Levee Failure Along the Mississippi River

Yamar Ba
Prof. Williams
STAT 793

INTRODUCTION

Data on levee failures throughout the past century along much of the Middle Mississippi River will be analyzed with the goal of testing the relative importance of geologic, geomorphic, and other physical factors that have caused it.

A levee is a wall, either artificial or natural, used to keep water (river, lake or ocean) out or in. The adjacent land is kept dry for human development or as a means of safety for human inhabitants. They are extremely useful and vital in many communities in the United States. Including communities along the Mississippi River. Figure 1 is an illustration of the evolution of a particular levee along the Middle Mississippi river, increasing in height and width over time as water levels increase.

Fig. 1

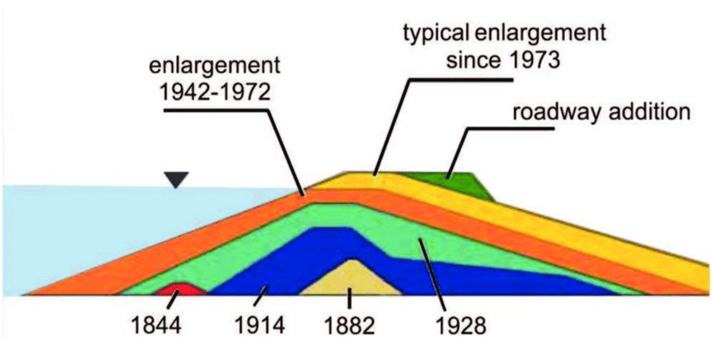


Fig. 2 Levee maintained by the US Army Corps of Engineers.

Levees were a critical part of flood-control strategy in the United States throughout the 19th and much of the 20th centuries. There were widespread levee failures along the Mississippi River in 1927 that inundated a vast area of land that killed at least 500 people and displaced around another 900,000. Other methods of flood control were explored and

Fig. 3 Levee breach

employed given that there were obvious limitations of levees. However, levees are still an important part of flood control in America. Over 40% of the U.S. population live in a county with at least one levee. In 2005 levees in New Orleans failed. A hurricane was thought to have caused the levees to fail, flooding 80% of the city and causing massive human displacement and the death of about 1,500 people. Understanding the underlying causes and failures of levees remains relevant today.

DATA

Data collected from 1880 to 1998 along the Middle Mississippi river of potential levee failure causing variables will be analyzed to determine if factors leading to levee failures can be identified. Data from levee failure sites was created using information from scientific reports, historic maps and print media sources. Non-levee failure sites were randomly generated by a random number generator and normalized to the total number of river-miles to get a random river-mile location. The number of non-failure and failure sites are equal in the data. An additional random number was generated to select either the right bank or left bank at each location. If a levee was not present at the generated location, the random function ran again until it coincided with a levee.

The levee dataset analyzed was retrieved from the College of Liberal Arts and Science at the University of Florida. The data set can be found on their user website. A link to the data will also be provided in the references section of this paper.

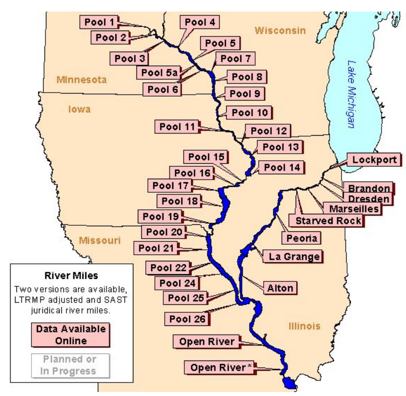
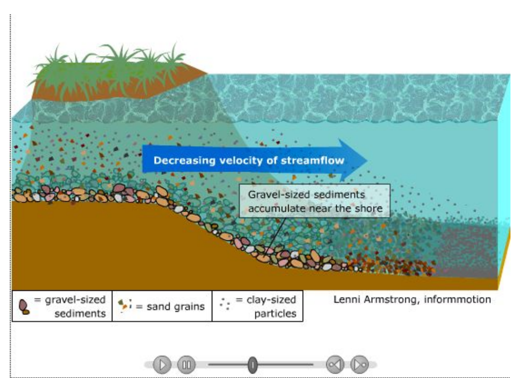
VARIABLES

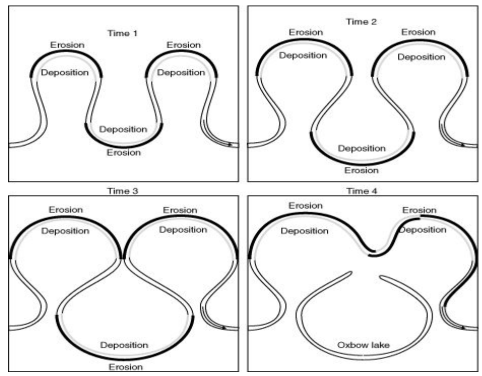
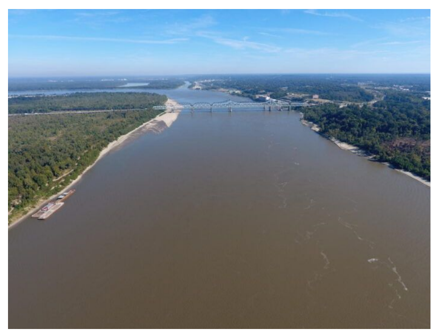
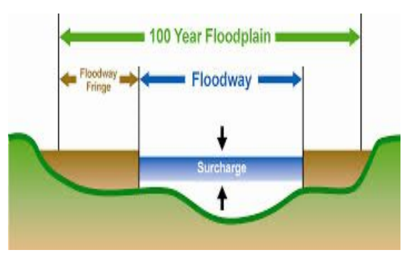
In general, there are three main causes of levee failure:

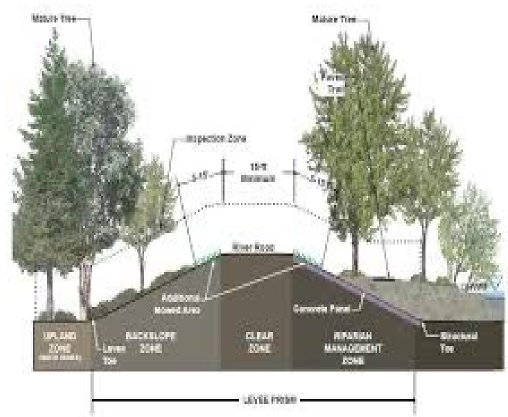
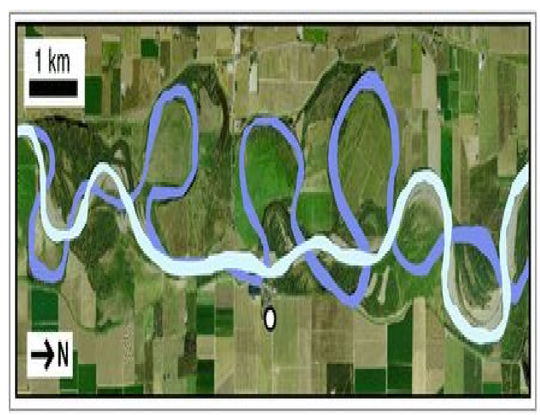
1. Foundation failure of a levee
2. Erosion and damage of a levee
3. Overtopping or water overflow of a levee.

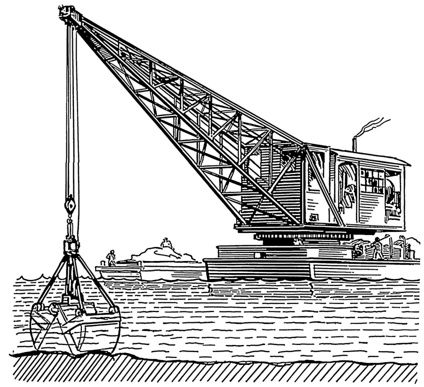
A storm surge can cause any one of the three mentioned above, if not all. However, there are usually underlying variables. Variables of interest present in the data are described in table 1. Many are multilevel factors. Two level factors indicate presence or absence.

Table 1

	Variable	**Data Type**	**Description**	**Illustration**
1	Failure	2 Level Factor	1 = levee failed, 0 = levee did not fail	See illustrations 1&2
2	Year	4 Level Factor	Observations made from 1890 to 1998.	
3	River Mile	Numeric	A measure of distance in miles along a river from its mouth.	
4	Sediments	2 Level Factor	Eroded rocks, sand and other matter that settles to the bottom of the river.	

	Variable	Data Type	Description	Illustration
5	Borrow Pit	2 Level Factor	An area where soil, gravel or sand has been dug for use at another location	
6	Meander	4 level Factor	A winding curve or bend of a river. 1=Inside bend, 2=outside bend, 3=chute, 4=straight	
7	Channel Width	Numeric	Distance across a river channel.	
8	Floodway Width	Numeric	Adjacent land to a river or stream designated for water overflow.	
9	Constriction Factor	Numeric	Constriction of the floodway over time.	Same as above.

	Variable	Data Type	Description	Illustration
10	Land Cover	4 Level Factor	Four groups: agriculture, forest, grass, and open water. These are the predominant land-cover types between the levee and the main channel at each levee location.	
11	Vegetation Width	Numeric	Vegetated buffer between the channel and the levee.	
12	Sinuosity	Numeric	Channel sinuosity is equal to the line of lowest elevation within a valley or watercourse length divided by the valley length. Channel segments typically 10 river miles up-valley and down-valley.	

	Variable	Data Type	Description	Illustration
13	Dredging	Integer	Dredging is the removal of material from the bottom of lakes, rivers, harbors and other water bodies. Dredging intensity for each location was calculated using the total volumes of material dredged from the closest river mile.	
14	Revetement	2 Level Factor	The purpose of bank revetments is to reduce erosion along the bank and ultimately reduce the risk of levee failure. It will either present or not present.	

METHODS OF ANALYSIS AND PROCEDURES

Model

To model the probability of a levee failure we will use logistic regression since the dependent variable is a binary response, absence or presence of failure. We will perform multinomial logistic regression and our final model will be multinomial.

The models are as follows, respectively:

logit(p)=ln(p/1-p)=$\beta_0+\beta_1X_1$ OR Antilog p/1-p=EXP($\beta_0+\beta_1X_1$) *binomial
logit(p)=ln(p/1-p)=$\beta_0+\beta_1X_1$ OR Antilog p/1-p=EXP($\beta_0+\beta_1X_1+....+\beta_nX_n$) *multinomial

Data exploration

There are many variables to consider and we begin by exploring them. The variables in question are continuous or categorical. The frequency of the categorical variables and the pairwise relationship between the continuous variables will be explored.

Methods

We will perform logistic regression on the data in R using two methods on the same dataset. In the first method, we will use the classic logistic regression analysis model and glm function where all of the dependent variables are used to predict the dependent variable. The variables of significance will then be put into a new better fitting model. On the second method, we will use machine learning techniques with a training and testing dataset to find our predicted model. Variables of significance will be determined using the Boruta and Random Forest algorithms.

Comparisons will be made within and between models along the way. Their respective AIC's, BIC's, Pseudo R-Squareds and P-values will be compared. Will our final models have the same variables of interest? Which one will be the better, tighter fitting model? We will go through the procedures and answer these questions.

ANALYSIS

The dataset has 6 factor variables and 7 continuous or numeric variables. The variable Year does not add any useful information to our models so we will exclude the variable from analysis. See structure of data below from R output.

```
'data.frame':	70 obs. of  14 variables:
 $ Fail    : Factor w/ 2 levels "Fail","not Fail": 2 2 2 2 2 2 2 2 2 2 ...
 $ Yr      : int  1880 1908 1908 1908 1908 1908 1908 1948 1948 1948 ...
 $ R. Mile : num  188 190 174 147 143 ...
 $ Seds    : Factor w/ 2 levels "0","1": 1 2 1 2 1 1 1 1 1 2 ...
 $ B.Pit   : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
 $ Mea     : Factor w/ 4 levels "1","2","3","4": 2 1 1 1 4 4 1 1 1 3 ...
 $ C.Width : num  2513 1271 920 1115 1032 ...
 $ F.Width : num  6991 4344 3396 2105 2513 ...
 $ C.Factor: num  1 3.058 1.001 0.949 0.958 ...
 $ L.Cover : Factor w/ 4 levels "1","2","3","4": 1 3 3 1 4 3 2 4 4 4 ...
 $ V.Width : num  0 183 2411 0 305 ...
 $ Sin     : num  1.231 1.24 0.994 1.107 0.997 ...
 $ Dredg   : int  0 0 19354 34968 46540 0 0 218751 218751 0 ...
 $ Rev     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

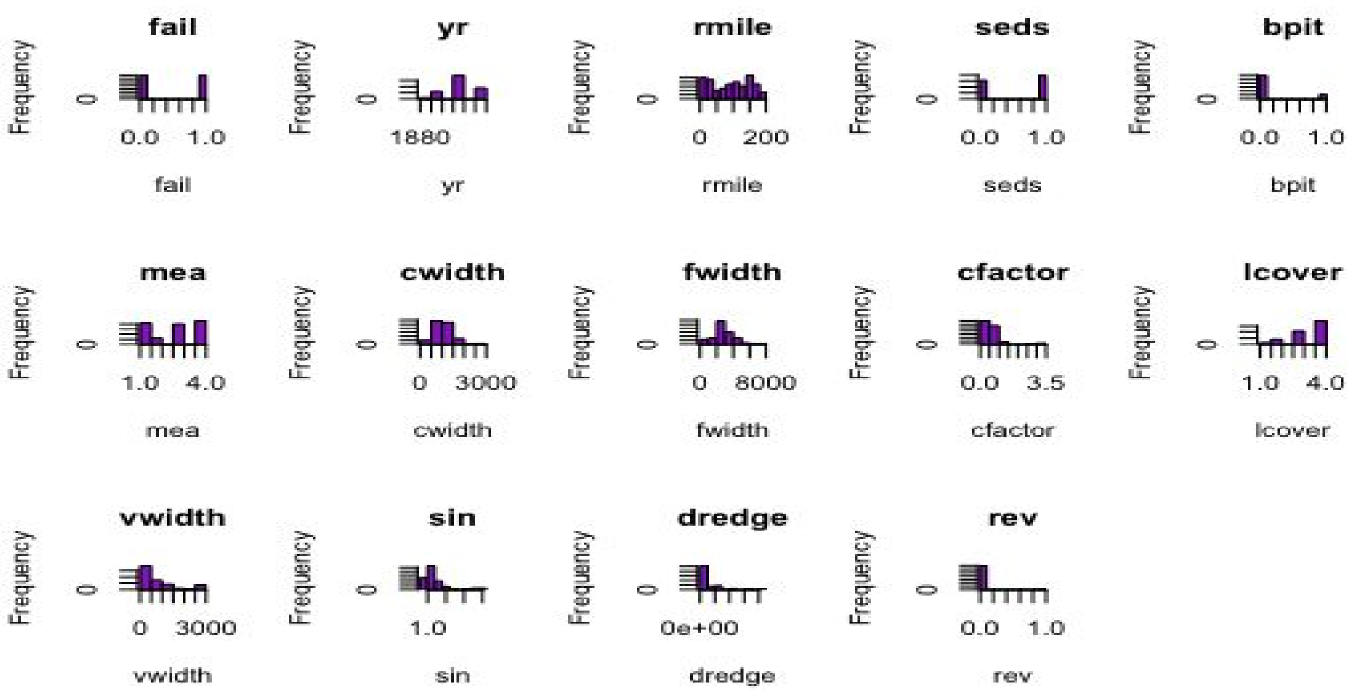
A glance at the head and tail of the data shows us what the dataset looks like.

```
  fail    yr rmile seds bpit mea   cwidth   fwidth cfactor lcover   vwidth    sin dredge
rev
1    1 1880 188.4    0    0   2 2512.91 6990.65  1.0000      1    0.00 1.2307      0
2    1 1908 190.0    1    0   1 1270.84 4343.60  3.0576      3  182.96 1.2405      0
3    1 1908 174.2    0    0   1  920.22 3395.81  1.0012      3 2410.73 0.9939  19354

  fail    yr  rmile seds bpit mea   cwidth   fwidth cfactor lcover  vwidth    sin dredge

68    0 1998  92.94    0    1   4  568.06 3196.81  0.3268      3 2577.53 1.0966  27980
69    0 1998  92.66    0    1   4  646.59 3393.07  0.3268      3 2714.91 1.0803      0
70    0 1998   6.50    1    0   3  783.30 3786.90  0.2078      3  471.00 1.2376 211688
```

To get a better sense of the data, let's see the frequency of each of the variables.



We also use a frequency table to see how the data is spread out for our factor variables.

Fail	Seds 0	1
Fail	22	13
not Fail	9	26

Fail	B.Pit 0	1
Fail	29	6
not Fail	30	5

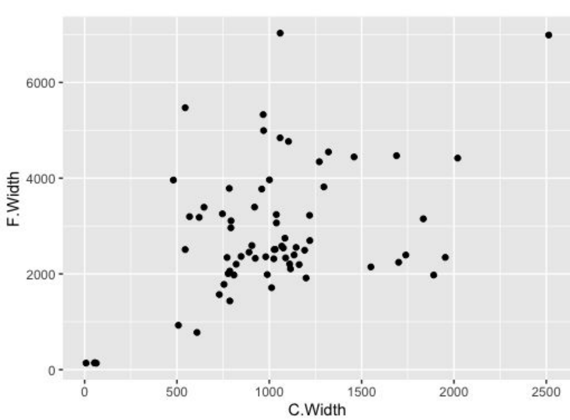
Fail	Mea 1	2	3	4
Fail	8	4	8	15
not Fail	13	2	12	8

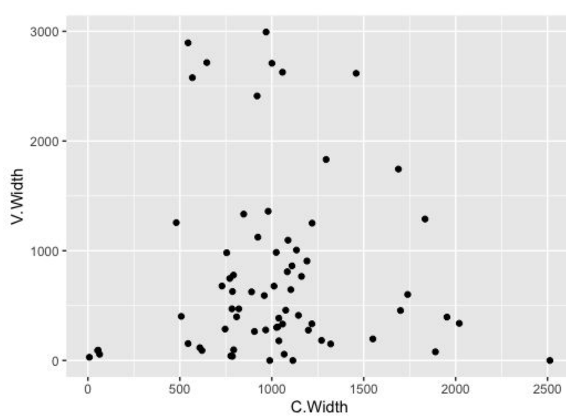
Fail	L.Cover 1	2	3	4
Fail	0	4	13	18
not Fail	3	4	8	20

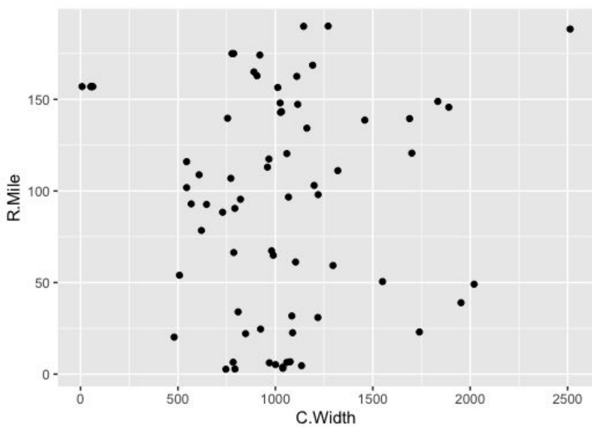
Fail	Rev 0	1
Fail	34	1
not Fail	35	0

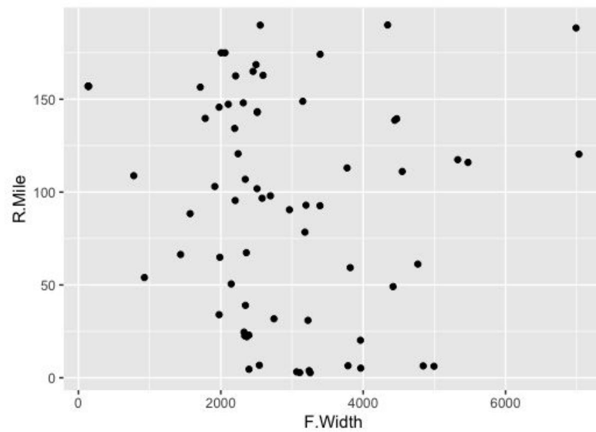
Land cover and Revetement both have zeros in their frequency tables. Indicating that the variables are not very well distributed. In addition, only one levee failure is represented in level 2 of Revetement compared to 34 at its level 1. Similarly, Land Cover has a zero at its level 1. These deep contrasts in the data may pose a problem in finding the best fitting line.

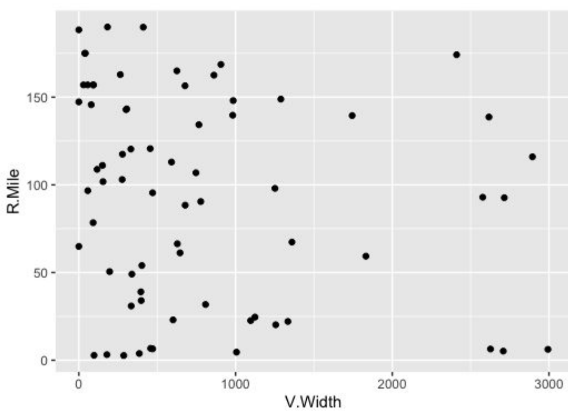
Similar to what was done on the factor level. The continuous variables, Channel Width, River Mile, Floodway Width, Constriction Factor, Vegetation Width, Sinuosity and Dredging are plotted against each other to have a better visual sense of the distribution of the data.

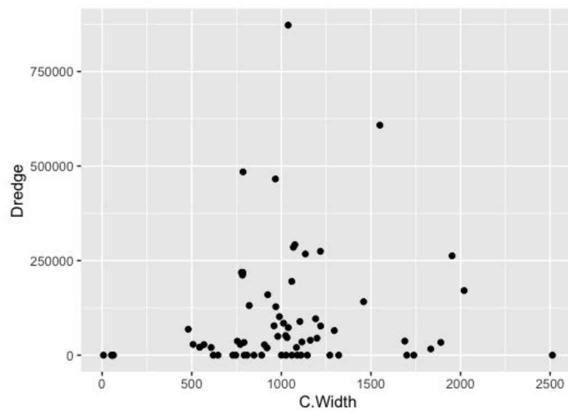


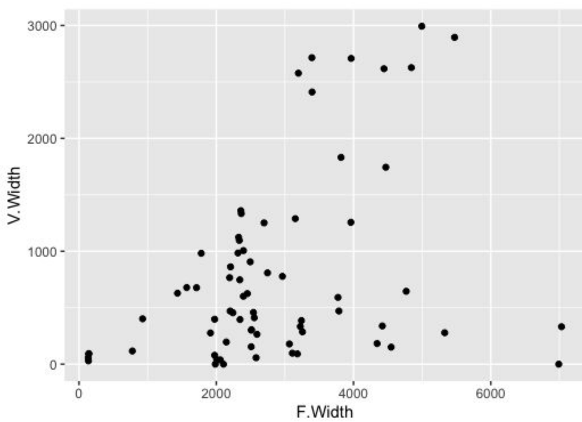


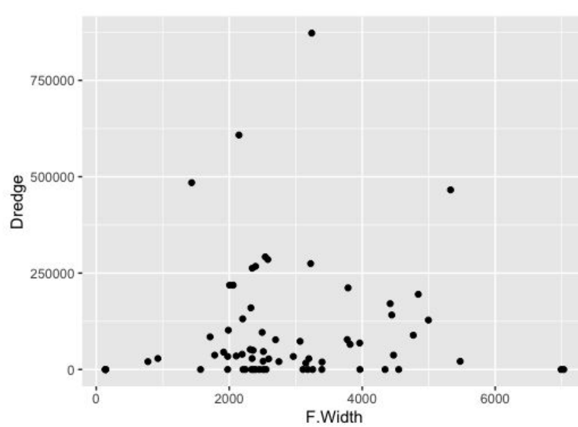

R.Mile
C.Width
0
50
100
150
500
1000
1500
2000
2500

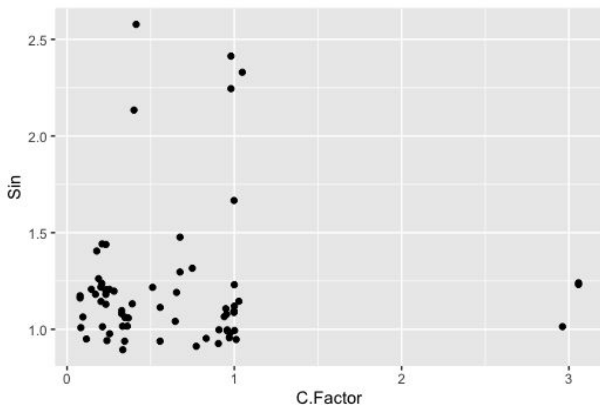

R.Mile
F.Width
0
50
100
150
2000
4000
6000

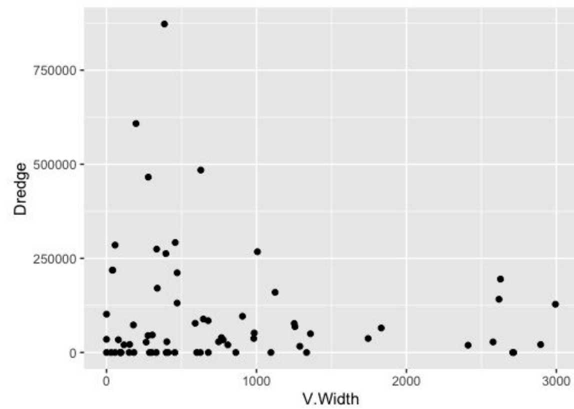

R.Mile
V.Width
0
50
100
150
1000
2000
3000

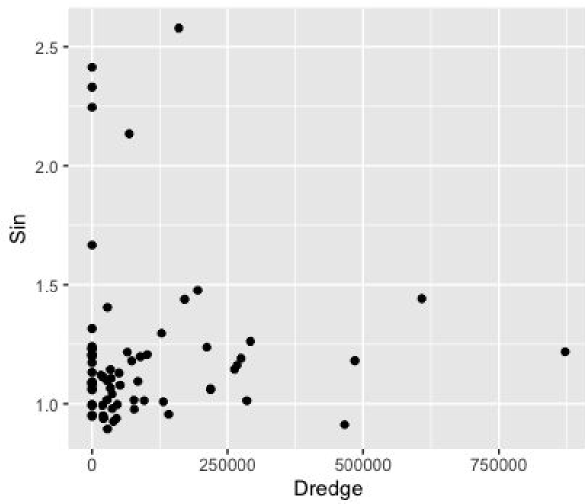

Dredge
C.Width
0
250000
500000
750000
500
1000
1500
2000
2500

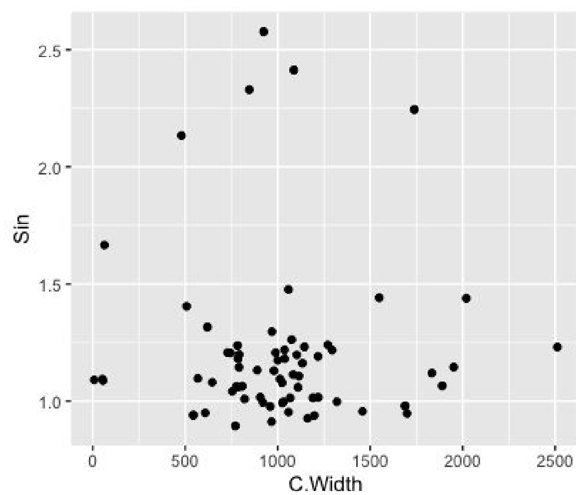

V.Width
F.Width
0
1000
2000
3000
2000
4000
6000

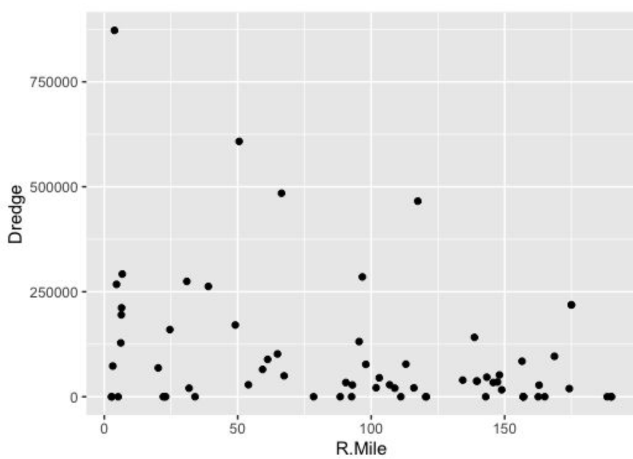

Dredge
F.Width
0
250000
500000
750000
2000
4000
6000

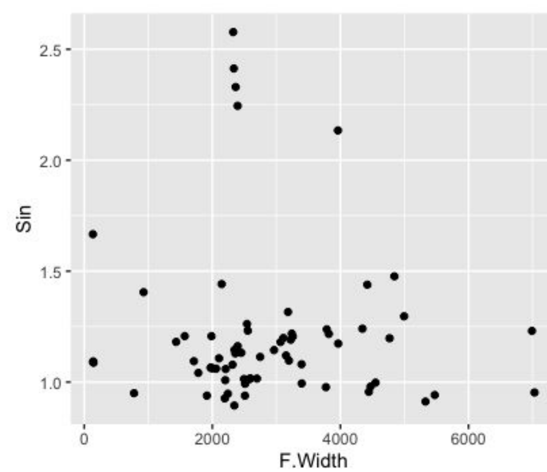

Sin
2.5
2.0
1.5
1.0
0
1
2
3
C.Factor

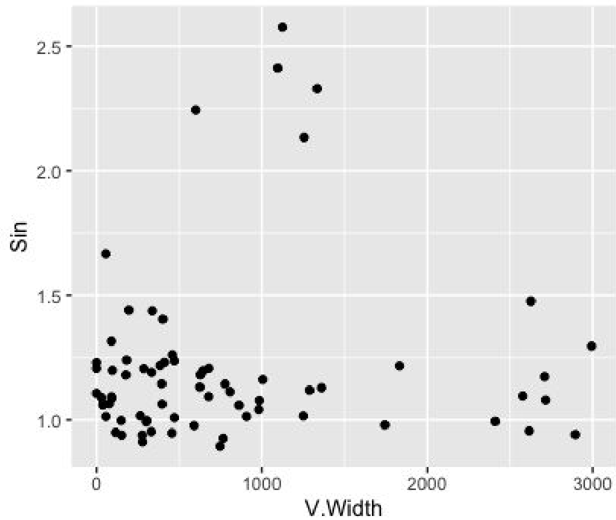

Dredge
750000
500000
250000
0
0
1000
2000
3000
V.Width


Sin
2.5
2.0
1.5
1.0
0
250000
500000
750000
Dredge

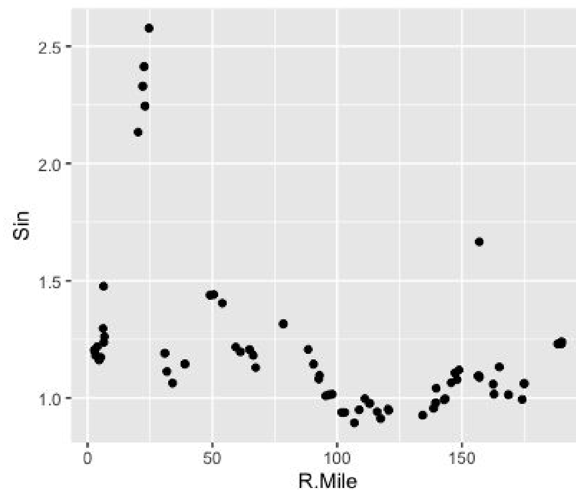

Sin
2.5
2.0
1.5
1.0
0
500
1000
1500
2000
2500
C.Width


Dredge
750000
500000
250000
0
0
50
100
150
R.Mile

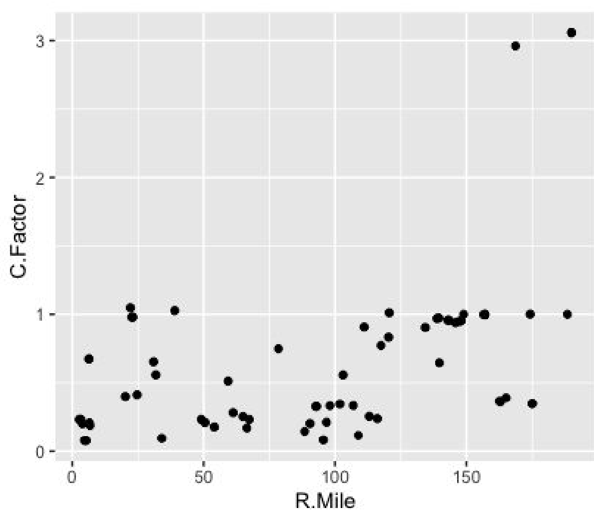

Sin
2.5
2.0
1.5
1.0
0
2000
4000
6000
F.Width



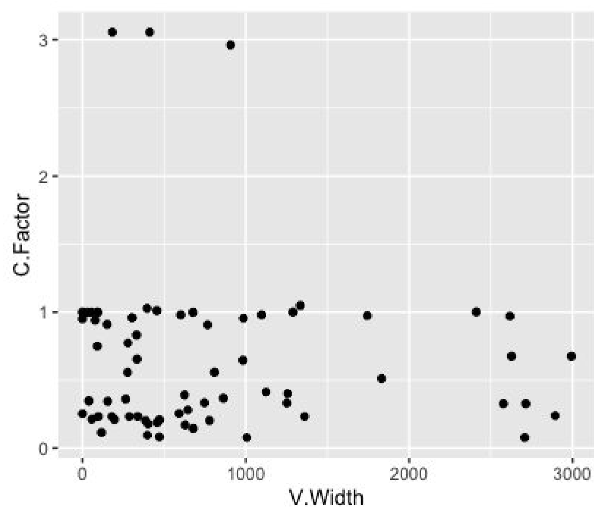
Sin
V.Width



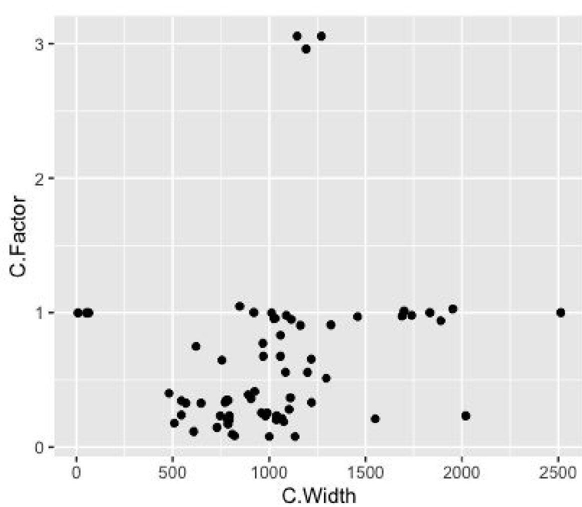
Sin
R.Mile



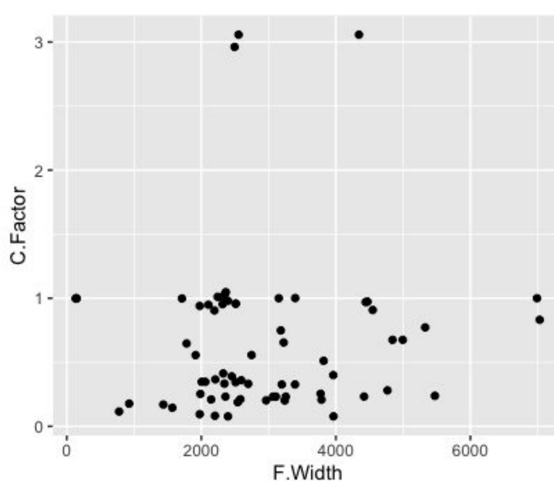
C.Factor
R.Mile



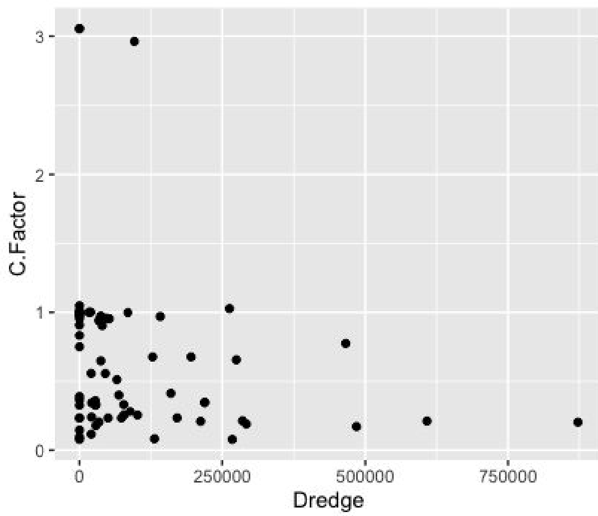
C.Factor
V.Width



C.Factor
C.Width



C.Factor
F.Width



From the pairwise plots of the continuous variables we can see that all of the continuous variables do not produce any visual relationship. The data for most of the variables are spread out with no visible pattern. Others have data concentrated on certain indexes but still with no visible or significant pattern. The lack of correlation among the continuous predictor variables is a positive indication that our multiple regression model is unlikely to be negatively impacted by multicollinearity. High correlations among predictor variables lead to unreliable and unstable estimates of regression coefficients.

METHOD 1

If we were predicting levee failure when there are sediments present or absent, at ≥95% (p-value .05) confidence we would get the following model, coefficients and equation.

```
Call:
glm(formula = Fail ~ Seds, family = "binomial", data = levee)
                          :
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.8938     0.3957  -2.259  0.02389 *
Seds1         1.5870     0.5215   3.043  0.00234 **
                          :

Model Equation-> Levee failure = -.0894 + 1.587Seds
```

Therefore, The log odds that a levee fails when sediments are not present is -.0894. The log odds that a levee fails when sediments are present is 1.498, which is a one unit increase in the odds of sediments causing levee failure. The presence of sediments increase the risk of levee failure.

To make our model more interesting, we include all variables to our glm call function.

```
Call:
glm(formula = Fail ~ ., family = "binomial", data = levee)

Deviance Residuals:
    Min       1Q    Median       3Q      Max
-2.3645  -0.7916  -0.0001   0.8543   2.0908

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  5.102e+01  1.629e+03   0.031   0.9750
Yr          -1.407e-02  1.447e-02  -0.972   0.3308
`R. Mile`   -4.685e-03  1.015e-02  -0.461   0.6446
Seds1        1.898e+00  7.643e-01   2.483   0.0130 *
B.Pit1       1.502e+00  1.443e+00   1.041   0.2980
Mea2        -3.133e+00  2.804e+00  -1.117   0.2639
Mea3        -1.389e+00  1.136e+00  -1.222   0.2216
Mea4        -1.563e+00  9.426e-01  -1.658   0.0973 .
C.Width     -1.815e-03  1.015e-03  -1.789   0.0737 .
F.Width     -4.074e-05  3.144e-04  -0.130   0.8969
C.Factor     1.299e-02  7.330e-01   0.018   0.9859
L.Cover2    -2.173e+01  1.628e+03  -0.013   0.9894
L.Cover3    -2.140e+01  1.628e+03  -0.013   0.9895
L.Cover4    -2.050e+01  1.628e+03  -0.013   0.9900
V.Width     -3.521e-04  4.707e-04  -0.748   0.4545
Sin          5.613e-02  1.107e+00   0.051   0.9596
Dredg       -2.771e-06  2.191e-06  -1.265   0.2060
Rev1        -1.673e+01  3.956e+03  -0.004   0.9966
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 97.041  on 69  degrees of freedom
Residual deviance: 68.606  on 52  degrees of freedom
AIC: 104.61

Number of Fisher Scoring iterations: 16
> BIC(logistic2)
[1] 145.0785
```

At ≥95% significance, we only have Sediments or Seds as significant. At ≥80% significance we have an additional three variables. As highlighted above, they are Meander/Mea and Channel Width/C.Width at ≥90% significance and Dredge at 80% significance.

We model on the variables that were significant at ≥80% significance.

```
Call:
glm(formula = Fail ~ Seds + Mea + C.Width + Dredge, family = "binomial",
    data = levee)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.05726  -0.83896  -0.03084   0.85944   1.87766

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.912e-01  8.748e-01   0.790   0.4295
Seds1        1.978e+00  6.813e-01   2.903   0.0037 **
Mea2         4.795e-02  1.320e+00   0.036   0.9710
Mea3        -9.716e-01  8.456e-01  -1.149   0.2506
Mea4        -1.370e+00  7.343e-01  -1.866   0.0621 .
C.Width     -7.229e-04  7.205e-04  -1.003   0.3157
Dredge      -3.221e-06  2.035e-06  -1.583   0.1135
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 97.041  on 69  degrees of freedom
Residual deviance: 78.094  on 63  degrees of freedom
AIC: 92.094

Number of Fisher Scoring iterations: 4

> BIC(logistic3)
[1] 107.8331
```

Of the variables that were included in our new model, logistic3, at ≥80%, C.Width is no longer significant.

Comparing the old model, logistic2 with all of the variables included, to the new model, logistic3, with only the significant variables, the AIC and BIC is much lower. See table 1 below:

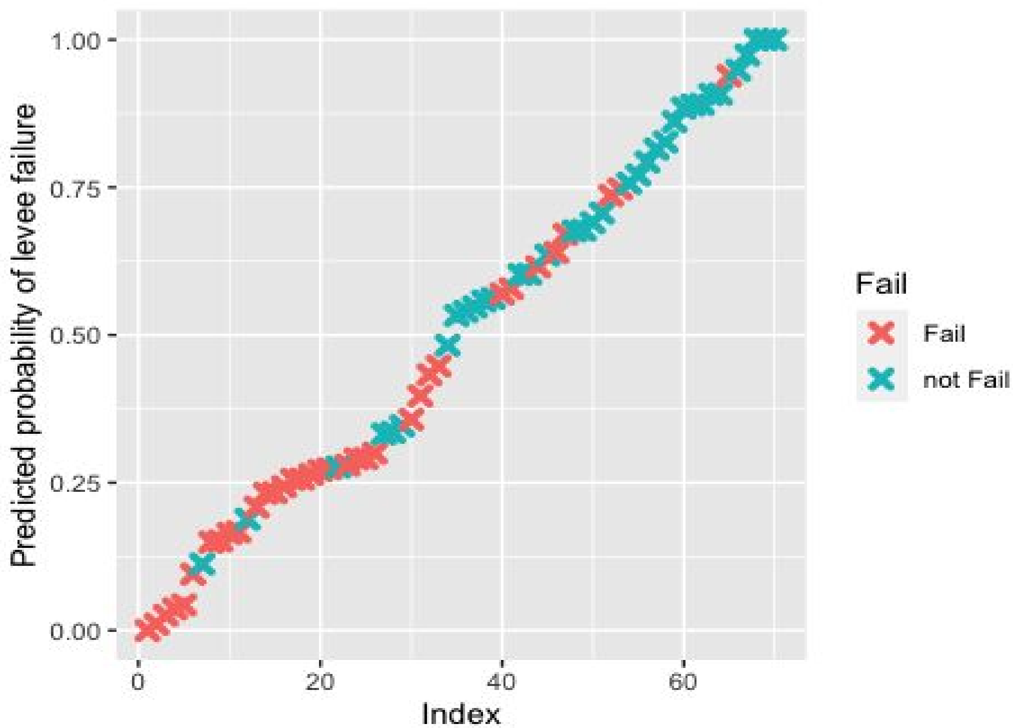
Table 1

Model	AIC	BIC
Model 1/logistic2	104.61	145.0785
Model 2/logistic3	92.094	107.8331

We plot predicted probabilities that each levee fails along with their actual observations. See plot 1.

```
> ggplot(data=predicted.data, aes(x=rank, y=probability.of.fail)) +
      geom_point(aes(color=Fail), alpha=1, shape=4,stroke=2) + xlab("Index") +
      ylab("Predicted probability of levee failure")
```

Plot 1



METHOD 2

We use a second method to confirm our variables form the first method. We use Random Forest and Boruta algorithms to determine our variables of significance.

Random Forest ranks the variables of interest below. The higher the number, the more important. The top 4 ranking variables are highlighted.

```
output.forest1 <- randomForest(fail ~ rmile + seds + bpit + mea + cwidth +
                         fwidth + cfactor + lcover + vwidth + sin + dredge  +
                         rev, data = leveee)
randomForest::importance(output.forest1)
```

```
        IncNodePurity
rmile      1.93052774
seds       1.35883457
bpit       0.16039293
mea        0.59379127
cwidth     1.53316754
fwidth     1.65744883
cfactor    1.70009233
lcover     0.42899567
vwidth     2.10762809
sin        1.55149412
dredge     2.43553204
rev        0.03581666
```

To confirm our variables we use the Boruta method. Below is the output:

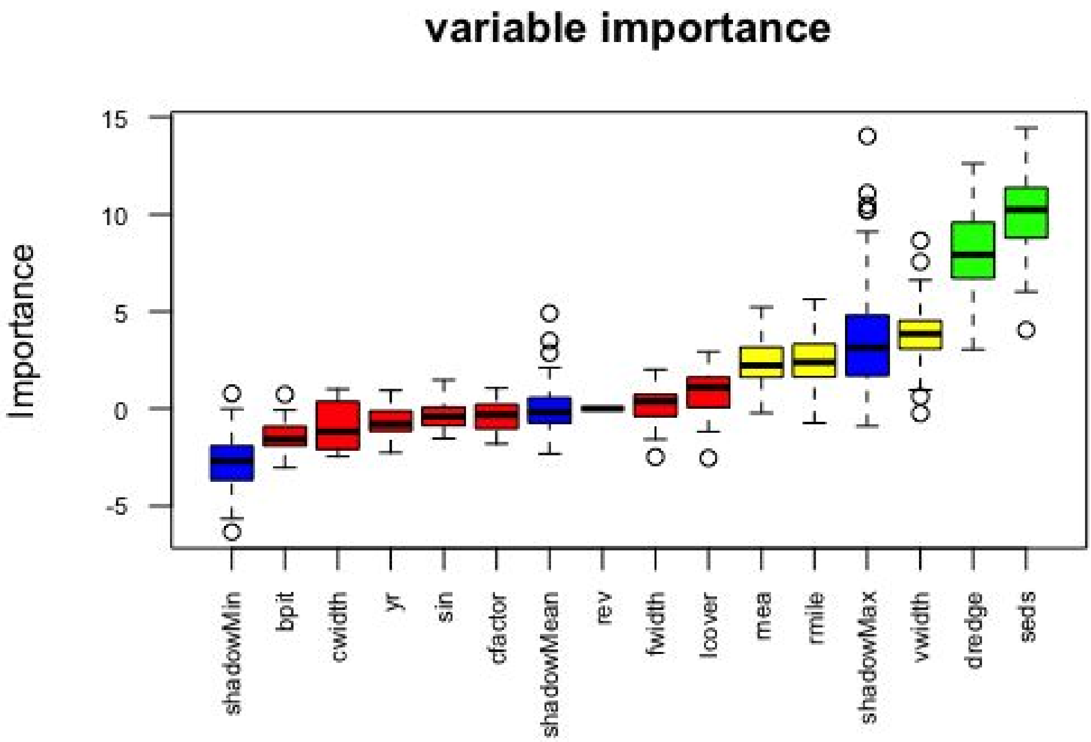
```
> print(boruta_output) #lists important and unimportant  variables
Boruta performed 99 iterations in 3.656029 secs.
 2 attributes confirmed important: dredge, seds;
 8 attributes confirmed unimportant: bpit, cfactor, cwidth, fwidth,
lcover and 3 more;
 3 tentative attributes left: mea, rmile, vwidth;

>print(boruta_signif)
[1] "rmile"  "seds"   "mea"    "vwidth" "dredge"
```

As highlighted above we have the variables Dredge and Seds confirmed significant along with 3 additional variables. They are River Mile/rmile, Meander/mea and Vegetation Width/ width.

We plot the variable of importance in plot 2. The green box plots are important and the indeterminant variables are red and blue. The yellow boxplots are secondary. The green and yellow boxplots match variables of importance in the Boruta output.

Plot 2



We split our dataset for training and testing. 75% and 25%, respectively. We use the training dataset in our third logistic model, logisticModel1, using all of the variables in this alternative method and get our AIC and BIC from the model.

```
> logisticModel1<- glm(fail ~ rmile + seds + bpit + mea + cwidth + fwidth +
                        cfactor + lcover + vwidth + sin + dredge + rev,
                        data = trainingData, family = binomial(link="logit"))

> AIC(logisticModel1)
[1] 64.8665
> BIC(logisticModel1)
[1] 88.28143
```

We then build our fourth model, logisticModel2, using the variables of significance from the Boruta output. Those variables were rmile, seds, mea, vwidth and dredge.

```
> summary(logisticModel2)
Call:
glm(formula = fail ~ rmile + seds + mea + vwidth + dredge, family =
binomial(link = "logit"),
    data = trainingData)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.20713  -0.69441  -0.01734   0.68176   1.80800

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.174e+00  1.571e+00   1.384  0.16646
rmile       -5.590e-03  7.023e-03  -0.796  0.42606
seds         2.350e+00  9.126e-01   2.575  0.01004 *
mea         -9.298e-01  3.567e-01  -2.607  0.00914 **
vwidth      -1.933e-04  4.317e-04  -0.448  0.65428
dredge      -3.568e-06  2.333e-06  -1.530  0.12607
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 72.087  on 51  degrees of freedom
Residual deviance: 49.631  on 46  degrees of freedom
AIC: 61.631

Number of Fisher Scoring iterations: 5

> BIC(logisticModel2)
[1] 73.33797
```

We can see from the summary output of our fourth model that at ≥95% significance seds and mea are significant. At 80% significant dredge becomes significant. Excluded from this model is rmile and width.

We compare the two models from our training set. See table 1 below:

Table 2

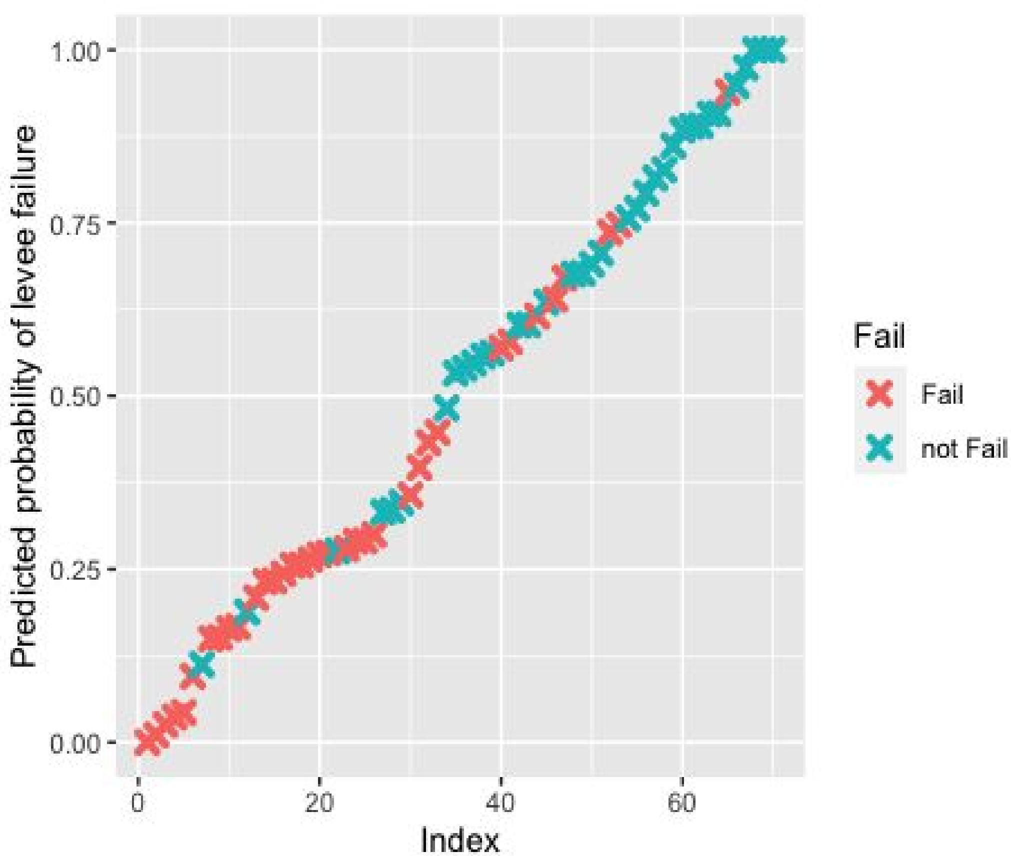
Model	AIC	BIC
Model 3/logisticModel1	64.8665	88.28143
Model 4/logisticModel2	61.631	73.33797

The lower AIC and BIC confirms that model 4 is significantly better than model 3 when the insignificant variables are excluded.

We plot predicted probabilities that each levee fails along with their actual observations using the training dataset. The graph is identical to the graph of model 1 from the original dataset. See graph 2 below.

```
ggplot(data=predicted.data, aes(x=rank, y=probability.of.fail)) +
      geom_point(aes(color=Fail), alpha=1, shape=4,stroke=2) + xlab("Index") +
      ylab("Predicted probability of levee failure")
```

Plot 3



We create frequency tables to compare our predicted values with the observed values from the original data set in method 1 and from the training data set from method 2.

Observed vs Predicted from Original Dataset

```
> mytable1 <- table(levee$Fail, levee$Predict)
> mytable1 #observed vs predicted

           Pred. neg Pred. pos
  Obs. neg        25        10
  Obs. pos        11        24
```

Observed vs Predicted from Training Dataset

```
mytable1 <- table(levee$fail, leveee$Predict)
> mytable1 #observed vs predicted

           Pred. neg Pred. pos
  Obs. neg        19         7
  Obs. pos         9        17
```

Proportionally, the frequency tables are nearly identical.

Lastly, we tested and compared the efficiency or the accuracy of model 2, the model with the significant variables from the original dataset, to model 4, the model with the significant variables from the training set. The pseudo R-squared and p-values for both models are also compared.

We use McFadden's pseudo R-squared because an R-Squared value does not exist in logistic regression. The logit regression relies on maximum likelihood estimates, they are not calculated to minimize variance. However, we can use McFadden's pseudo R-Squared to achieve a similar goodness of fit metric. McFadden's Pseudo R-squared is defined as $R2_{adj} = 1 - [\ln LL(\hat{M}_{full})-K]/[\ln LL(\hat{M}_{intercept})]$., where k is the number of regressors in the model.

Table 3

Model	R-squared	Accuracy	P-value
Model 2	0.1952479	0.7	0.004254134
Model 4	0.3115222	0.6923077	0.0004285961

The R-Squared in table 2 tells us that the proportion of the variance in the dependent variable that is predictable from the independent variable in Model 4 is higher than that of Model 2. The accuracy of the two models are identical but the p-value of Model 4 is significantly lower than that of Model 2.

The variables of interest or variables of significance from our original dataset and from the training set are identical at ≥80% significance. They are seds, mea and dredge. When we lower the significance level lower than 80%, the additional variables included in the models begin to differ. Of the two models, defined as Model 2 and Model 4, Model 4 is the better model because of its higher R-Squared and lower p-value.

CONCLUSION

We wanted to know if some, if not all, of the geologic, geomorphic, and other physical factors variables collected could help us predict the occurrence of levee failure along the Middle Mississippi River. Or

H*o*: Variables do not help us to predict the occurrence of levee failures.

H*a*: Variables help us to predict the occurrence of levee failures.

We used several methods to analyze the data. Using the original dataset, we first used the glm function on all of the dependent variables to predict the dependent variable. Variables with significant p-values were then used in a new model.

We also used machine learning algorithms to determine our final model. We split our data into training and testing, 75% and 25%, respectively. Then created a new training data set which was used to analyze the data. We determined our variables of significance using Randomforest and the Boruta method. Similar to the previous method, we then used the significant variables in a new model from which significant variables were used in the final equation.

The final variables of interest in both methods were identical. To choose between the models with the best fit we chose the model with the best pseudo R-squared, accuracy test and p-value.

In conclusion, we reject H*o* and conclude H*a*. Our final equation is:

Log Odds of Levee Failure = *2.174* + *2.350***seds** - *.9298***mea** - *.000003568***dredge**

REFERENCES

Roger, J. David, Phd, P.E., R.G.. "Evolution of the Levee System Along the Lower Mississippi River." *Natural Hazards Mitigation Institute Department of Geological Engineering University of Missouri-Rolla*. https://web.mst.edu/~rogersda/levees/Evolution%20of%20the%20Levee%20System%20Along%20the%20Mississippi.pdf

'Hurricane Katrina' (2020) Wikipedia. Available at: https://en.wikipedia.org/wiki/Hurricane_Katrina (Accessed: 28 June 2020)

'Great Mississipi Flood of 1927' (2020) Wikipedia. Available at: https://en.wikipedia.org/wiki/Great_Mississippi_Flood_of_1927 (Accessed: 28 June 2020)

Senol Cabi, Nalan, Phd. Weiner, Ann. Ma, Nan. *"The United States' Aging River Levees."* November 20, 2004. https://www.air-worldwide.com/publications/air-currents/2014/the-united-states-aging-river-levees/

Flor, Andrew. Pinter, Nicholas. Remo, Jonathan W.F.. "*Evaluating Levee Failure Susceptibility on the Mississippi River Using Logistic Regression Analysis.*" Engineering Geology: An International Journal. 31 March 2010

Link to Data: http://users.stat.ufl.edu/~winner/data/mmr_levee.dat
Link to Data Description: http://users.stat.ufl.edu/~winner/data/mmr_levee.txt

LINKS to IMAGES

Levee Failure
https://www.youtube.com/watch?v=aOAv02ivFz4
Meander
https://www.geol.umd.edu/~jmerck/geol100/lectures/29.html
Dredging
https://en.wikipedia.org/wiki/Dredging
Levee
https://www.researchgate.net/publication/319529755_FLOODS_NEW_CONCEPTS_EMERGE_-_OLD_PROBLEMS_REMAIN/figures?lo=1
Channel Width
https://www.wxxv25.com/2020/01/27/mississippi-river-water-levels-rising-army-corps-activate-phase-ii-flood-fight-procedures/
Floodway Width
https://www.fema.gov/media-library-data/1578062957793-0274cb6a7a3801a07a3db7916e64e80d/FloodwayAnalysis_and_Mapping_Nov_2019.pdf
Revetment
https://en.wikipedia.org/wiki/Revetment
Borrow Pit
https://www.researchgate.net/publication/301469590_EXAMPLES_OF_ALTERNATIVE_METHODS_IN_INTENSIFICATION_OF_DESLUDGING_SITES/figures?lo=1

Sinuosity
https://www.google.com/search?q=Sinuosity+mississipi+river&tbm=isch&ved=2ahUKEwjBhIGt_oHqAhXMFN8KHcCdCdkQ2-cCegQIABAA&oq=Sinuosity+mississipi+river&gs_lcp=CgNpbWcQAzoCCAA6BggAEAgQHjoGCAAQChAYOgQIABAYUK7KBFjEkgVg3pMFaARwAHgAgAG6AYgBogmSAQQxNy4xmAEAoAEBqgELZ3dzLXdpei1pbWc&sclient=img&ei=WnPmXsFvzKn8BsC7psgN&bih=812&biw=1440&client=safari&safe=active#imgrc=8Dl0HTO7WMHuXM
River Mile
https://www.umesc.usgs.gov/data_library/aqa_feat_bath_str/river_miles.html
Vegetation Width
https://www.co.pierce.wa.us/ArchiveCenter/ViewFile/Item/4622
Sediments
http://thebritishgeographer.weebly.com/river-processes.html