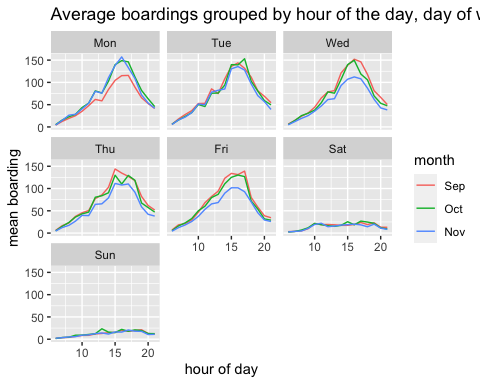
Exercise 2

Yuting Huang, Jipeng Cheng, Weidi Hou

3/5/2022

# Problem 1: visualization

## 1.1 Plot 1: average boarding line graph



According to the graph, it shows that the hour of peak boardings do not change from day to day, and it broadly similar showing the peak hour at 15-17.

**Does the hour of peak boarding change from day to day, or is it broadly similar across days?**

According to the graph, it shows that the hour of peak boardings do not change from day to day, and it broadly similar showing the peak hour at 15-17.

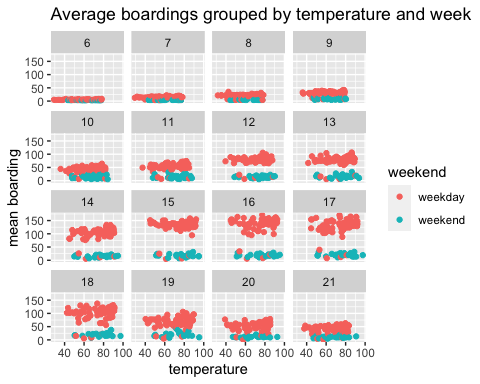
**Why do you think average boarding on Mondays in September look lower, compared to other days and months?**

We can guess that there are less average boardings on September because the beginning of Fall semester. On the other hand, students may not prefer choosing the courses on Monday due to “Monday Blue”.

**Similarly, why do you think average boardings on Weds/Thurs/Fri in November look lower?**

Average boarding on Weds/Thurs/Fri in November look lower because students may have to prepare for the midterm exam, they would like to stay home rather than go outside.

## 1.2 Plot 2: scatter plots showing boardings vs. temperature



**When we hold hour of day and weekend status constant, does temperature seem to have a noticeable effect on the number of UT students riding the bus?**

According to above graph, temperature seem to have no noticeable effect on the number of UT students riding the bus.

# Problem 2: Saratoga house prices

## 2.1 The best linear model

## errs\_lm1 errs\_lm2 errs\_lm3 errs\_lm4   
## 77191.87 66544.64 68911.86 65028.20

**Build the best linear model for price that you can. It should clearly outperform the “medium” model that we considered in class. Use any combination of transformations, engineering features, polynomial terms, and interactions that you want; and use any strategy for selecting the model that you want.**

The result suggests that lm4 is the best, since it has the lowest cross validation rmse. The formular of linear model 4 is as below. The cross-validation rmse of lm4 is smaller than that of medium model. Therefore, our model overperfom the medium model.

$price=livingarea+centralair+bathrooms+fuel+lotsize+bedrooms+rooms+livingareacentralair+livingarea bathrooms+livingarea fuel+livingarea rooms+bathrooms bedrooms+centralair fuel+bathroom fuel+fuel lotsize+centralair bathrooms+bedrooms rooms $

## 2.2 The Best KNN

We will find optimal k and features simultaneously.

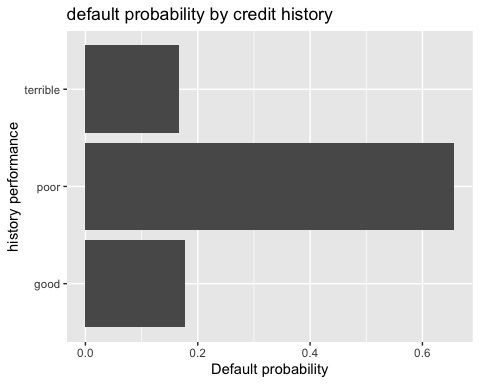
## k err std\_err  
## knn1 40 78057.66 1311.066  
## knn2 23 69015.87 1355.464  
## knn3 33 69419.95 1182.717

**Which model seems to do better at achieving lower out-of-sample mean-squared error?**

Analysis: According to our results, the cross-validation error is lower for linear model, so linear model seems to do better at achieving lower out of sample mean squared error. However, the best variables and cross-validation error for knn and linear model is pretty similar, the difference is , which is very small compared to their value.

# problem 3: Classification and retrospective sampling

## 3.1 bar plot of default probability by credit history



## 3.2 logistic regression model

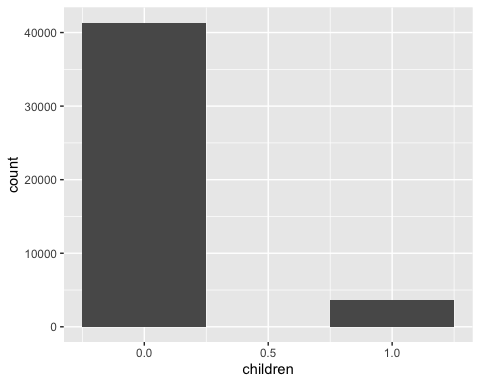
g\_c\_glm = glm(formula = Default ~ duration + amount + installment + age +   
 history + purpose + foreign, family = "binomial", data = g\_c)

**What do you notice about the history variable vis-a-vis predicting defaults? What do you think is going on here? In light of what you see here, do you think this data set is appropriate for building a predictive model of defaults, if the purpose of the model is to screen prospective borrowers to classify them into “high” versus “low” probability of default? Why or why not—and if not, would you recommend any changes to the bank’s sampling scheme?**

Based on the data, the default probability of people with good history is higher than the default probability of people with poor history. There is a sampling problem in this model, since this model based on the “case-control” design. To be specific, if we want to choose three features as factors in the model, but only two of them have been chosen in to the model but one feature has been omitted, it will lead to more incorrect analysis for those two features. It’s a type of problem of selection bias and misrepresentation. Based on our previous analysis, this data set is inappropriate for building a predictive model of defaults. We recommend the bank change their sampling scheme in order to solve the unbalanced sample problem, such as using random sampling method.

# problem 4: Children and hotel reservations

## 4.1 build model

Count the dependent variable: 

### 4.1.1 Build baseline model

baseline1 = glm(children ~ market\_segment + adults + customer\_type + is\_repeated\_guest,  
 data = hotel\_dev\_train, family = "binomial")  
baseline2 = glm(children ~ . - arrival\_date , data = hotel\_dev\_train, family = "binomial")

### 4.2 Build best model - Feature engineering with LASSO

idea: use LASSO to find main effects + interaction by eyeballing

hotel\_lasso\_x\_main = model.matrix(children ~ (.-1-arrival\_date), data=hotel\_dev\_train)  
hotel\_lasso\_x\_itac = model.matrix(children ~ (.-1-arrival\_date)^2, data=hotel\_dev\_train)  
hotel\_lasso\_y = hotel\_dev\_train$children  
#//see https://cran.r-project.org/web/packages/gamlr/gamlr.pdf to see more  
hotel\_lasso\_main = cv.gamlr(hotel\_lasso\_x\_main, hotel\_lasso\_y, nfold=10, verb=TRUE, family="binomial")

## fold 1,2,3,4,5,6,7,8,9,10,done.

hotel\_lasso\_itac = cv.gamlr(hotel\_lasso\_x\_itac, hotel\_lasso\_y, nfold=10, verb=TRUE, family="binomial")

## fold 1,2,3,4,5,6,7,8,9,10,done.

**Extract strong single covariates:**

## 49 x 1 sparse Matrix of class "dgCMatrix"  
## seg100  
## intercept -4.4425241336  
## hotelCity\_Hotel 0.6886883677  
## hotelResort\_Hotel -0.0010064992  
## lead\_time 0.0005849715  
## stays\_in\_weekend\_nights 0.0501041212  
## stays\_in\_week\_nights -0.0012659695  
## adults -0.5016266001  
## mealFB 0.6450670848  
## mealHB .   
## mealSC -1.0797159048  
## mealUndefined .   
## market\_segmentComplementary 0.1050831206  
## market\_segmentCorporate -1.0223156529  
## market\_segmentDirect .   
## market\_segmentGroups -1.0297104242  
## market\_segmentOffline\_TA/TO .   
## market\_segmentOnline\_TA 0.1054001264  
## distribution\_channelDirect 0.1483963646  
## distribution\_channelGDS -1.3398380948  
## distribution\_channelTA/TO .   
## is\_repeated\_guest -0.8017789974  
## previous\_cancellations .   
## previous\_bookings\_not\_canceled -0.0851655563  
## reserved\_room\_typeB 1.6789215162  
## reserved\_room\_typeC 2.8379369774  
## reserved\_room\_typeD -1.0442876331  
## reserved\_room\_typeE .   
## reserved\_room\_typeF 1.6234661219  
## reserved\_room\_typeG 2.3176525342  
## reserved\_room\_typeH 2.7850887569  
## assigned\_room\_typeB 0.2223686225  
## assigned\_room\_typeC 1.3842601784  
## assigned\_room\_typeD 0.9858630083  
## assigned\_room\_typeE 0.4866210803  
## assigned\_room\_typeF 0.7967735311  
## assigned\_room\_typeG 1.0180119152  
## assigned\_room\_typeH 1.5585023992  
## assigned\_room\_typeI 1.2422884759  
## assigned\_room\_typeK .   
## booking\_changes 0.2422534895  
## deposit\_typeNon\_Refund .   
## deposit\_typeRefundable .   
## days\_in\_waiting\_list .   
## customer\_typeGroup .   
## customer\_typeTransient 0.2079811998  
## customer\_typeTransient-Party -0.4523303579  
## average\_daily\_rate 0.0106105320  
## required\_car\_parking\_spacesparking 0.0790750643  
## total\_of\_special\_requests 0.4627108625

From the output, we see that the variable **deposit\_type** is insignificant from zero. Therefore, we rule out it.

**Extract strong interactions:**

## strong\_interaction\_name  
## 1 market\_segmentOnline\_TA:reserved\_room\_typeB  
## 2 reserved\_room\_typeD:assigned\_room\_typeB  
## 3 reserved\_room\_typeB:assigned\_room\_typeI  
## 4 market\_segmentOffline\_TA/TO:reserved\_room\_typeH  
## 5 reserved\_room\_typeF:assigned\_room\_typeK  
## 6 hotelResort\_Hotel:reserved\_room\_typeB  
## 7 mealHB:reserved\_room\_typeF  
## 8 mealUndefined:reserved\_room\_typeG  
## 9 reserved\_room\_typeB:assigned\_room\_typeB  
## 10 market\_segmentComplementary:reserved\_room\_typeF  
## 11 reserved\_room\_typeB:assigned\_room\_typeG  
## 12 mealSC:reserved\_room\_typeG  
## 13 mealUndefined:assigned\_room\_typeB  
## 14 market\_segmentDirect:reserved\_room\_typeF  
## 15 market\_segmentComplementary:reserved\_room\_typeC  
## 16 mealSC:reserved\_room\_typeF  
## 17 reserved\_room\_typeH:assigned\_room\_typeF  
## 18 reserved\_room\_typeB:assigned\_room\_typeH  
## 19 adults:previous\_bookings\_not\_canceled  
## 20 reserved\_room\_typeE  
## 21 reserved\_room\_typeB:assigned\_room\_typeF  
## 22 market\_segmentDirect:reserved\_room\_typeC  
## 23 assigned\_room\_typeH:required\_car\_parking\_spacesparking  
## 24 mealFB:assigned\_room\_typeI  
## 25 hotelResort\_Hotel:reserved\_room\_typeE  
## 26 reserved\_room\_typeF  
## 27 is\_repeated\_guest:reserved\_room\_typeE  
## 28 reserved\_room\_typeF:assigned\_room\_typeB  
## 29 mealHB:reserved\_room\_typeD  
## 30 mealUndefined:assigned\_room\_typeD  
## strong\_interaction\_beta abs\_beta  
## 1 -4.05270622352351 4.052706  
## 2 3.93023580386564 3.930236  
## 3 3.80360626103647 3.803606  
## 4 3.19899477005572 3.198995  
## 5 3.18612384092973 3.186124  
## 6 3.11391986157249 3.113920  
## 7 3.01295371737258 3.012954  
## 8 -2.9743065251412 2.974307  
## 9 -2.84578478707202 2.845785  
## 10 2.44406993328313 2.444070  
## 11 2.40879558537554 2.408796  
## 12 2.38054515773994 2.380545  
## 13 -2.34513063376819 2.345131  
## 14 2.32677126069374 2.326771  
## 15 2.30581640077205 2.305816  
## 16 -2.29435972473223 2.294360  
## 17 2.22115056290171 2.221151  
## 18 2.07398976642851 2.073990  
## 19 -2.0573340603283 2.057334  
## 20 2.02762017809684 2.027620  
## 21 2.01653150491817 2.016532  
## 22 2.00502301665241 2.005023  
## 23 1.95331345421732 1.953313  
## 24 -1.93188627292425 1.931886  
## 25 -1.92243654062527 1.922437  
## 26 1.82557160489881 1.825572  
## 27 1.73680308773075 1.736803  
## 28 1.71702004164993 1.717020  
## 29 1.64006485914289 1.640065  
## 30 -1.63888517151056 1.638885

From the results, we pick these terms: , , , ,, , , , , by eyeballing, since they are significant from zero.

**Rule out non-converged covariates & interactions:**

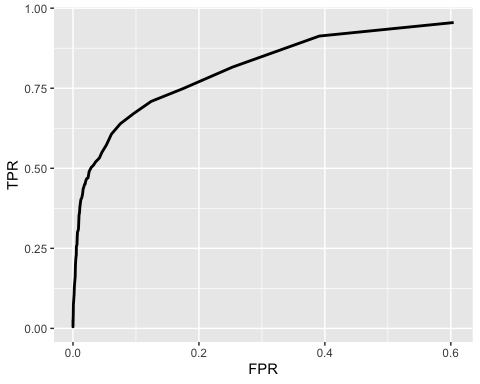
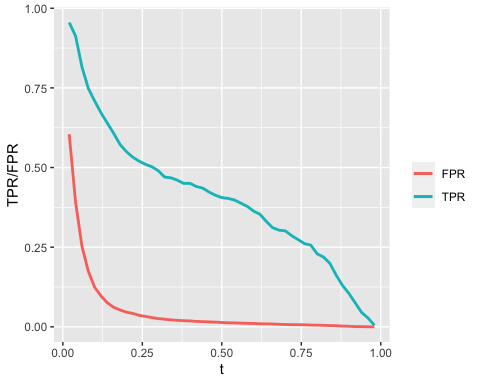
## (Intercept)   
## -5.107927e+15   
## hotelResort\_Hotel   
## 1.820967e+14   
## lead\_time   
## 9.560505e+11   
## stays\_in\_weekend\_nights   
## -3.792517e+14   
## stays\_in\_week\_nights   
## 2.115145e+13   
## adults   
## -2.414171e+13   
## mealFB   
## 3.397841e+14   
## mealHB   
## 1.339467e+15   
## mealSC   
## 2.148706e+15   
## mealUndefined   
## -1.162702e+14   
## market\_segmentComplementary   
## 2.647540e+15   
## market\_segmentCorporate   
## -2.551907e+14   
## market\_segmentDirect   
## 6.977011e+14   
## market\_segmentGroups   
## -1.827543e+15   
## market\_segmentOffline\_TA/TO   
## -1.376313e+15   
## market\_segmentOnline\_TA   
## -9.762417e+14   
## distribution\_channelDirect   
## 1.441851e+15   
## distribution\_channelGDS   
## 2.560627e+15   
## distribution\_channelTA/TO   
## 2.294321e+15   
## is\_repeated\_guest   
## 2.698821e+14   
## previous\_cancellations   
## -1.384021e+13   
## previous\_bookings\_not\_canceled   
## 4.995294e+13   
## reserved\_room\_typeB   
## 2.498519e+15   
## reserved\_room\_typeC   
## 2.525762e+15   
## reserved\_room\_typeD   
## 6.246599e+14   
## reserved\_room\_typeE   
## 1.104843e+15   
## reserved\_room\_typeF   
## 5.222617e+15   
## reserved\_room\_typeG   
## 7.406528e+14   
## reserved\_room\_typeH   
## 2.676534e+15   
## assigned\_room\_typeB   
## 4.787397e+14   
## assigned\_room\_typeC   
## 8.887079e+14   
## assigned\_room\_typeD   
## 6.927153e+14   
## assigned\_room\_typeE   
## 5.264769e+14   
## assigned\_room\_typeF   
## 5.074982e+14   
## assigned\_room\_typeG   
## 1.560769e+14   
## assigned\_room\_typeH   
## -8.092147e+13   
## assigned\_room\_typeI   
## 4.966296e+14   
## assigned\_room\_typeK   
## 6.256111e+14   
## booking\_changes   
## 3.727485e+13   
## days\_in\_waiting\_list   
## -1.941049e+11   
## customer\_typeGroup   
## 9.025728e+14   
## customer\_typeTransient   
## 3.560760e+14   
## customer\_typeTransient-Party   
## 4.484622e+14   
## average\_daily\_rate   
## 2.863693e+12   
## required\_car\_parking\_spacesparking   
## 6.586663e+14   
## total\_of\_special\_requests   
## 6.296324e+13   
## mealFB:reserved\_room\_typeB   
## NA   
## mealHB:reserved\_room\_typeB   
## -3.886331e+14   
## mealSC:reserved\_room\_typeB   
## -3.625277e+14   
## mealUndefined:reserved\_room\_typeB   
## NA   
## mealFB:reserved\_room\_typeC   
## -1.657095e+15   
## mealHB:reserved\_room\_typeC   
## -2.440390e+14   
## mealSC:reserved\_room\_typeC   
## -2.982761e+15   
## mealUndefined:reserved\_room\_typeC   
## -1.236603e+15   
## mealFB:reserved\_room\_typeD   
## 1.234026e+15   
## mealHB:reserved\_room\_typeD   
## -7.440971e+14   
## mealSC:reserved\_room\_typeD   
## -2.676844e+15   
## mealUndefined:reserved\_room\_typeD   
## 9.761676e+14   
## mealFB:reserved\_room\_typeE   
## -4.608181e+14   
## mealHB:reserved\_room\_typeE   
## -1.589608e+15   
## mealSC:reserved\_room\_typeE   
## -1.853442e+15   
## mealUndefined:reserved\_room\_typeE   
## -2.548205e+14   
## mealFB:reserved\_room\_typeF   
## -2.035338e+15   
## mealHB:reserved\_room\_typeF   
## -7.988601e+14   
## mealSC:reserved\_room\_typeF   
## -1.371471e+15   
## mealUndefined:reserved\_room\_typeF   
## -2.616674e+15   
## mealFB:reserved\_room\_typeG   
## -2.023297e+15   
## mealHB:reserved\_room\_typeG   
## -9.270201e+14   
## mealSC:reserved\_room\_typeG   
## -2.632502e+15   
## mealUndefined:reserved\_room\_typeG   
## 1.277713e+15   
## mealFB:reserved\_room\_typeH   
## -4.885814e+15   
## mealHB:reserved\_room\_typeH   
## -2.154198e+13   
## mealSC:reserved\_room\_typeH   
## NA   
## mealUndefined:reserved\_room\_typeH   
## NA   
## reserved\_room\_typeB:assigned\_room\_typeB   
## -1.557465e+15   
## reserved\_room\_typeC:assigned\_room\_typeB   
## 1.465351e+14   
## reserved\_room\_typeD:assigned\_room\_typeB   
## -1.445211e+15   
## reserved\_room\_typeE:assigned\_room\_typeB   
## 5.765552e+15   
## reserved\_room\_typeF:assigned\_room\_typeB   
## -8.979064e+14   
## reserved\_room\_typeG:assigned\_room\_typeB   
## 8.001474e+15   
## reserved\_room\_typeH:assigned\_room\_typeB   
## NA   
## reserved\_room\_typeB:assigned\_room\_typeC   
## NA   
## reserved\_room\_typeC:assigned\_room\_typeC   
## 3.870905e+15   
## reserved\_room\_typeD:assigned\_room\_typeC   
## -4.209341e+14   
## reserved\_room\_typeE:assigned\_room\_typeC   
## -2.972519e+14   
## reserved\_room\_typeF:assigned\_room\_typeC   
## NA   
## reserved\_room\_typeG:assigned\_room\_typeC   
## 5.528663e+15   
## reserved\_room\_typeH:assigned\_room\_typeC   
## NA   
## reserved\_room\_typeB:assigned\_room\_typeD   
## 6.244548e+14   
## reserved\_room\_typeC:assigned\_room\_typeD   
## 3.313446e+15   
## reserved\_room\_typeD:assigned\_room\_typeD   
## -1.100168e+15   
## reserved\_room\_typeE:assigned\_room\_typeD   
## -3.131536e+14   
## reserved\_room\_typeF:assigned\_room\_typeD   
## NA   
## reserved\_room\_typeG:assigned\_room\_typeD   
## NA   
## reserved\_room\_typeH:assigned\_room\_typeD   
## -2.665117e+15   
## reserved\_room\_typeB:assigned\_room\_typeE   
## 1.164791e+15   
## reserved\_room\_typeC:assigned\_room\_typeE   
## 2.362162e+15   
## reserved\_room\_typeD:assigned\_room\_typeE   
## 5.171731e+13   
## reserved\_room\_typeE:assigned\_room\_typeE   
## -1.158549e+15   
## reserved\_room\_typeF:assigned\_room\_typeE   
## -3.072317e+14   
## reserved\_room\_typeG:assigned\_room\_typeE   
## 5.734799e+15   
## reserved\_room\_typeH:assigned\_room\_typeE   
## NA   
## reserved\_room\_typeB:assigned\_room\_typeF   
## -7.746544e+14   
## reserved\_room\_typeC:assigned\_room\_typeF   
## 2.712517e+15   
## reserved\_room\_typeD:assigned\_room\_typeF   
## -2.298227e+13   
## reserved\_room\_typeE:assigned\_room\_typeF   
## -1.453710e+15   
## reserved\_room\_typeF:assigned\_room\_typeF   
## 5.103820e+14   
## reserved\_room\_typeG:assigned\_room\_typeF   
## 6.002946e+15   
## reserved\_room\_typeH:assigned\_room\_typeF   
## NA   
## reserved\_room\_typeB:assigned\_room\_typeG   
## 2.911614e+15   
## reserved\_room\_typeC:assigned\_room\_typeG   
## 9.030179e+15   
## reserved\_room\_typeD:assigned\_room\_typeG   
## -7.757586e+13   
## reserved\_room\_typeE:assigned\_room\_typeG   
## -2.167847e+15   
## reserved\_room\_typeF:assigned\_room\_typeG   
## 1.428091e+15   
## reserved\_room\_typeG:assigned\_room\_typeG   
## 4.501701e+15   
## reserved\_room\_typeH:assigned\_room\_typeG   
## 2.502855e+15   
## reserved\_room\_typeB:assigned\_room\_typeH   
## NA   
## reserved\_room\_typeC:assigned\_room\_typeH   
## 5.084263e+15   
## reserved\_room\_typeD:assigned\_room\_typeH   
## 9.914732e+13   
## reserved\_room\_typeE:assigned\_room\_typeH   
## -2.599982e+14   
## reserved\_room\_typeF:assigned\_room\_typeH   
## 1.383247e+15   
## reserved\_room\_typeG:assigned\_room\_typeH   
## 3.631496e+15   
## reserved\_room\_typeH:assigned\_room\_typeH   
## 3.051394e+15   
## reserved\_room\_typeB:assigned\_room\_typeI   
## NA   
## reserved\_room\_typeC:assigned\_room\_typeI   
## 5.045900e+15   
## reserved\_room\_typeD:assigned\_room\_typeI   
## 1.499050e+13   
## reserved\_room\_typeE:assigned\_room\_typeI   
## -1.733832e+15   
## reserved\_room\_typeF:assigned\_room\_typeI   
## 1.517186e+14   
## reserved\_room\_typeG:assigned\_room\_typeI   
## 3.277429e+15   
## reserved\_room\_typeH:assigned\_room\_typeI   
## NA   
## reserved\_room\_typeB:assigned\_room\_typeK   
## -1.399068e+15   
## reserved\_room\_typeC:assigned\_room\_typeK   
## NA   
## reserved\_room\_typeD:assigned\_room\_typeK   
## -1.539321e+15   
## reserved\_room\_typeE:assigned\_room\_typeK   
## -1.922533e+15   
## reserved\_room\_typeF:assigned\_room\_typeK   
## -2.622003e+15   
## reserved\_room\_typeG:assigned\_room\_typeK   
## 3.646695e+15   
## reserved\_room\_typeH:assigned\_room\_typeK   
## NA   
## hotelResort\_Hotel:reserved\_room\_typeB   
## NA   
## hotelResort\_Hotel:reserved\_room\_typeC   
## -2.421784e+15   
## hotelResort\_Hotel:reserved\_room\_typeD   
## -6.212235e+14   
## hotelResort\_Hotel:reserved\_room\_typeE   
## 1.772308e+13   
## hotelResort\_Hotel:reserved\_room\_typeF   
## -3.510272e+15   
## hotelResort\_Hotel:reserved\_room\_typeG   
## -5.376024e+14   
## hotelResort\_Hotel:reserved\_room\_typeH   
## NA   
## market\_segmentComplementary:reserved\_room\_typeB   
## -1.092031e+15   
## market\_segmentCorporate:reserved\_room\_typeB   
## 2.630130e+15   
## market\_segmentDirect:reserved\_room\_typeB   
## -2.504049e+15   
## market\_segmentGroups:reserved\_room\_typeB   
## -3.508738e+15   
## market\_segmentOffline\_TA/TO:reserved\_room\_typeB   
## -6.257062e+14   
## market\_segmentOnline\_TA:reserved\_room\_typeB   
## NA   
## market\_segmentComplementary:reserved\_room\_typeC   
## -4.875634e+15   
## market\_segmentCorporate:reserved\_room\_typeC   
## 1.966726e+15   
## market\_segmentDirect:reserved\_room\_typeC   
## -1.481392e+15   
## market\_segmentGroups:reserved\_room\_typeC   
## -1.683002e+15   
## market\_segmentOffline\_TA/TO:reserved\_room\_typeC   
## -1.437382e+15   
## market\_segmentOnline\_TA:reserved\_room\_typeC   
## NA   
## market\_segmentComplementary:reserved\_room\_typeD   
## -6.130681e+14   
## market\_segmentCorporate:reserved\_room\_typeD   
## 1.364377e+13   
## market\_segmentDirect:reserved\_room\_typeD   
## -9.458363e+14   
## market\_segmentGroups:reserved\_room\_typeD   
## -5.089717e+14   
## market\_segmentOffline\_TA/TO:reserved\_room\_typeD   
## 8.038606e+14   
## market\_segmentOnline\_TA:reserved\_room\_typeD   
## 1.320351e+15   
## market\_segmentComplementary:reserved\_room\_typeE   
## -8.694407e+14   
## market\_segmentCorporate:reserved\_room\_typeE   
## 1.430568e+14   
## market\_segmentDirect:reserved\_room\_typeE   
## -1.497208e+15   
## market\_segmentGroups:reserved\_room\_typeE   
## -5.317258e+14   
## market\_segmentOffline\_TA/TO:reserved\_room\_typeE   
## 7.870310e+13   
## market\_segmentOnline\_TA:reserved\_room\_typeE   
## 1.026677e+15   
## market\_segmentComplementary:reserved\_room\_typeF   
## -3.558147e+15   
## market\_segmentCorporate:reserved\_room\_typeF   
## -1.001084e+15   
## market\_segmentDirect:reserved\_room\_typeF   
## -3.253980e+15   
## market\_segmentGroups:reserved\_room\_typeF   
## -4.820032e+15   
## market\_segmentOffline\_TA/TO:reserved\_room\_typeF   
## -3.628825e+15   
## market\_segmentOnline\_TA:reserved\_room\_typeF   
## NA   
## market\_segmentComplementary:reserved\_room\_typeG   
## -6.861807e+15   
## market\_segmentCorporate:reserved\_room\_typeG   
## -3.662587e+15   
## market\_segmentDirect:reserved\_room\_typeG   
## -3.720594e+15   
## market\_segmentGroups:reserved\_room\_typeG   
## -5.823385e+15   
## market\_segmentOffline\_TA/TO:reserved\_room\_typeG   
## -2.046066e+15   
## market\_segmentOnline\_TA:reserved\_room\_typeG   
## NA   
## market\_segmentComplementary:reserved\_room\_typeH   
## NA   
## market\_segmentCorporate:reserved\_room\_typeH   
## NA   
## market\_segmentDirect:reserved\_room\_typeH   
## -4.870903e+15   
## market\_segmentGroups:reserved\_room\_typeH   
## NA   
## market\_segmentOffline\_TA/TO:reserved\_room\_typeH   
## NA   
## market\_segmentOnline\_TA:reserved\_room\_typeH   
## NA   
## mealFB:is\_repeated\_guest   
## NA   
## mealHB:is\_repeated\_guest   
## -8.510117e+14   
## mealSC:is\_repeated\_guest   
## -9.461369e+14   
## mealUndefined:is\_repeated\_guest   
## -8.695337e+13   
## adults:previous\_bookings\_not\_canceled   
## -1.915090e+13   
## mealFB:previous\_bookings\_not\_canceled   
## 3.347937e+14   
## mealHB:previous\_bookings\_not\_canceled   
## 2.979656e+14   
## mealSC:previous\_bookings\_not\_canceled   
## 4.489150e+13   
## mealUndefined:previous\_bookings\_not\_canceled   
## 1.979878e+15   
## market\_segmentComplementary:customer\_typeGroup   
## -3.195503e+15   
## market\_segmentCorporate:customer\_typeGroup   
## 1.872189e+15   
## market\_segmentDirect:customer\_typeGroup   
## -1.948419e+14   
## market\_segmentGroups:customer\_typeGroup   
## 1.367252e+14   
## market\_segmentOffline\_TA/TO:customer\_typeGroup   
## -1.203849e+15   
## market\_segmentOnline\_TA:customer\_typeGroup   
## NA   
## market\_segmentComplementary:customer\_typeTransient   
## -7.133043e+14   
## market\_segmentCorporate:customer\_typeTransient   
## 5.708077e+13   
## market\_segmentDirect:customer\_typeTransient   
## 8.397086e+14   
## market\_segmentGroups:customer\_typeTransient   
## 2.100600e+15   
## market\_segmentOffline\_TA/TO:customer\_typeTransient   
## 7.922138e+14   
## market\_segmentOnline\_TA:customer\_typeTransient   
## -3.296102e+13   
## market\_segmentComplementary:customer\_typeTransient-Party   
## -1.936333e+15   
## market\_segmentCorporate:customer\_typeTransient-Party   
## -2.407040e+14   
## market\_segmentDirect:customer\_typeTransient-Party   
## 1.166229e+15   
## market\_segmentGroups:customer\_typeTransient-Party   
## 2.420970e+15   
## market\_segmentOffline\_TA/TO:customer\_typeTransient-Party   
## 8.309730e+13   
## market\_segmentOnline\_TA:customer\_typeTransient-Party   
## NA   
## is\_repeated\_guest:assigned\_room\_typeB   
## -2.237900e+14   
## is\_repeated\_guest:assigned\_room\_typeC   
## -1.910581e+15   
## is\_repeated\_guest:assigned\_room\_typeD   
## -3.618319e+14   
## is\_repeated\_guest:assigned\_room\_typeE   
## -2.881088e+14   
## is\_repeated\_guest:assigned\_room\_typeF   
## -3.116607e+14   
## is\_repeated\_guest:assigned\_room\_typeG   
## -7.681218e+14   
## is\_repeated\_guest:assigned\_room\_typeH   
## -8.367698e+14   
## is\_repeated\_guest:assigned\_room\_typeI   
## -6.491084e+14   
## is\_repeated\_guest:assigned\_room\_typeK   
## -5.660068e+13   
## assigned\_room\_typeB:required\_car\_parking\_spacesparking   
## -6.896684e+14   
## assigned\_room\_typeC:required\_car\_parking\_spacesparking   
## -4.986269e+14   
## assigned\_room\_typeD:required\_car\_parking\_spacesparking   
## 1.594190e+14   
## assigned\_room\_typeE:required\_car\_parking\_spacesparking   
## -7.327335e+14   
## assigned\_room\_typeF:required\_car\_parking\_spacesparking   
## -4.339228e+14   
## assigned\_room\_typeG:required\_car\_parking\_spacesparking   
## -7.610148e+14   
## assigned\_room\_typeH:required\_car\_parking\_spacesparking   
## -4.043553e+14   
## assigned\_room\_typeI:required\_car\_parking\_spacesparking   
## -2.417985e+14   
## assigned\_room\_typeK:required\_car\_parking\_spacesparking   
## -7.815233e+14

We rule out all the variables and interactions with **NA**, since they are non-converged.

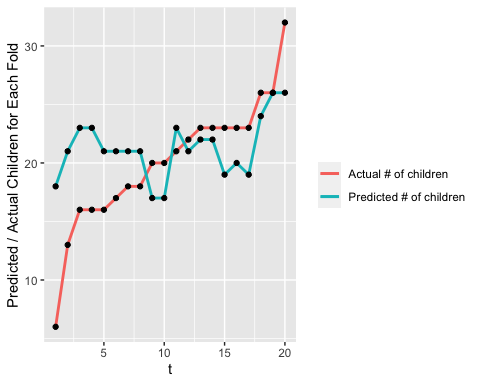
## 4.3 Out-of-sample performance evaluation

## [,1] [,2] [,3] [,4]   
## measurement "Deviance" "TPR" "FPR" "FDR"   
## eval\_baseline1 "3291.583" "0" "0" "NaN"   
## eval\_baseline2 "2334.185" "0.344" "0.013" "0.304"  
## eval\_lasso\_selected "2311.875" "0.378" "0.012" "0.273"

## 4.4 Model Validation: Step 1



## 4.5 Model Validation: Step 2



From the plot, the difference of each fold for actual data and predicted model is small, which is almost smaller than 5. Therefore, our model do well at predicting the total number of bookings with children in a group of 250 bookings.