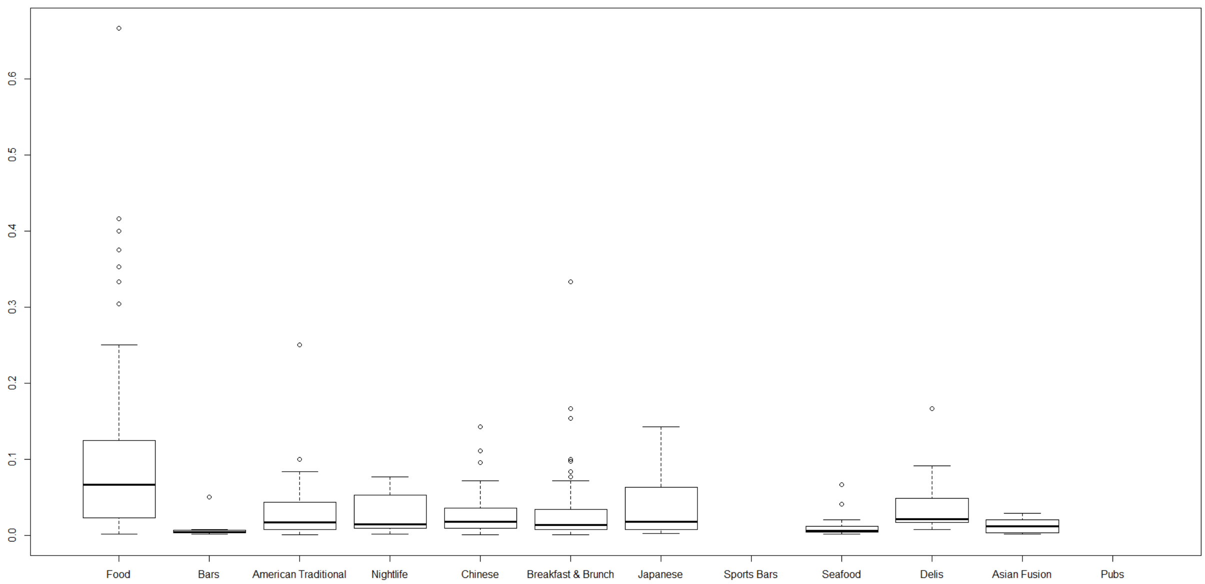
Part III

1. Choosing restaurant.

After we find our final model, we all interested in when people in which kind of restaurant, they are more care about the Wi-Fi. So, we use the N-way ANOVA to calculate the difference of the mean of the number of the commons which refer to wifi. But, we realized for some kinds of the restaurant, which only has one or two commons but refers to wifi, the response of this restaurant is one or two; meanwhile, if for some restaurant which has thousands of commons the number of the commons referring to wifi is also one or two, then this restaurant’s response is the same as previous one, which is obviously unfair. So, we decided to use the number of reviews of wifi over the totally number of reviews as our response.



The graph above shows the boxplot of our response to each kind of restaurant. As we can see, Food has a very high difference from other restaurant. But, we cannot choose some specific kinds.

Then, we chose N-way ANOVA to analyzed the data.

```
> summary(aov)
                                               Df Sum Sq Mean Sq F value  Pr(>F)    
newsub0$Food                                    1  0.515  0.5153 125.732 < 2e-16 ***
newsub0$Bars                                    1  0.012  0.0117   2.852 0.09158 .  
newsub0$`American Traditional`                  1  0.010  0.0098   2.382 0.12307    
newsub0$Nightlife                               1  0.034  0.0342   8.335 0.00398 ** 
newsub0$`Breakfast & Brunch`                    1  0.032  0.0318   7.765 0.00544 ** 
newsub0$Japanese                                1  0.000  0.0000   0.000 0.99426    
newsub0$`Sports Bars`                           1  0.001  0.0012   0.282 0.59545    
newsub0$Seafood                                 1  0.028  0.0279   6.809 0.00922 ** 
newsub0$Delis                                   1  0.008  0.0080   1.941 0.16390    
newsub0$`Asian Fusion`                          1  0.014  0.0136   3.307 0.06931 .  
newsub0$Pubs                                    1  0.002  0.0022   0.541 0.46216    
newsub0$Chinese                                 1  0.026  0.0265   6.458 0.01121 *  
newsub0$Food:newsub0$Bars                       1  0.027  0.0268   6.530 0.01077 *  
newsub0$Food:newsub0$`American Traditional`     1  0.032  0.0321   7.824 0.00527 ** 
newsub0$Bars:newsub0$`American Traditional`     1  0.003  0.0031   0.750 0.38676    
```

By the table shows above, we can find Food, Nightlife, Breakfast, Seafood, and Chinese restaurants have a significant difference from the average of all restaurant. But, we don’t know which restaurant is above the mean, which one is less. So, we used Tukey method to find the diff of each kind.

```
$`newsub0$Food`
             diff        lwr        upr p adj
Food-0 0.04690805 0.03869788 0.05511822     0

$`newsub0$Bars`
               diff         lwr         upr     p adj
Bars-0 -0.007115777 -0.01596286 0.001731302 0.1147949

$`newsub0$`American Traditional``
                               diff         lwr         upr     p adj
American Traditional-0 -0.007463098 -0.01744675 0.002520555 0.1426978

$`newsub0$Nightlife`
                   diff          lwr        upr     p adj
Nightlife-0 0.008087991 -0.001375567 0.01755155 0.0938259

$`newsub0$`Breakfast & Brunch``
                            diff         lwr          upr     p adj
Breakfast & Brunch-0 -0.01324319 -0.02286955 -0.003616831 0.0070636

$`newsub0$Japanese`
                  diff         lwr        upr     p adj
Japanese-0 5.71922e-05 -0.01785069 0.01796508 0.9950004

$`newsub0$`Sports Bars``
                     diff         lwr        upr     p adj
Sports Bars-0 0.003901779 -0.01297436 0.02077792 0.6501174

$`newsub0$Seafood`
                 diff         lwr          upr     p adj
Seafood-0 -0.02329715 -0.04189601 -0.004698287 0.0141434

$`newsub0$Delis`
               diff         lwr         upr     p adj
Delis-0 -0.01244716 -0.03035505 0.005460725 0.1728673

$`newsub0$`Asian Fusion``
                      diff         lwr         upr     p adj
Asian Fusion-0 -0.01965471 -0.04190647 0.002597059 0.0833424

$`newsub0$Pubs`
               diff         lwr        upr     p adj
Pubs-0 -0.005938527 -0.02368628 0.01180923 0.5115449

$`newsub0$Chinese`
                 diff         lwr          upr     p adj
Chinese-0 -0.01875562 -0.03537094 -0.002140291 0.0269808
```

By the table shows above, we selected Food and Nightlife to test whether people will give a higher star when they stay in those kinds of restaurant with wifi service.

2. Comparing the selected restaurant with all restaurants.

We decided to use ridge regression first. Since we used Lasso regression at the previous chapter and also ridge regression won’t kick wifi out if wifi is not a factor.

```
                                                    cityLas Vegas                          cityothers                           cityPhoenix
                3.9078818203                         -0.0215124969                       -0.0346255860                         -0.0225972813
               cityPittsburgh  `attributes.Accepts Credit Cards`  `attributes.Good For Groups`TRUE  `attributes.Outdoor Seating`TRUE
               -0.0422141088                         -0.3826373422                        0.0039997256                          0.0871862563
      `attributes.Price Range`   `attributes.Good for Kids`TRUE          attributes.Alcoholfull_bar                attributes.Alcoholnone
               -0.0158140437                         -0.1034633002                       -0.2408499873                         -0.0834540358
   `attributes.Noise Level`loud   `attributes.Noise Level`quiet  `attributes.Noise Level`very_loud             `attributes.Has TV`TRUE
               -0.1555335044                          0.1475491373                       -0.3731702087                         -0.0256886359
        attributes.Attiredressy             attributes.Attireformal              attributes.DeliveryTRUE           `attributes.Take-out`TRUE
                0.0009856363                         -0.5468063698                        0.1702427243                         -0.1576075253
`attributes.Takes Reservations`TRUE  `attributes.Waiter Service`TRUE                  attributes.wififree                   attributes.wifipaid
                0.0683921287                          0.0279705740                       -0.0085485186                         -0.2681905851
  attributes.Ambience.romanticTRUE   attributes.Ambience.intimateTRUE    attributes.Ambience.classyTRUE        attributes.Ambience.diveyTRUE
                0.2091580950                          0.2309000133                        0.3008382004                          0.1493611032
  attributes.Ambience.touristyTRUE     attributes.Ambience.trendyTRUE    attributes.Ambience.upscaleTRUE      attributes.Ambience.casualTRUE
               -0.2577760537                          0.1911879867                        0.3417336466                          0.1499238638
 `attributes.Good For.dessert`TRUE `attributes.Good For.latenight`TRUE   `attributes.Good For.lunch`TRUE  `attributes.Good For.dinner`TRUE
                0.0260925120                         -0.0797058576                        0.0252188725                         -0.0125492315
`attributes.Good For.breakfast`TRUE  `attributes.Good For.brunch`TRUE     attributes.Parking.streetTRUE  attributes.Parking.validatedTRUE
                0.0119443064                          0.1417137841                        0.3331161519                         -0.0034687671
       attributes.Parking.lotTRUE       attributes.Parking.valetTRUE                                nrev                             nrev_wifi
                0.2072709029                         -0.1122131392                        0.0007787030                          0.0087700770
```

The table above shows the result of ridge regression. We only have 2883 variables. (The restaurants relate to Food and Nightlife.)

The coefficient of wifi-free is -0.008545 and of wifi-paid is -0.268190. We were confused about why wifi still gives a negative effect? Then we only tested the Food restaurant.

```
Coefficients:
                        (Intercept)                       cityLas Vegas                          cityothers
                           3.728311                           -0.150318                           -0.096833
                        cityPhoenix                      cityPittsburgh   `attributes.Accepts Credit Cards`
                          -0.116253                           -0.099072                           -0.440583
   `attributes.Good For Groups`TRUE    `attributes.Outdoor Seating`TRUE            `attributes.Price Range`
                           0.008172                            0.162510                            0.026675
     `attributes.Good for Kids`TRUE          attributes.Alcoholfull_bar               attributes.Alcoholnone
                          -0.064822                           -0.294453                           -0.041645
       `attributes.Noise Level`loud       `attributes.Noise Level`quiet   `attributes.Noise Level`very_loud
                          -0.140774                            0.172795                           -0.467158
            `attributes.Has TV`TRUE             attributes.Attiredressy             attributes.Attireformal
                           0.034465                            0.215915                           -0.450234
            attributes.DeliveryTRUE           `attributes.Take-out`TRUE `attributes.Takes Reservations`TRUE
                           0.139506                           -0.069342                            0.046127
    `attributes.Waiter Service`TRUE                 attributes.wififree                 attributes.wifipaid
                           0.066672                           -0.057475                           -0.330179
   attributes.Ambience.romanticTRUE    attributes.Ambience.intimateTRUE      attributes.Ambience.classyTRUE
                           0.139702                            0.435240                            0.070560
      attributes.Ambience.diveyTRUE    attributes.Ambience.touristyTRUE      attributes.Ambience.trendyTRUE
                           0.215185                           -0.206806                            0.247001
    attributes.Ambience.upscaleTRUE      attributes.Ambience.casualTRUE   `attributes.Good For.dessert`TRUE
                           0.101286                            0.179496                            0.081361
`attributes.Good For.latenight`TRUE     `attributes.Good For.lunch`TRUE    `attributes.Good For.dinner`TRUE
                          -0.193475                            0.098909                           -0.078657
`attributes.Good For.breakfast`TRUE    `attributes.Good For.brunch`TRUE       attributes.Parking.streetTRUE
                           0.011696                            0.172884                            0.428519
   attributes.Parking.validatedTRUE          attributes.Parking.lotTRUE        attributes.Parking.valetTRUE
                           0.053148                            0.198924                           -0.072224
                               nrev                           nrev_wifi
                           0.001023                            0.007894
```

We used the step forward method first. We chose wifi-free and response ave-star as our null model. And forward step to the full model. Now, the coefficient of free-wifi is -0.057475. This is very wired. Does wifi hurt customers’ feeling? We don’t know. Then, we decided to use ridge regression to analyze it again.

```
                                                 cityLas Vegas                                    cityothers
                  3.6351894732                   -0.0843934793                                 -0.0310623261
                   cityPhoenix                  cityPittsburgh          `attributes.Accepts Credit Cards`
                 -0.0465659935                   -0.0124067152                                 -0.4025814058
  `attributes.Good For Groups`TRUE  `attributes.Outdoor Seating`TRUE        `attributes.Price Range`
                 -0.0044528029                    0.1502564625                                  0.0300642854
   `attributes.Good for Kids`TRUE      attributes.Alcoholfull_bar             attributes.Alcoholnone
                 -0.0499367750                   -0.2180878805                                 -0.0314468627
     `attributes.Noise Level`loud   `attributes.Noise Level`quiet  `attributes.Noise Level`very_loud
                 -0.1259170312                    0.1480647421                                 -0.4285114555
          `attributes.Has TV`TRUE         attributes.Attiredressy             attributes.Attireformal
                  0.0225792338                    0.1531586027                                 -0.4022831500
        attributes.DeliveryTRUE      `attributes.Take-out`TRUE `attributes.Takes Reservations`TRUE
                  0.1357208180                   -0.0560093248                                  0.0416024923
  `attributes.Waiter Service`TRUE           attributes.wififree                 attributes.wifipaid
                  0.0616544348                   -0.0419745213                                 -0.3058201475
 attributes.Ambience.romanticTRUE attributes.Ambience.intimateTRUE    attributes.Ambience.classyTRUE
                  0.1180199933                    0.3491271252                                  0.0906382694
    attributes.Ambience.diveyTRUE attributes.Ambience.touristyTRUE    attributes.Ambience.trendyTRUE
                  0.1923024902                   -0.2073836470                                  0.2011613450
  attributes.Ambience.upscaleTRUE   attributes.Ambience.casualTRUE  `attributes.Good For.dessert`TRUE
                  0.0675734083                    0.1723704213                                  0.0854177539
`attributes.Good For.latenight`TRUE  `attributes.Good For.lunch`TRUE  `attributes.Good For.dinner`TRUE
                 -0.1648334336                    0.1005129125                                 -0.0607125136
`attributes.Good For.breakfast`TRUE `attributes.Good For.brunch`TRUE   attributes.Parking.streetTRUE
                  0.0133343613                    0.1602929335                                  0.3662694991
 attributes.Parking.validatedTRUE      attributes.Parking.lotTRUE     attributes.Parking.valetTRUE
                  0.0995737152                    0.1706992817                                 -0.0632751725
                          nrev                       nrev_wifi
                  0.0008848963                    0.0082066882
```

The table above shows the result. The coefficient of free- wifi is -0.04197. Still a negative number.

```
Call:
lm(formula = exp_ave_star ~ ., data = df6, weights = sqrt(nrev))

Weighted Residuals:
    Min      1Q  Median      3Q     Max 
-8.8919 -1.0772 -0.1477  0.8963  9.1122 

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)                        2.300e+00  7.692e-02  29.895  < 2e-16 ***
cityLas Vegas                      1.147e-01  2.451e-02   4.679 2.93e-06 ***
cityothers                         6.777e-02  2.253e-02   3.007 0.002643 ** 
cityPhoenix                        8.969e-02  2.572e-02   3.487 0.000491 ***
cityPittsburgh                    -4.224e-02  3.360e-02  -1.257 0.208833    
attributes.Accepts_Credit_Cards   -4.188e-01  4.118e-02 -10.170  < 2e-16 ***
attributes.Good_For_GroupsTRUE    -1.609e-01  2.349e-02  -6.851 7.76e-12 ***
attributes.Outdoor_SeatingTRUE    -8.540e-02  1.289e-02  -6.623 3.70e-11 ***
attributes.Price_Range            -4.921e-03  1.498e-02  -0.329 0.742466    
attributes.Good_for_KidsTRUE      -8.511e-02  2.026e-02  -4.201 2.68e-05 ***
attributes.Alcoholfull_bar        -1.739e-01  1.894e-02  -9.185  < 2e-16 ***
attributes.Alcoholnone             3.298e-03  1.909e-02   0.173 0.862833    
attributes.Noise_Levelloud         2.012e-02  2.239e-02   0.899 0.368897    
attributes.Noise_Levelquiet        1.188e-02  1.785e-02   0.666 0.505663    
attributes.Noise_Levelvery_loud   -2.134e-02  3.922e-02  -0.544 0.586394    
attributes.Has_TVTRUE             -6.433e-02  1.354e-02  -4.749 2.07e-06 ***
attributes.Attiredressy            9.864e-02  3.736e-02   2.640 0.008302 ** 
attributes.Attireformal            3.993e-01  2.355e-01   1.696 0.089976 .  
attributes.DeliveryTRUE           -2.334e-02  1.858e-02  -1.256 0.208975    
attributes.Take_outTRUE            5.416e-02  2.456e-02   2.205 0.027444 *  
attributes.Takes_ReservationsTRUE -1.446e-02  1.575e-02  -0.918 0.358451    
attributes.Waiter_ServiceTRUE     -2.117e-02  1.872e-02  -1.131 0.258034    
attributes.wififree               -8.959e-03  1.344e-02  -0.666 0.505197    
attributes.wifipaid               -1.157e-01  6.095e-02  -1.898 0.057758 .  
attributes.Ambience.romanticTRUE   3.197e-02  4.512e-02   0.709 0.478606    
attributes.Ambience.intimateTRUE   1.697e-01  5.104e-02   3.325 0.000889 ***
attributes.Ambience.classyTRUE     1.008e-01  3.718e-02   2.712 0.006695 ** 
attributes.Ambience.hipsterTRUE    2.383e-01  4.101e-02   5.811 6.41e-09 ***
attributes.Ambience.diveyTRUE      2.863e-01  3.210e-02   8.921  < 2e-16 ***
attributes.Ambience.trendyTRUE     2.596e-02  3.077e-02   0.844 0.398794    
attributes.Ambience.upscaleTRUE    1.421e-01  5.362e-02   2.649 0.008079 ** 
attributes.Ambience.casualTRUE     9.730e-02  1.980e-02   4.914 9.04e-07 ***
attributes.Good_For.dessertTRUE   -1.086e-02  4.194e-02  -0.259 0.795744    
attributes.Good_For.latenightTRUE  1.516e-04  2.473e-02   0.006 0.995108    
attributes.Good_For.lunchTRUE     -4.240e-02  1.468e-02  -2.889 0.003878 ** 
attributes.Good_For.dinnerTRUE    -3.917e-02  1.485e-02  -2.639 0.008337 ** 
attributes.Good_For.breakfastTRUE -8.694e-02  2.433e-02  -3.573 0.000355 ***
attributes.Good_For.brunchTRUE    -5.658e-02  2.257e-02  -2.507 0.012179 *  
attributes.Parking.garageTRUE     -2.455e-01  2.241e-02 -10.953  < 2e-16 ***
attributes.Parking.streetTRUE      1.935e-01  2.000e-02   9.672  < 2e-16 ***
attributes.Parking.lotTRUE         5.112e-02  1.771e-02   2.887 0.003897 ** 
attributes.Parking.valetTRUE      -6.004e-02  2.617e-02  -2.294 0.021809 *  
nrev                               4.281e-04  2.219e-05  19.298  < 2e-16 ***
```

The table above is the coefficient of our final model. By comparing with our final model, we found the coefficient of wifi-free in the selected restaurant gives more negative effects to the average start than all kinds of restaurant, which is very strange.

We were really very confused about this result.

Why wifi gives a negative effect on the average stars? People hate free wifi? The next chapter will give you the answer.

Codes:

```
Library(MASS)

new.sub.2 <- subset(new.sub.1, rowSums(new.sub.1) > 2)

new.sub.3 <- subset(new.sub.1, rowSums(new.sub.1) <= 2)

 new.sub.0 <- subset(new.sub.1, rowSums(new.sub.1) > 1)

 new.sub.2 <- subset(new.sub.0, rowSums(new.sub.0) > 2)

 new.sub.3 <- subset(new.sub.0, rowSums(new.sub.0) <= 2)

save(new.sub.0, file = "new.sub.0.cvs")

saveRDS(foo, file="data.Rda")

bar <- readRDS(file="data.Rda")

newsub0<-new.sub.0


newsub0$Food[which(newsub0$Food=="1")]<-"Food"

newsub0$Bars[which(newsub0$Bars=="1")]<-"Bars"

newsub0$`American Traditional`[which(newsub0$`American
Traditional`=="1")]<-"American Traditional"

newsub0$Nightlife[which(newsub0$Nightlife=="1")]<-"Nightlife"

newsub0$`Breakfast & Brunch`[which(newsub0$`Breakfast &
Brunch`=="1")]<-"Breakfast & Brunch"

newsub0$Japanese[which(newsub0$Japanese=="1")]<-"Japanese"

newsub0$`Sports Bars`[which(newsub0$`Sports Bars`=="1")]<-"Sports Bars"

newsub0$Seafood[which(newsub0$Seafood=="1")]<-"Seafood"

newsub0$Delis[which(newsub0$Delis=="1")]<-"Delis"

newsub0$`Asian Fusion`[which(newsub0$`Asian Fusion`=="1")]<-"Asian Fusion"

newsub0$Pubs[which(newsub0$Pubs=="1")]<-"Pubs"


newsub0$Chinese[which(newsub0$Chinese=="1")]<-"chinese"

aov <- aov(newsub0$ave.wifi ~
newsub0$Food*newsub0$Bars*newsub0$`American
Traditional`*newsub0$Nightlife*newsub0$`Breakfast &
Brunch`*newsub0$Japanese*newsub0$`Sports
```

```
Bars`*newsub0$Seafood*newsub0$Delis*newsub0$`Asian
Fusion`*newsub0$Pubs*newsub0$Chinese, data=newsub0)

summary(aov)

TukeyHSD(aov)

> df2.sub<-df2[,-36:-38]

> View(df2.sub)

> df2.sub1<-df2.sub[,-37]

> View(df2.sub1)

df2.sub2<-subset(df2.sub1,df2.sub1$Food>0|df2.sub1$Nightlife>0|
df2.sub1$`Breakfast & Brunch`>0|df2.sub1$Seafood>0|df2.sub1$Chinese>0)

df2.sub3<-df2.sub2[,-37:-48]

 fit0<-aov(df2.sub3$ave_star~df2.sub3$attributes.wifi)

glmmod<-glmnet(x=df2.sub3[,-
35],y=as.factor(df2.sub3$ave_sta),alpha=1,family='binomial')

lm<-lm(df2.sub4$ave_star~.,df2.sub4)

step(lm,direction="forward",keep=df2.sub4$attributes.wifi,steps=1000,)

df2.3<-subset(df2,df2$Food>0|df2$Nightlife>0|df2$`Breakfast & Brunch`>0|
df2$Seafood>0|df2$Chinese>0)

df2.3.1<-df2.3[,-36:-38]

df2.3.2<-df2.3.1[,-37]

df2.3.3<-df2.3.2[,-36:-48]

full2.3=lm(ave_star~.,data=df2.3.3)

null2.3=lm(ave_star~attributes.wifi,data=df2.3.3)

step(full2.3,direction="forward",scope=c(null2.3,full2.3),steps=1000)

df2.0<-df2[,-36:-38]

df2.0.1<-df2.0[,-37]

df2.0.2<-df2.0.1[,-36:-48]


full2.0=lm(ave_star~.,data=df2)

null2.0=lm(ave_star~attributes.wifi,data=df2)

step(full2.0,direction="forward",scope=c(null2.0,full2.0),steps=1000)
```

```
library(MASS)

model.ridge <- lm.ridge(ave_star~ ., data=df2.3.3, lambda = seq(0,1000,0.1))

plot(seq(0,1000,0.1), model.ridge$GCV, main="GCV of Ridge Regression",
type="l", xlab=expression(lambda), ylab="GCV")

lambda.ridge <- seq(0,10,0.1)[which.min(model.ridge$GCV)]

beta.ridge <- coef(model.ridge)[which.min(model.ridge$GCV),]

resid.ridge <- train$ave_star - beta.ridge[1] - as.matrix(train[,1:8])%*
%beta.ridge[2:9]

resid.ridge <- train$lpsa - beta.ridge[1] - as.matrix(train[,1:8])%*
%beta.ridge[2:9]

sum(resid.ridge)

model.ridge1 <- lm.ridge(ave_star~ ., data=df2, lambda = seq(0,1000,0.1))

beta.ridge1 <- coef(model.ridge1)[which.min(model.ridge1$GCV),]

beta.ridge1

model.ridge2 <- lm.ridge(ave_star~ ., data=df2.sub5, lambda =
seq(0,1000,0.1))

beta.ridge2 <- coef(model.ridge2)[which.min(model.ridge2$GCV),]

beta.ridge2

model.ridge3 <- lm.ridge(ave_star~ ., data=df2.1, lambda = seq(0,1000,0.1))

beta.ridge3 <- coef(model.ridge3)[which.min(model.ridge3$GCV),]

beta.ridge3

lm3<-lm(ave_star~ ., data=df2.sub5)

summary(lm3)

boxplotlist <- list()

for(i in 1:(ncol(new.sub.3)-1)){

    boxplotlist[[i]] <- (new.sub.3[,i] * new.sub.3$ave.wifi)[new.sub.3[,i] *
new.sub.3$ave.wifi>0]

}

df3<-df2[,-36:-38]

df3.1<-df3[,-37]

df3.2<-subset(df3.1,df3.1$Food>0|df3.1$Nightlife>0)
```

```
df3.3<-df3.2[,-37:-48]

model.ridge3.3 <- lm.ridge(ave_star~ ., data=df3.3, lambda = seq(0,1000,0.1))

beta.ridge3.3 <- coef(model.ridge3.3)[which.min(model.ridge3.3$GCV),]

beta.ridge3.3

df2.sub.all<-subset(df2,df2$Food>0)

df2.sub.all.1<-df2.sub.all[,-41:-52]

model.ridge.all.1 <- lm.ridge(ave_star~ ., data=df2.sub.all.1, lambda =
seq(0,1000,0.1))

beta.ridge.all.1 <- coef(model.ridge.all.1)[which.min(model.ridge.all.1$GCV),]

beta.ridge.all.1
```