liquor regression and k-means

2024-04-17

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

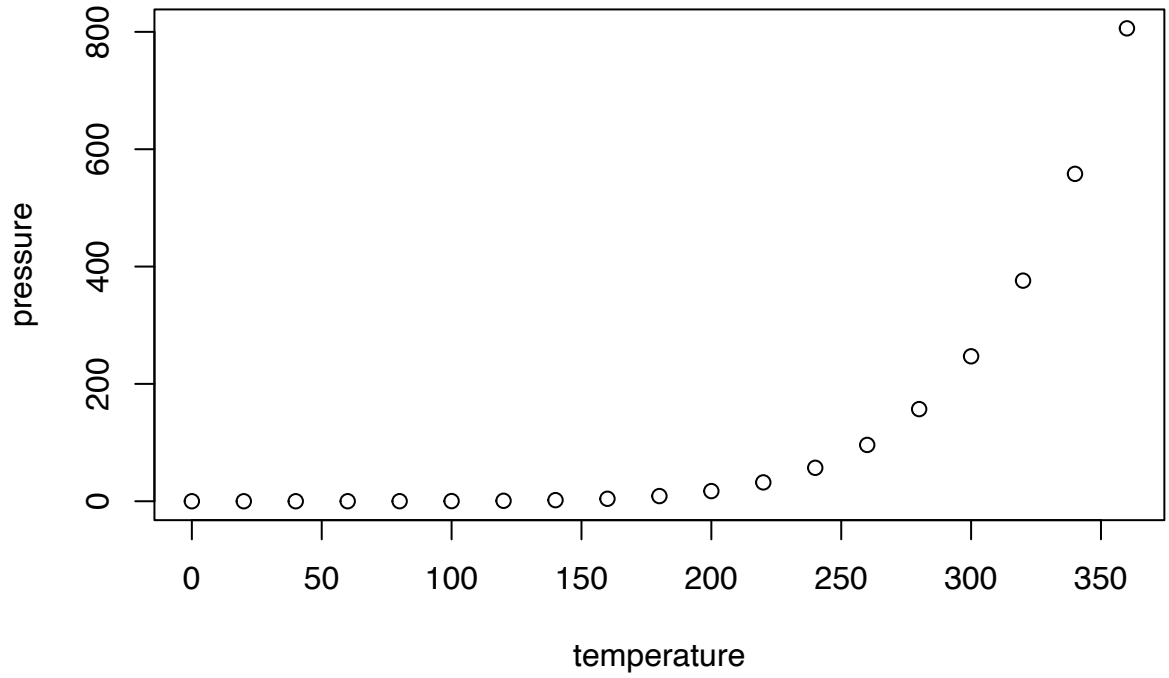
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
liquor_all <- read.csv("/Users/hebeyuan/Desktop/bc/7900 spring/Iowa_AllLiquorSales_Education.csv")
head(liquor_all)
```

```
##   X county            store.name       city       date category
## 1 1  Adair        Fareway Stores Greenfield 2016-01-19      Gin
## 2 2  Adair        Fareway Stores Greenfield 2016-06-28      Gin
## 3 3  Adair        Fareway Stores Greenfield 2016-08-19      Gin
## 4 4  Adair        Fareway Stores Greenfield 2017-01-13      Gin
## 5 5  Adair        Fareway Stores Greenfield 2017-02-17      Gin
## 6 6  Adair Casey's General Store Greenfield 2016-03-15      Gin
##         subcategory                   item volume retail.price sale.bottles
## 1 American Dry Gins      New Amsterdam Gin    750        10.35            2
## 2 American Dry Gins Seagrams Extra Dry Gin    750         9.74            4
## 3 American Dry Gins      New Amsterdam Gin    750        10.35            1
## 4 Imported Dry Gins          Tanqueray Gin    750        19.50            2
## 5 American Dry Gins      New Amsterdam Gin    750        11.25            2
## 6 American Dry Gins       Five O'clock Gin    750         5.19           12
##   sale.volume sale.dollars weekday month quarter High.school.graduate.or.higher
## 1        1.50        20.70     Tue   Jan      Q1                           0.95
## 2        3.00        38.96     Tue   Jun      Q2                           0.95
## 3        0.75        10.35     Fri   Aug      Q3                           0.95
## 4        1.50        19.50     Fri   Jan      Q1                           0.95
## 5        1.50        11.25     Fri   Feb      Q1                           0.95
```

```
## 6        9.00        62.28     Tue   Mar      Q1                           0.95
##   Bachelor.s.degree.or.higher
## 1                       0.195
## 2                       0.195
## 3                       0.195
## 4                       0.195
## 5                       0.195
## 6                       0.195
```

```
summary(liquor_all)
```

```
##        X             county           store.name            city
##  Min.   :     1   Length:159128      Length:159128      Length:159128
##  1st Qu.: 39783   Class :character   Class :character   Class :character
##  Median : 79564   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 79564
##  3rd Qu.:119346
##  Max.   :159128
##
##      date              category         subcategory            item
##  Length:159128      Length:159128      Length:159128      Length:159128
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##      volume        retail.price    sale.bottles       sale.volume
##  Min.   :  50.0   Min.   : 1.46   Min.   :   0.000   Min.   :  0.030
##  1st Qu.: 750.0   1st Qu.: 6.35   1st Qu.:   1.000   1st Qu.:  1.000
##  Median : 750.0   Median :10.67   Median :   2.000   Median :  1.750
##  Mean   : 964.9   Mean   :13.41   Mean   :   4.132   Mean   :  3.503
##  3rd Qu.:1000.0   3rd Qu.:19.50   3rd Qu.:   4.000   3rd Qu.:  3.500
##  Max.   :1750.0   Max.   :60.15   Max.   :1440.000   Max.   :945.000
##
##   sale.dollars        weekday             month              quarter
##  Min.   :    0.00   Length:159128      Length:159128      Length:159128
##  1st Qu.:   15.75   Class :character   Class :character   Class :character
##  Median :   59.94   Mode  :character   Mode  :character   Mode  :character
##  Mean   :   78.59
##  3rd Qu.:  107.94
##  Max.   :20244.60
##
##  High.school.graduate.or.higher Bachelor.s.degree.or.higher
##  Min.   :0.7790                 Min.   :0.1430
##  1st Qu.:0.9180                 1st Qu.:0.2210
##  Median :0.9260                 Median :0.3220
##  Mean   :0.9263                 Mean   :0.3178
##  3rd Qu.:0.9460                 3rd Qu.:0.3760
##  Max.   :0.9680                 Max.   :0.5430
##  NA's   :804                    NA's   :804
```

```r
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
```

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.5
## v ggplot2   3.5.0     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become err
```

```r
library(ggplot2)
library(grid)
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
library(ggpubr)
```

```
##
## Attaching package: 'ggpubr'
##
## The following object is masked from 'package:forecast':
##
##     gghistogram
```

```r
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
```

```
## The following object is masked from 'package:purrr':
##
##     some
##
## The following object is masked from 'package:dplyr':
##
##     recode
```

To assess the relationship between the liquor sales and the education level, prior to organize current variables.

```
liquor1 <- lm(sale.dollars ~ subcategory + item , data = liquor_all)
summary(liquor1)
```

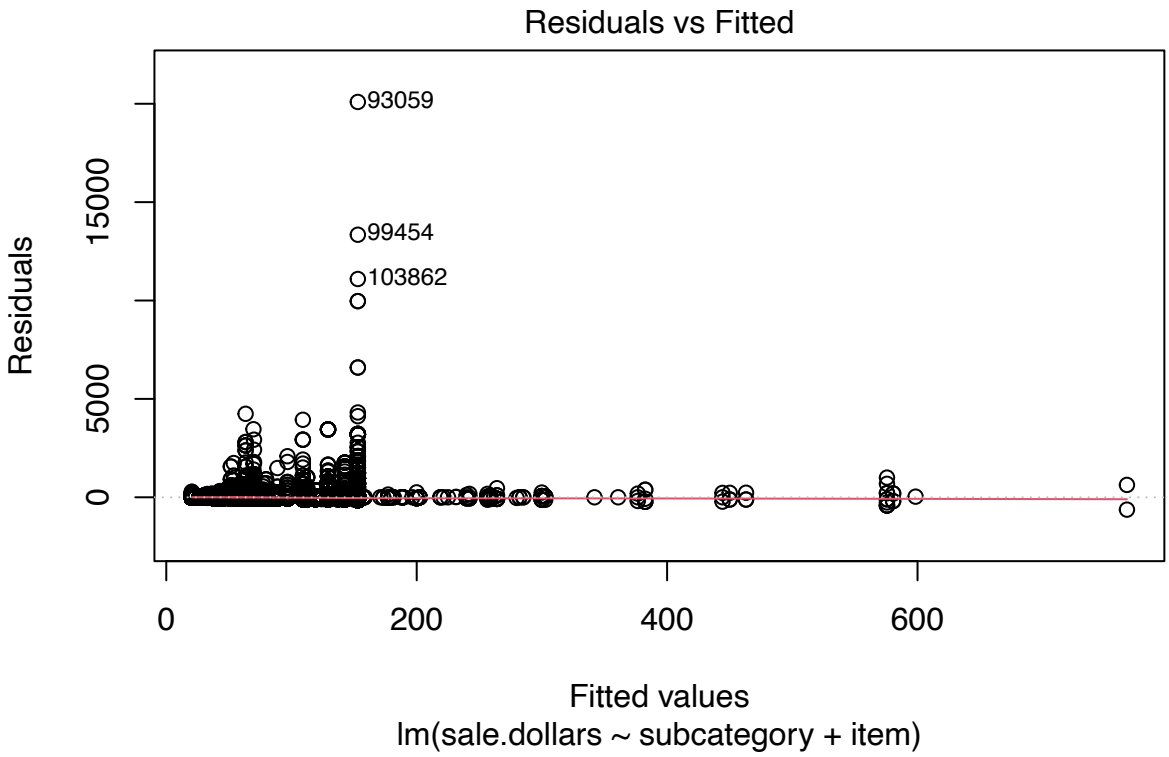
```
##
## Call:
## lm(formula = sale.dollars ~ subcategory + item, data = liquor_all)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -627.7   -53.2    -5.2    39.0 20091.6
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           23.011     14.567   1.580 0.114186
## subcategoryAmerican Sloe Gins        296.871    149.698   1.983 0.047354 *
## subcategoryFlavored Gins            -128.303    142.166  -0.902 0.366800
## subcategoryImported Dry Gins          66.517    100.130   0.664 0.506490
## itemArrow Sloe Gin                  -293.101    150.789  -1.944 0.051923 .
## itemAviation American Gin             62.205     15.863   3.921 8.80e-05 ***
## itemBarton Gin                        46.618     14.756   3.159 0.001582 **
## itemBeefeater """"24""""              22.912    167.517   0.137 0.891210
## itemBeefeater Gin                     19.568    101.197   0.193 0.846672
## itemBellringer Gin                   119.549     95.520   1.252 0.210732
## itemBlaum Bros. Gin                   54.527     20.480   2.662 0.007759 **
## itemBluecoat American Dry Gin        744.240     95.520   7.791 6.67e-15 ***
## itemBols Genever Gin                 152.532    117.486   1.298 0.194185
## itemBombay Dry Gin                   -36.572    101.235  -0.361 0.717904
## itemBombay Sapphire East             -33.166    101.290  -0.327 0.743337
## itemBombay Sapphire Gin               39.671    101.192   0.392 0.695033
## itemBombay Sapphire Gin Minis        150.472    167.517   0.898 0.369055
## itemBoodles British Gin London Dry   -23.034    101.260  -0.227 0.820059
## itemBoodles Mulberry British Gin     363.292    171.275   2.121 0.033915 *
## itemBoomsma Genever Jonge              8.032    138.384   0.058 0.953716
## itemBroker's London Dry Gin          -15.270    101.314  -0.151 0.880195
## itemBrokers Gin Mini                  30.172    138.384   0.218 0.827405
## itemBrooklyn Gin                     156.989     78.444   2.001 0.045363 *
## itemBulldog Gin                      -22.312    101.461  -0.220 0.825942
## itemBurnett's Gin London Dry          37.279     14.765   2.525 0.011575 *
## itemCaliber Gin                       37.218     16.405   2.269 0.023291 *
## itemCalvert Gin                       32.337     15.825   2.043 0.041019 *
## itemCitadelle Gin                     40.972    167.517   0.245 0.806779
## itemClearheart Gin                    47.758     16.207   2.947 0.003213 **
## itemCrater Lake Gin                  -98.482    201.111  -0.490 0.624354
```

```
## itemDeath''s Door Gin                 110.490    95.520   1.157 0.247392
## itemDistiller's Gin No. 11            173.549   134.298   1.292 0.196267
## itemDripping Springs Artisan Gin      179.489   134.298   1.336 0.181389
## itemDrumshanbo Gunpowder Irish Gin     98.032   138.384   0.708 0.478694
## itemEdinburgh Gin                     -54.090   106.546  -0.508 0.611685
## itemFifty Pounds                       92.032   138.384   0.665 0.506020
## itemFive O'clock                       16.113    15.392   1.047 0.295186
## itemFive O'clock Gin                   19.685    14.635   1.345 0.178612
## itemFleischmann's Dry Gin              13.380    15.149   0.883 0.377124
## itemFleischmann's Gin                  29.235    14.738   1.984 0.047299 *
## itemFleischmann's Gin Mini             66.990    95.520   0.701 0.483110
## itemFrey Ranch Gin                    147.989   134.298   1.102 0.270487
## itemGenevieve Gin                      57.989    95.520   0.607 0.543792
## itemGilbey's Gin London Dry            41.312    14.678   2.814 0.004887 **
## itemGordon's Gin London Dry            36.103    14.762   2.446 0.014460 *
## itemGordon's Gin London Dry - Pet      63.070    14.765   4.271 1.94e-05 ***
## itemHawkeye Gin                        28.590    14.709   1.944 0.051923 .
## itemHaymans Old Tom Gin               141.917   131.303   1.081 0.279770
## itemHaymans Sloe Gin                  190.262   142.910   1.331 0.183078
## itemHendrick Gin                      271.372   167.517   1.620 0.105242
## itemHendrick's Gin                     53.038   101.205   0.524 0.600234
## itemHendricks Gin Mini                390.412   171.275   2.279 0.022642 *
## itemIndian Summer                     152.173   144.681   1.052 0.292901
## itemIvy City Gin                      120.990   134.298   0.901 0.367643
## itemJunipero Gin                      241.077    42.808   5.632 1.79e-08 ***
## itemKirkland Signature London Dry Gin 105.870    95.520   1.108 0.267714
## itemKoval Dry Gin                     165.989    56.417   2.942 0.003259 **
## itemLangley's No. 8 Gin               287.032   127.198   2.257 0.024036 *
## itemLetherbee Gin                     259.110    95.520   2.713 0.006676 **
## itemMagellan Gin                      292.972   114.930   2.549 0.010800 *
## itemMahon Gin                         112.912   167.517   0.674 0.500292
## itemMalfy Gin                         -45.500   104.395  -0.436 0.662948
## itemMartin Millers Gin                166.972   117.486   1.421 0.155258
## itemMaster's Gin                      -13.858   106.546  -0.130 0.896515
## itemMcCormick Gin                      33.636    15.543   2.164 0.030465 *
## itemMcCormick Gin Pet                  31.802    15.254   2.085 0.037080 *
## itemMerrylegs Genever Style Gin       318.990   134.298   2.375 0.017539 *
## itemMonkey 47                         354.712   127.198   2.789 0.005293 **
## itemMr Boston Sloe Gin               -179.662   159.976  -1.123 0.261415
## itemMr. Boston Riva Gin                47.550    95.520   0.498 0.618629
## itemNew Amsterdam Gin                  47.044    14.626   3.217 0.001298 **
## itemNo. 3 London Dry Gin              129.472   127.198   1.018 0.308739
## itemNolet's Silver Gin                110.419   114.930   0.961 0.336678
## itemOpihr Gin                          -4.043   101.439  -0.040 0.968207
## itemParamount Gin                      30.851    14.610   2.112 0.034717 *
## itemParamount Gin Traveler             26.744    17.771   1.505 0.132348
## itemParamount Sloe Gin               -287.041   150.441  -1.908 0.056393 .
## itemPhillips Gin                       54.218    14.970   3.622 0.000293 ***
## itemPillar 136 Gin                      5.750    38.540   0.149 0.881409
## itemPinnacle Gin                       22.944   101.436   0.226 0.821057
## itemPlymouth Gin                       29.749   101.341   0.294 0.769100
## itemPlymouth Navy Strength Gin        150.392   127.198   1.182 0.237072
## itemPrairie Organic Gin                56.682    15.693   3.612 0.000304 ***
## itemRansom Old Tom Gin                557.490    68.324   8.160 3.39e-16 ***
```

```
## itemRiver Rose Gin                      73.790     15.540   4.748 2.05e-06 ***
## itemRogue Pink Spruce Gin              440.029     78.444   5.609 2.03e-08 ***
## itemRogue Spruce Gin                   279.599     68.324   4.092 4.27e-05 ***
## itemSeagram's Distillers Reserve Gin   135.389     95.520   1.417 0.156371
## itemSeagram's Lime Twisted Gin         307.612    162.371   1.894 0.058161 .
## itemSeagrams Extra Dry Gin              40.361     14.590   2.766 0.005669 **
## itemSeagrams Extra Dry Gin Mini         -2.590     14.640  -0.177 0.859593
## itemSeagrams Extra Dry Gin Pet          40.598     15.034   2.700 0.006926 **
## itemSeagrams Lime Twisted Gin          151.014    142.929   1.057 0.290711
## itemSeagrams Peach Twisted Gin         135.872    143.026   0.950 0.342122
## itemSipsmith London Dry Gin            -23.856    102.896  -0.232 0.816660
## itemSt. George Botanivore Gin          262.792    148.015   1.775 0.075828 .
## itemSt. George Dry Rye Gin             134.489     78.444   1.714 0.086446 .
## itemSt. George Terroir Gin             154.177     49.398   3.121 0.001802 **
## itemTanqueray Bloomsbury                34.772    109.638   0.317 0.751128
## itemTanqueray Bloomsbury Gin            51.722    121.219   0.427 0.669612
## itemTanqueray Gin                       63.476    101.188   0.627 0.530460
## itemTanqueray Gin Mini                 172.192    167.517   1.028 0.303995
## itemTanqueray No. 10                   360.472    127.198   2.834 0.004598 **
## itemTanqueray No. Ten                   -0.766    101.252  -0.008 0.993964
## itemTanqueray Rangpur                  135.472    167.517   0.809 0.418686
## itemTanqueray Rangpur Gin              -35.401    101.232  -0.350 0.726559
## itemTattersall Gin                      41.358     21.960   1.883 0.059661 .
## itemThe Botanist Islay Dry Gin         210.372    108.275   1.943 0.052025 .
## itemTru Gin                            235.739     68.324   3.450 0.000560 ***
## itemUncle Val's Botanical Gin          255.714    145.525   1.757 0.078889 .
## itemUncle Val's Restorative Gin        278.567    163.122   1.708 0.087689 .
## itemUncle Vals Peppered Gin            249.292    171.275   1.456 0.145531
## itemYahara Bay Barrell Mellowed Gin     63.613     46.825   1.359 0.174304
## itemYahara Bay Gin                      65.654     68.324   0.961 0.336589
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 133.5 on 159014 degrees of freedom
## Multiple R-squared:  0.08142,    Adjusted R-squared:  0.08076
## F-statistic: 124.7 on 113 and 159014 DF,  p-value: < 2.2e-16
```

From the first step of doing the regression of the liquor sales on the subcategory of gins, item name, high school graduation and bachelor graduation percent, it's not hard to find several variables are significant than other. For example, in the 'subcategory', the 'American Sloe Gins' shows siginifance. The high school graduation degress and bachelor degree also show relative strong significance on the liquor sales, which will be discussed further.

```
plot(liquor1,1)
```




```
# the multicollinearity exists or not; the multicollinearity is not a problem here
vif(liquor1)
```

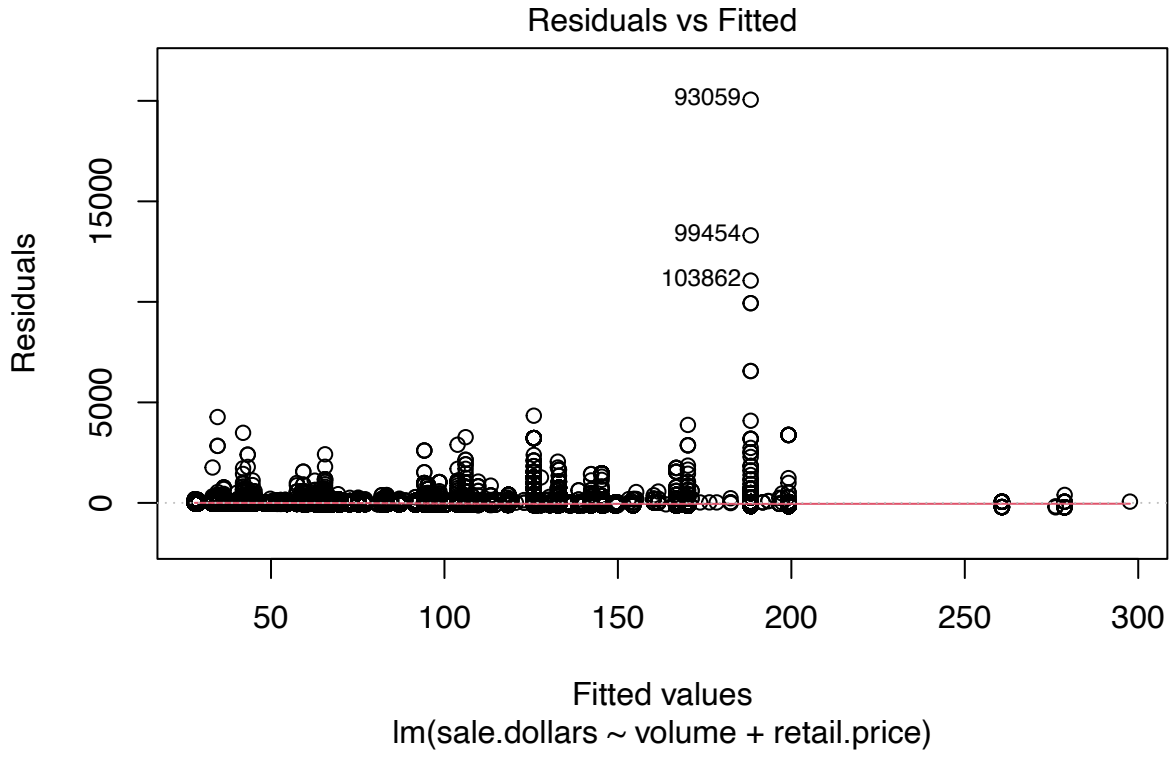
```
##                   GVIF  Df GVIF^(1/(2*Df))
## subcategory 9094903873   3       45.687741
## item        9094903873 110        1.109858
```

```
# regress liquor sales on other numerical variables
liquor2 <- lm(sale.dollars ~ volume + retail.price, data = liquor_all)
summary(liquor2)
```

```
##
## Call:
## lm(formula = sale.dollars ~ volume + retail.price, data = liquor_all)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -222.5   -52.9     2.9    32.3 20056.3
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  21.826384   0.794564   27.47   <2e-16 ***
## volume       -0.008249   0.000758  -10.88   <2e-16 ***
## retail.price  4.824897   0.043990  109.68   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 133.7 on 159125 degrees of freedom
## Multiple R-squared:  0.07849,    Adjusted R-squared:  0.07848
## F-statistic:  6777 on 2 and 159125 DF,  p-value: < 2.2e-16
```

```
plot(liquor2,1)
```




```
#filter a specific county to see if the regression is
liquor_benton <- filter(liquor_all, county =="Benton")
liquor_johnson <- filter(liquor_all,county =="Johnson")
```

Based on the regression of liquor sales on different variables, filter specific counties to see whether the significant difference changes.

```
Johnson <- glm(sale.dollars ~ subcategory + item + retail.price + volume, data = liquor_johnson)
summary(Johnson)
```

```
##
## Call:
## glm(formula = sale.dollars ~ subcategory + item + retail.price +
##     volume, data = liquor_johnson)
##
## Coefficients: (2 not defined because of singularities)
```

```
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             -6.218e+01  2.243e+01  -2.772 0.005585
## subcategoryAmerican Sloe Gins            5.471e+02  1.790e+02   3.056 0.002251
## subcategoryFlavored Gins                 4.410e+00  1.661e+02   0.027 0.978819
## subcategoryImported Dry Gins            -1.967e+01  1.663e+02  -0.118 0.905850
## itemBarton Gin                           1.340e+02  2.671e+01   5.016 5.34e-07
## itemBeefeater Gin                        8.843e+01  1.654e+02   0.535 0.592916
## itemBellringer Gin                       1.244e+02  1.185e+02   1.050 0.293810
## itemBlaum Bros. Gin                     -3.689e+01  5.580e+01  -0.661 0.508598
## itemBluecoat American Dry Gin            7.425e+00  1.661e+02   0.045 0.964348
## itemBombay Dry Gin                       6.819e+01  1.660e+02   0.411 0.681229
## itemBombay Sapphire East                -1.009e+01  1.660e+02  -0.061 0.951543
## itemBombay Sapphire Gin                  6.813e+01  1.652e+02   0.412 0.680049
## itemBoodles British Gin London Dry       7.396e+01  1.657e+02   0.446 0.655355
## itemBroker's London Dry Gin              2.416e+01  1.663e+02   0.145 0.884514
## itemBrooklyn Gin                        -1.971e+01  9.747e+01  -0.202 0.839778
## itemBulldog Gin                          2.679e+01  1.666e+02   0.161 0.872264
## itemBurnett's Gin London Dry             8.469e+01  2.537e+01   3.338 0.000847
## itemCaliber Gin                          1.106e+02  5.574e+01   1.984 0.047244
## itemCalvert Gin                          1.092e+02  7.836e+01   1.394 0.163412
## itemCitadelle Gin                        3.312e+01  2.334e+02   0.142 0.887168
## itemClearheart Gin                       6.032e+00  2.555e+01   0.236 0.813399
## itemDistiller's Gin No. 11              -3.080e+01  1.664e+02  -0.185 0.853147
## itemDripping Springs Artisan Gin        -3.478e+01  1.664e+02  -0.209 0.834482
## itemEdinburgh Gin                       -1.230e+02  1.905e+02  -0.646 0.518604
## itemFive O'clock                         9.389e+01  2.752e+01   3.412 0.000646
## itemFive O'clock Gin                     8.458e+01  2.436e+01   3.471 0.000520
## itemFleischmann's Gin                    8.607e+01  2.838e+01   3.033 0.002424
## itemFrey Ranch Gin                      -1.368e+01  1.662e+02  -0.082 0.934423
## itemGenevieve Gin                       -3.862e+02  1.683e+02  -2.294 0.021815
## itemGilbey's Gin London Dry              8.595e+01  2.519e+01   3.412 0.000647
## itemGordon's Gin London Dry              9.769e+01  2.413e+01   4.049 5.17e-05
## itemGordon's Gin London Dry - Pet        8.399e+01  2.562e+01   3.278 0.001048
## itemHawkeye Gin                          1.097e+02  2.796e+01   3.925 8.71e-05
## itemHaymans Old Tom Gin                  1.614e+02  2.334e+02   0.691 0.489334
## itemHaymans Sloe Gin                     1.614e+02  2.334e+02   0.691 0.489334
## itemHendrick's Gin                       3.426e+01  1.652e+02   0.207 0.835657
## itemIndian Summer                       -9.856e+01  1.723e+02  -0.572 0.567350
## itemJunipero Gin                        -2.075e+01  1.663e+02  -0.125 0.900686
## itemKirkland Signature London Dry Gin    6.613e+01  1.664e+02   0.397 0.691015
## itemKoval Dry Gin                       -2.573e+01  1.663e+02  -0.155 0.877041
## itemMahon Gin                           -2.175e+00  2.333e+02  -0.009 0.992564
## itemMalfy Gin                           -8.916e+01  1.906e+02  -0.468 0.639920
## itemMartin Millers Gin                   1.676e+02  2.021e+02   0.829 0.407063
## itemMaster's Gin                        -9.892e+01  2.021e+02  -0.489 0.624509
## itemMcCormick Gin                        1.498e+02  3.420e+01   4.380 1.20e-05
## itemMcCormick Gin Pet                    1.066e+02  4.273e+01   2.496 0.012589
## itemMr. Boston Riva Gin                  2.156e+02  2.344e+02   0.920 0.357563
## itemNew Amsterdam Gin                    6.457e+01  2.205e+01   2.928 0.003419
## itemOpihr Gin                           -1.375e+01  1.664e+02  -0.083 0.934183
## itemParamount Gin                        9.894e+01  2.613e+01   3.786 0.000154
## itemParamount Gin Traveler               1.051e+02  4.997e+01   2.104 0.035382
## itemParamount Sloe Gin                  -4.514e+02  1.804e+02  -2.503 0.012337
## itemPhillips Gin                         1.440e+02  2.747e+01   5.240 1.63e-07
```

```
## itemPinnacle Gin                        9.621e+01  1.718e+02   0.560 0.575434
## itemPlymouth Gin                        6.188e+00  1.657e+02   0.037 0.970201
## itemPlymouth Navy Strength Gin         -2.639e-10  2.333e+02   0.000 1.000000
## itemPrairie Organic Gin                 2.570e+01  2.597e+01   0.989 0.322481
## itemRansom Old Tom Gin                  1.648e+02  1.664e+02   0.990 0.322039
## itemRiver Rose Gin                     -7.144e+01  3.370e+01  -2.120 0.034033
## itemRogue Pink Spruce Gin               1.582e+02  1.184e+02   1.336 0.181621
## itemRogue Spruce Gin                    1.580e+02  1.662e+02   0.950 0.342004
## itemSeagrams Extra Dry Gin              8.173e+01  2.225e+01   3.673 0.000240
## itemSeagrams Extra Dry Gin Mini         7.439e+01  2.313e+01   3.216 0.001304
## itemSeagrams Extra Dry Gin Pet          6.678e+01  2.580e+01   2.588 0.009656
## itemSeagrams Lime Twisted Gin           3.774e+01  1.658e+02   0.228 0.819957
## itemSeagrams Peach Twisted Gin          3.835e+01  1.697e+02   0.226 0.821200
## itemSipsmith London Dry Gin            -8.298e+01  1.782e+02  -0.466 0.641476
## itemSt. George Botanivore Gin          -9.044e+00  2.333e+02  -0.039 0.969082
## itemSt. George Dry Rye Gin             -4.633e+00  1.662e+02  -0.028 0.977755
## itemSt. George Terroir Gin             -4.633e+00  1.662e+02  -0.028 0.977755
## itemTanqueray Bloomsbury Gin           -3.661e+01  2.333e+02  -0.157 0.875339
## itemTanqueray Gin                       1.154e+02  1.652e+02   0.699 0.484811
## itemTanqueray No. Ten                  -2.515e+01  1.659e+02  -0.152 0.879564
## itemTanqueray Rangpur Gin              -1.171e+00  1.657e+02  -0.007 0.994360
## itemTattersall Gin                     -3.081e+01  8.478e+01  -0.363 0.716333
## itemThe Botanist Islay Dry Gin                 NA         NA      NA       NA
## itemUncle Val's Botanical Gin          -6.293e+00  1.696e+02  -0.037 0.970405
## itemUncle Val's Restorative Gin        -3.243e+01  1.918e+02  -0.169 0.865747
## itemUncle Vals Peppered Gin                    NA         NA      NA       NA
## retail.price                            1.002e+01  8.347e-01  12.003  < 2e-16
## volume                                 -5.159e-02  9.718e-03  -5.309 1.12e-07
##
## (Intercept)                            **
## subcategoryAmerican Sloe Gins          **
## subcategoryFlavored Gins
## subcategoryImported Dry Gins
## itemBarton Gin                         ***
## itemBeefeater Gin
## itemBellringer Gin
## itemBlaum Bros. Gin
## itemBluecoat American Dry Gin
## itemBombay Dry Gin
## itemBombay Sapphire East
## itemBombay Sapphire Gin
## itemBoodles British Gin London Dry
## itemBroker's London Dry Gin
## itemBrooklyn Gin
## itemBulldog Gin
## itemBurnett's Gin London Dry           ***
## itemCaliber Gin                        *
## itemCalvert Gin
## itemCitadelle Gin
## itemClearheart Gin
## itemDistiller's Gin No. 11
## itemDripping Springs Artisan Gin
## itemEdinburgh Gin
## itemFive O'clock                       ***
```

```
## itemFive O'clock Gin                  ***
## itemFleischmann's Gin                 **
## itemFrey Ranch Gin
## itemGenevieve Gin                     *
## itemGilbey's Gin London Dry           ***
## itemGordon's Gin London Dry           ***
## itemGordon's Gin London Dry - Pet     **
## itemHawkeye Gin                       ***
## itemHaymans Old Tom Gin
## itemHaymans Sloe Gin
## itemHendrick's Gin
## itemIndian Summer
## itemJunipero Gin
## itemKirkland Signature London Dry Gin
## itemKoval Dry Gin
## itemMahon Gin
## itemMalfy Gin
## itemMartin Millers Gin
## itemMaster's Gin
## itemMcCormick Gin                     ***
## itemMcCormick Gin Pet                 *
## itemMr. Boston Riva Gin
## itemNew Amsterdam Gin                 **
## itemOpihr Gin
## itemParamount Gin                     ***
## itemParamount Gin Traveler            *
## itemParamount Sloe Gin                *
## itemPhillips Gin                      ***
## itemPinnacle Gin
## itemPlymouth Gin
## itemPlymouth Navy Strength Gin
## itemPrairie Organic Gin
## itemRansom Old Tom Gin
## itemRiver Rose Gin                    *
## itemRogue Pink Spruce Gin
## itemRogue Spruce Gin
## itemSeagrams Extra Dry Gin            ***
## itemSeagrams Extra Dry Gin Mini       **
## itemSeagrams Extra Dry Gin Pet        **
## itemSeagrams Lime Twisted Gin
## itemSeagrams Peach Twisted Gin
## itemSipsmith London Dry Gin
## itemSt. George Botanivore Gin
## itemSt. George Dry Rye Gin
## itemSt. George Terroir Gin
## itemTanqueray Bloomsbury Gin
## itemTanqueray Gin
## itemTanqueray No. Ten
## itemTanqueray Rangpur Gin
## itemTattersall Gin
## itemThe Botanist Islay Dry Gin
## itemUncle Val's Botanical Gin
## itemUncle Val's Restorative Gin
## itemUncle Vals Peppered Gin
```

```
## retail.price                            ***
## volume                                  ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 27217.74)
##
##     Null deviance: 373837849  on 11789  degrees of freedom
## Residual deviance: 318746945  on 11711  degrees of freedom
## AIC: 153934
##
## Number of Fisher Scoring iterations: 2
```

```
benton <- glm(sale.dollars ~ subcategory + item + retail.price + volume, data = liquor_benton)
summary(benton)
```

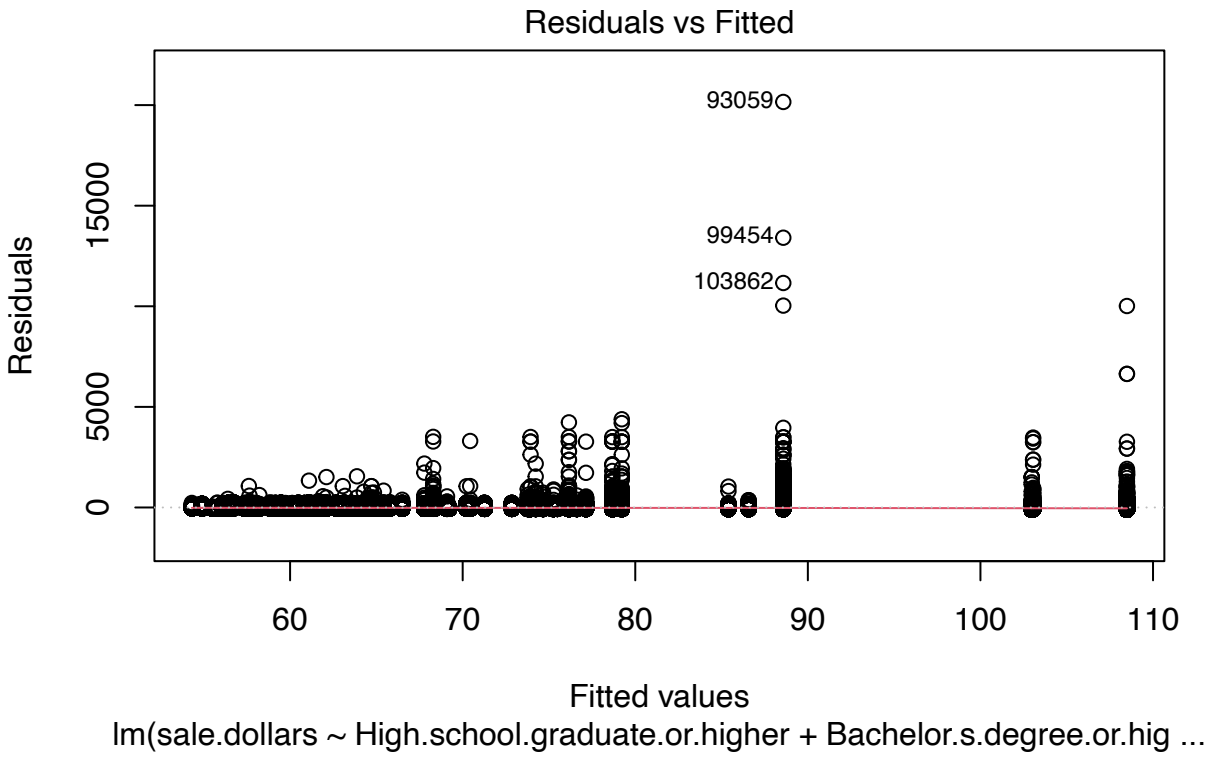
```
##
## Call:
## glm(formula = sale.dollars ~ subcategory + item + retail.price +
##     volume, data = liquor_benton)
##
## Coefficients: (3 not defined because of singularities)
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         7.61649   15.79419   0.482   0.6299
## subcategoryAmerican Sloe Gins     -44.91670   42.33864  -1.061   0.2893
## subcategoryFlavored Gins           35.86760   45.29190   0.792   0.4288
## subcategoryImported Dry Gins       29.02519   22.41894   1.295   0.1961
## itemBombay Dry Gin                -17.73376   31.20201  -0.568   0.5701
## itemBombay Sapphire Gin             0.45183   17.04025   0.027   0.9789
## itemBroker's London Dry Gin       -20.51679   41.76250  -0.491   0.6235
## itemBurnett's Gin London Dry       -1.83848   18.43216  -0.100   0.9206
## itemFive O'clock                  -39.25908   42.43140  -0.925   0.3553
## itemFive O'clock Gin              -36.35026   16.80290  -2.163   0.0310 *
## itemFleischmann's Gin             -28.72210   22.70971  -1.265   0.2066
## itemGilbey's Gin London Dry        19.84475   17.44365   1.138   0.2559
## itemGordon's Gin London Dry       -29.92381   42.16769  -0.710   0.4783
## itemGordon's Gin London Dry - Pet -59.58314   25.22846  -2.362   0.0186 *
## itemHawkeye Gin                   -41.90358   27.78917  -1.508   0.1323
## itemHendrick's Gin                 19.37026   33.04804   0.586   0.5581
## itemIndian Summer                        NA         NA      NA       NA
## itemNew Amsterdam Gin              9.93951   16.17783   0.614   0.5393
## itemParamount Gin                 -36.70555   17.59563  -2.086   0.0375 *
## itemParamount Sloe Gin                   NA         NA      NA       NA
## itemPhillips Gin                  -66.55738   32.45942  -2.050   0.0409 *
## itemPrairie Organic Gin            6.44736   42.84829   0.150   0.8805
## itemSeagrams Extra Dry Gin         7.17517   15.47023   0.464   0.6430
## itemSeagrams Extra Dry Gin Mini  -22.69010   24.90133  -0.911   0.3627
## itemSeagrams Extra Dry Gin Pet           NA         NA      NA       NA
## itemTanqueray Gin                 15.76817   15.10599   1.044   0.2971
## itemTanqueray No. Ten             16.78386   32.79944   0.512   0.6091
## itemTanqueray Rangpur Gin         -2.38436   31.18214  -0.076   0.9391
## retail.price                      -3.04124    1.19263  -2.550   0.0111 *
## volume                             0.07016    0.01133   6.190 1.36e-09 ***
## ---
```

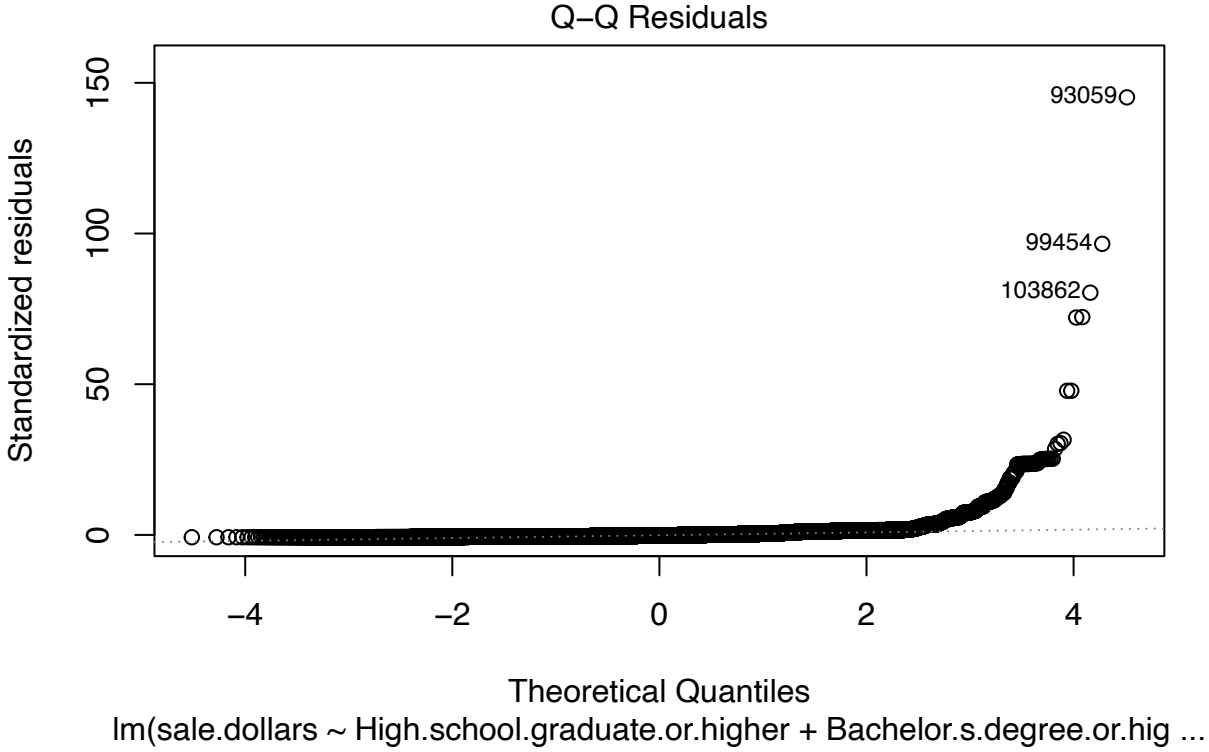
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1549.605)
##
##     Null deviance: 985879  on 472  degrees of freedom
## Residual deviance: 691124  on 446  degrees of freedom
## AIC: 4845.1
##
## Number of Fisher Scoring iterations: 2
```

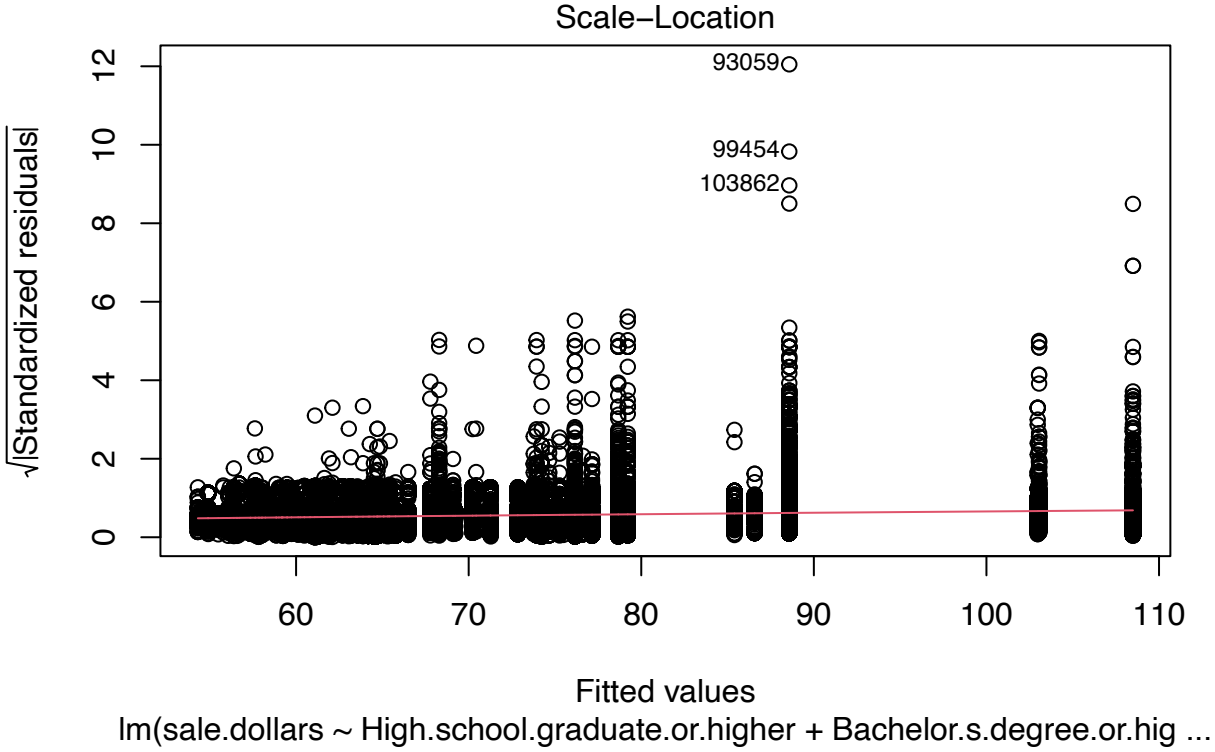
```
#only include the educational factors:
liquor_education <- lm(data = liquor_all, sale.dollars ~ High.school.graduate.or.higher + Bachelor.s.
summary(liquor_education)
```

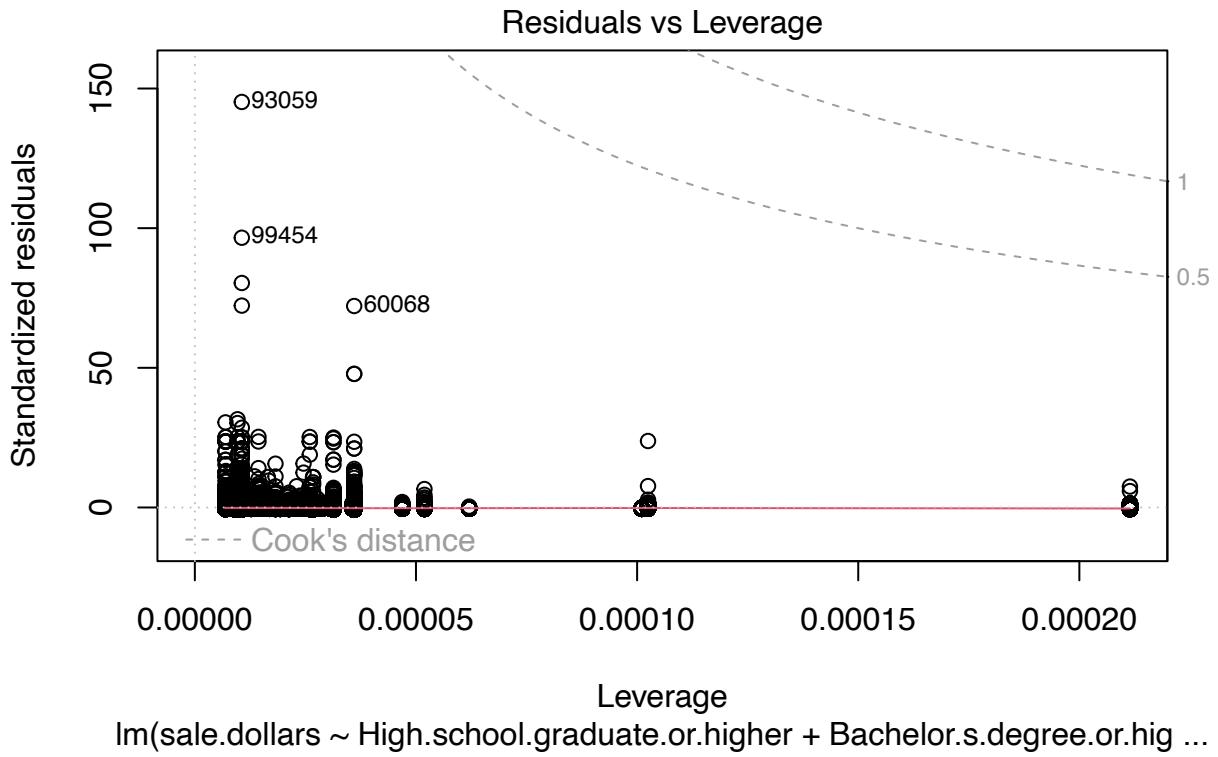
```
##
## Call:
## lm(formula = sale.dollars ~ High.school.graduate.or.higher +
##     Bachelor.s.degree.or.higher, data = liquor_all)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -107.0   -59.0   -24.6    28.0 20156.0
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      177.57      13.18   13.47   <2e-16 ***
## High.school.graduate.or.higher  -159.32      14.96  -10.65   <2e-16 ***
## Bachelor.s.degree.or.higher      153.58       4.05   37.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 138.8 on 158321 degrees of freedom
##   (804 observations deleted due to missingness)
## Multiple R-squared:  0.01002,    Adjusted R-squared:  0.01001
## F-statistic: 801.3 on 2 and 158321 DF,  p-value: < 2.2e-16
```

```
plot(liquor_education)
```


Residuals vs Fitted
93059
99454
103862
Residuals
0
5000
15000
60
70
80
90
100
110
Fitted values
lm(sale.dollars ~ High.school.graduate.or.higher + Bachelor.s.degree.or.hig ...


Q–Q Residuals
Standardized residuals
0
50
100
150
93059
99454
103862
–4
–2
0
2
4
Theoretical Quantiles
lm(sale.dollars ~ High.school.graduate.or.higher + Bachelor.s.degree.or.hig ...


Scale−Location
93059
99454
103862
√|Standardized residuals|
0
2
4
6
8
10
12
60
70
80
90
100
110
Fitted values
lm(sale.dollars ~ High.school.graduate.or.higher + Bachelor.s.degree.or.hig ...



By selecting two seperate counties, we could feature some changes between the subcategory and specific item influences on the sales of liquor. What if doing the t-test to find higher/lower education level?

```
library(stats)
as.numeric(liquor_all$Bachelor.s.degree.or.higher)
```

```
##     [1] 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195
##    [13] 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195
##    [25] 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195
##    [37] 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195
##    [49] 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195
##    [61] 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195
##    [73] 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195
##    [85] 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195
##    [97] 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195
##   [109] 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195 0.195
##   [121] 0.195 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173
##   [133] 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173
##   [145] 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173
##   [157] 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173
##   [169] 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173
##   [181] 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173
##   [193] 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173
##   [205] 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173
##   [217] 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173
##   [229] 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173 0.173
```