Final Project\_Models

David Aarhus

4/29/2020

Students must upload to Canvas a compiled Rmarkdown document that shows one model(s) estimated against the data. Code and replication files are required.

#Libraries

library("tidyverse")

## ── Attaching packages ────────────────────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.2.1 ✓ purrr 0.3.3  
## ✓ tibble 2.1.3 ✓ dplyr 0.8.3  
## ✓ tidyr 1.0.0 ✓ stringr 1.4.0  
## ✓ readr 1.3.1 ✓ forcats 0.5.0

## ── Conflicts ───────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library("ggplot2")  
library("caret")

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

library("doBy")  
library("glmnet")

## Loading required package: Matrix

##   
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack

## Loaded glmnet 3.0-2

library("glmnetUtils")

##   
## Attaching package: 'glmnetUtils'

## The following objects are masked from 'package:glmnet':  
##   
## cv.glmnet, glmnet

library("leaps")

rm(list = ls()) #removing all variables

#Loads in Data

suicide <- read.csv("/Users/DavidAarhus/Documents/310 R/Datasets/suicide.csv")

#Data Transformation

#remove meaning less columns (HDI.for.year)  
suicide$HDI.for.year <- NULL  
suicide$country.year <- NULL  
#create new column (gdpM, gdpB, suicide\_hvy)  
suicide$gdpM <- suicide$gdp\_for\_year..../1e+6  
suicide$gdpB <- suicide$gdp\_for\_year..../1e+9  
suicide$suicide\_hvy <- ifelse(suicide$suicides\_no > 365, 1, 0)

set.seed(310)  
train\_indx <- sample(1:nrow(suicide), 0.70 \* nrow(suicide), replace=FALSE)  
suicide\_train <- suicide[train\_indx, ]  
suicide\_test <- suicide[-train\_indx, ]

#Linear Regression

# We are using a linear model to see the affect of certain variables on the number of suicides in different countries  
Linear\_mod <- lm(suicides\_no ~ ., suicide\_train)  
summary(Linear\_mod)

##   
## Call:  
## lm(formula = suicides\_no ~ ., data = suicide\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3683.5 -97.2 23.1 108.0 16351.4   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.302e+03 2.199e+03 1.502 0.133221   
## countryAntigua and Barbuda 3.910e+01 5.978e+01 0.654 0.513119   
## countryArgentina -3.441e+02 5.928e+01 -5.805 6.52e-09 \*\*\*  
## countryArmenia -2.414e+00 6.208e+01 -0.039 0.968983   
## countryAruba -9.100e+01 7.283e+01 -1.249 0.211531   
## countryAustralia -2.007e+02 6.223e+01 -3.226 0.001259 \*\*   
## countryAustria -2.161e+02 6.137e+01 -3.521 0.000431 \*\*\*  
## countryAzerbaijan -3.620e+01 7.083e+01 -0.511 0.609337   
## countryBahamas 1.413e+01 6.446e+01 0.219 0.826527   
## countryBahrain 1.556e+01 6.336e+01 0.246 0.806014   
## countryBarbados 1.789e+01 6.212e+01 0.288 0.773333   
## countryBelarus -2.314e+02 6.470e+01 -3.576 0.000350 \*\*\*  
## countryBelgium -1.943e+02 6.158e+01 -3.155 0.001605 \*\*   
## countryBelize -1.005e+01 5.979e+01 -0.168 0.866553   
## countryBosnia and Herzegovina 1.926e+01 1.504e+02 0.128 0.898116   
## countryBrazil -1.454e+03 6.707e+01 -21.680 < 2e-16 \*\*\*  
## countryBulgaria -1.471e+02 5.958e+01 -2.469 0.013554 \*   
## countryCabo Verde -8.138e+01 1.922e+02 -0.423 0.671986   
## countryCanada -2.830e+02 6.266e+01 -4.517 6.31e-06 \*\*\*  
## countryChile -1.370e+02 5.860e+01 -2.337 0.019451 \*   
## countryColombia -3.559e+02 5.928e+01 -6.003 1.97e-09 \*\*\*  
## countryCosta Rica -3.777e+01 5.860e+01 -0.645 0.519217   
## countryCroatia -1.583e+02 6.441e+01 -2.458 0.013979 \*   
## countryCuba -1.980e+02 6.242e+01 -3.173 0.001511 \*\*   
## countryCyprus -3.685e+00 6.948e+01 -0.053 0.957697   
## countryCzech Republic -1.605e+02 6.183e+01 -2.595 0.009467 \*\*   
## countryDenmark -1.561e+02 6.899e+01 -2.263 0.023643 \*   
## countryDominica 2.409e+01 2.678e+02 0.090 0.928345   
## countryEcuador -8.674e+01 5.895e+01 -1.471 0.141199   
## countryEl Salvador -7.597e+01 6.198e+01 -1.226 0.220281   
## countryEstonia -2.342e+02 6.376e+01 -3.673 0.000240 \*\*\*  
## countryFiji 5.474e+00 8.017e+01 0.068 0.945557   
## countryFinland -2.041e+02 6.293e+01 -3.243 0.001186 \*\*   
## countryFrance -2.115e+02 6.490e+01 -3.259 0.001120 \*\*   
## countryGeorgia -2.047e+01 6.385e+01 -0.321 0.748482   
## countryGermany -3.676e+02 6.846e+01 -5.369 7.99e-08 \*\*\*  
## countryGreece -9.317e+01 5.919e+01 -1.574 0.115529   
## countryGrenada 3.328e+01 6.121e+01 0.544 0.586697   
## countryGuatemala -7.224e+01 5.890e+01 -1.227 0.219990   
## countryGuyana -1.682e+02 6.239e+01 -2.696 0.007027 \*\*   
## countryHungary -2.526e+02 6.227e+01 -4.057 5.00e-05 \*\*\*  
## countryIceland -1.207e+02 6.222e+01 -1.940 0.052397 .   
## countryIreland -1.073e+02 6.148e+01 -1.745 0.080940 .   
## countryIsrael -9.349e+01 6.027e+01 -1.551 0.120866   
## countryItaly -4.885e+02 6.177e+01 -7.908 2.76e-15 \*\*\*  
## countryJamaica 2.447e+01 6.747e+01 0.363 0.716793   
## countryJapan 2.701e+02 7.341e+01 3.679 0.000235 \*\*\*  
## countryKazakhstan -2.460e+02 6.207e+01 -3.963 7.44e-05 \*\*\*  
## countryKiribati -2.109e+01 7.945e+01 -0.266 0.790613   
## countryKuwait -4.302e+00 6.296e+01 -0.068 0.945533   
## countryKyrgyzstan -1.062e+02 6.040e+01 -1.759 0.078613 .   
## countryLatvia -2.437e+02 6.580e+01 -3.703 0.000213 \*\*\*  
## countryLithuania -3.159e+02 6.569e+01 -4.808 1.53e-06 \*\*\*  
## countryLuxembourg -2.154e+02 7.054e+01 -3.054 0.002264 \*\*   
## countryMacau -1.340e+02 1.924e+02 -0.696 0.486145   
## countryMaldives 4.484e+01 7.979e+01 0.562 0.574088   
## countryMalta -1.281e+01 5.968e+01 -0.215 0.830005   
## countryMauritius -7.330e+01 5.805e+01 -1.263 0.206713   
## countryMexico -8.963e+02 6.209e+01 -14.435 < 2e-16 \*\*\*  
## countryMongolia -1.826e+02 2.453e+02 -0.744 0.456620   
## countryMontenegro -3.315e+00 7.853e+01 -0.042 0.966324   
## countryNetherlands -1.639e+02 6.130e+01 -2.674 0.007504 \*\*   
## countryNew Zealand -1.188e+02 6.097e+01 -1.949 0.051346 .   
## countryNicaragua -3.221e+01 9.500e+01 -0.339 0.734607   
## countryNorway -1.526e+02 6.696e+01 -2.279 0.022656 \*   
## countryOman 1.465e+00 1.208e+02 0.012 0.990325   
## countryPanama -2.910e+01 6.159e+01 -0.473 0.636546   
## countryParaguay -2.639e+01 6.024e+01 -0.438 0.661410   
## countryPhilippines -7.505e+02 7.334e+01 -10.233 < 2e-16 \*\*\*  
## countryPoland -1.979e+02 6.276e+01 -3.152 0.001622 \*\*   
## countryPortugal -1.160e+02 6.158e+01 -1.884 0.059572 .   
## countryPuerto Rico -8.452e+01 5.979e+01 -1.414 0.157489   
## countryQatar -4.716e+01 7.890e+01 -0.598 0.550005   
## countryRepublic of Korea -2.758e+02 6.002e+01 -4.596 4.34e-06 \*\*\*  
## countryRomania -1.750e+02 5.966e+01 -2.934 0.003355 \*\*   
## countryRussian Federation 1.513e+03 6.807e+01 22.226 < 2e-16 \*\*\*  
## countrySaint Kitts and Nevis 4.450e+01 1.398e+02 0.318 0.750322   
## countrySaint Lucia -3.574e+01 5.938e+01 -0.602 0.547237   
## countrySaint Vincent and Grenadines 8.385e+00 6.077e+01 0.138 0.890252   
## countrySan Marino -7.288e+01 1.345e+02 -0.542 0.587933   
## countrySerbia -1.681e+02 6.595e+01 -2.549 0.010814 \*   
## countrySeychelles -4.301e+01 6.565e+01 -0.655 0.512427   
## countrySingapore -1.727e+02 6.283e+01 -2.748 0.005997 \*\*   
## countrySlovakia -8.378e+01 6.334e+01 -1.323 0.185908   
## countrySlovenia -2.291e+02 6.590e+01 -3.477 0.000508 \*\*\*  
## countrySouth Africa -4.579e+02 6.585e+01 -6.953 3.68e-12 \*\*\*  
## countrySpain -3.602e+02 6.065e+01 -5.939 2.92e-09 \*\*\*  
## countrySri Lanka -2.397e+02 7.672e+01 -3.124 0.001787 \*\*   
## countrySuriname -1.732e+02 5.992e+01 -2.891 0.003842 \*\*   
## countrySweden -1.417e+02 6.306e+01 -2.246 0.024687 \*   
## countrySwitzerland -1.918e+02 7.202e+01 -2.662 0.007764 \*\*   
## countryThailand -4.986e+02 6.082e+01 -8.198 2.60e-16 \*\*\*  
## countryTrinidad and Tobago -1.018e+02 6.122e+01 -1.663 0.096237 .   
## countryTurkey -6.881e+02 9.081e+01 -7.578 3.68e-14 \*\*\*  
## countryTurkmenistan -5.342e+01 5.921e+01 -0.902 0.366941   
## countryUkraine -8.663e+01 6.217e+01 -1.394 0.163448   
## countryUnited Arab Emirates -5.228e+01 9.597e+01 -0.545 0.585927   
## countryUnited Kingdom -4.696e+02 6.339e+01 -7.408 1.33e-13 \*\*\*  
## countryUnited States -6.475e+02 1.149e+02 -5.633 1.79e-08 \*\*\*  
## countryUruguay -1.380e+02 6.058e+01 -2.277 0.022793 \*   
## countryUzbekistan -1.887e+02 6.415e+01 -2.942 0.003269 \*\*   
## year -1.706e+00 1.108e+00 -1.541 0.123395   
## sexmale 6.982e+01 9.824e+00 7.107 1.23e-12 \*\*\*  
## age25-34 years 3.039e+01 1.782e+01 1.706 0.088105 .   
## age35-54 years 4.912e+01 2.814e+01 1.746 0.080841 .   
## age5-14 years -3.050e+01 1.866e+01 -1.634 0.102175   
## age55-74 years 7.751e+01 4.180e+01 1.854 0.063712 .   
## age75+ years 1.321e+01 4.856e+01 0.272 0.785645   
## population 1.474e-04 2.583e-06 57.077 < 2e-16 \*\*\*  
## suicides.100k.pop 1.080e+01 3.285e-01 32.863 < 2e-16 \*\*\*  
## gdp\_for\_year.... -2.448e-11 8.446e-12 -2.899 0.003751 \*\*   
## gdp\_per\_capita.... 1.393e-03 5.938e-04 2.345 0.019016 \*   
## generationG.I. Generation -2.732e+01 3.639e+01 -0.751 0.452854   
## generationGeneration X -2.290e+01 2.140e+01 -1.070 0.284626   
## generationGeneration Z 1.762e+01 4.693e+01 0.375 0.707375   
## generationMillenials -8.299e+00 3.279e+01 -0.253 0.800200   
## generationSilent -1.191e+01 2.372e+01 -0.502 0.615527   
## gdpM NA NA NA NA   
## gdpB NA NA NA NA   
## suicide\_hvy 3.867e+02 1.887e+01 20.493 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 589.3 on 19356 degrees of freedom  
## Multiple R-squared: 0.5876, Adjusted R-squared: 0.5851   
## F-statistic: 235.7 on 117 and 19356 DF, p-value: < 2.2e-16

#This linear model helps point out the factors other than countries that play a key role in the amount of suicides  
Linear\_mod\_noCountries <- lm(suicides\_no ~ year + sex + age   
 + population + suicides.100k.pop   
 + gdp\_for\_year.... + gdp\_per\_capita....  
 + generation + gdpM + gdpB, suicide\_train)  
summary(Linear\_mod\_noCountries)

##   
## Call:  
## lm(formula = suicides\_no ~ year + sex + age + population + suicides.100k.pop +   
## gdp\_for\_year.... + gdp\_per\_capita.... + generation + gdpM +   
## gdpB, data = suicide\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3044.9 -136.6 26.8 157.8 18016.2   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.619e+03 2.228e+03 1.175 0.239896   
## year -1.434e+00 1.124e+00 -1.276 0.202053   
## sexmale 6.736e+01 1.050e+01 6.414 1.45e-10 \*\*\*  
## age25-34 years 3.639e+01 2.002e+01 1.817 0.069170 .   
## age35-54 years 9.190e+01 3.144e+01 2.923 0.003475 \*\*   
## age5-14 years -4.753e+01 2.080e+01 -2.285 0.022298 \*   
## age55-74 years 9.890e+01 4.693e+01 2.107 0.035101 \*   
## age75+ years -2.024e+01 5.446e+01 -0.372 0.710190   
## population 1.450e-04 1.838e-06 78.847 < 2e-16 \*\*\*  
## suicides.100k.pop 1.405e+01 3.014e-01 46.617 < 2e-16 \*\*\*  
## gdp\_for\_year.... -1.904e-11 5.034e-12 -3.782 0.000156 \*\*\*  
## gdp\_per\_capita.... 8.278e-04 2.829e-04 2.926 0.003432 \*\*   
## generationG.I. Generation -6.146e+01 4.087e+01 -1.504 0.132623   
## generationGeneration X -2.397e+01 2.406e+01 -0.997 0.318971   
## generationGeneration Z 2.714e+01 5.272e+01 0.515 0.606726   
## generationMillenials -1.703e+00 3.684e+01 -0.046 0.963125   
## generationSilent -2.289e+01 2.664e+01 -0.859 0.390167   
## gdpM NA NA NA NA   
## gdpB NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 663.5 on 19457 degrees of freedom  
## Multiple R-squared: 0.4746, Adjusted R-squared: 0.4742   
## F-statistic: 1099 on 16 and 19457 DF, p-value: < 2.2e-16

##Logistic Regression

# this logistic regression model helps identify what variables contribute to a country being "suicide heavy" (365 or more suicides per year)  
logit\_fit <- glm(suicide\_hvy ~ ., family = binomial, suicide\_train)

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(logit\_fit)

##   
## Call:  
## glm(formula = suicide\_hvy ~ ., family = binomial, data = suicide\_train)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.006895 0.000000 0.000000 0.000000 0.007046   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -4.147e+02 1.204e+05 -0.003 0.997  
## countryAntigua and Barbuda 4.662e+00 1.568e+04 0.000 1.000  
## countryArgentina -1.052e+03 1.927e+04 -0.055 0.956  
## countryArmenia 1.986e+01 1.606e+04 0.001 0.999  
## countryAruba -1.057e+01 1.883e+04 -0.001 1.000  
## countryAustralia -1.045e+03 1.808e+04 -0.058 0.954  
## countryAustria -1.099e+03 1.845e+04 -0.060 0.953  
## countryAzerbaijan 2.304e+01 1.615e+04 0.001 0.999  
## countryBahamas -1.883e+01 1.647e+04 -0.001 0.999  
## countryBahrain -3.064e+00 1.684e+04 0.000 1.000  
## countryBarbados 1.654e+00 1.615e+04 0.000 1.000  
## countryBelarus -1.086e+03 1.682e+05 -0.006 0.995  
## countryBelgium -1.077e+03 1.846e+04 -0.058 0.953  
## countryBelize 1.318e+01 1.548e+04 0.001 0.999  
## countryBosnia and Herzegovina 2.668e+01 4.293e+04 0.001 1.000  
## countryBrazil -9.127e+02 1.857e+04 -0.049 0.961  
## countryBulgaria -1.062e+03 2.119e+04 -0.050 0.960  
## countryCabo Verde 2.635e+01 4.761e+04 0.001 1.000  
## countryCanada -1.068e+03 1.758e+04 -0.061 0.952  
## countryChile -1.051e+03 2.082e+04 -0.050 0.960  
## countryColombia -9.924e+02 2.759e+04 -0.036 0.971  
## countryCosta Rica -9.100e+01 2.090e+04 -0.004 0.997  
## countryCroatia -6.942e+02 2.256e+04 -0.031 0.975  
## countryCuba -1.061e+03 1.815e+04 -0.058 0.953  
## countryCyprus -1.275e+01 1.999e+04 -0.001 0.999  
## countryCzech Republic -9.686e+02 2.292e+04 -0.042 0.966  
## countryDenmark -5.870e+02 2.433e+04 -0.024 0.981  
## countryDominica 3.749e+02 2.554e+04 0.015 0.988  
## countryEcuador -5.644e+02 2.135e+04 -0.026 0.979  
## countryEl Salvador -1.365e+02 2.158e+04 -0.006 0.995  
## countryEstonia -4.950e+02 2.405e+04 -0.021 0.984  
## countryFiji 1.379e+01 2.162e+04 0.001 0.999  
## countryFinland -9.925e+02 2.053e+04 -0.048 0.961  
## countryFrance -9.816e+02 2.292e+04 -0.043 0.966  
## countryGeorgia 1.832e+01 1.586e+04 0.001 0.999  
## countryGermany -9.798e+02 1.680e+04 -0.058 0.954  
## countryGreece -2.698e+02 2.153e+04 -0.013 0.990  
## countryGrenada 1.292e+01 1.589e+04 0.001 0.999  
## countryGuatemala -1.118e+02 2.119e+04 -0.005 0.996  
## countryGuyana 5.128e+00 1.562e+04 0.000 1.000  
## countryHungary -1.064e+03 1.868e+04 -0.057 0.955  
## countryIceland -5.821e+01 1.770e+04 -0.003 0.997  
## countryIreland -4.145e+02 2.181e+04 -0.019 0.985  
## countryIsrael -1.415e+02 2.114e+04 -0.007 0.995  
## countryItaly -9.729e+02 1.666e+04 -0.058 0.953  
## countryJamaica 1.471e+01 1.839e+04 0.001 0.999  
## countryJapan -1.014e+03 1.686e+04 -0.060 0.952  
## countryKazakhstan -1.071e+03 1.838e+04 -0.058 0.954  
## countryKiribati 1.819e+01 2.424e+04 0.001 0.999  
## countryKuwait -3.113e+01 1.487e+04 -0.002 0.998  
## countryKyrgyzstan -3.577e+02 2.181e+04 -0.016 0.987  
## countryLatvia -9.576e+02 2.433e+04 -0.039 0.969  
## countryLithuania -1.007e+03 1.955e+04 -0.051 0.959  
## countryLuxembourg -1.231e+02 1.762e+04 -0.007 0.994  
## countryMacau -3.452e+00 6.733e+04 0.000 1.000  
## countryMaldives 1.627e+01 2.265e+04 0.001 0.999  
## countryMalta -3.459e+00 1.871e+04 0.000 1.000  
## countryMauritius 4.208e+00 1.477e+04 0.000 1.000  
## countryMexico -1.014e+03 1.848e+04 -0.055 0.956  
## countryMongolia -1.188e+02 9.580e+04 -0.001 0.999  
## countryMontenegro 1.541e+01 2.301e+04 0.001 0.999  
## countryNetherlands -9.303e+02 3.494e+04 -0.027 0.979  
## countryNew Zealand -1.941e+02 2.106e+04 -0.009 0.993  
## countryNicaragua -5.991e+00 2.633e+04 0.000 1.000  
## countryNorway -3.170e+02 2.071e+04 -0.015 0.988  
## countryOman 7.909e+00 3.724e+04 0.000 1.000  
## countryPanama 1.069e+01 1.453e+04 0.001 0.999  
## countryParaguay -3.333e+01 2.061e+04 -0.002 0.999  
## countryPhilippines -9.701e+02 2.070e+04 -0.047 0.963  
## countryPoland -1.085e+03 1.837e+04 -0.059 0.953  
## countryPortugal -8.315e+02 2.189e+04 -0.038 0.970  
## countryPuerto Rico -1.203e+02 2.134e+04 -0.006 0.996  
## countryQatar -7.719e+01 2.041e+04 -0.004 0.997  
## countryRepublic of Korea -1.041e+03 1.778e+04 -0.059 0.953  
## countryRomania -1.043e+03 1.782e+04 -0.059 0.953  
## countryRussian Federation -1.384e+03 1.944e+04 -0.071 0.943  
## countrySaint Kitts and Nevis 1.324e+01 4.504e+04 0.000 1.000  
## countrySaint Lucia 1.393e+01 1.552e+04 0.001 0.999  
## countrySaint Vincent and Grenadines 1.412e+01 1.632e+04 0.001 0.999  
## countrySan Marino -6.821e+01 3.571e+04 -0.002 0.998  
## countrySerbia -9.898e+02 1.851e+04 -0.053 0.957  
## countrySeychelles 6.307e+00 1.642e+04 0.000 1.000  
## countrySingapore -6.067e+01 1.695e+04 -0.004 0.997  
## countrySlovakia -7.699e+02 2.312e+04 -0.033 0.973  
## countrySlovenia -3.737e+02 2.402e+04 -0.016 0.988  
## countrySouth Africa -2.333e+02 2.348e+04 -0.010 0.992  
## countrySpain -1.006e+03 1.738e+04 -0.058 0.954  
## countrySri Lanka -1.003e+03 1.711e+05 -0.006 0.995  
## countrySuriname 8.329e+00 1.880e+04 0.000 1.000  
## countrySweden -1.078e+03 1.884e+04 -0.057 0.954  
## countrySwitzerland -1.096e+03 2.394e+04 -0.046 0.963  
## countryThailand -1.046e+03 1.810e+04 -0.058 0.954  
## countryTrinidad and Tobago 8.200e-01 1.707e+04 0.000 1.000  
## countryTurkey -6.998e+02 2.091e+04 -0.033 0.973  
## countryTurkmenistan -1.380e+02 1.964e+04 -0.007 0.994  
## countryUkraine -1.048e+03 1.996e+04 -0.052 0.958  
## countryUnited Arab Emirates -2.475e+01 2.506e+04 -0.001 0.999  
## countryUnited Kingdom -1.005e+03 1.741e+04 -0.058 0.954  
## countryUnited States -5.108e+02 1.652e+04 -0.031 0.975  
## countryUruguay -2.038e+02 2.134e+04 -0.010 0.992  
## countryUzbekistan -1.047e+03 1.798e+04 -0.058 0.954  
## year 7.088e-03 5.764e+01 0.000 1.000  
## sexmale -2.776e+01 1.748e+03 -0.016 0.987  
## age25-34 years -8.979e+00 6.068e+02 -0.015 0.988  
## age35-54 years -3.385e+00 1.615e+03 -0.002 0.998  
## age5-14 years 3.408e+02 1.304e+04 0.026 0.979  
## age55-74 years -9.437e-01 1.587e+03 -0.001 1.000  
## age75+ years -3.902e+01 6.657e+03 -0.006 0.995  
## suicides\_no 4.112e+00 1.805e+01 0.228 0.820  
## population -7.040e-06 4.883e-04 -0.014 0.988  
## suicides.100k.pop -2.367e-01 5.171e+01 -0.005 0.996  
## gdp\_for\_year.... -4.488e-11 1.572e-09 -0.029 0.977  
## gdp\_per\_capita.... 1.281e-03 4.294e-02 0.030 0.976  
## generationG.I. Generation 4.055e+00 6.131e+03 0.001 0.999  
## generationGeneration X -1.040e+01 4.000e+02 -0.026 0.979  
## generationGeneration Z 6.062e+00 1.139e+04 0.001 1.000  
## generationMillenials 1.734e+01 1.168e+03 0.015 0.988  
## generationSilent -1.057e+01 5.924e+02 -0.018 0.986  
## gdpM NA NA NA NA  
## gdpB NA NA NA NA  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1.5039e+04 on 19473 degrees of freedom  
## Residual deviance: 3.0780e-04 on 19356 degrees of freedom  
## AIC: 236  
##   
## Number of Fisher Scoring iterations: 25

# this logistic regression model helps identify what variables contribute being "suicide heavy" (365 or more suicides per year)  
logit\_fit\_noCountries <- glm(suicide\_hvy ~ year + sex + age   
 + population + suicides.100k.pop   
 + gdp\_for\_year.... + gdp\_per\_capita....  
 + generation + gdpM + gdpB , family = binomial, suicide\_train)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(logit\_fit\_noCountries)

##   
## Call:  
## glm(formula = suicide\_hvy ~ year + sex + age + population + suicides.100k.pop +   
## gdp\_for\_year.... + gdp\_per\_capita.... + generation + gdpM +   
## gdpB, family = binomial, data = suicide\_train)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.7466 -0.2399 -0.0605 0.0000 7.0701   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 5.127e+01 1.617e+01 3.171 0.001521 \*\*   
## year -2.920e-02 8.165e-03 -3.577 0.000348 \*\*\*  
## sexmale 2.953e+00 1.273e-01 23.189 < 2e-16 \*\*\*  
## age25-34 years 2.023e-01 1.504e-01 1.346 0.178450   
## age35-54 years 1.065e+00 2.252e-01 4.729 2.26e-06 \*\*\*  
## age5-14 years -2.799e+01 1.353e+00 -20.681 < 2e-16 \*\*\*  
## age55-74 years 9.563e-01 3.317e-01 2.883 0.003935 \*\*   
## age75+ years -1.322e-01 3.966e-01 -0.333 0.738827   
## population 7.001e-07 2.223e-08 31.497 < 2e-16 \*\*\*  
## suicides.100k.pop 4.196e-02 1.739e-03 24.132 < 2e-16 \*\*\*  
## gdp\_for\_year.... 1.513e-12 8.656e-14 17.477 < 2e-16 \*\*\*  
## gdp\_per\_capita.... -2.328e-05 3.439e-06 -6.769 1.30e-11 \*\*\*  
## generationG.I. Generation -4.915e-01 2.824e-01 -1.741 0.081765 .   
## generationGeneration X 4.882e-02 1.638e-01 0.298 0.765668   
## generationGeneration Z -2.054e+01 2.111e+02 -0.097 0.922502   
## generationMillenials 1.978e-02 2.741e-01 0.072 0.942464   
## generationSilent -2.972e-01 1.735e-01 -1.712 0.086821 .   
## gdpM NA NA NA NA   
## gdpB NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 15039.0 on 19473 degrees of freedom  
## Residual deviance: 5330.3 on 19457 degrees of freedom  
## AIC: 5364.3  
##   
## Number of Fisher Scoring iterations: 17

#Forward Stepwise Model

# This forward stepwise model helps identify which variables contribute the most to the amount of suicides.   
forward\_mod <- regsubsets(suicides\_no ~ year + sex   
 + gdp\_for\_year.... + gdp\_per\_capita....  
 + generation + gdpM + gdpB, data = suicide\_train, nvmax = 6, method = "forward")

## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in =  
## force.in, : 2 linear dependencies found

summary(forward\_mod)

## Subset selection object  
## Call: regsubsets.formula(suicides\_no ~ year + sex + gdp\_for\_year.... +   
## gdp\_per\_capita.... + generation + gdpM + gdpB, data = suicide\_train,   
## nvmax = 6, method = "forward")  
## 11 Variables (and intercept)  
## Forced in Forced out  
## year FALSE FALSE  
## sexmale FALSE FALSE  
## gdp\_for\_year.... FALSE FALSE  
## gdp\_per\_capita.... FALSE FALSE  
## generationG.I. Generation FALSE FALSE  
## generationGeneration X FALSE FALSE  
## generationGeneration Z FALSE FALSE  
## generationMillenials FALSE FALSE  
## generationSilent FALSE FALSE  
## gdpM FALSE FALSE  
## gdpB FALSE FALSE  
## 1 subsets of each size up to 6  
## Selection Algorithm: forward  
## year sexmale gdp\_for\_year.... gdp\_per\_capita....  
## 1 ( 1 ) " " " " "\*" " "   
## 2 ( 1 ) " " "\*" "\*" " "   
## 3 ( 1 ) " " "\*" "\*" " "   
## 4 ( 1 ) " " "\*" "\*" " "   
## 5 ( 1 ) " " "\*" "\*" "\*"   
## 6 ( 1 ) " " "\*" "\*" "\*"   
## generationG.I. Generation generationGeneration X  
## 1 ( 1 ) " " " "   
## 2 ( 1 ) " " " "   
## 3 ( 1 ) " " " "   
## 4 ( 1 ) " " " "   
## 5 ( 1 ) " " " "   
## 6 ( 1 ) " " "\*"   
## generationGeneration Z generationMillenials generationSilent gdpM gdpB  
## 1 ( 1 ) " " " " " " " " " "   
## 2 ( 1 ) " " " " " " " " " "   
## 3 ( 1 ) " " "\*" " " " " " "   
## 4 ( 1 ) "\*" "\*" " " " " " "   
## 5 ( 1 ) "\*" "\*" " " " " " "   
## 6 ( 1 ) "\*" "\*" " " " " " "