STAT 5371 final project—-WiFi matters?

Li Sun

November 19, 2015

Introduction:

I tried to find out, should restaurants' owners provide wifi (free or paid) to customers to make them feel better. Part I we finished converting 3 json files into a single data frame containing most relevant information.

Methods and Data:

Data source are from 'http://www.yelp.com/dataset_challenge'. Data comes in json format, and there are 5 files: user, business, checkin, tip, and reviews. To answer my specific question: 'Does wifi access really helps improve customers' experience in restuarants?' I choose only three data sets from these five, user, business, and reviews. And I used regression method to study effects of different variables on response, stars. I want to see if free wifi can have positive effects to stars of that business or is just irrelevant, in a good model considering all relevant variables

1. Missing values and data types

First thing is get rid of missing values. I didnt impute missing values in this case because most of the missingness occur in factor like variables, which could not to be imputed easily. Also some of the columns are in data types which are not good for later analysis like "array", "character". We will convert them to "numeric" and "factor". Identity columns are separated to a new data frame "busi id df". Finally, we also remove those variables with very low variance, which means over 99% of all observations of those columns are unique value. Because the coeficients of those variables are going to be very large.

```
## [1] 10217
```

```
## [1] "logical"   "character" "integer"   "numeric"
```

EDA

What people say about wifi
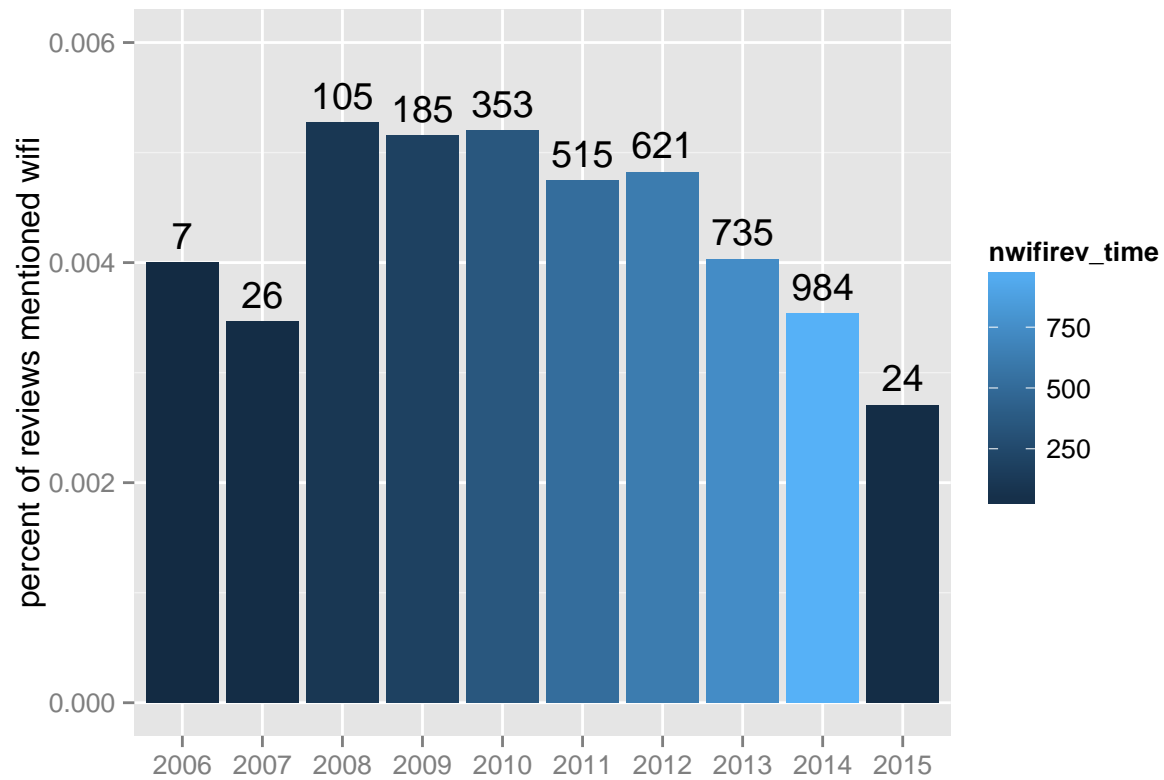

```
#png(file="2.png",width=2500,height=1700,res=350)
mywc(stem_wifi_text)
```



```
#dev.off()
#png(file="3.png",width=2500,height=1700,res=350)
mywc(gram2)
```

free wifi

```
#dev.off()
```

City and category reduction

First we found the cities var contains too many values, and not distributed uniformly. So in order to increase the efficincy of regression, lower the number of dummy variables in the final model, we will group some of the citis

We still have too many variables specifying types of restuarants. Some of the types are rare, with less occurance, that would be of our less interest. So I want to cut the type of restuarants down to several majors types. The standard of maintainning certain types depends on occurance and also we want to cover almost all the business in list.

Based on this figuer, we will choose first 13 types which cover 0.7888813 of all observations. And store full categories in categ data frame
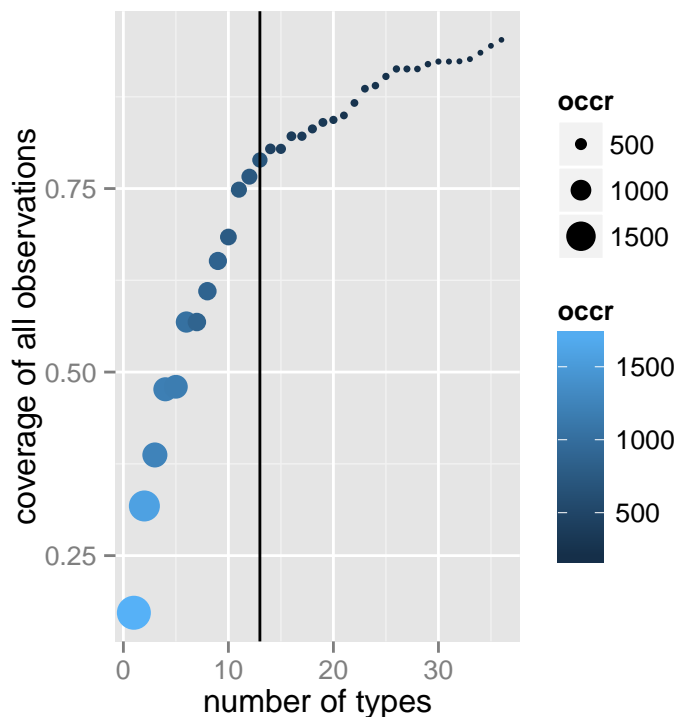
Check the correlations

Highly correlated predictor can cause problems, so lets check the correlations

```
## pdf
##   2
```

This tell us: 1. stars which is from business table are highly correlated to average stars we want to analyze, if we include this in the model, we will get model with high R square but can provide no information about what people really like. So I will exclude this business star from our model 2. two variables are almost the same, "review_count", and "nrev", they are talking about the same thing, so I will remove "review count" . Similarly, "nightlife" and "bars" overlap significantly, so remove "nightlife"

Finally, lets look at some numeric variables distributions

Transform skewed the data to lower the skewness
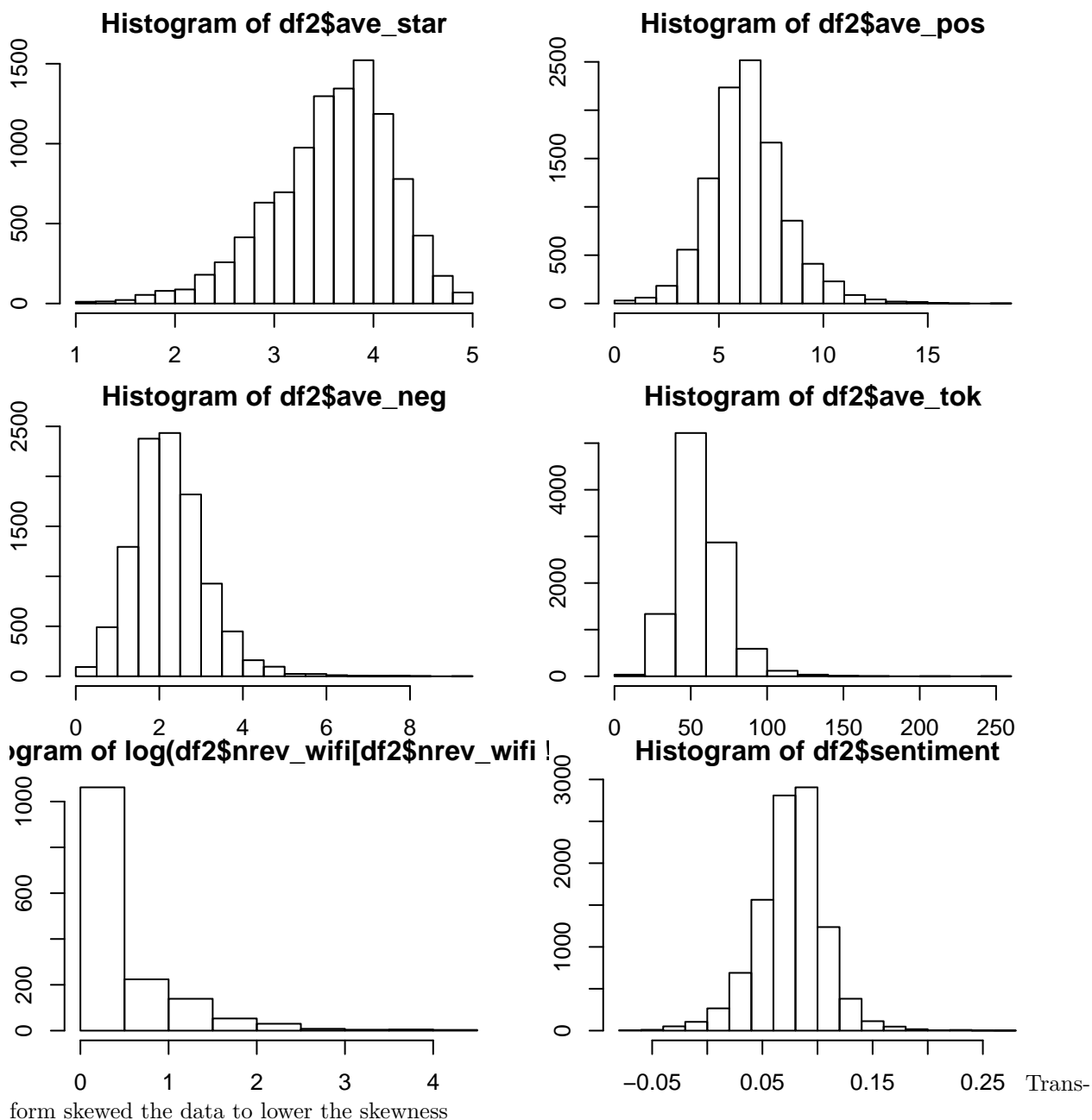
Exploratory modeling

full model

Let's throw all variables into the regular linear model and see how this full model perform and compare it to weighted model.

Compare to weighted lm model: weighted model

From here, we know we need to use weighted model because weighted model give us much higher R square. In both models we see four variables with pretty high vif indicating multicolinearity. To deal with this, first let's try to center numeric vaiables to reduce multicolinearity

Above all,

```
##                Rsqr      sigma vif.ave_pos vif.sentiment vif.ave_tok
## summarym 0.6047803 0.6467686    9.262988      7.493427    5.748540
##          0.6483797 1.5148387   15.615383     10.652487    9.239138
##          0.6047803 0.6286650    9.262988      7.493427    5.748540
##          0.6483797 1.4724370   15.615383     10.652487    9.239138
##          vif.ave_neg vif.attributes.Alcohol attributes.wififree
## summarym    5.006452               3.309733         -0.02051956
##             6.559921               3.319564         -0.02163445
##             5.006452               3.309733         -0.01994520
##             6.559921               3.319564         -0.02102889
##          attributes.wifipaid
## summarym          -0.1509753
##                   -0.1199546
##                   -0.1467494
##                   -0.1165970
```

1. Center numeric data is barely improving anything.
2. Weighted model significantly increase adjusted R square but also increase the vif. Considering even without weights, the vif of the top 4 are so high that we have to remove one or several. So we will stick to weighted and non-centered model from now on. We can either remove variable "sentiment" or both "ave pos" and "ave neg". Let's try both

From here we find when we remove "ave pos" and "ave neg", full model has higher R squared and smaller maximum vif. So we will do so from now on.

Do we have interactions here?

We have lots of categorical data, so a natural question to ask is do any of those interact.

Given : ave_tok

sentiment
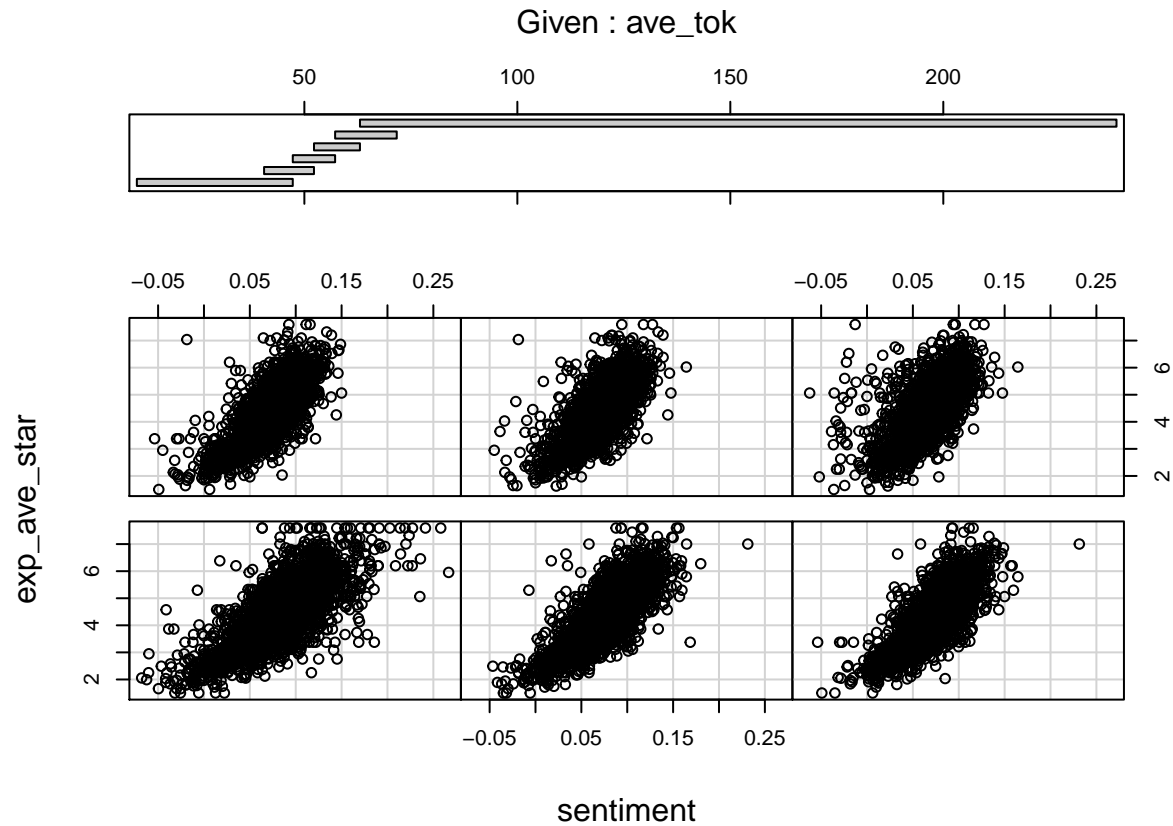
```
## 
## Call:
## lm(formula = exp_ave_star ~ attributes.wifi * ., data = df4,
##     weights = sqrt(nrev))
## 
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -7.5348 -1.0648 -0.1355  0.8795  8.9339
## 
## Coefficients: (1 not defined because of singularities)
##                                            Estimate
## (Intercept)                                2.283e+00
## attributes.wififree                        1.131e-01
## attributes.wifipaid                        3.009e-01
## cityLas Vegas                              1.096e-01
## cityothers                                 8.663e-02
## cityPhoenix                                1.136e-01
## cityPittsburgh                            -1.822e-02
## attributes.Accepts_Credit_Cards          -4.141e-01
## attributes.Good_For_GroupsTRUE           -1.431e-01
## attributes.Outdoor_SeatingTRUE           -7.257e-02
## attributes.Price_Range                    -8.303e-03
## attributes.Good_for_KidsTRUE             -9.603e-02
## attributes.Alcoholfull_bar               -1.922e-01
## attributes.Alcoholnone                    1.387e-02
## attributes.Noise_Levelloud                3.367e-02
## attributes.Noise_Levelquiet              -8.194e-03
```

```
## attributes.Noise_Levelvery_loud                           -2.734e-03
## attributes.Has_TVTRUE                                      -4.932e-02
## attributes.Attiredressy                                     9.790e-02
## attributes.Attireformal                                     4.471e-01
## attributes.DeliveryTRUE                                     -3.709e-02
## attributes.Take_outTRUE                                     5.632e-02
## attributes.Takes_ReservationsTRUE                           -7.595e-03
## attributes.Waiter_ServiceTRUE                               -2.496e-02
## attributes.Ambience.romanticTRUE                            3.075e-02
## attributes.Ambience.intimateTRUE                            1.662e-01
## attributes.Ambience.classyTRUE                              1.170e-01
## attributes.Ambience.hipsterTRUE                             2.668e-01
## attributes.Ambience.diveyTRUE                               2.531e-01
## attributes.Ambience.trendyTRUE                              -3.282e-02
## attributes.Ambience.upscaleTRUE                             1.491e-01
## attributes.Ambience.casualTRUE                              8.600e-02
## attributes.Good_For.dessertTRUE                             -5.915e-02
## attributes.Good_For.latenightTRUE                           3.122e-02
## attributes.Good_For.lunchTRUE                               -3.092e-02
## attributes.Good_For.dinnerTRUE                              -1.603e-02
## attributes.Good_For.breakfastTRUE                           -9.312e-02
## attributes.Good_For.brunchTRUE                              -8.889e-02
## attributes.Parking.garageTRUE                               -2.452e-01
## attributes.Parking.streetTRUE                               2.124e-01
## attributes.Parking.lotTRUE                                  6.597e-02
## attributes.Parking.valetTRUE                                -6.402e-02
## nrev                                                         3.758e-04
## ave_tok                                                      1.181e-02
## log_nrev_wifi                                                3.576e-02
## sentiment                                                    2.677e+01
## Food                                                         2.387e-01
## Bars                                                        -4.455e-02
## American_Traditional                                        -8.283e-02
## Mexican                                                     -4.088e-02
## Pizza                                                        6.785e-02
## Fast_Food                                                   -2.487e-01
## Sandwiches                                                  -3.665e-02
## American_New                                                -9.402e-03
## Italian                                                     -6.190e-02
## Chinese                                                     -7.607e-02
## Burgers                                                      6.039e-02
## `Breakfast_&_Brunch`                                         5.077e-02
## attributes.wififree:cityLas Vegas                           9.882e-03
## attributes.wifipaid:cityLas Vegas                          -1.033e-01
## attributes.wififree:cityothers                             -4.755e-02
## attributes.wifipaid:cityothers                             -3.497e-01
## attributes.wififree:cityPhoenix                            -5.979e-02
## attributes.wifipaid:cityPhoenix                            -4.721e-01
## attributes.wififree:cityPittsburgh                         -4.895e-02
## attributes.wifipaid:cityPittsburgh                          2.394e-01
## attributes.wififree:attributes.Accepts_Credit_Cards         3.715e-02
## attributes.wifipaid:attributes.Accepts_Credit_Cards         1.373e+00
## attributes.wififree:attributes.Good_For_GroupsTRUE         -4.233e-02
## attributes.wifipaid:attributes.Good_For_GroupsTRUE         -2.004e-01
```

```
## attributes.wififree:attributes.Outdoor_SeatingTRUE     -4.511e-02
## attributes.wifipaid:attributes.Outdoor_SeatingTRUE     -2.725e-01
## attributes.wififree:attributes.Price_Range              2.137e-02
## attributes.wifipaid:attributes.Price_Range             -3.468e-01
## attributes.wififree:attributes.Good_for_KidsTRUE        1.334e-02
## attributes.wifipaid:attributes.Good_for_KidsTRUE       -1.937e-01
## attributes.wififree:attributes.Alcoholfull_bar          4.687e-02
## attributes.wifipaid:attributes.Alcoholfull_bar          3.534e-01
## attributes.wififree:attributes.Alcoholnone             -3.749e-02
## attributes.wifipaid:attributes.Alcoholnone             -3.584e-01
## attributes.wififree:attributes.Noise_Levelloud         -3.408e-02
## attributes.wifipaid:attributes.Noise_Levelloud          2.683e-01
## attributes.wififree:attributes.Noise_Levelquiet         7.359e-02
## attributes.wifipaid:attributes.Noise_Levelquiet         7.885e-02
## attributes.wififree:attributes.Noise_Levelvery_loud    -5.137e-02
## attributes.wifipaid:attributes.Noise_Levelvery_loud    -6.442e-01
## attributes.wififree:attributes.Has_TVTRUE              -5.855e-02
## attributes.wifipaid:attributes.Has_TVTRUE              -5.598e-02
## attributes.wififree:attributes.Attiredressy            -3.152e-02
## attributes.wifipaid:attributes.Attiredressy            -2.094e-01
## attributes.wififree:attributes.Attireformal            -5.780e-02
## attributes.wifipaid:attributes.Attireformal                     NA
## attributes.wififree:attributes.DeliveryTRUE             4.030e-02
## attributes.wifipaid:attributes.DeliveryTRUE             8.817e-01
## attributes.wififree:attributes.Take_outTRUE            -8.048e-02
## attributes.wifipaid:attributes.Take_outTRUE            2.194e-01
## attributes.wififree:attributes.Takes_ReservationsTRUE -2.102e-02
## attributes.wifipaid:attributes.Takes_ReservationsTRUE  5.801e-03
## attributes.wififree:attributes.Waiter_ServiceTRUE      2.341e-02
## attributes.wifipaid:attributes.Waiter_ServiceTRUE     -1.745e-01
## attributes.wififree:attributes.Ambience.romanticTRUE  -3.503e-02
## attributes.wifipaid:attributes.Ambience.romanticTRUE   4.304e-01
## attributes.wififree:attributes.Ambience.intimateTRUE  -7.657e-03
## attributes.wifipaid:attributes.Ambience.intimateTRUE  -1.320e-01
## attributes.wififree:attributes.Ambience.classyTRUE    -8.807e-02
## attributes.wifipaid:attributes.Ambience.classyTRUE     1.339e-01
## attributes.wififree:attributes.Ambience.hipsterTRUE   -4.180e-02
## attributes.wifipaid:attributes.Ambience.hipsterTRUE   -2.869e+00
## attributes.wififree:attributes.Ambience.diveyTRUE      6.488e-02
## attributes.wifipaid:attributes.Ambience.diveyTRUE     -1.523e-02
## attributes.wififree:attributes.Ambience.trendyTRUE     1.020e-01
## attributes.wifipaid:attributes.Ambience.trendyTRUE     3.872e-01
## attributes.wififree:attributes.Ambience.upscaleTRUE    2.170e-02
## attributes.wifipaid:attributes.Ambience.upscaleTRUE   -2.586e-01
## attributes.wififree:attributes.Ambience.casualTRUE     7.507e-03
## attributes.wifipaid:attributes.Ambience.casualTRUE     2.837e-02
## attributes.wififree:attributes.Good_For.dessertTRUE    1.101e-01
## attributes.wifipaid:attributes.Good_For.dessertTRUE    1.074e+00
## attributes.wififree:attributes.Good_For.latenightTRUE -7.912e-02
## attributes.wifipaid:attributes.Good_For.latenightTRUE  1.342e-01
## attributes.wififree:attributes.Good_For.lunchTRUE     -3.489e-02
## attributes.wifipaid:attributes.Good_For.lunchTRUE     -4.394e-02
## attributes.wififree:attributes.Good_For.dinnerTRUE    -5.566e-02
## attributes.wifipaid:attributes.Good_For.dinnerTRUE    -2.709e-01
```

```
## attributes.wififree:attributes.Good_For.breakfastTRUE    1.219e-02
## attributes.wifipaid:attributes.Good_For.breakfastTRUE    1.974e-02
## attributes.wififree:attributes.Good_For.brunchTRUE       9.584e-02
## attributes.wifipaid:attributes.Good_For.brunchTRUE       4.652e-01
## attributes.wififree:attributes.Parking.garageTRUE        4.345e-02
## attributes.wifipaid:attributes.Parking.garageTRUE       -3.065e-01
## attributes.wififree:attributes.Parking.streetTRUE       -5.834e-02
## attributes.wifipaid:attributes.Parking.streetTRUE       -5.794e-01
## attributes.wififree:attributes.Parking.lotTRUE          -2.421e-02
## attributes.wifipaid:attributes.Parking.lotTRUE          -1.232e-01
## attributes.wififree:attributes.Parking.valetTRUE        -3.875e-02
## attributes.wifipaid:attributes.Parking.valetTRUE         3.997e-01
## attributes.wififree:nrev                                 3.096e-04
## attributes.wifipaid:nrev                                 1.242e-03
## attributes.wififree:ave_tok                             -1.261e-03
## attributes.wifipaid:ave_tok                             -9.099e-03
## attributes.wififree:log_nrev_wifi                       -4.024e-02
## attributes.wifipaid:log_nrev_wifi                       -7.716e-02
## attributes.wififree:sentiment                            1.413e+00
## attributes.wifipaid:sentiment                           -3.223e+00
## attributes.wififree:Food                                -1.150e-01
## attributes.wifipaid:Food                                 5.759e-01
## attributes.wififree:Bars                                 3.152e-02
## attributes.wifipaid:Bars                                 2.504e-02
## attributes.wififree:American_Traditional                -3.252e-02
## attributes.wifipaid:American_Traditional                -4.271e-02
## attributes.wififree:Mexican                             -8.403e-02
## attributes.wifipaid:Mexican                              5.626e-03
## attributes.wififree:Pizza                                9.903e-03
## attributes.wifipaid:Pizza                               -6.211e-01
## attributes.wififree:Fast_Food                            3.129e-02
## attributes.wifipaid:Fast_Food                           -1.011e+00
## attributes.wififree:Sandwiches                          -3.812e-02
## attributes.wifipaid:Sandwiches                          -5.461e-01
## attributes.wififree:American_New                        -1.219e-01
## attributes.wifipaid:American_New                         2.508e-01
## attributes.wififree:Italian                             -5.583e-02
## attributes.wifipaid:Italian                              3.427e-01
## attributes.wififree:Chinese                             -7.752e-02
## attributes.wifipaid:Chinese                             -5.155e-01
## attributes.wififree:Burgers                             -1.173e-01
## attributes.wifipaid:Burgers                              6.003e-01
## attributes.wififree:`Breakfast_&_Brunch`                -9.683e-02
## attributes.wifipaid:`Breakfast_&_Brunch`                -2.856e-01
##                                                         Std. Error t value
## (Intercept)                                              9.116e-02  25.042
## attributes.wififree                                      1.963e-01   0.576
## attributes.wifipaid                                      1.404e+00   0.214
## cityLas Vegas                                            3.271e-02   3.350
## cityothers                                               3.066e-02   2.825
## cityPhoenix                                              3.455e-02   3.289
## cityPittsburgh                                           4.307e-02  -0.423
## attributes.Accepts_Credit_Cards                          4.408e-02  -9.393
## attributes.Good_For_GroupsTRUE                           2.688e-02  -5.323
```

```
## attributes.Outdoor_SeatingTRUE                 1.583e-02  -4.584
## attributes.Price_Range                         1.859e-02  -0.447
## attributes.Good_for_KidsTRUE                    2.752e-02  -3.490
## attributes.Alcoholfull_bar                      2.305e-02  -8.341
## attributes.Alcoholnone                          2.307e-02   0.601
## attributes.Noise_Levelloud                      2.849e-02   1.182
## attributes.Noise_Levelquiet                     2.146e-02  -0.382
## attributes.Noise_Levelvery_loud                 4.663e-02  -0.059
## attributes.Has_TVTRUE                           1.647e-02  -2.994
## attributes.Attiredressy                         4.612e-02   2.123
## attributes.Attireformal                         2.663e-01   1.679
## attributes.DeliveryTRUE                         2.400e-02  -1.545
## attributes.Take_outTRUE                         3.017e-02   1.867
## attributes.Takes_ReservationsTRUE               2.051e-02  -0.370
## attributes.Waiter_ServiceTRUE                   2.312e-02  -1.079
## attributes.Ambience.romanticTRUE                5.514e-02   0.558
## attributes.Ambience.intimateTRUE                5.950e-02   2.793
## attributes.Ambience.classyTRUE                  4.661e-02   2.510
## attributes.Ambience.hipsterTRUE                 6.199e-02   4.304
## attributes.Ambience.diveyTRUE                   3.774e-02   6.707
## attributes.Ambience.trendyTRUE                  4.027e-02  -0.815
## attributes.Ambience.upscaleTRUE                 6.545e-02   2.278
## attributes.Ambience.casualTRUE                  2.436e-02   3.530
## attributes.Good_For.dessertTRUE                 5.649e-02  -1.047
## attributes.Good_For.latenightTRUE               3.219e-02   0.970
## attributes.Good_For.lunchTRUE                   1.827e-02  -1.692
## attributes.Good_For.dinnerTRUE                  1.835e-02  -0.873
## attributes.Good_For.breakfastTRUE               3.305e-02  -2.817
## attributes.Good_For.brunchTRUE                  3.043e-02  -2.922
## attributes.Parking.garageTRUE                   3.047e-02  -8.047
## attributes.Parking.streetTRUE                   2.775e-02   7.654
## attributes.Parking.lotTRUE                      2.344e-02   2.814
## attributes.Parking.valetTRUE                    3.452e-02  -1.854
## nrev                                            2.611e-05  14.391
## ave_tok                                         6.380e-04  18.510
## log_nrev_wifi                                   1.192e-02   2.999
## sentiment                                       3.172e-01  84.390
## Food                                            2.963e-02   8.055
## Bars                                            2.292e-02  -1.944
## American_Traditional                            2.521e-02  -3.286
## Mexican                                         2.435e-02  -1.679
## Pizza                                           2.937e-02   2.310
## Fast_Food                                       4.235e-02  -5.871
## Sandwiches                                      3.203e-02  -1.144
## American_New                                    2.889e-02  -0.325
## Italian                                         2.822e-02  -2.194
## Chinese                                         2.871e-02  -2.650
## Burgers                                         3.154e-02   1.914
## `Breakfast_&_Brunch`                            3.608e-02   1.407
## attributes.wififree:cityLas Vegas               5.027e-02   0.197
## attributes.wifipaid:cityLas Vegas               4.517e-01  -0.229
## attributes.wififree:cityothers                  4.543e-02  -1.047
## attributes.wifipaid:cityothers                  3.540e-01  -0.988
## attributes.wififree:cityPhoenix                 5.200e-02  -1.150
```

```
## attributes.wifipaid:cityPhoenix                           5.151e-01  -0.917
## attributes.wififree:cityPittsburgh                        7.039e-02  -0.695
## attributes.wifipaid:cityPittsburgh                        6.434e-01   0.372
## attributes.wififree:attributes.Accepts_Credit_Cards       1.312e-01   0.283
## attributes.wifipaid:attributes.Accepts_Credit_Cards       1.206e+00   1.138
## attributes.wififree:attributes.Good_For_GroupsTRUE        5.626e-02  -0.752
## attributes.wifipaid:attributes.Good_For_GroupsTRUE        4.516e-01  -0.444
## attributes.wififree:attributes.Outdoor_SeatingTRUE        2.757e-02  -1.636
## attributes.wifipaid:attributes.Outdoor_SeatingTRUE        2.100e-01  -1.298
## attributes.wififree:attributes.Price_Range                3.171e-02   0.674
## attributes.wifipaid:attributes.Price_Range                2.526e-01  -1.373
## attributes.wififree:attributes.Good_for_KidsTRUE          4.112e-02   0.324
## attributes.wifipaid:attributes.Good_for_KidsTRUE          3.126e-01  -0.620
## attributes.wififree:attributes.Alcoholfull_bar            4.123e-02   1.137
## attributes.wifipaid:attributes.Alcoholfull_bar            3.609e-01   0.979
## attributes.wififree:attributes.Alcoholnone                4.167e-02  -0.900
## attributes.wifipaid:attributes.Alcoholnone                3.715e-01  -0.965
## attributes.wififree:attributes.Noise_Levelloud            4.642e-02  -0.734
## attributes.wifipaid:attributes.Noise_Levelloud            2.931e-01   0.916
## attributes.wififree:attributes.Noise_Levelquiet           3.922e-02   1.876
## attributes.wifipaid:attributes.Noise_Levelquiet           2.399e-01   0.329
## attributes.wififree:attributes.Noise_Levelvery_loud       8.866e-02  -0.579
## attributes.wifipaid:attributes.Noise_Levelvery_loud       4.285e-01  -1.504
## attributes.wififree:attributes.Has_TVTRUE                 2.949e-02  -1.985
## attributes.wifipaid:attributes.Has_TVTRUE                 2.093e-01  -0.267
## attributes.wififree:attributes.Attiredressy               8.063e-02  -0.391
## attributes.wifipaid:attributes.Attiredressy               4.938e-01  -0.424
## attributes.wififree:attributes.Attireformal               5.704e-01  -0.101
## attributes.wifipaid:attributes.Attireformal                      NA      NA
## attributes.wififree:attributes.DeliveryTRUE               3.820e-02   1.055
## attributes.wifipaid:attributes.DeliveryTRUE               3.685e-01   2.393
## attributes.wififree:attributes.Take_outTRUE               5.515e-02  -1.459
## attributes.wifipaid:attributes.Take_outTRUE               2.723e-01   0.806
## attributes.wififree:attributes.Takes_ReservationsTRUE     3.250e-02  -0.647
## attributes.wifipaid:attributes.Takes_ReservationsTRUE     2.410e-01   0.024
## attributes.wififree:attributes.Waiter_ServiceTRUE         4.004e-02   0.585
## attributes.wifipaid:attributes.Waiter_ServiceTRUE         2.673e-01  -0.653
## attributes.wififree:attributes.Ambience.romanticTRUE      9.789e-02  -0.358
## attributes.wifipaid:attributes.Ambience.romanticTRUE      5.679e-01   0.758
## attributes.wififree:attributes.Ambience.intimateTRUE      1.168e-01  -0.066
## attributes.wifipaid:attributes.Ambience.intimateTRUE      7.797e-01  -0.169
## attributes.wififree:attributes.Ambience.classyTRUE        7.849e-02  -1.122
## attributes.wifipaid:attributes.Ambience.classyTRUE        5.541e-01   0.242
## attributes.wififree:attributes.Ambience.hipsterTRUE       8.446e-02  -0.495
## attributes.wifipaid:attributes.Ambience.hipsterTRUE       1.552e+00  -1.848
## attributes.wififree:attributes.Ambience.diveyTRUE         7.345e-02   0.883
## attributes.wifipaid:attributes.Ambience.diveyTRUE         4.987e-01  -0.031
## attributes.wififree:attributes.Ambience.trendyTRUE        6.354e-02   1.606
## attributes.wifipaid:attributes.Ambience.trendyTRUE        6.715e-01   0.577
## attributes.wififree:attributes.Ambience.upscaleTRUE       1.193e-01   0.182
## attributes.wifipaid:attributes.Ambience.upscaleTRUE       6.687e-01  -0.387
## attributes.wififree:attributes.Ambience.casualTRUE        4.261e-02   0.176
## attributes.wifipaid:attributes.Ambience.casualTRUE        2.305e-01   0.123
## attributes.wififree:attributes.Good_For.dessertTRUE       8.463e-02   1.301
```

```
## attributes.wifipaid:attributes.Good_For.dessertTRUE        1.226e+00    0.876
## attributes.wififree:attributes.Good_For.latenightTRUE      5.078e-02   -1.558
## attributes.wifipaid:attributes.Good_For.latenightTRUE      5.750e-01    0.233
## attributes.wififree:attributes.Good_For.lunchTRUE          3.138e-02   -1.112
## attributes.wifipaid:attributes.Good_For.lunchTRUE          2.176e-01   -0.202
## attributes.wififree:attributes.Good_For.dinnerTRUE         3.167e-02   -1.757
## attributes.wifipaid:attributes.Good_For.dinnerTRUE         2.305e-01   -1.175
## attributes.wififree:attributes.Good_For.breakfastTRUE      5.043e-02    0.242
## attributes.wifipaid:attributes.Good_For.breakfastTRUE      2.309e-01    0.085
## attributes.wififree:attributes.Good_For.brunchTRUE         4.653e-02    2.060
## attributes.wifipaid:attributes.Good_For.brunchTRUE         2.639e-01    1.763
## attributes.wififree:attributes.Parking.garageTRUE          4.654e-02    0.934
## attributes.wifipaid:attributes.Parking.garageTRUE          2.573e-01   -1.191
## attributes.wififree:attributes.Parking.streetTRUE          4.057e-02   -1.438
## attributes.wifipaid:attributes.Parking.streetTRUE          5.293e-01   -1.095
## attributes.wififree:attributes.Parking.lotTRUE             3.668e-02   -0.660
## attributes.wifipaid:attributes.Parking.lotTRUE             2.062e-01   -0.598
## attributes.wififree:attributes.Parking.valetTRUE           5.446e-02   -0.712
## attributes.wifipaid:attributes.Parking.valetTRUE           2.684e-01    1.489
## attributes.wififree:nrev                                   5.219e-05    5.933
## attributes.wifipaid:nrev                                   1.540e-03    0.806
## attributes.wififree:ave_tok                                1.104e-03   -1.142
## attributes.wifipaid:ave_tok                                6.946e-03   -1.310
## attributes.wififree:log_nrev_wifi                          1.215e-02   -3.312
## attributes.wifipaid:log_nrev_wifi                          4.436e-02   -1.739
## attributes.wififree:sentiment                              5.436e-01    2.598
## attributes.wifipaid:sentiment                              3.328e+00   -0.968
## attributes.wififree:Food                                   4.383e-02   -2.624
## attributes.wifipaid:Food                                   4.779e-01    1.205
## attributes.wififree:Bars                                   3.500e-02    0.901
## attributes.wifipaid:Bars                                   2.844e-01    0.088
## attributes.wififree:American_Traditional                   3.958e-02   -0.821
## attributes.wifipaid:American_Traditional                   2.865e-01   -0.149
## attributes.wififree:Mexican                                4.643e-02   -1.810
## attributes.wifipaid:Mexican                                4.491e-01    0.013
## attributes.wififree:Pizza                                  4.882e-02    0.203
## attributes.wifipaid:Pizza                                  5.021e-01   -1.237
## attributes.wififree:Fast_Food                              8.019e-02    0.390
## attributes.wifipaid:Fast_Food                              7.506e-01   -1.347
## attributes.wififree:Sandwiches                             4.784e-02   -0.797
## attributes.wifipaid:Sandwiches                             4.387e-01   -1.245
## attributes.wififree:American_New                           4.064e-02   -2.999
## attributes.wifipaid:American_New                           2.861e-01    0.876
## attributes.wififree:Italian                                5.002e-02   -1.116
## attributes.wifipaid:Italian                                2.962e-01    1.157
## attributes.wififree:Chinese                                6.676e-02   -1.161
## attributes.wifipaid:Chinese                                5.789e-01   -0.890
## attributes.wififree:Burgers                                5.113e-02   -2.295
## attributes.wifipaid:Burgers                                3.907e-01    1.536
## attributes.wififree:`Breakfast_&_Brunch`                   5.087e-02   -1.904
## attributes.wifipaid:`Breakfast_&_Brunch`                   2.933e-01   -0.974
##                                                            Pr(>|t|)
## (Intercept)                                                < 2e-16 ***
## attributes.wififree                                        0.564545
```

```
## attributes.wifipaid                      0.830308
## cityLas Vegas                             0.000810 ***
## cityothers                                0.004731 **
## cityPhoenix                               0.001008 **
## cityPittsburgh                            0.672250
## attributes.Accepts_Credit_Cards           < 2e-16 ***
## attributes.Good_For_GroupsTRUE           1.04e-07 ***
## attributes.Outdoor_SeatingTRUE           4.61e-06 ***
## attributes.Price_Range                    0.655215
## attributes.Good_for_KidsTRUE             0.000486 ***
## attributes.Alcoholfull_bar                < 2e-16 ***
## attributes.Alcoholnone                    0.547821
## attributes.Noise_Levelloud                0.237321
## attributes.Noise_Levelquiet               0.702656
## attributes.Noise_Levelvery_loud           0.953253
## attributes.Has_TVTRUE                     0.002761 **
## attributes.Attiredressy                   0.033788 *
## attributes.Attireformal                   0.093209 .
## attributes.DeliveryTRUE                    0.122284
## attributes.Take_outTRUE                    0.061984 .
## attributes.Takes_ReservationsTRUE          0.711139
## attributes.Waiter_ServiceTRUE             0.280447
## attributes.Ambience.romanticTRUE          0.576991
## attributes.Ambience.intimateTRUE          0.005238 **
## attributes.Ambience.classyTRUE            0.012083 *
## attributes.Ambience.hipsterTRUE          1.70e-05 ***
## attributes.Ambience.diveyTRUE            2.10e-11 ***
## attributes.Ambience.trendyTRUE            0.415105
## attributes.Ambience.upscaleTRUE           0.022771 *
## attributes.Ambience.casualTRUE            0.000417 ***
## attributes.Good_For.dessertTRUE           0.295097
## attributes.Good_For.latenightTRUE         0.332100
## attributes.Good_For.lunchTRUE             0.090666 .
## attributes.Good_For.dinnerTRUE            0.382524
## attributes.Good_For.breakfastTRUE         0.004855 **
## attributes.Good_For.brunchTRUE            0.003491 **
## attributes.Parking.garageTRUE            9.44e-16 ***
## attributes.Parking.streetTRUE            2.13e-14 ***
## attributes.Parking.lotTRUE                0.004895 **
## attributes.Parking.valetTRUE              0.063708 .
## nrev                                       < 2e-16 ***
## ave_tok                                    < 2e-16 ***
## log_nrev_wifi                             0.002717 **
## sentiment                                  < 2e-16 ***
## Food                                     8.85e-16 ***
## Bars                                      0.051957 .
## American_Traditional                      0.001020 **
## Mexican                                   0.093148 .
## Pizza                                     0.020903 *
## Fast_Food                                4.46e-09 ***
## Sandwiches                                0.252531
## American_New                              0.744842
## Italian                                   0.028274 *
## Chinese                                   0.008070 **
```

```
## Burgers                                               0.055588 .
## `Breakfast_&_Brunch`                                  0.159408
## attributes.wififree:cityLas Vegas                     0.844154
## attributes.wifipaid:cityLas Vegas                     0.819120
## attributes.wififree:cityothers                        0.295244
## attributes.wifipaid:cityothers                        0.323196
## attributes.wififree:cityPhoenix                       0.250308
## attributes.wifipaid:cityPhoenix                       0.359409
## attributes.wififree:cityPittsburgh                    0.486849
## attributes.wifipaid:cityPittsburgh                    0.709778
## attributes.wififree:attributes.Accepts_Credit_Cards   0.777090
## attributes.wifipaid:attributes.Accepts_Credit_Cards   0.254955
## attributes.wififree:attributes.Good_For_GroupsTRUE    0.451885
## attributes.wifipaid:attributes.Good_For_GroupsTRUE    0.657313
## attributes.wififree:attributes.Outdoor_SeatingTRUE    0.101829
## attributes.wifipaid:attributes.Outdoor_SeatingTRUE    0.194441
## attributes.wififree:attributes.Price_Range            0.500476
## attributes.wifipaid:attributes.Price_Range            0.169849
## attributes.wififree:attributes.Good_for_KidsTRUE      0.745682
## attributes.wifipaid:attributes.Good_for_KidsTRUE      0.535488
## attributes.wififree:attributes.Alcoholfull_bar        0.255596
## attributes.wifipaid:attributes.Alcoholfull_bar        0.327495
## attributes.wififree:attributes.Alcoholnone            0.368214
## attributes.wifipaid:attributes.Alcoholnone            0.334661
## attributes.wififree:attributes.Noise_Levelloud        0.462792
## attributes.wifipaid:attributes.Noise_Levelloud        0.359942
## attributes.wififree:attributes.Noise_Levelquiet       0.060653 .
## attributes.wifipaid:attributes.Noise_Levelquiet       0.742440
## attributes.wififree:attributes.Noise_Levelvery_loud   0.562342
## attributes.wifipaid:attributes.Noise_Levelvery_loud   0.132727
## attributes.wififree:attributes.Has_TVTRUE             0.047147 *
## attributes.wifipaid:attributes.Has_TVTRUE             0.789157
## attributes.wififree:attributes.Attiredressy           0.695866
## attributes.wifipaid:attributes.Attiredressy           0.671558
## attributes.wififree:attributes.Attireformal           0.919282
## attributes.wifipaid:attributes.Attireformal                 NA
## attributes.wififree:attributes.DeliveryTRUE           0.291404
## attributes.wifipaid:attributes.DeliveryTRUE           0.016751 *
## attributes.wififree:attributes.Take_outTRUE           0.144572
## attributes.wifipaid:attributes.Take_outTRUE           0.420300
## attributes.wififree:attributes.Takes_ReservationsTRUE 0.517774
## attributes.wifipaid:attributes.Takes_ReservationsTRUE 0.980801
## attributes.wififree:attributes.Waiter_ServiceTRUE     0.558807
## attributes.wifipaid:attributes.Waiter_ServiceTRUE     0.513837
## attributes.wififree:attributes.Ambience.romanticTRUE  0.720452
## attributes.wifipaid:attributes.Ambience.romanticTRUE  0.448496
## attributes.wififree:attributes.Ambience.intimateTRUE  0.947753
## attributes.wifipaid:attributes.Ambience.intimateTRUE  0.865520
## attributes.wififree:attributes.Ambience.classyTRUE    0.261858
## attributes.wifipaid:attributes.Ambience.classyTRUE    0.809104
## attributes.wififree:attributes.Ambience.hipsterTRUE   0.620718
## attributes.wifipaid:attributes.Ambience.hipsterTRUE   0.064601 .
## attributes.wififree:attributes.Ambience.diveyTRUE     0.377127
## attributes.wifipaid:attributes.Ambience.diveyTRUE     0.975633
```

```
## attributes.wififree:attributes.Ambience.trendyTRUE      0.108396
## attributes.wifipaid:attributes.Ambience.trendyTRUE      0.564198
## attributes.wififree:attributes.Ambience.upscaleTRUE     0.855624
## attributes.wifipaid:attributes.Ambience.upscaleTRUE     0.698950
## attributes.wififree:attributes.Ambience.casualTRUE      0.860171
## attributes.wifipaid:attributes.Ambience.casualTRUE      0.902031
## attributes.wififree:attributes.Good_For.dessertTRUE     0.193393
## attributes.wifipaid:attributes.Good_For.dessertTRUE     0.380813
## attributes.wififree:attributes.Good_For.latenightTRUE 0.119237
## attributes.wifipaid:attributes.Good_For.latenightTRUE 0.815436
## attributes.wififree:attributes.Good_For.lunchTRUE       0.266194
## attributes.wifipaid:attributes.Good_For.lunchTRUE       0.840019
## attributes.wififree:attributes.Good_For.dinnerTRUE      0.078867 .
## attributes.wifipaid:attributes.Good_For.dinnerTRUE      0.239966
## attributes.wififree:attributes.Good_For.breakfastTRUE 0.808966
## attributes.wifipaid:attributes.Good_For.breakfastTRUE 0.931883
## attributes.wififree:attributes.Good_For.brunchTRUE      0.039451 *
## attributes.wifipaid:attributes.Good_For.brunchTRUE      0.078003 .
## attributes.wififree:attributes.Parking.garageTRUE       0.350489
## attributes.wifipaid:attributes.Parking.garageTRUE       0.233641
## attributes.wififree:attributes.Parking.streetTRUE       0.150440
## attributes.wifipaid:attributes.Parking.streetTRUE       0.273650
## attributes.wififree:attributes.Parking.lotTRUE          0.509144
## attributes.wifipaid:attributes.Parking.lotTRUE          0.550151
## attributes.wififree:attributes.Parking.valetTRUE        0.476697
## attributes.wifipaid:attributes.Parking.valetTRUE        0.136510
## attributes.wififree:nrev                                3.08e-09 ***
## attributes.wifipaid:nrev                                0.420016
## attributes.wififree:ave_tok                             0.253636
## attributes.wifipaid:ave_tok                             0.190206
## attributes.wififree:log_nrev_wifi                       0.000930 ***
## attributes.wifipaid:log_nrev_wifi                       0.081990 .
## attributes.wififree:sentiment                           0.009380 **
## attributes.wifipaid:sentiment                           0.332847
## attributes.wififree:Food                                0.008694 **
## attributes.wifipaid:Food                                0.228205
## attributes.wififree:Bars                                0.367793
## attributes.wifipaid:Bars                                0.929864
## attributes.wififree:American_Traditional                0.411385
## attributes.wifipaid:American_Traditional                0.881492
## attributes.wififree:Mexican                             0.070348 .
## attributes.wifipaid:Mexican                             0.990005
## attributes.wififree:Pizza                               0.839261
## attributes.wifipaid:Pizza                               0.216141
## attributes.wififree:Fast_Food                           0.696379
## attributes.wifipaid:Fast_Food                           0.178137
## attributes.wififree:Sandwiches                          0.425581
## attributes.wifipaid:Sandwiches                          0.213203
## attributes.wififree:American_New                        0.002718 **
## attributes.wifipaid:American_New                        0.380815
## attributes.wififree:Italian                             0.264420
## attributes.wifipaid:Italian                             0.247232
## attributes.wififree:Chinese                             0.245616
## attributes.wifipaid:Chinese                             0.373251
```

```
## attributes.wififree:Burgers                          0.021782 *
## attributes.wifipaid:Burgers                          0.124448
## attributes.wififree:`Breakfast_&_Brunch`             0.056995 .
## attributes.wifipaid:`Breakfast_&_Brunch`             0.330246
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.549 on 10050 degrees of freedom
## Multiple R-squared:  0.6385, Adjusted R-squared:  0.6325
## F-statistic: 106.9 on 166 and 10050 DF,  p-value: < 2.2e-16
```

Given : as.factor(attributes.wifi)

exp_ave_star

log_nrev_wifi
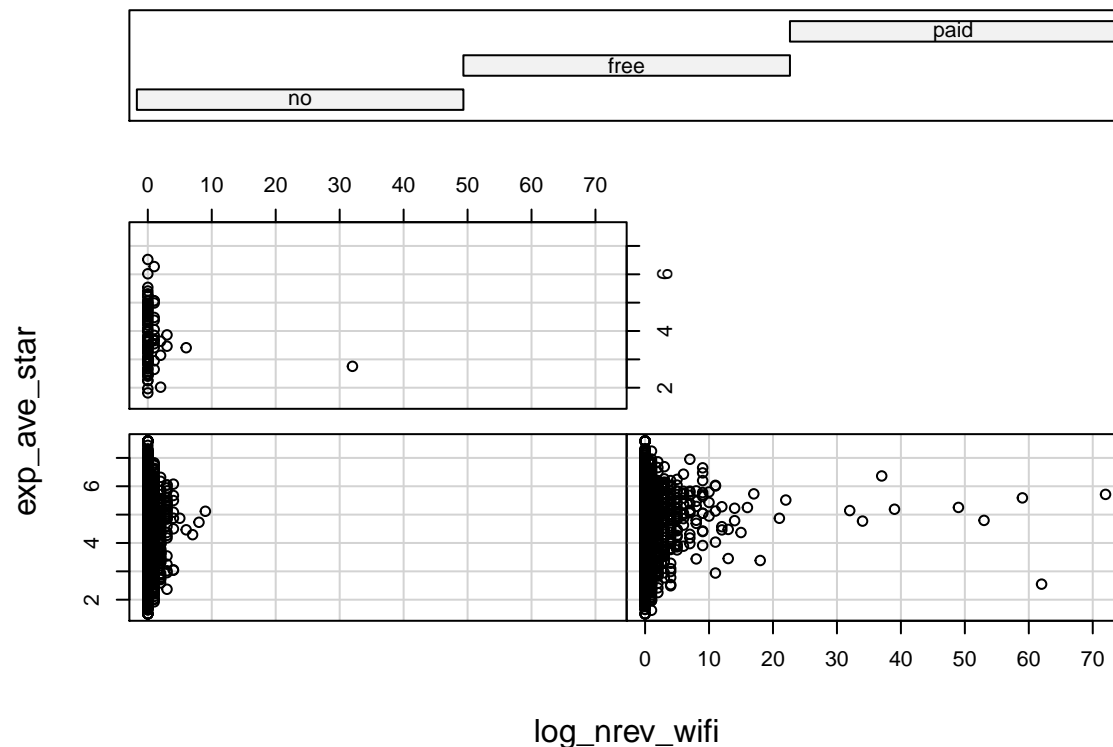
We do find
a little interaction between 1. number of reviews and number of reviews talking about wifi with wifi status 2.
ave_tok and sentiment So we will include those 3 interacting term in my model by add interacting variables
in the dataset.

Comparing all full models

```
summarym
```

```
##                        Rsqr     sigma      vif       vif      vif
## original          0.6047803 0.6467686  9.262988  7.493427 5.748540
## weighted          0.6483797 1.5148387 15.615383 10.652487 9.239138
## centered          0.6047803 0.6286650  9.262988  7.493427 5.748540
## centered_weighted 0.6483797 1.4724370 15.615383 10.652487 9.239138
## -sentiment        0.6235987 1.5673104  5.583675  3.841143 3.318218
## -pos and -neg     0.6284231 1.5572337  3.305844  3.186477 2.706598
## +interaction1     0.6308897 1.5520564 13.641470 10.169595 7.018457
## +interaction2     0.6285856 1.5568932  3.308806  3.186555 2.708015
```

16

```
##                          vif      vif attributes.wififree
## original           5.006452 3.309733         -0.02051956
## weighted           6.559921 3.319564         -0.02163445
## centered           5.006452 3.309733         -0.01994520
## centered_weighted  6.559921 3.319564         -0.02102889
## -sentiment         3.186557 2.711912         -0.02428633
## -pos and -neg      2.296321 2.191786         -0.01494762
## +interaction1      3.308822 3.187737         -0.01122650
## +interaction2      2.301848 2.193567         -0.00895865
##                       attributes.wifipaid
## original                       -0.1509753
## weighted                       -0.1199546
## centered                       -0.1467494
## centered_weighted              -0.1165970
## -sentiment                     -0.1238664
## -pos and -neg                  -0.1393911
## +interaction1                  -0.1164799
## +interaction2                  -0.1156762
```
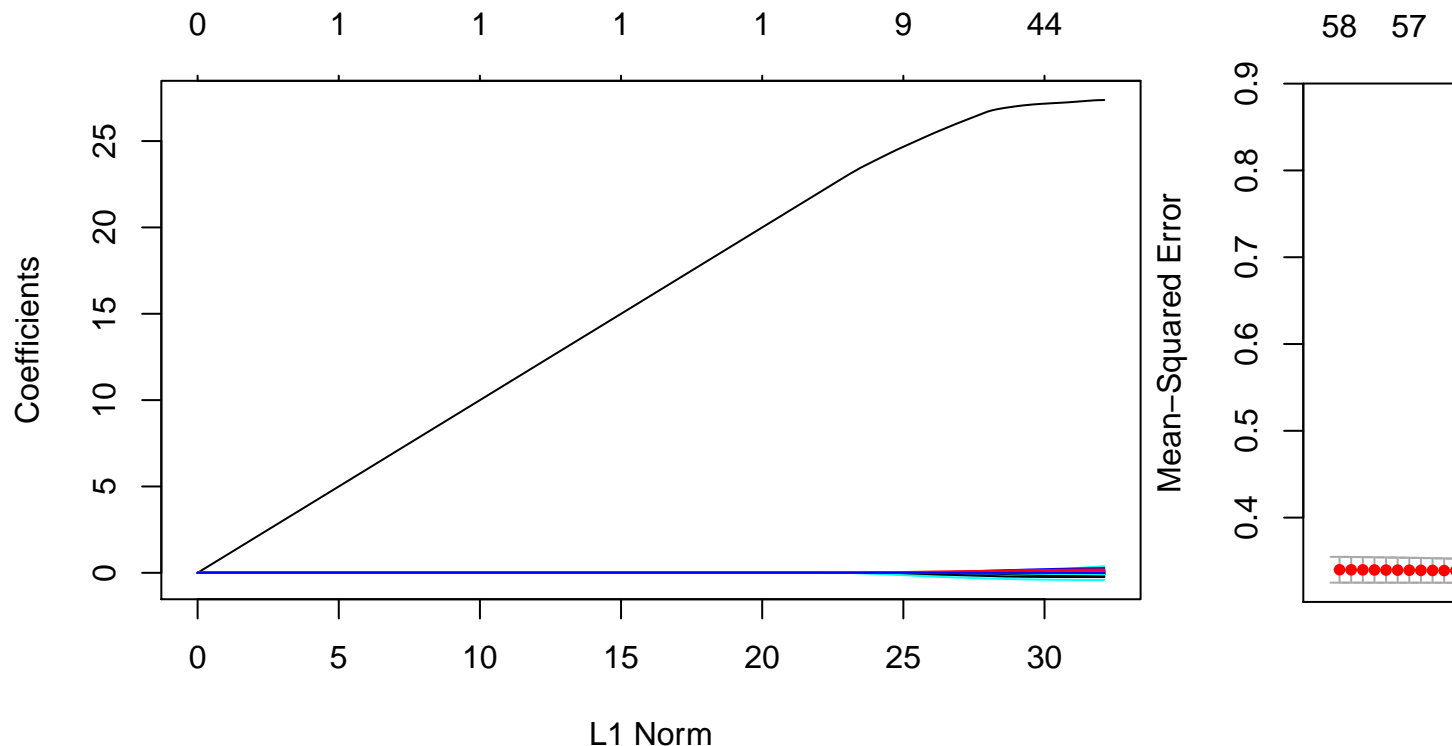
```r
write.csv(summarym, "summarym.csv")
```

Model selection

After we have our full model, several methods are used to choose best parsimonious model by lasso


This tell us: To explain number of stars, variable sentiment almost explained all of it. And all other variables combined have only marginal effects after we include sentiment in the model. And also sentiment of customers reflect overall experience to that restuarant, which means all the attributes we gave them are likely actually contributing to sentiment and then to stars. Our focus is about wifi access. So let's leave R square aside and try model without sentiment.
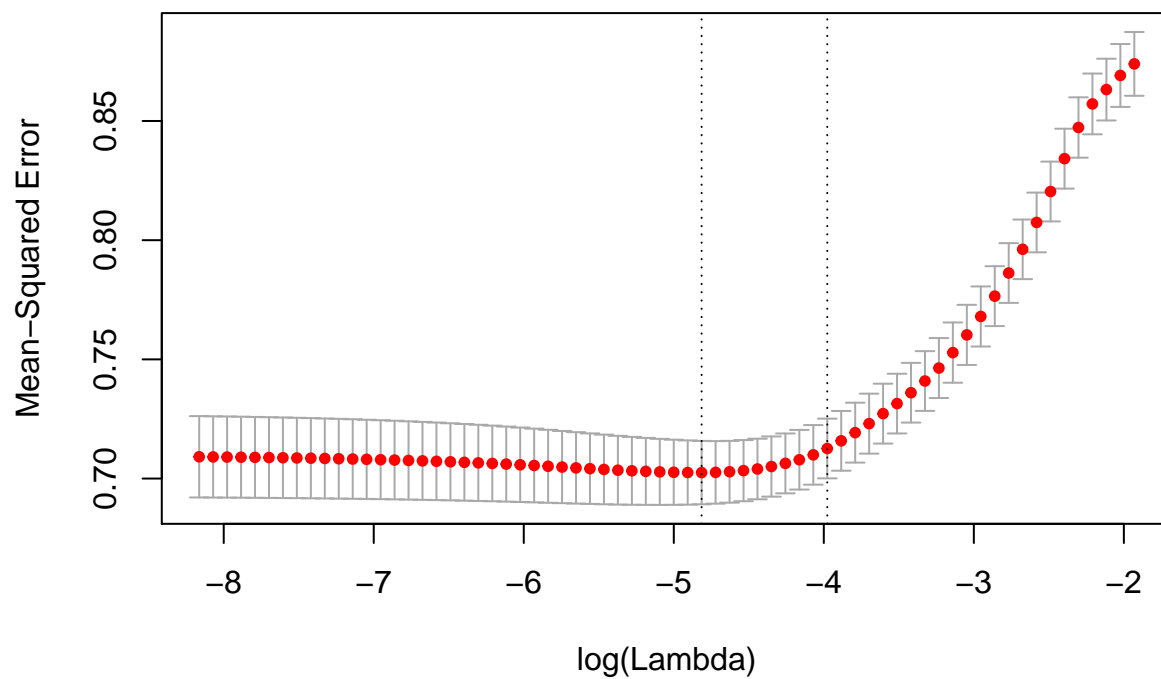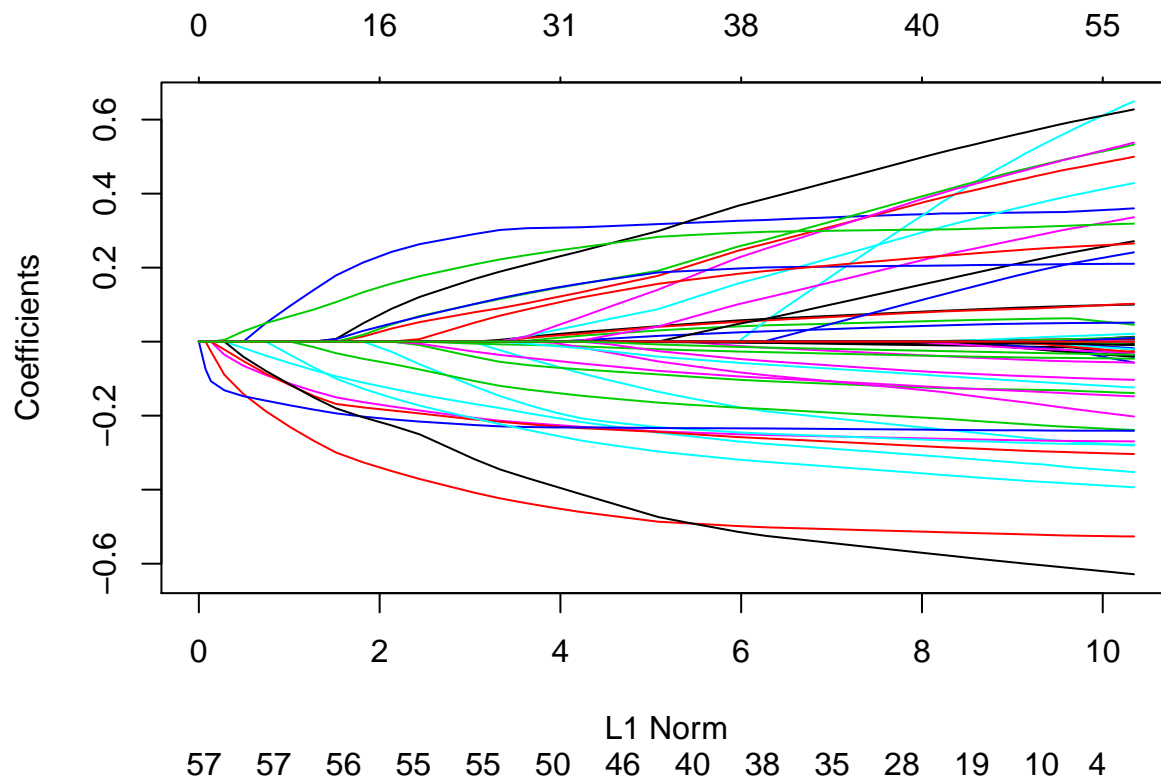
```
#png(file="12.png",width=2500,height=1700,res=200)
suppressWarnings(ggplot(data=data.frame("Chosen_Var"=factor(paranames2, levels=paranames2), "Fit_Value"=
```

Warning: Stacking not well defined when ymin != 0

Coeficients of Lasso Regression Model Fit
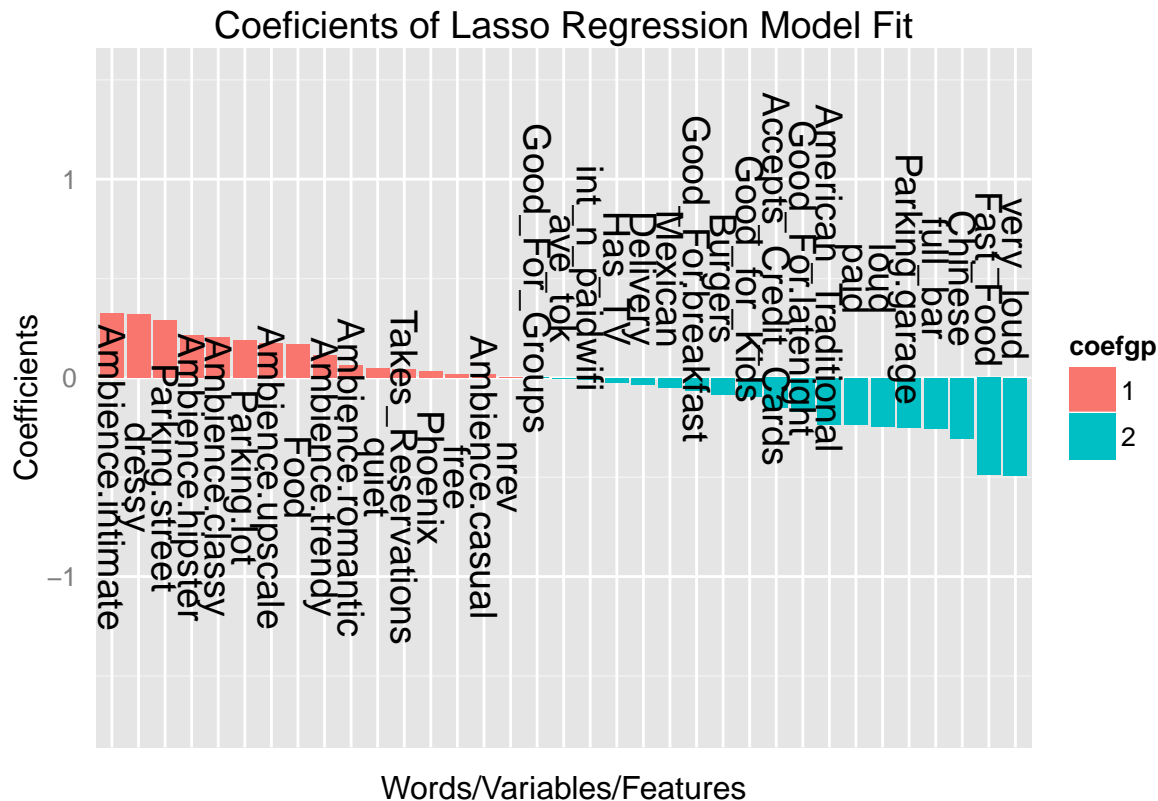
```
#dev.off()
```

Results:

In this data set, variable sentiment is too good to explain the ratings been give. And this sentiment are also related to all other variables to a certain extend. So makes all the coeficients of others not so significant. If we exclude sentiment and just analyze the contributions of other attributes. R square dropped dramatically but there are still 20% been accounted. When we include sentiment, coefficient of wifi free is always below zero. Which can be explained by multicolinearity. When sentiment excluded, the coefficient of wifi free become positive and wifi paid remaind negative. And both are significant.

Discussion:

Wifi are not as important as I think. People most likely will not give higher stars because of free wifi. However, paid wifi is hurting people's feeling from various evidence. So if you want to have your own restuarant, please provide free wifi or don't provide wifi access. One reason that wifi is irrelevant might be the wireless data is getting cheaper. Because less and less people mention wifi along the past 5 or 6 years. BIG BIG confounder here!! Nothing about service and food quality considered, which cannot be easily quantified unless extracting from review. Lots of work could be done to do more text mining and also we can study which type of restuarants needs free wifi than others.

Thank you for reading!!