6-7RandallPlylerAfter

Randall Plyler

2/12/2022

Predicting Airfare on New Routes. The following problem takes place in the United States in the late 1990s, when many major US cities were facing issues with airport congestion, partly as a result of the 1978 deregulation of airlines. Both fares and routes were freed from regulation, and low-fare carriers such as Southwest (SW) began competing on existing routes and starting nonstop service on routes that previously lacked it.

Building completely new airports is generally not feasible, but sometimes decommissioned military bases or smaller municipal airports can be reconfigured as regional or larger commercial airports. There are numerous players and interests involved in the issue (airlines, city, state and federal authorities, civic groups, the military, airport operators), and an aviation consulting firm is seeking advisory contracts with these players. The firm needs predictive models to support its consulting service. One thing the firm might want to be able to predict is fares, in the event a new airport is brought into service. The firm starts with the file Airfares.csv, which contains real data that were collected between Q3-1996 and Q2-1997.

```
Airfares.csv <- read.csv("C:/Users/randa/Dropbox/Masters/Winter/TBANLT 560 Data Mining/Files/DMBA-R-data
#show(Airfares.csv)
```

The variables in these data are listed in Table 6.11, and are believed to be important in predicting FARE. Some airport-to-airport data are available, but most are at the city-to-city level. One question that will be of interest in the analysis is the effect that the presence or absence of Southwest has on FARE.

a. Explore the numerical predictors and response (FARE) by creating a correlation table and examining some scatterplots between FARE and those predictors. What seems to be the best single predictor of FARE?

```
airfares.df <- Airfares.csv
#show(airfares.df)
AirFareDataSet2 <- airfares.df[, c('COUPON','NEW','HI','S_INCOME','E_INCOME','S_POP','E_POP','DISTANCE'
corA <- cor(AirFareDataSet2)
corA
```

```
##               COUPON         NEW          HI    S_INCOME   E_INCOME       S_POP
## COUPON    1.00000000  0.02022307 -0.34725207 -0.08840265  0.0468892 -0.10776336
## NEW       0.02022307  1.00000000  0.05414685  0.02659673  0.1133766 -0.01667212
## HI       -0.34725207  0.05414685  1.00000000 -0.02738221  0.0823926 -0.17249541
## S_INCOME -0.08840265  0.02659673 -0.02738221  1.00000000 -0.1388642  0.51718718
## E_INCOME  0.04688920  0.11337664  0.08239260 -0.13886420  1.0000000 -0.14405857
## S_POP    -0.10776336 -0.01667212 -0.17249541  0.51718718 -0.1440586  1.00000000
## E_POP     0.09496994  0.05856818 -0.06245600 -0.27228027  0.4584181 -0.28014283
## DISTANCE  0.74680521  0.08096520 -0.31237457  0.02815334  0.1765307  0.01843667
## PAX      -0.33697358  0.01049527 -0.16896078  0.13819710  0.2599611  0.28461056
## FARE      0.49653696  0.09172969  0.02519492  0.20913485  0.3260923  0.14509708
##                E_POP    DISTANCE         PAX        FARE
## COUPON    0.09496994  0.74680521 -0.33697358  0.49653696
## NEW       0.05856818  0.08096520  0.01049527  0.09172969
## HI       -0.06245600 -0.31237457 -0.16896078  0.02519492
## S_INCOME -0.27228027  0.02815334  0.13819710  0.20913485
## E_INCOME  0.45841806  0.17653074  0.25996105  0.32609229
## S_POP    -0.28014283  0.01843667  0.28461056  0.14509708
## E_POP     1.00000000  0.11563970  0.31469750  0.28504299
## DISTANCE  0.11563970  1.00000000 -0.10248160  0.67001599
## PAX       0.31469750 -0.10248160  1.00000000 -0.09070541
## FARE      0.28504299  0.67001599 -0.09070541  1.00000000
```

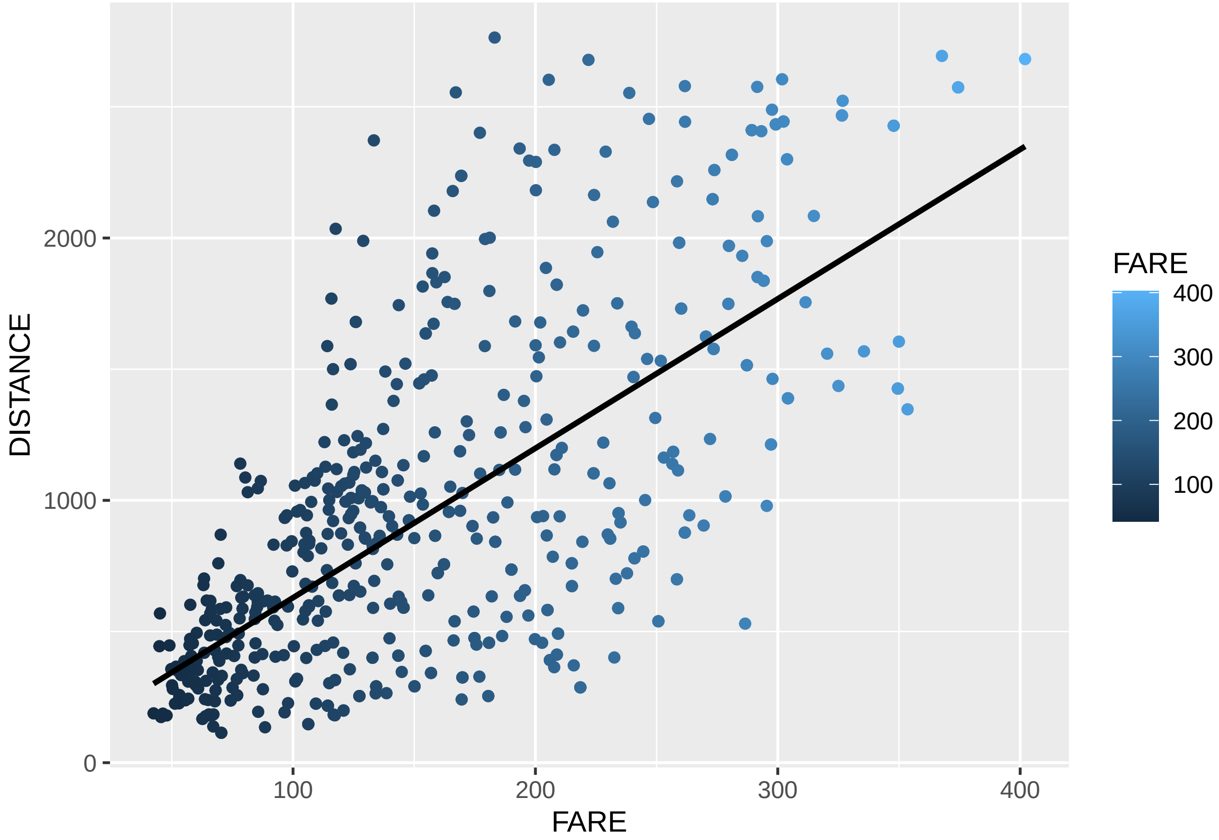
```r
set.seed(100)
train.index <- sample(row.names(AirFareDataSet2), 0.7*dim(AirFareDataSet2)[1])
valid.index <- setdiff(row.names(AirFareDataSet2), train.index)
trainingdataset <- AirFareDataSet2[train.index, ]
valid.df <- AirFareDataSet2[valid.index, ]

AirplaneLM <- lm(FARE ~ ., data = trainingdataset)

options(scipen = 100)
summary(AirplaneLM)
```

```
##
## Call:
## lm(formula = FARE ~ ., data = trainingdataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -122.730  -28.716    3.858   26.354  132.302
##
## Coefficients:
##                    Estimate      Std. Error t value             Pr(>|t|)
## (Intercept) -220.6145688733   31.8566395630  -6.925     0.00000000001567 ***
## COUPON        15.4272946184   17.3341783685   0.890                0.374
## NEW           -2.1483841737    2.9266580547  -0.734                0.463
## HI             0.0094469062    0.0013857511   6.817     0.00000000003105 ***
## S_INCOME       0.0049002658    0.0007020682   6.980     0.00000000001107 ***
## E_INCOME       0.0028676671    0.0005358116   5.352     0.00000014088061 ***
## S_POP          0.0000062412    0.0000008687   7.185     0.00000000000294 ***
## E_POP          0.0000091960    0.0000009684   9.496 < 0.0000000000000002 ***
## DISTANCE       0.0713436254    0.0050519730  14.122 < 0.0000000000000002 ***
## PAX           -0.0012596326    0.0002015209  -6.251     0.00000000097450 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 43.36 on 436 degrees of freedom
## Multiple R-squared:  0.6778, Adjusted R-squared:  0.6711
## F-statistic: 101.9 on 9 and 436 DF,  p-value: < 0.00000000000000022
```

```r
#The best predictor of fare is distance, then next is coupon.
```

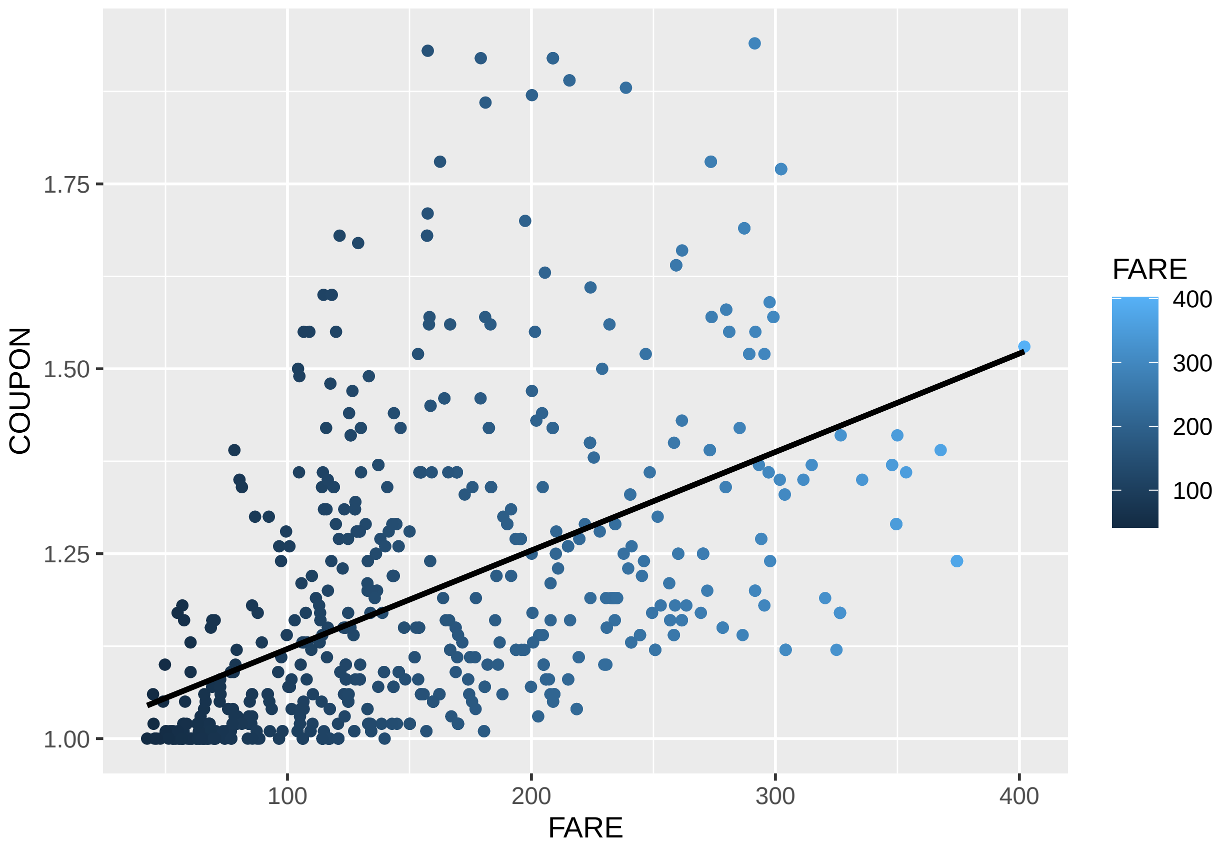
```
library(ggplot2)
ggplot(AirFareDataSet2, aes(x=FARE, y=DISTANCE, colour=FARE)) + geom_point(size=1.5) +stat_smooth(metho
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



```
ggplot(AirFareDataSet2, aes(x=FARE, y=COUPON, colour=FARE)) + geom_point(size=1.5) +stat_smooth(method=
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



b. Explore the categorical predictors (excluding the first four) by computing the percentage of flights in each category. Create a pivot table with the average fare in each category. Which categorical predictor seems best for predicting FARE?

```
#prop.table(table(x), 1)
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(dbplyr)
```

```
##
## Attaching package: 'dbplyr'
## The following objects are masked from 'package:dplyr':
##
##     ident, sql
```

```
AirFareDataSet3 <- airfares.df[, c(7,8,14,15,18)]
store_variable_vaca <- count(AirFareDataSet3,'VACATION')

no_len=length(AirFareDataSet3[AirFareDataSet3$VACATION=="No", ]$VACATION )
yes_len=length(AirFareDataSet3[AirFareDataSet3$VACATION=="Yes", ]$VACATION )
summary(yes_len)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     170     170     170     170     170     170
```

```
summary(no_len)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     468     468     468     468     468     468
```

```
airfare.calc.no=sum(AirFareDataSet3$FARE)
airfare.calc.no
```

```
## [1] 102639.3
```

```
avg.fare.no=airfare.calc.no/no_len
avg.fare.no
```

```
## [1] 219.3148
```

```
AirFareDataSet3[AirFareDataSet3$VACATION=="No", ]
```

```
##     VACATION  SW       SLOT        GATE   FARE
## 1          No Yes       Free        Free  64.11
## 2          No  No       Free        Free 174.47
## 3          No  No       Free        Free 207.76
## 4          No Yes Controlled        Free  85.47
## 5          No Yes       Free        Free  85.47
## 6          No Yes       Free        Free  56.76
## 7          No  No       Free        Free 228.00
## 9          No Yes       Free        Free 172.63
## 10         No Yes       Free        Free 114.76
## 12         No Yes       Free        Free 228.99
## 13         No Yes       Free        Free  79.17
## 14         No Yes       Free        Free 132.05
## 15         No  No Controlled        Free 117.23
## 16         No  No Controlled        Free 117.23
## 17         No  No       Free Constrained 117.23
## 19         No Yes       Free        Free 181.16
## 20         No Yes       Free        Free 157.50
## 21         No Yes       Free        Free 200.20
## 22        No  No       Free        Free 246.85
## 23        No Yes       Free        Free  99.70
## 25        No Yes Controlled        Free 113.50
## 26        No Yes       Free        Free 113.50
## 27        No Yes       Free        Free  69.12
## 28        No  No       Free        Free 210.00
## 29        No  No Controlled        Free 134.30
## 30        No  No Controlled        Free 134.30
## 31        No  No       Free Constrained 134.30
## 32        No  No Controlled        Free 118.95
## 33        No  No Controlled        Free 118.95
## 34        No  No       Free Constrained 118.95
## 35        No  No       Free Constrained  97.96
## 36        No  No       Free Constrained 237.80
## 37        No  No Controlled Constrained 234.15
## 38        No  No       Free Constrained 234.15
## 39        No  No       Free Constrained 203.17
## 40        No  No Controlled Constrained 250.73
## 41        No  No Controlled Constrained 250.73
## 42        No  No       Free Constrained 250.73
## 43        No  No Controlled        Free 106.60
## 44        No  No       Free        Free 106.60
## 45        No Yes Controlled        Free 136.27
## 46        No Yes       Free        Free 136.27
## 47        No  No Controlled        Free 230.87
## 48        No  No       Free        Free 230.87
## 49        No  No Controlled Constrained 180.56
## 50        No  No       Free Constrained 180.56
## 51        No  No       Free Constrained 215.83
## 52        No  No       Free        Free 197.10
## 53        No Yes Controlled        Free  69.10
## 54        No Yes       Free        Free  69.10
## 55        No  No       Free        Free  91.83
## 56        No  No       Free        Free 111.66
## 58        No  No       Free        Free 104.72
## 59        No  No Controlled        Free 154.74
## 60        No  No Controlled        Free 154.74
## 61        No  No       Free Constrained 154.74
## 62        No  No       Free        Free  77.98
## 63        No  No       Free        Free 157.20
## 64        No  No Controlled        Free 157.20
## 65        No  No       Free        Free 113.20
## 66        No  No       Free        Free 143.59
## 67        No Yes Controlled        Free  75.07
## 68        No Yes       Free        Free  75.07
## 69        No Yes       Free        Free  84.46
## 70        No  No       Free        Free 113.99
## 71        No Yes       Free        Free  67.17
## 72        No  No       Free        Free 320.37
## 73        No  No Controlled        Free 244.50
## 74        No  No       Free        Free 244.50
## 75        No Yes       Free        Free  78.62
## 76        No  No       Free        Free 210.90
## 77        No  No       Free        Free 311.46
## 78        No  No Controlled        Free 174.06
## 79        No  No       Free        Free 174.06
## 80        No  No       Free        Free 155.81
## 81        No  No       Free Constrained 106.56
## 82        No  No       Free Constrained 110.42
## 83        No Yes Controlled Constrained  74.28
## 84        No Yes       Free Constrained  74.28
## 85        No  No       Free Constrained 245.28
## 86        No  No       Free Constrained 256.48
## 87        No Yes       Free        Free  84.23
## 98        No  No Controlled        Free 180.85
## 99        No  No Controlled        Free 180.85
## 100       No  No       Free Constrained 180.85
## 101       No  No       Free        Free 175.81
## 102       No  No Controlled        Free 240.88
## 103       No  No       Free        Free 240.88
## 108       No  No       Free        Free 233.16
## 109       No Yes       Free        Free  67.10
## 110       No  No       Free        Free 349.97
## 111       No Yes Controlled        Free 139.56
## 112       No Yes       Free        Free 139.56
## 113       No Yes       Free        Free 191.63
## 114       No Yes       Free        Free  65.31
## 115       No Yes       Free        Free  67.78
## 116       No  No       Free        Free 204.68
## 117       No Yes       Free Constrained 177.22
## 118       No  No       Free        Free 109.78
## 119       No Yes Controlled        Free  62.63
## 120       No Yes       Free        Free  62.63
## 121       No Yes       Free Constrained 169.58
## 124       No  No       Free        Free 153.50
## 125       No  No Controlled        Free 195.64
## 126       No  No Controlled        Free 195.64
## 127       No  No       Free Constrained 195.64
## 129       No Yes       Free        Free 138.08
## 130       No  No       Free        Free 157.45
## 136       No  No       Free        Free 116.18
## 137       No Yes Controlled        Free  75.71
## 138       No Yes       Free        Free  75.71
## 139       No  No       Free        Free 116.57
## 140       No  No       Free        Free 110.25
## 141       No Yes       Free Constrained 123.27
## 142       No Yes       Free        Free 127.78
## 144       No Yes       Free        Free 116.00
## 160       No Yes       Free        Free  59.77
## 161       No Yes       Free        Free  76.79
## 162       No  No       Free        Free 225.56
## 163       No  No       Free        Free 301.79
## 164       No  No Controlled        Free 233.78
## 165       No  No       Free        Free 233.78
## 166       No  No       Free        Free 231.97
## 167       No  No       Free        Free 179.23
## 168       No  No       Free        Free 272.06
## 169       No  No       Free        Free 129.80
## 170       No  No       Free Constrained 295.49
## 172       No Yes       Free        Free 195.28
## 173       No  No       Free        Free 101.68
## 174       No Yes       Free        Free  60.26
## 175       No Yes Controlled        Free  68.06
## 176       No Yes       Free        Free  68.06
## 177       No Yes       Free        Free  60.87
## 178       No  No       Free        Free  70.62
## 187       No  No Controlled        Free 190.09
## 188       No  No Controlled        Free 190.09
## 189       No  No       Free Constrained 190.09
## 191       No  No       Free        Free 154.06
## 192       No  No       Free Constrained 269.43
## 193       No  No       Free Constrained 258.85
## 194       No  No Controlled Constrained 156.93
## 195       No  No       Free Constrained 156.93
## 196       No  No       Free Constrained 230.71
## 197       No  No       Free Constrained 133.50
## 198       No  No       Free Constrained 286.54
## 199       No  No       Free Constrained 232.55
## 201       No  No       Free Constrained 246.10
## 202       No Yes Controlled        Free  84.21
## 203       No Yes       Free        Free  84.21
## 204       No  No       Free        Free 181.99
## 205       No Yes       Free Constrained 166.25
## 206       No Yes       Free        Free 107.86
## 207       No Yes       Free        Free 181.02
## 208       No  No Controlled        Free 215.01
## 209       No  No Controlled        Free 215.01
## 210       No  No       Free Constrained 215.01
## 212       No  No       Free        Free 120.70
## 213       No Yes Controlled        Free 132.85
## 214       No Yes       Free        Free 132.85
## 215       No Yes       Free        Free  84.53
## 216       No Yes       Free        Free  76.81
## 217       No Yes       Free        Free 202.00
## 218       No  No Controlled        Free 208.79
## 219       No  No Controlled        Free 208.79
## 220       No  No       Free Constrained 208.79
## 221       No  No Controlled        Free 162.28
## 222       No  No Controlled        Free 162.28
## 223       No  No       Free Constrained 162.28
## 224       No  No Controlled        Free 287.23
## 225       No  No Controlled        Free 287.23
## 226       No  No       Free Constrained 287.23
## 227       No  No Controlled        Free 116.78
## 228       No  No Controlled        Free 116.78
## 229       No  No       Free Constrained 116.78
## 230       No  No Controlled        Free 159.71
## 231       No  No Controlled        Free 159.71
## 232       No  No Controlled Constrained 159.71
## 233       No  No Controlled        Free 159.71
## 234       No  No Controlled        Free 159.71
## 235       No  No       Free Constrained 159.71
## 236       No  No Controlled Constrained 205.00
## 237       No  No Controlled Constrained 205.00
## 238       No  No       Free Constrained 205.00
## 239       No  No Controlled        Free 143.44
## 240       No  No Controlled        Free 143.44
## 241       No  No       Free Constrained 143.44
## 242       No  No Controlled        Free 174.87
## 243       No  No Controlled        Free 174.87
## 244       No  No       Free Constrained 174.87
## 245       No  No Controlled        Free 304.18
## 246       No  No Controlled        Free 304.18
## 247       No  No       Free Constrained 304.18
## 248       No  No Controlled        Free 270.36
## 249       No  No Controlled        Free 270.36
## 250       No  No       Free Constrained 270.36
## 251       No  No Controlled Constrained 209.35
## 252       No  No Controlled Constrained 209.35
## 253       No  No       Free Constrained 209.35
## 260       No  No Controlled        Free 144.60
## 261       No  No Controlled        Free 144.60
## 262       No  No       Free Constrained 144.60
## 263       No  No Controlled        Free 349.53
## 264       No  No Controlled        Free 349.53
## 265       No  No       Free Constrained 349.53
## 269       No  No Controlled        Free 223.99
## 270       No  No Controlled        Free 223.99
## 271       No  No       Free Constrained 223.99
## 275       No  No Controlled        Free 326.47
## 276       No  No Controlled        Free 326.47
## 277       No  No       Free Constrained 326.47
## 278       No  No Controlled        Free 234.31
## 279       No  No Controlled        Free 234.31
## 280       No  No       Free Constrained 234.31
## 284       No  No Controlled Constrained 278.39
## 285       No  No Controlled Constrained 278.39
## 286       No  No       Free Constrained 278.39
## 287       No  No Controlled        Free 208.71
## 288       No  No Controlled        Free 208.71
## 289       No  No       Free Constrained 208.71
## 290       No  No Controlled        Free 150.13
## 291       No  No Controlled        Free 150.13
## 292       No  No       Free Constrained 150.13
## 293       No Yes       Free        Free  56.43
## 295       No Yes       Free        Free  53.80
## 296       No Yes       Free        Free  66.14
## 297       No Yes       Free        Free  96.18
## 298       No Yes Controlled        Free  67.77
## 299       No Yes       Free        Free  67.77
## 300       No  No       Free        Free 139.81
## 324       No  No       Free        Free 125.09
## 325       No  No       Free        Free 138.56
## 326       No  No Controlled        Free 215.06
## 327       No  No       Free        Free 215.06
## 328       No  No       Free        Free 249.45
## 329        No  No       Free        Free 335.55
## 330        No  No       Free Constrained 175.66
## 332        No  No       Free        Free 353.56
## 334        No  No       Free        Free 293.21
## 336        No  No       Free Constrained 295.46
## 338        No Yes       Free        Free  57.05
## 339        No  No       Free        Free 200.09
## 340        No Yes       Free        Free 105.41
## 341        No  No       Free        Free 197.42
## 342        No Yes       Free        Free  57.33
## 343        No Yes Controlled        Free 152.10
## 344        No Yes       Free        Free 152.10
## 345        No Yes       Free        Free 166.66
## 346        No Yes       Free        Free 158.00
## 347        No  No       Free        Free 229.84
## 348        No  No       Free        Free 133.04
## 349        No Yes       Free Constrained 191.66
## 350        No Yes       Free        Free  60.73
## 351        No Yes       Free        Free 148.28
## 352        No Yes       Free        Free  85.48
## 354        No Yes       Free        Free  51.73
## 355        No  No       Free Constrained 195.91
## 356        No  No Controlled        Free 273.12
## 357        No  No Controlled        Free 273.12
## 358        No  No       Free Constrained 273.12
## 359        No Yes       Free        Free  84.15
## 360        No Yes       Free        Free  81.32
## 362        No  No       Free        Free 291.78
## 363        No  No       Free Constrained  93.55
## 364        No  No       Free Constrained 186.28
## 365        No  No Controlled Constrained 208.86
## 366        No  No       Free Constrained 208.86
## 367        No  No Controlled Constrained 169.90
## 368        No  No Controlled Constrained 169.90
## 369        No  No       Free Constrained 169.90
## 370        No  No       Free Constrained 169.90
## 372        No  No       Free Constrained 134.09
## 373        No Yes       Free        Free  66.88
## 374        No  No Controlled        Free 279.61
## 375        No  No       Free        Free 279.61
## 376        No  No       Free        Free 188.46
## 379        No Yes       Free        Free  91.97
## 380        No  No Controlled        Free 302.33
## 381        No  No Controlled        Free 302.33
## 382        No  No       Free Constrained 302.33
## 383        No Yes       Free        Free  63.76
## 384        No Yes       Free        Free 114.95
## 385        No Yes       Free        Free  65.80
## 386        No Yes       Free        Free  79.48
## 387        No Yes       Free        Free  96.58
## 388        No  No       Free        Free  68.41
## 389        No Yes       Free        Free  65.84
## 390        No  No       Free        Free  88.46
## 391        No Yes       Free        Free  50.38
## 392       No  No       Free        Free 176.88
## 393       No  No Controlled        Free 219.38
## 394       No  No       Free        Free 219.38
## 395       No  No Controlled        Free 106.29
## 396       No  No Controlled        Free 106.29
## 397       No  No       Free Constrained 106.29
## 399       No  No       Free        Free 123.44
## 400       No  No       Free        Free 140.07
## 401       No  No Controlled        Free 193.67
## 402       No  No       Free        Free 193.67
## 403       No  No       Free        Free 230.56
## 404       No  No Controlled        Free 154.73
## 405       No  No Controlled        Free 154.73
## 406       No  No       Free Constrained 154.73
## 407       No  No       Free        Free 144.86
## 408       No  No       Free        Free 109.44
## 409       No  No Controlled        Free 109.44
## 420       No  No Controlled        Free 218.54
## 421       No  No Controlled        Free 218.54
## 422       No  No       Free Constrained 218.54
## 423       No  No Controlled        Free 127.38
## 424       No  No Controlled        Free 127.38
## 425       No  No       Free Constrained 127.38
## 426       No Yes       Free        Free  52.53
## 428       No Yes       Free        Free  56.79
## 429       No Yes       Free        Free  78.67
## 430       No Yes       Free        Free  57.57
## 431       No Yes       Free        Free  64.39
## 432       No  No       Free        Free 179.13
## 433       No Yes Controlled        Free 185.65
## 434       No Yes       Free        Free 185.65
## 435       No  No       Free        Free  69.60
## 437       No Yes       Free        Free  66.46
## 438       No  No Controlled        Free 259.32
## 439       No  No Controlled        Free 259.32
## 440       No  No       Free Constrained 259.32
## 441       No Yes       Free        Free  69.95
## 442       No Yes       Free        Free  73.69
## 443       No Yes       Free        Free  72.22
## 444       No Yes       Free        Free  65.91
## 445       No  No       Free        Free  85.19
## 446       No Yes       Free        Free  78.30
## 447       No Yes Controlled        Free 128.36
## 448       No Yes       Free        Free 128.36
## 449       No Yes       Free        Free  63.69
## 450       No Yes       Free        Free  63.92
## 452       No Yes       Free        Free 130.09
## 453       No  No Controlled        Free 273.53
## 454       No  No Controlled        Free 273.53
## 455       No  No       Free Constrained 273.53
## 457       No  No       Free        Free 204.35
## 458       No  No       Free        Free 291.51
## 459       No Yes Controlled        Free 219.63
## 460       No Yes       Free        Free 219.63
## 461       No  No       Free        Free 252.97
## 462       No  No       Free        Free 134.79
## 464       No  No       Free        Free  70.41
## 465       No  No       Free        Free  70.41
## 466       No  No       Free Constrained 251.73
## 467       No  No Controlled        Free 299.17
## 468       No  No Controlled        Free 299.17
## 469       No  No       Free Constrained 299.17
## 470       No Yes       Free        Free  57.29
## 471       No Yes       Free        Free  50.10
## 472       No Yes       Free        Free  58.68
## 473       No Yes       Free        Free 100.80
## 474       No  No       Free        Free 248.49
## 475       No  No       Free        Free 367.72
## 476       No  No       Free        Free  83.74
## 477       No  No Controlled        Free 291.66
## 478       No  No       Free        Free 291.66
## 479       No  No       Free        Free 297.83
## 480       No  No       Free        Free 168.96
## 481       No  No       Free Constrained 314.88
## 482       No Yes       Free        Free 241.04
## 483       No  No       Free        Free 116.52
## 485       No  No       Free        Free  79.23
## 487       No  No       Free Constrained 224.17
## 488       No  No Controlled        Free 374.40
## 489       No  No Controlled        Free 374.40
## 490       No  No       Free Constrained 374.40
## 492       No  No       Free        Free 326.76
## 493       No Yes       Free        Free  85.52
## 494       No  No       Free        Free  81.28
## 495       No Yes       Free        Free 200.41
## 496       No  No       Free        Free 402.02
## 497       No Yes       Free        Free  59.80
## 498       No  No Controlled        Free 294.18
## 499       No  No       Free        Free 294.18
## 500       No  No       Free        Free 325.02
## 501       No  No       Free        Free 263.48
## 503       No Yes       Free        Free  58.98
## 504       No Yes       Free        Free  92.57
## 505       No Yes       Free        Free  63.39
## 506       No Yes       Free        Free  63.30
## 510       No  No       Free        Free 142.83
## 511       No  No       Free        Free 200.20
## 512       No  No       Free        Free 297.61
## 513       No Yes       Free        Free  97.46
## 514       No  No Controlled        Free 260.16
## 515       No  No       Free        Free 260.16
## 516       No  No       Free        Free 239.66
## 517       No  No       Free        Free 169.92
## 518       No  No       Free Constrained 285.34
## 520       No Yes       Free        Free 101.64
## 521       No  No       Free Constrained 186.96
## 522       No  No Controlled        Free 289.25
## 523       No  No Controlled        Free 289.25
## 524       No  No       Free Constrained 289.25
## 525       No Yes       Free        Free  63.06
## 527       No Yes       Free        Free 110.00
## 528       No Yes       Free        Free  51.30
## 529       No  No       Free        Free 199.80
## 530       No Yes Controlled        Free  76.96
## 531       No Yes       Free        Free  76.96
## 532       No Yes       Free        Free  77.62
## 533       No  No       Free        Free 188.11
## 534       No  No       Free        Free 207.17
## 535       No Yes       Free Constrained  77.46
## 536       No Yes       Free        Free 105.13
## 537       No Yes       Free        Free  56.80
## 539       No Yes       Free        Free 210.16
## 540       No  No       Free Constrained 202.77
## 541       No  No Controlled        Free 261.63
## 542       No  No Controlled        Free 261.63
## 543       No  No       Free Constrained 261.63
## 545       No Yes       Free        Free 137.20
## 546       No  No Controlled        Free 120.84
## 547       No  No Controlled        Free 120.84
## 548       No  No       Free Constrained 120.84
## 570       No Yes       Free        Free  49.02
## 571       No Yes       Free        Free  54.96
## 572       No Yes       Free        Free  64.97
## 573       No Yes       Free        Free 100.36
## 574       No  No       Free        Free 215.57
## 575       No  No Controlled        Free 215.57
## 576       No  No       Free        Free 166.67
## 577       No  No Controlled        Free 166.67
## 578       No  No       Free        Free 132.79
## 579       No  No Controlled        Free 132.79
## 580       No  No Controlled        Free 145.61
## 581       No  No Controlled        Free 145.61
## 582       No  No       Free        Free 145.61
## 583       No  No Controlled        Free 145.61
## 584       No  No       Free        Free 100.95
## 585       No  No Controlled        Free 100.95
## 586       No  No       Free        Free 256.86
## 587       No  No Controlled        Free 256.86
## 588       No  No       Free        Free 240.48
## 589       No  No Controlled        Free 240.48
## 590       No  No       Free Constrained 205.96
## 591       No  No Controlled Constrained 205.96
## 594       No  No       Free        Free 117.35
## 595       No  No Controlled        Free 117.35
## 596       No  No       Free        Free 297.20
## 597       No  No Controlled        Free 297.20
## 598       No  No       Free        Free 182.56
## 599       No  No Controlled        Free 182.56
## 600       No  No       Free        Free 303.82
## 601       No  No Controlled        Free 303.82
## 604       No  No       Free Constrained 235.10
## 605       No  No Controlled Constrained 235.10
## 606       No  No       Free        Free 164.30
## 607       No  No Controlled        Free 164.30
## 608       No  No Controlled        Free 114.35
## 609       No  No Controlled        Free 114.35
## 610       No  No Controlled        Free 114.35
## 611       No  No Controlled        Free 114.35
## 612       No  No       Free Constrained 114.35
## 613       No  No Controlled Constrained 114.35
## 616       No  No       Free        Free 279.83
## 617       No  No Controlled        Free 279.83
## 618       No  No       Free        Free 273.83
## 619       No  No Controlled        Free 273.83
## 620       No  No       Free        Free 347.82
## 621       No  No Controlled        Free 347.82
## 622       No  No       Free        Free 281.06
## 623       No  No Controlled        Free 281.06
## 624       No  No       Free        Free 258.37
## 625       No  No Controlled        Free 258.37
```

```
AirFareDataSet3[AirFareDataSet3$VACATION=="Yes", ]
```

```
##     VACATION  SW       SLOT        GATE   FARE
## 8        Yes Yes       Free        Free 116.54
## 11       Yes Yes       Free        Free 158.20
## 18       Yes Yes       Free        Free 106.11
## 24       Yes Yes       Free        Free 106.77
## 57       Yes  No       Free        Free  57.62
## 88       Yes  No       Free        Free 105.10
## 89       Yes  No       Free        Free 121.09
## 90       Yes Yes Controlled        Free 153.95
## 91       Yes Yes       Free        Free 153.95
## 92       Yes  No       Free        Free 207.84
## 93       Yes  No       Free Constrained 113.39
## 94       Yes  No       Free        Free 126.62
## 95       Yes  No Controlled        Free 136.68
## 96       Yes  No       Free        Free 136.68
## 97       Yes  No       Free Constrained 108.15
## 104      Yes  No       Free        Free 183.19
## 105      Yes  No       Free        Free 167.16
## 106      Yes  No       Free        Free 177.09
## 107      Yes  No       Free        Free 221.89
## 122      Yes  No       Free        Free 105.73
## 123      Yes  No       Free        Free 114.13
## 128      Yes Yes       Free        Free  97.36
## 131      Yes Yes       Free        Free  99.43
## 132      Yes  No       Free        Free  87.59
## 133      Yes  No Controlled        Free 158.63
## 134      Yes  No       Free        Free 158.63
## 135      Yes  No       Free        Free 114.93
## 143      Yes Yes       Free        Free  78.24
## 145      Yes Yes       Free        Free  72.43
## 146      Yes  No       Free        Free 143.62
## 147      Yes Yes       Free        Free  80.31
## 148      Yes  No       Free        Free 133.35
## 149      Yes Yes       Free        Free  52.92
## 150      Yes Yes Controlled        Free 123.74
## 151      Yes Yes       Free        Free 123.74
## 152      Yes Yes       Free        Free 159.12
## 153      Yes  No       Free        Free 115.84
## 154      Yes  No       Free        Free 164.88
## 155      Yes  No       Free        Free  89.47
## 156      Yes  No       Free Constrained 163.78
## 157      Yes Yes       Free        Free 112.99
## 158      Yes Yes       Free        Free  55.57
## 159      Yes Yes       Free        Free  55.57
## 171      Yes  No       Free        Free 193.50
## 179      Yes  No       Free        Free  97.93
## 180      Yes  No       Free        Free 158.50
## 181      Yes  No Controlled        Free 168.92
## 182      Yes  No       Free        Free 168.92
## 183      Yes  No       Free        Free 185.11
## 184      Yes  No       Free Constrained 133.98
## 185      Yes  No       Free        Free 207.83
## 186      Yes  No       Free        Free 146.36
## 190      Yes  No       Free        Free 104.87
## 200      Yes  No       Free Constrained 171.67
## 211      Yes Yes       Free        Free  87.80
## 254      Yes  No Controlled        Free 123.18
## 255      Yes  No Controlled        Free 123.18
## 256      Yes  No       Free Constrained 123.18
## 257      Yes  No Controlled        Free 143.20
## 258      Yes  No Controlled        Free 143.20
## 259      Yes  No       Free Constrained 143.20
## 266      Yes  No Controlled        Free 183.43
## 267      Yes  No Controlled        Free 183.43
## 268      Yes  No       Free Constrained 183.43
## 272      Yes  No Controlled        Free 169.41
## 273      Yes  No Controlled        Free 169.41
## 274      Yes  No       Free Constrained 169.41
## 281      Yes  No Controlled        Free 124.92
## 282      Yes  No Controlled        Free 124.92
## 283      Yes  No       Free Constrained 124.92
## 294      Yes Yes       Free        Free  58.03
## 301      Yes  No       Free        Free  92.78
## 302      Yes  No       Free        Free 117.97
## 303      Yes Yes Controlled        Free 121.67
## 304      Yes Yes       Free        Free 121.67
## 305      Yes  No       Free Constrained 138.88
## 306      Yes  No       Free        Free 140.90
## 307      Yes Yes       Free        Free 104.33
## 308      Yes  No       Free        Free 153.58
## 309      Yes  No       Free        Free 201.43
## 310      Yes  No       Free Constrained 102.95
## 311      Yes Yes       Free        Free  45.55
## 312      Yes  No       Free        Free 114.50
## 313      Yes Yes       Free        Free 150.04
## 314      Yes Yes       Free        Free 121.35
## 315      Yes  No       Free        Free 117.59
## 316      Yes  No       Free        Free 258.43
## 317      Yes  No       Free        Free  96.53
## 318      Yes  No       Free        Free  96.53
## 319      Yes  No       Free Constrained 204.62
## 320      Yes Yes       Free        Free  92.35
## 321      Yes  No Controlled        Free 123.97
## 322      Yes  No Controlled        Free 123.97
## 323      Yes  No       Free Constrained 123.97
## 331      Yes  No       Free        Free 132.77
## 333      Yes  No       Free        Free 165.90
## 335      Yes  No       Free        Free 152.67
## 337      Yes  No       Free        Free 114.28
## 353      Yes Yes       Free        Free  53.07
## 361      Yes Yes       Free        Free 162.53
## 371      Yes  No       Free Constrained 122.62
## 377      Yes  No       Free        Free 205.51
## 378      Yes Yes       Free        Free  69.19
## 398      Yes  No       Free        Free 108.96
## 410      Yes  No Controlled        Free 125.90
## 411      Yes  No       Free        Free 125.90
## 412      Yes Yes       Free        Free  54.38
## 413      Yes Yes       Free        Free  60.28
## 414      Yes Yes       Free        Free  47.85
## 415      Yes Yes       Free        Free  72.42
## 416      Yes Yes       Free        Free  44.89
## 417      Yes Yes       Free        Free  68.59
## 418      Yes Yes       Free        Free  42.47
## 419      Yes Yes       Free        Free  45.11
## 427      Yes Yes       Free        Free  56.91
## 436      Yes Yes       Free        Free  57.40
## 451      Yes Yes       Free        Free  86.71
## 456      Yes Yes       Free        Free 118.17
## 463      Yes Yes       Free        Free  53.14
## 484      Yes Yes       Free        Free  72.58
## 486      Yes  No       Free        Free 261.67
## 491      Yes  No       Free        Free 261.74
## 502      Yes Yes       Free        Free  55.16
## 507      Yes  No Controlled        Free 137.25
## 508      Yes  No Controlled        Free 137.25
## 509      Yes  No       Free Constrained 137.25
## 519      Yes Yes       Free        Free  70.16
## 526      Yes  No       Free        Free 238.73
## 538      Yes Yes       Free        Free 141.48
## 544      Yes Yes       Free        Free 142.98
## 549      Yes  No       Free        Free 119.90
## 550      Yes  No       Free        Free 105.45
## 551      Yes  No       Free        Free  87.35
## 552      Yes  No       Free        Free 124.82
## 553      Yes Yes Controlled        Free 127.06
## 554      Yes Yes       Free        Free 127.06
## 555      Yes Yes       Free        Free 106.65
## 556      Yes  No       Free        Free 200.69
## 557      Yes  No       Free Constrained 107.51
## 558      Yes Yes       Free        Free  46.32
## 559      Yes  No       Free        Free 125.25
## 560      Yes Yes       Free        Free 128.97
## 561      Yes  No       Free        Free 224.21
## 562      Yes  No       Free        Free  85.62
## 563      Yes  No Controlled        Free 123.89
## 564      Yes  No Controlled        Free 123.89
## 565      Yes  No       Free Constrained 123.89
## 566      Yes  No       Free        Free 122.99
## 567      Yes  No       Free Constrained 119.84
## 568      Yes Yes       Free        Free 135.76
## 569      Yes Yes       Free        Free  49.77
## 592      Yes  No       Free        Free 127.67
## 593      Yes  No Controlled        Free 127.67
## 602      Yes  No       Free        Free 147.80
## 603      Yes  No Controlled        Free 147.80
## 614      Yes  No       Free        Free 125.80
## 615      Yes  No Controlled        Free 125.80
## 626      Yes  No       Free        Free 132.94
## 627      Yes  No Controlled        Free 132.94
## 628      Yes  No       Free        Free 104.11
## 629      Yes  No       Free        Free 127.83
## 630      Yes  No Controlled        Free 145.53
## 631      Yes  No       Free        Free 145.53
## 632      Yes  No       Free        Free 130.15
## 633      Yes  No Controlled        Free 129.63
## 634      Yes  No Controlled        Free 129.63
## 635      Yes  No       Free Constrained 129.63
## 636      Yes  No       Free        Free 124.87
## 637      Yes  No       Free        Free 129.62
## 638      Yes  No Controlled        Free 129.62
```

```
no_len=length(AirFareDataSet3[AirFareDataSet3$VACATION=="No", ]$VACATION )

yes_len=length(AirFareDataSet3[AirFareDataSet3$VACATION=="Yes", ]$VACATION )

vac.no=subset(AirFareDataSet3, VACATION=="No" )
length(vac.no$VACATION)
```

```
## [1] 468
```

```
no.df=AirFareDataSet3[AirFareDataSet3$VACATION=="No", ]

yes.df=AirFareDataSet3[AirFareDataSet3$VACATION=="Yes", ]

airfare.calc.no=sum(no.df$FARE)
airfare.calc.no
```

```
## [1] 81222.57
```

```
avg.fare.no=airfare.calc.no/no_len

AirFareDataSet3$VACATION <- ifelse(AirFareDataSet3$VACATION == "Yes", 1 , 0)
```

```
AirFareDataSet3$SW <- ifelse(AirFareDataSet3$SW == "Yes", 1 , 0)
AirFareDataSet3$SLOT <- ifelse(AirFareDataSet3$SLOT == "Controlled", 1 , 0)
AirFareDataSet3$GATE <- ifelse(AirFareDataSet3$GATE == "Constrained", 1 , 0)
str(AirFareDataSet3)
```

```
## 'data.frame':    638 obs. of  5 variables:
##  $ VACATION: num  0 0 0 0 0 0 0 1 0 0 ...
##  $ SW      : num  1 0 0 1 1 1 0 1 1 1 ...
##  $ SLOT    : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ GATE    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ FARE    : num  64.1 174.5 207.8 85.5 85.5 ...
```

```
freq.Vacation <- table(AirFareDataSet3$VACATION)
round(prop.table(freq.Vacation),4)*100
```

```
##
##     0     1
## 73.35 26.65
```

```
show(freq.Vacation)
```

```
##
##   0   1
## 468 170
```

```
freq.SW <- table(AirFareDataSet3$SW)
round(prop.table(freq.SW),4)*100
```

```
##
##     0     1
## 69.59 30.41
```

```
show(freq.SW)
```

```
##
##   0   1
## 444 194
```

```
freq.GATE <- table(AirFareDataSet3$GATE)
round(prop.table(freq.GATE),4)*100
```

```
##
##     0     1
## 80.56 19.44
```

```
show(freq.GATE)
```

```
##
##   0   1
## 514 124
```

```r
freq.SLOT <- table(AirFareDataSet3$SLOT)
round(prop.table(freq.SLOT),4)*100
```

```
##
##     0     1
## 71.47 28.53
```

```r
show(freq.SLOT)
```

```
##
##   0   1
## 456 182
```

```r
aggmean <- aggregate(AirFareDataSet3$FARE, list(AirFareDataSet3$VACATION), mean)
show(aggmean)
```

```
##   Group.1        x
## 1       0 173.5525
## 2       1 125.9809
```

```r
aggmean <- aggregate(AirFareDataSet3$FARE, list(AirFareDataSet3$SW), mean)
show(aggmean)
```

```
##   Group.1         x
## 1       0 188.18279
## 2       1  98.38227
```

```r
aggmean <- aggregate(AirFareDataSet3$FARE, list(AirFareDataSet3$GATE), mean)
show(aggmean)
```

```
##   Group.1       x
## 1       0 153.096
## 2       1 193.129
```

```r
aggmean <- aggregate(AirFareDataSet3$FARE, list(AirFareDataSet3$SLOT), mean)
show(aggmean)
```

```
##   Group.1        x
## 1       0 150.8257
## 2       1 186.0594
```

Looking at the output from the modeling of the categorical variables, the highest is SLOT and the lowest is SW.

---

c. Find a model for predicting the average fare on a new route:
d. Convert categorical variables (e.g., SW) into dummy variables. Then, partition the data into training and validation sets. The model will be fit to the training and evaluated on the validation set.

```
ntotal <- length(airfares.df$FARE)

ntrain <- round(ntotal * 0.6)
nvalid <- ntotal - ntrain
set.seed(202)
ntrain.index <- sort(sample(ntotal, ntrain))
trainingdataset <- airfares.df[ntrain.index, ]
valid.df <- airfares.df[-ntrain.index, ]
```

ii. Use stepwise regression to reduce the number of predictors. You can ignore the first four predictors (S_CODE, S_CITY, E_CODE, E_CITY). Report the estimated model selected.

```
library(leaps)

search <- regsubsets(FARE ~ .,data = trainingdataset,nbest = 1,nvmax = dim(trainingdataset)[2],method =
```

```
## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in =
## force.in, : 10 linear dependencies found
```

```
## Reordering variables and trying again:
```

```
sum <- summary(search)
t(t(sum$adjr2))
```

```
##             [,1]
##  [1,] 0.45814970
##  [2,] 0.61669714
##  [3,] 0.71488909
##  [4,] 0.73796442
##  [5,] 0.75382031
##  [6,] 0.76935648
##  [7,] 0.77944428
##  [8,] 0.79099337
##  [9,] 0.80141893
## [10,] 0.81006567
## [11,] 0.81489309
## [12,] 0.81898660
## [13,] 0.82306915
## [14,] 0.82707533
## [15,] 0.83059180
## [16,] 0.83368268
## [17,] 0.02557734
## [18,] 0.85567346
## [19,] 0.85820310
```

```
models <-  order(sum$adjr2, decreasing = T)[1:3]
```

iii. Repeat (ii) using exhaustive search instead of stepwise regression. Compare the resulting best model to the one you obtained in (ii) in terms of the predictors that are in the model.

```
library(leaps)

#search <- regsubsets(FARE ~ .,data = trainingdataset,nbest = 1,nvmax = dim(trainingdataset)[2],method
```

NOTE: Dr Davalos, I've made multiple attempts to run this method with no success. The amount of memory (even on my high-end machine) cannot allow for the running of really.big and exhaustive method. I understand that this method gives the best prediction but cannot knit or complete the next three steps as I cannot achieve a successful run of this portion. Note: 4 hours of running the code and no success (on multiple attempts). This will impact the completeness and unsuccesful portions of the quesiton. I would have like to continue to finish however the program will not allow and I cannot continue on without the information.

d. In competitive industries, a new entrant with a novel business plan can have a disruptive effect on existing firms. If a new entrant's business model is sustainable, other players are forced to respond by changing their business practices. If the goal of the analysis was to evaluate the effect of Southwest Airlines' presence on the airline industry rather than predicting fares on new routes, how would the analysis be different? Describe technical and conceptual aspects.

I would perform a SWOT analysis on the industry. A SWOT analysis composes of strengths, weaknesses, opportunities, and threats. When understanding the strengths portion, I would look at what characteristics the company is good at. In southwest airlines for instance, I would analyze factors that drive down fare seeing as they are a low cost carrier. Second, for weaknesses, I would also look at fare and understand what may be driving up the fare. Maybe the correlation of a variable like S_INCOME or E_INCOME could impact the fares price in a negative way. Opportunities I would look at analyzing ways to improve the fare of a ticket based on insights of variables like coupon or city. Finally threats I would analyze data like vacation or S_POP to understand if there is a correlation that could be threatening the business model. Ultimately these analysis are conducted on the basis that they should help improve and or assist the businesses success.