STAT 5371 final project—-WiFi matters?

Li Sun, Yang Cai, Hong Li

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.

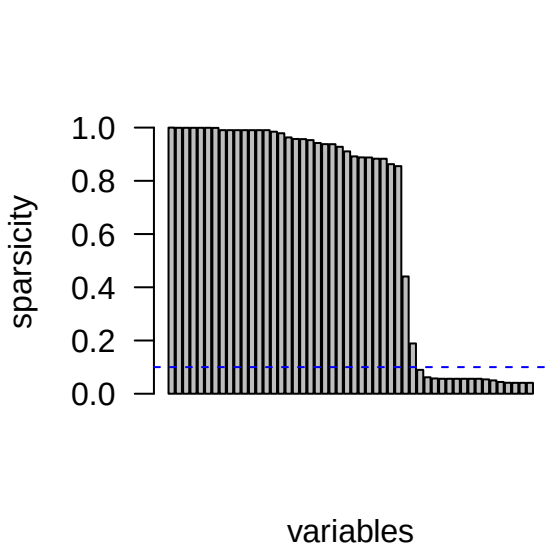
Methods and Data:

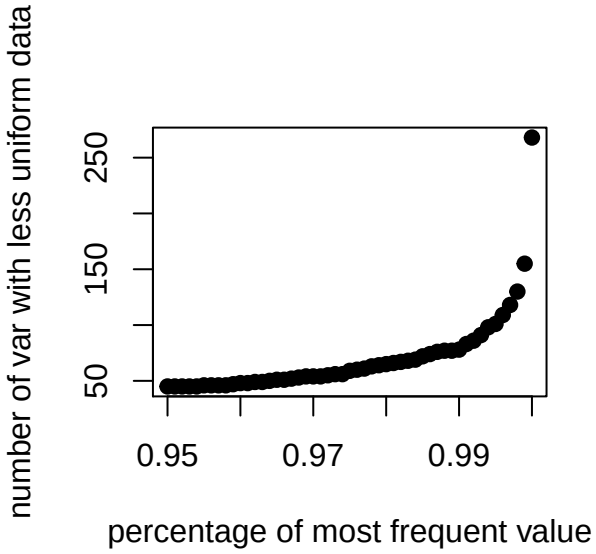
Data source are from 'http://www.yelp.com/dataset_challenge'. Data comes in json format, and there are 5 files: user, business, checkin, tip, and reviews. To answer my specific question: 'Does wifi access really helps improve customers' experience in restuarants?' I choose only three data sets from these 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.] Preliminary data process: 1. Read in Json data 2. Clean data and merge data sets 3. get rid of irrelevant infomation like geographic variables These processes resulted in a single data frame "busi3".

First load needed packages and read in busi3 data frame.

1. Missing values and data types

Next 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.

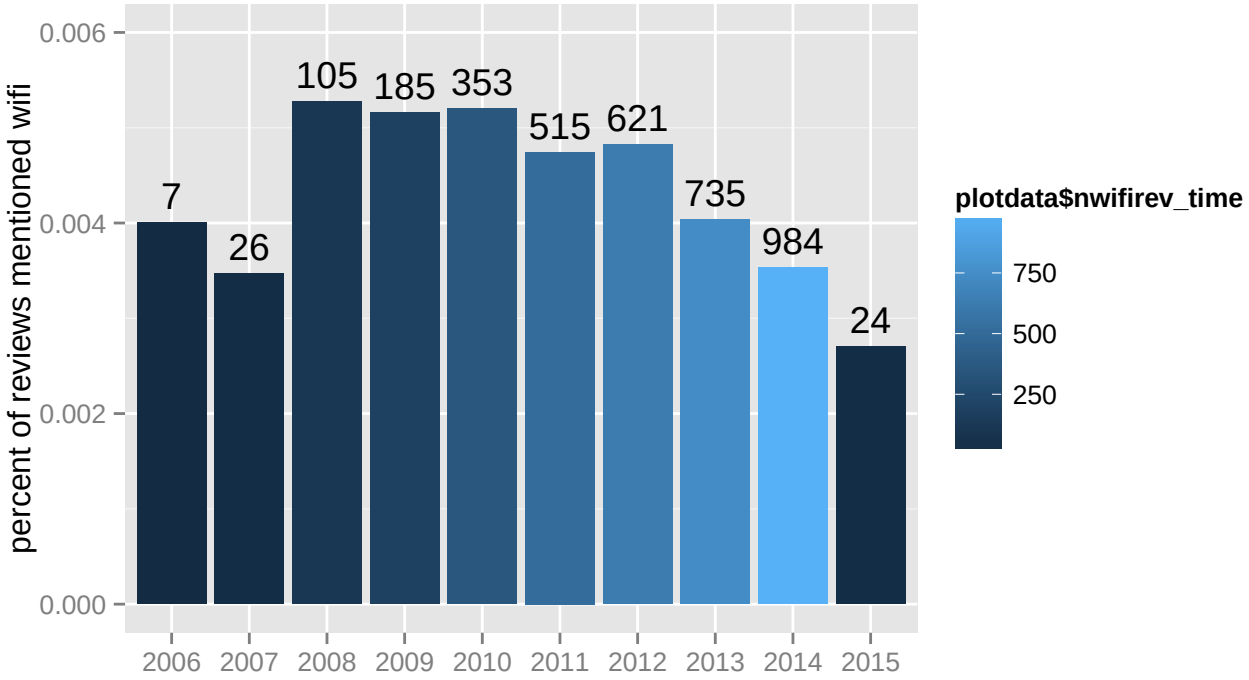




Exploratory Data Analysis

What people say about wifi

In this section, we look at the data generally to explore the posibility of answering our question about wifi. So we are looking at how many people are commenting on wifi along the years and what are they saying.



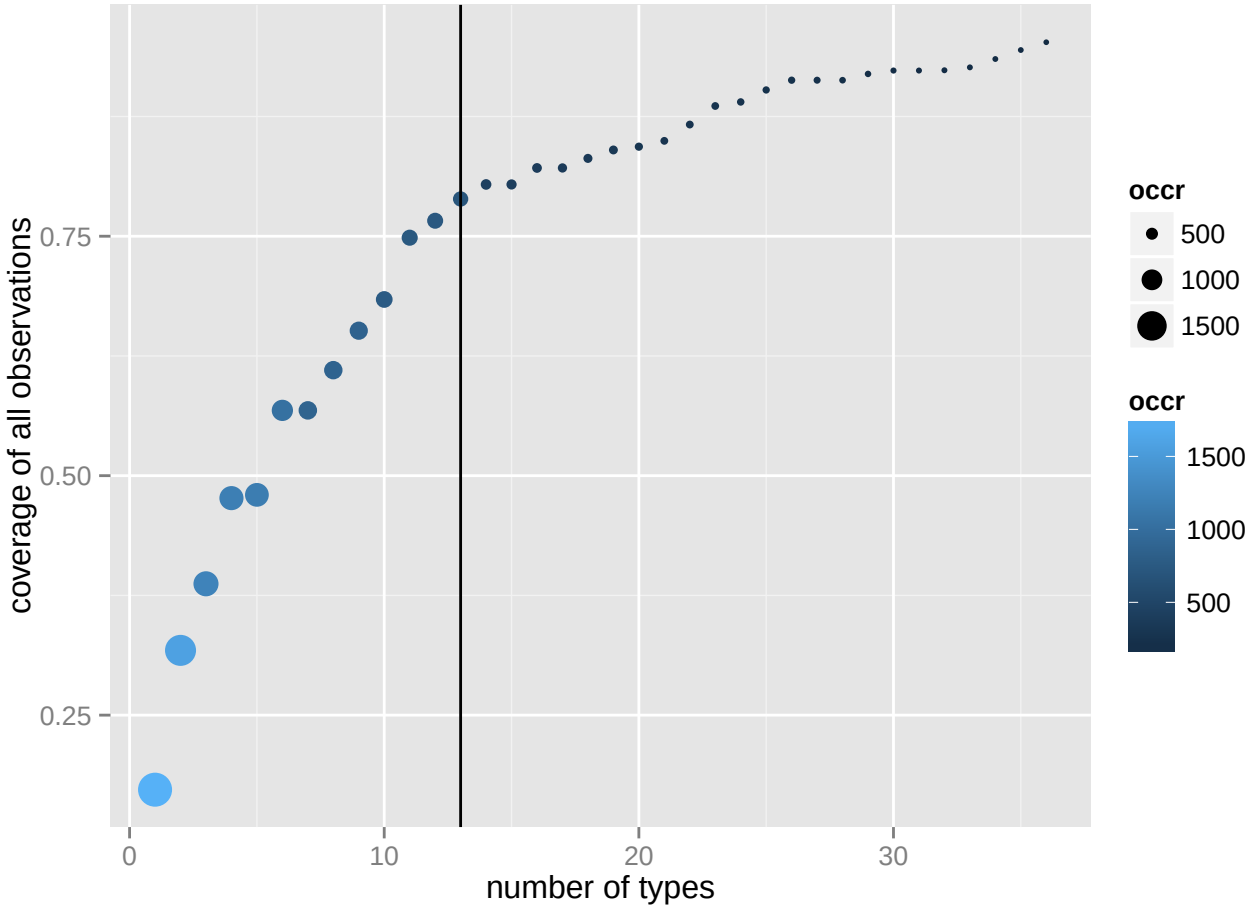
People are less and less caring about wifi since 2008 according to frequency of reviews mentioning “wifi”, or “internet”

It is always straightforward to see wordcloud. The left one is 1-gram token and right one is 2-gram tokens.

City and category reduction

Then 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 into “others”.

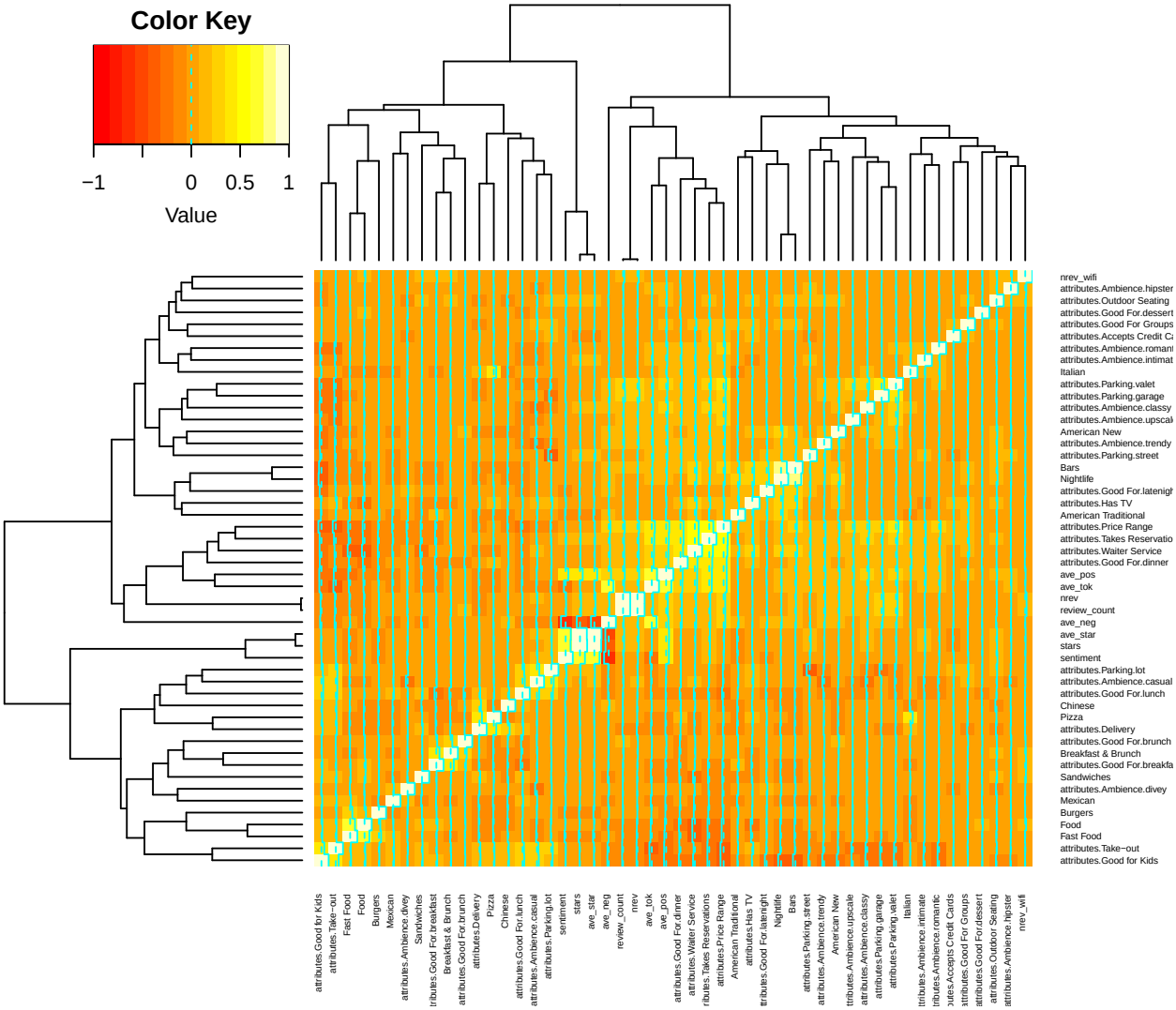
We still have too many variables specifying types of restuarants. Some of the types are rare, with less occurance, that would be of our less interest. So I want to cut the type of restuarants down to several majors types. The standard of maintainning certain types depends on occurance and also we want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

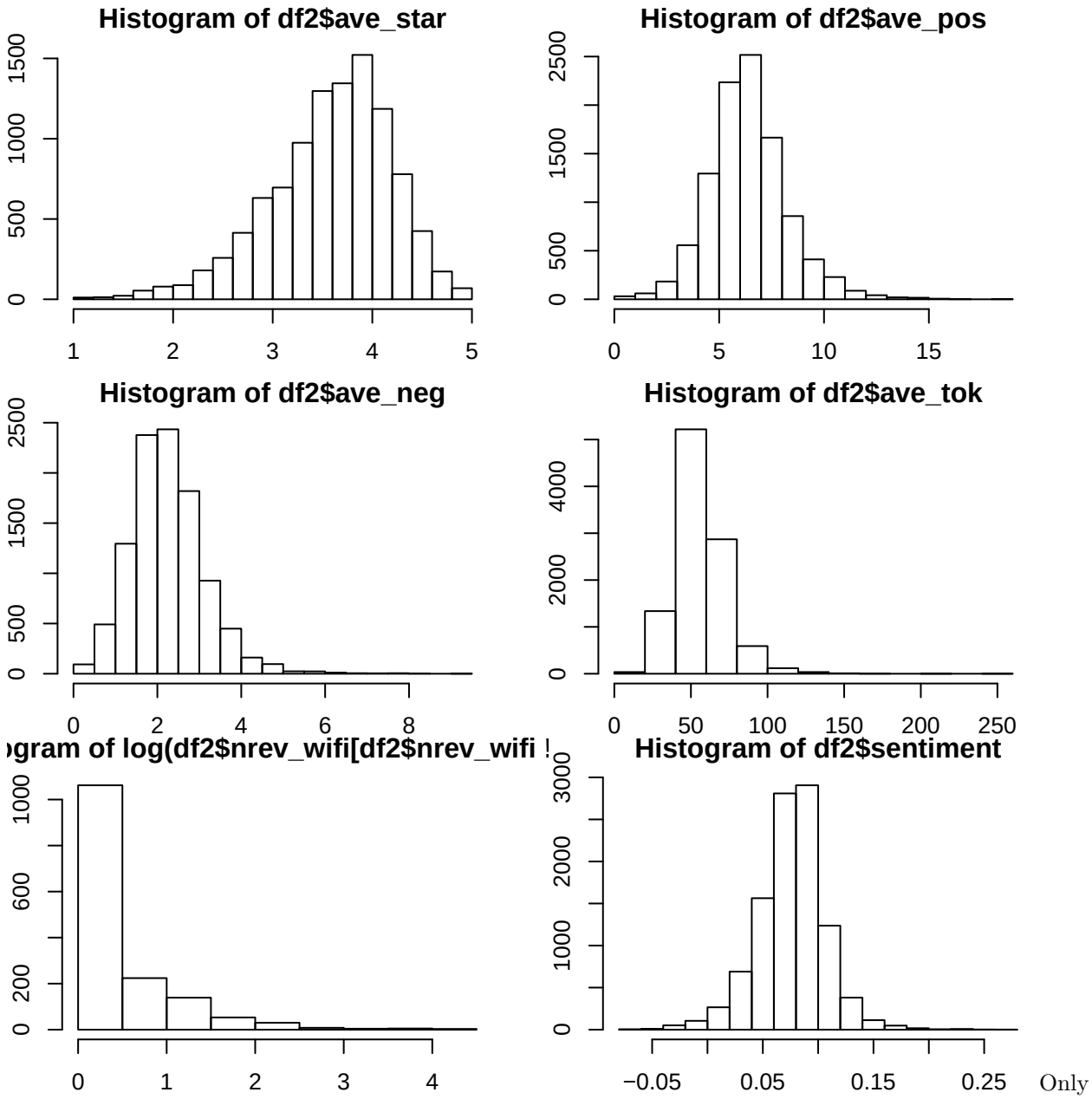


This tell us:

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



Only mild skewness were found and no transformation were applied. for number of reviews which strongly skewed to right, due to large number of 0s, even transformation will not help a lot. So we just dont do any transformation here.

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.

```
##
## Call:
```

```
## lm(formula = ave_star ~ ., data = df2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1317 -0.2183  0.0082  0.2161  2.2872
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 2.990e+00  5.136e-02  58.217  < 2e-16 ***
## cityLas Vegas               3.926e-02  1.573e-02   2.496 0.012582 *
## cityothers                  3.226e-02  1.406e-02   2.295 0.021727 *
## cityPhoenix                 3.620e-02  1.645e-02   2.201 0.027793 *
## cityPittsburgh              1.400e-02  2.082e-02   0.673 0.501251
## `Accepts Credit Cards`     -3.097e-01  2.498e-02 -12.400  < 2e-16 ***
## `Good For Groups`TRUE      -2.609e-02  1.408e-02  -1.853 0.063890 .
## `Outdoor Seating`TRUE      -1.877e-02  8.353e-03  -2.247 0.024632 *
## `Price Range`              -2.039e-02  9.542e-03  -2.137 0.032602 *
## `Good for Kids`TRUE        -5.225e-02  1.377e-02  -3.793 0.000149 ***
## Alcoholfull_bar            -8.469e-02  1.304e-02  -6.496 8.61e-11 ***
## Alcoholnone                -3.840e-02  1.237e-02  -3.104 0.001917 **
## `Noise Level`loud          -1.768e-02  1.414e-02  -1.250 0.211322
## `Noise Level`quiet          9.203e-03  1.033e-02   0.891 0.373141
## `Noise Level`very_loud     -3.862e-02  2.295e-02  -1.683 0.092436 .
## `Has TV`TRUE               -2.723e-02  8.812e-03  -3.090 0.002005 **
## Attiredressy                4.445e-02  2.761e-02   1.610 0.107454
## Attireformal                6.383e-02  1.339e-01   0.477 0.633666
## DeliveryTRUE                1.466e-02  1.165e-02   1.258 0.208267
## `Take-out`TRUE              2.941e-03  1.707e-02   0.172 0.863169
## `Takes Reservations`TRUE -1.625e-03  1.078e-02  -0.151 0.880175
## `Waiter Service`TRUE        2.624e-02  1.154e-02   2.274 0.022989 *
## wififree                   -2.490e-02  8.633e-03  -2.884 0.003930 **
## wifipaid                   -1.117e-01  3.696e-02  -3.021 0.002528 **
## Ambience.romanticTRUE       6.383e-02  3.250e-02   1.964 0.049545 *
## Ambience.intimateTRUE       9.299e-02  3.527e-02   2.636 0.008390 **
## Ambience.classyTRUE         4.727e-02  2.697e-02   1.753 0.079644 .
## Ambience.hipsterTRUE        1.169e-01  2.856e-02   4.094 4.28e-05 ***
## Ambience.diveyTRUE          1.907e-01  1.891e-02  10.086  < 2e-16 ***
## Ambience.trendyTRUE         3.385e-02  2.214e-02   1.529 0.126309
## Ambience.upscaleTRUE        8.379e-02  4.162e-02   2.013 0.044103 *
## Ambience.casualTRUE         5.007e-02  1.082e-02   4.629 3.73e-06 ***
## `Good For.dessert`TRUE      1.718e-02  2.709e-02   0.634 0.526003
## `Good For.latenight`TRUE    1.441e-02  1.685e-02   0.855 0.392470
## `Good For.lunch`TRUE       -1.151e-03  9.312e-03  -0.124 0.901596
## `Good For.dinner`TRUE      -7.328e-03  9.859e-03  -0.743 0.457317
## `Good For.breakfast`TRUE -4.236e-02  1.491e-02  -2.842 0.004496 **
## `Good For.brunch`TRUE      -3.792e-03  1.531e-02  -0.248 0.804404
## Parking.garageTRUE         -1.021e-01  1.604e-02  -6.365 2.04e-10 ***
## Parking.streetTRUE          1.100e-01  1.276e-02   8.620  < 2e-16 ***
## Parking.lotTRUE             4.465e-02  1.047e-02   4.264 2.03e-05 ***
## Parking.valetTRUE          -3.574e-02  2.142e-02  -1.668 0.095263 .
## nrev                        4.068e-04  2.826e-05  14.397  < 2e-16 ***
## ave_pos                     3.666e-02  6.010e-03   6.100 1.10e-09 ***
## ave_neg                    -2.374e-01  9.433e-03 -25.168  < 2e-16 ***
## ave_tok                     8.937e-03  5.405e-04  16.535  < 2e-16 ***
```

```
## nrev_wifi                -2.561e-03  2.178e-03  -1.176 0.239796
## sentiment                 9.729e+00  3.285e-01  29.619  < 2e-16 ***
## Food                      1.208e-01  1.440e-02   8.393  < 2e-16 ***
## Bars                     -1.329e-02  1.218e-02  -1.091 0.275181
## `American Traditional`   -6.664e-02  1.283e-02  -5.196 2.08e-07 ***
## Mexican                  -9.669e-03  1.323e-02  -0.731 0.464850
## Pizza                    -8.736e-03  1.484e-02  -0.589 0.556035
## `Fast Food`              -1.705e-01  2.001e-02  -8.519  < 2e-16 ***
## Sandwiches               -4.535e-03  1.444e-02  -0.314 0.753450
## `American New`           -2.035e-02  1.461e-02  -1.393 0.163547
## Italian                  -4.551e-02  1.584e-02  -2.874 0.004066 **
## Chinese                  -3.289e-03  1.635e-02  -0.201 0.840521
## Burgers                  -1.403e-02  1.579e-02  -0.889 0.374236
## `Breakfast & Brunch`     -6.375e-03  1.736e-02  -0.367 0.713397
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3754 on 10157 degrees of freedom
## Multiple R-squared:  0.6273, Adjusted R-squared:  0.6252
## F-statistic: 289.8 on 59 and 10157 DF,  p-value: < 2.2e-16
```

R square of over 62% were achieved and this is impressing in this case with all the variables I used. But the because we are regressing on the business which contain average data from different number of reviews. So let's compare to weighted lm model: weighted model

	Rsqr	sigma	vif.ave_pos	vif.sentiment	vif.ave_tok	vif.ave_neg	vif.attr
summarym	0.6251679	0.3753570	9.262988	7.493427	5.748540	5.006452	
	0.6656115	0.8582954	15.615383	10.652487	9.239138	6.559921	
The first r	ow is un-wei	ghted model	and 2nd row is	weighted model			
From here,	we know we n	eed to use w	eighted model	because weighted	model give us	much higher R	

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 see if this helps.

	Rsqr	sigma	vif.ave_pos	vif.sentiment	vif.ave_tok	vif.ave_neg	vif.attributes
summarym	0.6251679	0.3753570	9.262988	7.493427	5.748540	5.006452	3
	0.6656115	0.8582954	15.615383	10.652487	9.239138	6.559921	3
	0.6251679	0.6122354	9.262988	7.493427	5.748540	5.006452	3
	0.6656115	1.3999441	15.615383	10.652487	9.239138	6.559921	3
3rd row is	centered un-	weighted ful	l model, 4th r	ow is centered a	nd weighted fu	ll model.	

Above all,

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.

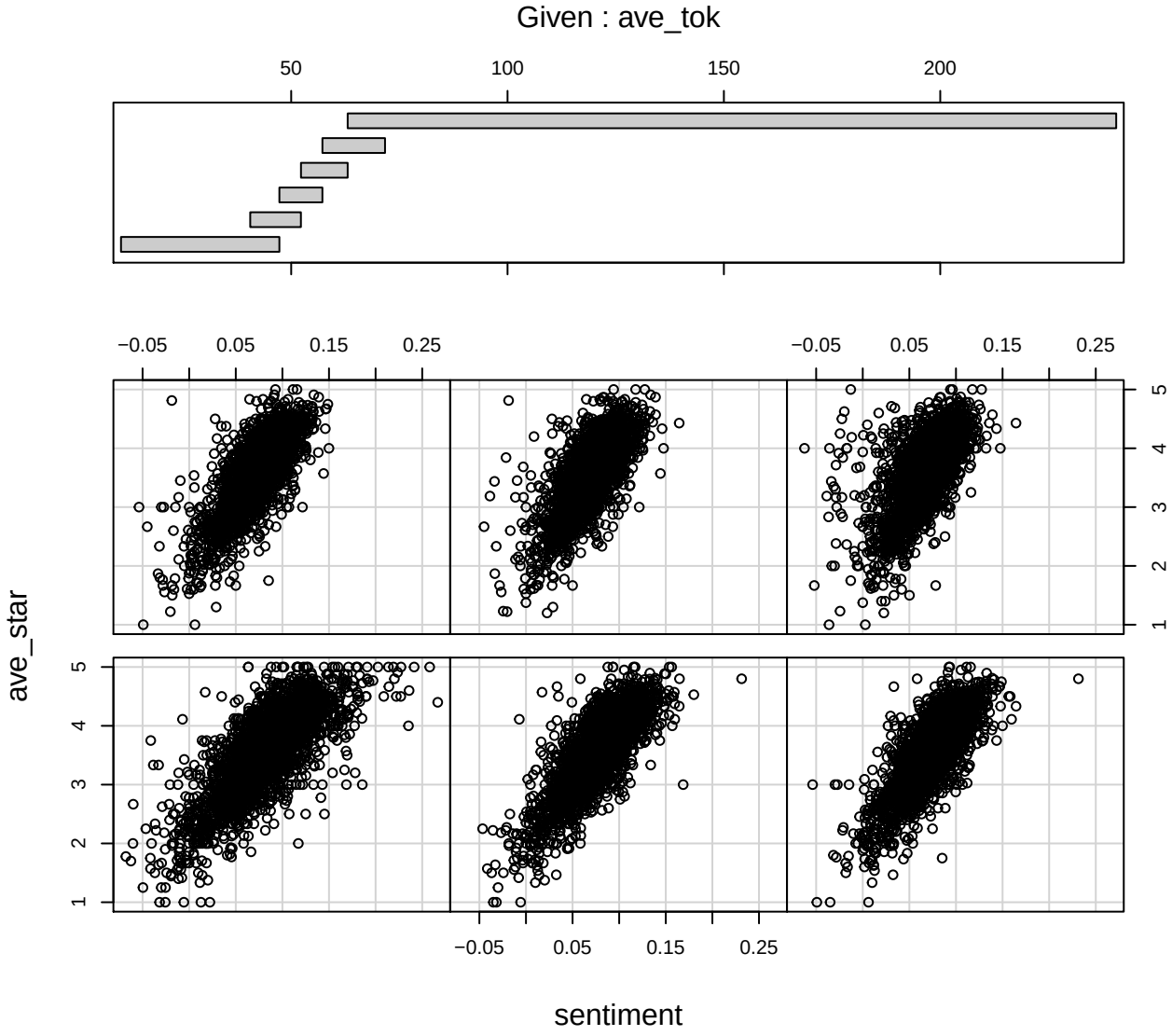
To reduce the vif, We can either remove variable "sentiment" or both "ave pos" and "ave neg". Let's try both

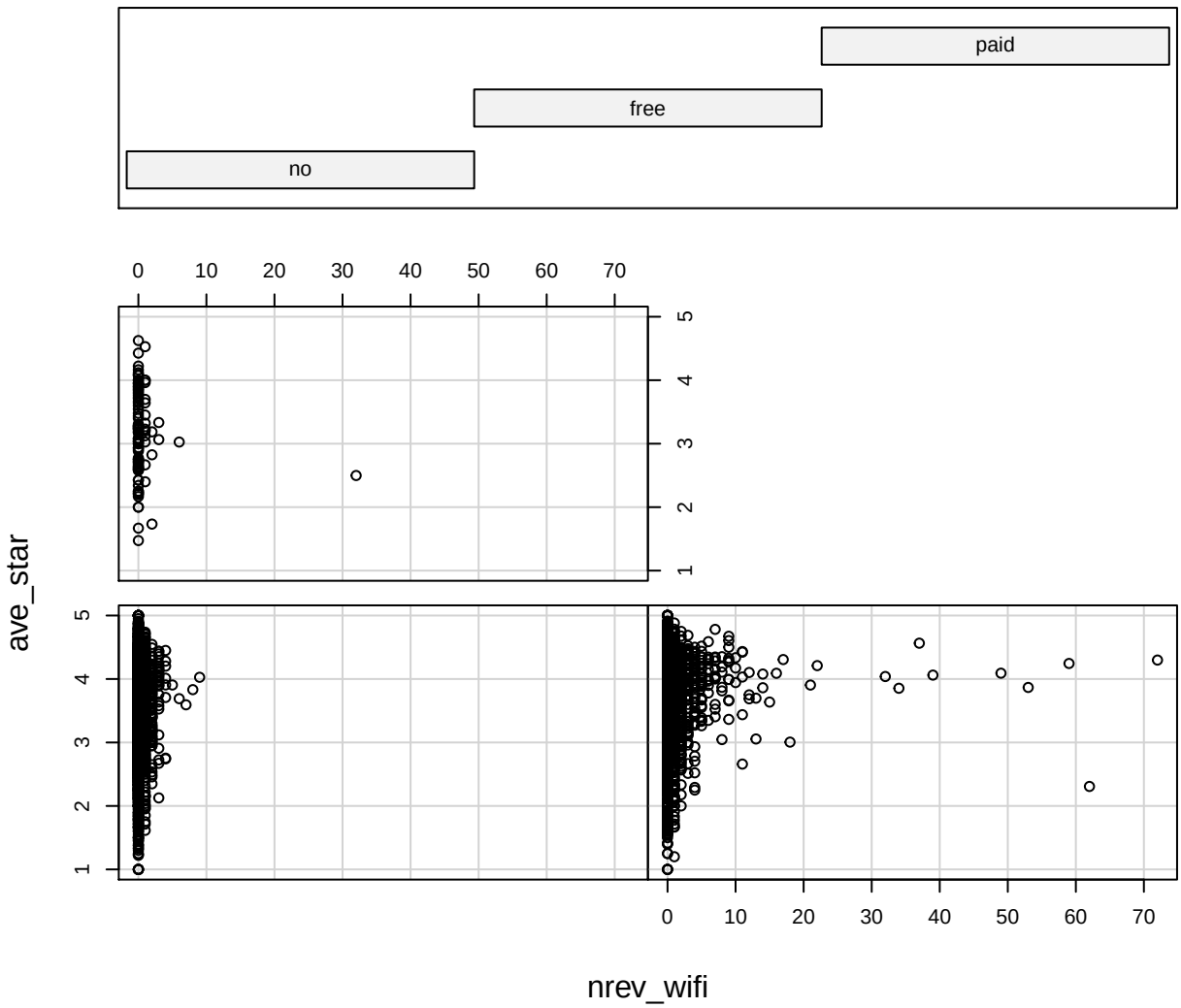
	Rsqr	sigma	vif.ave_pos	vif.sentiment	vif.ave_tok	vif.ave_neg	vif.attributes.Alcoh
original	0.6251679	0.3753570	9.262988	7.493427	5.748540	5.006452	3.3097
weighted	0.6656115	0.8582954	15.615383	10.652487	9.239138	6.559921	3.3195
centered	0.6251679	0.6122354	9.262988	7.493427	5.748540	5.006452	3.3097
centered_weighted	0.6656115	1.3999441	15.615383	10.652487	9.239138	6.559921	3.3195
-sentiment	0.6422074	0.8878238	5.583675	3.841143	3.318218	3.186557	2.7119
-pos and -neg	0.6439048	0.8857153	3.305844	3.186476	2.706598	2.296321	2.1917

From above comparison 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. We start off with checking all possible interactions and followed by checking interaction with only wifi variable.





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

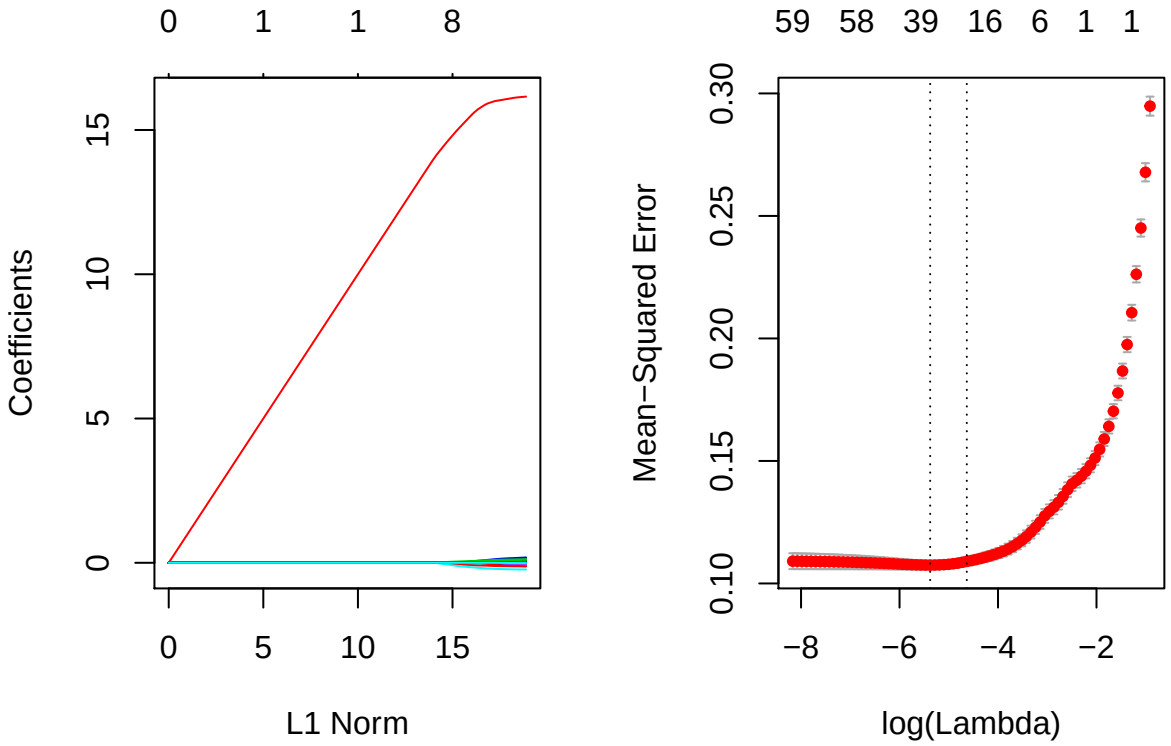
```
kable(summarym)
```

	Rsqr	sigma	vif	vif	vif	vif	vif	attributes.wififree
original	0.6251679	0.3753570	9.262988	7.493427	5.748540	5.006452	3.309733	-0.0249008
weighted	0.6656115	0.8582954	15.615383	10.652487	9.239138	6.559921	3.319564	-0.0233355
centered	0.6251679	0.6122354	9.262988	7.493427	5.748540	5.006452	3.309733	-0.0406150
centered_weighted	0.6656115	1.3999441	15.615383	10.652487	9.239138	6.559921	3.319564	-0.0380619
-sentiment	0.6422074	0.8878238	5.583675	3.841143	3.318218	3.186557	2.711912	-0.0248328
-pos and -neg	0.6439048	0.8857153	3.305844	3.186476	2.706598	2.296321	2.191786	-0.0191361
+interaction1	0.6473138	0.8814656	13.641470	10.169595	7.018457	3.308822	3.187737	-0.0169510
+interaction2	0.6441657	0.8853908	3.308806	3.186555	2.708015	2.301848	2.193567	-0.0154144

According to all the full models we have tried above, the last one give us most feasible results to proceed to variable selection. The last one is: * weighted * un-centered * without variable positive words number * without variable negative words number * with interaction between number of review mentioning 'wifi' and wifi By the way, outliers checked and no significant outliers found.

Model selection

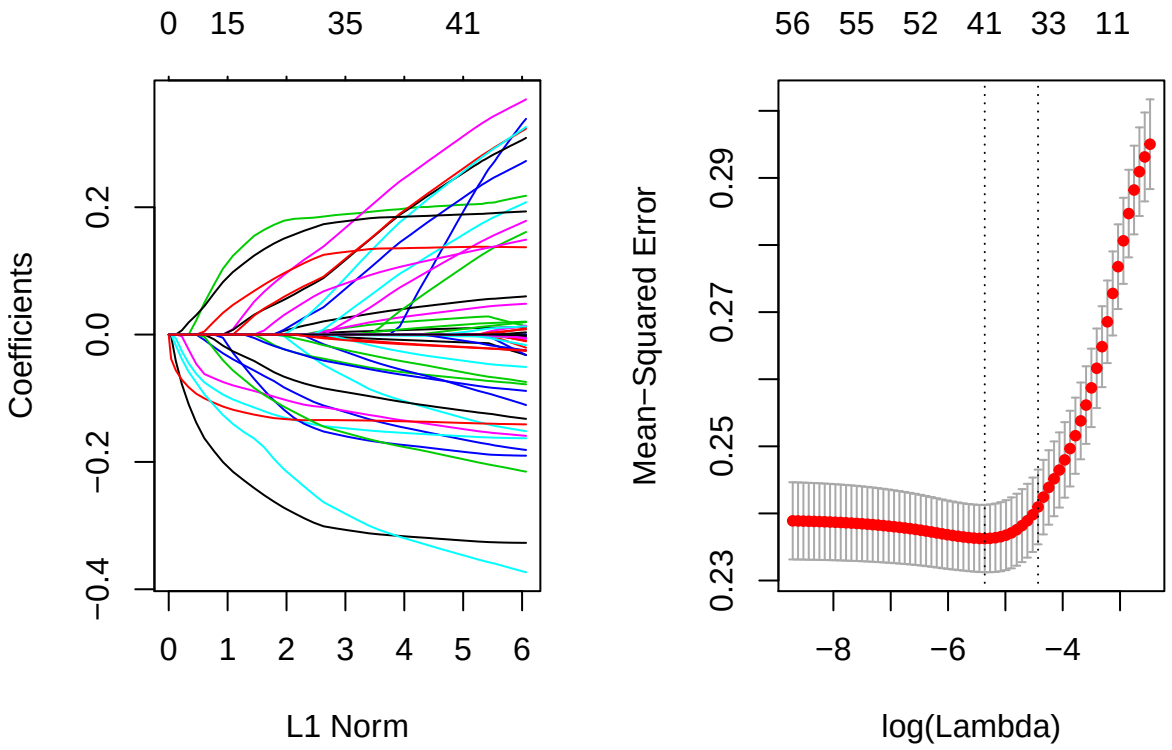
After we have our full model, several methods are used to choose best parsimonious model by lasso



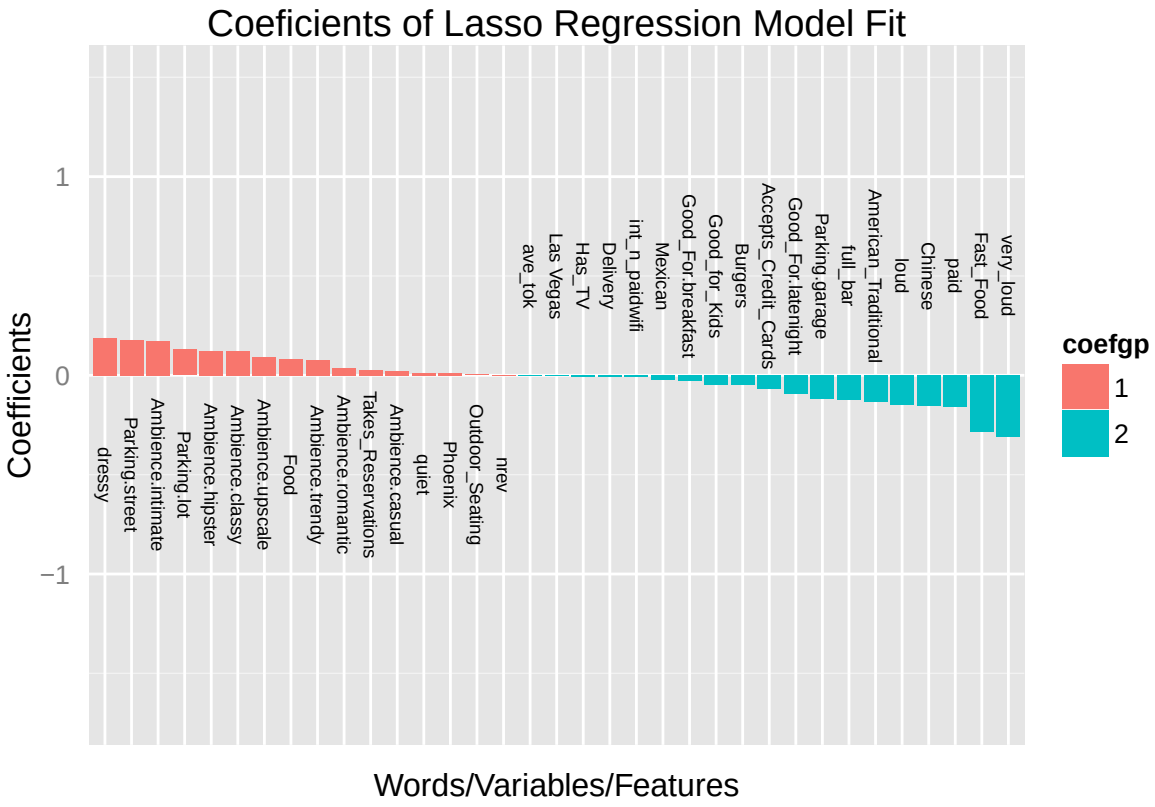

```
## 60 x 1 sparse Matrix of class "dgCMatrix"
##                                      1
## (Intercept)            2.3040059363
## Las Vegas              0.0008622253
## others                 .
## Phoenix                .
## Pittsburgh             .
## Accepts_Credit_Cards -0.1920699647
## Good_For_Groups      -0.0486372804
## Outdoor_Seating      -0.0222347462
## Price_Range            .
## Good_for_Kids          .
## beer_and_wine          .
## full_bar             -0.0803476773
## loud                   .
## quiet                  .
## very_loud              .
## Has_TV               -0.0254455898
## dressy                 0.0019795779
## formal                 .
```

```
## Delivery              .
## Take_out              .
## Takes_Reservations    .
## Waiter_Service        .
## free                  .
## paid                 -0.0537357499
## Ambience.romantic     .
## Ambience.intimate     .
## Ambience.classy       .
## Ambience.hipster      0.0383345204
## Ambience.divey        0.0920055671
## Ambience.trendy       .
## Ambience.upscale      .
## Ambience.casual       0.0034543005
## Good_For.dessert      .
## Good_For.latenight    .
## Good_For.lunch        .
## Good_For.dinner       .
## Good_For.breakfast    .
## Good_For.brunch       .
## Parking.garage       -0.0970832334
## Parking.street        0.0836376105
## Parking.lot           0.0192858709
## Parking.valet         .
## nrev                  0.0002132371
## ave_tok               0.0063824986
## sentiment            15.8862784813
## Food                  0.0349706985
## Bars                  .
## American_Traditional -0.0269583075
## Mexican               .
## Pizza                 .
## Fast_Food            -0.0330428282
## Sandwiches            .
## American_New          .
## Italian               .
## Chinese               .
## Burgers               .
## Breakfast_&_Brunch    .
## int_n_nowifi          0.0044532818
## int_n_freewifi        .
## int_n_paidwifi        .
```

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.




```
## Warning: Stacking not well defined when ymin != 0
```



Results 1:

In this data set, variable sentiment is too good to explain the ratings been given. And this sentiment is also related to all other variables to a certain extend. So all other coeficients become not so significant. If we exclude sentiment and just analyze the contributions from other attributes. R square dropped dramatically but there are still over 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.

Unfortunately, we can not say anything about wifi so far, because the coeficient is too close to zero and it changes from negative to positive between different models. Even it is significant in our last model, but the model only explains 20% variance in ratings and the data size is not small. So we cannot make any conclusion about wifi's impact on customers' experience. Interestingly, paid wifi always has negative coeficient and it is always more significant comparing to free wifi. So we can make conclusion that paid wifi is hurting people's feeling.

In next part, I looked into the data more deeply to find out why we saw negative wifi coeficients in previous analysis because I believe that free stuff never hurts.

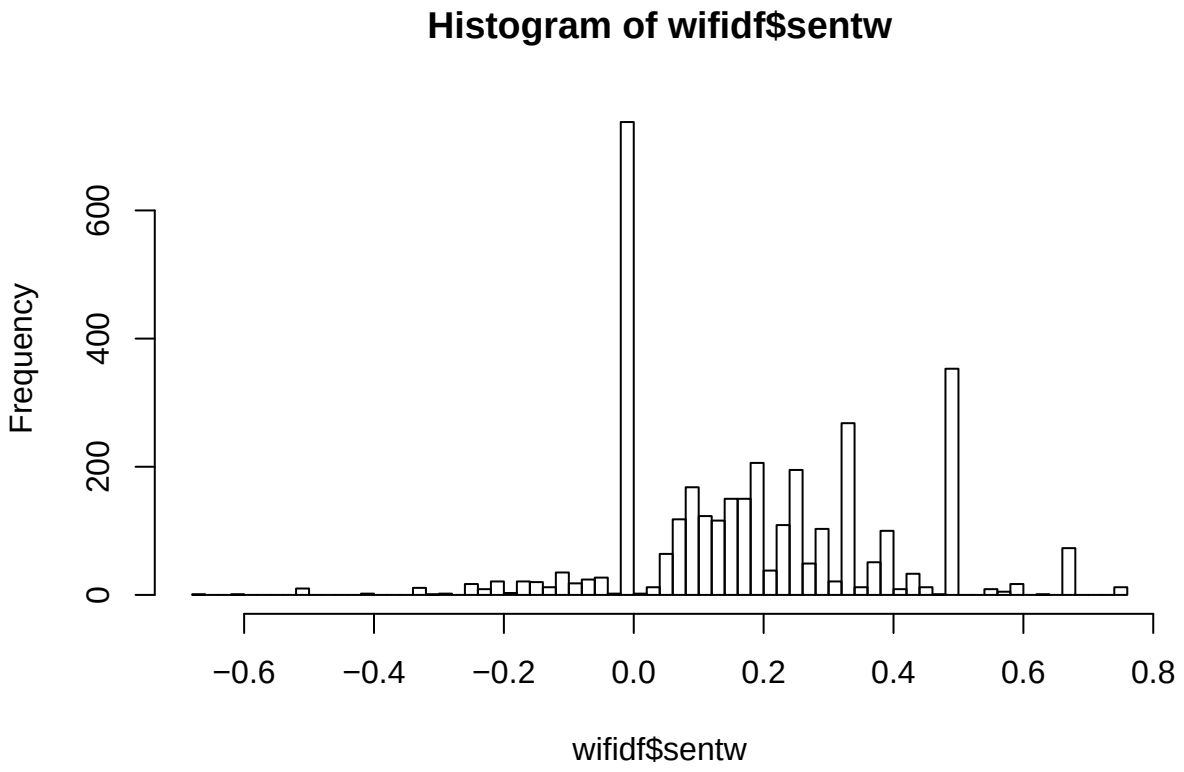
PartII, Why free wifi is hurting people?

In this part, I am trying to use small portion of the data whose review contains wifi or internet, that is using the data frame I made "wifidf", to do regression on average stars the business got. I believe people's words contains much more information than just a dummy variable of 0 and 1s. Also, by using this model, we are assuming we can separate people's feeling about wifi and all other things by exploring the review text. In this way, sentiment variable might not be associate with people's feeling about wifi anymore. Thus the coeficients here might be more reliable to make conclusions about if people like or hate wifi.

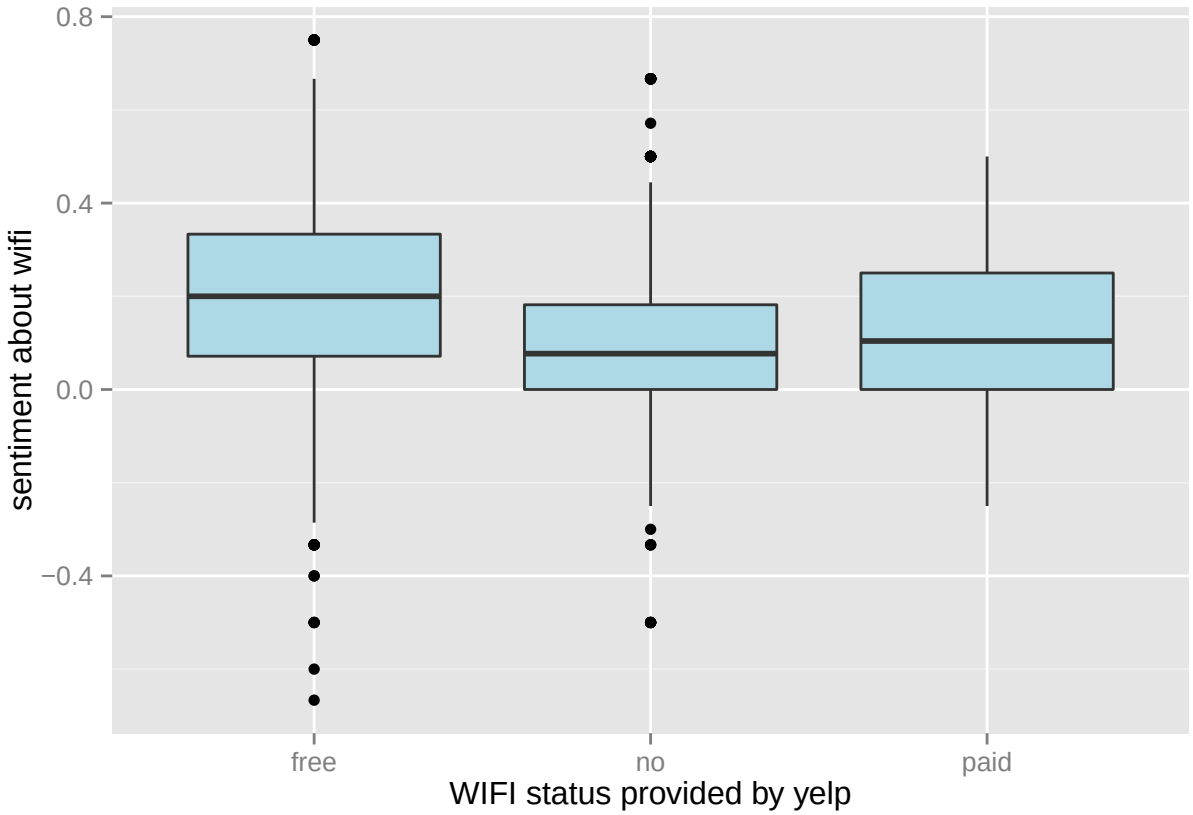
read in data frame

Text mining

Here we try to seperate people's words about wifi and all other stuff. And extract the sentiments.



Histgram told us most people are happy when they were talking about wifi



In this boxplot, we found some interesting things * why people are complaining when there is free wifi? * why people are happy when there is no wifi?

Let's take a look at what people say then

What people say in restaurants with "free wifi"?

```
##  [1] "+ wifi with no problem"
##  [2] "The wifi is pretty terrible"
##  [3] " Bad wi-fi mojo"
##  [4] " Be warned, if you are here to surf the web - internet here is very unstable"
##  [5] " The WiFi is very poor"
##  [6] "Wifi is somewhat slow, which can be an issue"
##  [7] " Unfortunately no Wifi"
##  [8] " Lose the internet jukebox"
##  [9] "The Internet SUCKS,   I've been here several times and the Internet is very intermittent"
## [10] "I hate to disagree with so many of my fellow yelpers, but Montesano's does not live up to the
## [11] "  The prices are not outrageous and they have wifi"
## [12] "The internet can be so deceiving"
## [13] " The internet is slower than dial up"
## [14] "Wifi is VERY slow"
## [15] "The only downside is that their wi-fi is really slow and unstable"
## [16] " WiFi very intermittent and weak"
```

What people say in restaurants with "no wifi"?

```
##  [1] " has free Wifi if you're interested"
##  [2] "An added bonus was the free wifi that is faster than Starbucks"
##  [3] "  Free wifi that's a plus"
##  [4] "Free wifi, yay"
##  [5] " They have free wifi, if you are wondering"
##  [6] "the wifi is awesome and fast as it should be"
##  [7] " Free wifi available too"
##  [8] "Excellent and fresh sushi, great service and free wifi"
##  [9] " Would be nice if they had free wifi"
## [10] "  The free wifi was nice too"
## [11] "  Free wifi is a plus"
## [12] " They have free wi-fi which is cool"
## [13] "Working wifi would be a plus"
## [14] " Free wifi a bonus"
## [15] " Great they have free wi-fi"
```

What about paid wifi? Let's check the word clouds

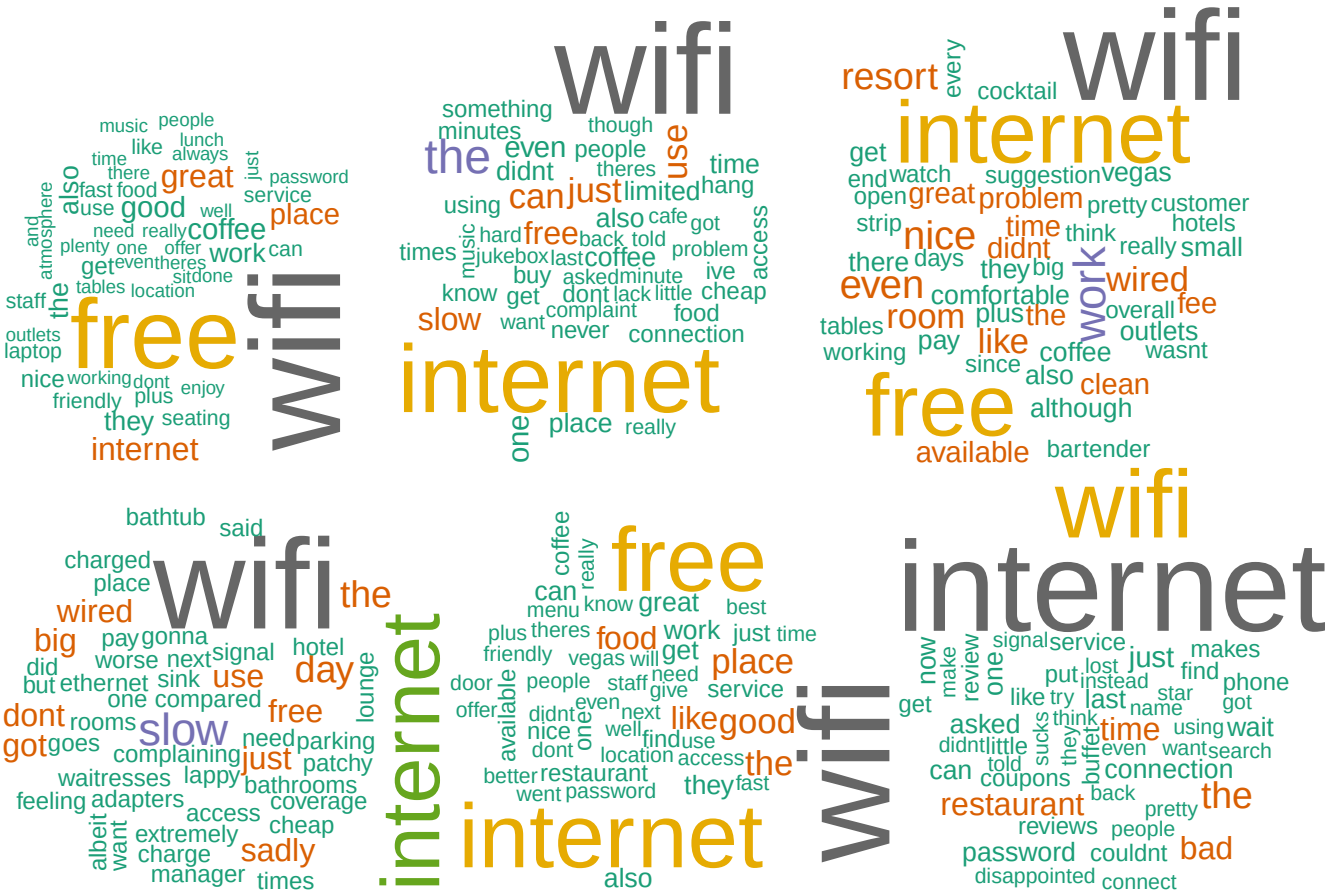
row1: good comments

row2: bad comments

column1: free wifi

column2: paid wifi

column3: no wifi



Now things become interesting and also annoying. We can understand that people would complaining the free wifi when the quality of wifi connection is poor. But why people are telling there were free wifi even the yelp's data told us there were no wifi? So we believe that the data about wifi of different business is outdated or wrong.

However, out of over 800000 reviews, only 3555 reviews mentioned wifi, so there is no way to check the ratio of mistakes in the wifi data. What we can do is just stay with this 3555 reviews and try regression on them to see if there is any conclusion we can reach to answer our primary question: "should restaurants' owner provide wifi to increasing customers' experience"

Our strategy is definitive, the original wifi variable is not accurate, and also, the connection quality matters a lot. So we will use sentiment of review about wifi as our target variable to see:

"When people are happy about the wifi, are they happy with the restaurants too?"

If yes, that means a good wifi can help improve customers' experience If no, that means wifi really dosen't matter

Modeling

Modeling on individual reviews

```
##
## Call:
## lm(formula = stars ~ ., data = wifidf6)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -3.4829 -0.5433  0.1122  0.6967  3.9973
##
## Coefficients: (15 not defined because of singularities)
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       3.3503684  0.2399179  13.965  < 2e-16
## cityLas Vegas                     0.1400661  0.0947660   1.478  0.13951
## cityothers                        0.0244350  0.0907916   0.269  0.78785
## cityPhoenix                       0.0094974  0.0986242   0.096  0.92329
## cityScottsdale                    0.0546494  0.1102499   0.496  0.62015
## `attributes.Accepts Credit Cards` -0.0573528  0.1291837  -0.444  0.65710
## `attributes.Good For Groups`TRUE  -0.0006469  0.0921882  -0.007  0.99440
## `attributes.Outdoor Seating`TRUE  -0.0365136  0.0502665  -0.726  0.46765
## `attributes.Price Range`          -0.0420064  0.0552318  -0.761  0.44699
## `attributes.Good for Kids`TRUE    -0.0137817  0.0717744  -0.192  0.84774
## attributes.Alcoholfull_bar        -0.0491332  0.0750469  -0.655  0.51271
## attributes.Alcoholnone            -0.2080487  0.0663552  -3.135  0.00173
## `attributes.Noise Level`loud      -0.0170142  0.0824163  -0.206  0.83646
## `attributes.Noise Level`quiet      0.0145900  0.0679168   0.215  0.82992
## `attributes.Noise Level`very_loud -0.2558613  0.1603413  -1.596  0.11065
## `attributes.Has TV`TRUE           -0.0838808  0.0508806  -1.649  0.09934
## attributes.Attiredressy            0.5060259  0.2067569   2.447  0.01444
## attributes.DeliveryTRUE            0.0526822  0.0742310   0.710  0.47794
## `attributes.Take-out`TRUE          0.0865010  0.1070540   0.808  0.41915
## `attributes.Takes Reservations`TRUE -0.0044201  0.0643276  -0.069  0.94522
## `attributes.Waiter Service`TRUE   -0.0586733  0.0628134  -0.934  0.35033
## attributes.wifino                 -0.1038499  0.0541632  -1.917  0.05529
## attributes.wifipaid               -0.3751040  0.1799417  -2.085  0.03719
## attributes.Ambience.romanticTRUE   0.3726286  0.2229579   1.671  0.09477
## attributes.Ambience.intimateTRUE   0.2452278  0.3591291   0.683  0.49476
## attributes.Ambience.classyTRUE     0.0339058  0.1739276   0.195  0.84545
## attributes.Ambience.hipsterTRUE    0.1425419  0.0892404   1.597  0.11031
## attributes.Ambience.touristyTRUE  -0.4530349  0.2335906  -1.939  0.05254
## attributes.Ambience.trendyTRUE     0.2153178  0.1176950   1.829  0.06743
## attributes.Ambience.upscaleTRUE    0.1165233  0.2966195   0.393  0.69447
## attributes.Ambience.casualTRUE     0.0908374  0.0646323   1.405  0.15999
## `attributes.Good For.dessert`TRUE  0.1244077  0.1169074   1.064  0.28734
## `attributes.Good For.latenight`TRUE -0.1956664  0.0857842  -2.281  0.02262
## `attributes.Good For.lunch`TRUE   -0.0573208  0.0582395  -0.984  0.32508
## `attributes.Good For.dinner`TRUE   0.0633532  0.0652764   0.971  0.33186
## `attributes.Good For.breakfast`TRUE -0.0232817  0.0640671  -0.363  0.71633
## `attributes.Good For.brunch`TRUE  -0.1290693  0.0663410  -1.946  0.05180
## attributes.Parking.garageTRUE     -0.0835398  0.0857096  -0.975  0.32979
## attributes.Parking.streetTRUE      0.0968487  0.0676378   1.432  0.15228
## attributes.Parking.validatedTRUE  -0.1355050  0.3101606  -0.437  0.66222
## attributes.Parking.lotTRUE         0.1680280  0.0668168   2.515  0.01196
## attributes.Parking.valetTRUE      -0.0747250  0.1092414  -0.684  0.49400
## sentw                              0.7853078  0.0959635   8.183 4.01e-16
## sentiment                          3.0720965  0.1645385  18.671  < 2e-16
## Barbeque                          -0.1607229  0.1995849  -0.805  0.42072
## `Fast Food`                        0.0782781  0.1269423   0.617  0.53752
## Mexican                           -0.1438120  0.1155971  -1.244  0.21357
## Bakeries                          -0.2405391  0.1298550  -1.852  0.06407
## Food                               0.1876900  0.1443482   1.300  0.19361
```

```
## Grocery                          0.4486198  0.5586480   0.803  0.42201
## Cafes                            0.1856821  0.0722625   2.570  0.01023
## `Coffee & Tea`                  -0.0052722  0.1323725  -0.040  0.96823
## `Breakfast & Brunch`             0.1815753  0.0645765   2.812  0.00496
## `Ice Cream & Frozen Yogurt`      0.0703449  0.3131474   0.225  0.82228
## Sandwiches                       0.0849794  0.0689913   1.232  0.21814
## Italian                         -0.1736122  0.1137066  -1.527  0.12690
## Japanese                        -0.0743754  0.1723573  -0.432  0.66612
## Bars                             0.2564609  0.4560918   0.562  0.57395
## `Wine Bars`                      0.0945736  0.1921098   0.492  0.62255
## `Tapas/Small Plates`            -0.4656401  0.4243609  -1.097  0.27261
## Nightlife                       -0.1215580  0.4611753  -0.264  0.79212
## Burgers                         -0.1160824  0.0973153  -1.193  0.23302
## `Sports Bars`                   -0.1446913  0.1502458  -0.963  0.33561
## `American (New)`                 0.0204803  0.0761693   0.269  0.78804
## `Chicken Wings`                 -0.1923592  0.1823883  -1.055  0.29166
## Vegetarian                       0.2952660  0.1220146   2.420  0.01558
## Mediterranean                    0.2625123  0.1674258   1.568  0.11700
## Pizza                           -0.0004880  0.1103794  -0.004  0.99647
## Vegan                           -0.1005194  0.1461112  -0.688  0.49153
## Korean                           0.3341588  0.2882093   1.159  0.24637
## Desserts                        -0.0135406  0.2254468  -0.060  0.95211
## Donuts                          -0.0004915  0.5038851  -0.001  0.99922
## Russian                          0.5944462  0.7263297   0.818  0.41318
## `Music Venues`                  -0.0461824  0.9303197  -0.050  0.96041
## `Arts & Entertainment`           0.1008587  0.8495839   0.119  0.90551
## Gastropubs                       0.1057746  0.1799574   0.588  0.55673
## Scandinavian                    -0.0357485  0.3006509  -0.119  0.90536
## `American (Traditional)`        -0.1070007  0.0818177  -1.308  0.19104
## `Middle Eastern`                 0.2947866  0.2703410   1.090  0.27561
## Casinos                         -0.5223241  0.8248074  -0.633  0.52661
## `Event Planning & Services`      0.0686915  0.5982656   0.115  0.90860
## `Hotels & Travel`                0.5495692  1.0239694   0.537  0.59151
## Hotels                          -1.0204579  1.2161732  -0.839  0.40149
## `Tea Rooms`                      0.1826055  0.1858377   0.983  0.32588
## Diners                          -0.0856993  0.1702561  -0.503  0.61475
## Vietnamese                      -0.0938453  0.2121012  -0.442  0.65819
## Soup                            -0.2890018  0.1874664  -1.542  0.12327
## Salad                            0.2331102  0.1686256   1.382  0.16695
## French                           0.1413424  0.1873452   0.754  0.45064
## Chinese                          0.2832227  0.1451421   1.951  0.05111
## British                          0.2076090  0.2332721   0.890  0.37354
## Greek                           -0.3211623  0.2199058  -1.460  0.14427
## Pubs                             0.1328705  0.1398411   0.950  0.34211
## Bagels                          -0.1207813  0.2022542  -0.597  0.55043
## Cinema                          -0.2512067  0.4046395  -0.621  0.53477
## Filipino                         0.0241308  0.7229139   0.033  0.97337
## `Juice Bars & Smoothies`        -0.0804660  0.2189563  -0.367  0.71327
## `Gluten-Free`                    0.0979368  0.1859736   0.527  0.59850
## `Irish Pub`                      0.3479730  0.5290758   0.658  0.51078
## `Food Trucks`                   -0.2293215  0.3595914  -0.638  0.52370
## German                           0.4627673  0.5303071   0.873  0.38293
## Brasseries                       0.1829332  0.2991945   0.611  0.54097
## `Hot Pot`                        1.0176426  0.7380975   1.379  0.16808
```

```
## `Sushi Bars`                0.0544154  0.1907189   0.285  0.77542
## Creperies                  -0.0089413  0.1718451  -0.052  0.95851
## Gelato                     -0.2505426  0.2717751  -0.922  0.35667
## `Asian Fusion`              0.2382515  0.1862333   1.279  0.20088
## Lounges                     0.0611248  0.2167274   0.282  0.77794
## Delis                       0.0798696  0.1344170   0.594  0.55243
## Spanish                    -1.1379418  1.1728815  -0.970  0.33202
## Steakhouses                 0.0288036  0.1713857   0.168  0.86654
## Basque                     -0.8035910  1.5567772  -0.516  0.60576
## Thai                       -0.0843094  0.1657530  -0.509  0.61104
## Portuguese                  0.0006199  0.7248202   0.001  0.99932
## `Active Life`               0.7909316  1.1114932   0.712  0.47677
## `Kids Activities`           0.0544715  1.2214732   0.045  0.96443
## Laotian                            NA         NA      NA       NA
## `Venues & Event Spaces`    -0.5396840  0.7668661  -0.704  0.48164
## Ramen                       0.3727050  0.3578410   1.042  0.29771
## Hawaiian                    0.0195293  0.3008151   0.065  0.94824
## `Tex-Mex`                  -0.4871251  0.3042576  -1.601  0.10947
## Seafood                    -0.0742192  0.1517769  -0.489  0.62488
## Taiwanese                  -0.1258989  0.2543363  -0.495  0.62063
## `Dance Clubs`              -1.1441085  0.7716010  -1.483  0.13824
## Buffets                    -0.0128281  0.1902635  -0.067  0.94625
## `Beer, Wine & Spirits`     -0.0633080  0.2738098  -0.231  0.81717
## Poutineries                -0.2163823  0.4534396  -0.477  0.63325
## Irish                       0.1467717  0.2472222   0.594  0.55277
## `Ethnic Food`               0.6855747  0.6928191   0.990  0.32248
## `Specialty Food`           -0.5467305  0.5483121  -0.997  0.31879
## `Comfort Food`             -0.2885728  0.3158314  -0.914  0.36095
## `Latin American`           -0.0035702  0.4627715  -0.008  0.99385
## `Food Court`                1.3799012  1.0176071   1.356  0.17519
## `Internet Cafes`            0.2593172  0.2822740   0.919  0.35834
## `Cocktail Bars`             0.2154385  0.3454424   0.624  0.53290
## Cheesesteaks               -0.2100379  0.7453723  -0.282  0.77812
## `Hot Dogs`                  0.6651808  0.3744962   1.776  0.07580
## Indian                     -0.0906315  0.4285901  -0.211  0.83254
## Breweries                   0.1012991  0.2744069   0.369  0.71204
## `Home Services`             1.1812638  1.1356907   1.040  0.29836
## `Shared Office Spaces`             NA         NA      NA       NA
## `Real Estate`                      NA         NA      NA       NA
## `Adult Entertainment`              NA         NA      NA       NA
## `Canadian (New)`            0.2226446  0.2518127   0.884  0.37668
## Belgian                     0.6589604  0.7385175   0.892  0.37232
## `Cajun/Creole`              0.8869923  0.4087824   2.170  0.03010
## `Modern European`           0.1317560  0.4307584   0.306  0.75972
## `Shaved Ice`                0.0342958  0.5157583   0.066  0.94699
## Shopping                    0.9764845  1.0155438   0.962  0.33636
## Drugstores                 -1.4265053  1.2315761  -1.158  0.24684
## `Hookah Bars`               0.1240181  0.4123827   0.301  0.76364
## `Swimming Pools`                   NA         NA      NA       NA
## `Bubble Tea`                0.2282762  0.5606937   0.407  0.68394
## Szechuan                   -0.3099278  0.4636222  -0.668  0.50387
## `Fish & Chips`              0.0592486  0.5152553   0.115  0.90846
## `Dive Bars`                 0.4136212  0.7448618   0.555  0.57873
## Golf                        0.7907914  1.5068151   0.525  0.59975
```

```
## Resorts                          -1.9182944  1.1660002  -1.645  0.10003
## Caribbean                        -0.2542240  0.4462845  -0.570  0.56896
## Brazilian                         0.1475691  0.4719195   0.313  0.75453
## `Pool Halls`                             NA         NA      NA       NA
## `Jazz & Blues`                    0.4407509  1.4106807   0.312  0.75473
## `Persian/Iranian`                 0.7547911  0.7325201   1.030  0.30290
## Halal                            -0.2618130  1.0400959  -0.252  0.80127
## Tours                                    NA         NA      NA       NA
## Peruvian                          0.5505302  1.0152443   0.542  0.58768
## `Champagne Bars`                 -0.8505829  0.7763447  -1.096  0.27333
## Southern                         -0.6181028  0.3475143  -1.779  0.07540
## `Seafood Markets`                 1.4636878  1.2795921   1.144  0.25277
## Arcades                           0.0609518  0.9851133   0.062  0.95067
## Caterers                                 NA         NA      NA       NA
## `Farmers Market`                  0.3656957  0.7463137   0.490  0.62417
## `Dim Sum`                         0.6492008  0.5565376   1.166  0.24350
## Cantonese                        -0.6395172  0.7717034  -0.829  0.40733
## Turkish                           0.7181074  0.5873235   1.223  0.22155
## `Live/Raw Food`                   0.3592842  0.7407740   0.485  0.62770
## Afghan                            0.0291594  0.6175050   0.047  0.96234
## `Shopping Centers`                0.2876900  1.4440740   0.199  0.84210
## Pakistani                         0.4776342  1.1021791   0.433  0.66479
## Kosher                            0.3496520  0.5387283   0.649  0.51637
## Wineries                          0.2882879  1.1575183   0.249  0.80333
## Cuban                             0.4852653  1.0129915   0.479  0.63194
## `Chocolatiers & Shops`            1.2260052  0.8173848   1.500  0.13374
## `Food Delivery Services`          0.8038906  1.0279400   0.782  0.43425
## Cafeteria                         1.2532383  1.0546420   1.188  0.23481
## `Tapas Bars`                      0.7508405  0.4053074   1.853  0.06405
## Malaysian                         1.0684755  1.0147078   1.053  0.29243
## `Candy Stores`                   -1.2591647  0.7665306  -1.643  0.10055
## `Arts & Crafts`                          NA         NA      NA       NA
## `RV Parks`                        1.5807010  0.6471656   2.442  0.01464
## Fondue                           -2.5518773  1.0389389  -2.456  0.01410
## Butcher                           0.4717959  1.1995673   0.393  0.69412
## `Street Vendors`                  0.7725529  1.1374614   0.679  0.49707
## `Car Wash`                        0.1764729  0.5556714   0.318  0.75082
## Automotive                               NA         NA      NA       NA
## Scottish                                 NA         NA      NA       NA
## Ethiopian                         0.4557038  1.0192531   0.447  0.65484
## `Soul Food`                      -0.4885638  0.7397470  -0.660  0.50902
## Cambodian                         0.7372957  1.0406033   0.709  0.47867
## Bowling                                  NA         NA      NA       NA
## Indonesian                               NA         NA      NA       NA
## Bistros                                  NA         NA      NA       NA
## `Beer Bar`                       -0.9025884  1.0669798  -0.846  0.39766
## `Food Stands`                    -1.4054630  1.0210984  -1.376  0.16879
## Mongolian                         0.2335649  1.0147864   0.230  0.81798
## Salvadoran                               NA         NA      NA       NA
## `Cultural Center`                -0.5420724  1.3395374  -0.405  0.68575
## Karaoke                           0.2630278  1.1436772   0.230  0.81812
##
## (Intercept)                      ***
## cityLas Vegas
```

```
## cityothers
## cityPhoenix
## cityScottsdale
## `attributes.Accepts Credit Cards`
## `attributes.Good For Groups`TRUE
## `attributes.Outdoor Seating`TRUE
## `attributes.Price Range`
## `attributes.Good for Kids`TRUE
## attributes.Alcoholfull_bar
## attributes.Alcoholnone              **
## `attributes.Noise Level`loud
## `attributes.Noise Level`quiet
## `attributes.Noise Level`very_loud
## `attributes.Has TV`TRUE             .
## attributes.Attiredressy             *
## attributes.DeliveryTRUE
## `attributes.Take-out`TRUE
## `attributes.Takes Reservations`TRUE
## `attributes.Waiter Service`TRUE
## attributes.wifino                   .
## attributes.wifipaid                 *
## attributes.Ambience.romanticTRUE    .
## attributes.Ambience.intimateTRUE
## attributes.Ambience.classyTRUE
## attributes.Ambience.hipsterTRUE
## attributes.Ambience.touristyTRUE    .
## attributes.Ambience.trendyTRUE      .
## attributes.Ambience.upscaleTRUE
## attributes.Ambience.casualTRUE
## `attributes.Good For.dessert`TRUE
## `attributes.Good For.latenight`TRUE *
## `attributes.Good For.lunch`TRUE
## `attributes.Good For.dinner`TRUE
## `attributes.Good For.breakfast`TRUE
## `attributes.Good For.brunch`TRUE    .
## attributes.Parking.garageTRUE
## attributes.Parking.streetTRUE
## attributes.Parking.validatedTRUE
## attributes.Parking.lotTRUE          *
## attributes.Parking.valetTRUE
## sentw                               ***
## sentiment                           ***
## Barbeque
## `Fast Food`
## Mexican
## Bakeries                            .
## Food
## Grocery
## Cafes                               *
## `Coffee & Tea`
## `Breakfast & Brunch`                **
## `Ice Cream & Frozen Yogurt`
## Sandwiches
## Italian
```

```
## Japanese
## Bars
## `Wine Bars`
## `Tapas/Small Plates`
## Nightlife
## Burgers
## `Sports Bars`
## `American (New)`
## `Chicken Wings`
## Vegetarian                          *
## Mediterranean
## Pizza
## Vegan
## Korean
## Desserts
## Donuts
## Russian
## `Music Venues`
## `Arts & Entertainment`
## Gastropubs
## Scandinavian
## `American (Traditional)`
## `Middle Eastern`
## Casinos
## `Event Planning & Services`
## `Hotels & Travel`
## Hotels
## `Tea Rooms`
## Diners
## Vietnamese
## Soup
## Salad
## French
## Chinese                             .
## British
## Greek
## Pubs
## Bagels
## Cinema
## Filipino
## `Juice Bars & Smoothies`
## `Gluten-Free`
## `Irish Pub`
## `Food Trucks`
## German
## Brasseries
## `Hot Pot`
## `Sushi Bars`
## Creperies
## Gelato
## `Asian Fusion`
## Lounges
## Delis
## Spanish
```

```
## Steakhouses
## Basque
## Thai
## Portuguese
## `Active Life`
## `Kids Activities`
## Laotian
## `Venues & Event Spaces`
## Ramen
## Hawaiian
## `Tex-Mex`
## Seafood
## Taiwanese
## `Dance Clubs`
## Buffets
## `Beer, Wine & Spirits`
## Poutineries
## Irish
## `Ethnic Food`
## `Specialty Food`
## `Comfort Food`
## `Latin American`
## `Food Court`
## `Internet Cafes`
## `Cocktail Bars`
## Cheesesteaks
## `Hot Dogs`                          .
## Indian
## Breweries
## `Home Services`
## `Shared Office Spaces`
## `Real Estate`
## `Adult Entertainment`
## `Canadian (New)`
## Belgian
## `Cajun/Creole`                      *
## `Modern European`
## `Shaved Ice`
## Shopping
## Drugstores
## `Hookah Bars`
## `Swimming Pools`
## `Bubble Tea`
## Szechuan
## `Fish & Chips`
## `Dive Bars`
## Golf
## Resorts
## Caribbean
## Brazilian
## `Pool Halls`
## `Jazz & Blues`
## `Persian/Iranian`
## Halal
```

```
## Tours
## Peruvian
## `Champagne Bars`
## Southern                          .
## `Seafood Markets`
## Arcades
## Caterers
## `Farmers Market`
## `Dim Sum`
## Cantonese
## Turkish
## `Live/Raw Food`
## Afghan
## `Shopping Centers`
## Pakistani
## Kosher
## Wineries
## Cuban
## `Chocolatiers & Shops`
## `Food Delivery Services`
## Cafeteria
## `Tapas Bars`                      .
## Malaysian
## `Candy Stores`
## `Arts & Crafts`
## `RV Parks`                        *
## Fondue                            *
## Butcher
## `Street Vendors`
## `Car Wash`
## Automotive
## Scottish
## Ethiopian
## `Soul Food`
## Cambodian
## Bowling
## Indonesian
## Bistros
## `Beer Bar`
## `Food Stands`
## Mongolian
## Salvadoran
## `Cultural Center`
## Karaoke
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.006 on 3045 degrees of freedom
## Multiple R-squared:  0.287,  Adjusted R-squared:  0.2421
## F-statistic: 6.385 on 192 and 3045 DF,  p-value: < 2.2e-16

##           2.5 %    97.5 %
## sentw 0.6141534 0.9754185
```

Modeling by weighted regression on only business

```
##
## Call:
## lm(formula = ave_star ~ ave_sent + ave_sentw + Food + Casinos +
##     `RV Parks` + Vegetarian + `Tapas Bars` + Chinese + `Candy Stores` +
##     `Cajun/Creole` + attributes.Alcohol + attributes.Ambience.touristy +
##     Basque + `attributes.Good for Kids` + `Tex-Mex` + Fondue +
##     `attributes.Has TV` + attributes.Parking.lot + Bakeries +
##     Cafes + Southern + `Persian/Iranian` + `Hot Dogs` + `American (Traditional)` +
##     `Event Planning & Services` + attributes.Parking.street +
##     `Asian Fusion` + `Food Stands` + attributes.Attire + Pubs +
##     `Hot Pot` + `Food Court` + `Tapas/Small Plates` + attributes.Ambience.romantic +
##     Grocery + Italian + Mexican, data = wifidf8, weights = sqrt(nrev))
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5825 -0.5816  0.1241  0.6863  3.8951
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       3.22392    0.10622  30.351  < 2e-16 ***
## ave_sent                          4.48640    0.26718  16.792  < 2e-16 ***
## ave_sentw                         1.00577    0.13629   7.380  2.6e-13 ***
## Food                              0.23759    0.06387   3.720 0.000207 ***
## Casinos                          -0.66922    0.23022  -2.907 0.003704 **
## `RV Parks`                        2.82664    0.81959   3.449 0.000578 ***
## Vegetarian                        0.31992    0.12885   2.483 0.013143 *
## `Tapas Bars`                      0.86073    0.39880   2.158 0.031062 *
## Chinese                           0.25262    0.12912   1.956 0.050597 .
## `Candy Stores`                   -1.64934    0.67513  -2.443 0.014680 *
## `Cajun/Creole`                    1.05859    0.39162   2.703 0.006946 **
## attributes.Alcoholfull_bar       -0.08897    0.07237  -1.229 0.219141
## attributes.Alcoholnone           -0.19845    0.06815  -2.912 0.003646 **
## attributes.Ambience.touristyTRUE -0.44273    0.23929  -1.850 0.064475 .
## Basque                           -2.11764    1.00439  -2.108 0.035162 *
## `attributes.Good for Kids`TRUE   -0.07080    0.06981  -1.014 0.310600
## `Tex-Mex`                        -0.54605    0.30693  -1.779 0.075428 .
## Fondue                           -2.46976    1.02605  -2.407 0.016201 *
## `attributes.Has TV`TRUE          -0.10155    0.05263  -1.929 0.053863 .
## attributes.Parking.lotTRUE        0.20977    0.05791   3.622 0.000302 ***
## Bakeries                         -0.25272    0.10727  -2.356 0.018604 *
## Cafes                             0.16720    0.07477   2.236 0.025490 *
## Southern                         -0.74289    0.34499  -2.153 0.031450 *
## `Persian/Iranian`                 1.33090    0.71167   1.870 0.061662 .
## `Hot Dogs`                        0.70734    0.37076   1.908 0.056605 .
## `American (Traditional)`         -0.16260    0.07513  -2.164 0.030603 *
## `Event Planning & Services`      -0.43129    0.22105  -1.951 0.051233 .
## attributes.Parking.streetTRUE     0.12282    0.06563   1.871 0.061478 .
## `Asian Fusion`                    0.29766    0.17254   1.725 0.084695 .
## `Food Stands`                    -1.55509    1.00466  -1.548 0.121860
## attributes.Attiredressy           0.27862    0.16584   1.680 0.093148 .
## Pubs                              0.21026    0.13047   1.611 0.107285
## `Hot Pot`                         1.18732    0.84442   1.406 0.159905
```

```
## `Food Court`                        1.44510    1.00505   1.438 0.150685
## `Tapas/Small Plates`               -0.63242    0.39874  -1.586 0.112935
## attributes.Ambience.romanticTRUE  0.32620    0.21717   1.502 0.133291
## Grocery                            0.56520    0.39961   1.414 0.157452
## Italian                           -0.16576    0.10905  -1.520 0.128712
## Mexican                           -0.14978    0.10650  -1.406 0.159805
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.001 on 1513 degrees of freedom
## Multiple R-squared:  0.2889, Adjusted R-squared:  0.271
## F-statistic: 16.17 on 38 and 1513 DF,  p-value: < 2.2e-16

##               2.5 %   97.5 %
## ave_sentw 0.7384369 1.273102
```

Conclusion 2:

Wifi do help improve rating on different restaurants. But it really depends on the quality of the wifi. In this part, regression are applied and sentiment of comment about wifi are used as our indicator variable. The coeficient of this variable stay well above 0 and has very low p value. Which simply means the happier the people about wifi, the higher rating they tend to give to the restaurants.

However, this analysis could be biased, because we only use the 3555 reviews containing “wifi” or “internet”. Which means the reviewers might value more on wifi access than general public. Their might be a way to assess this bias by using user data set provieded by yelp. Also it would be interensting to check distribution of type of restaurants in the 3555 reviews comparing to all business we used before. This might shed some light on what kind of restaurants might need wifi more than other.

Above all, we have following conclusion The positive affects on ratings are as followed Good quality free wifi > no wifi > paid wifi or low quality wifi So, if you are going to have restaurant, provide only free and good quality wifi or provide nothing!

Thanks for reading.