Krummholz        897 (FN)            13460 (TP)
## [1] "Sensitivity= 0.937521766385735 (True positive rate of Krummholz = TP/(TP+FN) = 13460 /( 13460 +
## [1] "Specificity= 0.918557825625969 (True negative rate of Other = TN/(TN+FP) = 360398 /( 360398 + 3
## [1] "Sens^2+Spec^2=1.722"
## [1] "Baseline (Other) Accuracy=0.964699"
## [1] "Logistic Accuracy=0.919227"
```

The accuracy for the best threshold on the testing set for Krummholz using all individuated data is shown below.

```r
result = calcLogisticModelAccuracy (forestTest$Krummholz, Krumm_Ind_Test_predict,
                        curThresh, curThresh, 1, "Krummholz", "Other", 3,
                        saveFile=saveFileName, desc="Krummholz All Individualized Vars",
```

```
                        AIC=Krumm_Ind_All_aic, AUC=Krumm_Ind_All_ROC_AUC)
```

```
## [1] "Model Performance for threshold= 0.034"
## [1] "predicted performance="
##                       Predicted
## Actual                 FALSE=Predict:Other TRUE=Predict:Krummholz
##   0=Actual:Other           154148 (TN)         14002 (FP)
##   1=Actual:Krummholz       390 (FN)            5763 (TP)
## [1] "Sensitivity= 0.936616284739152 (True positive rate of Krummholz = TP/(TP+FN) = 5763 /( 5763 + 3
## [1] "Specificity= 0.916729110912875 (True negative rate of Other = TN/(TN+FP) = 154148 /( 154148 + 1
## [1] "Sens^2+Spec^2=1.717"
## [1] "Baseline (Other) Accuracy=0.964699"
## [1] "Logistic Accuracy=0.917431"
```

```
list[RC, Krumm_Ind_All_model_acc, Krumm_Ind_All_baseline_acc,
      TN, FN, FP, TP, Krumm_Ind_All_sens, Krumm_Ind_All_spec] <- result
  if (RC != "OK") {
    print(paste("Error - terminating:",RC))
    knitr:knit_exit()
  }
  Krumm_Ind_All_model_acc = as.integer(as.numeric(Krumm_Ind_All_model_acc)*1000)/10
  Krumm_Ind_All_baseline_acc = as.integer(as.numeric(Krumm_Ind_All_baseline_acc)*1000)/10
  Krumm_Ind_All_sens = as.integer(as.numeric(Krumm_Ind_All_sens)*1000)/10
  Krumm_Ind_All_spec = as.integer(as.numeric(Krumm_Ind_All_spec)*1000)/10
```

The Krummholz aggregated model accuracy on the test data is 77.15% compared to 77.12% for the individuated data model, essentially identical. Both are ~ 14% better than the baseline model.

## Krummholz Logistic Regression - Significant Variables

### Create Krummholz Logistic Model - Sig Vars

Now create the logistic model for the Aggregated Soil data using just the significant variables and compare to the previous models.

### Krummholz Logistic Model using Significant Aggregated Data

Variables that have been removed are commented out in the code below.

```
  Krumm_Agg_LogMod =
    glm(Krummholz ~
          Elev +     # Elevation in meters of cell
          Aspect +   # Direction in degrees slope faces
          Slope +    # Slope / steepness of hill in degrees (0 to 90)
          H2OHD +    # Horizontal distance in meters to nearest water
          H2OVD +    # Vertical distance in meters to nearest water
          # RoadHD +   # Horizontal distance in meters to nearest road # removed 2nd pass
          FirePtHD + # Horizontal distance in meters to nearest fire point
          Shade9AM + Shade12PM + Shade3PM + # Amount of shade at 9am, 12pm and 3pm
          # Wilderness areas:
            # RWwild + NEwild + CMwild + CPwild +
          # Aggregated Soil type:
            # ST01 + ST02 + ST03 +
            ST04 +
```

```r
            # ST05 + ST06 + ST07 +
            #ST08 + ST09 + ST10 + ST11 + ST12 +
            ST13 +
            # ST14 + ST15 +
            # ST16 + ST17 + ST18 +
            ST19 +
            # ST20 +
            # ST21 +  # removed 2nd pass
            ST22 + ST23 + ST24 +
            #ST25 + ST26 +
            # ST27 + # removed 2nd pass
            # ST28 +
            ST29 +
            # ST30 + # removed 2nd pass
            ST31 + ST32 + ST33 +
            # ST34 +
            ST35 +
            ST36 +
            # ST37 +
            ST38 + ST39 ,
            # + ST40 ,
          data=forestTrain, family=binomial)

# save model for later use
Krumm_Agg_Sig_LogMod = Krumm_Agg_LogMod
save("Krumm_Agg_Sig_LogMod", file="Krumm_Agg_Sig_LogMod.Rdata")

Krumm_Agg_Sig_aic<-as.integer(Krumm_Agg_LogMod$aic)
Krumm_Agg_Sig_aic
```

```
## [1] 55786
```

Check the coefficients of the Krummholz model using significant aggregated data.

```r
summary(Krumm_Agg_LogMod)
```

```
##
## Call:
## glm(formula = Krummholz ~ Elev + Aspect + Slope + H2OHD + H2OVD +
##     FirePtHD + Shade9AM + Shade12PM + Shade3PM + ST04 + ST13 +
##     ST19 + ST22 + ST23 + ST24 + ST29 + ST31 + ST32 + ST33 + ST35 +
##     ST36 + ST38 + ST39, family = binomial, data = forestTrain)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3633  -0.1065  -0.0380  -0.0104   4.7337
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.764e+01  7.067e-01 -67.414  < 2e-16 ***
## Elev         1.438e-02  1.419e-04 101.303  < 2e-16 ***
## Aspect       1.586e-03  1.440e-04  11.018  < 2e-16 ***
## Slope       -2.996e-02  3.331e-03  -8.997  < 2e-16 ***
## H2OHD       -1.375e-03  6.503e-05 -21.149  < 2e-16 ***
## H2OVD       -1.232e-03  2.475e-04  -4.977 6.44e-07 ***
```

```
## FirePtHD     1.379e-04  1.045e-05  13.191  < 2e-16 ***
## Shade9AM    -5.414e-03  3.125e-03  -1.733   0.0832 .
## Shade12PM    9.646e-03  2.594e-03   3.719   0.0002 ***
## Shade3PM    -1.757e-02  2.515e-03  -6.985 2.85e-12 ***
## ST04         4.996e+00  1.606e-01  31.106  < 2e-16 ***
## ST13        -2.084e+00  4.502e-01  -4.630 3.65e-06 ***
## ST19        -3.954e+00  1.001e+00  -3.950 7.81e-05 ***
## ST22        -2.140e+00  1.081e-01 -19.795  < 2e-16 ***
## ST23        -7.108e-01  5.968e-02 -11.911  < 2e-16 ***
## ST24        -5.079e-01  9.476e-02  -5.360 8.33e-08 ***
## ST29        -5.591e-01  5.698e-02  -9.814  < 2e-16 ***
## ST31        -8.842e-01  9.045e-02  -9.776  < 2e-16 ***
## ST32        -1.034e+00  5.424e-02 -19.054  < 2e-16 ***
## ST33        -6.648e-01  6.253e-02 -10.631  < 2e-16 ***
## ST35         8.164e-01  6.899e-02  11.834  < 2e-16 ***
## ST36         1.850e+00  2.314e-01   7.995 1.29e-15 ***
## ST38         1.178e+00  3.807e-02  30.948  < 2e-16 ***
## ST39         1.543e+00  3.818e-02  40.396  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 124217  on 406708  degrees of freedom
## Residual deviance:  55739  on 406685  degrees of freedom
## AIC: 55787
##
## Number of Fisher Scoring iterations: 10
```

The intercept looks much more reasonable. Some soil types that were significant previously are no longer significant.

### Krummholz Logistic Model using Significant Individuated Data

Create a logistic model for the significant individuated variables.

Again, the non-significant variables have been commented out.

```
Krumm_Ind_LogMod =
  glm(Krummholz ~
        Elev +      # Elevation in meters of cell
        Aspect +    # Direction in degrees slope faces
        Slope +     # Slope / steepness of hill in degrees (0 to 90)
        H2OHD +     # Horizontal distance in meters to nearest water
        H2OVD +     # Vertical distance in meters to nearest water
        RoadHD +    # Horizontal distance in meters to nearest road
        FirePtHD + # Horizontal distance in meters to nearest fire point
        Shade9AM + Shade12PM + Shade3PM + # Amount of shade at 9am, 12pm and 3pm
        # Wilderness areas:
          # RWwild + NEwild + CMwild + CPwild +
        # Climate Zone:
        # ClimateName +
          # Montane_low + Montane +
          #SubAlpine + Alpine +
          # Dry + Non_Dry +
```

```r
          # Geology Zone:
          # GeoName +
            #Alluvium + Glacial +
            # Sed_mix + Ign_Meta +
          # Soil Family:
            #Aquolis_cmplx +
            # Argiborolis_Pachic +
            #Borohemists_cmplx + Bross +
            #Bullwark + Bullwark_Cmplx +
            Catamount + Catamount_cmplx +
            # Cathedral + Como +
            #Cryaquepts_cmplx + Cryaquepts_Typic + Cryaquolls +
            Cryaquolls_cmplx +
            # Cryaquolls_Typic +
            Cryaquolls_Typic_cmplx +
            # Cryoborolis_cmplx +
            #Cryorthents +
            # Cryorthents_cmplx + Cryumbrepts + Cryumbrepts_cmplx + Gateview +
            # Gothic + Granile + Haploborolis +
            #Legault +
            # Legault_cmplx +
            #Leighcan + Leighcan_cmplx + Leighcan_warm +
            # Moran + Ratake + Ratake_cmplx + Rogert + Supervisor_Limber_cmplx +
            # Troutville + Unspecified + Vanet + Wetmore +
          # Soil Rock composition:
            # Bouldery_ext +
            # Rock_Land +  # removed 2nd pass
            # Rock_Land_cmplx + Rock_Outcrop +
            Rock_Outcrop_cmplx ,
            # Rubbly + Stony + Stony_extreme + Stony_very + Till_Substratum ,
          data=forestTrain, family=binomial)

# save model for later use
Krumm_Ind_Sig_LogMod = Krumm_Ind_LogMod
save("Krumm_Ind_Sig_LogMod", file="Krumm_Ind_Sig_LogMod.Rdata")

Krumm_Ind_Sig_aic<-as.integer(Krumm_Ind_LogMod$aic)
Krumm_Ind_Sig_aic
```

```
## [1] 60618
```

```r
summary(Krumm_Ind_LogMod)
```

```
##
## Call:
## glm(formula = Krummholz ~ Elev + Aspect + Slope + H2OHD + H2OVD +
##     RoadHD + FirePtHD + Shade9AM + Shade12PM + Shade3PM + Catamount +
##     Catamount_cmplx + Cryaquolls_cmplx + Cryaquolls_Typic_cmplx +
##     Rock_Outcrop_cmplx, family = binomial, data = forestTrain)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6079  -0.1268  -0.0346  -0.0069   4.6917
##
```

```
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)    
## (Intercept)            -5.235e+01  6.487e-01 -80.702  < 2e-16 ***
## Elev                    1.621e-02  1.265e-04 128.121  < 2e-16 ***
## Aspect                  1.438e-03  1.392e-04  10.333  < 2e-16 ***
## Slope                  -1.087e-02  3.116e-03  -3.490 0.000483 ***
## H2OHD                  -1.488e-03  6.399e-05 -23.244  < 2e-16 ***
## H2OVD                  -1.952e-03  2.430e-04  -8.034 9.41e-16 ***
## RoadHD                 -3.341e-05  7.923e-06  -4.216 2.48e-05 ***
## FirePtHD                1.060e-04  9.941e-06  10.661  < 2e-16 ***
## Shade9AM               -1.057e-02  2.927e-03  -3.610 0.000306 ***
## Shade12PM               1.131e-02  2.436e-03   4.643 3.43e-06 ***
## Shade3PM               -2.000e-02  2.351e-03  -8.508  < 2e-16 ***
## Catamount              -1.398e+00  4.727e-02 -29.574  < 2e-16 ***
## Catamount_cmplx        -1.084e+00  8.627e-02 -12.570  < 2e-16 ***
## Cryaquolls_cmplx        9.015e-01  2.777e-02  32.461  < 2e-16 ***
## Cryaquolls_Typic_cmplx -8.374e-01  5.081e-02 -16.483  < 2e-16 ***
## Rock_Outcrop_cmplx      5.011e-01  6.602e-02   7.590 3.19e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 124217  on 406708  degrees of freedom
## Residual deviance:  60587  on 406693  degrees of freedom
## AIC: 60619
## 
## Number of Fisher Scoring iterations: 9
```

Again the intercept looks much better. Also a few variables have become non-significant.

#### Predict Krummholz Logistic Model Probabilities - Sig Vars

##### Krummholz Probabilities using Significant Aggregated Data

Predict the probability of Krummholz for aggregated Data - significant variables.

```r
# Predict Krummholz Agg Data - significant variables

  Krumm_Agg_Train_predict= predict(Krumm_Agg_LogMod, type="response")
  summary(Krumm_Agg_Train_predict)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
## 0.0000000 0.0000877 0.0009488 0.0353004 0.0077579 0.9964300
```

```r
  Krumm_Agg_Test_predict= predict(Krumm_Agg_LogMod, type="response",newdata=forestTest)
  summary(Krumm_Agg_Test_predict)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
## 0.0000000 0.0000882 0.0009531 0.0355783 0.0078629 0.9944804
```

##### Krummholz Probabilities using Significant Individuated Data

Predict the probability of Krummholz using significant Individuated Data.

```r
Krumm_Ind_Train_predict= predict(Krumm_Ind_LogMod, type="response")
summary(Krumm_Ind_Train_predict)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000000 0.0000424 0.0008317 0.0353004 0.0115607 0.9987682
```

```r
Krumm_Ind_Test_predict= predict(Krumm_Ind_LogMod, type="response",newdata=forestTest)
summary(Krumm_Ind_Test_predict)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000000 0.0000421 0.0008370 0.0357083 0.0116658 0.9981250
```

```r
print(paste("ROCR graph 2 completed at",curTime))
```

```
## [1] "ROCR graph 2 completed at 2018-08-12 18:50:33"
```

#### Krummholz Receiver Operating Characteristic (ROC) - Sig Vars

Look at the True Positive and False Positive rates based on threshold value.

```r
if (calcROC) {
  ROCpred_Krumm_Agg = prediction(Krumm_Agg_Train_predict, forestTrain$Krummholz)
  summary(ROCpred_Krumm_Agg)

  ROCperf_Krumm_Agg = performance(ROCpred_Krumm_Agg, "tpr", "fpr")
  summary(ROCperf_Krumm_Agg)

  Krumm_Agg_Sig_ROC_AUC = as.numeric(performance(ROCpred_Krumm_Agg, "auc")@y.values)
  Krumm_Agg_Sig_ROC_AUC=as.integer(as.numeric(Krumm_Agg_Sig_ROC_AUC)*1000)/10
  Krumm_Agg_Sig_ROC_AUC

  jpeg(filename="Fig-ROCR_perf_Krumm_Agg_Sig.jpg")
  plot(ROCperf_Krumm_Agg, colorize=TRUE, print.cutoffs.at=seq(0,1,0.1), text.adj=c(-0.2,1.7))
  dev.off()
} else {
  Krumm_Agg_Sig_ROC_AUC = 83.7
}
```

```
## pdf
##   2
```

```r
if (calcROC) {
  curTime=Sys.time()
  print(paste("ROCR graph 2 started at",curTime))

  ROCpred_Krumm_Ind = prediction(Krumm_Ind_Train_predict, forestTrain$Krummholz)
  summary(ROCpred_Krumm_Ind)

  ROCperf_Krumm_Ind = performance(ROCpred_Krumm_Ind, "tpr", "fpr")
  summary(ROCperf_Krumm_Ind)

  Krumm_Ind_Sig_ROC_AUC = as.numeric(performance(ROCpred_Krumm_Ind, "auc")@y.values)
  Krumm_Ind_Sig_ROC_AUC=as.integer(as.numeric(Krumm_Ind_Sig_ROC_AUC)*1000)/10
  Krumm_Ind_Sig_ROC_AUC

  jpeg(filename="Fig-ROC_perf_Krumm_Ind_Sig.jpg")
```

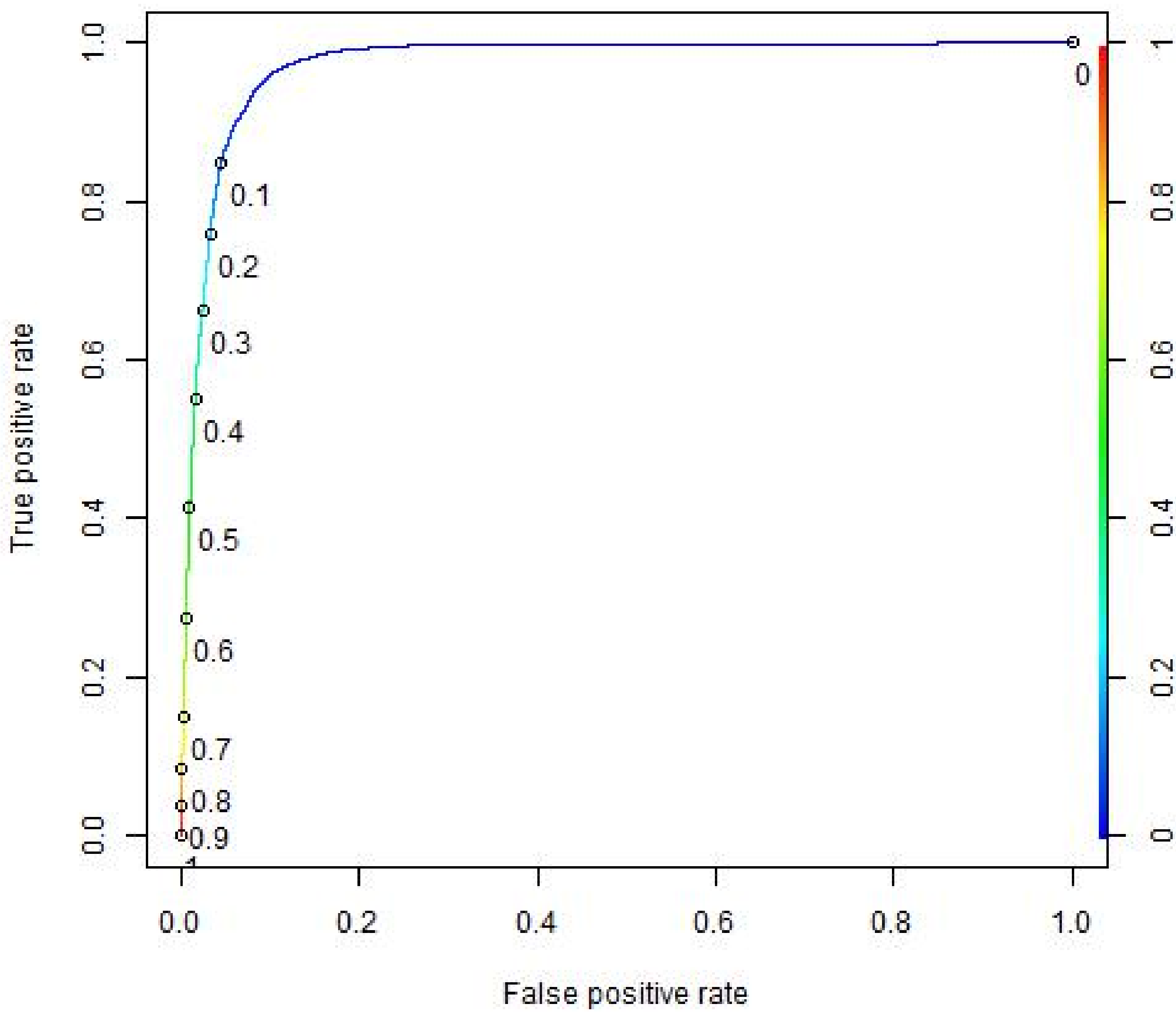


Figure 3: Krummholz ROC for Aggregated Significant Data

```
    plot(ROCperf_Krumm_Ind, colorize=TRUE, print.cutoffs.at=seq(0,1,0.1), text.adj=c(-0.2,1.7))
    dev.off()
  } else {
    Krumm_Ind_Sig_ROC_AUC = 83.8
  }
```

```
## [1] "ROCR graph 2 started at 2018-08-12 18:55:29"

## pdf
##   2
```

The threshold graphs are essentially identical. This is making me think that there is not much difference between the two models. The AIC score for the Soil Type model is AIC: 351676 and for the individuated variables is: AIC: 351839. The Soil type model AIC score is 0.046% better than the individuated model.

### Calculate Accuracy of Krummholz Logisitic Model - Sig Vars

#### Calculate Krummholz Aggregated Data Logisitic Model Accuracy - Significant Vars

Find best Krummholz threshold for Aggregated Data using significant variables.

```
result = calcLogisticModelAccuracy (forestTrain$Krummholz, Krumm_Agg_Train_predict,
                        0.0, 1, 10, "Krummholz", "Other", 1,1)
```

```
## [1] "Searching for threshold producing best Sensitivity_Specificity"
## [1] "start= 0 end= 1 inc= 0.1"
## [1] "Thresh=0, Accuracy=3.5%, BaseAcc(Other)=96.4%, Sens=100%, Spec=0%, Sens^2+Spec^2=-2"
## [1] "Thresh=0.1, Accuracy=95.1%, BaseAcc(Other)=96.4%, Sens=84.8%, Spec=95.5%, Sens^2+Spec^2=1.632"
## [1] "Thresh=0.2, Accuracy=96%, BaseAcc(Other)=96.4%, Sens=75.8%, Spec=96.8%, Sens^2+Spec^2=1.511"
## [1] "Thresh=0.3, Accuracy=96.5%, BaseAcc(Other)=96.4%, Sens=66.3%, Spec=97.6%, Sens^2+Spec^2=1.394"
## [1] "Thresh=0.4, Accuracy=96.9%, BaseAcc(Other)=96.4%, Sens=54.9%, Spec=98.4%, Sens^2+Spec^2=1.27"
## [1] "Thresh=0.5, Accuracy=97%, BaseAcc(Other)=96.4%, Sens=41.2%, Spec=99%, Sens^2+Spec^2=1.151"
## [1] "Thresh=0.6, Accuracy=96.9%, BaseAcc(Other)=96.4%, Sens=27.5%, Spec=99.4%, Sens^2+Spec^2=1.065"
## [1] "Thresh=0.7, Accuracy=96.7%, BaseAcc(Other)=96.4%, Sens=15%, Spec=99.7%, Sens^2+Spec^2=1.016"
## [1] "Thresh=0.8, Accuracy=96.6%, BaseAcc(Other)=96.4%, Sens=8.3%, Spec=99.9%, Sens^2+Spec^2=1.005"
## [1] "Thresh=0.9, Accuracy=96.6%, BaseAcc(Other)=96.4%, Sens=3.8%, Spec=99.9%, Sens^2+Spec^2=1.001"
## [1] "Thresh=1, Accuracy=96.4%, BaseAcc(Other)=96.4%, Sens=0%, Spec=100%, Sens^2+Spec^2=-2"
## [1] "Best Sensitivity_Specificity threshold= 0.1 inc= 0.1"
## [1] "========================================"
## [1] "start= 0 end= 0.2 inc= 0.01"
## [1] "Thresh=0, Accuracy=3.5%, BaseAcc(Other)=96.4%, Sens=100%, Spec=0%, Sens^2+Spec^2=-2"
## [1] "Thresh=0.01, Accuracy=81%, BaseAcc(Other)=96.4%, Sens=99.2%, Spec=80.4%, Sens^2+Spec^2=1.631"
## [1] "Thresh=0.02, Accuracy=87.6%, BaseAcc(Other)=96.4%, Sens=97.5%, Spec=87.2%, Sens^2+Spec^2=1.712"
## [1] "Thresh=0.03, Accuracy=90.5%, BaseAcc(Other)=96.4%, Sens=95.6%, Spec=90.3%, Sens^2+Spec^2=1.731"
## [1] "Thresh=0.04, Accuracy=92.1%, BaseAcc(Other)=96.4%, Sens=93.4%, Spec=92%, Sens^2+Spec^2=1.72"
## [1] "Thresh=0.05, Accuracy=93.1%, BaseAcc(Other)=96.4%, Sens=90.9%, Spec=93.2%, Sens^2+Spec^2=1.697"
## [1] "Thresh=0.06, Accuracy=93.9%, BaseAcc(Other)=96.4%, Sens=89.3%, Spec=94.1%, Sens^2+Spec^2=1.684"
## [1] "Thresh=0.07, Accuracy=94.4%, BaseAcc(Other)=96.4%, Sens=87.9%, Spec=94.6%, Sens^2+Spec^2=1.669"
## [1] "Thresh=0.08, Accuracy=94.7%, BaseAcc(Other)=96.4%, Sens=86.7%, Spec=95%, Sens^2+Spec^2=1.655"
## [1] "Thresh=0.09, Accuracy=94.9%, BaseAcc(Other)=96.4%, Sens=85.7%, Spec=95.3%, Sens^2+Spec^2=1.644"
## [1] "Thresh=0.1, Accuracy=95.1%, BaseAcc(Other)=96.4%, Sens=84.8%, Spec=95.5%, Sens^2+Spec^2=1.632"
## [1] "Thresh=0.11, Accuracy=95.3%, BaseAcc(Other)=96.4%, Sens=83.8%, Spec=95.7%, Sens^2+Spec^2=1.619"
## [1] "Thresh=0.12, Accuracy=95.4%, BaseAcc(Other)=96.4%, Sens=82.6%, Spec=95.9%, Sens^2+Spec^2=1.602"
## [1] "Thresh=0.13, Accuracy=95.5%, BaseAcc(Other)=96.4%, Sens=81.5%, Spec=96%, Sens^2+Spec^2=1.588"
## [1] "Thresh=0.14, Accuracy=95.6%, BaseAcc(Other)=96.4%, Sens=80.5%, Spec=96.2%, Sens^2+Spec^2=1.574"
```

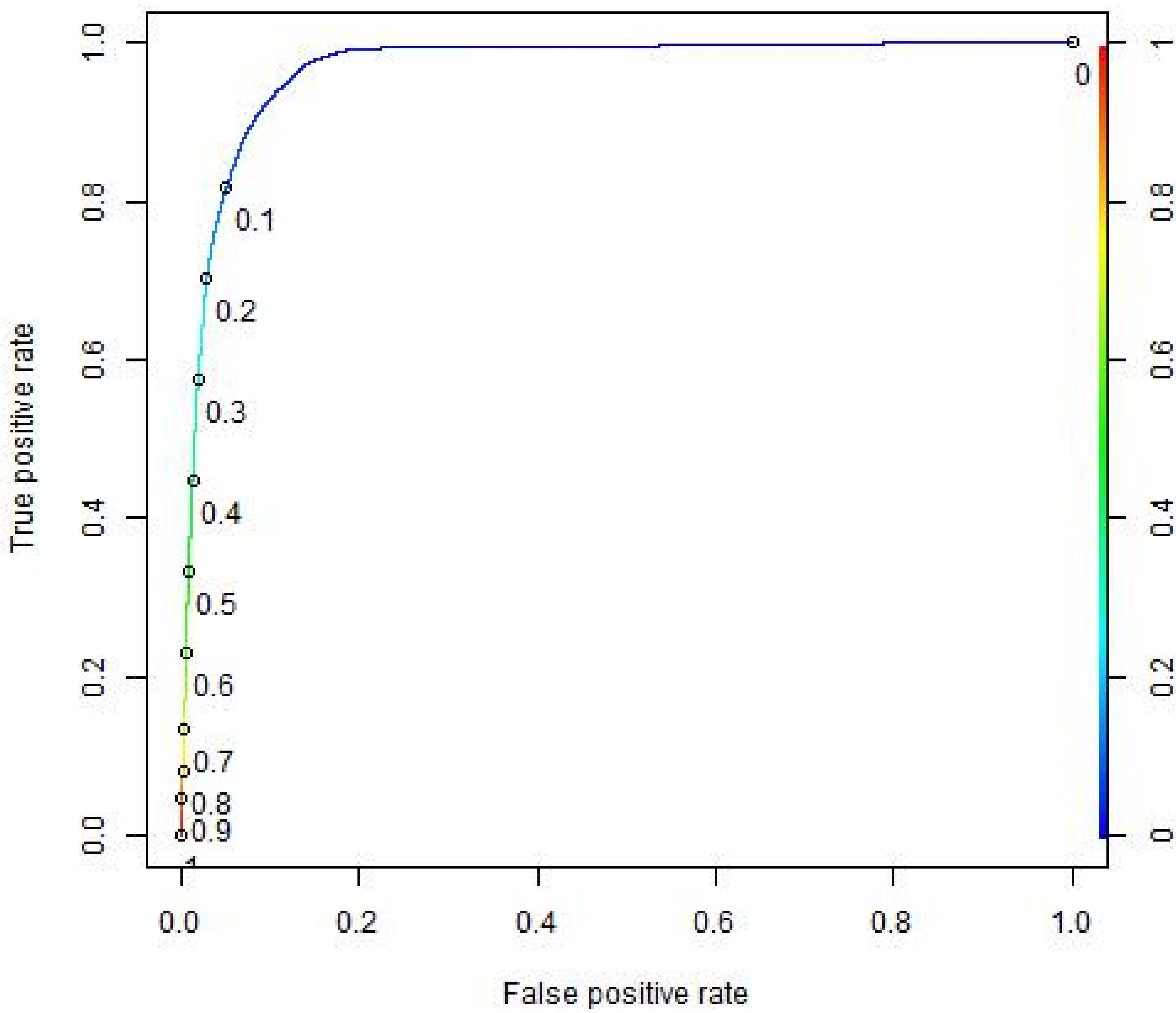


Figure 4: Krummholz ROC for Individuated Significant Data

```
## [1] "Thresh=0.15, Accuracy=95.7%, BaseAcc(Other)=96.4%, Sens=79.6%, Spec=96.3%, Sens^2+Spec^2=1.562"
## [1] "Thresh=0.16, Accuracy=95.8%, BaseAcc(Other)=96.4%, Sens=78.8%, Spec=96.4%, Sens^2+Spec^2=1.552"
## [1] "Thresh=0.17, Accuracy=95.8%, BaseAcc(Other)=96.4%, Sens=78%, Spec=96.5%, Sens^2+Spec^2=1.54"
## [1] "Thresh=0.18, Accuracy=95.9%, BaseAcc(Other)=96.4%, Sens=77.3%, Spec=96.6%, Sens^2+Spec^2=1.532"
## [1] "Thresh=0.19, Accuracy=96%, BaseAcc(Other)=96.4%, Sens=76.5%, Spec=96.7%, Sens^2+Spec^2=1.521"
## [1] "Best Sensitivity_Specificity threshold= 0.03 inc= 0.01"
## [1] "========================================"
## [1] "start= 0.02 end= 0.04 inc= 0.001"
## [1] "Thresh=0.02, Accuracy=87.6%, BaseAcc(Other)=96.4%, Sens=97.5%, Spec=87.2%, Sens^2+Spec^2=1.712"
## [1] "Thresh=0.021, Accuracy=88%, BaseAcc(Other)=96.4%, Sens=97.3%, Spec=87.6%, Sens^2+Spec^2=1.715"
## [1] "Thresh=0.022, Accuracy=88.3%, BaseAcc(Other)=96.4%, Sens=97.1%, Spec=88%, Sens^2+Spec^2=1.719"
## [1] "Thresh=0.023, Accuracy=88.7%, BaseAcc(Other)=96.4%, Sens=96.9%, Spec=88.4%, Sens^2+Spec^2=1.722
## [1] "Thresh=0.024, Accuracy=89%, BaseAcc(Other)=96.4%, Sens=96.7%, Spec=88.7%, Sens^2+Spec^2=1.724"
## [1] "Thresh=0.025, Accuracy=89.3%, BaseAcc(Other)=96.4%, Sens=96.6%, Spec=89%, Sens^2+Spec^2=1.727"
## [1] "Thresh=0.026, Accuracy=89.6%, BaseAcc(Other)=96.4%, Sens=96.4%, Spec=89.3%, Sens^2+Spec^2=1.728
## [1] "Thresh=0.027, Accuracy=89.8%, BaseAcc(Other)=96.4%, Sens=96.2%, Spec=89.6%, Sens^2+Spec^2=1.73"
## [1] "Thresh=0.028, Accuracy=90.1%, BaseAcc(Other)=96.4%, Sens=96.1%, Spec=89.8%, Sens^2+Spec^2=1.731
## [1] "Thresh=0.029, Accuracy=90.3%, BaseAcc(Other)=96.4%, Sens=95.8%, Spec=90.1%, Sens^2+Spec^2=1.731
## [1] "Thresh=0.03, Accuracy=90.5%, BaseAcc(Other)=96.4%, Sens=95.6%, Spec=90.3%, Sens^2+Spec^2=1.731"
## [1] "Thresh=0.031, Accuracy=90.7%, BaseAcc(Other)=96.4%, Sens=95.3%, Spec=90.5%, Sens^2+Spec^2=1.73"
## [1] "Thresh=0.032, Accuracy=90.8%, BaseAcc(Other)=96.4%, Sens=95.1%, Spec=90.7%, Sens^2+Spec^2=1.728
## [1] "Thresh=0.033, Accuracy=91%, BaseAcc(Other)=96.4%, Sens=94.9%, Spec=90.9%, Sens^2+Spec^2=1.727"
## [1] "Thresh=0.034, Accuracy=91.2%, BaseAcc(Other)=96.4%, Sens=94.6%, Spec=91.1%, Sens^2+Spec^2=1.726
## [1] "Thresh=0.035, Accuracy=91.3%, BaseAcc(Other)=96.4%, Sens=94.4%, Spec=91.2%, Sens^2+Spec^2=1.726
## [1] "Thresh=0.036, Accuracy=91.5%, BaseAcc(Other)=96.4%, Sens=94.2%, Spec=91.4%, Sens^2+Spec^2=1.725
## [1] "Thresh=0.037, Accuracy=91.6%, BaseAcc(Other)=96.4%, Sens=94.1%, Spec=91.5%, Sens^2+Spec^2=1.724
## [1] "Thresh=0.038, Accuracy=91.8%, BaseAcc(Other)=96.4%, Sens=93.8%, Spec=91.7%, Sens^2+Spec^2=1.722
## [1] "Thresh=0.039, Accuracy=91.9%, BaseAcc(Other)=96.4%, Sens=93.6%, Spec=91.8%, Sens^2+Spec^2=1.722
## [1] "========================================"
## [1] "Best Threshold=0.029"
## [1] "Best Sensitivity_Specificity=1.73185080788257"
```

```r
curThresh = as.numeric(result[bestThreshIndex])
Krumm_Agg_Sig_threshold = curThresh
```

The accuracy for the best threshold on the training set for Krummholz using significant aggregated data is shown below.

```r
result = calcLogisticModelAccuracy (forestTrain$Krummholz, Krumm_Agg_Train_predict,
                       curThresh, curThresh, 1, "Krummholz", "Other", 3)
```

```
## [1] "Model Performance for threshold= 0.029"
## [1] "predicted performance="
##                        Predicted
## Actual                  FALSE=Predict:Other TRUE=Predict:Krummholz
##   0=Actual:Other         353600 (TN)         38752 (FP)
##   1=Actual:Krummholz     589 (FN)            13768 (TP)
## [1] "Sensitivity= 0.95897471616633 (True positive rate of Krummholz = TP/(TP+FN) = 13768 /( 13768 + 5
## [1] "Specificity= 0.901231547182122 (True negative rate of Other = TN/(TN+FP) = 353600 /( 353600 + 38
## [1] "Sens^2+Spec^2=1.731"
## [1] "Baseline (Other) Accuracy=0.964699"
## [1] "Logistic Accuracy=0.903269"
```

The accuracy for the best threshold on the testing set for Krummholz using significant aggregated data is shown below.

```
result = calcLogisticModelAccuracy (forestTest$Krummholz, Krumm_Agg_Test_predict,
                        curThresh, curThresh, 1, "Krummholz", "Other", 3,
                        saveFile=saveFileName, desc="Krummholz Sig Aggregate Vars",
                        AIC=Krumm_Agg_Sig_aic, AUC=Krumm_Agg_Sig_ROC_AUC)
```

```
## [1] "Model Performance for threshold= 0.029"
## [1] "predicted performance="
##                      Predicted
## Actual                FALSE=Predict:Other TRUE=Predict:Krummholz
##   0=Actual:Other          151308 (TN)         16842 (FP)
##   1=Actual:Krummholz      289 (FN)            5864 (TP)
## [1] "Sensitivity= 0.953031041768243 (True positive rate of Krummholz = TP/(TP+FN) = 5864 /( 5864 + 28
## [1] "Specificity= 0.899839429081178 (True negative rate of Other = TN/(TN+FP) = 151308 /( 151308 + 16
## [1] "Sens^2+Spec^2=1.717"
## [1] "Baseline (Other) Accuracy=0.964699"
## [1] "Logistic Accuracy=0.901717"
```

```
list[RC, Krumm_Agg_Sig_model_acc, Krumm_Agg_Sig_baseline_acc,
      TN, FN, FP, TP, Krumm_Agg_Sig_sens, Krumm_Agg_Sig_spec] <- result
  if (RC != "OK") {
    print(paste("Error - terminating:",RC))
    knitr:knit_exit()
  }
  Krumm_Agg_Sig_model_acc = as.integer(as.numeric(Krumm_Agg_Sig_model_acc)*1000)/10
  Krumm_Agg_Sig_baseline_acc = as.integer(as.numeric(Krumm_Agg_Sig_baseline_acc)*1000)/10
  Krumm_Agg_Sig_sens = as.integer(as.numeric(Krumm_Agg_Sig_sens)*1000)/10
  Krumm_Agg_Sig_spec = as.integer(as.numeric(Krumm_Agg_Sig_spec)*1000)/10
```

### Calculate Krummholz Individuated Data Logisitic Model Accuracy - Significant Vars

Find best Krummholz threshold for Inividuated Data using significant variables.

```
result = calcLogisticModelAccuracy (forestTrain$Krummholz, Krumm_Ind_Train_predict,
                        0.0, 1, 10, "Krummholz", "Other", 1,1)
```

```
## [1] "Searching for threshold producing best Sensitivity_Specificity"
## [1] "start= 0 end= 1 inc= 0.1"
## [1] "Thresh=0, Accuracy=3.5%, BaseAcc(Other)=96.4%, Sens=100%, Spec=0%, Sens^2+Spec^2=-2"
## [1] "Thresh=0.1, Accuracy=94.4%, BaseAcc(Other)=96.4%, Sens=81.7%, Spec=94.9%, Sens^2+Spec^2=1.569"
## [1] "Thresh=0.2, Accuracy=96.2%, BaseAcc(Other)=96.4%, Sens=70.1%, Spec=97.1%, Sens^2+Spec^2=1.436"
## [1] "Thresh=0.3, Accuracy=96.6%, BaseAcc(Other)=96.4%, Sens=57.6%, Spec=98.1%, Sens^2+Spec^2=1.295"
## [1] "Thresh=0.4, Accuracy=96.8%, BaseAcc(Other)=96.4%, Sens=44.8%, Spec=98.7%, Sens^2+Spec^2=1.176"
## [1] "Thresh=0.5, Accuracy=96.8%, BaseAcc(Other)=96.4%, Sens=33.4%, Spec=99.1%, Sens^2+Spec^2=1.095"
## [1] "Thresh=0.6, Accuracy=96.7%, BaseAcc(Other)=96.4%, Sens=22.9%, Spec=99.4%, Sens^2+Spec^2=1.041"
## [1] "Thresh=0.7, Accuracy=96.6%, BaseAcc(Other)=96.4%, Sens=13.3%, Spec=99.6%, Sens^2+Spec^2=1.011"
## [1] "Thresh=0.8, Accuracy=96.6%, BaseAcc(Other)=96.4%, Sens=8.1%, Spec=99.8%, Sens^2+Spec^2=1.003"
## [1] "Thresh=0.9, Accuracy=96.5%, BaseAcc(Other)=96.4%, Sens=4.8%, Spec=99.9%, Sens^2+Spec^2=1.001"
## [1] "Thresh=1, Accuracy=96.4%, BaseAcc(Other)=96.4%, Sens=0%, Spec=100%, Sens^2+Spec^2=-2"
## [1] "Best Sensitivity_Specificity threshold= 0.1 inc= 0.1"
## [1] "========================================"
## [1] "start= 0 end= 0.2 inc= 0.01"
## [1] "Thresh=0, Accuracy=3.5%, BaseAcc(Other)=96.4%, Sens=100%, Spec=0%, Sens^2+Spec^2=-2"
## [1] "Thresh=0.01, Accuracy=77%, BaseAcc(Other)=96.4%, Sens=99.4%, Spec=76.2%, Sens^2+Spec^2=1.569"
## [1] "Thresh=0.02, Accuracy=83.6%, BaseAcc(Other)=96.4%, Sens=98.4%, Spec=83.1%, Sens^2+Spec^2=1.66"
## [1] "Thresh=0.03, Accuracy=87%, BaseAcc(Other)=96.4%, Sens=96.6%, Spec=86.6%, Sens^2+Spec^2=1.685"
```

```
## [1] "Thresh=0.04, Accuracy=89.2%, BaseAcc(Other)=96.4%, Sens=94%, Spec=89%, Sens^2+Spec^2=1.678"
## [1] "Thresh=0.05, Accuracy=90.8%, BaseAcc(Other)=96.4%, Sens=91.8%, Spec=90.7%, Sens^2+Spec^2=1.667"
## [1] "Thresh=0.06, Accuracy=91.9%, BaseAcc(Other)=96.4%, Sens=89.8%, Spec=92%, Sens^2+Spec^2=1.654"
## [1] "Thresh=0.07, Accuracy=92.8%, BaseAcc(Other)=96.4%, Sens=87.7%, Spec=93%, Sens^2+Spec^2=1.634"
## [1] "Thresh=0.08, Accuracy=93.4%, BaseAcc(Other)=96.4%, Sens=85.7%, Spec=93.7%, Sens^2+Spec^2=1.614"
## [1] "Thresh=0.09, Accuracy=93.9%, BaseAcc(Other)=96.4%, Sens=83.7%, Spec=94.3%, Sens^2+Spec^2=1.592"
## [1] "Thresh=0.1, Accuracy=94.4%, BaseAcc(Other)=96.4%, Sens=81.7%, Spec=94.9%, Sens^2+Spec^2=1.569"
## [1] "Thresh=0.11, Accuracy=94.8%, BaseAcc(Other)=96.4%, Sens=80.1%, Spec=95.3%, Sens^2+Spec^2=1.551"
## [1] "Thresh=0.12, Accuracy=95%, BaseAcc(Other)=96.4%, Sens=78.7%, Spec=95.6%, Sens^2+Spec^2=1.535"
## [1] "Thresh=0.13, Accuracy=95.3%, BaseAcc(Other)=96.4%, Sens=77.6%, Spec=95.9%, Sens^2+Spec^2=1.524"
## [1] "Thresh=0.14, Accuracy=95.5%, BaseAcc(Other)=96.4%, Sens=76.4%, Spec=96.2%, Sens^2+Spec^2=1.51"
## [1] "Thresh=0.15, Accuracy=95.6%, BaseAcc(Other)=96.4%, Sens=75.3%, Spec=96.4%, Sens^2+Spec^2=1.497"
## [1] "Thresh=0.16, Accuracy=95.8%, BaseAcc(Other)=96.4%, Sens=74.2%, Spec=96.6%, Sens^2+Spec^2=1.484"
## [1] "Thresh=0.17, Accuracy=95.9%, BaseAcc(Other)=96.4%, Sens=73.1%, Spec=96.7%, Sens^2+Spec^2=1.472"
## [1] "Thresh=0.18, Accuracy=96%, BaseAcc(Other)=96.4%, Sens=72.1%, Spec=96.9%, Sens^2+Spec^2=1.459"
## [1] "Thresh=0.19, Accuracy=96.1%, BaseAcc(Other)=96.4%, Sens=71%, Spec=97%, Sens^2+Spec^2=1.446"
## [1] "Best Sensitivity_Specificity threshold= 0.03 inc= 0.01"
## [1] "========================================"
## [1] "start= 0.02 end= 0.04 inc= 0.001"
## [1] "Thresh=0.02, Accuracy=83.6%, BaseAcc(Other)=96.4%, Sens=98.4%, Spec=83.1%, Sens^2+Spec^2=1.66"
## [1] "Thresh=0.021, Accuracy=84%, BaseAcc(Other)=96.4%, Sens=98.3%, Spec=83.5%, Sens^2+Spec^2=1.666"
## [1] "Thresh=0.022, Accuracy=84.4%, BaseAcc(Other)=96.4%, Sens=98.2%, Spec=83.9%, Sens^2+Spec^2=1.67"
## [1] "Thresh=0.023, Accuracy=84.8%, BaseAcc(Other)=96.4%, Sens=98%, Spec=84.4%, Sens^2+Spec^2=1.674"
## [1] "Thresh=0.024, Accuracy=85.2%, BaseAcc(Other)=96.4%, Sens=97.9%, Spec=84.7%, Sens^2+Spec^2=1.678
## [1] "Thresh=0.025, Accuracy=85.5%, BaseAcc(Other)=96.4%, Sens=97.7%, Spec=85.1%, Sens^2+Spec^2=1.679
## [1] "Thresh=0.026, Accuracy=85.8%, BaseAcc(Other)=96.4%, Sens=97.5%, Spec=85.4%, Sens^2+Spec^2=1.682
## [1] "Thresh=0.027, Accuracy=86.1%, BaseAcc(Other)=96.4%, Sens=97.4%, Spec=85.7%, Sens^2+Spec^2=1.684
## [1] "Thresh=0.028, Accuracy=86.4%, BaseAcc(Other)=96.4%, Sens=97.1%, Spec=86%, Sens^2+Spec^2=1.685"
## [1] "Thresh=0.029, Accuracy=86.7%, BaseAcc(Other)=96.4%, Sens=96.9%, Spec=86.3%, Sens^2+Spec^2=1.685
## [1] "Thresh=0.03, Accuracy=87%, BaseAcc(Other)=96.4%, Sens=96.6%, Spec=86.6%, Sens^2+Spec^2=1.685"
## [1] "Thresh=0.031, Accuracy=87.2%, BaseAcc(Other)=96.4%, Sens=96.3%, Spec=86.9%, Sens^2+Spec^2=1.685
## [1] "Thresh=0.032, Accuracy=87.5%, BaseAcc(Other)=96.4%, Sens=96%, Spec=87.2%, Sens^2+Spec^2=1.683"
## [1] "Thresh=0.033, Accuracy=87.7%, BaseAcc(Other)=96.4%, Sens=95.7%, Spec=87.4%, Sens^2+Spec^2=1.682
## [1] "Thresh=0.034, Accuracy=87.9%, BaseAcc(Other)=96.4%, Sens=95.5%, Spec=87.7%, Sens^2+Spec^2=1.682
## [1] "Thresh=0.035, Accuracy=88.2%, BaseAcc(Other)=96.4%, Sens=95.2%, Spec=87.9%, Sens^2+Spec^2=1.68"
## [1] "Thresh=0.036, Accuracy=88.4%, BaseAcc(Other)=96.4%, Sens=94.9%, Spec=88.1%, Sens^2+Spec^2=1.68"
## [1] "Thresh=0.037, Accuracy=88.6%, BaseAcc(Other)=96.4%, Sens=94.7%, Spec=88.4%, Sens^2+Spec^2=1.679
## [1] "Thresh=0.038, Accuracy=88.8%, BaseAcc(Other)=96.4%, Sens=94.4%, Spec=88.6%, Sens^2+Spec^2=1.678
## [1] "Thresh=0.039, Accuracy=89%, BaseAcc(Other)=96.4%, Sens=94.2%, Spec=88.8%, Sens^2+Spec^2=1.678"
## [1] "========================================"
## [1] "Best Threshold=0.03"
## [1] "Best Sensitivity_Specificity=1.68597013766091"
```

```
curThresh = as.numeric(result[bestThreshIndex])
Krumm_Ind_Sig_threshold = curThresh
```

The accuracy for the best threshold on the training set for Krummholz using significant individuated data is shown below.

```
result = calcLogisticModelAccuracy (forestTrain$Krummholz, Krumm_Ind_Train_predict,
                       curThresh, curThresh, 1, "Krummholz", "Other", 3)
```

```
## [1] "Model Performance for threshold= 0.03"
## [1] "predicted performance="
##                      Predicted
```

```
## Actual               FALSE=Predict:Other TRUE=Predict:Krummholz
##   0=Actual:Other         340054 (TN)         52298 (FP)
##   1=Actual:Krummholz     476 (FN)            13881 (TP)
## [1] "Sensitivity= 0.966845441248172 (True positive rate of Krummholz = TP/(TP+FN) = 13881 /( 13881 +
## [1] "Specificity= 0.866706426881984 (True negative rate of Other = TN/(TN+FP) = 340054 /( 340054 + 5
## [1] "Sens^2+Spec^2=1.685"
## [1] "Baseline (Other) Accuracy=0.964699"
## [1] "Logistic Accuracy=0.870241"
```

The accuracy for the best threshold on the testing set for Krummholz using significant individuated data is shown below.

```
result = calcLogisticModelAccuracy (forestTest$Krummholz, Krumm_Ind_Test_predict,
                       curThresh, curThresh, 1, "Krummholz", "Other", 3,
                       saveFile=saveFileName, desc="Krummholz Sig Individualized Vars",
                       AIC=Krumm_Ind_Sig_aic, AUC=Krumm_Ind_Sig_ROC_AUC)
```

```
## [1] "Model Performance for threshold= 0.03"
## [1] "predicted performance="
##                       Predicted
## Actual               FALSE=Predict:Other TRUE=Predict:Krummholz
##   0=Actual:Other         145565 (TN)         22585 (FP)
##   1=Actual:Krummholz     223 (FN)            5930 (TP)
## [1] "Sensitivity= 0.963757516658541 (True positive rate of Krummholz = TP/(TP+FN) = 5930 /( 5930 + 2
## [1] "Specificity= 0.865685399940529 (True negative rate of Other = TN/(TN+FP) = 145565 /( 145565 + 2
## [1] "Sens^2+Spec^2=1.678"
## [1] "Baseline (Other) Accuracy=0.964699"
## [1] "Logistic Accuracy=0.869147"
```

```
list[RC, Krumm_Ind_Sig_model_acc, Krumm_Ind_Sig_baseline_acc,
      TN, FN, FP, TP, Krumm_Ind_Sig_sens, Krumm_Ind_Sig_spec] <- result
  if (RC != "OK") {
    print(paste("Error - terminating:",RC))
    knitr:knit_exit()
  }
  Krumm_Ind_Sig_model_acc = as.integer(as.numeric(Krumm_Ind_Sig_model_acc)*1000)/10
  Krumm_Ind_Sig_baseline_acc = as.integer(as.numeric(Krumm_Ind_Sig_baseline_acc)*1000)/10
  Krumm_Ind_Sig_sens = as.integer(as.numeric(Krumm_Ind_Sig_sens)*1000)/10
  Krumm_Ind_Sig_spec = as.integer(as.numeric(Krumm_Ind_Sig_spec)*1000)/10

############# End End End End End End End End End End End End ##################
```

The accuracy of the models is shown below:

| Logistic Model | Accuracy | Sens | Spec | AIC | AUC | Threshhold |
|---|---|---|---|---|---|---|
| Krummholz Aggregate All Vars | 91.8% | 93.5% | 91.8% | 46631 | 98% | 0.035 |
| Krummholz Individual All Vars | 91.7% | 93.6% | 91.6% | 46642 | 98% | 0.034 |
| Krummholz Aggregate Sig Vars | 90.1% | 95.3% | 89.9% | 55786 | 97.4% | 0.029 |
| Krummholz Individual Sig Vars | 86.9% | 96.3% | 86.5% | 60618 | 96.8% | 0.03 |

There is a slight degradation in the accuracy with insignificant variables eliminated, but not by much.

## Conclusion

It is beginning to look like there is no advantage to dis-aggregating the Soil Type variables into their component parts. I was hoping there would be some improvement by allowing the individual variables to be "more finely" tuned. There is probably a mathematical explanation that proves there is no advantage of breaking out aggregated variables. I have to think about that more.

The logistic regression results for Krummholz are 7% better than the original paper this project was modeled after. These tests need to be done for the remaining 6 forest cover types to see how regression does overall.

```
curTime=Sys.time()
print(paste("Forest Cover Logistic script ended at",curTime))
```

```
## [1] "Forest Cover Logistic script ended at 2018-08-12 18:59:23"
```