Final Project中Logistic Regression，Bagging，Boosting的备忘录

1. 记得把data exploration里面的关于variables的description写清楚
2. Research question里面把具体我们想要用哪些method写清楚
3. 还有abstract里面对于project的introduction也改一下，根据实际用的method，还有对游戏的稍加详细的介绍。
4. 改动一下Analysis Plan，根据我们实际用的method和plot改
5. 把reference填了
6. 所有引用的library可以都放在Reference下面那个chunk里
7. 记得每换一种新的method重新导入一下training和test dataset，具体可参考我bagging和boosting里面怎么做的，这样防止如果在某个method里面改动了某个column或者row不会影响到其他method的正常运行
8. 我做的Logistic和Bagging的应该没有问题，error rate看着都正常。但是需要订正一下每个每个function里面的context。
   1. logistics regression部分，glm.fit <- glm(…)中的family = binomial，需要查一下针对我们的dataset他是不是应该是binomial。
   2. 对于coefficient interpretation部分可以举例对significant的variables再稍加interpret
   3. Bagging部分，randomForest function里mtry、ntree具体值的选用，ntree我选500的原因是因为在验证过不加ntree项和ntree=700后，ntree=500得到的error rate最低。
   4. 另外bagging的plot下面那个legend我是完全按照lab写的，你们看一下如果需要改就改一下，从目前给的图我觉得还行。
   5. Boosting部分，gbm function里的“Bernoulli distribution“应该是对的，因为Bernoulli是binary classification problem，gaussion是regression problem，这个你们再商讨研究一下。因为我现在boosting按照lab的方式作出来得到的error rate=0.9997。这肯定是不对的，你们看一下有没有办法改，没有的话可以把boosting这个部分删了，不用了。（我试过将gbm里面的Bernoulli改为gaussion了，error rate还是一样的）
9. 很多chunk可以选择include=FALSE或者eval=FALSE这些指令来在knit后的PDF里隐藏code或者coding的output，这个lab0里面有各种各样的指令，有些不需要加入的code或者output可以隐藏掉。你们记得改一下
10. 最好的是，中午前你俩弄完能去一下TA的office hour问一下我提到的一些问题或者你们自己的部分做出来的问题。周三早上十点和下午五点六点都有office hour。
11. 总之目前我做出来的logistic regression和bagging的都是可以用的，boosting不知道是哪个地方出问题了，可以office hour问一下。实在不行就先不用了，如果不用了记得先找个备用rmd文件把那些coding写存起来。你们做的其他部分如果也遇到我boosting遇到的问题，你俩商量下，没办法解决的话可以先不用了。