Code to Extract the ABC file

The code below can be used to extrat the ABC file base on data length

```r
rm(list=ls(all=TRUE)) #start with empty workspace
startTime <- Sys.time()

temp <- tempfile()

starTime <- Sys.time()

download.file("http://www.abc.ca.gov/datport/ABC_Data_Export.zip",temp)
con <- unz(temp, "m_tape460.LST")
data <-  read.fwf(con, fill = TRUE, comment.char = "",
  widths = c(2,8,3,8,11,11,8,3,1,2,4,2,50,50,50,25,2,10,50,50,50,25,2,10,16,7))
unlink(temp)

# add the variable names to the file

names(data) <- c("License Type","File Number","License Or Application",
                "Type Status","Type Orig Issue Dates","Expiration Dates",
                "Fee Codes","Duplicate Counts","Master Indicator",
                "Term in Months","Geo Code","District Office",
                "Primary Name","Premise Street Address1",
                "Premise Street Address2","Premise City",
                "Premise State","Premise Zip","DBA name",
                "Mail Street Address1","Mail Street Address2",
                "Mail City","Mail State","Mail Zip","Premise County","Premise Census Tract")

# take a look at the file

summary(data)
```

```
##    License Type     File Number      License Or Application
## 41     :25853   00270771:    20   0:0:      1
## 47     :17155   00252014:    18   APP:10108
## 20     :16250   00331049:    15   LIC:99776
## 21     :15216   00378663:    15
## 02     : 6148   00337074:    14
## 58     : 5537   00355808:    12
## (Other):23726   (Other) :109791
##    Type Status      Type Orig Issue Dates    Expiration Dates
## ACTIVE :100732                  :10179    30-JUN-2015:20193
## PEND   :  6050   01-JAN-1994:   581                  :10179
## SUREND :  2615   22-JUN-2009:   376    31-MAY-2015: 7623
## REVPEN :   211   08-NOV-2013:   335    29-FEB-2016: 7622
## SUSPEN :   193   20-MAY-2008:   174    31-AUG-2015: 7315
## R64B   :    79   (Other)    :98239    (Other)    :56952
## (Other):     5   NA's       :    1    NA's       :    1
##    Fee Codes     Duplicate Counts Master Indicator Term in Months
## P40    :69676   Min.   :  0.00   N  :13391        Min.   :12
## P0     :26393   1st Qu.:  1.00   Y  :96493        1st Qu.:12
```

```
##  P20     : 8657  Median :  1.00  NA's:    1       Median :12
##  PO-GL5K : 2163  Mean   :  1.41                   Mean   :12
##  PO-GL20K:  703  3rd Qu.:  1.00                   3rd Qu.:12
##  (Other) : 2292  Max.   :583.00                   Max.   :12
##  NA's    :    1  NA's   :96650                     NA's   :1
##     Geo Code     District Office
##  Min.   : 100   Min.   : 2.00
##  1st Qu.:1933   1st Qu.: 7.00
##  Median :3100   Median :13.00
##  Mean   :3057   Mean   :16.53
##  3rd Qu.:3902   3rd Qu.:24.00
##  Max.   :9999   Max.   :75.00
##  NA's   :119    NA's   :2
##                                       Primary Name
##  7 ELEVEN INC                           :  1593
##  GARFIELD BEACH CVS LLC                 :   865
##  SAFEWAY INC                            :   599
##  THRIFTY PAYLESS INC                    :   586
##  WALGREEN CO                            :   457
##  (Other)                                :105784
##  NA's                                   :     1
##                                  Premise Street Address1
##                                         :   118
##  930 MCLAUGHLIN AVE                     :    81
##  20580 8TH ST E                         :    65
##  875 HANNA DR                           :    61
##  SF INTL AIRPORT                        :    58
##  (Other)                                :109501
##  NA's                                   :     1
##                                  Premise Street Address2
##                                         :89872
##  STE A                                  : 1563
##  STE B                                  :  916
##  STE C                                  :  592
##  STE 100                                :  493
##  (Other)                                :16448
##  NA's                                   :    1
##               Premise City   Premise State      Premise Zip
##  LOS ANGELES     : 5686   CA     :108393   92101   :   474
##  SAN FRANCISCO   : 5191          :   249   94558   :   337
##  SAN DIEGO       : 4212   WA     :   247           :   249
##  NAPA            : 2213   OR     :   222   93940   :   244
##  SAN JOSE        : 2096   NY     :   120   94574   :   243
##  (Other)         :90486   (Other):   653   (Other) :108337
##  NA's            :    1   NA's   :     1   NA's    :     1
##                                         DBA name
##                                         :  5215
##  ROUND TABLE PIZZA                      :   319
##  CHIPOTLE MEXICAN GRILL                 :   221
##  GROCERY OUTLET                         :   211
##  BEVMO                                  :   192
##  (Other)                                :103726
##  NA's                                   :     1
##                                    Mail Street Address1
```

```
##                                                 :51738
## PO BOX 219088                                   : 1579
## PO BOX 29096                                    :  892
## 1 CVS DR                                        :  846
## 2600 CAPITOL AVE                                :  579
## (Other)                                         :54250
## NA's                                            :    1
##                                Mail Street Address2
##                                       :89203
## ATT: 7 ELEVEN LICENSING               : 1574
## STE 300                               : 1252
## MAIL DROP 23062A                      :  849
## STE 100                               :  771
## (Other)                               :16235
## NA's                                  :    1
##             Mail City       Mail State        Mail Zip
##                 :51755          :51755           :51755
## LOS ANGELES     : 2690   CA     :45349   75221-9088: 1594
## DALLAS          : 1992   TX     : 2880   85038-9096:  892
## SAN FRANCISCO   : 1918   AZ     : 1298   02895-6146:  861
## SAN DIEGO       : 1405   RI     :  880   95816-5930:  581
## (Other)         :50124   (Other): 7722   (Other)   :54201
## NA's            :    1   NA's   :    1   NA's      :    1
##          Premise County  Premise Census Tract
## LOS ANGELES   :22153   Min.   :   0.60
## SAN DIEGO     : 8561   1st Qu.:  85.11
## ORANGE        : 7477   Median : 525.05
## SAN FRANCISCO : 5150   Mean   :1834.84
## SANTA CLARA   : 4675   3rd Qu.:3160.00
## (Other)       :61868   Max.   :9832.00
## NA's          :    1   NA's   :2218
```

str(data)

```
## 'data.frame':    109885 obs. of  26 variables:
##  $ License Type          : Factor w/ 70 levels "01","02","03",..: 20 20 30 30 17 18 9 21 21 30 ...
##  $ File Number           : Factor w/ 87637 levels "00000004","00000094",..: 2536 2537 2538 2539 2540
##  $ License Or Application : Factor w/ 3 levels "0:0","APP","LIC": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Type Status           : Factor w/ 8 levels "5:56.82","ACTIVE  ",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Type Orig Issue Dates : Factor w/ 8916 levels "             ",..: 5975 5014 2531 1476 5479 7165 547
##  $ Expiration Dates      : Factor w/ 37 levels "             ",..: 10 33 24 24 10 10 10 19 4 24 ...
##  $ Fee Codes             : Factor w/ 28 levels "        ","GL20K   ",..: 28 28 12 12 12 12 12 28 12
##  $ Duplicate Counts      : int  NA NA NA NA NA NA 1 NA NA NA ...
##  $ Master Indicator      : Factor w/ 2 levels "N","Y": 2 2 2 2 2 2 2 1 2 2 2 ...
##  $ Term in Months        : int  12 12 12 12 12 12 12 12 12 12 12 ...
##  $ Geo Code              : int  3309 3405 4902 1701 1401 1401 1401 1933 1008 4904 ...
##  $ District Office       : int  8 23 27 27 6 6 6 4 21 27 ...
##  $ Primary Name          : Factor w/ 63661 levels "
##  $ Premise Street Address1: Factor w/ 76193 levels "
##  $ Premise Street Address2: Factor w/ 4154 levels "
##  $ Premise City          : Factor w/ 1928 levels "                          ",..: 797 1491 731 912 15
##  $ Premise State         : Factor w/ 50 levels " ","AK","AL",..: 5 5 5 5 5 5 5 5 5 5 5 ...
##  $ Premise Zip           : Factor w/ 36759 levels "            ","01007     ",..: 12186 34983 32297 32
##  $ DBA name              : Factor w/ 69122 levels "
```

3

```
##  $ Mail Street Address1  : Factor w/ 20800 levels "
##  $ Mail Street Address2  : Factor w/ 2097 levels "
##  $ Mail City             : Factor w/ 1820 levels "                        ",..: 1 1 1464 1464 1 1 
##  $ Mail State            : Factor w/ 49 levels "  ","AK","AL",..: 1 1 6 6 1 1 1 1 1 44 ...
##  $ Mail Zip              : Factor w/ 12581 levels "            ","01035     ",..: 1 1 10107 10107 1 1
##  $ Premise County        : Factor w/ 59 levels "                   ",..: 34 35 50 18 15 15 15 20 11 50
##  $ Premise Census Tract  : num  455 38 1539 4 4 ...
```

```r
# export the data to a text file that is tab separated

write.table(data, file = "~/ABC_Data_Export-Weekly-Export/ABCdata.txt", sep="\t", row.names = FALSE)

# get the time
endTime <- Sys.time()
```

The analysis was completed on Sun Jun 07 2:00:55 PM 2015 in 2.550893 minutes.