

Airbnb London

Kecheng Liang

November 28, 2018

Read data

1) import data

```
initial <- read.csv("airbnblondon.csv")
```

2) remove unrelevant variables and zero reviews

```
LondonAirbnb <- initial[,c(-2,-5,-6,-13,-15,-16,-19)]  
#reviews = 0 is meaningless  
LondonAirbnb <- filter(LondonAirbnb,reviews>0)  
#rating cannot be 0  
LondonAirbnb <- filter(LondonAirbnb,overall_satisfaction>0)
```

3) overview of the data

```
summary(LondonAirbnb)
```

```
##      room_id          host_id          room_type  
##  Min.   : 9554  Min.   : 516  Entire home/apt:17313  
##  1st Qu.: 5716659 1st Qu.: 6463857  Private room    :14564  
##  Median :11943267  Median : 21851426  Shared room     : 320  
##  Mean   :10865973  Mean   : 35038646  
##  3rd Qu.:16250299  3rd Qu.: 50598300  
##  Max.   :20071981  Max.   :141506176  
##  
##           borough          neighborhood  
##  Westminster       : 3751  West End             : 520  
##  Tower Hamlets     : 3123  St. Peter's        : 475  
##  Hackney            : 2624  St. James's       : 417  
##  Camden              : 2518  Weavers            : 389  
##  Kensington and Chelsea: 2289  Lancaster Gate    : 376  
##  Southwark            : 2217  Holborn and Covent Garden: 350  
##  (Other)              :15675  (Other)           :29670  
##      reviews          overall_satisfaction  accommodates    bedrooms  
##  Min.   : 3.00  Min.   :1.000  Min.   : 1.00  Min.   : 0.000  
##  1st Qu.: 5.00  1st Qu.:4.500  1st Qu.: 2.00  1st Qu.: 1.000  
##  Median :11.00  Median :4.500  Median : 2.00  Median : 1.000  
##  Mean   :23.41  Mean   :4.661  Mean   : 3.16  Mean   : 1.339  
##  3rd Qu.:27.00  3rd Qu.:5.000  3rd Qu.: 4.00  3rd Qu.: 2.000  
##  Max.   :435.00  Max.   :5.000  Max.   :16.00  Max.   :10.000  
##  
##      price          latitude        longitude  
##  Min.   : 7.0  Min.   :51.31  Min.   :-0.41389  
##  1st Qu.: 57.0 1st Qu.:51.49  1st Qu.:-0.18583  
##  Median : 96.0  Median :51.51  Median :-0.12442  
##  Mean   :121.2  Mean   :51.51  Mean   :-0.12817  
##  3rd Qu.:149.0  3rd Qu.:51.54  3rd Qu.:-0.07311  
##  Max.   :2451.0  Max.   :51.68  Max.   : 0.20753
```

```

##  

4) convert the numeric to character  

LondonAirbnb$room_id <- as.character(LondonAirbnb$room_id)  

LondonAirbnb$host_id <- as.character(LondonAirbnb$host_id)  

LondonAirbnb$overall_satisfaction <- as.numeric(LondonAirbnb$overall_satisfaction)  

5) remove outliers?  

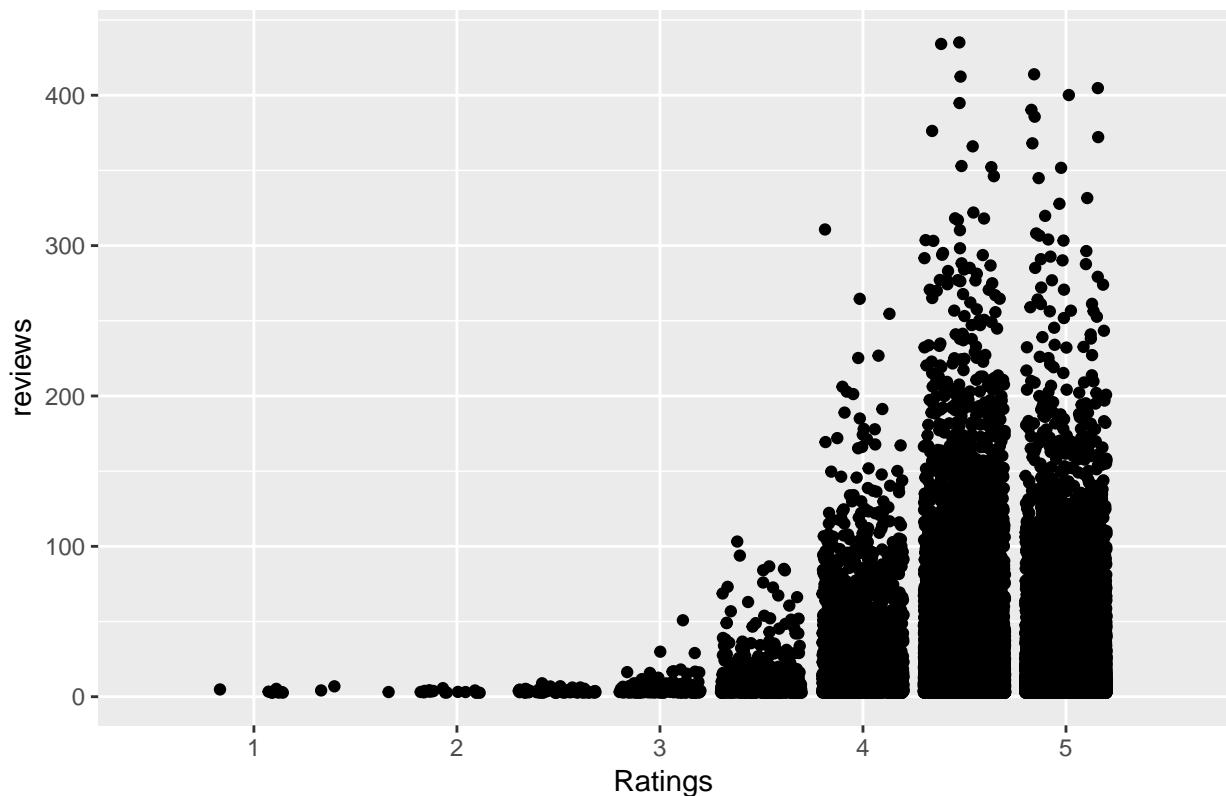
#not sure how to do this

```

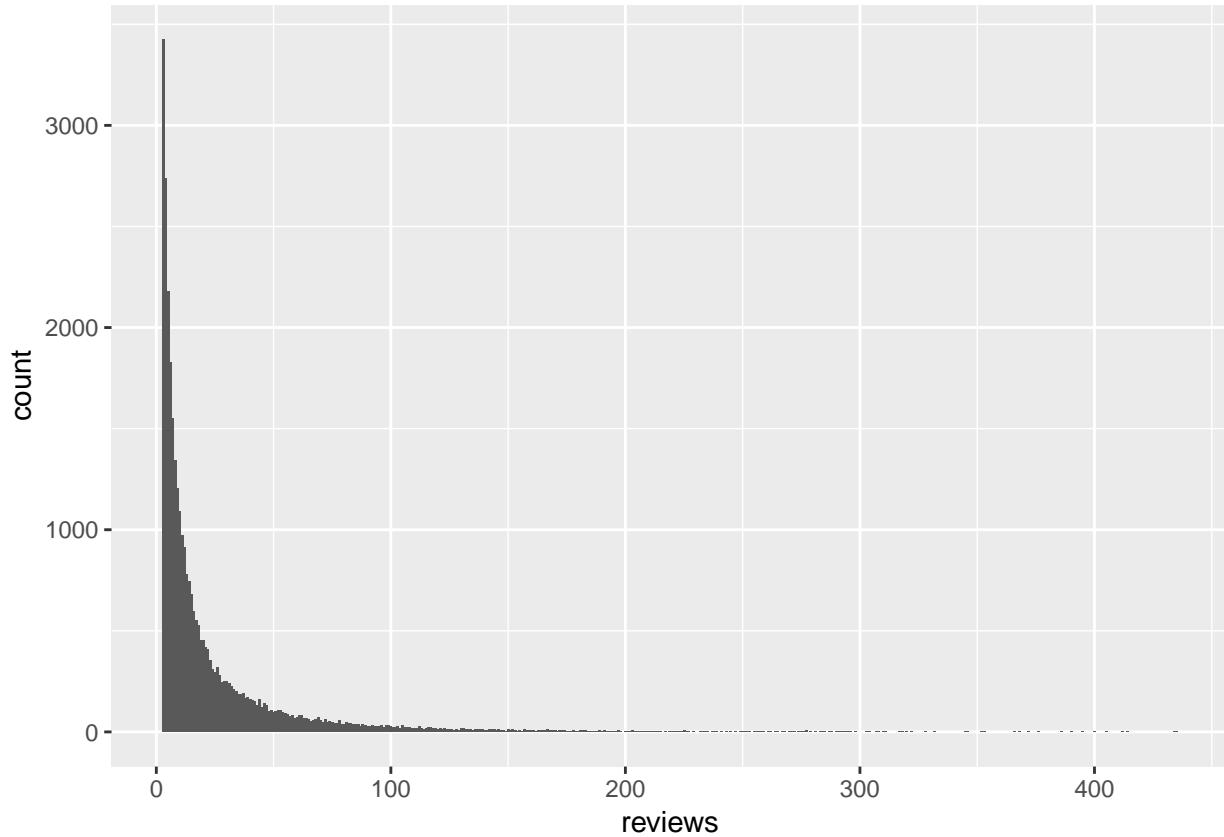
EDA

```
ggplot(data = LondonAirbnb,aes(x=overall_satisfaction,y=reviews))+geom_jitter()+scale_x_discrete(name =
```

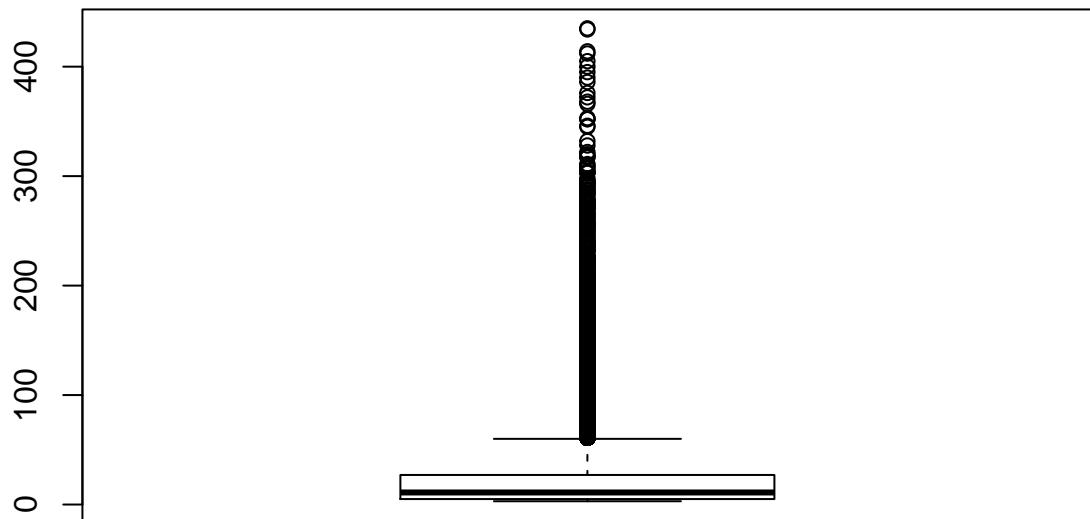
Ratings and Reviews



```
ggplot(data=LondonAirbnb,aes(x=reviews))+geom_bar()
```

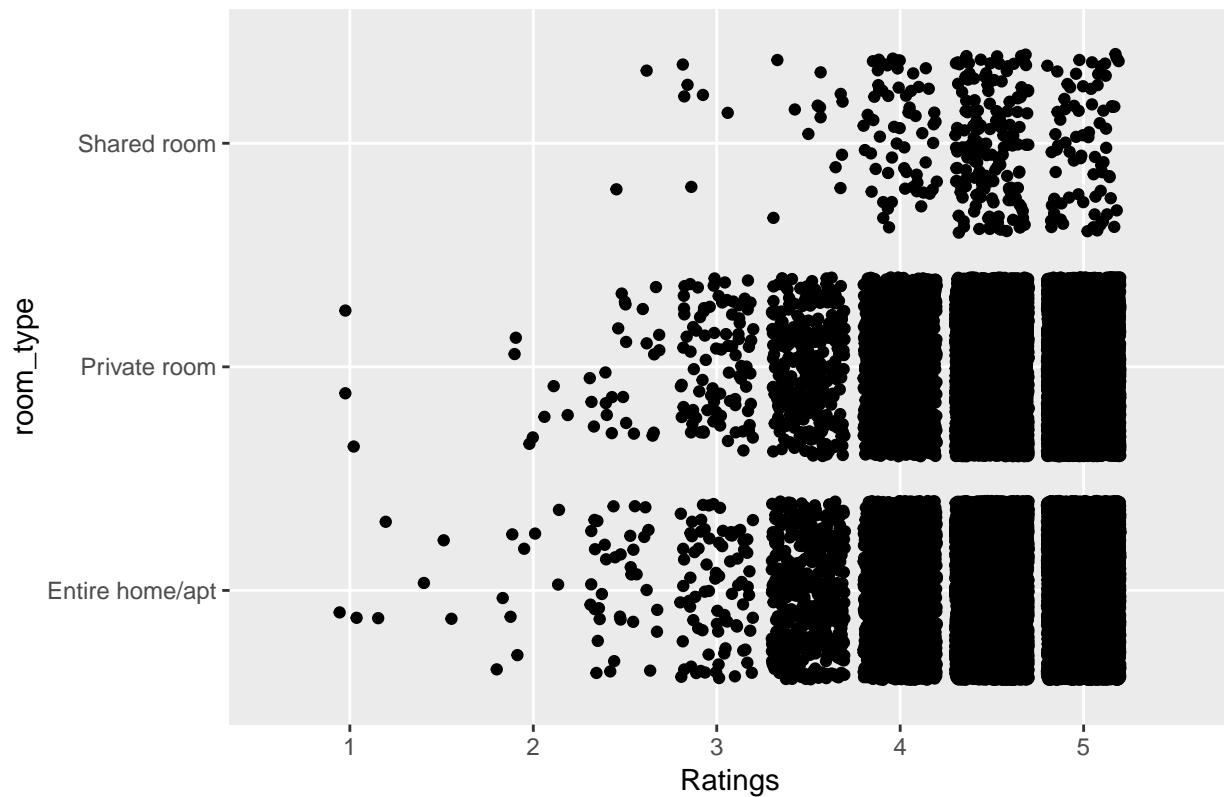


```
boxplot(LondonAirbnb$reviews)
```

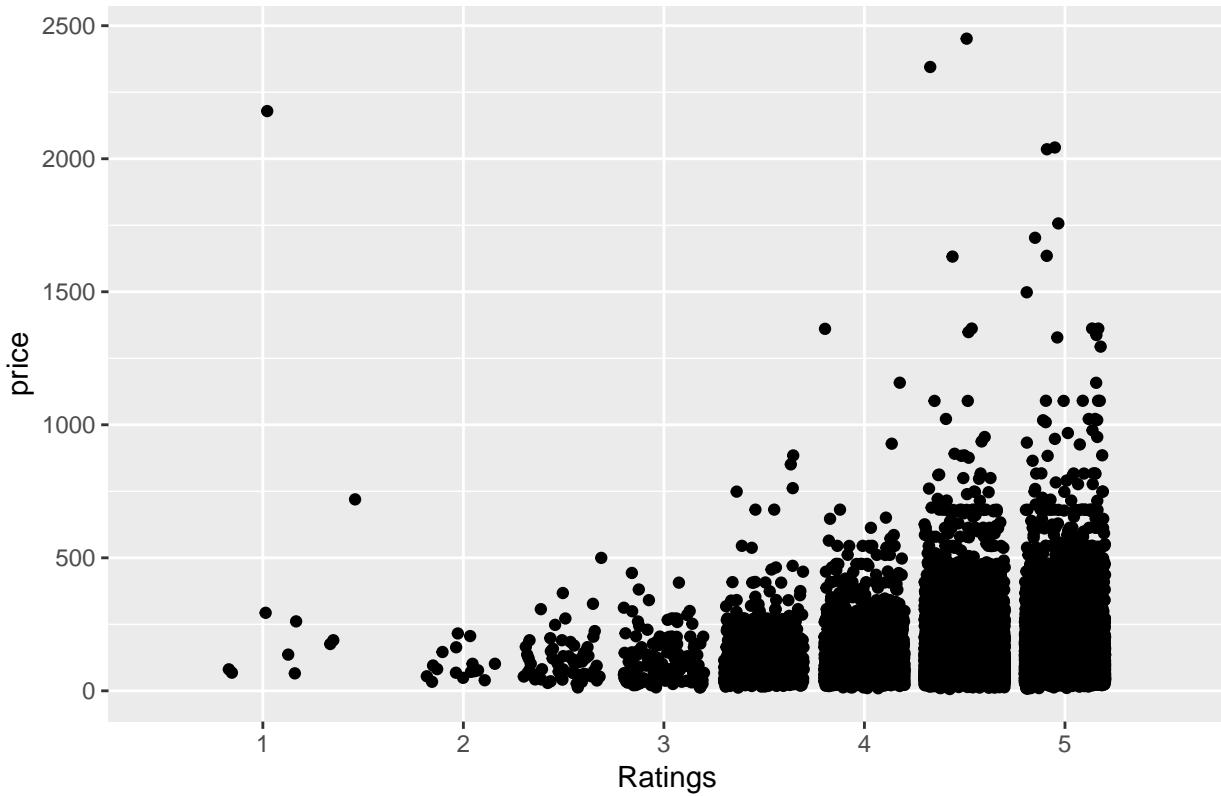


```
ggplot(data = LondonAirbnb,aes(x=overall_satisfaction,y=room_type))+geom_jitter()+scale_x_discrete(name
```

Ratings and Reviews

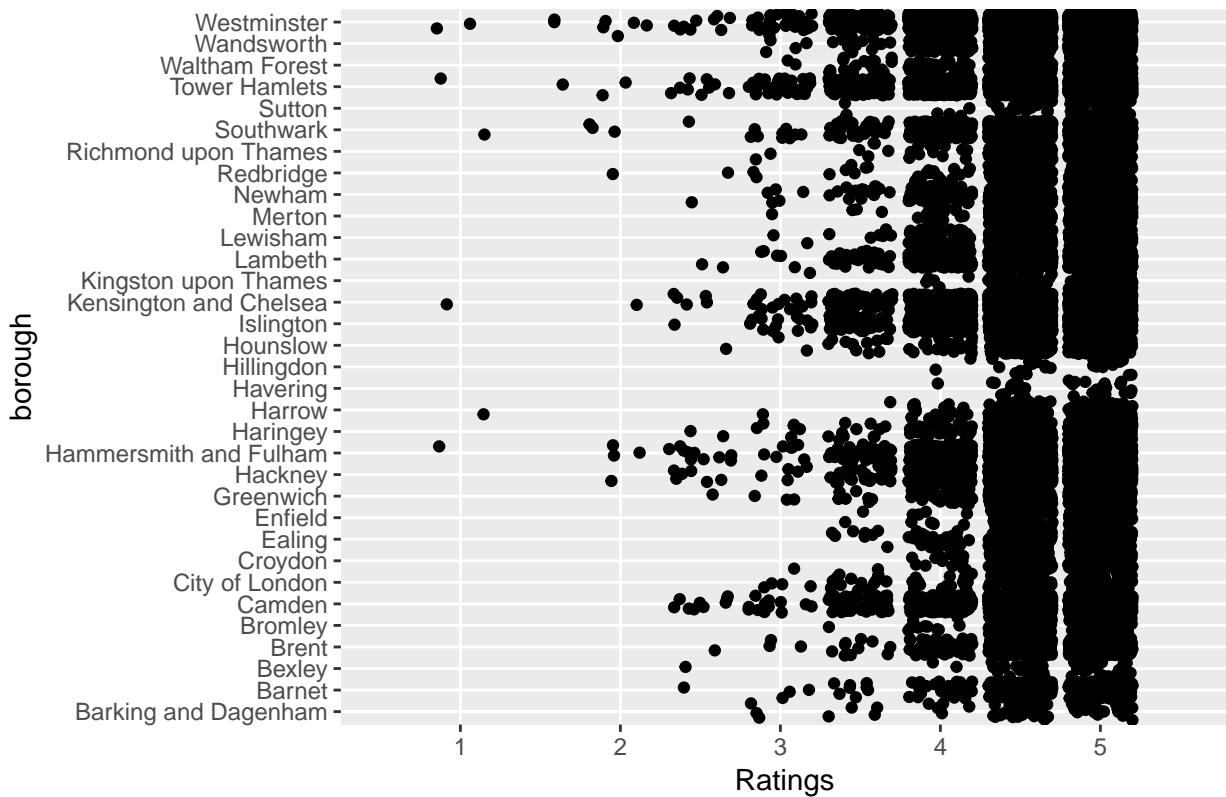


Ratings and Reviews

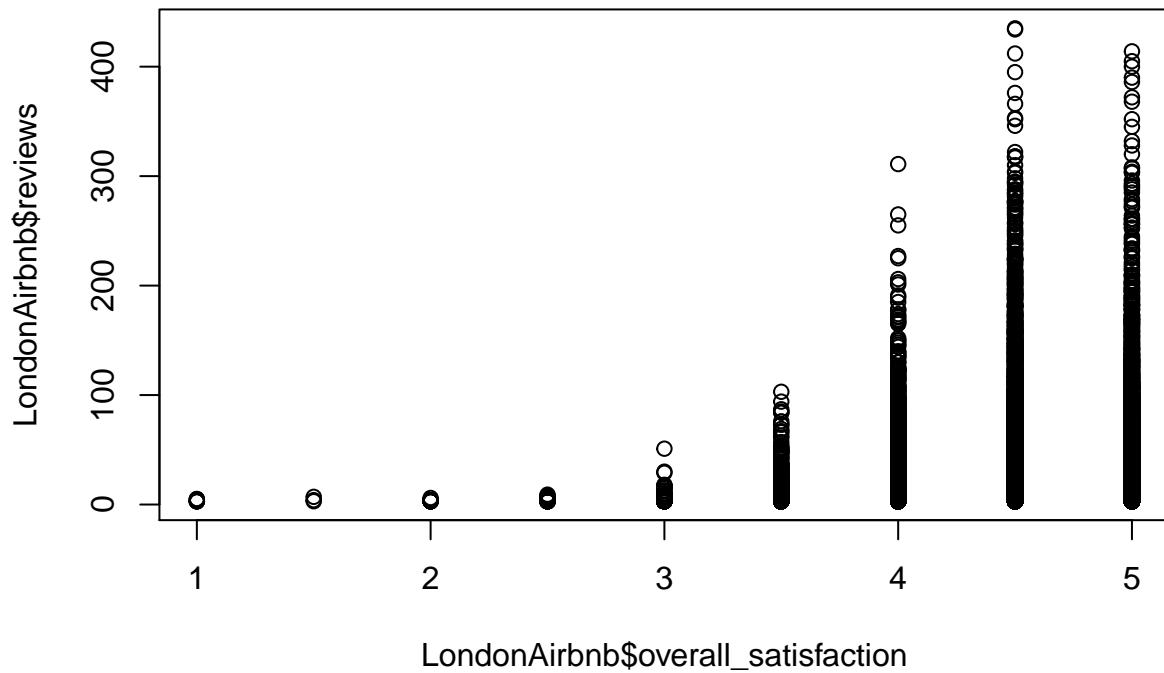


```
ggplot(data = LondonAirbnb,aes(x=overall_satisfaction,y=borough))+geom_jitter()+scale_x_discrete(name =
```

Ratings and Reviews



```
plot(LondonAirbnb$overall_satisfaction, LondonAirbnb$reviews)
```

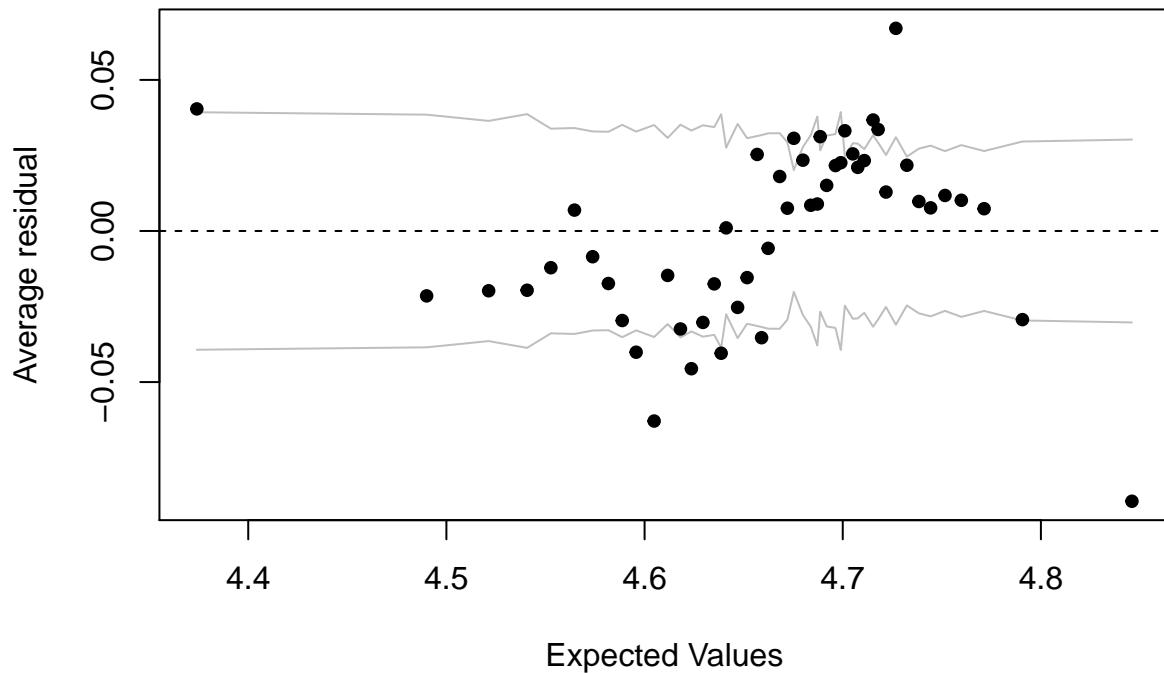


Models

```
#model1 no random effect
model1 <- glm(overall_satisfaction~room_type+reviews+accommodates+bedrooms+log(price),data = LondonAirbn
#remove reviews, no relationship

#model2 remove reviews
#ok d model
model2 <- glm(overall_satisfaction~factor(room_type)+accommodates+bedrooms+log(price),data = LondonAirbn
binnedplot(fitted(model2),residuals(model2),nclass=50)
```

Binned residual plot



```
#random intercept borough
model3 <- lmer(overall_satisfaction~factor(room_type)+accommodates+bedrooms+log(price)+(1|borough),data=airbnb)

#random slope
model4 <- lmer(overall_satisfaction~factor(room_type)+accommodates+bedrooms+log(price)+(0+log(price)|borough))

#random slope and intercept
model5 <- lmer(overall_satisfaction~factor(room_type)+accommodates+bedrooms+log(price)+(1+log(price)|borough))
```