Problem 1

Reshms Sekar

August 19, 2015

```
library('ggplot2')
library('XML')
air=read.csv("~/Desktop/ABIA.csv")
summary(air)
```

```
##       Year          Month         DayofMonth      DayOfWeek
##  Min.   :2008   Min.   : 1.00   Min.   : 1.00   Min.   :1.000
##  1st Qu.:2008   1st Qu.: 3.00   1st Qu.: 8.00   1st Qu.:2.000
##  Median :2008   Median : 6.00   Median :16.00   Median :4.000
##  Mean   :2008   Mean   : 6.29   Mean   :15.73   Mean   :3.902
##  3rd Qu.:2008   3rd Qu.: 9.00   3rd Qu.:23.00   3rd Qu.:6.000
##  Max.   :2008   Max.   :12.00   Max.   :31.00   Max.   :7.000
##
##     DepTime       CRSDepTime      ArrTime       CRSArrTime
##  Min.   :   1   Min.   :  55   Min.   :   1   Min.   :   5
##  1st Qu.: 917   1st Qu.: 915   1st Qu.:1107   1st Qu.:1115
##  Median :1329   Median :1320   Median :1531   Median :1535
##  Mean   :1329   Mean   :1320   Mean   :1487   Mean   :1505
##  3rd Qu.:1728   3rd Qu.:1720   3rd Qu.:1903   3rd Qu.:1902
##  Max.   :2400   Max.   :2346   Max.   :2400   Max.   :2400
##  NA's   :1413                  NA's   :1567
##  UniqueCarrier     FlightNum       TailNum      ActualElapsedTime
##  WN     :34876   Min.   :   1          : 1104   Min.   : 22.0
##  AA     :19995   1st Qu.: 640   N678CA :  195   1st Qu.: 57.0
##  CO     : 9230   Median :1465   N511SW :  180   Median :125.0
##  YV     : 4994   Mean   :1917   N526SW :  176   Mean   :120.2
##  B6     : 4798   3rd Qu.:2653   N528SW :  172   3rd Qu.:164.0
##  XE     : 4618   Max.   :9741   N520SW :  168   Max.   :506.0
##  (Other):20749                  (Other):97265   NA's   :1601
##  CRSElapsedTime     AirTime          ArrDelay          DepDelay

##  Min.   : 17.0   Min.   :  3.00   Min.   :-129.000   Min.   :-42.000

##  1st Qu.: 58.0   1st Qu.: 38.00   1st Qu.:  -9.000   1st Qu.: -4.000

##  Median :130.0   Median :105.00   Median :  -2.000   Median :  0.000

##  Mean   :122.1   Mean   : 99.81   Mean   :   7.065   Mean   :  9.171

##  3rd Qu.:165.0   3rd Qu.:142.00   3rd Qu.:  10.000   3rd Qu.:  8.000

##  Max.   :320.0   Max.   :402.00   Max.   : 948.000   Max.   :875.000
```

```
##  NA's   :11      NA's   :1601     NA's   :1601       NA's   :1413

##      Origin           Dest          Distance        TaxiIn
##  AUS    :49623   AUS    :49637   Min.   :  66   Min.   :  0.000
##  DAL    : 5583   DAL    : 5573   1st Qu.: 190   1st Qu.:  4.000
##  DFW    : 5508   DFW    : 5506   Median : 775   Median :  5.000
##  IAH    : 3704   IAH    : 3691   Mean   : 705   Mean   :  6.413
##  PHX    : 2786   PHX    : 2783   3rd Qu.:1085   3rd Qu.:  7.000
##  DEN    : 2719   DEN    : 2673   Max.   :1770   Max.   :143.000
##  (Other):29337   (Other):29397                  NA's   :1567
##     TaxiOut         Cancelled       CancellationCode    Diverted

##  Min.   :  1.00   Min.   :0.00000    :97840          Min.   :0.00000
0
##  1st Qu.:  9.00   1st Qu.:0.00000   A:  719          1st Qu.:0.00000
0
##  Median : 12.00   Median :0.00000   B:  605          Median :0.00000
0
##  Mean   : 13.96   Mean   :0.01431   C:   96          Mean   :0.00182
4
##  3rd Qu.: 16.00   3rd Qu.:0.00000                    3rd Qu.:0.00000
0
##  Max.   :305.00   Max.   :1.00000                    Max.   :1.00000
0
##  NA's   :1419

##   CarrierDelay     WeatherDelay       NASDelay      SecurityDelay
##  Min.   :  0.00   Min.   :  0.00   Min.   :  0.00   Min.   :  0.00
##  1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.00
##  Median :  0.00   Median :  0.00   Median :  2.00   Median :  0.00
##  Mean   : 15.39   Mean   :  2.24   Mean   : 12.47   Mean   :  0.07
##  3rd Qu.: 16.00   3rd Qu.:  0.00   3rd Qu.: 16.00   3rd Qu.:  0.00
##  Max.   :875.00   Max.   :412.00   Max.   :367.00   Max.   :199.00
##  NA's   :79513    NA's   :79513    NA's   :79513    NA's   :79513
##  LateAircraftDelay
##  Min.   :  0.00
##  1st Qu.:  0.00
##  Median :  6.00
##  Mean   : 22.97
##  3rd Qu.: 30.00
##  Max.   :458.00
##  NA's   :79513
```

```
names(air)
```

```
##  [1] "Year"              "Month"             "DayofMonth"
##  [4] "DayOfWeek"         "DepTime"           "CRSDepTime"
##  [7] "ArrTime"           "CRSArrTime"        "UniqueCarrier"
## [10] "FlightNum"         "TailNum"           "ActualElapsedTime"
```

```
## [13] "CRSElapsedTime"    "AirTime"           "ArrDelay"
## [16] "DepDelay"          "Origin"            "Dest"
## [19] "Distance"          "TaxiIn"            "TaxiOut"
## [22] "Cancelled"         "CancellationCode"  "Diverted"
## [25] "CarrierDelay"      "WeatherDelay"      "NASDelay"
## [28] "SecurityDelay"     "LateAircraftDelay"
```

```
attach(air)

Departureflights=subset(air,Origin=='AUS')
head(Departureflights,50)
```

```
##    Year Month DayofMonth DayOfWeek DepTime CRSDepTime ArrTime CRSArr
Time
## 2  2008     1          1         2     555        600     826
 835
## 3  2008     1          1         2     600        600     728
 729
## 4  2008     1          1         2     601        605     727
 750
## 5  2008     1          1         2     601        600     654
 700
## 6  2008     1          1         2     636        645     934
 932
## 7  2008     1          1         2     646        655     735
 750
## 10 2008     1          1         2     654        700    1117
1133
## 11 2008     1          1         2     712        705     805
 805
## 12 2008     1          1         2     715        715     826
 832
## 13 2008     1          1         2     722        726     819
 825
## 14 2008     1          1         2     725        730     844
 901
## 15 2008     1          1         2     735        740    1007
1010
## 16 2008     1          1         2     736        740     838
 840
## 17 2008     1          1         2     737        745     924
 943
## 18 2008     1          1         2     745        755     859
 905
## 21 2008     1          1         2     755        700     854
 821
## 22 2008     1          1         2     755        800    1018
1010
## 23 2008     1          1         2     809        815     859
 912
```

```
## 25 2008     1          1          2     819        825     943
 947
## 29 2008     1          1          2     825        830    1128
1136
## 32 2008     1          1          2     834        835     936
 935
## 33 2008     1          1          2     837        840    1008
1014
## 36 2008     1          1          2     850        850     957
1010
## 38 2008     1          1          2     857        900     946
1000
## 40 2008     1          1          2     903        900     952
1000
## 41 2008     1          1          2     907        915    1003
1013
## 44 2008     1          1          2     921        920    1019
1030
## 45 2008     1          1          2     922        925    1016
1032
## 48 2008     1          1          2     926        925    1307
1325
## 52 2008     1          1          2     957       1000    1053
1057
## 53 2008     1          1          2    1001        955    1124
1130
## 56 2008     1          1          2    1006       1005    1236
1235
## 57 2008     1          1          2    1007        945    1109
1045
## 59 2008     1          1          2    1023       1030    1219
1221
## 60 2008     1          1          2    1024       1025    1330
1344
## 62 2008     1          1          2    1035       1035    1336
1343
## 63 2008     1          1          2    1045       1050    1137
1145
## 64 2008     1          1          2    1045       1050    1149
1155
## 65 2008     1          1          2    1054       1055    1136
1140
## 69 2008     1          1          2    1105       1115    1226
1229
## 71 2008     1          1          2    1107       1110    1219
1230
## 72 2008     1          1          2    1118       1125    1206
1220
## 73 2008     1          1          2    1120       1120    1208
1220
```

```
## 74 2008     1          1         2    1129       1130    1321
1320
## 76 2008     1          1         2    1130       1115    1703
1553
## 79 2008     1          1         2    1136       1140    1413
1415
## 80 2008     1          1         2    1147       1150    1535
1535
## 81 2008     1          1         2    1155       1200    1327
1344
## 82 2008     1          1         2    1200       1200    1308
1325
## 83 2008     1          1         2    1206       1152    1555
1559
##    UniqueCarrier FlightNum TailNum ActualElapsedTime CRSElapsedTime
## 2             AA      1614  N438AA               151             155
## 3             YV      2883  N922FJ               148             149
## 4             9E      5743  89189E                86             105
## 5             AA      1157  N4XAAA                53              60
## 6             NW      1674   N967N               178             167
## 7             CO       340  N14604                49              55
## 10            B6      1060  N238JB               203             213
## 11            AA       652  N432AA                53              60
## 12            XE       519  N11194                71              77
## 13            CO      1573  N69351                57              59
## 14            XE       303  N13553               139             151
## 15            WN      3098  N750SA               152             150
## 16            AA      1958  N526AA                62              60
## 17            UA       657  N817UA               227             238
## 18            AA      1465  N4XPAA               194             190
## 21            XE         1  N12166               179             201
## 22            OO      4009  N368CA               203             190
## 23            CO       440  N17229                50              57
## 25            UA      1190  N374UA               144             142
## 29            OH      5124  N692CA               123             126
## 32            AA       450  N406AA                62              60
## 33            US       435  N158AW               151             154
## 36            AA       379  N455AA               187             200
## 38            AA      1743  N3CFAA                49              60
## 40            WN      2677  N496WN               169             180
## 41            CO       640  N79402                56              58
## 44            WN      2949  N611SW               178             190
## 45            XE       311  N18557               114             127
## 48            WN      3489  N462WN               161             180
## 52            CO      1572  N59338                56              57
## 53            WN       513  N505SW               143             155
## 56            AA       368  N4WTAA               150             150
## 57            AA       511  N404AA                62              60
## 59            XE       532  N11544               116             111
## 60            EV      4324  N707EV               126             139
```

```
## 62             EV      4338  N977EV               121             128
## 63             WN        61  N609SW                52              55
## 64             AA      2024  N474AA                64              65
## 65             WN      2577  N738CB                42              45
## 69             F9       215  N805FR               141             134
## 71             AA       813  N526AA               192             200
## 72             MQ      3488  N658AE                48              55
## 73             WN      1838  N247WN               168             180
## 74             OO      2931  N507CA               112             110
## 76             OH      5202  N679CA               273             218
## 79             AA      2486  N491AA               157             155
## 80             XE      2366  N27152               168             165
## 81             NW      1727  N605NW                92             104
## 82             WN      2985  N795SW               188             205
## 83             YV      7276  N509MJ               169             187
##    AirTime ArrDelay DepDelay Origin Dest Distance TaxiIn TaxiOut Can
celled
## 2      133       -9       -5    AUS  ORD      978      7      11
     0
## 3      125       -1        0    AUS  PHX      872      7      16
     0
## 4       70      -23       -4    AUS  MEM      559      4      12
     0
## 5       38       -6        1    AUS  DFW      190      5      10
     0
## 6      145        2       -9    AUS  MSP     1042     11      22
     0
## 7       28      -15       -9    AUS  IAH      140      6      15
     0
## 10     177      -16       -6    AUS  JFK     1522     13      13
     0
## 11      36        0        7    AUS  DFW      190      6      11
     0
## 12      56       -6        0    AUS  MSY      445      5      10
     0
## 13      28       -6       -4    AUS  IAH      140     13      16
     0
## 14     121      -17       -5    AUS  TUS      797      4      14
     0
## 15     123       -3       -5    AUS  MDW      972     15      14
     0
## 16      38       -2       -4    AUS  DFW      190     13      11
     0
## 17     203      -19       -8    AUS  SFO     1504      3      21
     0
## 18     179       -6      -10    AUS  SNA     1209      4      11
     0
## 21     163       33       55    AUS  ONT     1197      4      12
     0
## 22     171        8       -5    AUS  SLC     1085     18      14
```

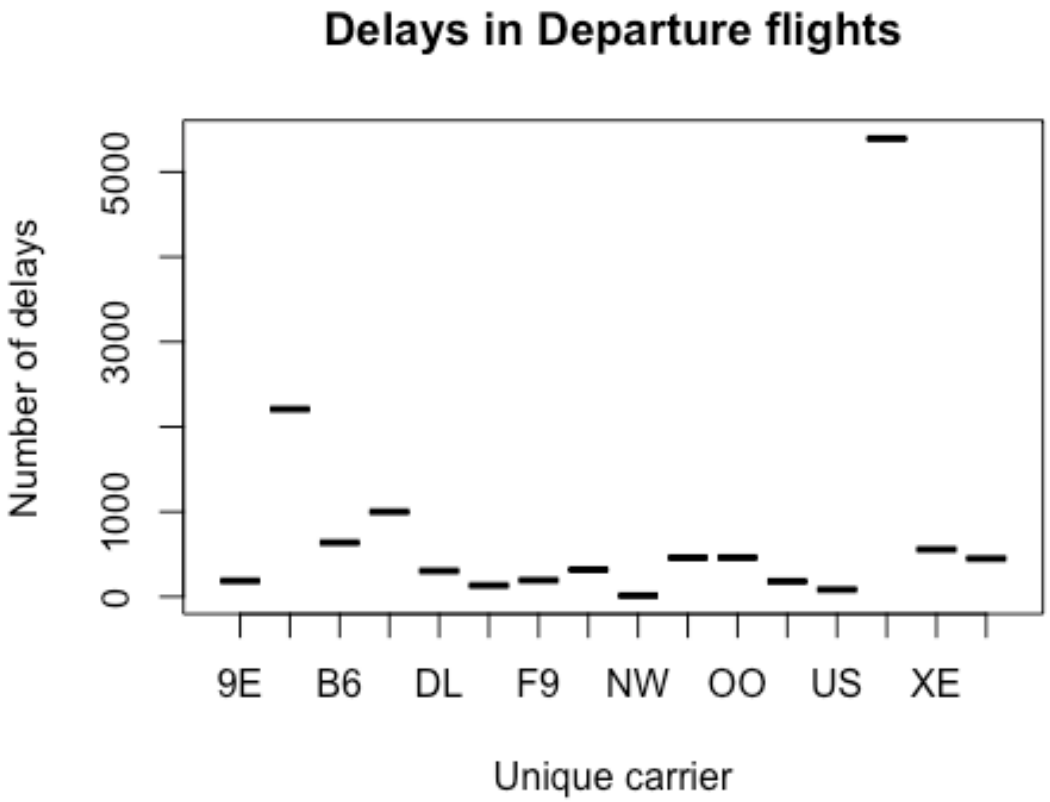
```
     0
## 23      31       -13        -6    AUS  IAH      140       5       14
     0
## 25     129        -4        -6    AUS  DEN      775       5       10
     0
## 29      93        -8        -5    AUS  ATL      813       8       22
     0
## 32      35         1        -1    AUS  DFW      190      13       14
     0
## 33     130        -6        -3    AUS  PHX      872       9       12
     0
## 36     173       -13         0    AUS  LAX     1242       4       10
     0
## 38      35       -14        -3    AUS  DFW      190       5        9
     0
## 40     151        -8         3    AUS  LAS     1090       7       11
     0
## 41      30       -10        -8    AUS  IAH      140       9       17
     0
## 44     163       -11         1    AUS  SAN     1164       3       12
     0
## 45     100       -16        -3    AUS  ABQ      619       4       10
     0
## 48     141       -18         1    AUS  BWI     1342      10       10
     0
## 52      32        -4        -3    AUS  IAH      140      10       14
     0
## 53     126        -6         6    AUS  PHX      872       8        9
     0
## 56     131         1         1    AUS  ORD      978       8       11
     0
## 57      34        24        22    AUS  DFW      190      15       13
     0
## 59     101        -2        -7    AUS  MCI      650       3       12
     0
## 60     112       -14        -1    AUS  CVG      958       5        9
     0
## 62      95        -7         0    AUS  ATL      813       9       17
     0
## 63      39        -8        -5    AUS  DAL      189       2       11
     0
## 64      38        -6        -5    AUS  DFW      190      10       16
     0
## 65      32        -4        -1    AUS  HOU      148       2        8
     0
## 69     126        -3       -10    AUS  DEN      775       5       10
     0
## 71     170       -11        -3    AUS  LAX     1242      12       10
     0
## 72      37       -14        -7    AUS  DAL      189       2        9
```

```
    0
## 73     154       -12         0    AUS  LAS     1090       5        9
    0
## 74      94         1        -1    AUS  MCI      650       5       13
    0
## 76     209        70        15    AUS  JFK     1522      48       16
    0
## 79     136        -2        -4    AUS  ORD      978       4       17
    0
## 80     151         0        -3    AUS  CLE     1174       6       11
    0
## 81      74       -17        -5    AUS  MEM      559       5       13
    0
## 82     170       -17         0    AUS  LAX     1242       8       10
    0
## 83     148        -4        14    AUS  IAD     1297       7       14
    0
##    CancellationCode Diverted CarrierDelay WeatherDelay NASDelay
## 2                           0           NA           NA       NA
## 3                           0           NA           NA       NA
## 4                           0           NA           NA       NA
## 5                           0           NA           NA       NA
## 6                           0           NA           NA       NA
## 7                           0           NA           NA       NA
## 10                          0           NA           NA       NA
## 11                          0           NA           NA       NA
## 12                          0           NA           NA       NA
## 13                          0           NA           NA       NA
## 14                          0           NA           NA       NA
## 15                          0           NA           NA       NA
## 16                          0           NA           NA       NA
## 17                          0           NA           NA       NA
## 18                          0           NA           NA       NA
## 21                          0           33            0        0
## 22                          0           NA           NA       NA
## 23                          0           NA           NA       NA
## 25                          0           NA           NA       NA
## 29                          0           NA           NA       NA
## 32                          0           NA           NA       NA
## 33                          0           NA           NA       NA
## 36                          0           NA           NA       NA
## 38                          0           NA           NA       NA
## 40                          0           NA           NA       NA
## 41                          0           NA           NA       NA
## 44                          0           NA           NA       NA
## 45                          0           NA           NA       NA
## 48                          0           NA           NA       NA
## 52                          0           NA           NA       NA
## 53                          0           NA           NA       NA
## 56                          0           NA           NA       NA
```

```
## 57                          0             0             0         2
## 59                          0            NA            NA        NA
## 60                          0            NA            NA        NA
## 62                          0            NA            NA        NA
## 63                          0            NA            NA        NA
## 64                          0            NA            NA        NA
## 65                          0            NA            NA        NA
## 69                          0            NA            NA        NA
## 71                          0            NA            NA        NA
## 72                          0            NA            NA        NA
## 73                          0            NA            NA        NA
## 74                          0            NA            NA        NA
## 76                          0            15             0        55
## 79                          0            NA            NA        NA
## 80                          0            NA            NA        NA
## 81                          0            NA            NA        NA
## 82                          0            NA            NA        NA
## 83                          0            NA            NA        NA
##    SecurityDelay LateAircraftDelay
## 2             NA                NA
## 3             NA                NA
## 4             NA                NA
## 5             NA                NA
## 6             NA                NA
## 7             NA                NA
## 10            NA                NA
## 11            NA                NA
## 12            NA                NA
## 13            NA                NA
## 14            NA                NA
## 15            NA                NA
## 16            NA                NA
## 17            NA                NA
## 18            NA                NA
## 21             0                 0
## 22            NA                NA
## 23            NA                NA
## 25            NA                NA
## 29            NA                NA
## 32            NA                NA
## 33            NA                NA
## 36            NA                NA
## 38            NA                NA
## 40            NA                NA
## 41            NA                NA
## 44            NA                NA
## 45            NA                NA
## 48            NA                NA
## 52            NA                NA
## 53            NA                NA
```

```
## 56             NA                 NA
## 57              0                 22
## 59             NA                 NA
## 60             NA                 NA
## 62             NA                 NA
## 63             NA                 NA
## 64             NA                 NA
## 65             NA                 NA
## 69             NA                 NA
## 71             NA                 NA
## 72             NA                 NA
## 73             NA                 NA
## 74             NA                 NA
## 76              0                  0
## 79             NA                 NA
## 80             NA                 NA
## 81             NA                 NA
## 82             NA                 NA
## 83             NA                 NA
```

```
Depflightdelay=subset(Departureflights,DepDelay>4)
departflightdelay=aggregate(Depflightdelay$DepDelay,list(Depflightdelay
$UniqueCarrier),length)
plot(departflightdelay,main='Delays in Departure flights',xlab='Unique
carrier',ylab='Number of delays')
```



Creating a subset of arrival flights

```
Arrivalflights=subset(air,Dest=='AUS')
head(Arrivalflights,50)
```

```
##      Year Month DayofMonth DayOfWeek DepTime CRSDepTime ArrTime CRSAr
rTime
## 1   2008     1          1         2     120       1935     309
 2130
## 8   2008     1          1         2     650        700     841
  857
## 9   2008     1          1         2     650        650    1139
 1145
## 19  2008     1          1         2     748        755    1007
 1020
## 20  2008     1          1         2     753        755    1137
 1125
## 24  2008     1          1         2     811        815    1048
 1100
## 26  2008     1          1         2     819        820     944
 1003
## 27  2008     1          1         2     822        810    1255
 1250
```

```
## 28  2008     1          1          2     823        825     916
  920
## 30  2008     1          1          2     827        830     958
 1009
## 31  2008     1          1          2     832        800     929
  900
## 34  2008     1          1          2     840        842    1138
 1118
## 35  2008     1          1          2     846        850    1317
 1330
## 37  2008     1          1          2     851        855    1119
 1109
## 39  2008     1          1          2     858        905    1053
 1100
## 42  2008     1          1          2     909        915    1205
 1210
## 43  2008     1          1          2     912        850    1305
 1235
## 46  2008     1          1          2     923        923    1111
 1119
## 47  2008     1          1          2     923        925    1014
 1025
## 49  2008     1          1          2     926        850    1402
 1332
## 50  2008     1          1          2     941        945    1042
 1050
## 51  2008     1          1          2     942        945    1116
 1130
## 54  2008     1          1          2    1001       1010    1045
 1100
## 55  2008     1          1          2    1001       1004    1058
 1101
## 58  2008     1          1          2    1016       1015    1501
 1505
## 61  2008     1          1          2    1028       1030    1344
 1352
## 66  2008     1          1          2    1055       1105    1329
 1349
## 67  2008     1          1          2    1058       1100    1155
 1157
## 68  2008     1          1          2    1059       1040    1150
 1140
## 70  2008     1          1          2    1105       1105    1406
 1420
## 75  2008     1          1          2    1129       1130    1226
 1230
## 77  2008     1          1          2    1133       1120    1356
 1345
## 78  2008     1          1          2    1135       1135    1221
 1225
```

```
## 84  2008     1          1         2    1206       1136    1739
 1701
## 86  2008     1          1         2    1209       1210    1356
 1410
## 90  2008     1          1         2    1226       1215    1739
 1735
## 92  2008     1          1         2    1235       1235    1348
 1335
## 95  2008     1          1         2    1241       1230    1512
 1515
## 97  2008     1          1         2    1243       1240    1716
 1720
## 98  2008     1          1         2    1244       1245    1329
 1340
## 101 2008     1          1         2    1302       1300    1436
 1438
## 102 2008     1          1         2    1308       1316    1614
 1638
## 103 2008     1          1         2    1309       1305    1800
 1805
## 104 2008     1          1         2    1316       1300    1504
 1500
## 106 2008     1          1         2    1320       1310    1526
 1508
## 107 2008     1          1         2    1334       1335    1700
 1649
## 109 2008     1          1         2    1340       1342    1715
 1725
## 111 2008     1          1         2    1347       1330    1433
 1415
## 112 2008     1          1         2    1350       1345    1637
 1645
## 114 2008     1          1         2    1355       1346    1451
 1439
##     UniqueCarrier FlightNum TailNum ActualElapsedTime CRSElapsedTime
## 1              9E      5746  84129E                109            115
## 8              XE       541  N18557                111            117
## 9              AA      1182  N4WAAA                169            175
## 19             WN        61  N609SW                 79             85
## 20             B6      1061  N179JB                284            270
## 24             AA      1199  N491AA                157            165
## 26             XE       511  N11544                 85            103
## 27             WN       297  N723SW                153            160
## 28             CO      1583  N59338                 53             55
## 30             EV      4677  N977EV                151            159
## 31             AA      1109  N404AA                 57             60
## 34             YV      7273  N509MJ                238            216
## 35             WN      3222  N639SW                151            160
## 37             XE      2252  N27152                208            194
## 39             OO      2930  N507CA                115            115
```

```
## 42              WN      3481  N704SW               236              235
## 43              B6      1263  N273JB               293              285
## 46              NW      1728  N605NW               108              116
## 47              AA       421  N526AA                51               60
## 49              XE         2  N12166               156              162
## 50              WN      1021  N247WN                61               65
## 51              WN      2985  N795SW               154              165
## 54              MQ      3355  N658AE                44               50
## 55              CO       441  N19621                57               57
## 58              WN       304  N638SW               165              170
## 61              US       225  N155AW               136              142
## 66              XE       531  N18557                94              104
## 67              CO       741  N14653                57               57
## 68              AA      1477  N533AA                51               60
## 70              WN      1105  N614SW               121              135
## 75              AA      1717  N484AA                57               60
## 77              WN      3152  N301SW                83               85
## 78              WN       501  N527SW                46               50
## 84              UA       374  N834UA               213              205
## 86              XE       308  N11194               167              180
## 90              AA      1024  N530AA               193              200
## 92              AA       668  N552AA                73               60
## 95              AA      1545  N580AA               151              165
## 97              WN       155  N394SW               153              160
## 98              MQ      3412  N649PP                45               55
## 101             OO      1991  N806SK               154              158
## 102             US       227  N311AW               126              142
## 103             AA      1308  N565AA               171              180
## 104             WN      1885  N688SW               168              180
## 106             OH      5339  N709CA               186              178
## 107             B6      1065  N265JB               266              254
## 109             OO      4052  N668CA               155              163
## 111             WN      2124  N364SW                46               45
## 112             WN      1856  N483WN               107              120
## 114             CO       241  N59630                56               53
##     AirTime ArrDelay DepDelay Origin Dest Distance TaxiIn TaxiOut
## 1        88      339      345    MEM  AUS      559      3      18
## 8        94      -16      -10    MCI  AUS      650      6      11
## 9       153       -6        0    LAX  AUS     1242      4      12
## 19       68      -13       -7    ELP  AUS      528      3       8
## 20      257       12       -2    JFK  AUS     1522      4      23
## 24      139      -12       -4    ORD  AUS      978      4      14
## 26       74      -19       -1    MSY  AUS      445      4       7
## 27      139        5       12    SAN  AUS     1164      4      10
## 28       28       -4       -2    IAH  AUS      140      5      20
## 30      131      -11       -3    ATL  AUS      813      3      17
## 31       36       29       32    DFW  AUS      190      4      17
## 34      218       20       -2    IAD  AUS     1297      4      16
## 35      136      -13       -4    LAS  AUS     1090      4      11
## 37      185       10       -4    CLE  AUS     1174      5      18
```

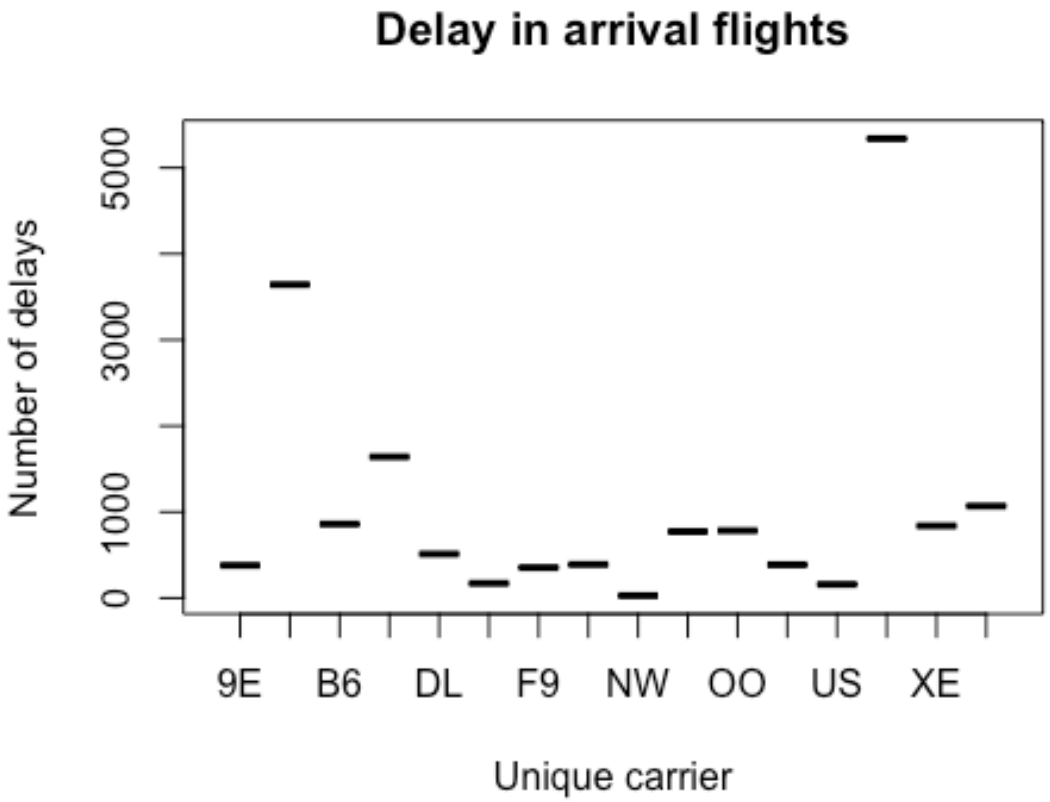
```
## 39       92       -7       -7    MCI  AUS      650      6      17
## 42      224       -5       -6    BWI  AUS     1342      5       7
## 43      268       30       22    BOS  AUS     1698      3      22
## 46       93       -8        0    MEM  AUS      559      3      12
## 47       37      -11       -2    DFW  AUS      190      2      12
## 49      144       30       36    ONT  AUS     1197      5       7
## 50       48       -8       -4    LBB  AUS      341      5       8
## 51      141      -14       -3    TPA  AUS      928      3      10
## 54       34      -15       -9    DAL  AUS      189      3       7
## 55       27       -3       -3    IAH  AUS      140      6      24
## 58      153       -4        1    LAX  AUS     1242      4       8
## 61      108       -8       -2    PHX  AUS      872      4      24
## 66       82      -20      -10    ABQ  AUS      619      5       7
## 67       28       -2       -2    IAH  AUS      140      5      24
## 68       33       10       19    DFW  AUS      190      3      15
## 70      107      -14        0    PHX  AUS      872      3      11
## 75       35       -4       -1    DFW  AUS      190      4      18
## 77       71       11       13    ELP  AUS      528      5       7
## 78       35       -4        0    DAL  AUS      189      4       7
## 84      185       38       30    SFO  AUS     1504      5      23
## 86      149      -14       -1    JAX  AUS      954      4      14
## 90      180        4       11    SJC  AUS     1476      3      10
## 92       36       13        0    DFW  AUS      190      4      33
## 95      136       -3       11    ORD  AUS      978      3      12
## 97      133       -4        3    LAS  AUS     1090      5      15
## 98       33      -11       -1    DAL  AUS      189      3       9
## 101     134       -2        2    ATL  AUS      813      5      15
## 102     108      -24       -8    PHX  AUS      872      4      14
## 103     151       -5        4    LAX  AUS     1242      7      13
## 104     151        4       16    MCO  AUS      993      3      14
## 106     155       18       10    CVG  AUS      958      7      24
## 107     246       11       -1    JFK  AUS     1522      3      17
## 109     128      -10       -2    SLC  AUS     1085      5      22
## 111      28       18       17    HOU  AUS      148      4      14
## 112      92       -8        5    DEN  AUS      775      4      11
## 114      29       12        9    IAH  AUS      140      6      21
##     Cancelled CancellationCode Diverted CarrierDelay WeatherDelay NA
SDelay
## 1           0                        0          339            0
     0
## 8           0                        0           NA           NA
    NA
## 9           0                        0           NA           NA
    NA
## 19          0                        0           NA           NA
    NA
## 20          0                        0           NA           NA
    NA
## 24          0                        0           NA           NA
    NA
```

```
## 26          0                       0          NA          NA
    NA
## 27          0                       0          NA          NA
    NA
## 28          0                       0          NA          NA
    NA
## 30          0                       0          NA          NA
    NA
## 31          0                       0          29           0
     0
## 34          0                       0           0           0
    20
## 35          0                       0          NA          NA
    NA
## 37          0                       0          NA          NA
    NA
## 39          0                       0          NA          NA
    NA
## 42          0                       0          NA          NA
    NA
## 43          0                       0          10           0
     8
## 46          0                       0          NA          NA
    NA
## 47          0                       0          NA          NA
    NA
## 49          0                       0          10           0
     0
## 50          0                       0          NA          NA
    NA
## 51          0                       0          NA          NA
    NA
## 54          0                       0          NA          NA
    NA
## 55          0                       0          NA          NA
    NA
## 58          0                       0          NA          NA
    NA
## 61          0                       0          NA          NA
    NA
## 66          0                       0          NA          NA
    NA
## 67          0                       0          NA          NA
    NA
## 68          0                       0          NA          NA
    NA
## 70          0                       0          NA          NA
    NA
## 75          0                       0          NA          NA
    NA
```

```
## 77           0                          0             NA             NA
    NA
## 78           0                          0             NA             NA
    NA
## 84           0                          0             21              0
     8
## 86           0                          0             NA             NA
    NA
## 90           0                          0             NA             NA
    NA
## 92           0                          0             NA             NA
    NA
## 95           0                          0             NA             NA
    NA
## 97           0                          0             NA             NA
    NA
## 98           0                          0             NA             NA
    NA
## 101          0                          0             NA             NA
    NA
## 102          0                          0             NA             NA
    NA
## 103          0                          0             NA             NA
    NA
## 104          0                          0             NA             NA
    NA
## 106          0                          0             10              0
     8
## 107          0                          0             NA             NA
    NA
## 109          0                          0             NA             NA
    NA
## 111          0                          0              4              0
     1
## 112          0                          0             NA             NA
    NA
## 114          0                          0             NA             NA
    NA
##     SecurityDelay LateAircraftDelay
## 1               0                 0
## 8              NA                NA
## 9              NA                NA
## 19             NA                NA
## 20             NA                NA
## 24             NA                NA
## 26             NA                NA
## 27             NA                NA
## 28             NA                NA
## 30             NA                NA
## 31              0                 0
```

```
## 34                0                  0
## 35               NA                 NA
## 37               NA                 NA
## 39               NA                 NA
## 42               NA                 NA
## 43                0                 12
## 46               NA                 NA
## 47               NA                 NA
## 49                0                 20
## 50               NA                 NA
## 51               NA                 NA
## 54               NA                 NA
## 55               NA                 NA
## 58               NA                 NA
## 61               NA                 NA
## 66               NA                 NA
## 67               NA                 NA
## 68               NA                 NA
## 70               NA                 NA
## 75               NA                 NA
## 77               NA                 NA
## 78               NA                 NA
## 84                0                  9
## 86               NA                 NA
## 90               NA                 NA
## 92               NA                 NA
## 95               NA                 NA
## 97               NA                 NA
## 98               NA                 NA
## 101              NA                 NA
## 102              NA                 NA
## 103              NA                 NA
## 104              NA                 NA
## 106               0                  0
## 107              NA                 NA
## 109              NA                 NA
## 111               0                 13
## 112              NA                 NA
## 114              NA                 NA
```

```
Arrflightdelay=subset(Arrivalflights,ArrDelay>4)
Arrflightdelay=aggregate(Arrflightdelay$ArrDelay,list(Arrflightdelay$Un
iqueCarrier),length)
plot(Arrflightdelay,main='Delay in arrival flights',xlab='Unique carrie
r',ylab='Number of delays')
```



From the above two plots we can infer which airlines has the most number of delays and which airlines has the least number of delays, which can help us choose an airline we favor for travel.

Problem 2

Reshms Sekar

August 19, 2015

```
library('tm')

## Loading required package: NLP

library('randomForest')

## randomForest 4.6-10
## Type rfNews() to see new features/changes/bug fixes.
```

```
library('e1071')
library('rpart')
library('ggplot2')

##
## Attaching package: 'ggplot2'
##
## The following object is masked from 'package:NLP':
##
##     annotate

library('caret')

## Loading required package: lattice

setwd("~/Desktop")
#reader function
readerPlain = function(fname){
  readPlain(elem=list(content=readLines(fname)), id=fname, language='en
') }

author_dirs = Sys.glob('~/Desktop/ReutersC50/C50train/*')
file_list = NULL
train_labels = NULL
for(author in author_dirs) {
  author_name = substring(author, first=23)
  files_to_add = Sys.glob(paste0(author, '/*.txt'))
  file_list = append(file_list, files_to_add)
  train_labels = append(train_labels, rep(author_name, length(files_to_
add)))
}

# Named conversion & cleanup
all_docs = lapply(file_list, readerPlain)
#names(all_docs) = file_list
#names(all_docs) = sub('.txt', '', names(all_docs))

#Initialize Training Corpus
train_corpus = Corpus(VectorSource(all_docs))
#names(train_corpus) = file_list

#Tokenization of training Corpus
train_corpus = tm_map(train_corpus, content_transformer(tolower))
train_corpus = tm_map(train_corpus, content_transformer(removeNumbers))

train_corpus = tm_map(train_corpus, content_transformer(removePunctuati
on))
train_corpus = tm_map(train_corpus, content_transformer(stripWhitespace
))
train_corpus = tm_map(train_corpus, content_transformer(removeWords), s
topwords("SMART"))
```

```
#Create training DTM & dense matrix
DTM_train = DocumentTermMatrix(train_corpus)
DTM_train = removeSparseTerms(DTM_train, 0.975)


author_dirs = Sys.glob('~/Desktop/ReutersC50/C50test/*')
file_list = NULL
test_labels = NULL
for(author in author_dirs) {
  author_name = substring(author, first=22)
  files_to_add = Sys.glob(paste0(author, '/*.txt'))
  file_list = append(file_list, files_to_add)
  test_labels = append(test_labels, rep(author_name, length(files_to_ad
d)))
}

# Named conversion & cleanup
all_docs = lapply(file_list, readerPlain)
#names(all_docs) = file_list
#names(all_docs) = sub('.txt', '', names(all_docs))

#Initialize Testing Corpus
test_corpus = Corpus(VectorSource(all_docs))
#names(test_corpus) = file_list

#Tokenization of Testing Corpus
test_corpus = tm_map(test_corpus, content_transformer(tolower))
test_corpus = tm_map(test_corpus, content_transformer(removeNumbers))
test_corpus = tm_map(test_corpus, content_transformer(removePunctuation
))
test_corpus = tm_map(test_corpus,content_transformer(stripWhitespace))
test_corpus = tm_map(test_corpus, content_transformer(removeWords), sto
pwords("SMART"))

reuters_dict = NULL
reuters_dict = dimnames(DTM_train)[[2]]

#Create testing DTM & matrix using dictionary words only
DTM_test = DocumentTermMatrix(test_corpus, list(dictionary=reuters_dict
))
DTM_test = removeSparseTerms(DTM_test, 0.975)
#DTM_test = as.matrix(DTM_test)
DTM_train_df = as.data.frame(inspect(DTM_train))
```

The above two models used are Naives Bayes and Random Forests. Naives Bayes gave an accuracy of 62% which is not as good as Random Forests, therefore Random Forests is a better model.

Problem 3

Reshms Sekar

August 19, 2015

PROBLEM 3

```
# Association rule mining
# Adapted from code by Matt Taddy
library('arules')  # has a big ecosystem of packages built around it
```

```
## Loading required package: Matrix
##
## Attaching package: 'arules'
##
## The following objects are masked from 'package:base':
##
##     %in%, write
```

```
# Read in playlists from users
groceries <- read.csv("~/Desktop/STA380-master/data/groceries.txt")


# First create a list of baskets: vectors of items by consumer
# Analagous to bags of words

# First split data into a list of artists for each user
groceries <- split(groceries,f=",")
## Remove duplicates ("de-dupe")
groceries <- lapply(groceries, unique)

## Cast this variable as a special arules "transactions" class.
groceriestrans <- read.transactions("~/Desktop/STA380-master/data/groce
ries.txt",format=c("basket"),sep=",",encoding="unknown",rm.duplicates=T
RUE)
# Now run the 'apriori' algorithm
# Look at rules with support > .01 & confidence >.5 & length (# artists
) <= 4
grocrules <- apriori(groceriestrans,
                     parameter=list(support=.01, confidence=.5, maxlen=
4))
```

```
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport support minlen ma
xlen
##         0.5    0.1    1 none FALSE            TRUE    0.01      1
   4
```

```
##  target   ext
##   rules FALSE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## apriori - find association rules with the apriori algorithm
## version 4.21 (2004.05.09)        (c) 1996-2004   Christian Borgelt
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [88 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [15 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
#detach(package:tm, unload)=TRUE
# Look at the output
inspect(grocrules)
```

```
##    lhs                     rhs                   support confidence
    lift
## 1  {curd,

##     yogurt}             => {whole milk}       0.01006609  0.5823529
2.279125
## 2  {butter,

##     other vegetables}   => {whole milk}       0.01148958  0.5736041
2.244885
## 3  {domestic eggs,

##     other vegetables}   => {whole milk}       0.01230300  0.5525114
2.162336
## 4  {whipped/sour cream,

##     yogurt}             => {whole milk}       0.01087951  0.5245098
2.052747
## 5  {other vegetables,

##     whipped/sour cream} => {whole milk}       0.01464159  0.5070423
1.984385
## 6  {other vegetables,

##     pip fruit}          => {whole milk}       0.01352313  0.5175097
2.025351
## 7  {citrus fruit,

##     root vegetables}    => {other vegetables} 0.01037112  0.5862069
```

```
3.029608
## 8  {root vegetables,

##     tropical fruit}     => {other vegetables} 0.01230300  0.5845411
3.020999
## 9  {root vegetables,

##     tropical fruit}     => {whole milk}       0.01199797  0.5700483
2.230969
## 10 {tropical fruit,

##     yogurt}             => {whole milk}       0.01514997  0.5173611
2.024770
## 11 {root vegetables,

##     yogurt}             => {other vegetables} 0.01291307  0.5000000
2.584078
## 12 {root vegetables,

##     yogurt}             => {whole milk}       0.01453991  0.5629921
2.203354
## 13 {rolls/buns,

##     root vegetables}    => {other vegetables} 0.01220132  0.5020921
2.594890
## 14 {rolls/buns,

##     root vegetables}    => {whole milk}       0.01270971  0.5230126
2.046888
## 15 {other vegetables,

##     yogurt}             => {whole milk}       0.02226741  0.5128806
2.007235

## Choose a subset
inspect(subset(grocrules, subset=support > .01 & confidence > 0.55))

##   lhs                   rhs                   support confidence
 lift
## 1 {curd,

##    yogurt}           => {whole milk}       0.01006609  0.5823529 2.2
79125
## 2 {butter,

##    other vegetables} => {whole milk}       0.01148958  0.5736041 2.2
44885
## 3 {domestic eggs,

##    other vegetables} => {whole milk}       0.01230300  0.5525114 2.1
```

```
62336
## 4 {citrus fruit,

##    root vegetables}  => {other vegetables} 0.01037112  0.5862069 3.0
29608
## 5 {root vegetables,

##    tropical fruit}   => {other vegetables} 0.01230300  0.5845411 3.0
20999
## 6 {root vegetables,

##    tropical fruit}   => {whole milk}       0.01199797  0.5700483 2.2
30969
## 7 {root vegetables,

##    yogurt}           => {whole milk}       0.01453991  0.5629921 2.2
03354
```

From the above we can infer that whole milk is brought very often in combination with curd, yoghurt and butter and vegetables, eggs and vegetables and other vegetables are brought a lot often in combination with root vegetables and fruits