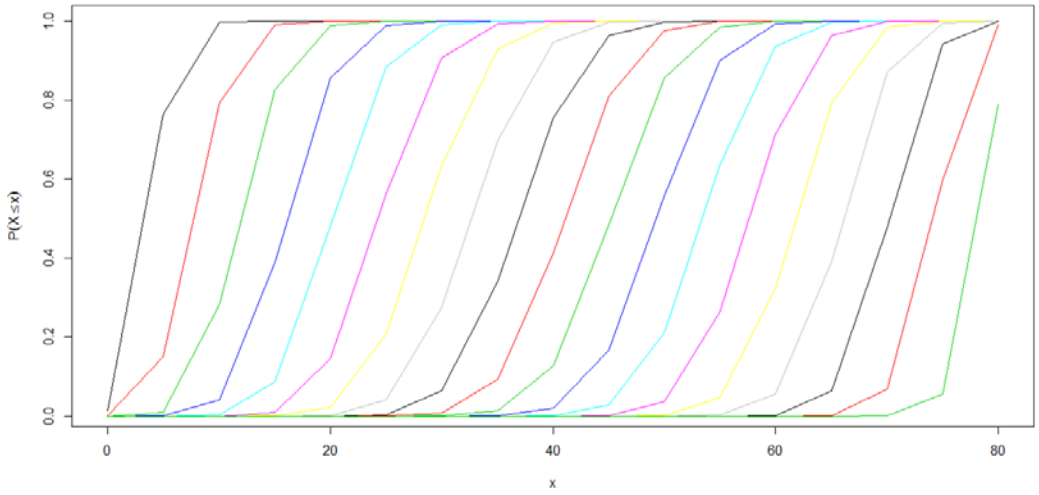
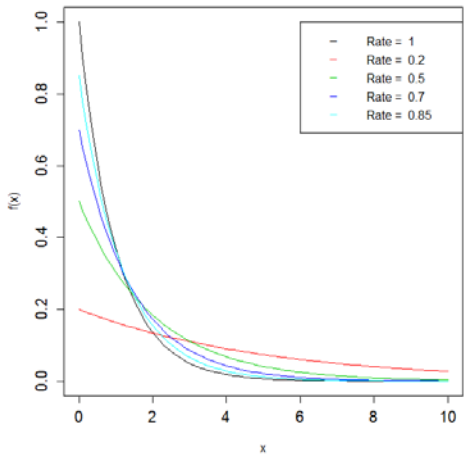
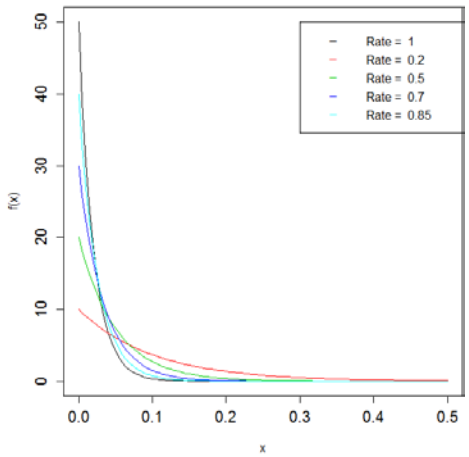
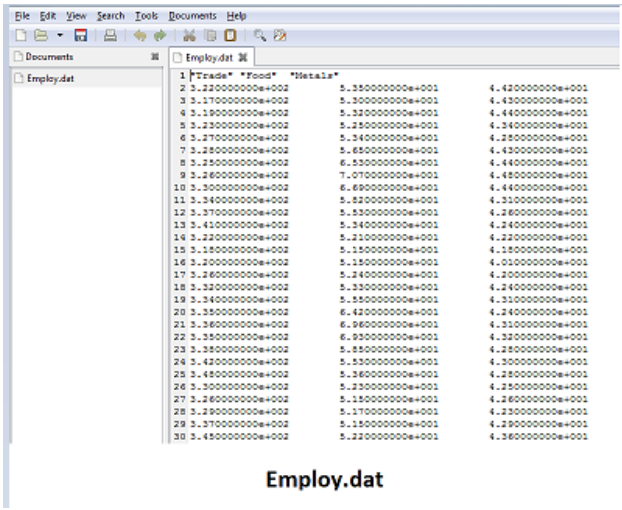
Chapter 1, Data Characteristics





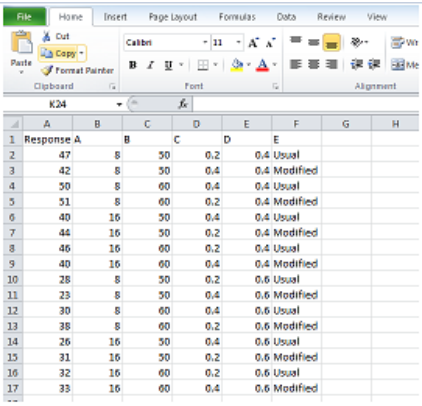


Chapter 2, Import/Export Data



Employ.dat

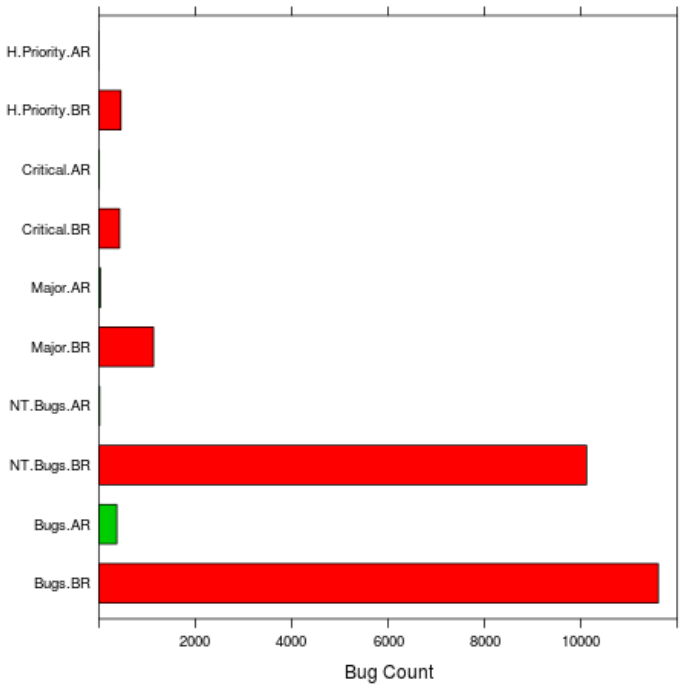
"Trade"	"Food"	"Metals"
3.220000000e+002	5.350000000e+001	4.420000000e+001
3.170000000e+002	5.300000000e+001	4.430000000e+001
3.190000000e+002	5.320000000e+001	4.440000000e+001
3.230000000e+002	5.280000000e+001	4.340000000e+001
3.270000000e+002	5.340000000e+001	4.280000000e+001
3.250000000e+002	5.650000000e+001	4.430000000e+001
3.250000000e+002	6.530000000e+001	4.440000000e+001
3.260000000e+002	7.070000000e+001	4.480000000e+001
3.300000000e+002	6.690000000e+001	4.440000000e+001
3.340000000e+002	5.820000000e+001	4.310000000e+001
3.370000000e+002	5.530000000e+001	4.260000000e+001
3.420000000e+002	5.340000000e+001	4.240000000e+001
3.220000000e+002	5.210000000e+001	4.220000000e+001
3.150000000e+002	5.150000000e+001	4.180000000e+001
3.200000000e+002	5.150000000e+001	4.010000000e+001
3.260000000e+002	5.240000000e+001	4.200000000e+001
3.320000000e+002	5.330000000e+001	4.240000000e+001
3.340000000e+002	5.550000000e+001	4.310000000e+001
3.350000000e+002	6.420000000e+001	4.240000000e+001
3.360000000e+002	6.960000000e+001	4.310000000e+001
3.350000000e+002	6.930000000e+001	4.320000000e+001
3.350000000e+002	5.850000000e+001	4.280000000e+001
3.420000000e+002	5.530000000e+001	4.300000000e+001
3.450000000e+002	5.360000000e+001	4.280000000e+001
3.300000000e+002	5.230000000e+001	4.250000000e+001
3.260000000e+002	5.150000000e+001	4.260000000e+001
3.290000000e+002	5.170000000e+001	4.230000000e+001
3.370000000e+002	5.150000000e+001	4.290000000e+001
3.450000000e+002	5.220000000e+001	4.360000000e+001

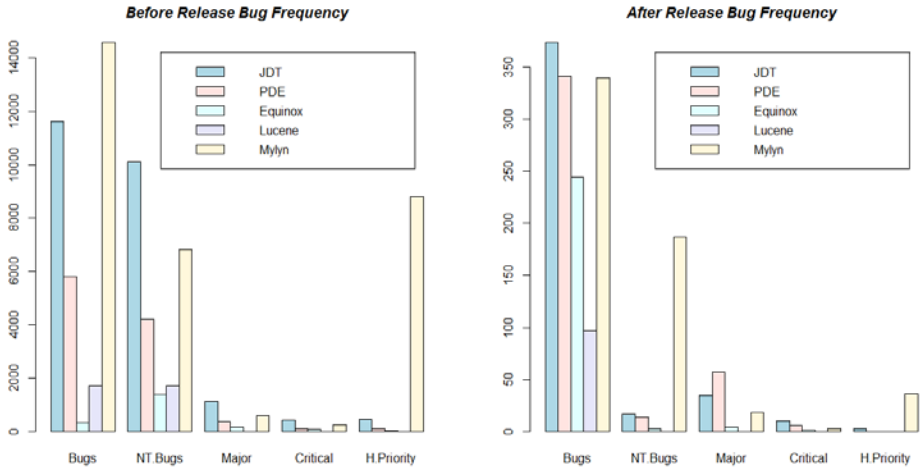


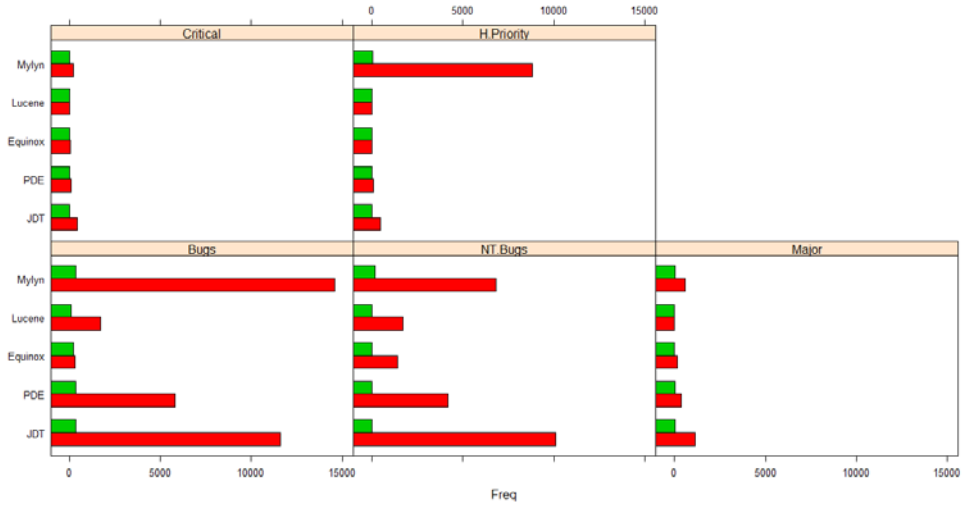
SCV.csv

Response	A	B	C	D	E
47	8	50	0.2	0.4	Usual
42	8	50	0.4	0.4	Modified
50	8	60	0.4	0.4	Usual
51	8	60	0.2	0.4	Modified
40	16	50	0.4	0.4	Usual
44	16	50	0.2	0.4	Modified
46	16	60	0.2	0.4	Usual
40	16	60	0.4	0.4	Modified
28	8	50	0.2	0.6	Usual
23	8	50	0.4	0.6	Modified
30	8	60	0.4	0.6	Usual
38	8	60	0.2	0.6	Modified
26	16	50	0.4	0.6	Usual
31	16	50	0.2	0.6	Modified
32	16	60	0.2	0.6	Usual
33	16	60	0.4	0.6	Modified

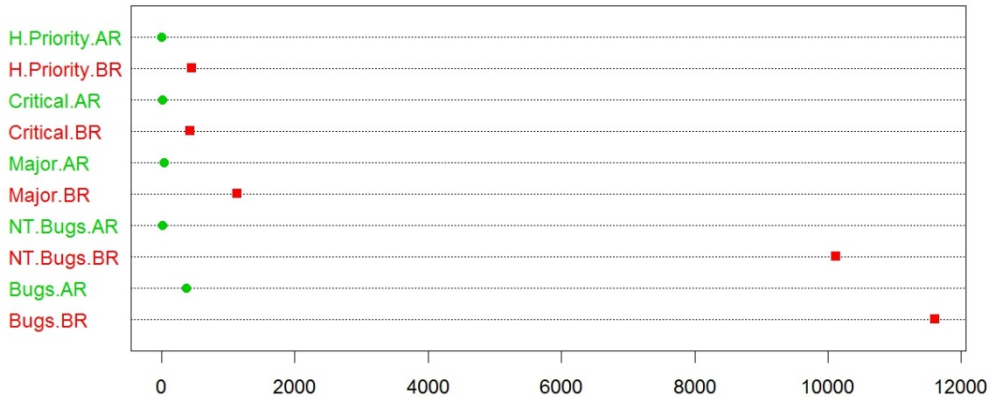
Chapter 3, Data Visualization

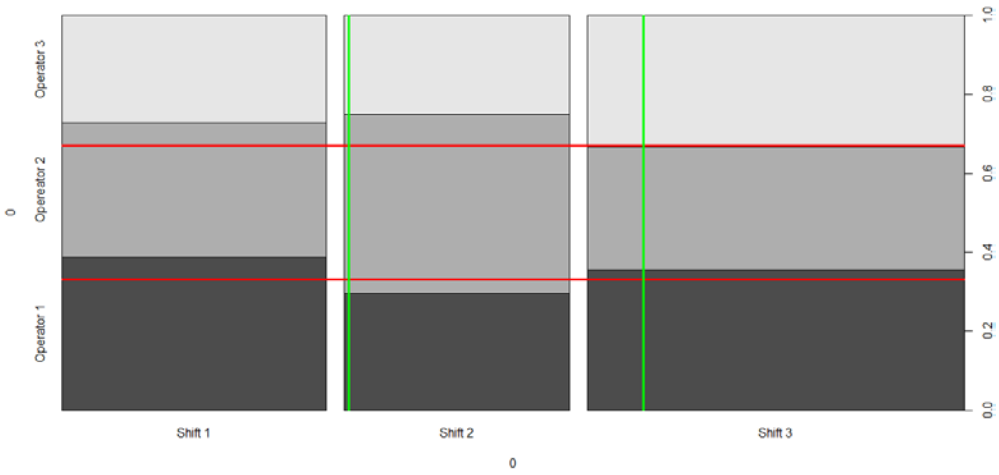


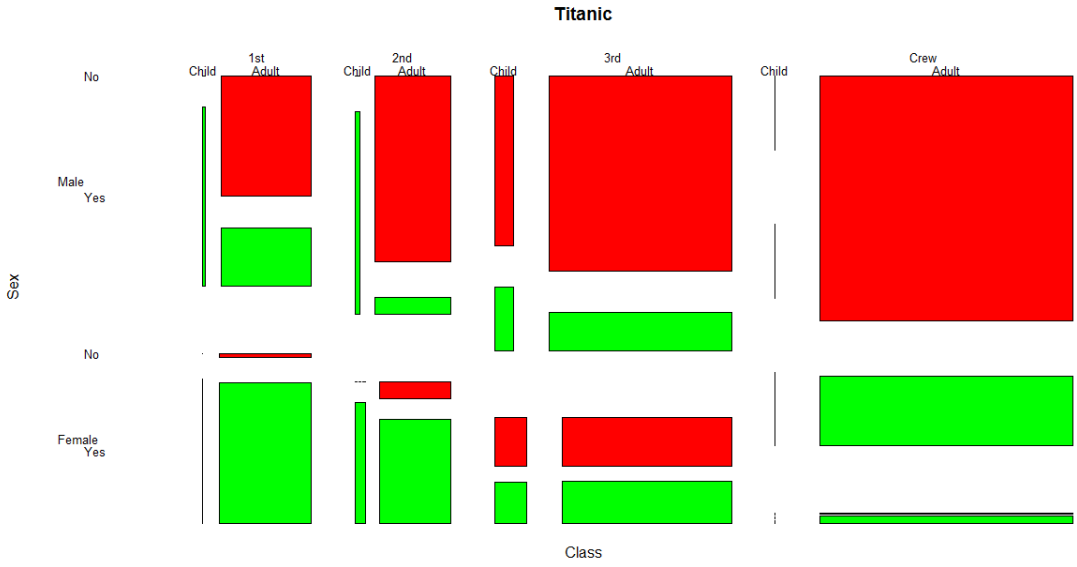


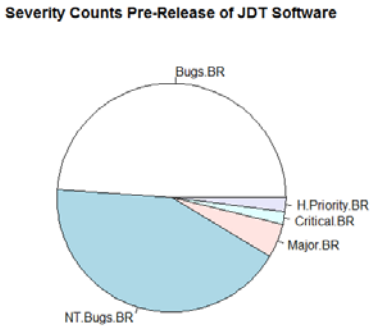


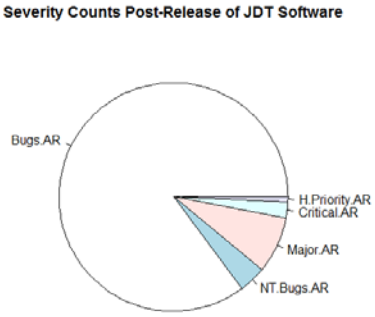
Dot Plot for the Before and After Release Bug Frequency

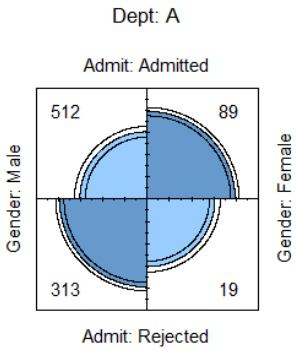


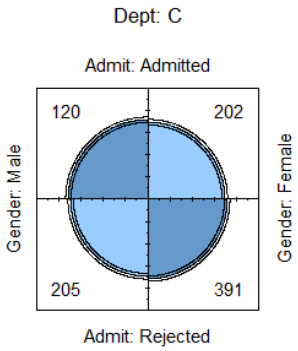

Operator 3
Operator 2
Operator 1
0
Shift 1
Shift 2
Shift 3
0
0.0
0.2
0.4
0.6
0.8
1.0

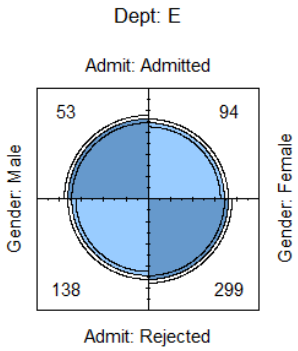

Titanic
1st
2nd
3rd
Crew
Child
Adult
No
Male
Yes
No
Female
Yes
Sex
Class

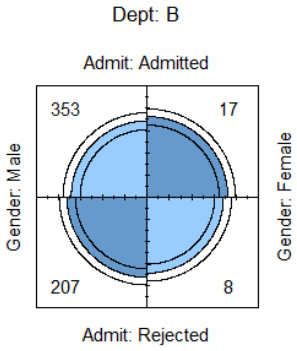

Severity Counts Pre-Release of JDT Software
Bugs.BR
H.Priority.BR
Critical.BR
Major.BR
NT.Bugs.BR

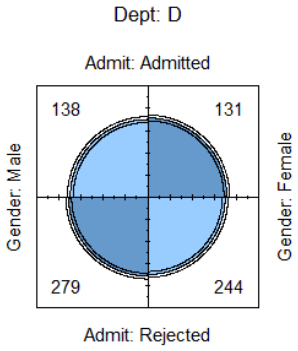

Severity Counts Post-Release of JDT Software
Bugs.AR
H.Priority.AR
Critical.AR
Major.AR
NT.Bugs.AR

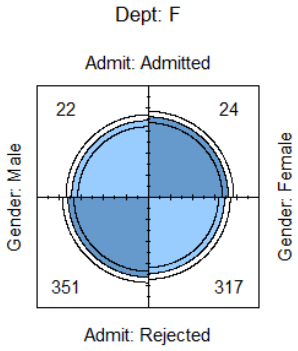

Dept: A
Admit: Admitted
512
89
313
19
Gender: Male
Gender: Female
Admit: Rejected

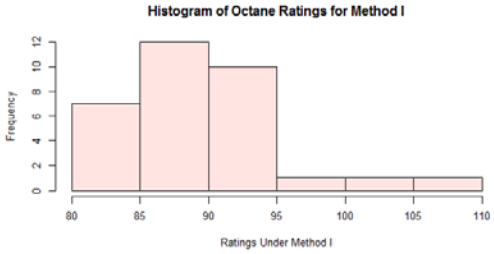

Dept: C
Admit: Admitted
120
202
205
391
Gender: Male
Gender: Female
Admit: Rejected

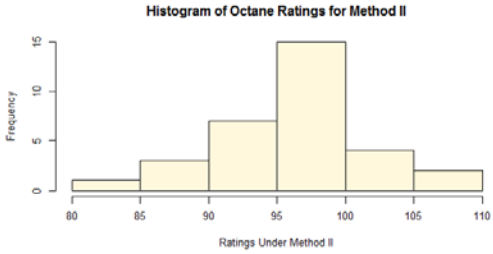

Dept: E
Admit: Admitted
53
94
138
299
Gender: Male
Gender: Female
Admit: Rejected

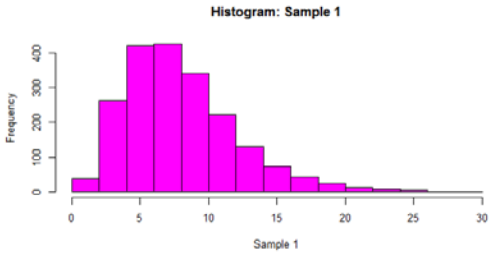

Dept: B
Admit: Admitted
353
17
207
8
Gender: Male
Gender: Female
Admit: Rejected

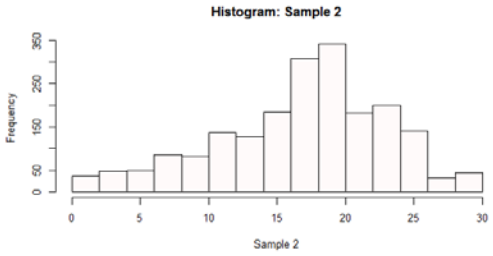

Dept: D
Admit: Admitted
138
131
279
244
Gender: Male
Gender: Female
Admit: Rejected

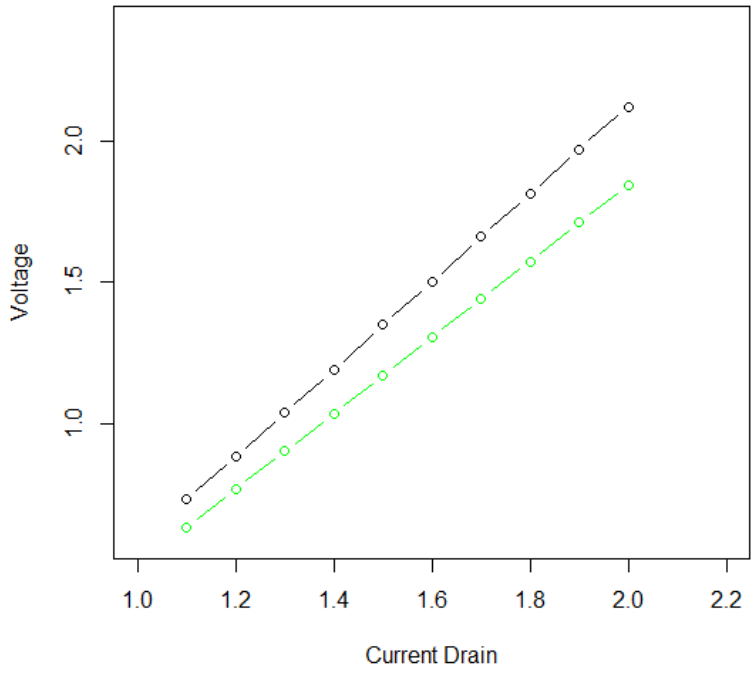

Dept: F
Admit: Admitted
22
24
351
317
Gender: Male
Gender: Female
Admit: Rejected

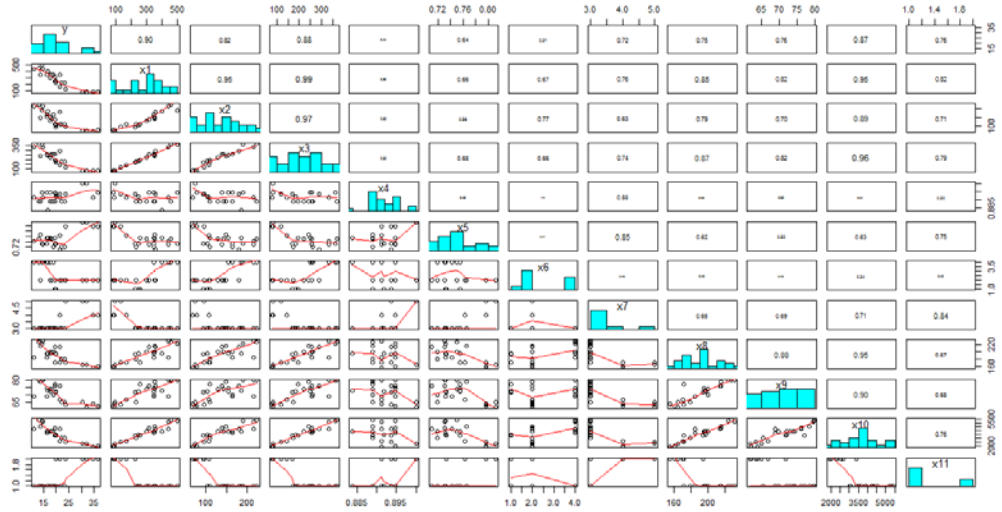

Histogram of Octane Ratings for Method I
Frequency
0
2
4
6
8
10
12
80
85
90
95
100
105
110
Ratings Under Method I

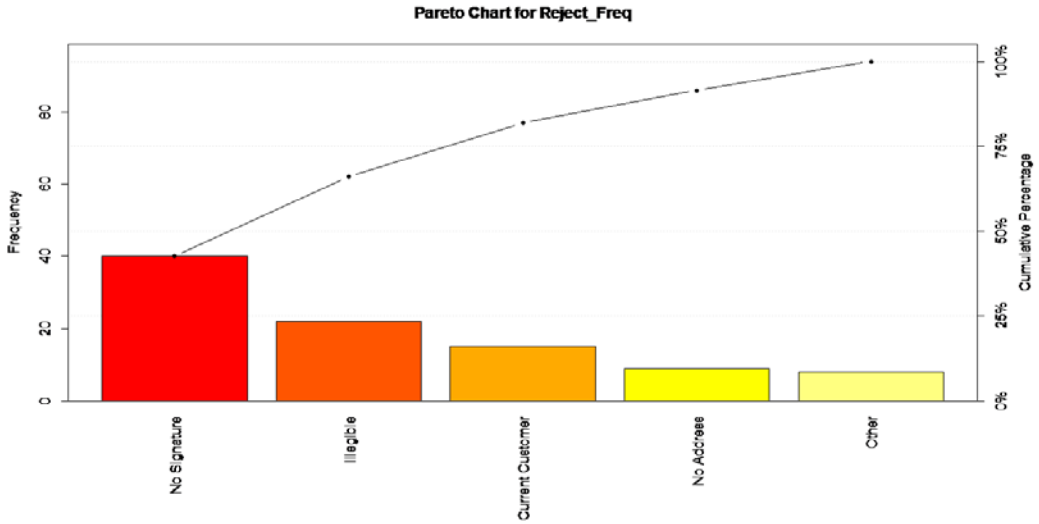

Histogram of Octane Ratings for Method II
Frequency
0
5
10
15
80
85
90
95
100
105
110
Ratings Under Method II


Histogram: Sample 1
Frequency
0
100
200
300
400
0
5
10
15
20
25
30
Sample 1

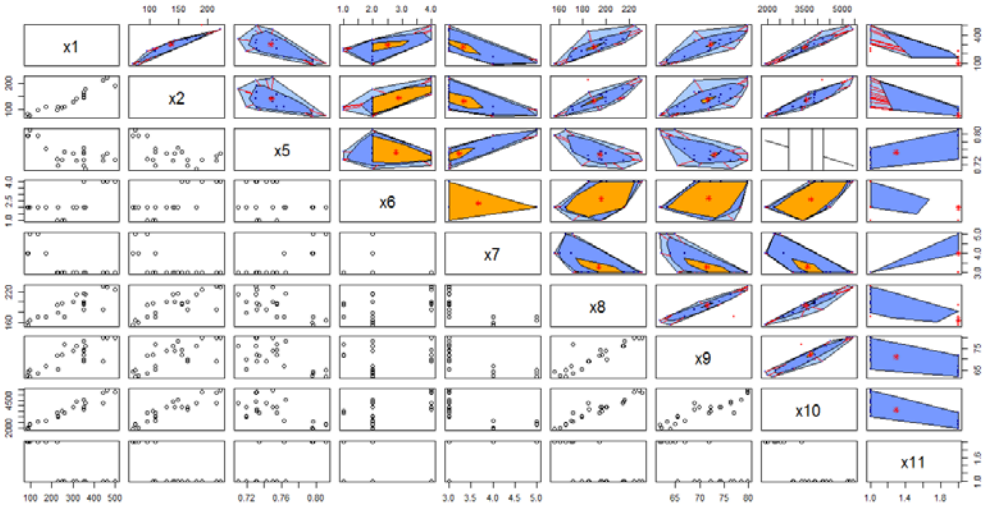

Histogram: Sample 2
Frequency
0
50
150
250
350
0
5
10
15
20
25
30
Sample 2

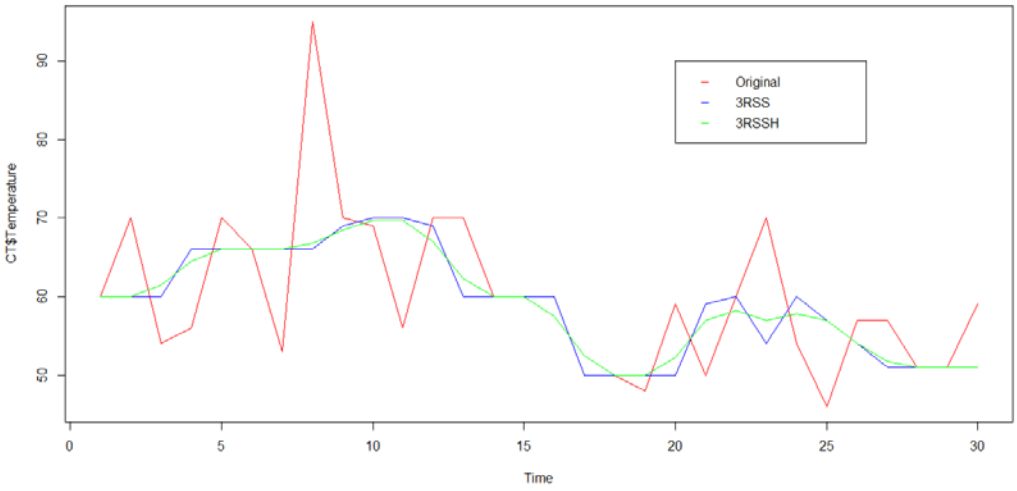

Voltage
1.0
1.5
2.0
1.0
1.2
1.4
1.6
1.8
2.0
2.2
Current Drain


y
x1
x2
x3
x4
x5
x6
x7
x8
x9
x10
x11

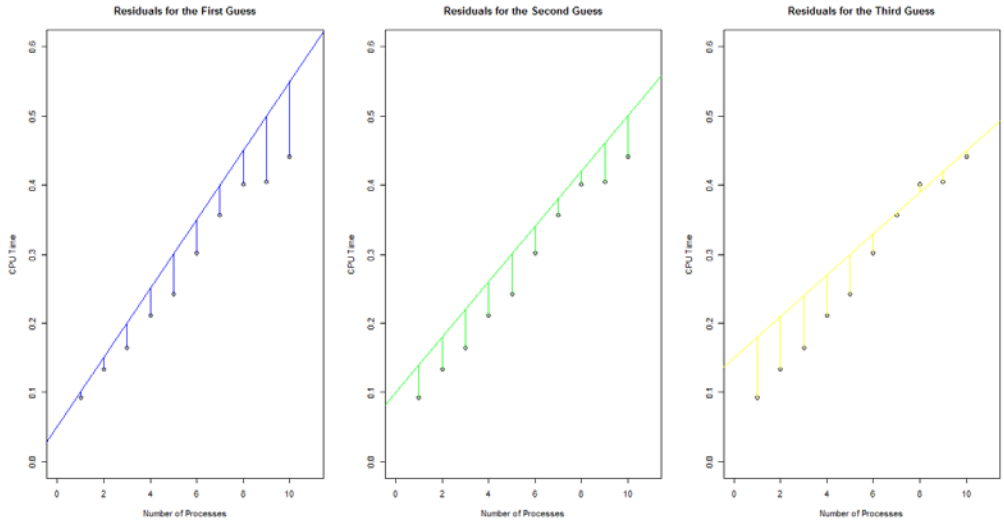

Pareto Chart for Reject_Freq
Frequency
Cumulative Percentage
0
20
40
60
80
0%
25%
50%
75%
100%
No Signature
Illegible
Current Customer
No Address
Other

Chapter 4, Exploratory Analysis



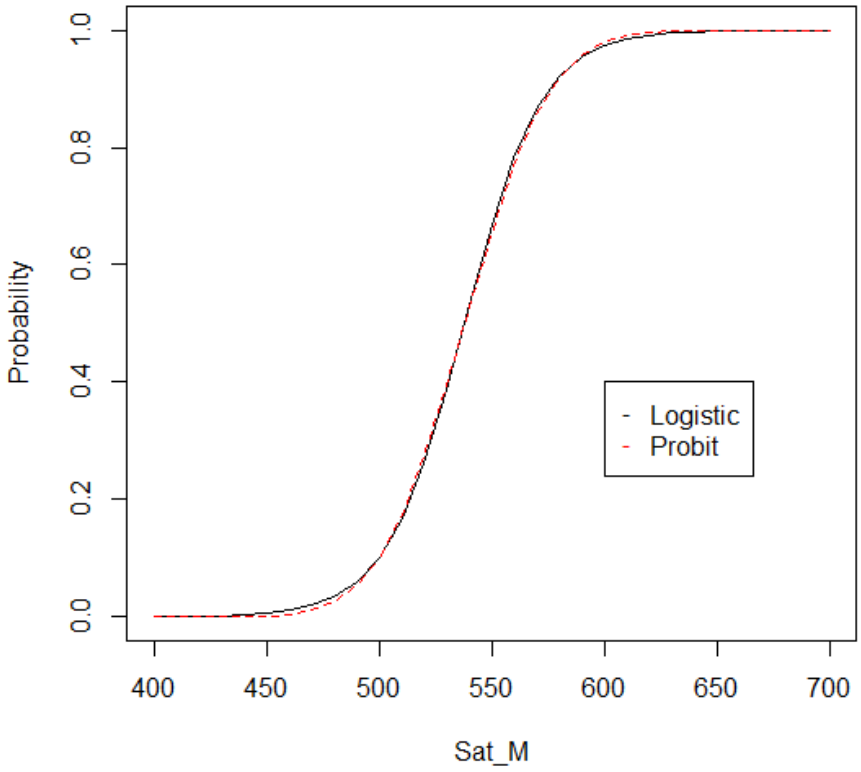


Chapter 6, Linear Regression Analysis



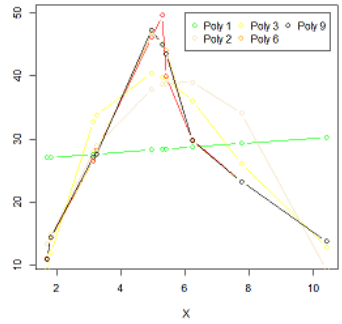
Chapter 7, The Logistic Regression Model

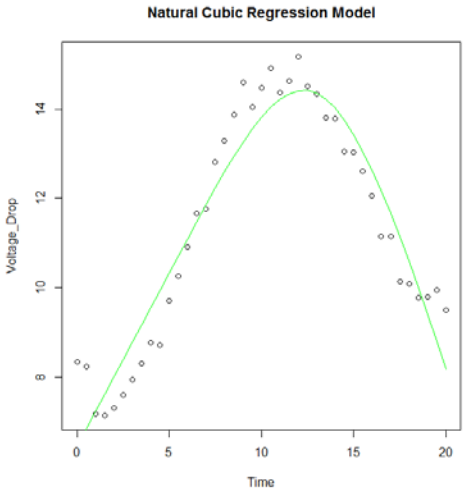
```
> sat_x <- seq(400,700, 10)
> pred_l <- predict(pass_logistic,newdata=list(Sat=sat_x),type="response")
> pred_p <- predict(pass_probit,newdata=list(Sat=sat_x),type="response")
> plot(sat_x,pred_l,type="l",ylab="Probability",xlab="Sat_M",col=1)
> lines(sat_x,pred_p,lty=2,col=2)
> legend(600,0.4,c("Logistic","Probit"),col=c(1:2),pch="-")
```

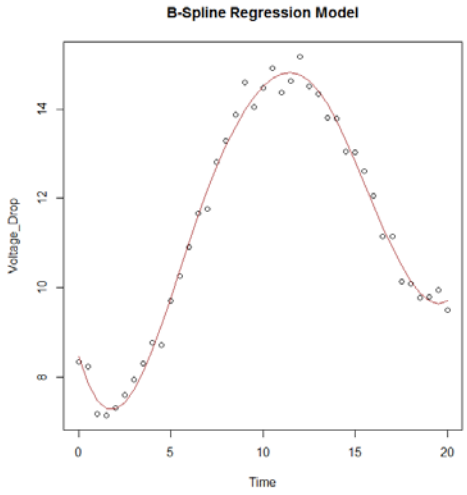


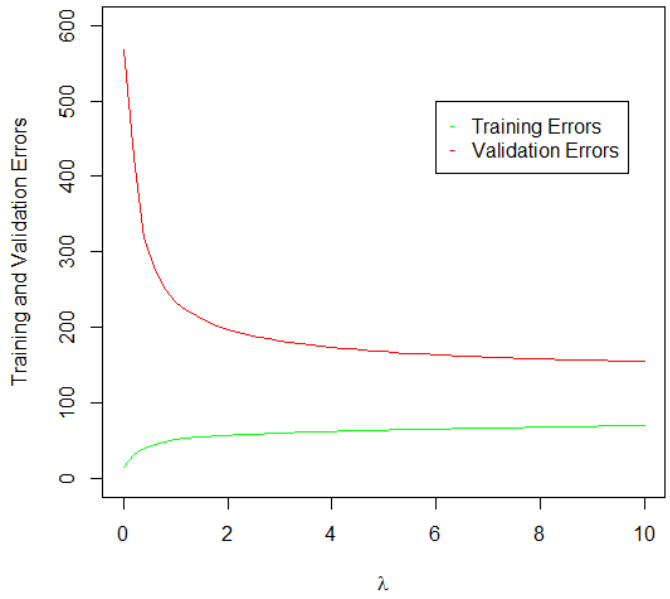
Chapter 8, Regression Models with Regularization

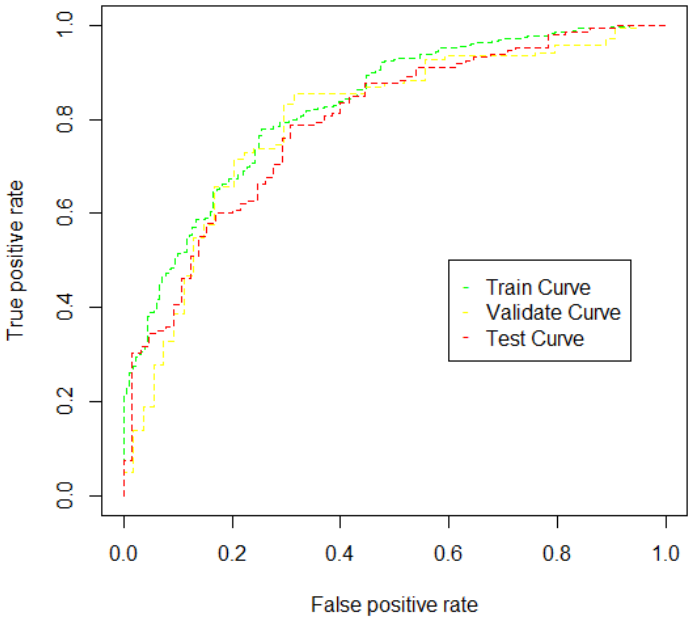
```
> data(OF)
> plot(OF$x, OF$Y,"b",col="red",xlab="X", ylab="Y")
> lines(OF$X,lm(Y~poly(X,1,raw=TRUE),data=OF)$fitted.values,"b",col="green")
> lines(OF$X,lm(Y~poly(X,2,raw=TRUE),data=OF)$fitted.values,"b",col="wheat")
> lines(OF$X,lm(Y~poly(X,3,raw=TRUE),data=OF)$fitted.values,"b",col="yellow")
> lines(OF$X,lm(Y~poly(X,6,raw=TRUE),data=OF)$fitted.values,"b",col="orange")
> lines(OF$X,lm(Y~poly(X,9,raw=TRUE),data=OF)$fitted.values,"b",col="black")
> legend(6,50,c("Poly 1 ","Poly 2 ","Poly 3 ","Poly 6 ","Poly 9"),
+ col=c('green','wheat','yellow','orange','black'),pch=1,ncol=3)
> R2 <- NULL; AdjR2 <- NULL; FStat <- NULL
> Mvar <- NULL; PolyOrder<-1:9
> for(i in 1:9){
+ temp <- summary(lm(Y~poly(X,i,raw=T),data=OF))
+ R2[i] <- temp$r.squared
+ AdjR2[i] <- temp$adj.r.squared
+ FStat[i] <- as.numeric(temp$fstatistic[1])
+ Mvar[i] <- temp$sigma
+ }
> cbind(PolyOrder,R2,AdjR2,FStat,Mvar)
      PolyOrder        R2     AdjR2     FStat    Mvar
 [1,]         1 0.005429 -0.1189  0.04367 14.274
 [2,]         2 0.744273  0.6712 10.18647  7.737
 [3,]         3 0.844078  0.7670 10.87646  6.513
 [4,]         4 0.905937  0.8307 12.03893  5.552
 [5,]         5 0.901566  0.9135 20.01489  3.968
 [6,]         6 0.976865  0.9306 21.11233  3.555
 [7,]         7 0.977085  0.8969 12.18247  4.333
 [8,]         8 0.977429  0.7969  5.41304  6.082
 [9,]         9 0.977429  0.7969  5.41304  6.082
```

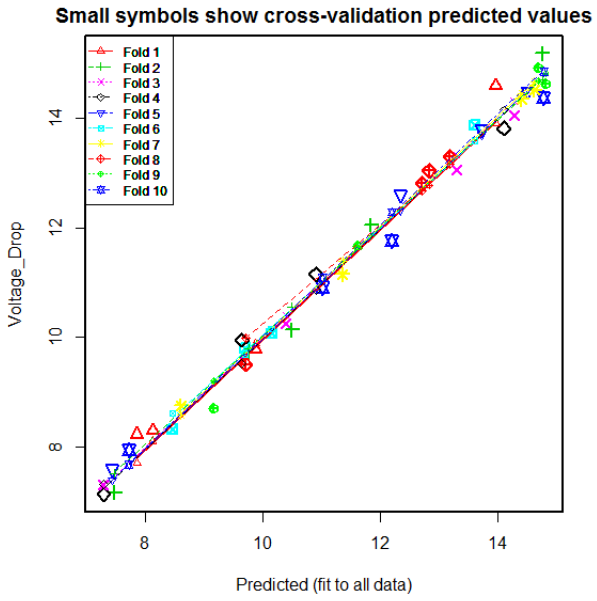




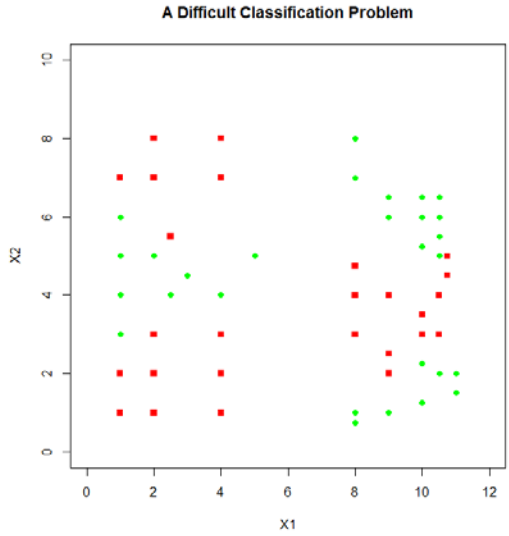


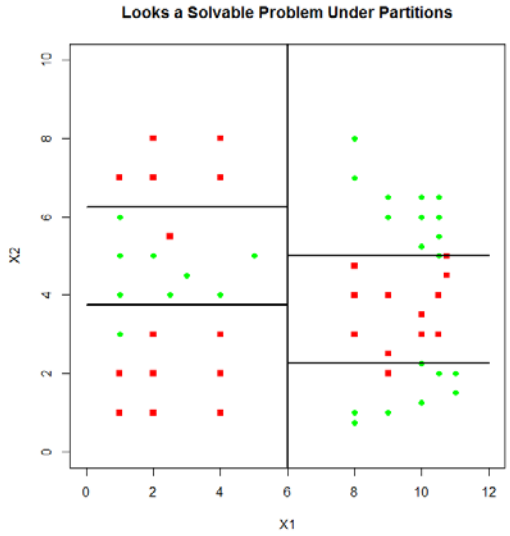

Training and Validation Errors
0
100
200
300
400
500
600
0
2
4
6
8
10
λ
Training Errors
Validation Errors

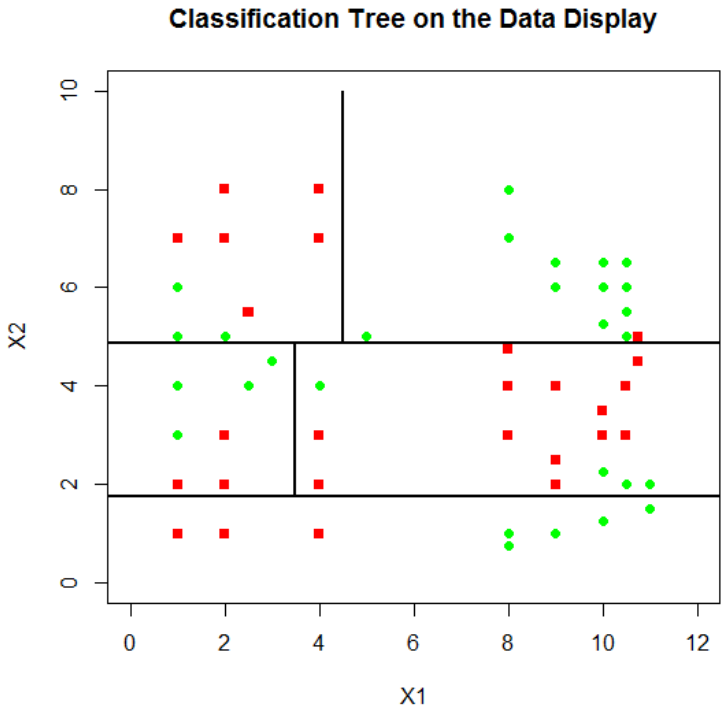

True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
Train Curve
Validate Curve
Test Curve

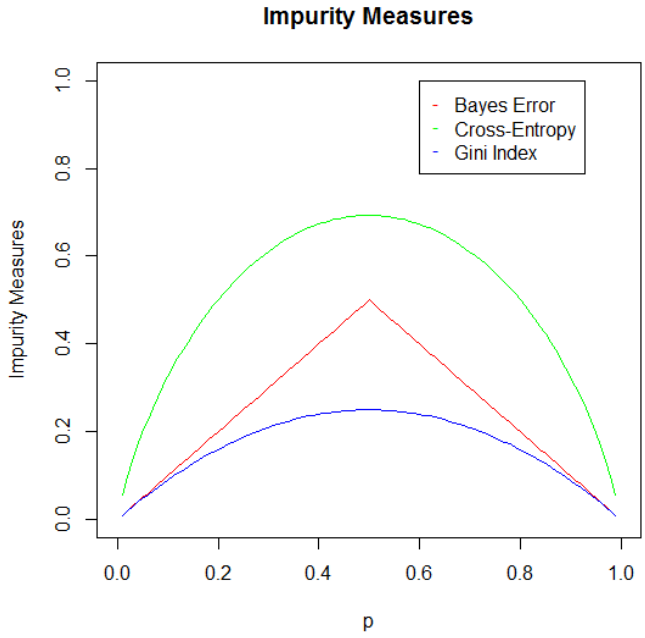


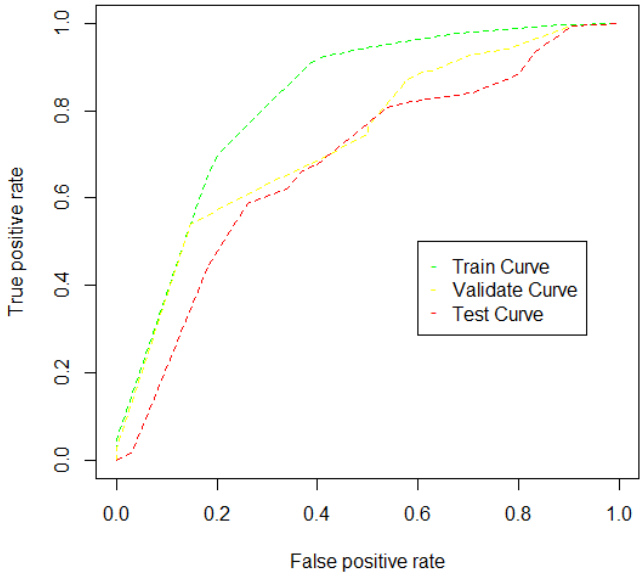
Chapter 9, Classification and Regression Trees

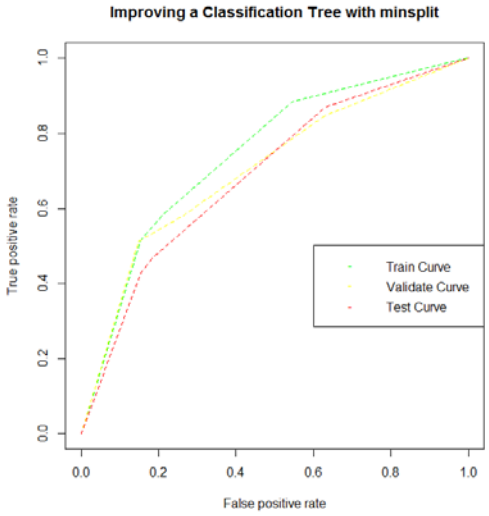


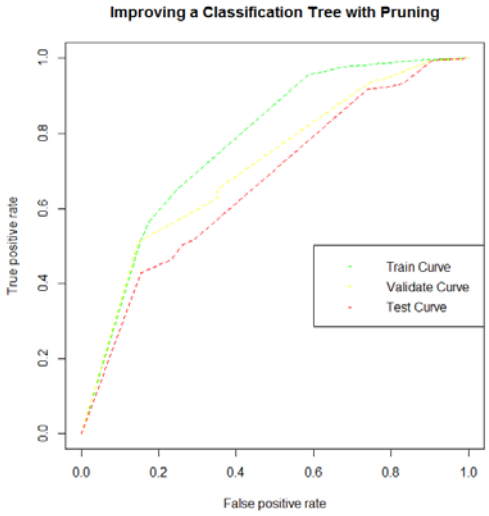



Classification Tree on the Data Display
X2
X1
0
2
4
6
8
10
12


Impurity Measures
Bayes Error
Cross-Entropy
Gini Index
Impurity Measures
p
0.0
0.2
0.4
0.6
0.8
1.0


True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Train Curve
Validate Curve
Test Curve


Improving a Classification Tree with minsplit
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Train Curve
Validate Curve
Test Curve


Improving a Classification Tree with Pruning
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Train Curve
Validate Curve
Test Curve

Chapter 10, CART and Beyond