MA678 Midterm Project

Shuai Jiang

1.Introduction

For P&C side insurance industry, ratemaking is an important process with respect to pricing. Given a dataset and the credibility information of the dataset, actuarial practitioners are responsible to pull the risk information along with the dataset and design the insurance products properly. Statistical models, in this case, are very helpful because generalized linear model or even more complicated models are able to tell which predictors are taking more significant roles and making more obvious impacts to the events causing insured make claims.

This project's dataset is provided by a Singaporean third-party travel insurance servicing company. The dataset contains standard information for each insurance policy and decent amount of observations (63326 records). The binary claim status outcome (Yes or No) is indicated to be target, and other information is assumed to be influential to claim status.

In order to figure out which predictors are making more significant impacts on the claim status, in this project, we firstly use GLM to check the overall significance based on each predictor. Then we selected some predictors based on it and EDA results, also considering the result generated by stepAIC. Then, we explore the influences made by destination as a grouping-criteria by using GLMER. We will apply some model-check tools and decide if the models are improving.

2.Data Cleaning

By using 'str', there's some problematic data needed to be fixed:

-The age that are older than 114 are deleted since the number looks not realistic.

-The travel duration of negative sign is deleted because the negative duration is impossible.

-There are 149 countries as travel destinations. But only about 25% of the countries are with travelers' frequency larger than 100. Due to the extreme distribution of the travel frequency among different countries, 10 countries with top travel insurance frequencies are selected.

```
str(ins)
```

```
## 'data.frame':    63326 obs. of  11 variables:
##  $ Agency              : Factor w/ 16 levels "ADM","ART","C2B",..: 4 4 7 7
 7 10 7 7 7 7 ...
##  $ Agency.Type         : Factor w/ 2 levels "Airlines","Travel Agency": 2
2 2 2 2 1 2 2 2 2 ...
##  $ Distribution.Channel: Factor w/ 2 levels "Offline","Online": 1 1 2 2 2
2 2 2 2 2 ...
##  $ Product.Name        : Factor w/ 26 levels "1 way Comprehensive Plan",..
: 13 13 17 17 17 26 17 17 17 17 ...
##  $ Claim               : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1
1 ...
##  $ Duration            : int  186 186 65 60 79 66 47 63 57 186 ...
##  $ Destination         : Factor w/ 149 levels "ALBANIA","ANGOLA",..: 80 80
 5 5 62 141 141 5 130 5 ...
##  $ Net.Sales           : num  -29 -29 -49.5 -39.6 -19.8 ...
##  $ Commision..in.value.: num  9.57 9.57 29.7 23.76 11.88 ...
##  $ Gender              : Factor w/ 3 levels "","F","M": 2 2 1 1 1 2 1 1 1
1 ...
##  $ Age                 : int  81 71 32 32 41 44 32 29 44 37 ...
```

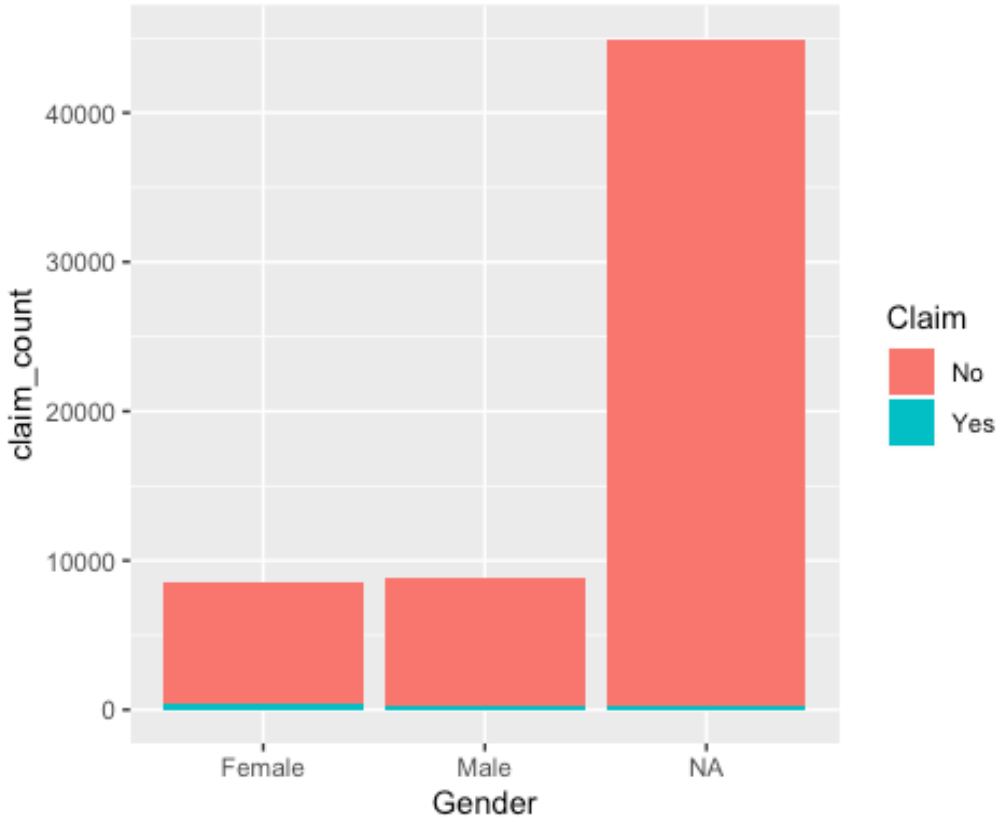
```
str(newins)
```

```
## 'data.frame':    45773 obs. of  11 variables:
##  $ Destination: Factor w/ 149 levels "ALBANIA","ANGOLA",..: 80 80 5 5 141
141 5 130 5 80 ...
##  $ Claim      : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Duration   : int  186 186 65 60 66 47 63 57 186 1 ...
##  $ Age        : int  81 71 32 32 44 32 29 44 37 47 ...
```

After filtering destination (countries), there are 45773 observations left.

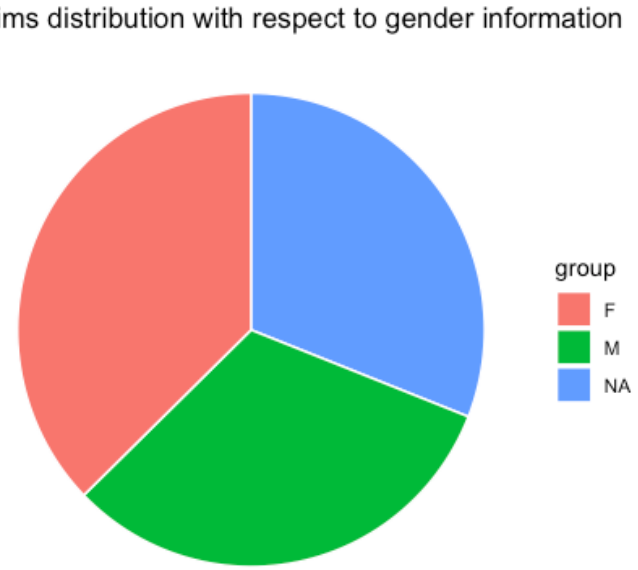
3.EDA

Figure 1. Distribution of total claim status with regard to gender information of 63326 observations.



From figure 1, we see more than 2/3 of the insured travelers didn't specify their gender, and it seems questionable. However, in practice, the risk item without specifying gender is also a piece of information.

Figure 2.



For insured who made claim, approximately, 1/3 of them are female, 1/3 of them are male, 1/3 of them didn't specify their gender.

Figure 3.

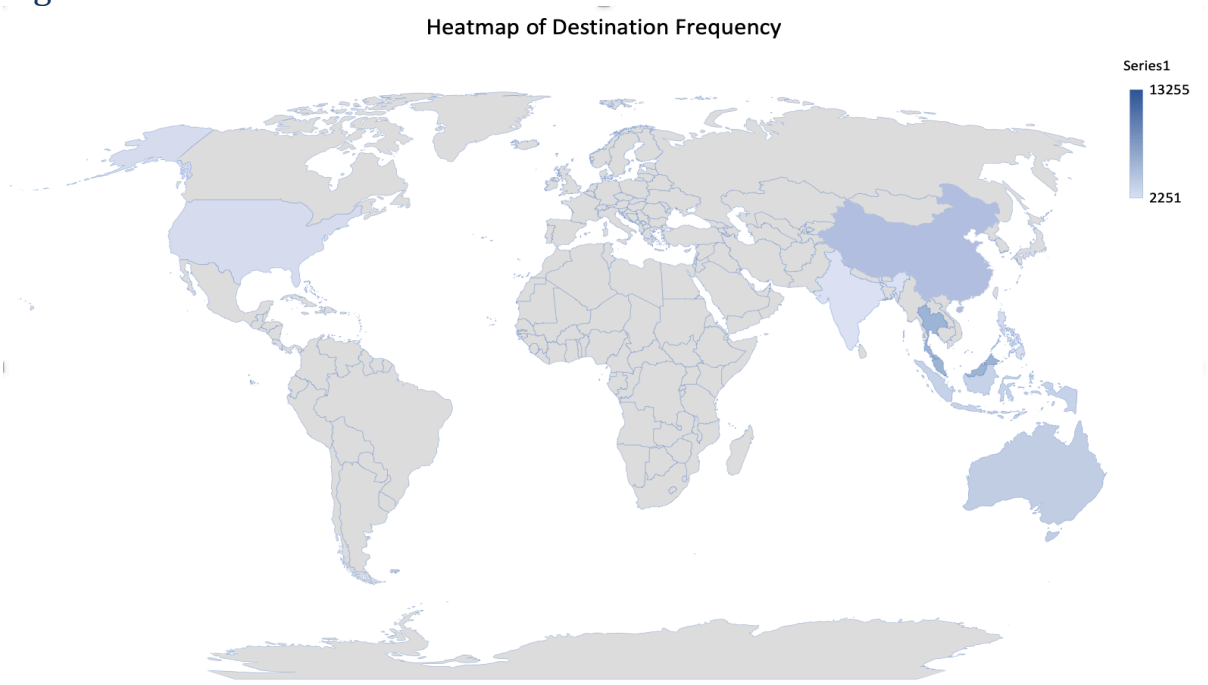


Figure 3 displays the 10 destination countries with travel insurance frequency.

Figure 4.

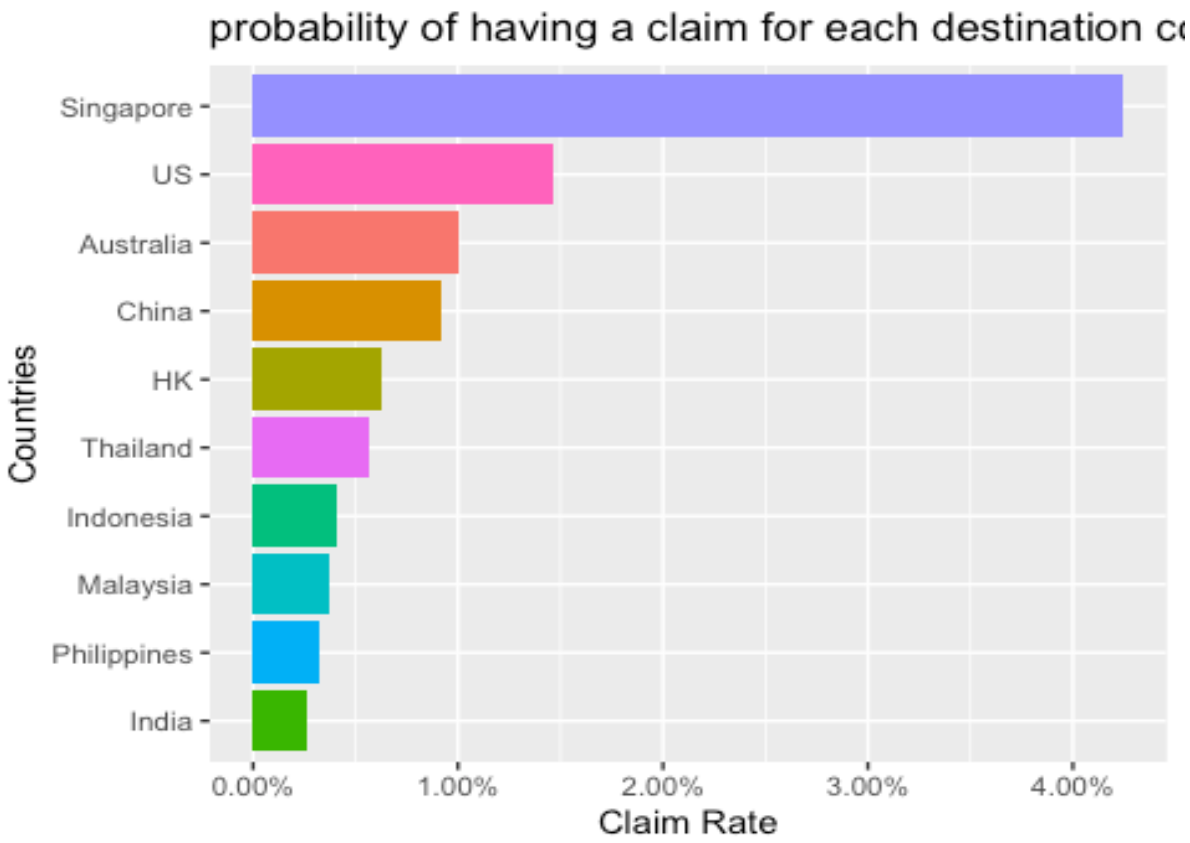


Figure 4 shows the claim rate for each country selected by travel insurance frequency. Singapore is with highest claim rate.

Figure 5. Distribution of insured travelers' age.

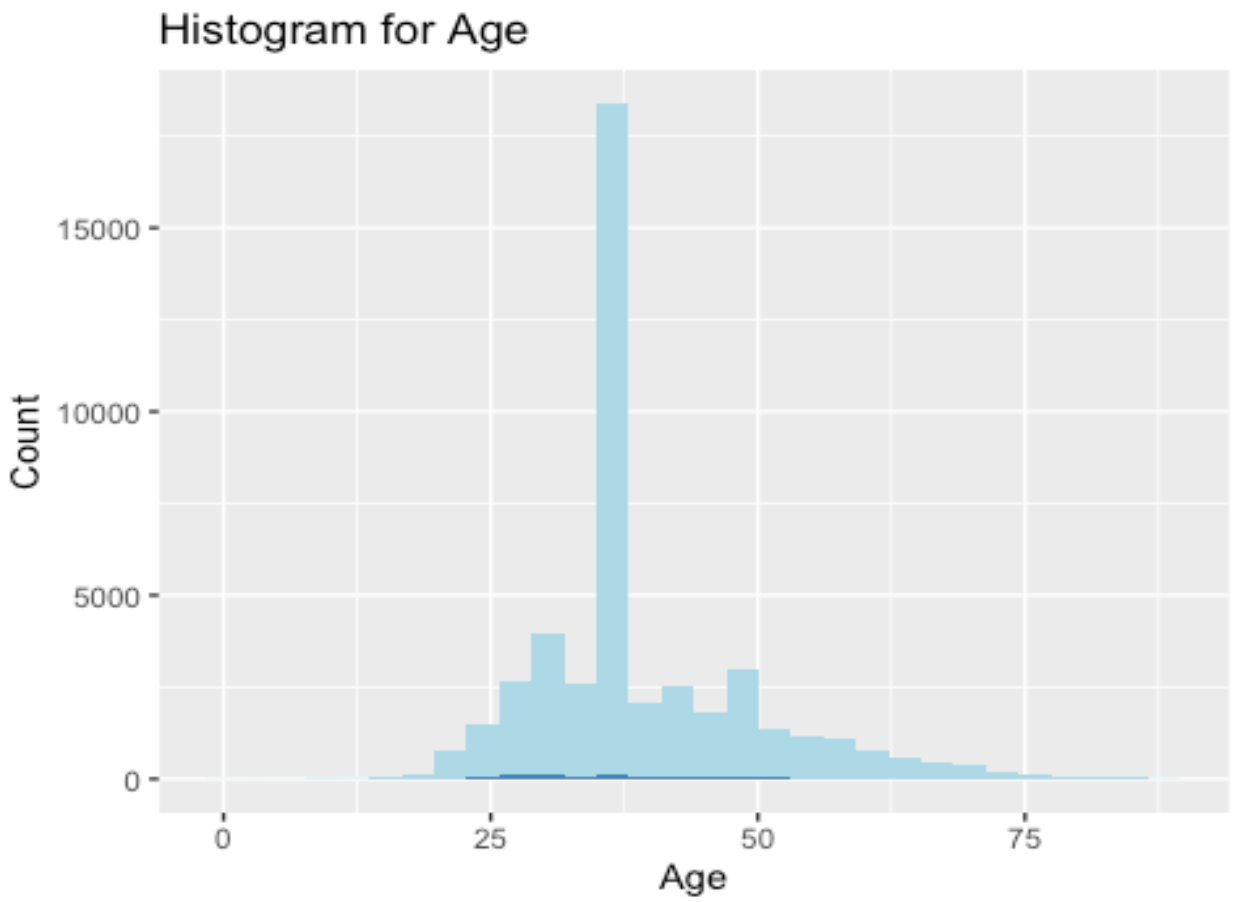


Figure 5 indicates that most travelers with travel insurance are within the age range from 25 years old to 50 years old.

The dark blue area within the histogram represents the count of claim made.

Figure 6. Claim count distribution country by country.

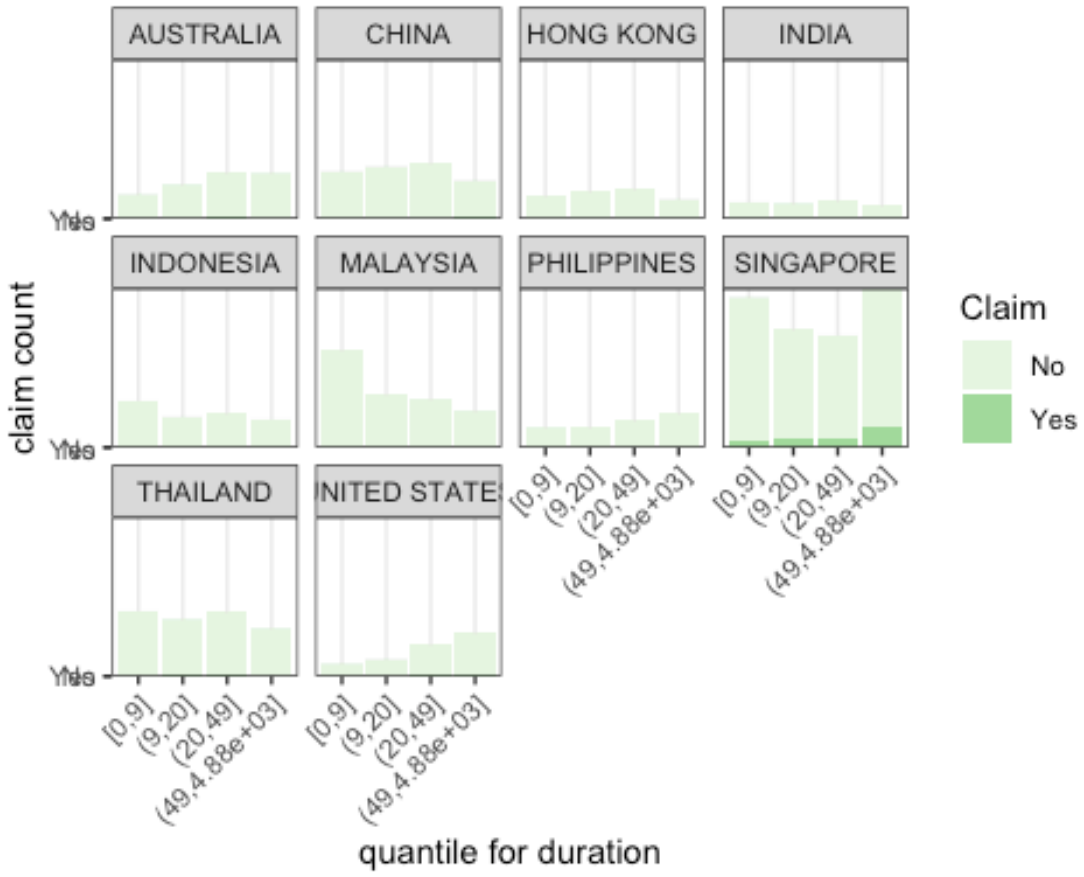


Figure 6's Y axis represents the claim count. The light green means that the claim status is "No", and the dark green means that the claim status is "Yes".

Figure 6's X axis represents the quantile for travel duration. The measurement for travel duration is day.

There might be some patterns, but the actual claims are little compared to the overall policy amount. Therefore, we may need to use figure 7. to obtain some ideas about the data.

Figure 7. Claim Percentage for each country based on Travel duration.

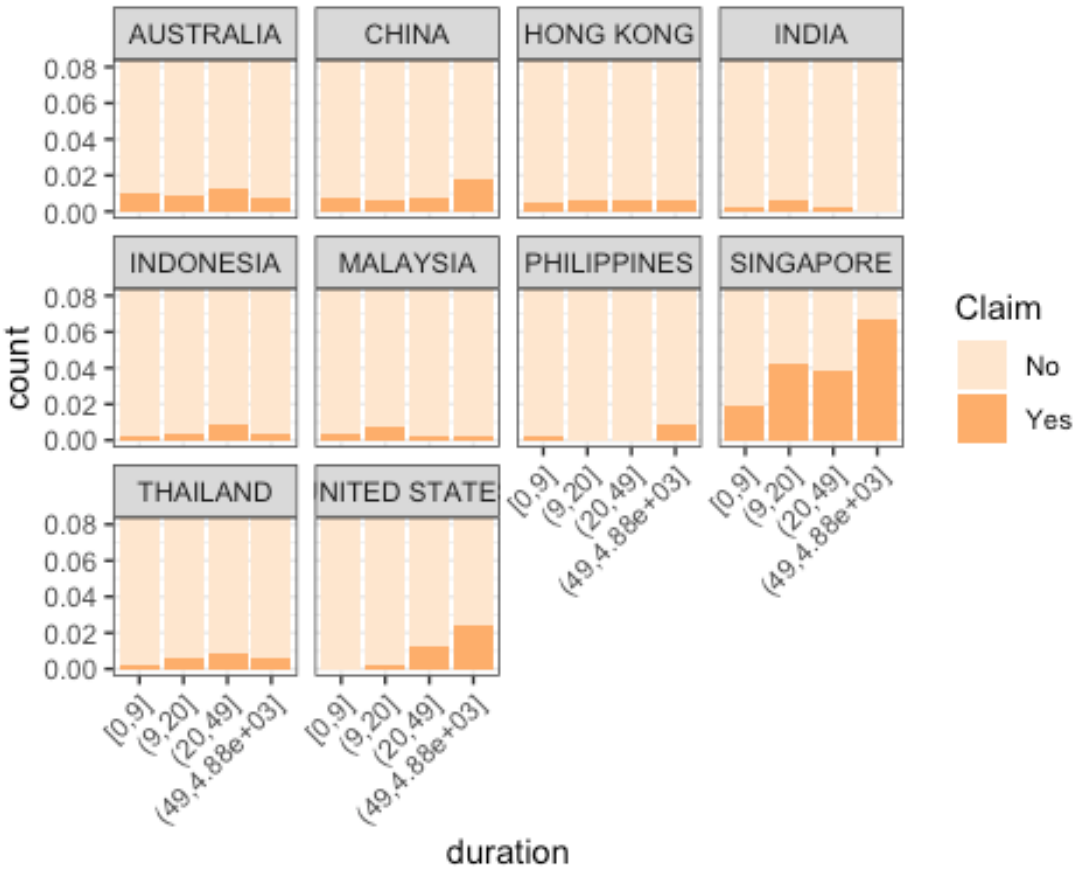
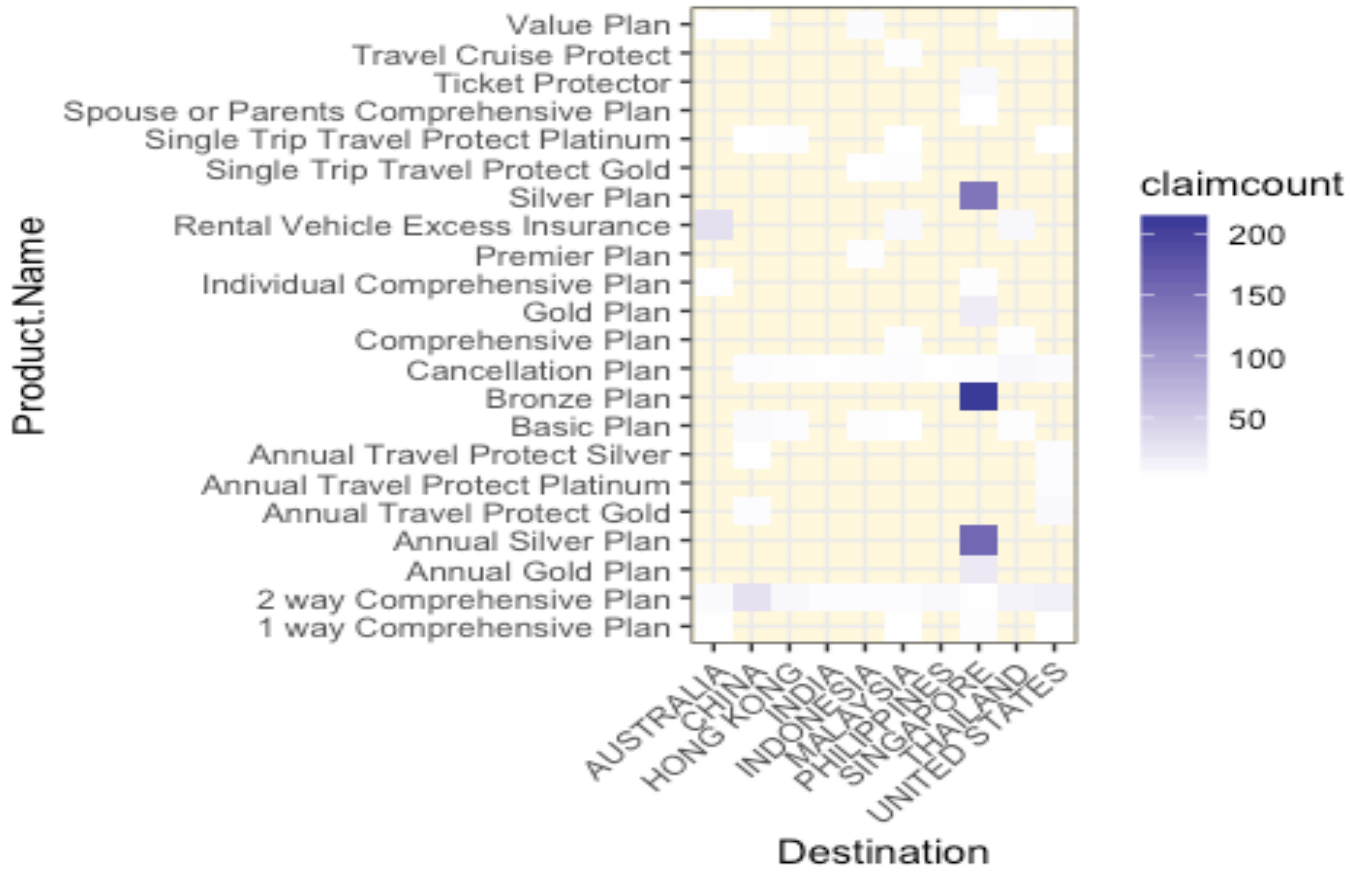


Figure 7 shows that for Singapore and the US, there's a tendency that as travel duration extends, the probability of occurring a claim is bigger. But other countries don't display the same obvious pattern.

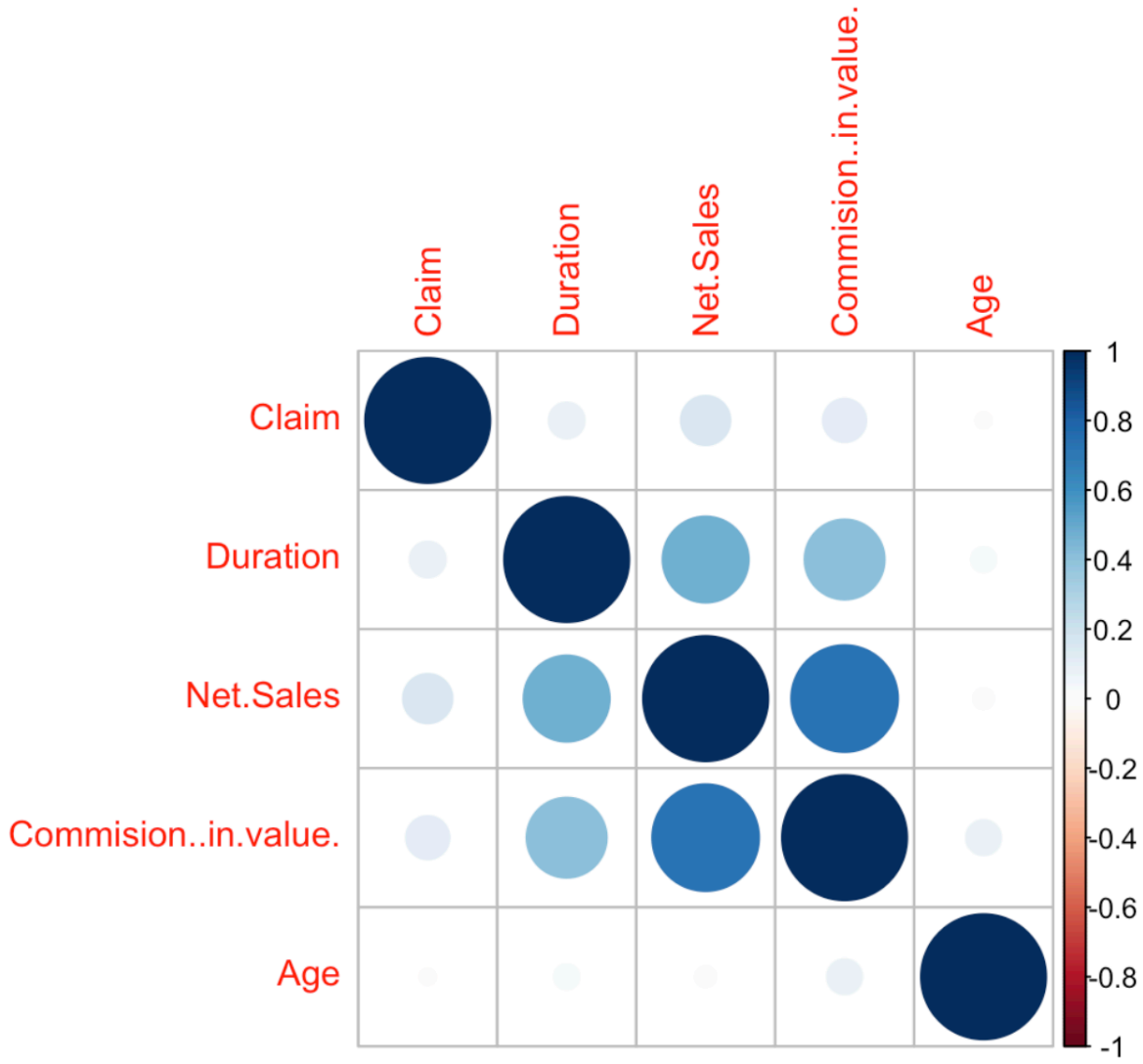
Besides, the claim percentages are very low in general.

Figure 8. Distribution of Product.Name among each Country.



Within figure 8, most block are white means that for the given country and product, only a few (possibly no more than 10) claims occurred.

Figure 9. Correlation plot for continuous variables



In figure 9, we can observe that variable (Net.Sales) and variable (Commision..in.value.) are strongly corelated. We may remove one of these two variables when we build models.

4.Model Used

```
md0 <- glm(Claim ~ Agency.Type+Distribution.Channel+Destination+Gender+Produc
t.Name+ Duration + Age+Net.Sales+Commision..in.value., data = newins2, family
=binomial(link="logit"))
display(md0)
```

```
## glm(formula = Claim ~ Agency.Type + Distribution.Channel + Destination +
##     Gender + Product.Name + Duration + Age + Net.Sales + Commision..in.val
ue.,
##     family = binomial(link = "logit"), data = newins2)
##                                                coef.est coef.se
## (Intercept)                                      -6.14     1.30
## Agency.TypeTravel Agency                          1.73     0.75
## Distribution.ChannelOnline                       -1.25     0.92
## DestinationCHINA                                  0.19     0.28
## DestinationHONG KONG                             -0.16     0.35
## DestinationINDIA                                 -0.74     0.55
## DestinationINDONESIA                             -0.84     0.40
## DestinationMALAYSIA                              -0.73     0.29
## DestinationPHILIPPINES                           -0.65     0.42
## DestinationSINGAPORE                             -1.15     0.44
## DestinationTHAILAND                              -0.35     0.27
## DestinationUNITED STATES                          0.10     0.28
## GenderF                                           0.28     2.37
## GenderM                                           0.30     2.38
## Product.Name2 way Comprehensive Plan              1.30     0.47
## Product.Name24 Protect                           -8.54   172.39
## Product.NameAnnual Gold Plan                      4.98     2.59
## Product.NameAnnual Silver Plan                    5.60     2.58
## Product.NameAnnual Travel Protect Gold            3.47     2.57
## Product.NameAnnual Travel Protect Platinum        2.83     2.69
## Product.NameAnnual Travel Protect Silver          2.73     2.57
## Product.NameBasic Plan                            1.89     2.55
## Product.NameBronze Plan                           5.45     2.55
## Product.NameCancellation Plan                    -0.04     0.49
## Product.NameChild Comprehensive Plan            -11.85   789.89
## Product.NameComprehensive Plan                    0.60     2.49
## Product.NameGold Plan                             5.18     2.56
## Product.NameIndividual Comprehensive Plan         1.54     2.54
## Product.NamePremier Plan                          2.05     2.47
## Product.NameRental Vehicle Excess Insurance       1.41     0.53
## Product.NameSilver Plan                           5.57     2.55
## Product.NameSingle Trip Travel Protect Gold       2.34     2.49
## Product.NameSingle Trip Travel Protect Platinum   3.75     2.47
## Product.NameSingle Trip Travel Protect Silver   -10.61   259.76
## Product.NameSpouse or Parents Comprehensive Plan  2.22     2.68
## Product.NameTicket Protector                      3.92     1.01
## Product.NameTravel Cruise Protect                -0.66     2.53
## Product.NameTravel Cruise Protect Family        -11.23  2399.55
```

```
## Product.NameValue Plan                                    2.74     2.47
## Duration                                                  0.00     0.00
## Age                                                      -0.01     0.00
## Net.Sales                                                 0.01     0.00
## Commision..in.value.                                     -0.01     0.01
## ---
##   n = 45773, k = 43
##   residual deviance = 6480.0, null deviance = 7866.7 (difference = 1386.8)
```

```
step = stepAIC(md0,direction = "both",trace = 0)
step$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## Claim ~ Agency.Type + Distribution.Channel + Destination + Gender +
##     Product.Name + Duration + Age + Net.Sales + Commision..in.value.
##
## Final Model:
## Claim ~ Agency.Type + Destination + Product.Name + Age + Net.Sales +
##     Commision..in.value.
##
##
##                      Step Df   Deviance Resid. Df Resid. Dev      AIC
## 1                                           45730   6479.984 6565.984
## 2               - Gender  2 0.04530269     45732   6480.029 6562.029
## 3             - Duration  1 0.42546263     45733   6480.455 6560.455
## 4 - Distribution.Channel  1 1.58063544     45734   6482.035 6560.035
```

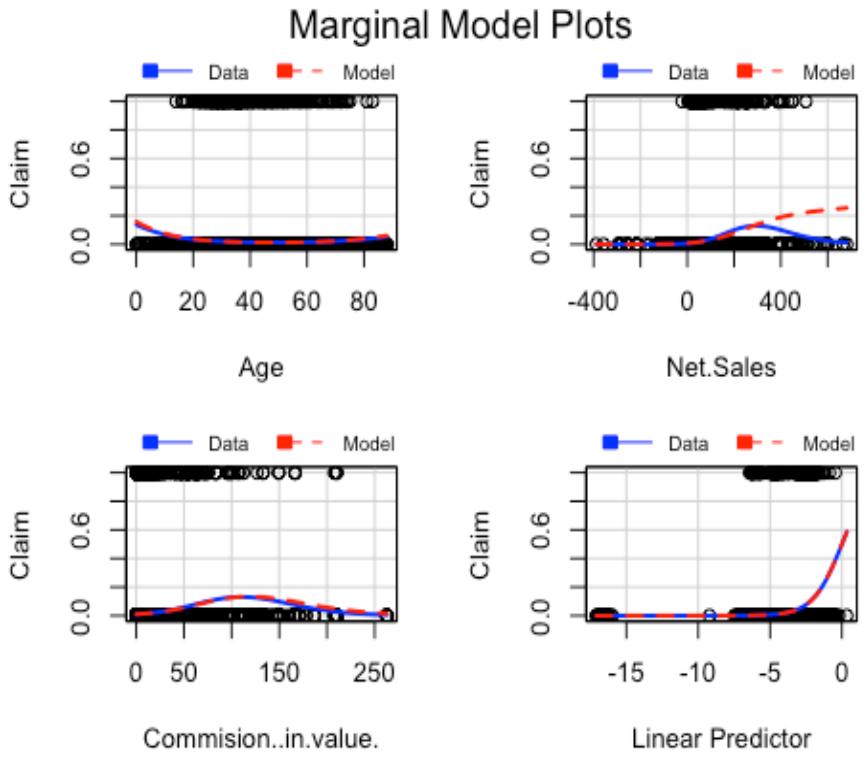
From the summary of md0, we see that the variables of statistical significance includes: Agency.Type, Destination, Product.Name, Age and Net.Sales.

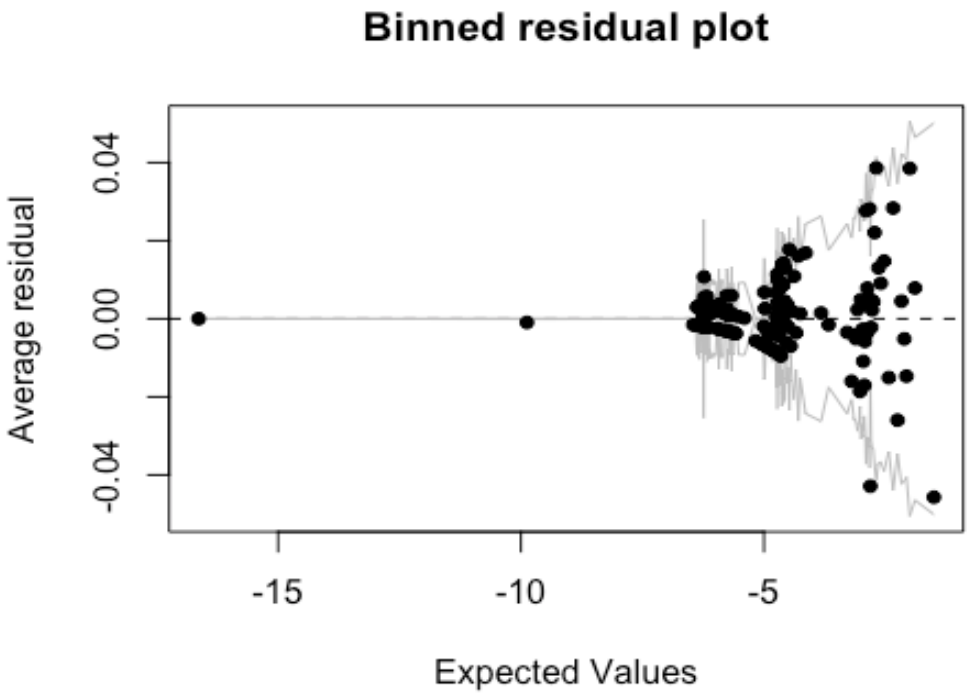
From the stepAIC function, the model selects Agency.Type, Destination, Product.Name, Age, Net.Sales and Commision..in.value as predictors.

According to the information above, we want to select some of the variables in or der to improve the model.

```r
md1 <- glm(Claim ~ Agency.Type+Product.Name + Age+Net.Sales+Commision..in.val
ue., data = newins2, family=binomial(link="logit"))


md2 <- glm(Claim ~ Agency.Type+Product.Name+ Age+ Net.Sales, data = newins2,
family=binomial(link="logit"))
summary(md2)

##
## Call:
## glm(formula = Claim ~ Agency.Type + Product.Name + Age + Net.Sales,
##     family = binomial(link = "logit"), data = newins2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3639  -0.1424  -0.1176  -0.0638   3.5732
##
## Coefficients:
##                                                Estimate Std. Error
## (Intercept)                                  -7.058e+00  8.131e-01
## Agency.TypeTravel Agency                      9.503e-01  6.693e-01
## Product.Name2 way Comprehensive Plan          1.533e+00  4.607e-01
## Product.Name24 Protect                       -9.293e+00  1.729e+02
## Product.NameAnnual Gold Plan                  3.453e+00  8.621e-01
## Product.NameAnnual Silver Plan                4.113e+00  8.162e-01
## Product.NameAnnual Travel Protect Gold        2.928e+00  5.776e-01
## Product.NameAnnual Travel Protect Platinum    1.912e+00  7.722e-01
## Product.NameAnnual Travel Protect Silver      2.321e+00  6.920e-01
## Product.NameBasic Plan                        1.514e+00  8.491e-01
## Product.NameBronze Plan                       4.223e+00  8.080e-01
## Product.NameCancellation Plan                 4.716e-02  4.888e-01
## Product.NameChild Comprehensive Plan         -1.074e+01  7.997e+02
## Product.NameComprehensive Plan                1.989e+00  6.753e-01
## Product.NameGold Plan                         3.912e+00  8.424e-01
## Product.NameIndividual Comprehensive Plan     2.775e+00  7.420e-01
## Product.NamePremier Plan                      2.489e+00  8.450e-01
## Product.NameRental Vehicle Excess Insurance   1.401e+00  4.723e-01
## Product.NameSilver Plan                       4.333e+00  8.094e-01
## Product.NameSingle Trip Travel Protect Gold   2.568e+00  7.367e-01
## Product.NameSingle Trip Travel Protect Platinum  3.960e+00  6.496e-01
## Product.NameSingle Trip Travel Protect Silver   -1.034e+01  2.615e+02
## Product.NameSpouse or Parents Comprehensive Plan  3.353e+00  1.129e+00
## Product.NameTicket Protector                  2.376e+00  8.918e-01
## Product.NameTravel Cruise Protect             7.417e-01  8.389e-01
## Product.NameTravel Cruise Protect Family     -1.023e+01  2.400e+03
## Product.NameValue Plan                        2.558e+00  6.717e-01
## Age                                          -7.161e-03  3.201e-03
## Net.Sales                                     5.047e-03  6.552e-04
##                                              z value Pr(>|z|)
## (Intercept)                                   -8.680  < 2e-16 ***
```

```
## Agency.TypeTravel Agency                                1.420 0.155631
## Product.Name2 way Comprehensive Plan                    3.328 0.000876 ***
## Product.Name24 Protect                                 -0.054 0.957129
## Product.NameAnnual Gold Plan                            4.005 6.19e-05 ***
## Product.NameAnnual Silver Plan                          5.039 4.69e-07 ***
## Product.NameAnnual Travel Protect Gold                  5.069 4.00e-07 ***
## Product.NameAnnual Travel Protect Platinum              2.476 0.013287 *
## Product.NameAnnual Travel Protect Silver                3.354 0.000796 ***
## Product.NameBasic Plan                                  1.783 0.074606 .
## Product.NameBronze Plan                                 5.227 1.73e-07 ***
## Product.NameCancellation Plan                           0.096 0.923148
## Product.NameChild Comprehensive Plan                   -0.013 0.989285
## Product.NameComprehensive Plan                          2.946 0.003224 **
## Product.NameGold Plan                                   4.644 3.42e-06 ***
## Product.NameIndividual Comprehensive Plan               3.740 0.000184 ***
## Product.NamePremier Plan                                2.946 0.003223 **
## Product.NameRental Vehicle Excess Insurance             2.967 0.003007 **
## Product.NameSilver Plan                                 5.353 8.64e-08 ***
## Product.NameSingle Trip Travel Protect Gold             3.485 0.000492 ***
## Product.NameSingle Trip Travel Protect Platinum         6.096 1.09e-09 ***
## Product.NameSingle Trip Travel Protect Silver          -0.040 0.968467
## Product.NameSpouse or Parents Comprehensive Plan        2.970 0.002978 **
## Product.NameTicket Protector                            2.664 0.007716 **
## Product.NameTravel Cruise Protect                       0.884 0.376635
## Product.NameTravel Cruise Protect Family               -0.004 0.996599
## Product.NameValue Plan                                  3.809 0.000140 ***
## Age                                                    -2.237 0.025303 *
## Net.Sales                                               7.703 1.33e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7866.7  on 45772  degrees of freedom
## Residual deviance: 6515.5  on 45744  degrees of freedom
## AIC: 6573.5
##
## Number of Fisher Scoring iterations: 15

library(arm)
library(alr3)
binnedplot(predict(md1),resid(md1,type = "response"))
```





The observations are locating around y = 0. Most of the points are within the interval.

Try to build md3 without pooling.

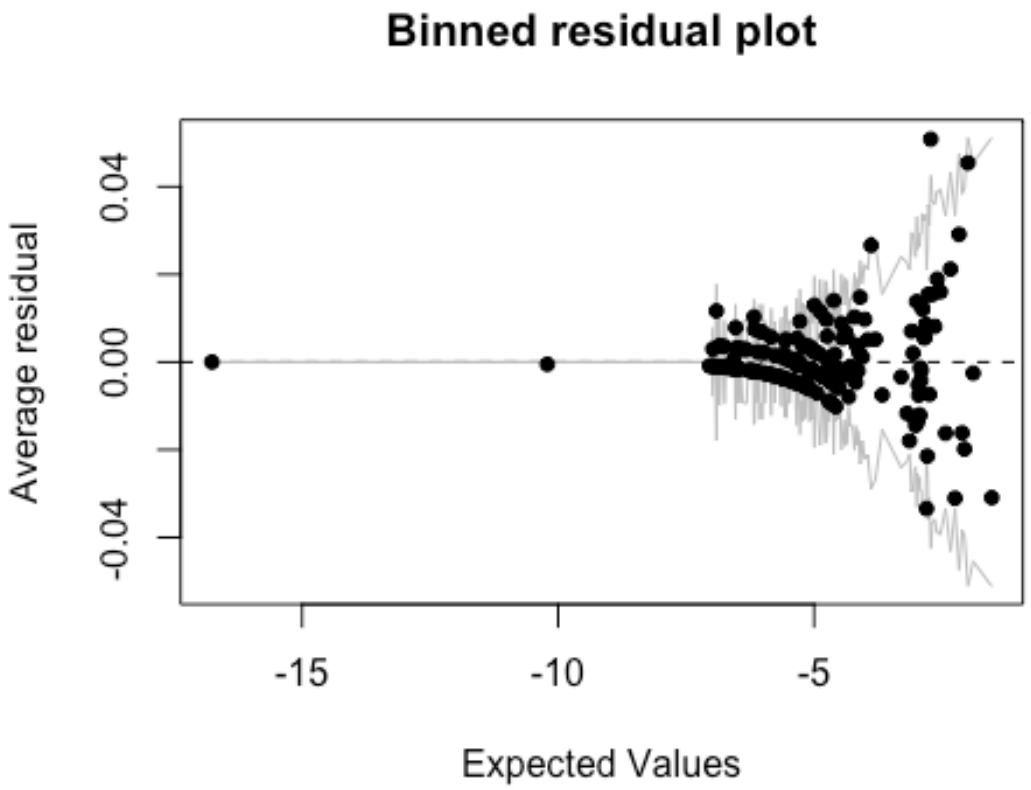
```
md3<- glm(Claim ~ Agency.Type+Product.Name+ Age+ Net.Sales + factor(Destinati
on), data = newins2, family=binomial(link="logit"))
summary(md3)
```

```
##
## Call:
## glm(formula = Claim ~ Agency.Type + Product.Name + Age + Net.Sales +
##     factor(Destination), family = binomial(link = "logit"), data = newins2
)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2845  -0.1602  -0.1001  -0.0633   3.7347
##
## Coefficients:
##                                                Estimate Std. Error
## (Intercept)                                  -7.484e+00  8.994e-01
## Agency.TypeTravel Agency                      1.861e+00  7.542e-01
## Product.Name2 way Comprehensive Plan          1.337e+00  4.660e-01
## Product.Name24 Protect                       -8.227e+00  1.723e+02
## Product.NameAnnual Gold Plan                  5.101e+00  9.871e-01
## Product.NameAnnual Silver Plan                5.720e+00  9.456e-01
## Product.NameAnnual Travel Protect Gold        2.365e+00  5.941e-01
## Product.NameAnnual Travel Protect Platinum    1.375e+00  7.889e-01
## Product.NameAnnual Travel Protect Silver      1.760e+00  7.036e-01
## Product.NameBasic Plan                        2.194e+00  9.244e-01
## Product.NameBronze Plan                       5.775e+00  9.363e-01
## Product.NameCancellation Plan                -5.567e-02  4.908e-01
## Product.NameChild Comprehensive Plan         -1.039e+01  7.906e+02
## Product.NameComprehensive Plan                1.990e+00  6.924e-01
## Product.NameGold Plan                         5.488e+00  9.668e-01
## Product.NameIndividual Comprehensive Plan     3.063e+00  7.652e-01
## Product.NamePremier Plan                      2.766e+00  8.786e-01
## Product.NameRental Vehicle Excess Insurance   1.117e+00  4.960e-01
## Product.NameSilver Plan                       5.892e+00  9.377e-01
## Product.NameSingle Trip Travel Protect Gold   2.376e+00  7.425e-01
## Product.NameSingle Trip Travel Protect Platinum  3.774e+00  6.583e-01
## Product.NameSingle Trip Travel Protect Silver   -1.056e+01  2.591e+02
## Product.NameSpouse or Parents Comprehensive Plan  3.701e+00  1.148e+00
## Product.NameTicket Protector                  3.921e+00  1.009e+00
## Product.NameTravel Cruise Protect             7.455e-01  8.534e-01
## Product.NameTravel Cruise Protect Family     -1.003e+01  2.400e+03
## Product.NameValue Plan                        2.979e+00  7.497e-01
## Age                                          -7.370e-03  3.204e-03
## Net.Sales                                     4.763e-03  6.661e-04
## factor(Destination)CHINA                      2.326e-01  2.781e-01
## factor(Destination)HONG KONG                 -1.443e-01  3.494e-01
## factor(Destination)INDIA                     -7.161e-01  5.536e-01
```

```
## factor(Destination)INDONESIA                          -8.901e-01  4.055e-01
## factor(Destination)MALAYSIA                           -6.822e-01  2.901e-01
## factor(Destination)PHILIPPINES                        -6.591e-01  4.235e-01
## factor(Destination)SINGAPORE                          -1.107e+00  4.397e-01
## factor(Destination)THAILAND                           -3.153e-01  2.749e-01
## factor(Destination)UNITED STATES                       1.534e-01  2.843e-01
##                                                      z value Pr(>|z|)
## (Intercept)                                           -8.321  < 2e-16 ***
## Agency.TypeTravel Agency                               2.468 0.013589 *
## Product.Name2 way Comprehensive Plan                   2.870 0.004109 **
## Product.Name24 Protect                                -0.048 0.961922
## Product.NameAnnual Gold Plan                           5.168 2.37e-07 ***
## Product.NameAnnual Silver Plan                         6.049 1.46e-09 ***
## Product.NameAnnual Travel Protect Gold                 3.981 6.85e-05 ***
## Product.NameAnnual Travel Protect Platinum             1.743 0.081356 .
## Product.NameAnnual Travel Protect Silver               2.501 0.012372 *
## Product.NameBasic Plan                                 2.373 0.017637 *
## Product.NameBronze Plan                                6.168 6.92e-10 ***
## Product.NameCancellation Plan                         -0.113 0.909694
## Product.NameChild Comprehensive Plan                  -0.013 0.989512
## Product.NameComprehensive Plan                         2.873 0.004060 **
## Product.NameGold Plan                                  5.676 1.38e-08 ***
## Product.NameIndividual Comprehensive Plan              4.003 6.26e-05 ***
## Product.NamePremier Plan                               3.148 0.001645 **
## Product.NameRental Vehicle Excess Insurance            2.251 0.024372 *
## Product.NameSilver Plan                                6.283 3.33e-10 ***
## Product.NameSingle Trip Travel Protect Gold            3.200 0.001372 **
## Product.NameSingle Trip Travel Protect Platinum        5.734 9.82e-09 ***
## Product.NameSingle Trip Travel Protect Silver         -0.041 0.967486
## Product.NameSpouse or Parents Comprehensive Plan       3.224 0.001264 **
## Product.NameTicket Protector                           3.884 0.000103 ***
## Product.NameTravel Cruise Protect                      0.874 0.382375
## Product.NameTravel Cruise Protect Family              -0.004 0.996666
## Product.NameValue Plan                                 3.973 7.09e-05 ***
## Age                                                   -2.301 0.021408 *
## Net.Sales                                              7.150 8.69e-13 ***
## factor(Destination)CHINA                               0.836 0.402999
## factor(Destination)HONG KONG                          -0.413 0.679632
## factor(Destination)INDIA                              -1.294 0.195828
## factor(Destination)INDONESIA                          -2.195 0.028146 *
## factor(Destination)MALAYSIA                           -2.352 0.018687 *
## factor(Destination)PHILIPPINES                        -1.556 0.119641
## factor(Destination)SINGAPORE                          -2.518 0.011818 *
## factor(Destination)THAILAND                           -1.147 0.251340
## factor(Destination)UNITED STATES                       0.540 0.589468
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 7866.7  on 45772  degrees of freedom
## Residual deviance: 6484.8  on 45735  degrees of freedom
## AIC: 6560.8
##
## Number of Fisher Scoring iterations: 15
```

```
binnedplot(predict(md3),resid(md3,type = "response"))
```



```r
library(lme4)
library(car)
md4 <- glmer(Claim ~ Agency.Type+Product.Name+ Age+ Net.Sales +(1|Destination
), data = newins2, family=binomial)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge: degenerate Hessian with 4
## negative eigenvalues
```

```r
summary(md4)
```

```
## Warning in vcov.merMod(object, use.hessian = use.hessian): variance-covari
ance matrix computed from finite-difference Hessian is
## not positive definite or contains NA values: falling back to var-cov estim
ated from RX
```

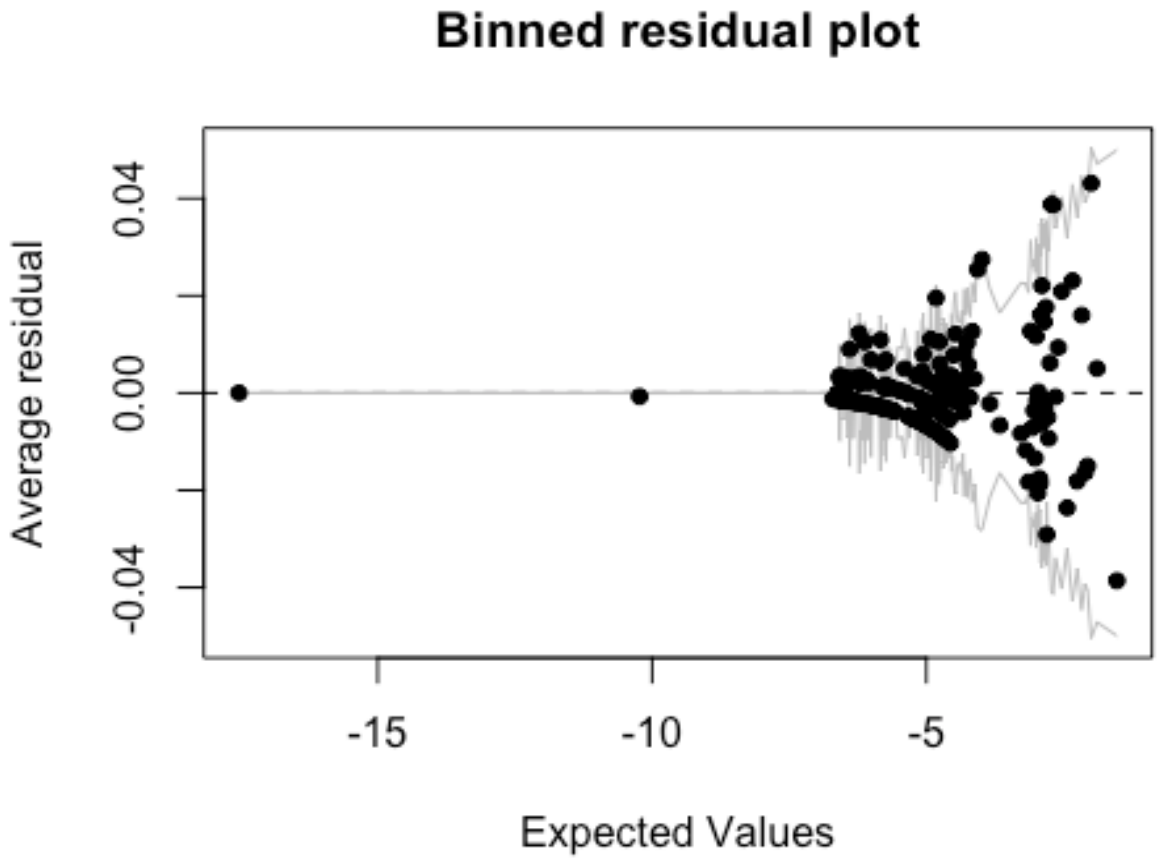
```
## Warning in vcov.merMod(object, correlation = correlation, sigm = sig): var
iance-covariance matrix computed from finite-difference Hessian is
## not positive definite or contains NA values: falling back to var-cov estim
ated from RX
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## Claim ~ Agency.Type + Product.Name + Age + Net.Sales + (1 | Destination)
##    Data: newins2
##
##      AIC      BIC   logLik deviance df.resid
##   6569.9   6831.8  -3254.9   6509.9    45743
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.1707 -0.1102 -0.0768 -0.0451 27.1991
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Destination (Intercept) 0.09638  0.3105
## Number of obs: 45773, groups:  Destination, 10
##
## Fixed effects:
##                                                Estimate Std. Error
## (Intercept)                                  -7.543e+00  8.555e-01
## Agency.TypeTravel Agency                      1.478e+00  7.094e-01
## Product.Name2 way Comprehensive Plan          1.426e+00  4.650e-01
## Product.Name24 Protect                       -9.577e+00  4.501e+02
## Product.NameAnnual Gold Plan                  4.336e+00  9.139e-01
## Product.NameAnnual Silver Plan                4.970e+00  8.697e-01
```

```
## Product.NameAnnual Travel Protect Gold            2.587e+00  5.912e-01
## Product.NameAnnual Travel Protect Platinum        1.596e+00  7.865e-01
## Product.NameAnnual Travel Protect Silver          1.974e+00  7.018e-01
## Product.NameBasic Plan                            1.909e+00  8.855e-01
## Product.NameBronze Plan                           5.046e+00  8.606e-01
## Product.NameCancellation Plan                    -6.760e-03  4.910e-01
## Product.NameChild Comprehensive Plan             -1.111e+01  1.727e+03
## Product.NameComprehensive Plan                    2.041e+00  6.878e-01
## Product.NameGold Plan                             4.750e+00  8.934e-01
## Product.NameIndividual Comprehensive Plan         2.922e+00  7.517e-01
## Product.NamePremier Plan                          2.651e+00  8.633e-01
## Product.NameRental Vehicle Excess Insurance       1.261e+00  4.884e-01
## Product.NameSilver Plan                           5.160e+00  8.621e-01
## Product.NameSingle Trip Travel Protect Gold       2.470e+00  7.420e-01
## Product.NameSingle Trip Travel Protect Platinum   3.877e+00  6.558e-01
## Product.NameSingle Trip Travel Protect Silver    -1.107e+01  5.862e+02
## Product.NameSpouse or Parents Comprehensive Plan  3.534e+00  1.135e+00
## Product.NameTicket Protector                      3.194e+00  9.397e-01
## Product.NameTravel Cruise Protect                 7.956e-01  8.499e-01
## Product.NameTravel Cruise Protect Family         -1.041e+01  4.780e+03
## Product.NameValue Plan                            2.783e+00  7.073e-01
## Age                                              -7.245e-03  3.205e-03
## Net.Sales                                         4.869e-03  6.620e-04
##                                                  z value Pr(>|z|)
## (Intercept)                                       -8.818  < 2e-16 ***
## Agency.TypeTravel Agency                           2.083 0.037226 *
## Product.Name2 way Comprehensive Plan               3.068 0.002158 **
## Product.Name24 Protect                            -0.021 0.983024
## Product.NameAnnual Gold Plan                       4.744 2.09e-06 ***
## Product.NameAnnual Silver Plan                     5.715 1.10e-08 ***
## Product.NameAnnual Travel Protect Gold             4.375 1.21e-05 ***
## Product.NameAnnual Travel Protect Platinum         2.029 0.042428 *
## Product.NameAnnual Travel Protect Silver           2.812 0.004919 **
## Product.NameBasic Plan                             2.156 0.031083 *
## Product.NameBronze Plan                            5.863 4.55e-09 ***
## Product.NameCancellation Plan                     -0.014 0.989016
## Product.NameChild Comprehensive Plan              -0.006 0.994866
## Product.NameComprehensive Plan                     2.967 0.003008 **
## Product.NameGold Plan                              5.316 1.06e-07 ***
## Product.NameIndividual Comprehensive Plan          3.887 0.000101 ***
## Product.NamePremier Plan                           3.070 0.002139 **
## Product.NameRental Vehicle Excess Insurance        2.582 0.009810 **
## Product.NameSilver Plan                            5.985 2.16e-09 ***
## Product.NameSingle Trip Travel Protect Gold        3.329 0.000872 ***
## Product.NameSingle Trip Travel Protect Platinum    5.913 3.37e-09 ***
## Product.NameSingle Trip Travel Protect Silver     -0.019 0.984928
## Product.NameSpouse or Parents Comprehensive Plan   3.112 0.001856 **
## Product.NameTicket Protector                       3.399 0.000677 ***
## Product.NameTravel Cruise Protect                  0.936 0.349255
## Product.NameTravel Cruise Protect Family          -0.002 0.998262
```

```
## Product.NameValue Plan                                  3.935 8.31e-05 ***
## Age                                                    -2.261 0.023789 *
## Net.Sales                                               7.355 1.92e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## convergence code: 0
## unable to evaluate scaled gradient
## Model failed to converge: degenerate  Hessian with 4 negative eigenvalues
```

```
binnedplot(predict(md4),resid(md4, type = "response"))
```



```
md5 <- glmer(Claim ~ Agency.Type+Product.Name+ Age+ Net.Sales +Destination+(1
|Duration), data = newins2, family=binomial)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.00951456
## (tol = 0.001, component 1)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkCon
v, : Model is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?;Model is nearly unidentifiable: large eigenvalue rat
io
##  - Rescale variables?
```

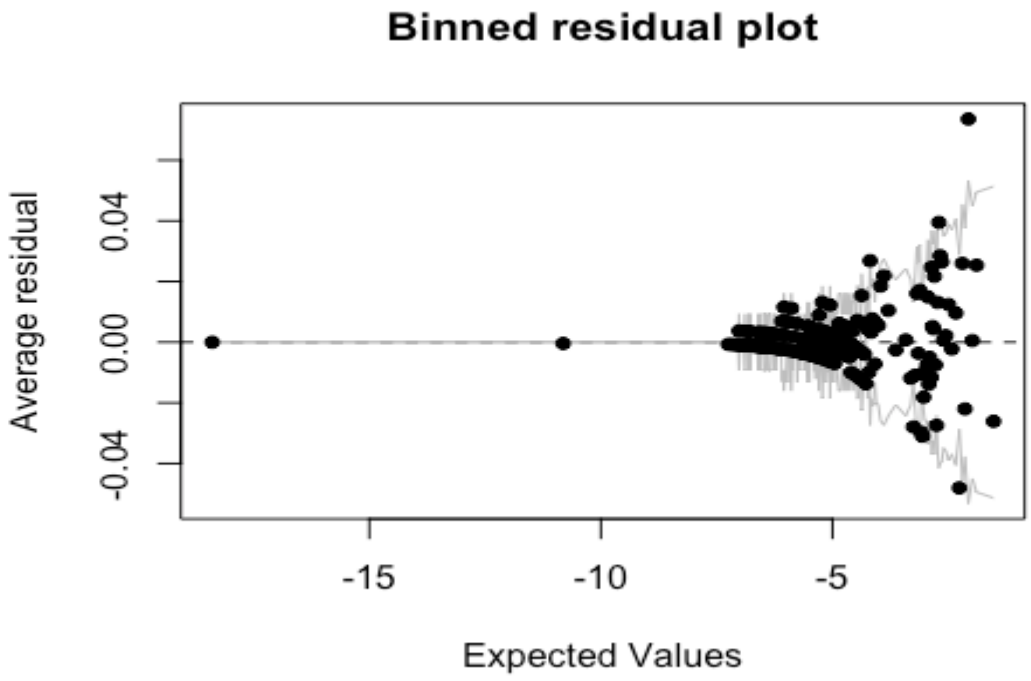
```
summary(md5)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## Claim ~ Agency.Type + Product.Name + Age + Net.Sales + Destination +
##     (1 | Duration)
##    Data: newins2
##
##      AIC      BIC   logLik deviance df.resid
##   6557.5   6898.1  -3239.8   6479.5    45734
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -1.048 -0.113 -0.070 -0.044 33.569
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Duration (Intercept) 0.05315  0.2305
## Number of obs: 45773, groups:  Duration, 440
##
## Fixed effects:
##                                                 Estimate Std. Error
## (Intercept)                                   -7.416e+00  9.030e-01
## Agency.TypeTravel Agency                       1.821e+00  7.560e-01
## Product.Name2 way Comprehensive Plan           1.292e+00  4.652e-01
## Product.Name24 Protect                        -9.910e+00  1.115e+02
## Product.NameAnnual Gold Plan                   5.017e+00  9.916e-01
## Product.NameAnnual Silver Plan                 5.608e+00  9.505e-01
## Product.NameAnnual Travel Protect Gold         2.311e+00  5.958e-01
## Product.NameAnnual Travel Protect Platinum     1.369e+00  7.900e-01
## Product.NameAnnual Travel Protect Silver       1.713e+00  7.043e-01
## Product.NameBasic Plan                         2.115e+00  9.274e-01
## Product.NameBronze Plan                        5.724e+00  9.390e-01
## Product.NameCancellation Plan                 -8.984e-02  4.895e-01
## Product.NameChild Comprehensive Plan          -1.211e+01  7.054e+01
```

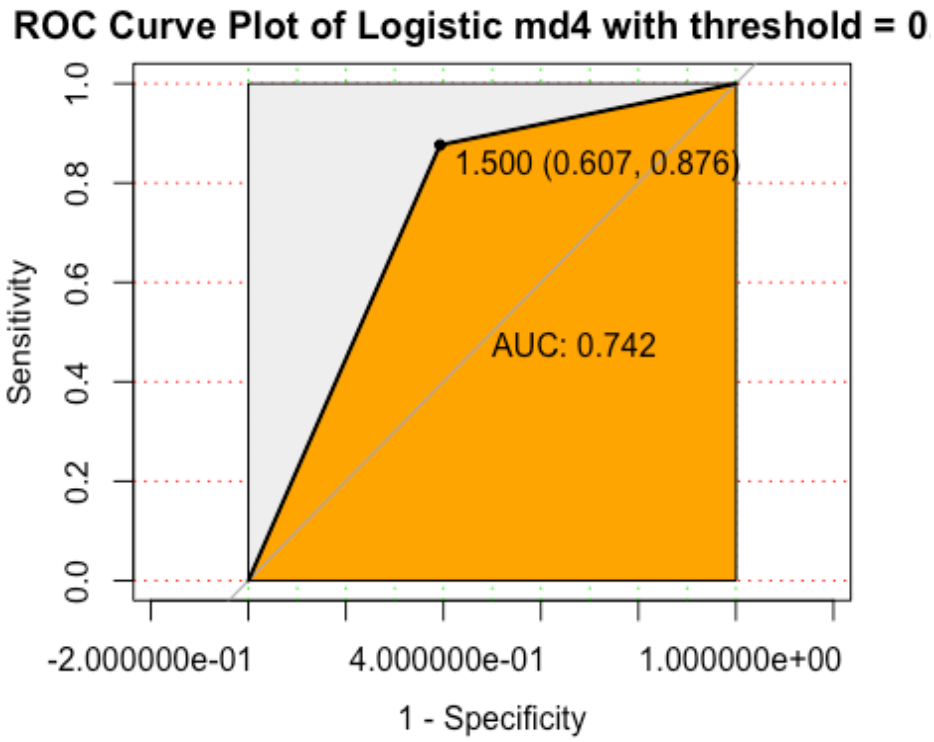
```
## Product.NameComprehensive Plan                      1.917e+00  6.921e-01
## Product.NameGold Plan                               5.457e+00  9.688e-01
## Product.NameIndividual Comprehensive Plan           2.962e+00  7.696e-01
## Product.NamePremier Plan                            2.719e+00  8.776e-01
## Product.NameRental Vehicle Excess Insurance         1.083e+00  4.949e-01
## Product.NameSilver Plan                             5.844e+00  9.404e-01
## Product.NameSingle Trip Travel Protect Gold         2.394e+00  7.417e-01
## Product.NameSingle Trip Travel Protect Platinum     3.791e+00  6.580e-01
## Product.NameSingle Trip Travel Protect Silver      -1.207e+01  2.423e+02
## Product.NameSpouse or Parents Comprehensive Plan    3.579e+00  1.150e+00
## Product.NameTicket Protector                        3.825e+00  1.013e+00
## Product.NameTravel Cruise Protect                   6.539e-01  8.536e-01
## Product.NameTravel Cruise Protect Family           -1.182e+01  1.615e+02
## Product.NameValue Plan                              2.911e+00  7.518e-01
## Age                                                -7.305e-03  3.205e-03
## Net.Sales                                           4.538e-03  6.848e-04
## DestinationCHINA                                    2.406e-01  2.782e-01
## DestinationHONG KONG                               -1.273e-01  3.493e-01
## DestinationINDIA                                   -7.065e-01  5.527e-01
## DestinationINDONESIA                               -8.543e-01  4.053e-01
## DestinationMALAYSIA                                -6.313e-01  2.909e-01
## DestinationPHILIPPINES                             -6.588e-01  4.231e-01
## DestinationSINGAPORE                               -1.098e+00  4.406e-01
## DestinationTHAILAND                                -2.895e-01  2.750e-01
## DestinationUNITED STATES                            1.523e-01  2.840e-01
##                                                   z value Pr(>|z|)
## (Intercept)                                        -8.212  < 2e-16 ***
## Agency.TypeTravel Agency                            2.409 0.015981 *
## Product.Name2 way Comprehensive Plan                2.777 0.005485 **
## Product.Name24 Protect                             -0.089 0.929185
## Product.NameAnnual Gold Plan                        5.059 4.21e-07 ***
## Product.NameAnnual Silver Plan                      5.900 3.64e-09 ***
## Product.NameAnnual Travel Protect Gold              3.879 0.000105 ***
## Product.NameAnnual Travel Protect Platinum          1.733 0.083123 .
## Product.NameAnnual Travel Protect Silver            2.432 0.015001 *
## Product.NameBasic Plan                              2.280 0.022581 *
## Product.NameBronze Plan                             6.096 1.09e-09 ***
## Product.NameCancellation Plan                      -0.184 0.854384
## Product.NameChild Comprehensive Plan               -0.172 0.863651
## Product.NameComprehensive Plan                      2.769 0.005619 **
## Product.NameGold Plan                               5.633 1.77e-08 ***
## Product.NameIndividual Comprehensive Plan           3.849 0.000118 ***
## Product.NamePremier Plan                            3.099 0.001944 **
## Product.NameRental Vehicle Excess Insurance         2.188 0.028690 *
## Product.NameSilver Plan                             6.214 5.16e-10 ***
## Product.NameSingle Trip Travel Protect Gold         3.227 0.001249 **
## Product.NameSingle Trip Travel Protect Platinum     5.761 8.35e-09 ***
## Product.NameSingle Trip Travel Protect Silver      -0.050 0.960279
## Product.NameSpouse or Parents Comprehensive Plan    3.113 0.001851 **
## Product.NameTicket Protector                        3.777 0.000159 ***
```

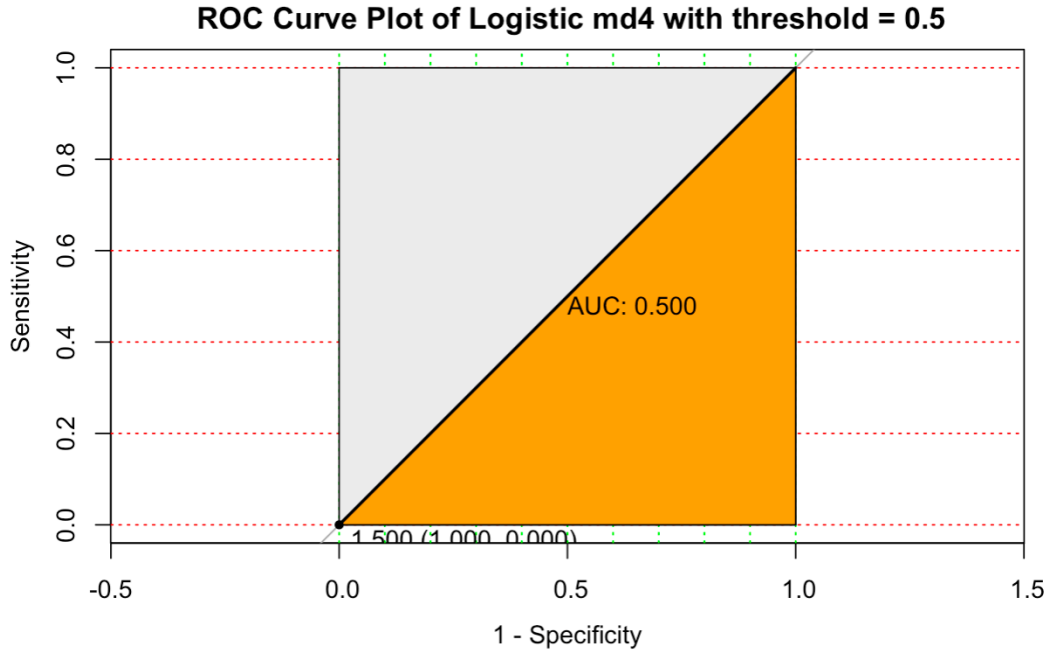
```
## Product.NameTravel Cruise Protect                    0.766 0.443681
## Product.NameTravel Cruise Protect Family            -0.073 0.941661
## Product.NameValue Plan                               3.872 0.000108 ***
## Age                                                 -2.279 0.022676 *
## Net.Sales                                            6.627 3.43e-11 ***
## DestinationCHINA                                     0.865 0.387224
## DestinationHONG KONG                                -0.364 0.715505
## DestinationINDIA                                    -1.278 0.201187
## DestinationINDONESIA                                -2.108 0.035032 *
## DestinationMALAYSIA                                 -2.170 0.029974 *
## DestinationPHILIPPINES                              -1.557 0.119501
## DestinationSINGAPORE                                -2.492 0.012700 *
## DestinationTHAILAND                                 -1.053 0.292508
## DestinationUNITED STATES                             0.536 0.591865
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## convergence code: 0
## Model failed to converge with max|grad| = 0.00951456 (tol = 0.001, compone
nt 1)
## Model is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?
## Model is nearly unidentifiable: large eigenvalue ratio
##  - Rescale variables?
```

```
binnedplot(predict(md5),resid(md5, type = "response"))
```



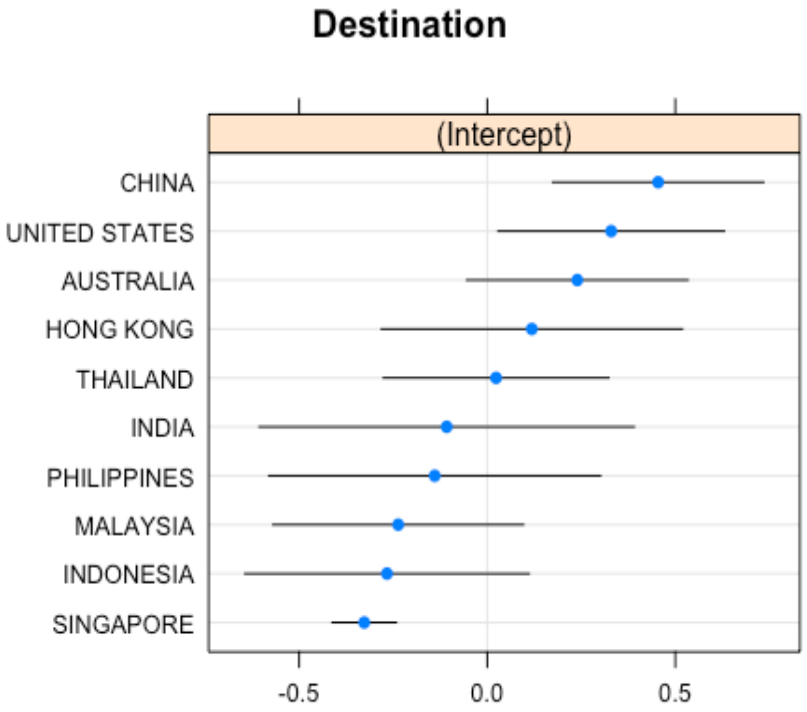
(threshold = 0.05)






```
lattice::dotplot(ranef(md4, condVar=TRUE))
```

```
## $Destination
```




```
anova(md5,md2,md3,md4)

## Data: newins2
## Models:
## md2: Claim ~ Agency.Type + Product.Name + Age + Net.Sales
## md4: Claim ~ Agency.Type + Product.Name + Age + Net.Sales + (1 | Destinati
on)
## md3: Claim ~ Agency.Type + Product.Name + Age + Net.Sales + factor(Destina
tion)
## md5: Claim ~ Agency.Type + Product.Name + Age + Net.Sales + Destination +
## md5:     (1 | Duration)
##     Df    AIC    BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)
## md2 29 6573.5 6826.7 -3257.7   6515.5
## md4 30 6569.9 6831.8 -3254.9   6509.9  5.5811      1   0.018155 *
## md3 38 6560.8 6892.6 -3242.4   6484.8 25.0373      8   0.001532 **
## md5 39 6557.5 6898.1 -3239.8   6479.5  5.3205      1   0.021076 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC(md2,md3,md4,md5)

##     df      AIC
## md2 29 6573.466
## md3 38 6560.848
## md4 30 6569.885
## md5 39 6557.528
```

From the binnedplot for md2,md3,md4,md5, we don't see there's an obvious improvement in each model.

From anova, md3 (without pooling) is of significance. And md3's AIC is relatively small. Multilevel models are not as better as expected.

Also, there are plots in Appendix that display the claim status and each continuous variable within the models respectively, grouping by destinations. We cannot see clear pattern within these 3 plots.

5.Discussion

Implication

According to the EDA results, the results of this study may have implication for potential claim making on the travel destination level, especially when we look at the group with destination in Singapore and the US. Compared to GLM, however, the models fitted with random intercepts cannot prove that there are clear improvements considering the AIC, ANOVA results and binned plots. Besides, the ROC curve doesn't imply that the model fits well.

Limitation

The models fitted are unexpectedly shocking because people would believe that some predictors such as travel duration and travel destination impact the probability of claim occurrence based on their commonsense. We are still not quite sure about what models have told us, not only because of our experience, but also because that the dataset itself has limitation. The amounts of binary outcomes are distributed extremely. Almost 99% of the insured didn't make a claim. So, the amount of observations with 0 claim are extremely huge. This issue makes it quite difficult for us to apply data visualization. Also, it confuses model very much.

Another fact is that, sometimes the claims are not made by insured even though the events covered by policies happened. Because the loss amount is not large enough, customers sometimes don't bother to make a claim. This is another reason to explain the very little amount of claim marked as "Yes".

Future Direction

As discussed above, in the future, we need to obtain more available data with more valid variables. So that we can fit more general models.

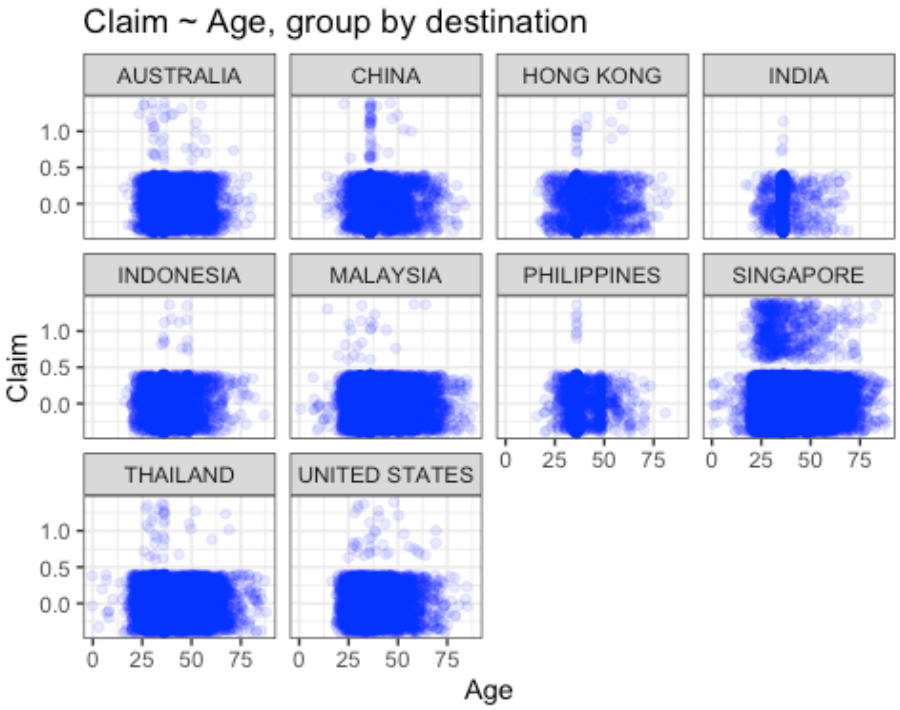
6.Acknowledgement

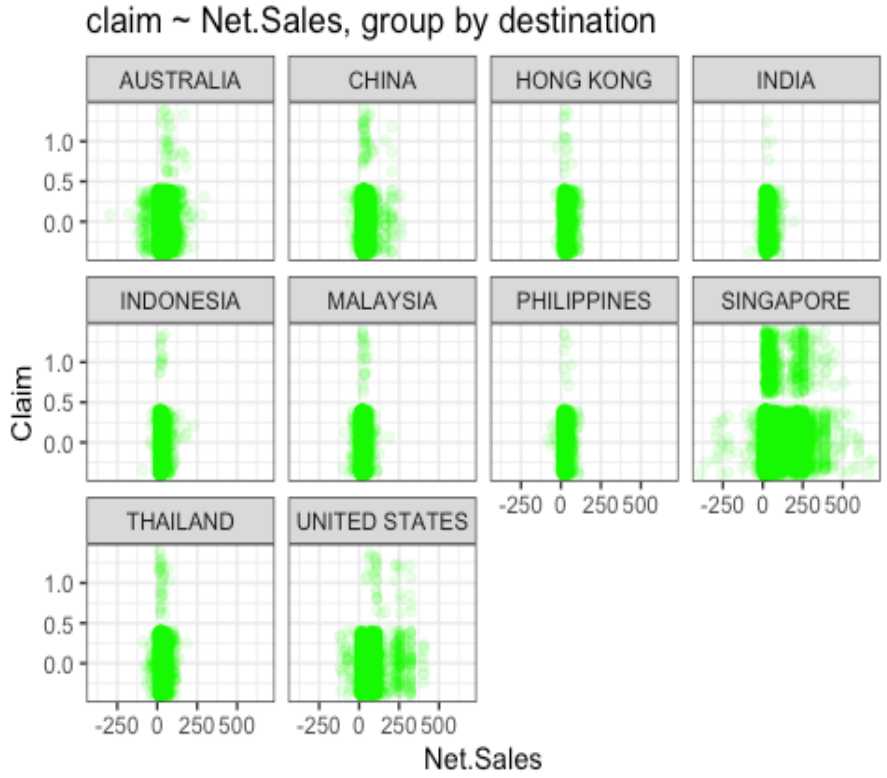
I would like to thank my classmate, Harry Cao, for answering questions and providing suggestions to help me better understand this project.

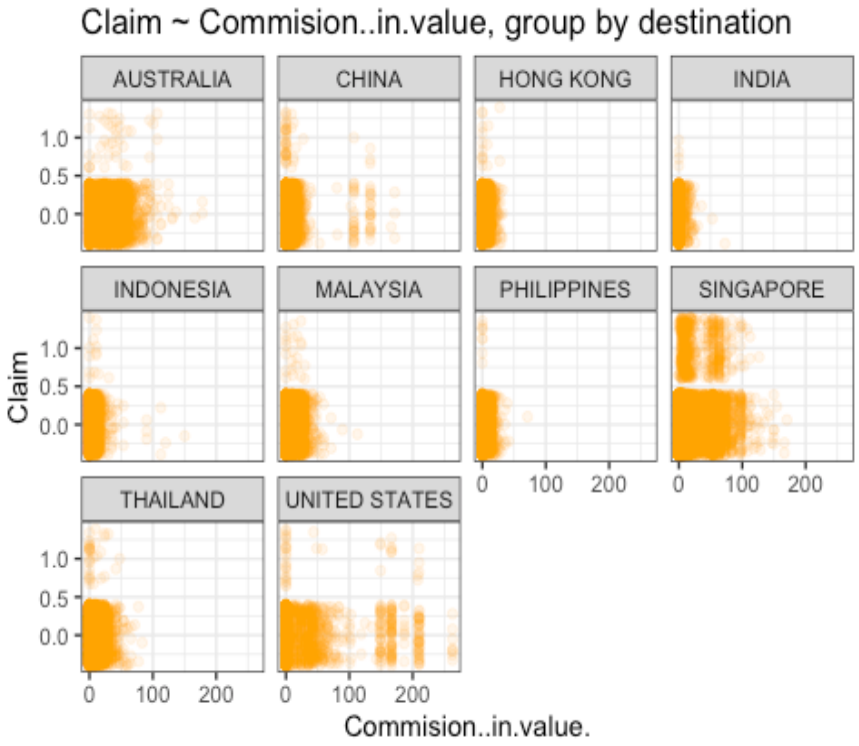
7.Reference

Travel Insurance, Zahier Nasrudin, *[Online] Available from:* *https://www.kaggle.com/mhdzahier/travel-insurance* *[Accesses 26th November 2019]*

8.Appendix


Claim ~ Age, group by destination
AUSTRALIA
CHINA
HONG KONG
INDIA
INDONESIA
MALAYSIA
PHILIPPINES
SINGAPORE
THAILAND
UNITED STATES
Claim
Age
1.0
0.5
0.0
0
25
50
75


claim ~ Net.Sales, group by destination
AUSTRALIA
CHINA
HONG KONG
INDIA
INDONESIA
MALAYSIA
PHILIPPINES
SINGAPORE
THAILAND
UNITED STATES
Claim
Net.Sales
1.0
0.5
0.0
-250
0
250
500


Claim ~ Commision..in.value, group by destination
AUSTRALIA
CHINA
HONG KONG
INDIA
INDONESIA
MALAYSIA
PHILIPPINES
SINGAPORE
THAILAND
UNITED STATES
Claim
1.0
0.5
0.0
0
100
200
Commision..in.value.