Project_Classification

Mason Cushing

2023-03-23

We begin by dividing the data into train and test sections. Note that the knn function requires different data than the other functions do, so I have two sets of divisions.

My data was too large, so I took a subset of it. It gives data about a flight and whether it was delayed. I will be trying to predict whether a flight was delayed based on this data.

The data is found here: https://www.kaggle.com/datasets/ulrikthygepedersen/airlines-delay (https://www.kaggle.com/datasets/ulrikthygepedersen/airlines-delay)

```
airlines_delay <- read.csv("~/Downloads/airlines_delay.csv")
df<-airlines_delay[airlines_delay$Airline == "DL",]
df<-subset(df,select=-c(Flight,Airline))
df$AirportFrom<-as.factor(df$AirportFrom)
df$AirportTo<-as.factor(df$AirportTo)
df$DayOfWeek<-as.factor(df$DayOfWeek)
df$Class<-as.factor(df$Class)
kdf<-subset(df,select=-c(Class))
kdf$AirportFrom<-as.numeric(kdf$AirportFrom)
kdf$AirportTo<-as.numeric(kdf$AirportTo)
kdf$DayOfWeek<-as.numeric(kdf$DayOfWeek)
set.seed(1969)
t <- sample(nrow(df),50000)
train<-df[t,]
test<-df[-t,]
ktrain<-kdf[t,]
ktest<-kdf[-t,]
```

We will then explore this data statistically and graphically.

```
mean(as.numeric(df$Class)-1)
```

```
## [1] 0.4504759
```

```
mean(df$Time)
```

```
## [1] 799.6362
```

```
median(df$Time)
```
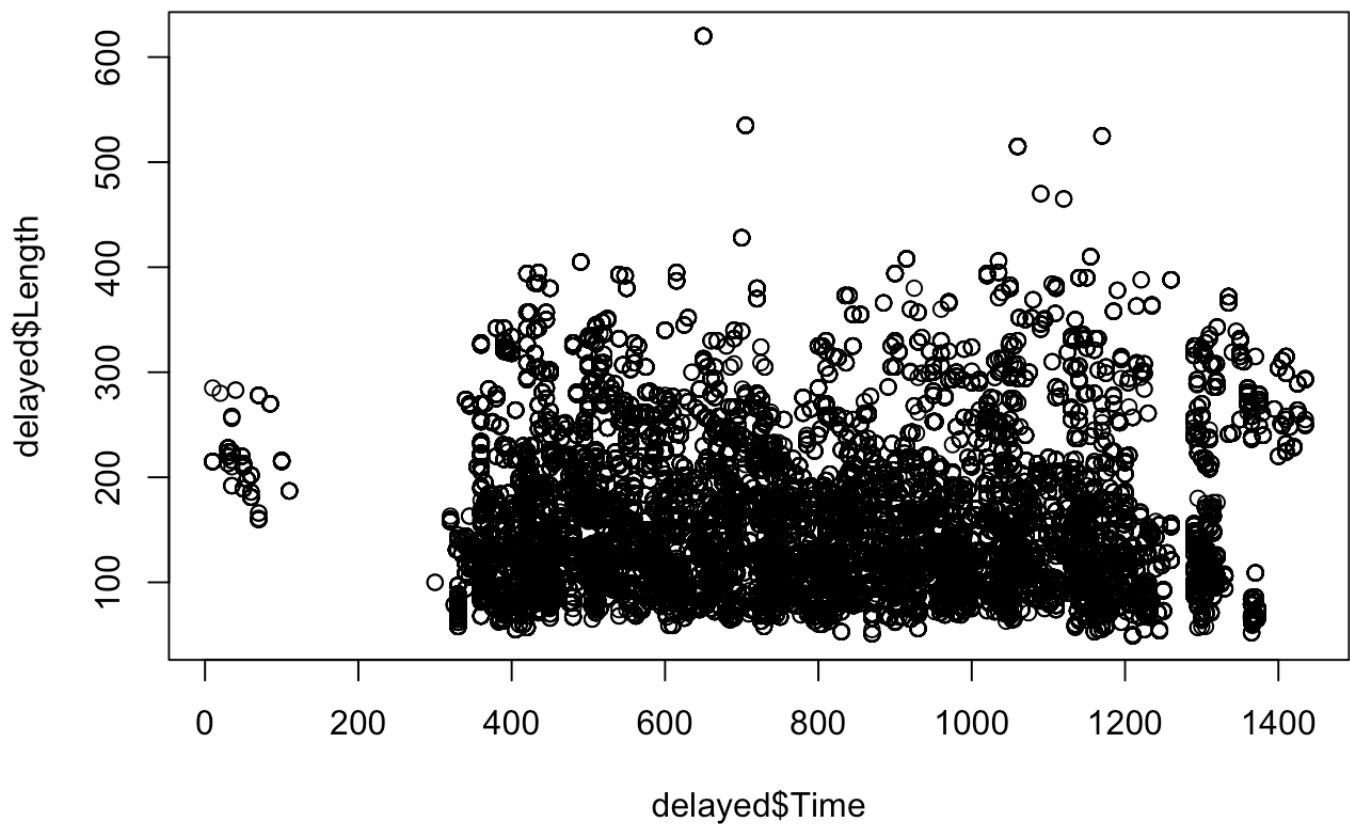
```
## [1] 795
```

```
mean(df$Length)
```
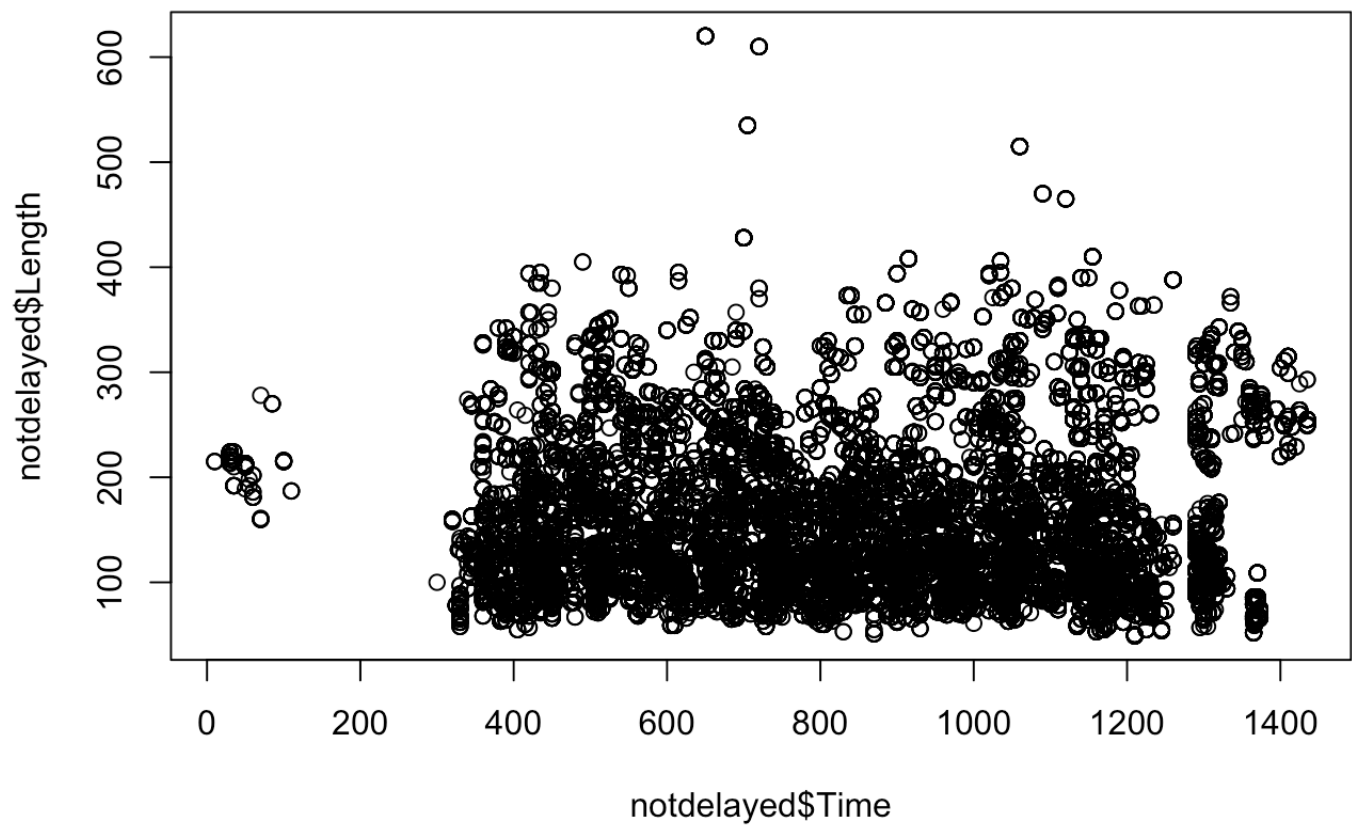
```
## [1] 157.5205
```
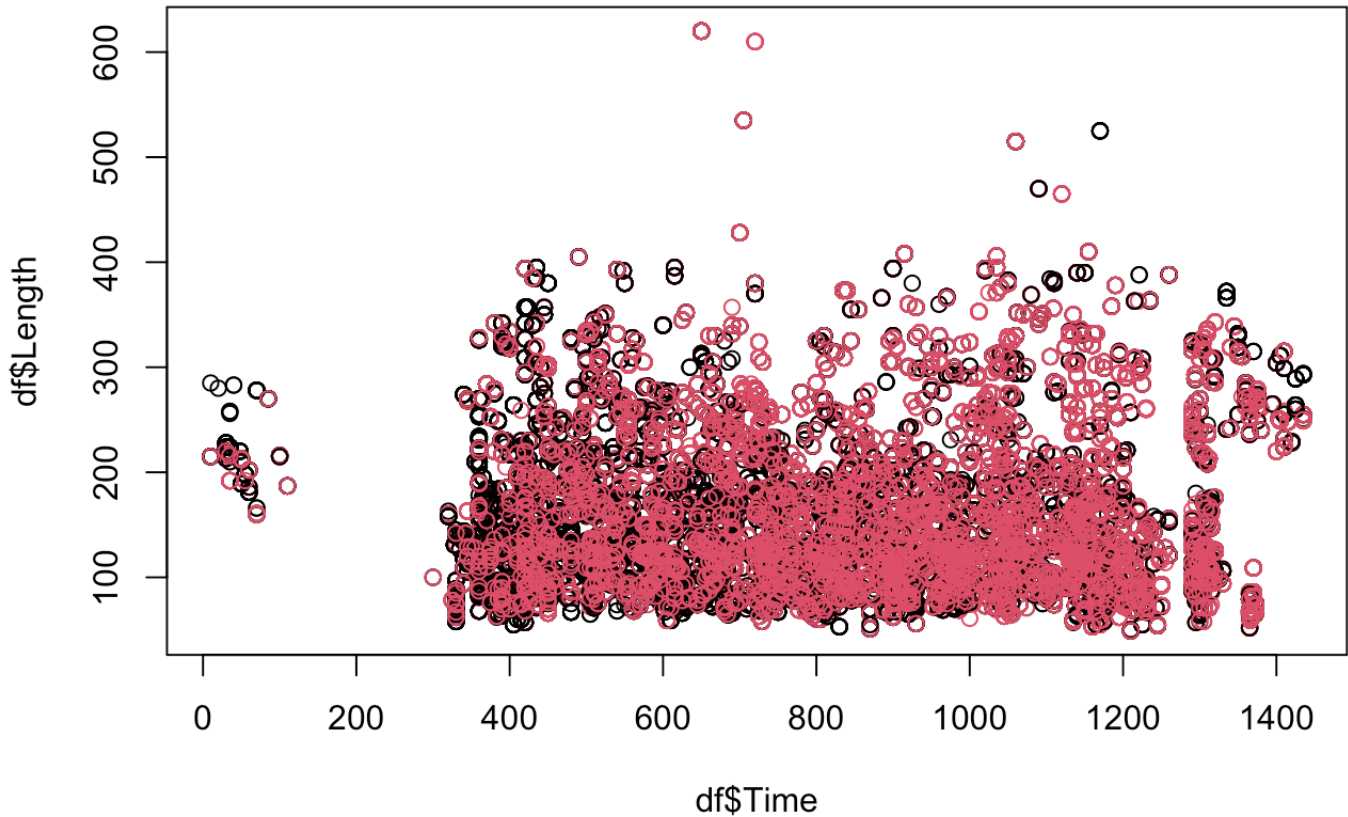
```
median(df$Length)
```

```
## [1] 134
```

```
delayed<-df[df$Class==0,]
notdelayed<-df[df$Class==1,]
plot(delayed$Time,delayed$Length)
```



```
plot(notdelayed$Time,notdelayed$Length)
```

```
plot(df$Time,df$Length,col=df$Class)
```

```
liechtenstein<-df[df$AirportFrom == "LIH" | df$AirportTo == "LIH",]
indonesia<-df[df$AirportFrom == "KOA" | df$AirportTo == "KOA",]
mean(as.numeric(liechtenstein$Class)-1)
```

```
## [1] 0.3064516
```

```
mean(as.numeric(indonesia$Class)-1)
```

```
## [1] 0.2741935
```

Note that the graphs look very similar for what we are trying to distinguish based on time and length. I would not expect accuracy to be great based on this data.

However, some airports are oddly consistent with not delaying flights, and Liechtenstein and Indonesia are among those airports. Average rates are shown.

We will next predict whether the flight was delayed using logistic regression, kNN, and decision trees in that order below.

```
model1<-glm(Class~.,data=train,family=binomial)
lpred<-predict(model1,newdata=test)
lpred[lpred<=0]<-0
lpred[lpred>=1]<-1
lpred<-round(lpred)
mean(lpred==test$Class)
```

```
## [1] 0.595064
```

```
library(class)
kmodel<-knn(train=ktrain,test=ktest,cl=train$Class,k=3)
kmodel<-as.numeric(kmodel)
mean((kmodel-1)==test$Class)
```

```
## [1] 0.6037477
```

```
library(rpart)
dtree<-rpart(Class~., data=train,method="class")
dpred<-predict(dtree,newdata=test,type="class")
mean(dpred==test$Class)
```

```
## [1] 0.6095064
```

```
summary(model1)
```

```
##
## Call:
## glm(formula = Class ~ ., family = binomial, data = train)
##
## Deviance Residuals:
##      Min       1Q    Median       3Q       Max
## -2.0265  -1.0427   -0.7242   1.1469    2.2255
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.575e+00  3.020e-01  -5.214 1.85e-07 ***
## Time              1.431e-03  3.621e-05  39.515  < 2e-16 ***
## Length            2.109e-04  2.079e-04   1.014 0.310537
## AirportFromALB    3.731e-01  4.533e-01   0.823 0.410428
## AirportFromANC   -2.340e-01  3.899e-01  -0.600 0.548454
## AirportFromATL    4.051e-01  2.125e-01   1.906 0.056621 .
## AirportFromAUS   -7.579e-01  2.733e-01  -2.773 0.005557 **
## AirportFromBDL    3.959e-01  2.354e-01   1.682 0.092577 .
## AirportFromBHM   -1.731e-02  3.373e-01  -0.051 0.959055
## AirportFromBIL    5.829e-01  6.626e-01   0.880 0.378992
```

```
## AirportFromBIS    1.698e-01   5.600e-01    0.303 0.761761
## AirportFromBNA    4.277e-01   2.659e-01    1.609 0.107720
## AirportFromBOI   -7.986e-01   3.604e-01   -2.216 0.026721 *
## AirportFromBOS    8.869e-02   2.211e-01    0.401 0.688320
## AirportFromBUF    4.671e-01   2.790e-01    1.674 0.094038 .
## AirportFromBWI   -1.004e-01   2.285e-01   -0.439 0.660519
## AirportFromBZN   -3.224e-01   4.178e-01   -0.772 0.440335
## AirportFromCAE   -5.426e-01   5.419e-01   -1.001 0.316724
## AirportFromCHS   -6.778e-01   2.813e-01   -2.410 0.015972 *
## AirportFromCLE    4.691e-02   7.412e-01    0.063 0.949538
## AirportFromCLT   -5.768e-02   2.357e-01   -0.245 0.806655
## AirportFromCMH   -3.641e-01   2.645e-01   -1.376 0.168681
## AirportFromCOS    8.534e-01   4.343e-01    1.965 0.049417 *
## AirportFromCVG    4.253e-01   2.238e-01    1.900 0.057458 .
## AirportFromDAB    2.853e-01   3.144e-01    0.908 0.364111
## AirportFromDAY   -1.482e-01   2.841e-01   -0.522 0.601931
## AirportFromDCA   -3.358e-02   2.206e-01   -0.152 0.879029
## AirportFromDEN    5.755e-01   2.289e-01    2.514 0.011932 *
## AirportFromDFW    1.319e-02   2.382e-01    0.055 0.955848
## AirportFromDTW    6.398e-01   2.137e-01    2.993 0.002759 **
## AirportFromECP   -7.291e-01   4.147e-01   -1.758 0.078715 .
## AirportFromEGE   -7.970e-01   7.011e-01   -1.137 0.255636
## AirportFromELP   -7.623e-01   5.878e-01   -1.297 0.194707
## AirportFromEWR    3.210e-01   2.323e-01    1.382 0.167077
## AirportFromEYW    2.631e-02   3.913e-01    0.067 0.946392
## AirportFromFAR    1.098e+00   5.083e-01    2.161 0.030688 *
## AirportFromFLL   -1.456e-01   2.218e-01   -0.657 0.511436
## AirportFromFNT    4.095e-01   3.361e-01    1.218 0.223034
## AirportFromFSD   -3.796e-01   8.458e-01   -0.449 0.653612
## AirportFromGEG    7.475e-01   3.069e-01    2.436 0.014868 *
## AirportFromGPT   -5.495e-01   3.817e-01   -1.440 0.149976
## AirportFromGRB    1.867e-01   4.706e-01    0.397 0.691541
## AirportFromGRR    2.956e-01   2.727e-01    1.084 0.278442
## AirportFromGSO   -1.018e+00   7.904e-01   -1.288 0.197700
## AirportFromGSP   -4.987e-02   4.974e-01   -0.100 0.920150
## AirportFromHDN    3.554e-01   6.174e-01    0.576 0.564883
## AirportFromHNL   -8.865e-01   2.626e-01   -3.376 0.000735 ***
## AirportFromHOU   -2.214e-01   3.679e-01   -0.602 0.547329
## AirportFromHSV   -2.000e-01   5.179e-01   -0.386 0.699345
## AirportFromIAD    6.445e-01   2.576e-01    2.502 0.012336 *
## AirportFromIAH   -2.631e-01   2.836e-01   -0.928 0.353602
## AirportFromIND    5.590e-02   2.350e-01    0.238 0.811948
## AirportFromJAC   -3.183e-01   3.543e-01   -0.898 0.368928
## AirportFromJAN   -1.532e+00   1.064e+00   -1.441 0.149713
## AirportFromJAX   -2.747e-01   2.427e-01   -1.132 0.257773
## AirportFromJFK   -2.993e-01   2.196e-01   -1.363 0.172851
## AirportFromKOA   -2.279e+00   5.419e-01   -4.206 2.60e-05 ***
## AirportFromLAS   -7.305e-02   2.218e-01   -0.329 0.741858
## AirportFromLAX    5.626e-01   2.188e-01    2.571 0.010137 *
## AirportFromLGA   -1.406e-01   2.165e-01   -0.649 0.516084
```

```
## AirportFromLIH -2.674e+00  6.571e-01  -4.069 4.72e-05 ***
## AirportFromLIT -3.466e-01  3.793e-01  -0.914 0.360847
## AirportFromMCI -3.160e-01  2.381e-01  -1.327 0.184354
## AirportFromMCO -2.699e-01  2.182e-01  -1.237 0.216126
## AirportFromMDW  1.013e+00  2.644e-01   3.831 0.000128 ***
## AirportFromMEM -1.168e-01  2.199e-01  -0.531 0.595367
## AirportFromMHT -8.853e-01  4.403e-01  -2.011 0.044362 *
## AirportFromMIA -1.556e-01  2.259e-01  -0.689 0.490921
## AirportFromMKE  1.279e-01  2.363e-01   0.541 0.588179
## AirportFromMLB -5.313e-01  3.315e-01  -1.602 0.109072
## AirportFromMSN -2.357e-01  3.529e-01  -0.668 0.504179
## AirportFromMSP  6.671e-01  2.140e-01   3.118 0.001821 **
## AirportFromMSY -5.479e-01  2.338e-01  -2.344 0.019103 *
## AirportFromMTJ  4.779e-01  6.449e-01   0.741 0.458685
## AirportFromOAK  4.044e-02  4.013e-01   0.101 0.919733
## AirportFromOGG -1.150e+00  3.692e-01  -3.114 0.001848 **
## AirportFromOKC -1.256e+00  6.472e-01  -1.941 0.052303 .
## AirportFromOMA  8.701e-02  3.369e-01   0.258 0.796196
## AirportFromONT -3.678e-01  5.223e-01  -0.704 0.481344
## AirportFromORD  1.047e+00  2.392e-01   4.375 1.21e-05 ***
## AirportFromORF -4.058e-01  2.830e-01  -1.434 0.151689
## AirportFromPBI -2.665e-01  2.347e-01  -1.135 0.256197
## AirportFromPDX -9.726e-02  2.444e-01  -0.398 0.690687
## AirportFromPHF -3.055e-01  3.687e-01  -0.829 0.407359
## AirportFromPHL -1.388e-01  2.307e-01  -0.602 0.547362
## AirportFromPHX  1.484e-02  2.263e-01   0.066 0.947720
## AirportFromPIT -1.304e-01  2.555e-01  -0.511 0.609687
## AirportFromPNS  7.391e-03  2.607e-01   0.028 0.977382
## AirportFromPSP -9.961e+00  1.195e+02  -0.083 0.933552
## AirportFromPVD -3.983e-01  3.293e-01  -1.210 0.226369
## AirportFromPWM -6.777e-02  4.038e-01  -0.168 0.866701
## AirportFromRAP  4.994e-01  4.802e-01   1.040 0.298342
## AirportFromRDU -1.652e-01  2.388e-01  -0.692 0.489005
## AirportFromRIC -2.115e-01  2.616e-01  -0.809 0.418673
## AirportFromRNO -6.407e-01  3.854e-01  -1.662 0.096430 .
## AirportFromROC  1.530e+00  4.616e-01   3.315 0.000917 ***
## AirportFromRSW -5.210e-01  2.337e-01  -2.230 0.025770 *
## AirportFromSAN -2.134e-01  2.348e-01  -0.909 0.363466
## AirportFromSAT -1.270e-02  2.449e-01  -0.052 0.958634
## AirportFromSAV -1.629e-01  2.724e-01  -0.598 0.549857
## AirportFromSDF -5.784e-02  3.657e-01  -0.158 0.874330
## AirportFromSEA  4.100e-01  2.262e-01   1.813 0.069894 .
## AirportFromSFO  4.057e-01  2.259e-01   1.796 0.072557 .
## AirportFromSJC  4.026e-01  3.512e-01   1.147 0.251584
## AirportFromSJU -8.932e-01  2.907e-01  -3.073 0.002120 **
## AirportFromSLC  6.285e-01  2.162e-01   2.907 0.003647 **
## AirportFromSMF  2.698e-01  2.703e-01   0.998 0.318137
## AirportFromSNA -4.799e-01  2.442e-01  -1.966 0.049355 *
## AirportFromSRQ -4.162e-01  2.693e-01  -1.546 0.122211
## AirportFromSTL  5.504e-01  2.408e-01   2.285 0.022294 *
```

```
## AirportFromSTT -1.834e+00  5.760e-01  -3.183 0.001455 **
## AirportFromSTX  1.357e+00  1.181e+00   1.150 0.250297
## AirportFromTLH -7.980e-01  5.879e-01  -1.357 0.174680
## AirportFromTPA -4.131e-01  2.241e-01  -1.844 0.065224 .
## AirportFromTUL  6.821e-01  4.350e-01   1.568 0.116873
## AirportFromTUS  2.018e-01  3.207e-01   0.629 0.529200
## AirportFromTYS -9.669e-01  7.904e-01  -1.223 0.221232
## AirportToALB   -1.239e-01  4.821e-01  -0.257 0.797236
## AirportToANC    5.265e-01  3.668e-01   1.435 0.151222
## AirportToATL    4.153e-01  2.108e-01   1.971 0.048762 *
## AirportToAUS   -4.129e-01  2.588e-01  -1.596 0.110555
## AirportToBDL   -1.276e-01  2.351e-01  -0.543 0.587330
## AirportToBHM    5.717e-02  3.173e-01   0.180 0.857029
## AirportToBIL    3.055e-01  6.424e-01   0.476 0.634368
## AirportToBIS   -1.436e+00  7.019e-01  -2.046 0.040747 *
## AirportToBNA   -4.102e-01  2.675e-01  -1.533 0.125231
## AirportToBOI    2.218e-01  3.337e-01   0.665 0.506278
## AirportToBOS   -1.036e-01  2.196e-01  -0.472 0.637151
## AirportToBUF   -5.130e-01  2.861e-01  -1.793 0.072953 .
## AirportToBWI   -4.044e-01  2.278e-01  -1.775 0.075909 .
## AirportToBZN    3.053e-01  4.048e-01   0.754 0.450769
## AirportToCAE   -1.565e+00  5.518e-01  -2.835 0.004576 **
## AirportToCHS   -4.546e-01  2.677e-01  -1.698 0.089526 .
## AirportToCLE    1.065e+00  8.197e-01   1.299 0.193901
## AirportToCLT   -4.710e-01  2.354e-01  -2.001 0.045441 *
## AirportToCMH   -5.072e-01  2.581e-01  -1.965 0.049426 *
## AirportToCOS    6.279e-01  5.006e-01   1.254 0.209710
## AirportToCVG    2.858e-01  2.229e-01   1.282 0.199739
## AirportToDAB    2.260e-01  3.192e-01   0.708 0.478793
## AirportToDAY   -5.927e-01  2.881e-01  -2.057 0.039678 *
## AirportToDCA   -1.746e-01  2.198e-01  -0.794 0.426992
## AirportToDEN    3.938e-01  2.266e-01   1.738 0.082203 .
## AirportToDFW    2.219e-02  2.343e-01   0.095 0.924556
## AirportToDTW    3.436e-01  2.120e-01   1.621 0.105024
## AirportToECP   -8.787e-01  4.380e-01  -2.006 0.044832 *
## AirportToEGE   -2.718e-01  6.386e-01  -0.426 0.670350
## AirportToELP    4.015e-01  4.866e-01   0.825 0.409340
## AirportToEWR    1.610e-01  2.305e-01   0.698 0.484933
## AirportToEYW   -4.066e-01  4.190e-01  -0.970 0.331918
## AirportToFAR    2.514e-02  4.615e-01   0.054 0.956563
## AirportToFLL    4.397e-01  2.196e-01   2.002 0.045274 *
## AirportToFNT    3.902e-01  3.498e-01   1.115 0.264647
## AirportToFSD    1.615e-02  7.644e-01   0.021 0.983150
## AirportToGEG    4.049e-01  3.104e-01   1.305 0.191990
## AirportToGPT   -6.084e-01  3.943e-01  -1.543 0.122832
## AirportToGRB   -1.315e-01  4.899e-01  -0.268 0.788414
## AirportToGRR    4.440e-01  2.726e-01   1.629 0.103320
## AirportToGSO    2.990e-01  5.985e-01   0.500 0.617403
## AirportToGSP    1.948e-01  4.464e-01   0.436 0.662614
## AirportToHDN    3.054e-01  7.052e-01   0.433 0.665012
```

```
## AirportToHNL      9.260e-01  2.664e-01   3.477 0.000507 ***
## AirportToHOU     -3.810e-01  3.703e-01  -1.029 0.303520
## AirportToHSV     -8.108e-02  4.980e-01  -0.163 0.870670
## AirportToIAD     -1.443e-01  2.575e-01  -0.560 0.575147
## AirportToIAH     -3.247e-01  2.767e-01  -1.173 0.240653
## AirportToIND     -1.799e-01  2.327e-01  -0.773 0.439437
## AirportToJAC      2.911e-01  3.454e-01   0.843 0.399360
## AirportToJAN     -8.670e-01  5.377e-01  -1.613 0.106847
## AirportToJAX     -3.646e-02  2.390e-01  -0.153 0.878752
## AirportToJFK     -3.952e-02  2.171e-01  -0.182 0.855529
## AirportToKOA     -3.737e-01  4.819e-01  -0.776 0.438003
## AirportToLAS      2.943e-01  2.190e-01   1.344 0.178891
## AirportToLAX      6.620e-01  2.170e-01   3.051 0.002278 **
## AirportToLGA      3.700e-02  2.153e-01   0.172 0.863540
## AirportToLIH     -1.132e-01  4.392e-01  -0.258 0.796657
## AirportToLIT     -2.223e-01  3.300e-01  -0.674 0.500585
## AirportToMCI     -2.449e-01  2.343e-01  -1.045 0.295953
## AirportToMCO      3.179e-01  2.158e-01   1.473 0.140789
## AirportToMDW      2.782e-01  2.605e-01   1.068 0.285558
## AirportToMEM      2.797e-01  2.183e-01   1.281 0.200120
## AirportToMHT     -8.460e-01  3.656e-01  -2.314 0.020677 *
## AirportToMIA      1.662e-02  2.242e-01   0.074 0.940910
## AirportToMKE      5.069e-02  2.345e-01   0.216 0.828852
## AirportToMLB     -2.844e-01  3.022e-01  -0.941 0.346604
## AirportToMSN     -2.377e-01  3.374e-01  -0.704 0.481152
## AirportToMSP      5.949e-01  2.121e-01   2.805 0.005030 **
## AirportToMSY     -4.413e-02  2.286e-01  -0.193 0.846902
## AirportToMTJ     -6.805e-02  6.629e-01  -0.103 0.918232
## AirportToOAK      1.116e+00  4.264e-01   2.616 0.008884 **
## AirportToOGG     -1.868e-01  3.628e-01  -0.515 0.606656
## AirportToOKC     -6.182e-01  4.418e-01  -1.399 0.161735
## AirportToOMA     -6.920e-01  3.342e-01  -2.070 0.038414 *
## AirportToONT      1.054e+00  5.986e-01   1.760 0.078332 .
## AirportToORD      1.724e-01  2.360e-01   0.731 0.464993
## AirportToORF     -2.492e-01  2.714e-01  -0.918 0.358657
## AirportToPBI      2.448e-01  2.316e-01   1.057 0.290501
## AirportToPDX      5.626e-01  2.425e-01   2.320 0.020361 *
## AirportToPHF     -5.399e-01  3.584e-01  -1.507 0.131931
## AirportToPHL     -2.319e-01  2.288e-01  -1.014 0.310629
## AirportToPHX      3.913e-01  2.243e-01   1.745 0.081051 .
## AirportToPIT     -2.619e-01  2.532e-01  -1.034 0.301064
## AirportToPNS     -3.377e-01  2.605e-01  -1.297 0.194763
## AirportToPSP      1.136e+01  1.195e+02   0.095 0.924248
## AirportToPVD     -1.064e-01  3.308e-01  -0.322 0.747829
## AirportToPWM     -1.705e+00  4.939e-01  -3.452 0.000557 ***
## AirportToRAP     -1.291e+00  4.688e-01  -2.755 0.005876 **
## AirportToRDU      3.109e-03  2.364e-01   0.013 0.989507
## AirportToRIC     -3.020e-01  2.572e-01  -1.174 0.240348
## AirportToRNO      1.812e-01  3.659e-01   0.495 0.620400
## AirportToROC     -5.750e-01  4.179e-01  -1.376 0.168806
```

```
## AirportToRSW     -2.082e-02  2.282e-01   -0.091 0.927317
## AirportToSAN      2.289e-01  2.318e-01    0.987 0.323460
## AirportToSAT      1.505e-02  2.403e-01    0.063 0.950053
## AirportToSAV     -5.080e-01  2.678e-01   -1.897 0.057847 .
## AirportToSDF      2.934e-01  3.675e-01    0.798 0.424765
## AirportToSEA      5.579e-01  2.252e-01    2.477 0.013248 *
## AirportToSFO      7.408e-01  2.252e-01    3.289 0.001005 **
## AirportToSJC      7.107e-01  3.510e-01    2.025 0.042916 *
## AirportToSJU      5.006e-01  2.766e-01    1.810 0.070311 .
## AirportToSLC      7.055e-01  2.144e-01    3.290 0.001001 **
## AirportToSMF      2.235e-01  2.742e-01    0.815 0.414973
## AirportToSNA     -2.041e-01  2.407e-01   -0.848 0.396448
## AirportToSRQ     -2.185e-01  2.648e-01   -0.825 0.409106
## AirportToSTL     -2.876e-01  2.401e-01   -1.198 0.230904
## AirportToSTT      8.046e-01  4.260e-01    1.889 0.058935 .
## AirportToSTX      1.133e+01  8.351e+01    0.136 0.892107
## AirportToTLH     -3.741e-01  4.474e-01   -0.836 0.403130
## AirportToTPA      3.017e-01  2.199e-01    1.372 0.170089
## AirportToTUL     -6.332e-01  4.900e-01   -1.292 0.196293
## AirportToTUS      9.516e-01  3.461e-01    2.749 0.005974 **
## AirportToTYS     -5.742e-01  5.812e-01   -0.988 0.323207
## DayOfWeek2       -2.352e-01  3.621e-02   -6.494 8.34e-11 ***
## DayOfWeek3       -2.174e-01  3.423e-02   -6.351 2.14e-10 ***
## DayOfWeek4       -2.890e-01  3.412e-02   -8.471  < 2e-16 ***
## DayOfWeek5       -4.934e-01  3.483e-02  -14.168  < 2e-16 ***
## DayOfWeek6       -7.209e-01  3.932e-02  -18.335  < 2e-16 ***
## DayOfWeek7       -1.424e-01  3.626e-02   -3.928 8.56e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 68810  on 49999  degrees of freedom
## Residual deviance: 64307  on 49763  degrees of freedom
## AIC: 64781
##
## Number of Fisher Scoring iterations: 9
```

```
summary(kmodel)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.427   2.000   2.000
```

```
plot(dtree,margin=0.2)
text(dtree)
```

```
## Warning in labels.rpart(x, minlength = minlength): more than 52 levels in a
## predicting factor, truncated for printout
```


Notice that the methods become progressively better, with logistic regression being the worst and decision trees being the most accurate, while they don't differ by more than two percentage points, so no method outperforms another by a significant margin.

Overall, it seems that no algorithm was able to get a great read on whether a flight is delayed. We would expect this given the data, if this were simple to predict with this data, the airlines would know about it and have problems they need to fix. More than likely, a weather field would have the information needed to get a higher accuracy than about sixty percent, but that wasn't given in the data. Thus, as these algorithms went through the data, as logistic regression looked for a linear classification separator, it was unable to find one, and decision trees would have ran into the same issue. The failure of kNN to find a trend reflects both the data's relative lack of one and the odd way its inner workings interact with factors, as they have to be arbitrarily ordered and kNN assigns their values mathematical meaning where there is none.