# STOR 455 Class 31 R Multiple Logistic Regression Again Agian

library(readr)  
library(bestglm)  
library(MASS)  
  
insurance <- read\_csv("https://raw.githubusercontent.com/JA-McLean/STOR455/master/data/insurance.csv")

**Issues with Insurance Data for bestglm**

head(insurance)

## # A tibble: 6 x 26  
## INDEX TARGET\_FLAG TARGET\_AMT KIDSDRIV AGE HOMEKIDS YOJ INCOME PARENT1  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr>   
## 1 1 0 0 0 60 0 11 $67,349 No   
## 2 2 0 0 0 43 0 11 $91,449 No   
## 3 4 0 0 0 35 1 10 $16,039 No   
## 4 5 0 0 0 51 0 14 <NA> No   
## 5 6 0 0 0 50 0 NA $114,986 No   
## 6 7 1 2946 0 34 1 12 $125,301 Yes   
## # ... with 17 more variables: HOME\_VAL <chr>, MSTATUS <chr>, SEX <chr>,  
## # EDUCATION <chr>, JOB <chr>, TRAVTIME <dbl>, CAR\_USE <chr>, BLUEBOOK <chr>,  
## # TIF <dbl>, CAR\_TYPE <chr>, RED\_CAR <chr>, OLDCLAIM <chr>, CLM\_FREQ <dbl>,  
## # REVOKED <chr>, MVR\_PTS <dbl>, CAR\_AGE <dbl>, URBANICITY <chr>

Predict if people are going to get into a car accident How likely that these people get in a car accident based on teh data we are working with

# If it sees a dollar sign or a comma, it changes things to a number like as.numeric   
# So this makes us just have numbers   
# Want 4 new variables in teh dataframe   
Currency\_Convert <- function(Field){  
 Field <- as.numeric(gsub("\\$|,","", Field))  
}  
  
#Change factors to numbers  
insurance$HOME\_VAL\_num = Currency\_Convert(insurance$HOME\_VAL)  
insurance$INCOME\_num = Currency\_Convert(insurance$INCOME)  
insurance$BLUEBOOK\_num = Currency\_Convert(insurance$BLUEBOOK)  
insurance$OLDCLAIM\_num = Currency\_Convert(insurance$OLDCLAIM)  
  
#remove unneeded variables  
# Got rid of them because it would be not known at teh time we wanted to make prediction   
# But we now technically have this data   
# So we are jus getting rid of it   
insurance.1 = within(insurance,   
 {INDEX = NULL  
 TARGET\_AMT = NULL  
 HOME\_VAL = NULL  
 INCOME = NULL   
 BLUEBOOK = NULL  
 OLDCLAIM = NULL})  
  
  
head(insurance.1)

## # A tibble: 6 x 24  
## TARGET\_FLAG KIDSDRIV AGE HOMEKIDS YOJ PARENT1 MSTATUS SEX EDUCATION   
## <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr> <chr>   
## 1 0 0 60 0 11 No z\_No M PhD   
## 2 0 0 43 0 11 No z\_No M z\_High School  
## 3 0 0 35 1 10 No Yes z\_F z\_High School  
## 4 0 0 51 0 14 No Yes M <High School   
## 5 0 0 50 0 NA No Yes z\_F PhD   
## 6 1 0 34 1 12 Yes z\_No z\_F Bachelors   
## # ... with 15 more variables: JOB <chr>, TRAVTIME <dbl>, CAR\_USE <chr>,  
## # TIF <dbl>, CAR\_TYPE <chr>, RED\_CAR <chr>, CLM\_FREQ <dbl>, REVOKED <chr>,  
## # MVR\_PTS <dbl>, CAR\_AGE <dbl>, URBANICITY <chr>, HOME\_VAL\_num <dbl>,  
## # INCOME\_num <dbl>, BLUEBOOK\_num <dbl>, OLDCLAIM\_num <dbl>

# Now we have numeric data where we need it

insurance.2 = insurance.1[,c(2:24,1)]   
# For best glm, we want the last column of the data as teh predicted, so we need that to be at t eh end of the dataframe   
# We want target flag at the end   
head(insurance.2)

## # A tibble: 6 x 24  
## KIDSDRIV AGE HOMEKIDS YOJ PARENT1 MSTATUS SEX EDUCATION JOB TRAVTIME  
## <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr> <chr> <chr> <dbl>  
## 1 0 60 0 11 No z\_No M PhD Profe~ 14  
## 2 0 43 0 11 No z\_No M z\_High Sc~ z\_Blu~ 22  
## 3 0 35 1 10 No Yes z\_F z\_High Sc~ Cleri~ 5  
## 4 0 51 0 14 No Yes M <High Sch~ z\_Blu~ 32  
## 5 0 50 0 NA No Yes z\_F PhD Doctor 36  
## 6 0 34 1 12 Yes z\_No z\_F Bachelors z\_Blu~ 46  
## # ... with 14 more variables: CAR\_USE <chr>, TIF <dbl>, CAR\_TYPE <chr>,  
## # RED\_CAR <chr>, CLM\_FREQ <dbl>, REVOKED <chr>, MVR\_PTS <dbl>, CAR\_AGE <dbl>,  
## # URBANICITY <chr>, HOME\_VAL\_num <dbl>, INCOME\_num <dbl>, BLUEBOOK\_num <dbl>,  
## # OLDCLAIM\_num <dbl>, TARGET\_FLAG <dbl>

#Sad trombone  
insurance.2 = as.data.frame(insurance.2)  
# bestglm(insurance.2, family=binomial)  
# This doesn't work because the structure of teh dataframe is maniuplated a bit weird; have to make it a dataframe; it still wont make it work all the time, but it might fix it   
# THis time it wont fix this issue

# So to fix that, we will use dplyr to make all the character vectors as factors   
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':  
##   
## select

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

insurance.2.1 = insurance.2 %>% mutate\_if(is.character, factor)  
head(insurance.2.1)

## KIDSDRIV AGE HOMEKIDS YOJ PARENT1 MSTATUS SEX EDUCATION JOB  
## 1 0 60 0 11 No z\_No M PhD Professional  
## 2 0 43 0 11 No z\_No M z\_High School z\_Blue Collar  
## 3 0 35 1 10 No Yes z\_F z\_High School Clerical  
## 4 0 51 0 14 No Yes M <High School z\_Blue Collar  
## 5 0 50 0 NA No Yes z\_F PhD Doctor  
## 6 0 34 1 12 Yes z\_No z\_F Bachelors z\_Blue Collar  
## TRAVTIME CAR\_USE TIF CAR\_TYPE RED\_CAR CLM\_FREQ REVOKED MVR\_PTS CAR\_AGE  
## 1 14 Private 11 Minivan yes 2 No 3 18  
## 2 22 Commercial 1 Minivan yes 0 No 0 1  
## 3 5 Private 4 z\_SUV no 2 No 3 10  
## 4 32 Private 7 Minivan yes 0 No 0 6  
## 5 36 Private 1 z\_SUV no 2 Yes 3 17  
## 6 46 Commercial 1 Sports Car no 0 No 0 7  
## URBANICITY HOME\_VAL\_num INCOME\_num BLUEBOOK\_num OLDCLAIM\_num  
## 1 Highly Urban/ Urban 0 67349 14230 4461  
## 2 Highly Urban/ Urban 257252 91449 14940 0  
## 3 Highly Urban/ Urban 124191 16039 4010 38690  
## 4 Highly Urban/ Urban 306251 NA 15440 0  
## 5 Highly Urban/ Urban 243925 114986 18000 19217  
## 6 Highly Urban/ Urban 0 125301 17430 0  
## TARGET\_FLAG  
## 1 0  
## 2 0  
## 3 0  
## 4 0  
## 5 0  
## 6 1

#Sadder trombone  
insurance.2.1 = as.data.frame(insurance.2.1)  
#bestglm(insurance.2.1, family=binomial)  
# When trying again, it doesn't appear to work still; so we have to try something different   
# NEw error, Error: p = 23, much be <= 15   
# tells us that there are too many variables, so we can't run it

When best glm doesn’t work, let’s use different stepwise. We will do AIC Stepwise

full\_insurance = glm(TARGET\_FLAG~., data=insurance.1, family="binomial")  
summary(full\_insurance)

##   
## Call:  
## glm(formula = TARGET\_FLAG ~ ., family = "binomial", data = insurance.1)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5538 -0.7033 -0.3906 0.6213 3.1736   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.474e-01 3.122e-01 -0.472 0.636775   
## KIDSDRIV 3.194e-01 7.059e-02 4.524 6.06e-06 \*\*\*  
## AGE -3.903e-03 4.681e-03 -0.834 0.404452   
## HOMEKIDS 2.313e-02 4.275e-02 0.541 0.588473   
## YOJ -9.356e-03 9.775e-03 -0.957 0.338476   
## PARENT1Yes 4.109e-01 1.266e-01 3.246 0.001170 \*\*   
## MSTATUSz\_No 4.266e-01 1.006e-01 4.240 2.24e-05 \*\*\*  
## SEXz\_F -2.076e-01 1.287e-01 -1.613 0.106647   
## EDUCATIONBachelors -3.762e-01 1.325e-01 -2.840 0.004508 \*\*   
## EDUCATIONMasters -4.433e-01 2.140e-01 -2.072 0.038312 \*   
## EDUCATIONPhD 9.120e-02 2.635e-01 0.346 0.729208   
## EDUCATIONz\_High School -1.126e-03 1.068e-01 -0.011 0.991590   
## JOBDoctor -8.849e-01 3.279e-01 -2.699 0.006958 \*\*   
## JOBHome Maker -3.239e-01 1.669e-01 -1.941 0.052309 .   
## JOBLawyer -1.491e-01 2.120e-01 -0.703 0.481916   
## JOBManager -1.075e+00 1.650e-01 -6.519 7.07e-11 \*\*\*  
## JOBProfessional -2.843e-01 1.418e-01 -2.006 0.044906 \*   
## JOBStudent -3.615e-01 1.528e-01 -2.365 0.018010 \*   
## JOBz\_Blue Collar -1.931e-01 1.204e-01 -1.603 0.108890   
## TRAVTIME 1.569e-02 2.192e-03 7.159 8.15e-13 \*\*\*  
## CAR\_USEPrivate -8.292e-01 1.060e-01 -7.822 5.18e-15 \*\*\*  
## TIF -5.225e-02 8.544e-03 -6.115 9.64e-10 \*\*\*  
## CAR\_TYPEPanel Truck 6.955e-01 1.948e-01 3.570 0.000357 \*\*\*  
## CAR\_TYPEPickup 5.556e-01 1.154e-01 4.816 1.47e-06 \*\*\*  
## CAR\_TYPESports Car 1.107e+00 1.466e-01 7.553 4.27e-14 \*\*\*  
## CAR\_TYPEVan 5.678e-01 1.497e-01 3.793 0.000149 \*\*\*  
## CAR\_TYPEz\_SUV 8.265e-01 1.257e-01 6.575 4.88e-11 \*\*\*  
## RED\_CARyes -2.278e-01 1.032e-01 -2.208 0.027231 \*   
## CLM\_FREQ 2.004e-01 3.320e-02 6.036 1.58e-09 \*\*\*  
## REVOKEDYes 8.521e-01 1.075e-01 7.930 2.19e-15 \*\*\*  
## MVR\_PTS 1.161e-01 1.587e-02 7.312 2.63e-13 \*\*\*  
## CAR\_AGE -3.920e-03 8.897e-03 -0.441 0.659514   
## URBANICITYz\_Highly Rural/ Rural -2.306e+00 1.244e-01 -18.537 < 2e-16 \*\*\*  
## HOME\_VAL\_num -1.420e-06 4.287e-07 -3.312 0.000927 \*\*\*  
## INCOME\_num -3.429e-06 1.432e-06 -2.394 0.016680 \*   
## BLUEBOOK\_num -2.257e-05 6.102e-06 -3.699 0.000216 \*\*\*  
## OLDCLAIM\_num -1.310e-05 4.574e-06 -2.864 0.004186 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 6990.9 on 6044 degrees of freedom  
## Residual deviance: 5362.3 on 6008 degrees of freedom  
## (2116 observations deleted due to missingness)  
## AIC: 5436.3  
##   
## Number of Fisher Scoring iterations: 5

#More sad trombone  
#stepAIC(full\_insurance)  
# Says that some blank values are messing us up, so lets get rid of them

insurance.forstepAIC = na.omit(insurance.1) # Gets rid of the na values   
  
full\_insurance.2 = glm(TARGET\_FLAG~., data=insurance.forstepAIC, family="binomial")  
  
# Given a full model - Backwards selection is the default  
stepAIC(full\_insurance.2)

## Start: AIC=5436.27  
## TARGET\_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + PARENT1 + MSTATUS +   
## SEX + EDUCATION + JOB + TRAVTIME + CAR\_USE + TIF + CAR\_TYPE +   
## RED\_CAR + CLM\_FREQ + REVOKED + MVR\_PTS + CAR\_AGE + URBANICITY +   
## HOME\_VAL\_num + INCOME\_num + BLUEBOOK\_num + OLDCLAIM\_num  
##   
## Df Deviance AIC  
## - CAR\_AGE 1 5362.5 5434.5  
## - HOMEKIDS 1 5362.6 5434.6  
## - AGE 1 5363.0 5435.0  
## - YOJ 1 5363.2 5435.2  
## <none> 5362.3 5436.3  
## - SEX 1 5364.9 5436.9  
## - RED\_CAR 1 5367.1 5439.1  
## - INCOME\_num 1 5368.1 5440.1  
## - OLDCLAIM\_num 1 5370.6 5442.6  
## - PARENT1 1 5372.8 5444.8  
## - HOME\_VAL\_num 1 5373.3 5445.3  
## - EDUCATION 4 5382.1 5448.1  
## - BLUEBOOK\_num 1 5376.2 5448.2  
## - MSTATUS 1 5380.1 5452.1  
## - KIDSDRIV 1 5382.6 5454.6  
## - CLM\_FREQ 1 5398.3 5470.3  
## - TIF 1 5401.0 5473.0  
## - JOB 7 5425.0 5485.0  
## - TRAVTIME 1 5413.7 5485.7  
## - MVR\_PTS 1 5416.3 5488.3  
## - REVOKED 1 5424.1 5496.1  
## - CAR\_USE 1 5424.5 5496.5  
## - CAR\_TYPE 5 5444.9 5508.9  
## - URBANICITY 1 5842.8 5914.8  
##   
## Step: AIC=5434.47  
## TARGET\_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + PARENT1 + MSTATUS +   
## SEX + EDUCATION + JOB + TRAVTIME + CAR\_USE + TIF + CAR\_TYPE +   
## RED\_CAR + CLM\_FREQ + REVOKED + MVR\_PTS + URBANICITY + HOME\_VAL\_num +   
## INCOME\_num + BLUEBOOK\_num + OLDCLAIM\_num  
##   
## Df Deviance AIC  
## - HOMEKIDS 1 5362.8 5432.8  
## - AGE 1 5363.2 5433.2  
## - YOJ 1 5363.4 5433.4  
## <none> 5362.5 5434.5  
## - SEX 1 5365.1 5435.1  
## - RED\_CAR 1 5367.3 5437.3  
## - INCOME\_num 1 5368.4 5438.4  
## - OLDCLAIM\_num 1 5370.8 5440.8  
## - PARENT1 1 5373.0 5443.0  
## - HOME\_VAL\_num 1 5373.3 5443.3  
## - BLUEBOOK\_num 1 5376.3 5446.3  
## - MSTATUS 1 5380.3 5450.3  
## - EDUCATION 4 5387.3 5451.3  
## - KIDSDRIV 1 5382.8 5452.8  
## - CLM\_FREQ 1 5398.4 5468.4  
## - TIF 1 5401.3 5471.3  
## - JOB 7 5425.2 5483.2  
## - TRAVTIME 1 5413.8 5483.8  
## - MVR\_PTS 1 5416.5 5486.5  
## - REVOKED 1 5424.2 5494.2  
## - CAR\_USE 1 5424.6 5494.6  
## - CAR\_TYPE 5 5445.2 5507.2  
## - URBANICITY 1 5843.0 5913.0  
##   
## Step: AIC=5432.76  
## TARGET\_FLAG ~ KIDSDRIV + AGE + YOJ + PARENT1 + MSTATUS + SEX +   
## EDUCATION + JOB + TRAVTIME + CAR\_USE + TIF + CAR\_TYPE + RED\_CAR +   
## CLM\_FREQ + REVOKED + MVR\_PTS + URBANICITY + HOME\_VAL\_num +   
## INCOME\_num + BLUEBOOK\_num + OLDCLAIM\_num  
##   
## Df Deviance AIC  
## - YOJ 1 5363.5 5431.5  
## - AGE 1 5364.0 5432.0  
## <none> 5362.8 5432.8  
## - SEX 1 5365.4 5433.4  
## - RED\_CAR 1 5367.6 5435.6  
## - INCOME\_num 1 5368.6 5436.6  
## - OLDCLAIM\_num 1 5371.1 5439.1  
## - HOME\_VAL\_num 1 5373.8 5441.8  
## - BLUEBOOK\_num 1 5376.6 5444.6  
## - PARENT1 1 5377.6 5445.6  
## - MSTATUS 1 5380.7 5448.7  
## - EDUCATION 4 5387.8 5449.8  
## - KIDSDRIV 1 5390.0 5458.0  
## - CLM\_FREQ 1 5398.8 5466.8  
## - TIF 1 5401.6 5469.6  
## - JOB 7 5425.4 5481.4  
## - TRAVTIME 1 5414.0 5482.0  
## - MVR\_PTS 1 5416.9 5484.9  
## - REVOKED 1 5424.8 5492.8  
## - CAR\_USE 1 5425.1 5493.1  
## - CAR\_TYPE 5 5445.7 5505.7  
## - URBANICITY 1 5843.1 5911.1  
##   
## Step: AIC=5431.49  
## TARGET\_FLAG ~ KIDSDRIV + AGE + PARENT1 + MSTATUS + SEX + EDUCATION +   
## JOB + TRAVTIME + CAR\_USE + TIF + CAR\_TYPE + RED\_CAR + CLM\_FREQ +   
## REVOKED + MVR\_PTS + URBANICITY + HOME\_VAL\_num + INCOME\_num +   
## BLUEBOOK\_num + OLDCLAIM\_num  
##   
## Df Deviance AIC  
## - AGE 1 5365.0 5431.0  
## <none> 5363.5 5431.5  
## - SEX 1 5366.1 5432.1  
## - RED\_CAR 1 5368.3 5434.3  
## - INCOME\_num 1 5369.8 5435.8  
## - OLDCLAIM\_num 1 5372.0 5438.0  
## - HOME\_VAL\_num 1 5374.5 5440.5  
## - BLUEBOOK\_num 1 5377.4 5443.4  
## - PARENT1 1 5377.9 5443.9  
## - EDUCATION 4 5388.3 5448.3  
## - MSTATUS 1 5382.8 5448.8  
## - KIDSDRIV 1 5390.4 5456.4  
## - CLM\_FREQ 1 5399.6 5465.6  
## - TIF 1 5402.6 5468.6  
## - JOB 7 5425.5 5479.5  
## - TRAVTIME 1 5414.6 5480.6  
## - MVR\_PTS 1 5417.9 5483.9  
## - REVOKED 1 5425.6 5491.6  
## - CAR\_USE 1 5426.2 5492.2  
## - CAR\_TYPE 5 5446.8 5504.8  
## - URBANICITY 1 5843.5 5909.5  
##   
## Step: AIC=5431  
## TARGET\_FLAG ~ KIDSDRIV + PARENT1 + MSTATUS + SEX + EDUCATION +   
## JOB + TRAVTIME + CAR\_USE + TIF + CAR\_TYPE + RED\_CAR + CLM\_FREQ +   
## REVOKED + MVR\_PTS + URBANICITY + HOME\_VAL\_num + INCOME\_num +   
## BLUEBOOK\_num + OLDCLAIM\_num  
##   
## Df Deviance AIC  
## <none> 5365.0 5431.0  
## - SEX 1 5367.1 5431.1  
## - RED\_CAR 1 5369.7 5433.7  
## - INCOME\_num 1 5371.0 5435.0  
## - OLDCLAIM\_num 1 5373.6 5437.6  
## - HOME\_VAL\_num 1 5376.9 5440.9  
## - BLUEBOOK\_num 1 5380.7 5444.7  
## - MSTATUS 1 5383.5 5447.5  
## - PARENT1 1 5383.9 5447.9  
## - EDUCATION 4 5390.0 5448.0  
## - KIDSDRIV 1 5391.4 5455.4  
## - CLM\_FREQ 1 5401.0 5465.0  
## - TIF 1 5404.0 5468.0  
## - TRAVTIME 1 5415.8 5479.8  
## - JOB 7 5428.4 5480.4  
## - MVR\_PTS 1 5420.1 5484.1  
## - REVOKED 1 5427.4 5491.4  
## - CAR\_USE 1 5427.8 5491.8  
## - CAR\_TYPE 5 5447.1 5503.1  
## - URBANICITY 1 5846.8 5910.8

##   
## Call: glm(formula = TARGET\_FLAG ~ KIDSDRIV + PARENT1 + MSTATUS + SEX +   
## EDUCATION + JOB + TRAVTIME + CAR\_USE + TIF + CAR\_TYPE + RED\_CAR +   
## CLM\_FREQ + REVOKED + MVR\_PTS + URBANICITY + HOME\_VAL\_num +   
## INCOME\_num + BLUEBOOK\_num + OLDCLAIM\_num, family = "binomial",   
## data = insurance.forstepAIC)  
##   
## Coefficients:  
## (Intercept) KIDSDRIV   
## -3.905e-01 3.307e-01   
## PARENT1Yes MSTATUSz\_No   
## 4.730e-01 4.146e-01   
## SEXz\_F EDUCATIONBachelors   
## -1.861e-01 -3.995e-01   
## EDUCATIONMasters EDUCATIONPhD   
## -4.969e-01 4.064e-02   
## EDUCATIONz\_High School JOBDoctor   
## -7.396e-03 -9.037e-01   
## JOBHome Maker JOBLawyer   
## -2.872e-01 -1.650e-01   
## JOBManager JOBProfessional   
## -1.091e+00 -2.958e-01   
## JOBStudent JOBz\_Blue Collar   
## -3.133e-01 -2.020e-01   
## TRAVTIME CAR\_USEPrivate   
## 1.557e-02 -8.324e-01   
## TIF CAR\_TYPEPanel Truck   
## -5.237e-02 7.121e-01   
## CAR\_TYPEPickup CAR\_TYPESports Car   
## 5.525e-01 1.090e+00   
## CAR\_TYPEVan CAR\_TYPEz\_SUV   
## 5.778e-01 8.112e-01   
## RED\_CARyes CLM\_FREQ   
## -2.236e-01 2.001e-01   
## REVOKEDYes MVR\_PTS   
## 8.551e-01 1.170e-01   
## URBANICITYz\_Highly Rural/ Rural HOME\_VAL\_num   
## -2.308e+00 -1.469e-06   
## INCOME\_num BLUEBOOK\_num   
## -3.471e-06 -2.370e-05   
## OLDCLAIM\_num   
## -1.326e-05   
##   
## Degrees of Freedom: 6044 Total (i.e. Null); 6012 Residual  
## Null Deviance: 6991   
## Residual Deviance: 5365 AIC: 5431

# trace=FALSE will show only the final model, not each step.  
# Looks at new insurance data   
# Big model at teh end   
stepAIC(full\_insurance.2, trace=FALSE)

##   
## Call: glm(formula = TARGET\_FLAG ~ KIDSDRIV + PARENT1 + MSTATUS + SEX +   
## EDUCATION + JOB + TRAVTIME + CAR\_USE + TIF + CAR\_TYPE + RED\_CAR +   
## CLM\_FREQ + REVOKED + MVR\_PTS + URBANICITY + HOME\_VAL\_num +   
## INCOME\_num + BLUEBOOK\_num + OLDCLAIM\_num, family = "binomial",   
## data = insurance.forstepAIC)  
##   
## Coefficients:  
## (Intercept) KIDSDRIV   
## -3.905e-01 3.307e-01   
## PARENT1Yes MSTATUSz\_No   
## 4.730e-01 4.146e-01   
## SEXz\_F EDUCATIONBachelors   
## -1.861e-01 -3.995e-01   
## EDUCATIONMasters EDUCATIONPhD   
## -4.969e-01 4.064e-02   
## EDUCATIONz\_High School JOBDoctor   
## -7.396e-03 -9.037e-01   
## JOBHome Maker JOBLawyer   
## -2.872e-01 -1.650e-01   
## JOBManager JOBProfessional   
## -1.091e+00 -2.958e-01   
## JOBStudent JOBz\_Blue Collar   
## -3.133e-01 -2.020e-01   
## TRAVTIME CAR\_USEPrivate   
## 1.557e-02 -8.324e-01   
## TIF CAR\_TYPEPanel Truck   
## -5.237e-02 7.121e-01   
## CAR\_TYPEPickup CAR\_TYPESports Car   
## 5.525e-01 1.090e+00   
## CAR\_TYPEVan CAR\_TYPEz\_SUV   
## 5.778e-01 8.112e-01   
## RED\_CARyes CLM\_FREQ   
## -2.236e-01 2.001e-01   
## REVOKEDYes MVR\_PTS   
## 8.551e-01 1.170e-01   
## URBANICITYz\_Highly Rural/ Rural HOME\_VAL\_num   
## -2.308e+00 -1.469e-06   
## INCOME\_num BLUEBOOK\_num   
## -3.471e-06 -2.370e-05   
## OLDCLAIM\_num   
## -1.326e-05   
##   
## Degrees of Freedom: 6044 Total (i.e. Null); 6012 Residual  
## Null Deviance: 6991   
## Residual Deviance: 5365 AIC: 5431

final\_model\_backwards=stepAIC(full\_insurance.2, trace=FALSE)   
  
summary(final\_model\_backwards)

##   
## Call:  
## glm(formula = TARGET\_FLAG ~ KIDSDRIV + PARENT1 + MSTATUS + SEX +   
## EDUCATION + JOB + TRAVTIME + CAR\_USE + TIF + CAR\_TYPE + RED\_CAR +   
## CLM\_FREQ + REVOKED + MVR\_PTS + URBANICITY + HOME\_VAL\_num +   
## INCOME\_num + BLUEBOOK\_num + OLDCLAIM\_num, family = "binomial",   
## data = insurance.forstepAIC)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5691 -0.7024 -0.3901 0.6201 3.1495   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.905e-01 2.287e-01 -1.707 0.087770 .   
## KIDSDRIV 3.307e-01 6.390e-02 5.176 2.27e-07 \*\*\*  
## PARENT1Yes 4.730e-01 1.089e-01 4.342 1.41e-05 \*\*\*  
## MSTATUSz\_No 4.146e-01 9.568e-02 4.334 1.47e-05 \*\*\*  
## SEXz\_F -1.861e-01 1.272e-01 -1.464 0.143274   
## EDUCATIONBachelors -3.995e-01 1.231e-01 -3.245 0.001174 \*\*   
## EDUCATIONMasters -4.969e-01 1.905e-01 -2.609 0.009094 \*\*   
## EDUCATIONPhD 4.064e-02 2.461e-01 0.165 0.868809   
## EDUCATIONz\_High School -7.396e-03 1.063e-01 -0.070 0.944554   
## JOBDoctor -9.037e-01 3.273e-01 -2.761 0.005755 \*\*   
## JOBHome Maker -2.872e-01 1.569e-01 -1.831 0.067098 .   
## JOBLawyer -1.650e-01 2.112e-01 -0.781 0.434854   
## JOBManager -1.091e+00 1.641e-01 -6.650 2.93e-11 \*\*\*  
## JOBProfessional -2.958e-01 1.410e-01 -2.098 0.035938 \*   
## JOBStudent -3.133e-01 1.446e-01 -2.166 0.030305 \*   
## JOBz\_Blue Collar -2.020e-01 1.201e-01 -1.682 0.092562 .   
## TRAVTIME 1.557e-02 2.189e-03 7.114 1.13e-12 \*\*\*  
## CAR\_USEPrivate -8.324e-01 1.059e-01 -7.861 3.82e-15 \*\*\*  
## TIF -5.237e-02 8.537e-03 -6.134 8.56e-10 \*\*\*  
## CAR\_TYPEPanel Truck 7.121e-01 1.944e-01 3.664 0.000249 \*\*\*  
## CAR\_TYPEPickup 5.525e-01 1.153e-01 4.793 1.65e-06 \*\*\*  
## CAR\_TYPESports Car 1.090e+00 1.451e-01 7.512 5.84e-14 \*\*\*  
## CAR\_TYPEVan 5.778e-01 1.494e-01 3.866 0.000110 \*\*\*  
## CAR\_TYPEz\_SUV 8.112e-01 1.247e-01 6.508 7.63e-11 \*\*\*  
## RED\_CARyes -2.236e-01 1.031e-01 -2.168 0.030141 \*   
## CLM\_FREQ 2.001e-01 3.318e-02 6.032 1.62e-09 \*\*\*  
## REVOKEDYes 8.551e-01 1.074e-01 7.965 1.65e-15 \*\*\*  
## MVR\_PTS 1.170e-01 1.585e-02 7.383 1.54e-13 \*\*\*  
## URBANICITYz\_Highly Rural/ Rural -2.308e+00 1.244e-01 -18.552 < 2e-16 \*\*\*  
## HOME\_VAL\_num -1.469e-06 4.267e-07 -3.443 0.000575 \*\*\*  
## INCOME\_num -3.471e-06 1.424e-06 -2.439 0.014744 \*   
## BLUEBOOK\_num -2.370e-05 6.031e-06 -3.930 8.51e-05 \*\*\*  
## OLDCLAIM\_num -1.326e-05 4.569e-06 -2.903 0.003700 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 6990.9 on 6044 degrees of freedom  
## Residual deviance: 5365.0 on 6012 degrees of freedom  
## AIC: 5431  
##   
## Number of Fisher Scoring iterations: 5

none = glm(TARGET\_FLAG~1, data=insurance.forstepAIC, family="binomial")  
  
# Tell it to do forward and stepwise below for an AIC model   
final\_model\_forwards = stepAIC(none, scope=list(upper=full\_insurance.2), direction="forward", trace=FALSE)  
final\_model\_both = stepAIC(none, scope=list(upper=full\_insurance.2), direction = "both", trace=FALSE)

summary(final\_model\_forwards)

##   
## Call:  
## glm(formula = TARGET\_FLAG ~ URBANICITY + JOB + MVR\_PTS + CAR\_TYPE +   
## MSTATUS + REVOKED + CAR\_USE + TRAVTIME + TIF + KIDSDRIV +   
## INCOME\_num + CLM\_FREQ + EDUCATION + BLUEBOOK\_num + PARENT1 +   
## HOME\_VAL\_num + OLDCLAIM\_num + RED\_CAR + SEX, family = "binomial",   
## data = insurance.forstepAIC)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5691 -0.7024 -0.3901 0.6201 3.1495   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.905e-01 2.287e-01 -1.707 0.087770 .   
## URBANICITYz\_Highly Rural/ Rural -2.308e+00 1.244e-01 -18.552 < 2e-16 \*\*\*  
## JOBDoctor -9.037e-01 3.273e-01 -2.761 0.005755 \*\*   
## JOBHome Maker -2.872e-01 1.569e-01 -1.831 0.067098 .   
## JOBLawyer -1.650e-01 2.112e-01 -0.781 0.434854   
## JOBManager -1.091e+00 1.641e-01 -6.650 2.93e-11 \*\*\*  
## JOBProfessional -2.958e-01 1.410e-01 -2.098 0.035938 \*   
## JOBStudent -3.133e-01 1.446e-01 -2.166 0.030305 \*   
## JOBz\_Blue Collar -2.020e-01 1.201e-01 -1.682 0.092562 .   
## MVR\_PTS 1.170e-01 1.585e-02 7.383 1.54e-13 \*\*\*  
## CAR\_TYPEPanel Truck 7.121e-01 1.944e-01 3.664 0.000249 \*\*\*  
## CAR\_TYPEPickup 5.525e-01 1.153e-01 4.793 1.65e-06 \*\*\*  
## CAR\_TYPESports Car 1.090e+00 1.451e-01 7.512 5.84e-14 \*\*\*  
## CAR\_TYPEVan 5.778e-01 1.494e-01 3.866 0.000110 \*\*\*  
## CAR\_TYPEz\_SUV 8.112e-01 1.247e-01 6.508 7.63e-11 \*\*\*  
## MSTATUSz\_No 4.146e-01 9.568e-02 4.334 1.47e-05 \*\*\*  
## REVOKEDYes 8.551e-01 1.074e-01 7.965 1.65e-15 \*\*\*  
## CAR\_USEPrivate -8.324e-01 1.059e-01 -7.861 3.82e-15 \*\*\*  
## TRAVTIME 1.557e-02 2.189e-03 7.114 1.13e-12 \*\*\*  
## TIF -5.237e-02 8.537e-03 -6.134 8.56e-10 \*\*\*  
## KIDSDRIV 3.307e-01 6.390e-02 5.176 2.27e-07 \*\*\*  
## INCOME\_num -3.471e-06 1.424e-06 -2.439 0.014744 \*   
## CLM\_FREQ 2.001e-01 3.318e-02 6.032 1.62e-09 \*\*\*  
## EDUCATIONBachelors -3.995e-01 1.231e-01 -3.245 0.001174 \*\*   
## EDUCATIONMasters -4.969e-01 1.905e-01 -2.609 0.009094 \*\*   
## EDUCATIONPhD 4.064e-02 2.461e-01 0.165 0.868809   
## EDUCATIONz\_High School -7.396e-03 1.063e-01 -0.070 0.944554   
## BLUEBOOK\_num -2.370e-05 6.031e-06 -3.930 8.51e-05 \*\*\*  
## PARENT1Yes 4.730e-01 1.089e-01 4.342 1.41e-05 \*\*\*  
## HOME\_VAL\_num -1.469e-06 4.267e-07 -3.443 0.000575 \*\*\*  
## OLDCLAIM\_num -1.326e-05 4.569e-06 -2.903 0.003700 \*\*   
## RED\_CARyes -2.236e-01 1.031e-01 -2.168 0.030141 \*   
## SEXz\_F -1.861e-01 1.272e-01 -1.464 0.143274   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 6990.9 on 6044 degrees of freedom  
## Residual deviance: 5365.0 on 6012 degrees of freedom  
## AIC: 5431  
##   
## Number of Fisher Scoring iterations: 5

summary(final\_model\_both)

##   
## Call:  
## glm(formula = TARGET\_FLAG ~ URBANICITY + JOB + MVR\_PTS + CAR\_TYPE +   
## MSTATUS + REVOKED + CAR\_USE + TRAVTIME + TIF + KIDSDRIV +   
## INCOME\_num + CLM\_FREQ + EDUCATION + BLUEBOOK\_num + PARENT1 +   
## HOME\_VAL\_num + OLDCLAIM\_num + RED\_CAR + SEX, family = "binomial",   
## data = insurance.forstepAIC)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5691 -0.7024 -0.3901 0.6201 3.1495   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.905e-01 2.287e-01 -1.707 0.087770 .   
## URBANICITYz\_Highly Rural/ Rural -2.308e+00 1.244e-01 -18.552 < 2e-16 \*\*\*  
## JOBDoctor -9.037e-01 3.273e-01 -2.761 0.005755 \*\*   
## JOBHome Maker -2.872e-01 1.569e-01 -1.831 0.067098 .   
## JOBLawyer -1.650e-01 2.112e-01 -0.781 0.434854   
## JOBManager -1.091e+00 1.641e-01 -6.650 2.93e-11 \*\*\*  
## JOBProfessional -2.958e-01 1.410e-01 -2.098 0.035938 \*   
## JOBStudent -3.133e-01 1.446e-01 -2.166 0.030305 \*   
## JOBz\_Blue Collar -2.020e-01 1.201e-01 -1.682 0.092562 .   
## MVR\_PTS 1.170e-01 1.585e-02 7.383 1.54e-13 \*\*\*  
## CAR\_TYPEPanel Truck 7.121e-01 1.944e-01 3.664 0.000249 \*\*\*  
## CAR\_TYPEPickup 5.525e-01 1.153e-01 4.793 1.65e-06 \*\*\*  
## CAR\_TYPESports Car 1.090e+00 1.451e-01 7.512 5.84e-14 \*\*\*  
## CAR\_TYPEVan 5.778e-01 1.494e-01 3.866 0.000110 \*\*\*  
## CAR\_TYPEz\_SUV 8.112e-01 1.247e-01 6.508 7.63e-11 \*\*\*  
## MSTATUSz\_No 4.146e-01 9.568e-02 4.334 1.47e-05 \*\*\*  
## REVOKEDYes 8.551e-01 1.074e-01 7.965 1.65e-15 \*\*\*  
## CAR\_USEPrivate -8.324e-01 1.059e-01 -7.861 3.82e-15 \*\*\*  
## TRAVTIME 1.557e-02 2.189e-03 7.114 1.13e-12 \*\*\*  
## TIF -5.237e-02 8.537e-03 -6.134 8.56e-10 \*\*\*  
## KIDSDRIV 3.307e-01 6.390e-02 5.176 2.27e-07 \*\*\*  
## INCOME\_num -3.471e-06 1.424e-06 -2.439 0.014744 \*   
## CLM\_FREQ 2.001e-01 3.318e-02 6.032 1.62e-09 \*\*\*  
## EDUCATIONBachelors -3.995e-01 1.231e-01 -3.245 0.001174 \*\*   
## EDUCATIONMasters -4.969e-01 1.905e-01 -2.609 0.009094 \*\*   
## EDUCATIONPhD 4.064e-02 2.461e-01 0.165 0.868809   
## EDUCATIONz\_High School -7.396e-03 1.063e-01 -0.070 0.944554   
## BLUEBOOK\_num -2.370e-05 6.031e-06 -3.930 8.51e-05 \*\*\*  
## PARENT1Yes 4.730e-01 1.089e-01 4.342 1.41e-05 \*\*\*  
## HOME\_VAL\_num -1.469e-06 4.267e-07 -3.443 0.000575 \*\*\*  
## OLDCLAIM\_num -1.326e-05 4.569e-06 -2.903 0.003700 \*\*   
## RED\_CARyes -2.236e-01 1.031e-01 -2.168 0.030141 \*   
## SEXz\_F -1.861e-01 1.272e-01 -1.464 0.143274   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 6990.9 on 6044 degrees of freedom  
## Residual deviance: 5365.0 on 6012 degrees of freedom  
## AIC: 5431  
##   
## Number of Fisher Scoring iterations: 5

anova(final\_model\_backwards, test="Chisq")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: TARGET\_FLAG  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 6044 6990.9   
## KIDSDRIV 1 42.25 6043 6948.6 8.032e-11 \*\*\*  
## PARENT1 1 121.66 6042 6826.9 < 2.2e-16 \*\*\*  
## MSTATUS 1 32.03 6041 6794.9 1.517e-08 \*\*\*  
## SEX 1 1.26 6040 6793.7 0.261667   
## EDUCATION 4 142.63 6036 6651.0 < 2.2e-16 \*\*\*  
## JOB 7 90.04 6029 6561.0 < 2.2e-16 \*\*\*  
## TRAVTIME 1 10.04 6028 6550.9 0.001530 \*\*   
## CAR\_USE 1 86.88 6027 6464.1 < 2.2e-16 \*\*\*  
## TIF 1 39.38 6026 6424.7 3.488e-10 \*\*\*  
## CAR\_TYPE 5 121.74 6021 6302.9 < 2.2e-16 \*\*\*  
## RED\_CAR 1 0.97 6020 6302.0 0.324811   
## CLM\_FREQ 1 230.56 6019 6071.4 < 2.2e-16 \*\*\*  
## REVOKED 1 80.01 6018 5991.4 < 2.2e-16 \*\*\*  
## MVR\_PTS 1 89.60 6017 5901.8 < 2.2e-16 \*\*\*  
## URBANICITY 1 476.35 6016 5425.5 < 2.2e-16 \*\*\*  
## HOME\_VAL\_num 1 26.75 6015 5398.7 2.317e-07 \*\*\*  
## INCOME\_num 1 9.38 6014 5389.3 0.002193 \*\*   
## BLUEBOOK\_num 1 15.73 6013 5373.6 7.304e-05 \*\*\*  
## OLDCLAIM\_num 1 8.60 6012 5365.0 0.003368 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

G = final\_model\_backwards$null.deviance - final\_model\_backwards$deviance  
Gdf = final\_model\_backwards$df.null - final\_model\_backwards$df.residual  
  
1-pchisq(G, Gdf)

## [1] 0

# how likely this random person that we made up would be to get into an accident   
  
some\_person = data.frame(   
   
 KIDSDRIV = 0,   
 KIDSDRIV = "No",  
 PARENT1 = "No",  
 MSTATUS = "z\_No",   
 SEX="z\_F",   
 EDUCATION = "Masters",   
 JOB= "Professional",   
 TRAVTIME = 15,   
 CAR\_USE = "Private",   
 TIF = 5,   
 CAR\_TYPE = "z\_SUV",   
 RED\_CAR = "no",   
 CLM\_FREQ = 0,   
 REVOKED = "No",   
 MVR\_PTS = 0,   
 URBANICITY = "Highly Urban/ Urban",   
 HOME\_VAL\_num = 258000,  
 INCOME\_num = 82000,   
 BLUEBOOK\_num = 16400,   
 OLDCLAIM\_num = 0  
)

# Predicts logodds  
# Using the final backwards model, this predicts how likely that the person we made up will get into an accident Put type = REsponse if you want probabilitiy   
  
predict(final\_model\_backwards, some\_person)

## 1   
## -2.05658

odds = exp(predict(final\_model\_backwards, some\_person))  
  
odds/(1+odds)

## 1   
## 0.1133892

# Gets teh odds of getting into an accident   
# Same as the predict probability function below

# Predicts probability  
# this predicts the probability that the person we made up will get into an accident  
predict(final\_model\_backwards, some\_person, type = "response")

## 1   
## 0.1133892

predict(final\_model\_forwards, some\_person, type = "response")

## 1   
## 0.1133892

predict(final\_model\_both, some\_person, type = "response")

## 1   
## 0.1133892