Advanced Statistics and Machine learning. Case study: Cancer death rate

Allan Kouidri

August 2022

Advanced Statistics and Machine learning

This project aim using the cancer_reg.csv https://data.world/exercises/linear-regression-exercise-1/workspace/file?filename=cancer_reg.csv to predict "TARGET_deathRate". Several models will be use such as ordinary least square, CART and Random Forest, and we will compare them

On this dataset, we first perform an ordinary least square model to explain the target_deathrate variable thanks to the numerical ones.

```
#Importing libraries
library(knitr)
library(missMDA)
library(glmnet)
library(MASS)
library(rpart)
library(rpart.plot)
library(randomForest)
library(VSURF)
```

```
#Importing the dataset
data <-
  read.csv(
    file = "./cancer_reg.csv",
    dec = ".",
    header = TRUE
    )
```

We first have a quick overview of the data sumarry statistic (output table not shown)

```
summary(data)
```

From this summary we can see:

(i) We have two character variables: "binnedinc" and "geography". They will be removed as we will work only with numerical ones.

(ii) We have missing values for three variables: pctsomecol18_24 pctsomecol18_24 and pctprivatecoveragealone

We will keep only numerical variables.

```
#Removing variables
data2 <- data
data2$binnedinc <- NULL
data2$geography <- NULL
```

Now we will impute missing values.

```
data3 <- data2
data3 <- as.data.frame(imputePCA(data2)$completeObs)

cat('number missing values in data2:',sum(is.na(data2)),'\n')
```

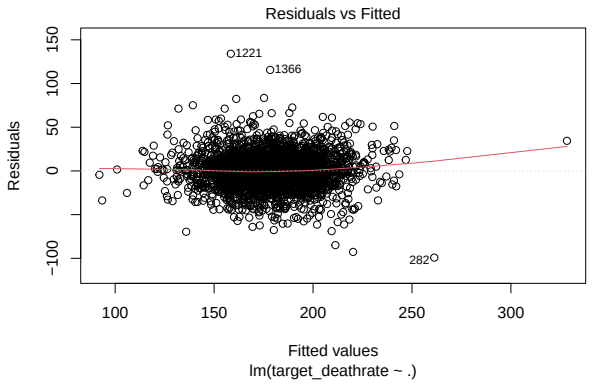
```
## number missing values in data2: 3046
```

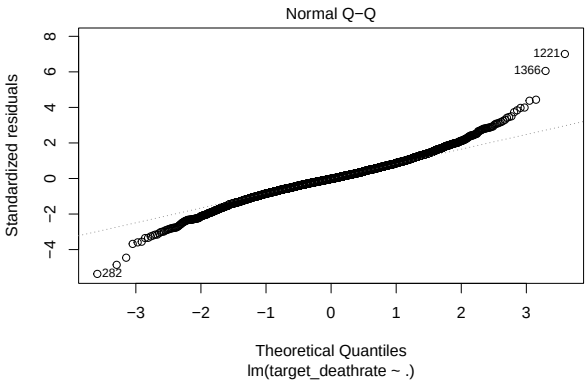
```
cat('number missing values in data3:',sum(is.na(data3)),'\n')
```

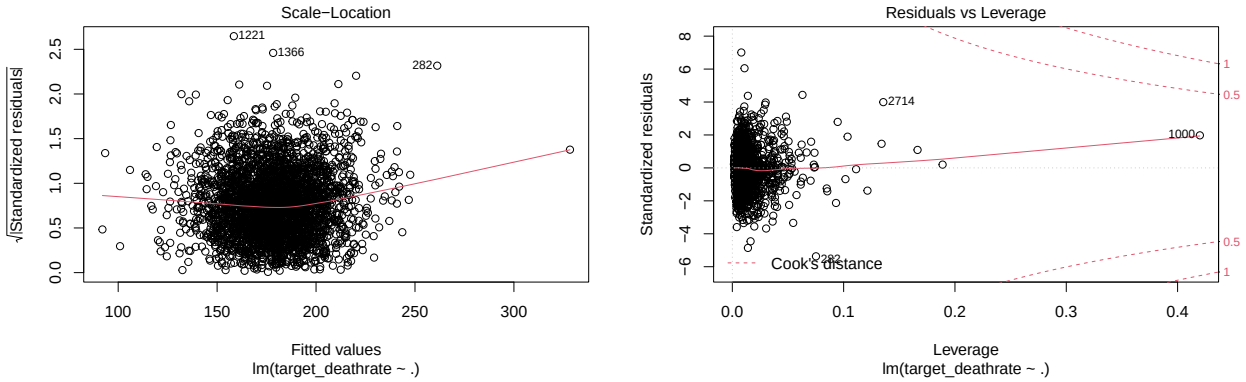
```
## number missing values in data3: 0
```

Let's look for potential outliers

```
model0 = lm(target_deathrate~.,data = data3)
plot(model0)
```







Looking at the model0 plot, Residuals vs Fitted, we have some potential outliers: points 282,1221,1366. We will suppress them.

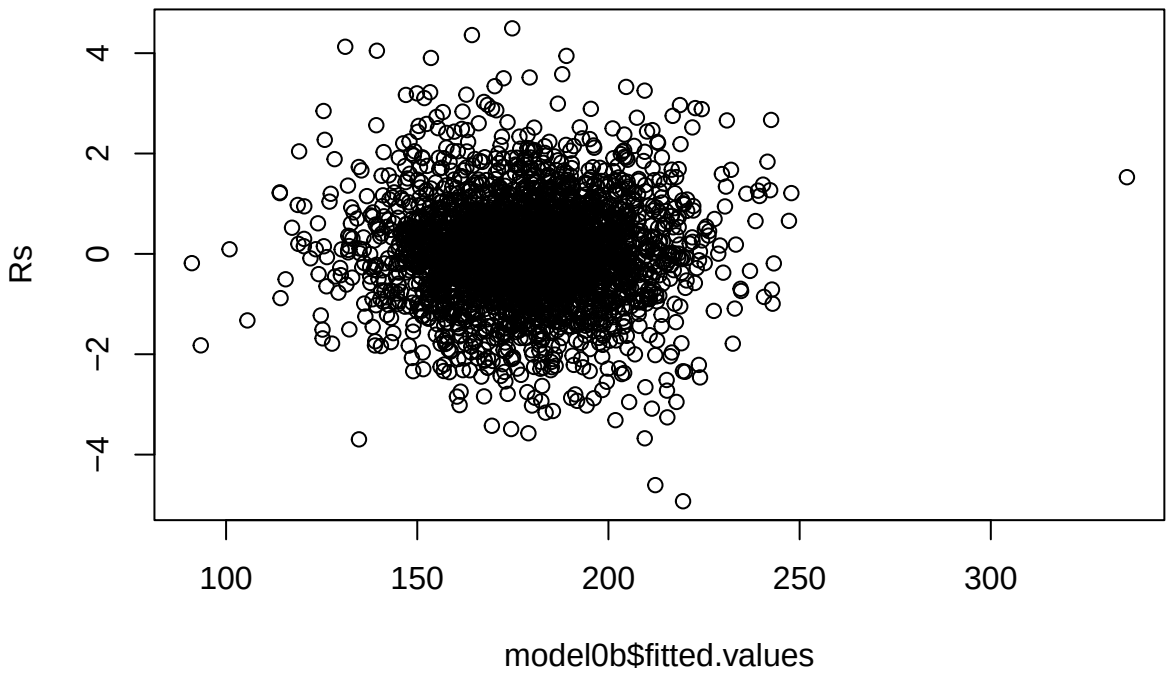
```
data4=data3[-c(282,1221,1366),]
```

Let's repeat the step after suppressing the selected observations.

```
model0b = lm(target_deathrate~.,data = data4)
```

More precisely if we want to see some outliers we should consider the studentized residuals.

```
Rs=rstudent(model0b)
plot(model0b$fitted.values, Rs)
```



We see several points >2 and <-2 that we would would like to suppress them from the studies.

```
#Getting the list of the points to remove
Rs_df<-(as.data.frame((Rs)))
row.names(Rs_df)[which(Rs_df>2)]
```

```
##  [1] "79"   "116"  "122"  "166"  "209"  "250"  "254"  "458"  "466"  "469"
## [11] "472"  "484"  "495"  "515"  "522"  "537"  "549"  "554"  "562"  "564"
## [21] "627"  "666"  "670"  "690"  "727"  "775"  "780"  "786"  "975"  "979"
## [31] "1000" "1076" "1174" "1204" "1217" "1236" "1261" "1276" "1297" "1310"
## [41] "1316" "1390" "1442" "1497" "1513" "1542" "1548" "1856" "1866" "1882"
## [51] "1884" "1897" "1914" "1958" "1962" "2001" "2016" "2027" "2036" "2040"
## [61] "2047" "2048" "2079" "2135" "2174" "2176" "2267" "2541" "2549" "2563"
## [71] "2587" "2590" "2596" "2598" "2600" "2637" "2682" "2714" "2726" "2727"
## [81] "2757" "2810" "2812" "2819" "2825" "2842" "2858" "3022" "3034" "3036"
## [91] "3040"
```

```
row.names(Rs_df)[which(Rs_df< -2)]
```

```
##  [1] "34"   "69"   "105"  "119"  "120"  "124"  "176"  "189"  "256"  "264"
## [11] "415"  "476"  "514"  "556"  "616"  "621"  "625"  "650"  "748"  "783"
## [21] "803"  "812"  "845"  "912"  "913"  "920"  "921"  "925"  "993"  "1048"
## [31] "1058" "1059" "1130" "1160" "1195" "1249" "1290" "1311" "1331" "1345"
## [41] "1405" "1429" "1445" "1560" "1568" "1580" "1686" "1701" "1708" "1777"
## [51] "1797" "1942" "1965" "1969" "2010" "2018" "2051" "2065" "2066" "2307"
## [61] "2311" "2312" "2318" "2328" "2344" "2351" "2353" "2386" "2404" "2427"
```

```
## [71] "2440" "2444" "2546" "2593" "2626" "2642" "2646" "2659" "2661" "2669"
## [81] "2674" "2696" "2720" "2734" "2741" "2789" "2809" "2822" "2985" "2988"
```

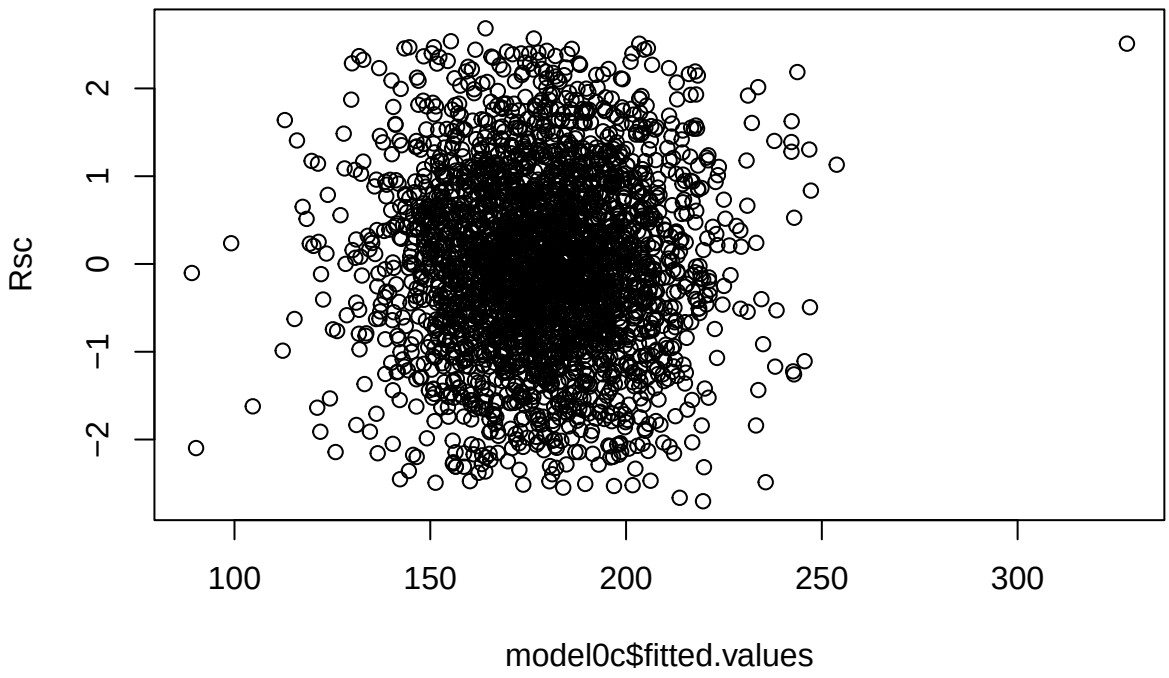
```
# Removing the columns
data5=data3[-c(1221,1366,282,2646, 31 , 79 , 116 , 122 ,166,  209 , 250 , 254 , 458 ,
               466 , 469 , 472 ,
               484 , 495 , 515 , 522 , 537 , 549 , 554 , 562 , 564 , 627 , 666 ,
               670 , 690 , 727 , 775 , 780 , 786 , 975 , 979 , 1000 , 1076 , 1174
               , 1204 , 1217 , 1221 , 1236 , 1261 , 1276 , 1297 , 1310 , 1316 ,
               1390 , 1442 , 1497 , 1513 , 1542 , 1548 , 1856 , 1866 , 1882 , 1884 ,
               1897 , 1914 , 1958 , 1962 , 2001 , 2016 , 2027 , 2036 , 2040 ,
               2048 , 2079 , 2135 , 2174 , 2176 , 2267 , 2549 , 2563 , 2587 ,
               2590 , 2596 , 2598 , 2600 , 2637 , 2673 , 2682 , 2714 , 2726 ,
               2727 , 2757 , 2810 , 2812 , 2819 , 2825 , 2842 , 2858 , 3022 ,
               3034 , 3036 , 3040 , 34 , 69 , 105 , 119 , 120 , 124 , 176 ,
               189 , 256 , 264 , 415 , 476 , 514 , 556 , 616 , 621 , 625 ,
               650 , 748 , 783 , 803 , 812 , 845 , 912 , 913 , 920 , 921 ,
               925 , 1048 , 1058 , 1059 , 1130 , 1160 , 1195 , 1249 , 1290 ,
               1311 , 1345 , 1405 , 1429 , 1445 , 1560 , 1568 , 1580 , 1686 ,
               1701 , 1708 , 1777 , 1797 , 1942 , 1965 , 1969 , 2010 , 2018 ,
               2051 , 2065 , 2066 , 2307 , 2311 , 2312 , 2318 , 2328 , 2344 ,
               2351 , 2353 , 2386 , 2404 , 2427 , 2440 , 2444 , 2546 , 2593 ,
               2626 , 2642 , 2659 , 2661 , 2669 , 2674 , 2696 , 2720 , 2734 ,
               2741 , 2789 , 2809 , 2822 , 2985 , 2988 ),]
```

```
#number of rows removed
nrow(data)-nrow(data5)
```

```
## [1] 182
```

Let's check again for potential outliers

```
model0c = lm(target_deathrate~.,data = data5)
Rsc=rstudent(model0c)
plot(model0c$fitted.values, Rsc)
```



Based on the plotting of the the studentized residuals, we do not see any potential outliers. We can processed to the modeling.

We split our dataset into train and test.

```
#Full dataset
set.seed(seed = 1703)
splitvector <-
  sample(
    x = c("learning", "test"),
    size = nrow(x = data3),
    replace = TRUE,
    prob = c(0.7, 0.3)
)

learningset3 <- data3[splitvector == "learning", ]
testset3 <- data3[splitvector == "test", ]
```

```
#Dataset with outliers removed
set.seed(seed = 1703)
splitvector <-
  sample(
    x = c("learning", "test"),
    size = nrow(x = data5),
    replace = TRUE,
    prob = c(0.7, 0.3)
)
```

```
learningset5 <- data5[splitvector == "learning", ]
testset5 <- data5[splitvector == "test", ]
```

Ordinary least square lm

We will test OLS models before and after removing potential outliers.

```
model1 = lm(target_deathrate~.,data = learningset3)
p1 = predict(model1, newdata = testset3)
cat('OLS with potential outliers:',sqrt(mean((p1-testset3$target_deathrate)**2)),'\n')
```

```
## OLS with potential outliers: 19.25947
```

```
model1b = lm(target_deathrate~.,data = learningset5)
p1b = predict(model1b, newdata = testset5)
cat('OLS without outliers:',sqrt(mean((p1b-testset5$target_deathrate)**2)),'\n')
```

```
## OLS without outliers: 15.17405
```

RMSE score after removing potential outliers is better. We will keep the the learningset5 and testset5 for making futher models.

Then we perform variable selection by using a step by step method at first and a penalized one then.

- Step by step:

```
model2=step(model1)
```

```
## Start:  AIC=12678.07
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + studypercap + medianage +
##     medianagemale + medianagefemale + percentmarried + pctnohs18_24 +
##     pcths18_24 + pctsomecol18_24 + pctbachdeg18_24 + pcths25_over +
##     pctbachdeg25_over + pctemployed16_over + pctunemployed16_over +
##     pctprivatecoverage + pctprivatecoveragealone + pctempprivcoverage +
##     pctpubliccoverage + pctpubliccoveragealone + pctwhite + pctblack +
##     pctasian + pctotherrace + pctmarriedhouseholds + birthrate
##
##                            Df Sum of Sq    RSS   AIC
## - studypercap               1         8 781141 12676
## - medianage                 1        31 781165 12676
## - pctasian                  1        37 781171 12676
## - pctpubliccoveragealone    1       117 781251 12676
## - pctbachdeg18_24           1       175 781308 12676
## - pctblack                  1       266 781399 12677
## - pctunemployed16_over      1       273 781406 12677
## - pctsomecol18_24           1       344 781478 12677
## - pctprivatecoveragealone   1       358 781491 12677
```

```
## - pctpubliccoverage        1       409 781543 12677
## - medianagefemale          1       604 781737 12678
## - povertypercent           1       612 781746 12678
## <none>                                 781133 12678
## - pctwhite                 1       929 782062 12679
## - medianagemale            1      1059 782192 12679
## - medincome                1      1305 782438 12680
## - popest2015               1      1621 782754 12680
## - pctempprivcoverage       1      2691 783825 12683
## - pctprivatecoverage       1      2718 783851 12684
## - pctnohs18_24             1      3061 784195 12684
## - avganncount              1      3417 784551 12685
## - avgdeathsperyear         1      4188 785321 12688
## - pcths18_24               1      4388 785521 12688
## - birthrate                1      4635 785768 12689
## - pctbachdeg25_over        1      7988 789121 12698
## - pctemployed16_over       1      8422 789555 12699
## - pcths25_over             1      9696 790829 12702
## - pctotherrace             1     10114 791248 12704
## - percentmarried           1     16433 797567 12721
## - pctmarriedhouseholds     1     18967 800100 12727
## - incidencerate            1    167692 948825 13092
##
## Step:  AIC=12676.09
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianage + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pctbachdeg18_24 + pcths25_over + pctbachdeg25_over +
##     pctemployed16_over + pctunemployed16_over + pctprivatecoverage +
##     pctprivatecoveragealone + pctempprivcoverage + pctpubliccoverage +
##     pctpubliccoveragealone + pctwhite + pctblack + pctasian +
##     pctotherrace + pctmarriedhouseholds + birthrate
##
##                           Df Sum of Sq    RSS   AIC
## - medianage                1        31 781172 12674
## - pctasian                 1        39 781180 12674
## - pctpubliccoveragealone   1       117 781258 12674
## - pctbachdeg18_24          1       176 781317 12675
## - pctblack                 1       265 781406 12675
## - pctunemployed16_over     1       274 781415 12675
## - pctsomecol18_24          1       343 781484 12675
## - pctprivatecoveragealone  1       360 781501 12675
## - pctpubliccoverage        1       407 781548 12675
## - medianagefemale          1       604 781744 12676
## - povertypercent           1       608 781749 12676
## <none>                                 781141 12676
## - pctwhite                 1       925 782066 12677
## - medianagemale            1      1060 782201 12677
## - medincome                1      1297 782438 12678
## - popest2015               1      1624 782765 12678
## - pctprivatecoverage       1      2714 783855 12682
## - pctempprivcoverage       1      2715 783856 12682
## - pctnohs18_24             1      3059 784200 12682
## - avganncount              1      3410 784551 12683
```

```
## - avgdeathsperyear         1      4183 785324 12686
## - pcths18_24               1      4410 785551 12686
## - birthrate                1      4628 785768 12687
## - pctbachdeg25_over        1      7982 789123 12696
## - pctemployed16_over       1      8414 789555 12697
## - pcths25_over             1      9692 790833 12700
## - pctotherrace             1     10112 791253 12702
## - percentmarried           1     16433 797574 12719
## - pctmarriedhouseholds     1     18960 800101 12725
## - incidencerate            1    168041 949182 13091
##
## Step:  AIC=12674.18
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pctbachdeg18_24 + pcths25_over + pctbachdeg25_over +
##     pctemployed16_over + pctunemployed16_over + pctprivatecoverage +
##     pctprivatecoveragealone + pctempprivcoverage + pctpubliccoverage +
##     pctpubliccoveragealone + pctwhite + pctblack + pctasian +
##     pctotherrace + pctmarriedhouseholds + birthrate
##
##                           Df Sum of Sq    RSS   AIC
## - pctasian                 1        38 781210 12672
## - pctpubliccoveragealone   1       117 781289 12672
## - pctbachdeg18_24          1       179 781351 12673
## - pctblack                 1       264 781436 12673
## - pctunemployed16_over     1       278 781450 12673
## - pctsomecol18_24          1       342 781514 12673
## - pctprivatecoveragealone  1       363 781535 12673
## - pctpubliccoverage        1       411 781582 12673
## - medianagefemale          1       599 781771 12674
## - povertypercent           1       609 781781 12674
## <none>                                 781172 12674
## - pctwhite                 1       920 782092 12675
## - medianagemale            1      1049 782221 12675
## - medincome                1      1301 782472 12676
## - popest2015               1      1620 782792 12677
## - pctprivatecoverage       1      2710 783882 12680
## - pctempprivcoverage       1      2714 783885 12680
## - pctnohs18_24             1      3064 784235 12680
## - avganncount              1      3415 784587 12682
## - avgdeathsperyear         1      4180 785352 12684
## - pcths18_24               1      4426 785597 12684
## - birthrate                1      4613 785785 12685
## - pctbachdeg25_over        1      8000 789172 12694
## - pctemployed16_over       1      8400 789571 12695
## - pcths25_over             1      9685 790857 12698
## - pctotherrace             1     10097 791269 12700
## - percentmarried           1     16435 797607 12717
## - pctmarriedhouseholds     1     18991 800162 12724
## - incidencerate            1    168083 949255 13089
##
## Step:  AIC=12672.28
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
```

```
##     medincome + popest2015 + povertypercent + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pctbachdeg18_24 + pcths25_over + pctbachdeg25_over +
##     pctemployed16_over + pctunemployed16_over + pctprivatecoverage +
##     pctprivatecoveragealone + pctempprivcoverage + pctpubliccoverage +
##     pctpubliccoveragealone + pctwhite + pctblack + pctotherrace +
##     pctmarriedhouseholds + birthrate
##
##                            Df Sum of Sq    RSS   AIC
## - pctpubliccoveragealone    1       116 781326 12671
## - pctbachdeg18_24           1       166 781375 12671
## - pctunemployed16_over      1       273 781483 12671
## - pctsomecol18_24           1       327 781536 12671
## - pctprivatecoveragealone   1       346 781556 12671
## - pctblack                  1       399 781609 12671
## - pctpubliccoverage         1       402 781611 12671
## - medianagefemale           1       597 781807 12672
## - povertypercent            1       633 781842 12672
## <none>                                  781210 12672
## - medianagemale             1      1046 782256 12673
## - pctwhite                  1      1293 782503 12674
## - medincome                 1      1397 782607 12674
## - popest2015                1      1587 782796 12675
## - pctprivatecoverage        1      2722 783932 12678
## - pctempprivcoverage        1      2732 783941 12678
## - pctnohs18_24              1      3107 784317 12679
## - avganncount               1      3430 784639 12680
## - avgdeathsperyear          1      4157 785366 12682
## - pcths18_24                1      4460 785670 12682
## - birthrate                 1      4669 785878 12683
## - pctbachdeg25_over         1      8021 789231 12692
## - pctemployed16_over        1      8512 789722 12694
## - pcths25_over              1      9647 790857 12696
## - pctotherrace              1     10190 791399 12698
## - percentmarried            1     16501 797710 12715
## - pctmarriedhouseholds      1     19119 800329 12722
## - incidencerate             1    168047 949257 13087
##
## Step:  AIC=12670.6
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pctbachdeg18_24 + pcths25_over + pctbachdeg25_over +
##     pctemployed16_over + pctunemployed16_over + pctprivatecoverage +
##     pctprivatecoveragealone + pctempprivcoverage + pctpubliccoverage +
##     pctwhite + pctblack + pctotherrace + pctmarriedhouseholds +
##     birthrate
##
##                            Df Sum of Sq    RSS   AIC
## - pctbachdeg18_24           1       144 781470 12669
## - pctprivatecoveragealone   1       257 781583 12669
## - pctunemployed16_over      1       286 781612 12669
## - pctsomecol18_24           1       337 781663 12670
## - pctblack                  1       378 781704 12670
```

```
## - pctpubliccoverage        1       443 781768 12670
## - povertypercent           1       697 782023 12670
## <none>                                 781326 12671
## - medianagefemale          1       751 782077 12671
## - medianagemale            1      1117 782443 12672
## - pctwhite                 1      1315 782641 12672
## - medincome                1      1516 782842 12673
## - popest2015               1      1589 782915 12673
## - pctnohs18_24             1      3171 784497 12677
## - pctempprivcoverage       1      3174 784500 12677
## - avganncount              1      3433 784759 12678
## - avgdeathsperyear         1      4172 785498 12680
## - pcths18_24               1      4515 785841 12681
## - birthrate                1      4634 785960 12681
## - pctprivatecoverage       1      6560 787886 12686
## - pctbachdeg25_over        1      7930 789256 12690
## - pctemployed16_over       1      8407 789733 12692
## - pcths25_over             1      9931 791257 12696
## - pctotherrace             1     10159 791485 12696
## - percentmarried           1     16392 797718 12713
## - pctmarriedhouseholds     1     19011 800337 12720
## - incidencerate            1    169834 951160 13089
##
## Step:  AIC=12668.99
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pcths25_over + pctbachdeg25_over + pctemployed16_over +
##     pctunemployed16_over + pctprivatecoverage + pctprivatecoveragealone +
##     pctempprivcoverage + pctpubliccoverage + pctwhite + pctblack +
##     pctotherrace + pctmarriedhouseholds + birthrate
##
##                            Df Sum of Sq    RSS   AIC
## - pctprivatecoveragealone   1       277 781748 12668
## - pctsomecol18_24           1       297 781767 12668
## - pctunemployed16_over      1       321 781792 12668
## - pctblack                  1       384 781855 12668
## - pctpubliccoverage         1       429 781899 12668
## - povertypercent            1       643 782114 12669
## <none>                                 781470 12669
## - medianagefemale           1       745 782215 12669
## - medianagemale             1      1177 782648 12670
## - pctwhite                  1      1328 782798 12671
## - medincome                 1      1395 782866 12671
## - popest2015                1      1554 783024 12671
## - pctnohs18_24              1      3029 784499 12675
## - pctempprivcoverage        1      3224 784694 12676
## - avganncount               1      3393 784863 12676
## - avgdeathsperyear          1      4083 785553 12678
## - birthrate                 1      4587 786058 12680
## - pcths18_24                1      5272 786743 12681
## - pctprivatecoverage        1      6577 788047 12685
## - pctemployed16_over        1      8409 789880 12690
## - pctbachdeg25_over         1      8461 789931 12690
```

```
## - pcths25_over             1      9810 791281 12694
## - pctotherrace             1     10371 791841 12695
## - percentmarried           1     16261 797731 12711
## - pctmarriedhouseholds     1     18916 800387 12718
## - incidencerate            1    169896 951367 13088
##
## Step:  AIC=12667.75
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pcths25_over + pctbachdeg25_over + pctemployed16_over +
##     pctunemployed16_over + pctprivatecoverage + pctempprivcoverage +
##     pctpubliccoverage + pctwhite + pctblack + pctotherrace +
##     pctmarriedhouseholds + birthrate
##
##                         Df Sum of Sq    RSS   AIC
## - pctpubliccoverage      1       268 782016 12666
## - pctunemployed16_over   1       341 782089 12667
## - pctsomecol18_24        1       343 782091 12667
## - pctblack               1       402 782150 12667
## - povertypercent         1       584 782332 12667
## - medianagefemale        1       730 782477 12668
## <none>                                781748 12668
## - medianagemale          1      1128 782875 12669
## - medincome              1      1298 783045 12669
## - pctwhite               1      1430 783178 12670
## - popest2015             1      1584 783331 12670
## - pctnohs18_24           1      2929 784676 12674
## - pctempprivcoverage     1      3005 784752 12674
## - avganncount            1      3394 785142 12675
## - avgdeathsperyear       1      4084 785832 12677
## - birthrate              1      4556 786303 12678
## - pcths18_24             1      5396 787144 12680
## - pctbachdeg25_over      1      8871 790619 12690
## - pctemployed16_over     1      9103 790851 12690
## - pcths25_over           1      9861 791609 12693
## - pctotherrace           1     10325 792073 12694
## - pctprivatecoverage     1     12187 793935 12699
## - percentmarried         1     16236 797984 12710
## - pctmarriedhouseholds   1     19130 800878 12717
## - incidencerate          1    170172 951920 13087
##
## Step:  AIC=12666.48
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pcths25_over + pctbachdeg25_over + pctemployed16_over +
##     pctunemployed16_over + pctprivatecoverage + pctempprivcoverage +
##     pctwhite + pctblack + pctotherrace + pctmarriedhouseholds +
##     birthrate
##
##                         Df Sum of Sq    RSS   AIC
## - pctunemployed16_over   1       262 782278 12665
## - pctsomecol18_24        1       303 782319 12665
```

```
## - pctblack              1       332 782347 12665
## - povertypercent        1       501 782516 12666
## <none>                              782016 12666
## - medianagefemale       1      1038 783054 12667
## - medianagemale         1      1200 783215 12668
## - medincome             1      1351 783367 12668
## - pctwhite              1      1435 783451 12668
## - popest2015            1      1519 783534 12669
## - pctnohs18_24          1      2704 784720 12672
## - pctempprivcoverage    1      3251 785267 12673
## - avganncount           1      3471 785486 12674
## - avgdeathsperyear      1      4016 786032 12675
## - birthrate             1      4639 786655 12677
## - pcths18_24            1      5786 787801 12680
## - pctbachdeg25_over     1      8627 790643 12688
## - pctemployed16_over    1      8845 790860 12688
## - pcths25_over          1      9723 791738 12691
## - pctotherrace          1     10473 792488 12693
## - pctprivatecoverage    1     12107 794122 12697
## - percentmarried        1     15980 797996 12708
## - pctmarriedhouseholds  1     18921 800936 12716
## - incidencerate         1    172516 954532 13091
##
## Step:  AIC=12665.2
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pcths25_over + pctbachdeg25_over + pctemployed16_over +
##     pctprivatecoverage + pctempprivcoverage + pctwhite + pctblack +
##     pctotherrace + pctmarriedhouseholds + birthrate
##
##                         Df Sum of Sq    RSS   AIC
## - pctsomecol18_24        1       308 782586 12664
## - pctblack               1       313 782591 12664
## - povertypercent         1       487 782765 12664
## <none>                               782278 12665
## - medianagemale          1      1111 783389 12666
## - medianagefemale        1      1202 783480 12666
## - medincome              1      1458 783736 12667
## - popest2015             1      1604 783882 12668
## - pctwhite               1      1607 783885 12668
## - pctnohs18_24           1      2918 785195 12671
## - pctempprivcoverage     1      3443 785720 12673
## - avganncount            1      3496 785773 12673
## - avgdeathsperyear       1      4202 786480 12675
## - birthrate              1      4676 786954 12676
## - pcths18_24             1      5802 788080 12679
## - pctbachdeg25_over      1      8984 791262 12688
## - pcths25_over           1      9492 791769 12689
## - pctotherrace           1     10902 793180 12693
## - pctemployed16_over     1     11250 793528 12694
## - pctprivatecoverage     1     13520 795798 12700
## - percentmarried         1     16149 798427 12707
## - pctmarriedhouseholds   1     19331 801609 12715
```

```
## - incidencerate         1    176045 958323 13097
##
## Step:  AIC=12664.04
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pcths25_over + pctbachdeg25_over + pctemployed16_over + pctprivatecoverage +
##     pctempprivcoverage + pctwhite + pctblack + pctotherrace +
##     pctmarriedhouseholds + birthrate
##
##                         Df Sum of Sq    RSS   AIC
## - pctblack               1       289 782875 12663
## - povertypercent         1       520 783106 12664
## <none>                               782586 12664
## - medianagemale          1      1120 783706 12665
## - medianagefemale        1      1130 783716 12665
## - medincome              1      1426 784012 12666
## - pctwhite               1      1545 784131 12666
## - popest2015             1      1586 784171 12666
## - pctnohs18_24           1      2610 785196 12669
## - pctempprivcoverage     1      3361 785947 12671
## - avganncount            1      3526 786112 12672
## - avgdeathsperyear       1      4138 786724 12673
## - birthrate              1      4758 787344 12675
## - pcths18_24             1      8415 791001 12685
## - pctbachdeg25_over      1      9142 791727 12687
## - pcths25_over           1      9711 792297 12688
## - pctotherrace           1     10959 793545 12692
## - pctemployed16_over     1     11359 793944 12693
## - pctprivatecoverage     1     13914 796500 12700
## - percentmarried         1     16544 799130 12707
## - pctmarriedhouseholds   1     19560 802146 12715
## - incidencerate          1    176966 959551 13098
##
## Step:  AIC=12662.83
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pcths25_over + pctbachdeg25_over + pctemployed16_over + pctprivatecoverage +
##     pctempprivcoverage + pctwhite + pctotherrace + pctmarriedhouseholds +
##     birthrate
##
##                         Df Sum of Sq    RSS   AIC
## - povertypercent         1       446 783321 12662
## <none>                               782875 12663
## - medianagemale          1      1042 783917 12664
## - medianagefemale        1      1334 784209 12664
## - popest2015             1      1563 784438 12665
## - medincome              1      1592 784467 12665
## - pctwhite               1      1690 784565 12665
## - pctnohs18_24           1      2692 785567 12668
## - pctempprivcoverage     1      3227 786102 12670
## - avganncount            1      3400 786274 12670
## - avgdeathsperyear       1      4038 786912 12672
```

```
## - birthrate             1      4630 787505 12673
## - pcths18_24            1      8496 791371 12684
## - pctbachdeg25_over     1      8943 791818 12685
## - pcths25_over          1      9888 792763 12688
## - pctotherrace          1     10805 793680 12690
## - pctemployed16_over    1     11398 794272 12692
## - pctprivatecoverage    1     14703 797578 12701
## - percentmarried        1     16687 799562 12706
## - pctmarriedhouseholds  1     19713 802588 12714
## - incidencerate         1    176982 959857 13097
##
## Step:  AIC=12662.05
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + medianagemale + medianagefemale +
##     percentmarried + pctnohs18_24 + pcths18_24 + pcths25_over +
##     pctbachdeg25_over + pctemployed16_over + pctprivatecoverage +
##     pctempprivcoverage + pctwhite + pctotherrace + pctmarriedhouseholds +
##     birthrate
##
##                         Df Sum of Sq    RSS   AIC
## <none>                               783321 12662
## - medincome              1      1152 784473 12663
## - medianagemale          1      1153 784474 12663
## - popest2015             1      1380 784702 12664
## - medianagefemale        1      1452 784773 12664
## - pctwhite               1      2200 785521 12666
## - pctnohs18_24           1      2767 786088 12668
## - avganncount            1      3354 786675 12669
## - pctempprivcoverage     1      3606 786927 12670
## - avgdeathsperyear       1      3753 787074 12670
## - birthrate              1      4567 787888 12672
## - pcths18_24             1      8234 791556 12682
## - pctbachdeg25_over      1      8726 792047 12684
## - pcths25_over           1      9639 792960 12686
## - pctotherrace           1     11527 794848 12691
## - pctemployed16_over     1     15072 798393 12701
## - percentmarried         1     16575 799897 12705
## - pctprivatecoverage     1     19731 803052 12713
## - pctmarriedhouseholds   1     20316 803637 12715
## - incidencerate          1    178057 961378 13098
```

```
p2=predict(model2, newdata = testset5)
```

```
cat('Step default parameters:',sqrt(mean((p2-testset5$target_deathrate)^2)),'\n')
```

```
## Step default parameters: 15.17311
```

```
stepforward=stepAIC(lm(target_deathrate~1,data=learningset5), scope=list(upper=model1b, lower=~1),test=
```

```
## Start:  AIC=12995.14
## target_deathrate ~ 1
##
```

```
##                            Df Sum of Sq     RSS   AIC F Value     Pr(F)
## + pctbachdeg25_over         1    414144  907950 12245  911.80 < 2.2e-16 ***
## + pctpubliccoveragealone    1    351669  970424 12378  724.41 < 2.2e-16 ***
## + incidencerate             1    323962  998132 12435  648.81 < 2.2e-16 ***
## + medincome                 1    321789 1000304 12439  643.06 < 2.2e-16 ***
## + povertypercent            1    320240 1001854 12442  638.97 < 2.2e-16 ***
## + pctemployed16_over        1    319930 1002164 12443  638.16 < 2.2e-16 ***
## + pctprivatecoveragealone   1    295694 1026400 12491  575.89 < 2.2e-16 ***
## + pcths25_over              1    290475 1031618 12501  562.86 < 2.2e-16 ***
## + pctpubliccoverage         1    285120 1036973 12511  549.63 < 2.2e-16 ***
## + pctprivatecoverage        1    279040 1043053 12523  534.78 < 2.2e-16 ***
## + pctunemployed16_over      1    264941 1057152 12550  500.99 < 2.2e-16 ***
## + pctsomecol18_24           1    202121 1119973 12665  360.76 < 2.2e-16 ***
## + pctbachdeg18_24           1    154200 1167893 12749  263.93 < 2.2e-16 ***
## + pctempprivcoverage        1    139295 1182798 12774  235.42 < 2.2e-16 ***
## + pctmarriedhouseholds      1    137423 1184671 12778  231.89 < 2.2e-16 ***
## + pctblack                  1    116766 1205327 12812  193.65 < 2.2e-16 ***
## + percentmarried            1    114794 1207299 12815  190.07 < 2.2e-16 ***
## + pcths18_24                1    105008 1217085 12832  172.47 < 2.2e-16 ***
## + pctasian                  1     70131 1251963 12888  111.98 < 2.2e-16 ***
## + pctotherrace              1     60870 1261223 12903   96.48 < 2.2e-16 ***
## + pctwhite                  1     51948 1270145 12917   81.76 < 2.2e-16 ***
## + avganncount               1     44374 1277719 12929   69.42 < 2.2e-16 ***
## + popest2015                1     30226 1291868 12951   46.77 1.057e-11 ***
## + pctnohs18_24              1     26142 1295952 12957   40.32 2.657e-10 ***
## + avgdeathsperyear          1     17417 1304676 12971   26.69 2.631e-07 ***
## + birthrate                 1     14623 1307470 12975   22.36 2.421e-06 ***
## + studypercap               1      1971 1320122 12994    2.98    0.0842 .
## <none>                                  1322093 12995
## + medianagefemale           1       635 1321458 12996    0.96    0.3270
## + medianage                 1       352 1321741 12997    0.53    0.4655
## + medianagemale             1       317 1321776 12997    0.48    0.4887
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=12245.2
## target_deathrate ~ pctbachdeg25_over
##
##                            Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + incidencerate             1    282742 625207 11501  903.57 < 2.2e-16 ***
## + pctmarriedhouseholds      1     97624 810325 12020  240.71 < 2.2e-16 ***
## + pctunemployed16_over      1     85647 822303 12049  208.10 < 2.2e-16 ***
## + percentmarried            1     78703 829247 12066  189.63 < 2.2e-16 ***
## + povertypercent            1     70201 837749 12086  167.43 < 2.2e-16 ***
## + pctblack                  1     65385 842564 12098  155.05 < 2.2e-16 ***
## + pctpubliccoveragealone    1     62053 845896 12106  146.57 < 2.2e-16 ***
## + pctotherrace              1     48485 859464 12137  112.71 < 2.2e-16 ***
## + pctwhite                  1     43412 864538 12149  100.33 < 2.2e-16 ***
## + pctemployed16_over        1     42342 865608 12152   97.73 < 2.2e-16 ***
## + birthrate                 1     34965 872985 12169   80.02 < 2.2e-16 ***
## + pctprivatecoverage        1     29367 878583 12181   66.78 5.318e-16 ***
## + medincome                 1     24379 883570 12193   55.13 1.662e-13 ***
## + pctpubliccoverage         1     22864 885086 12196   51.61 9.502e-13 ***
## + pctprivatecoveragealone   1     19567 888383 12204   44.01 4.202e-11 ***
```

```
## + pctnohs18_24             1     13429 894521 12217   30.00 4.878e-08 ***
## + medianagemale            1     13049 894901 12218   29.13 7.560e-08 ***
## + medianagefemale          1     10547 897403 12224   23.48 1.358e-06 ***
## + avgdeathsperyear         1      8640 899309 12228   19.20 1.241e-05 ***
## + pcths25_over             1      7197 900753 12231   15.96 6.690e-05 ***
## + pctsomecol18_24          1      5272 902678 12236   11.67 0.0006482 ***
## + popest2015               1      3636 904314 12239    8.03 0.0046396 **
## + pcths18_24               1      3566 904383 12239    7.88 0.0050497 **
## + pctasian                 1      1328 906622 12244    2.93 0.0872703 .
## + avganncount              1      1030 906919 12245    2.27 0.1320392
## <none>                                907950 12245
## + studypercap              1       693 907257 12246    1.53 0.2167610
## + pctempprivcoverage       1       560 907390 12246    1.23 0.2668803
## + medianage                1       440 907509 12246    0.97 0.3248616
## + pctbachdeg18_24          1       273 907676 12247    0.60 0.4378970
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=11500.62
## target_deathrate ~ pctbachdeg25_over + incidencerate
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + povertypercent           1     82876 542331 11218 305.171 < 2.2e-16 ***
## + pctprivatecoverage       1     69407 555800 11267 249.382 < 2.2e-16 ***
## + pctunemployed16_over     1     63809 561398 11287 226.982 < 2.2e-16 ***
## + pctmarriedhouseholds     1     59333 565874 11303 209.390 < 2.2e-16 ***
## + pctpubliccoveragealone   1     59313 565895 11303 209.310 < 2.2e-16 ***
## + pctemployed16_over       1     53810 571397 11322 188.063 < 2.2e-16 ***
## + percentmarried           1     51930 573277 11329 180.899 < 2.2e-16 ***
## + pctprivatecoveragealone  1     50014 575193 11336 173.643 < 2.2e-16 ***
## + pctwhite                 1     43944 581264 11357 150.974 < 2.2e-16 ***
## + pctblack                 1     42460 582747 11362 145.506 < 2.2e-16 ***
## + medincome                1     33949 591258 11391 114.664 < 2.2e-16 ***
## + pctpubliccoverage        1     20734 604473 11435  68.500 2.292e-16 ***
## + pctempprivcoverage       1     20172 605035 11437  66.581 5.874e-16 ***
## + birthrate                1     16085 609122 11450  52.734 5.448e-13 ***
## + medianagemale            1     12381 612826 11463  40.346 2.627e-10 ***
## + pctotherrace             1     10561 614646 11468  34.313 5.475e-09 ***
## + pctsomecol18_24          1      9830 615378 11471  31.898 1.857e-08 ***
## + medianagefemale          1      9655 615552 11472  31.323 2.487e-08 ***
## + pcths18_24               1      4823 620384 11487  15.527 8.415e-05 ***
## + avgdeathsperyear         1      1781 623426 11497   5.705   0.01701 *
## + pctbachdeg18_24          1      1033 624174 11499   3.305   0.06920 .
## + popest2015               1       909 624299 11500   2.906   0.08840 .
## + medianage                1       819 624389 11500   2.618   0.10580
## + avganncount              1       685 624522 11500   2.192   0.13891
## <none>                                625207 11501
## + pctasian                 1       504 624704 11501   1.610   0.20457
## + pctnohs18_24             1       371 624837 11501   1.185   0.27652
## + studypercap              1       344 624863 11502   1.100   0.29440
## + pcths25_over             1        20 625187 11503   0.063   0.80136
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Step:  AIC=11218.06
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent
##
##                            Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + pctotherrace              1   15355.1 526976 11163  58.160 3.715e-14 ***
## + pcths18_24                1   15316.6 527015 11163  58.010 4.000e-14 ***
## + pcths25_over              1   12767.4 529564 11172  48.122 5.391e-12 ***
## + birthrate                 1   10741.9 531589 11180  40.333 2.644e-10 ***
## + pctunemployed16_over      1   10453.2 531878 11181  39.228 4.605e-10 ***
## + pctsomecol18_24           1    5274.7 537056 11200  19.604 1.004e-05 ***
## + pctprivatecoverage        1    5195.5 537136 11201  19.307 1.172e-05 ***
## + pctemployed16_over        1    4506.4 537825 11203  16.724 4.494e-05 ***
## + pctmarriedhouseholds      1    4061.6 538270 11205  15.061 0.0001075 ***
## + pctblack                  1    3514.7 538816 11207  13.020 0.0003158 ***
## + pctpubliccoveragealone    1    3088.1 539243 11209  11.431 0.0007364 ***
## + pctwhite                  1    2527.1 539804 11211   9.344 0.0022664 **
## + pctnohs18_24              1    2167.1 540164 11212   8.008 0.0047046 **
## + pctprivatecoveragealone   1    1488.4 540843 11215   5.493 0.0191905 *
## + avganncount               1    1420.3 540911 11215   5.241 0.0221667 *
## + percentmarried            1     890.6 541441 11217   3.283 0.0701480 .
## + pctempprivcoverage        1     669.1 541662 11218   2.466 0.1165082
## <none>                                  542331 11218
## + studypercap               1     469.9 541861 11218   1.731 0.1884492
## + medincome                 1     444.9 541886 11218   1.639 0.2006494
## + medianagefemale           1     285.2 542046 11219   1.050 0.3055957
## + medianage                 1     236.5 542095 11219   0.871 0.3508218
## + medianagemale             1     234.7 542096 11219   0.864 0.3527207
## + avgdeathsperyear          1     117.1 542214 11220   0.431 0.5115890
## + pctbachdeg18_24           1      62.7 542268 11220   0.231 0.6310308
## + pctpubliccoverage         1      20.7 542310 11220   0.076 0.7825391
## + pctasian                  1      15.1 542316 11220   0.056 0.8136809
## + popest2015                1      14.4 542317 11220   0.053 0.8180698
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=11162.59
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace
##
##                            Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + pcths18_24                1   14001.0 512975 11111  54.451 2.327e-13 ***
## + pctprivatecoverage        1   11233.9 515742 11122  43.455 5.536e-11 ***
## + pctunemployed16_over      1   11165.4 515811 11122  43.184 6.339e-11 ***
## + birthrate                 1    9021.6 517954 11130  34.748 4.395e-09 ***
## + pctwhite                  1    6977.1 519999 11138  26.768 2.524e-07 ***
## + pctpubliccoveragealone    1    5053.4 521923 11145  19.316 1.166e-05 ***
## + pcths25_over              1    5040.7 521935 11145  19.267 1.196e-05 ***
## + pctmarriedhouseholds      1    4289.2 522687 11148  16.371 5.406e-05 ***
## + pctsomecol18_24           1    3672.4 523304 11151  14.000 0.0001879 ***
## + pctemployed16_over        1    2954.3 524022 11153  11.247 0.0008124 ***
## + pctblack                  1    2937.4 524039 11153  11.183 0.0008410 ***
## + medianagefemale           1    2865.3 524111 11154  10.907 0.0009751 ***
## + medianagemale             1    2749.1 524227 11154  10.462 0.0012385 **
## + medincome                 1    2632.1 524344 11155  10.014 0.0015766 **
```

```
## + percentmarried           1    2126.9 524849 11156   8.085 0.0045095 **
## + pctprivatecoveragealone  1    2016.8 524959 11157   7.665 0.0056835 **
## + avgdeathsperyear         1    1554.8 525421 11159   5.903 0.0151994 *
## + pctnohs18_24             1     937.2 526039 11161   3.554 0.0595299 .
## + popest2015               1     769.9 526206 11162   2.919 0.0877037 .
## + pctempprivcoverage       1     640.4 526336 11162   2.427 0.1193892
## <none>                                 526976 11163
## + studypercap              1     526.4 526450 11163   1.995 0.1579992
## + pctasian                 1     510.7 526465 11163   1.935 0.1643197
## + medianage                1     319.9 526656 11163   1.212 0.2711191
## + pctpubliccoverage        1     141.0 526835 11164   0.534 0.4650938
## + avganncount              1     127.9 526848 11164   0.484 0.4865057
## + pctbachdeg18_24          1     109.8 526866 11164   0.416 0.5190459
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=11110.71
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + birthrate                1    9867.8 503107 11074  39.110 4.888e-10 ***
## + pctunemployed16_over     1    9179.9 503795 11077  36.334 1.976e-09 ***
## + pctprivatecoverage       1    6698.2 506277 11086  26.381 3.076e-07 ***
## + pctmarriedhouseholds     1    6526.7 506448 11087  25.697 4.363e-07 ***
## + pctwhite                 1    6241.8 506733 11088  24.561 7.806e-07 ***
## + medianagemale            1    5116.1 507859 11093  20.087 7.818e-06 ***
## + medianagefemale          1    4900.4 508075 11094  19.232 1.218e-05 ***
## + percentmarried           1    3895.5 509080 11098  15.258 9.689e-05 ***
## + pctblack                 1    3019.0 509956 11101  11.805 0.0006029 ***
## + pctpubliccoveragealone   1    3008.2 509967 11101  11.762 0.0006167 ***
## + pcths25_over             1    2550.2 510425 11103   9.963 0.0016213 **
## + medincome                1    2017.2 510958 11105   7.872 0.0050687 **
## + pctempprivcoverage       1    1976.2 510999 11105   7.711 0.0055389 **
## + avgdeathsperyear         1    1568.8 511406 11107   6.117 0.0134735 *
## + pctemployed16_over       1    1329.7 511645 11108   5.182 0.0229273 *
## + popest2015               1     760.0 512215 11110   2.959 0.0855756 .
## + pctnohs18_24             1     648.5 512327 11110   2.524 0.1122933
## + pctpubliccoverage        1     603.4 512372 11110   2.348 0.1255886
## + pctasian                 1     514.2 512461 11111   2.001 0.1573807
## <none>                                 512975 11111
## + studypercap              1     408.4 512567 11111   1.589 0.2076292
## + medianage                1     396.3 512579 11111   1.542 0.2145012
## + pctprivatecoveragealone  1     354.7 512620 11111   1.380 0.2403009
## + pctbachdeg18_24          1     274.3 512701 11112   1.067 0.3018096
## + pctsomecol18_24          1     108.5 512867 11112   0.422 0.5161625
## + avganncount              1      29.0 512946 11113   0.113 0.7371425
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=11073.84
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate
##
```

```
##                            Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + medianagemale             1    7688.2 495419 11045 30.9285 3.038e-08 ***
## + pctunemployed16_over      1    7415.6 495692 11046 29.8155 5.345e-08 ***
## + medianagefemale           1    7409.6 495698 11046 29.7911 5.412e-08 ***
## + pctwhite                  1    6980.4 496127 11048 28.0411 1.319e-07 ***
## + pctprivatecoverage        1    6913.5 496194 11048 27.7686 1.515e-07 ***
## + pctmarriedhouseholds      1    5544.9 497562 11054 22.2101 2.613e-06 ***
## + pctpubliccoveragealone    1    2745.2 500362 11065 10.9345 0.0009607 ***
## + pctblack                  1    2504.8 500603 11066  9.9720 0.0016132 **
## + percentmarried            1    2470.2 500637 11066  9.8338 0.0017384 **
## + medincome                 1    2123.0 500984 11067  8.4458 0.0036993 **
## + pcths25_over              1    2004.8 501102 11068  7.9736 0.0047937 **
## + pctempprivcoverage        1    1268.0 501839 11071  5.0356 0.0249407 *
## + avgdeathsperyear          1    1217.8 501889 11071  4.8359 0.0279871 *
## + pctpubliccoverage         1    1016.4 502091 11072  4.0344 0.0447169 *
## + popest2015                1     596.6 502511 11074  2.3661 0.1241530
## <none>                                  503107 11074
## + medianage                 1     445.0 502662 11074  1.7645 0.1842183
## + pctasian                  1     405.6 502702 11074  1.6079 0.2049366
## + pctnohs18_24              1     337.9 502769 11074  1.3393 0.2472949
## + pctprivatecoveragealone   1     317.5 502790 11075  1.2587 0.2620386
## + pctemployed16_over        1     281.8 502826 11075  1.1168 0.2907273
## + studypercap               1     245.1 502862 11075  0.9713 0.3244871
## + pctsomecol18_24           1      90.5 503017 11076  0.3586 0.5493303
## + pctbachdeg18_24           1      89.2 503018 11076  0.3534 0.5522584
## + avganncount               1      40.4 503067 11076  0.1601 0.6891161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=11045.03
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale
##
##                            Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + pctprivatecoverage        1    9334.9 486084 11009  38.255 7.513e-10 ***
## + pctunemployed16_over      1    7175.9 488243 11018  29.277 7.029e-08 ***
## + pctpubliccoveragealone    1    5567.4 489852 11024  22.640 2.094e-06 ***
## + pctprivatecoveragealone   1    5202.6 490216 11026  21.141 4.535e-06 ***
## + pctmarriedhouseholds      1    5081.5 490338 11026  20.643 5.864e-06 ***
## + pctwhite                  1    4185.1 491234 11030  16.971 3.951e-05 ***
## + pctemployed16_over        1    3543.1 491876 11033  14.349 0.0001564 ***
## + pcths25_over              1    3158.3 492261 11034  12.780 0.0003586 ***
## + pctpubliccoverage         1    1649.3 493770 11040   6.654 0.0099650 **
## + pctblack                  1    1411.4 494008 11041   5.691 0.0171446 *
## + pctsomecol18_24           1    1386.9 494032 11041   5.592 0.0181368 *
## + avgdeathsperyear          1    1073.3 494346 11043   4.325 0.0376869 *
## <none>                                  495419 11045
## + medincome                 1     416.1 495003 11045   1.674 0.1958142
## + percentmarried            1     372.0 495047 11046   1.497 0.2212937
## + popest2015                1     362.0 495057 11046   1.457 0.2276193
## + studypercap               1     306.3 495113 11046   1.232 0.2670776
## + pctbachdeg18_24           1     212.4 495207 11046   0.854 0.3554181
## + medianagefemale           1     173.0 495246 11046   0.696 0.4042353
## + medianage                 1     154.1 495265 11046   0.620 0.4312847
```

```
## + pctasian                 1      77.2 495342 11047   0.311 0.5773430
## + pctempprivcoverage       1      74.9 495344 11047   0.301 0.5831659
## + avganncount              1      54.7 495364 11047   0.220 0.6392275
## + pctnohs18_24             1      13.2 495406 11047   0.053 0.8179940
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=11008.96
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + pcths25_over             1    6037.4 480047 10986 25.0401 6.108e-07 ***
## + pctmarriedhouseholds     1    4617.0 481467 10992 19.0924 1.309e-05 ***
## + pctempprivcoverage       1    3861.6 482223 10995 15.9437 6.762e-05 ***
## + pctunemployed16_over     1    3859.2 482225 10995 15.9337 6.798e-05 ***
## + pctwhite                 1    2701.7 483382 11000 11.1281 0.0008659 ***
## + pctemployed16_over       1    2342.7 483742 11001  9.6421 0.0019284 **
## + pctblack                 1    1629.6 484455 11004  6.6972 0.0097269 **
## + pctnohs18_24             1    1040.1 485044 11007  4.2694 0.0389328 *
## + pctbachdeg18_24          1     756.8 485327 11008  3.1046 0.0782225 .
## + medincome                1     737.3 485347 11008  3.0247 0.0821587 .
## + avgdeathsperyear         1     527.6 485557 11009  2.1633 0.1415006
## <none>                                 486084 11009
## + pctsomecol18_24          1     405.0 485679 11009  1.6602 0.1977272
## + percentmarried           1     314.1 485770 11010  1.2873 0.2566872
## + pctpubliccoveragealone   1     264.2 485820 11010  1.0826 0.2982466
## + studypercap              1     236.5 485848 11010  0.9692 0.3250108
## + medianage                1     170.7 485914 11010  0.6994 0.4030693
## + avganncount              1     157.2 485927 11010  0.6440 0.4223571
## + popest2015               1      86.6 485998 11011  0.3549 0.5514308
## + pctasian                 1      48.5 486036 11011  0.1988 0.6557320
## + pctprivatecoveragealone  1      39.0 486045 11011  0.1598 0.6894251
## + medianagefemale          1      21.5 486063 11011  0.0882 0.7664968
## + pctpubliccoverage        1       0.1 486084 11011  0.0005 0.9817823
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10985.95
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + pctunemployed16_over     1    5561.9 474485 10965 23.3267 1.471e-06 ***
## + pctmarriedhouseholds     1    4246.9 475800 10970 17.7625 2.615e-05 ***
## + pctemployed16_over       1    3869.7 476177 10972 16.1720 6.000e-05 ***
## + pctwhite                 1    3310.9 476736 10974 13.8203 0.0002067 ***
## + pctempprivcoverage       1    2463.0 477584 10978 10.2630 0.0013787 **
## + pctblack                 1    2041.0 478006 10979  8.4970 0.0035972 **
## + medincome                1    1688.6 478358 10981  7.0246 0.0081035 **
## + pctnohs18_24             1    1268.6 478778 10983  5.2729 0.0217635 *
## + avgdeathsperyear         1     572.7 479474 10986  2.3769 0.1233031
## <none>                                 480047 10986
```

```
## + pctbachdeg18_24          1     375.7 479671 10986  1.5586 0.2120238
## + pctsomecol18_24          1     290.3 479756 10987  1.2043 0.2725942
## + percentmarried           1     261.8 479785 10987  1.0859 0.2975026
## + pctasian                 1     234.1 479813 10987  0.9711 0.3245199
## + studypercap              1     196.5 479850 10987  0.8148 0.3668129
## + medianage                1     180.8 479866 10987  0.7498 0.3866595
## + pctpubliccoveragealone   1     129.0 479918 10987  0.5351 0.4645549
## + popest2015               1     124.8 479922 10987  0.5175 0.4719776
## + avganncount              1      94.5 479952 10988  0.3918 0.5314407
## + medianagefemale          1      16.1 480031 10988  0.0666 0.7964347
## + pctpubliccoverage        1       2.3 480045 10988  0.0096 0.9220244
## + pctprivatecoveragealone  1       1.3 480045 10988  0.0056 0.9406014
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10964.63
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + pctmarriedhouseholds     1    3278.7 471206 10953 13.8397 0.0002046 ***
## + pctempprivcoverage       1    1831.6 472653 10959  7.7077 0.0055501 **
## + pctwhite                 1    1543.9 472941 10960  6.4930 0.0109047 *
## + pctemployed16_over       1    1475.0 473010 10960  6.2024 0.0128388 *
## + pctnohs18_24             1    1084.5 473400 10962  4.5564 0.0329177 *
## + pctblack                 1     916.3 473569 10963  3.8486 0.0499259 *
## + medincome                1     730.0 473755 10964  3.0648 0.0801606 .
## + pctbachdeg18_24          1     481.9 474003 10965  2.0222 0.1551756
## <none>                                 474485 10965
## + pctsomecol18_24          1     470.0 474015 10965  1.9723 0.1603650
## + medianage                1     320.9 474164 10965  1.3462 0.2460746
## + avganncount              1     319.9 474165 10965  1.3418 0.2468527
## + studypercap              1     156.3 474329 10966  0.6552 0.4183533
## + avgdeathsperyear         1     152.5 474332 10966  0.6393 0.4240563
## + pctpubliccoverage        1     124.5 474360 10966  0.5221 0.4700315
## + pctasian                 1      31.7 474453 10966  0.1331 0.7153102
## + pctprivatecoveragealone  1      19.5 474465 10967  0.0817 0.7750822
## + pctpubliccoveragealone   1      10.3 474475 10967  0.0431 0.8354681
## + percentmarried           1       6.0 474479 10967  0.0251 0.8741481
## + medianagefemale          1       3.0 474482 10967  0.0124 0.9112090
## + popest2015               1       1.1 474484 10967  0.0045 0.9464117
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10952.76
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + percentmarried           1    6845.8 464360 10926 29.3080 6.921e-08 ***
## + pctemployed16_over       1    2353.9 468852 10945  9.9809  0.001606 **
## + pctempprivcoverage       1    1925.7 469281 10947  8.1578  0.004332 **
```

```
## + medincome                1    1523.6 469683 10948  6.4488  0.011178 *
## + pctnohs18_24             1     685.3 470521 10952  2.8956  0.088976 .
## + pctsomecol18_24          1     620.6 470586 10952  2.6217  0.105568
## + avganncount              1     620.4 470586 10952  2.6208  0.105632
## <none>                                 471206 10953
## + pctpubliccoverage        1     431.0 470775 10953  1.8199  0.177479
## + pctwhite                 1     403.9 470802 10953  1.7056  0.191709
## + medianage                1     351.6 470855 10953  1.4847  0.223192
## + studypercap              1     251.3 470955 10954  1.0606  0.303195
## + pctprivatecoveragealone  1     174.5 471032 10954  0.7365  0.390887
## + pctbachdeg18_24          1     145.0 471061 10954  0.6121  0.434093
## + pctblack                 1     144.8 471061 10954  0.6111  0.434474
## + pctpubliccoveragealone   1      82.1 471124 10954  0.3463  0.556305
## + medianagefemale          1      71.2 471135 10954  0.3003  0.583742
## + avgdeathsperyear         1      22.2 471184 10955  0.0936  0.759624
## + popest2015               1      18.5 471188 10955  0.0780  0.780067
## + pctasian                 1       5.2 471201 10955  0.0218  0.882581
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10925.47
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
##     percentmarried
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + pctemployed16_over       1    8192.8 456168 10892  35.686 2.739e-09 ***
## + medincome                1    4032.7 460328 10910  17.407 3.147e-05 ***
## + pctempprivcoverage       1    3656.0 460704 10912  15.768 7.415e-05 ***
## + pctwhite                 1    1442.4 462918 10921   6.191   0.01292 *
## + pctpubliccoverage        1     753.3 463607 10924   3.229   0.07251 .
## + pctblack                 1     733.0 463627 10924   3.141   0.07648 .
## + pctnohs18_24             1     546.2 463814 10925   2.340   0.12625
## + avganncount              1     542.5 463818 10925   2.324   0.12754
## <none>                                 464360 10926
## + pctsomecol18_24          1     395.9 463964 10926   1.696   0.19302
## + medianage                1     348.0 464012 10926   1.490   0.22234
## + studypercap              1     344.1 464016 10926   1.473   0.22496
## + pctprivatecoveragealone  1     324.4 464036 10926   1.389   0.23868
## + pctpubliccoveragealone   1     130.3 464230 10927   0.558   0.45525
## + pctasian                 1     103.2 464257 10927   0.442   0.50647
## + avgdeathsperyear         1      92.7 464268 10927   0.397   0.52888
## + pctbachdeg18_24          1      69.8 464291 10927   0.299   0.58468
## + medianagefemale          1      45.8 464315 10927   0.196   0.65812
## + popest2015               1       0.0 464360 10928   0.000   0.98869
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10891.86
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
##     percentmarried + pctemployed16_over
```

```
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + pctempprivcoverage       1    9094.2 447073 10854  40.398 2.562e-10 ***
## + medincome                1    6301.9 449866 10866  27.820 1.476e-07 ***
## + pctpubliccoverage        1    3195.6 452972 10880  14.011 0.0001869 ***
## + pctprivatecoveragealone  1    2804.0 453364 10882  12.283 0.0004672 ***
## + pctwhite                 1    2112.6 454055 10885   9.240 0.0023984 **
## + medianagefemale          1    1288.4 454879 10888   5.625 0.0178002 *
## + pctblack                 1    1160.7 455007 10889   5.066 0.0245047 *
## + pctnohs18_24             1     774.4 455393 10890   3.377 0.0662553 .
## <none>                                 456168 10892
## + avganncount              1     376.4 455791 10892   1.640 0.2004599
## + medianage                1     312.4 455855 10892   1.361 0.2434954
## + pctbachdeg18_24          1     274.6 455893 10893   1.196 0.2742261
## + studypercap              1     271.3 455896 10893   1.182 0.2770727
## + pctpubliccoveragealone   1     158.6 456009 10893   0.691 0.4060692
## + pctsomecol18_24          1     122.8 456045 10893   0.535 0.4646931
## + avgdeathsperyear         1     102.8 456065 10893   0.448 0.5034735
## + pctasian                 1      89.9 456078 10894   0.392 0.5315068
## + popest2015               1       2.6 456165 10894   0.011 0.9149167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10853.56
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
##     percentmarried + pctemployed16_over + pctempprivcoverage
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + pctwhite                 1   3093.31 443980 10842 13.8299 0.0002057 ***
## + medincome                1   2879.28 444194 10843 12.8668 0.0003426 ***
## + pctpubliccoverage        1   1836.25 445237 10847  8.1865 0.0042644 **
## + pctblack                 1   1553.96 445519 10849  6.9236 0.0085722 **
## + avganncount              1    862.64 446211 10852  3.8375 0.0502579 .
## + pctnohs18_24             1    850.02 446223 10852  3.7813 0.0519705 .
## + medianagefemale          1    805.12 446268 10852  3.5812 0.0585837 .
## + pctpubliccoveragealone   1    683.12 446390 10852  3.0377 0.0815058 .
## <none>                                 447073 10854
## + studypercap              1    396.70 446677 10854  1.7629 0.1844169
## + medianage                1    280.87 446793 10854  1.2479 0.2640975
## + pctsomecol18_24          1    225.63 446848 10855  1.0023 0.3168761
## + popest2015               1    132.22 446941 10855  0.5872 0.4435795
## + pctbachdeg18_24          1    104.15 446969 10855  0.4625 0.4965151
## + pctprivatecoveragealone  1     58.90 447015 10855  0.2615 0.6091244
## + avgdeathsperyear         1     10.66 447063 10856  0.0473 0.8277920
## + pctasian                 1      1.05 447072 10856  0.0047 0.9455581
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10841.67
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
```

```
##     percentmarried + pctemployed16_over + pctempprivcoverage +
##     pctwhite
##
##                            Df Sum of Sq    RSS   AIC F Value    Pr(F)
## + pctnohs18_24              1   1722.36 442258 10836  7.7266 0.005492 **
## + medincome                 1   1390.30 442590 10837  6.2323 0.012625 *
## + pctpubliccoverage         1   1121.33 442859 10839  5.0235 0.025115 *
## + medianagefemale           1   1049.46 442931 10839  4.7008 0.030267 *
## + avganncount               1   1045.68 442934 10839  4.6838 0.030567 *
## <none>                                  443980 10842
## + pctpubliccoveragealone    1    417.49 443563 10842  1.8674 0.171930
## + medianage                 1    319.69 443660 10842  1.4296 0.231972
## + studypercap               1    278.27 443702 10842  1.2443 0.264782
## + popest2015                1    225.31 443755 10843  1.0073 0.315666
## + pctasian                  1    199.69 443780 10843  0.8927 0.344851
## + pctsomecol18_24           1    166.99 443813 10843  0.7465 0.387695
## + pctbachdeg18_24           1     87.42 443893 10843  0.3907 0.531992
## + pctblack                  1     52.61 443928 10843  0.2351 0.627806
## + pctprivatecoveragealone   1     49.44 443931 10843  0.2209 0.638373
## + avgdeathsperyear          1     34.07 443946 10844  0.1523 0.696416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10835.89
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
##     percentmarried + pctemployed16_over + pctempprivcoverage +
##     pctwhite + pctnohs18_24
##
##                            Df Sum of Sq    RSS   AIC F Value    Pr(F)
## + medincome                 1   1747.41 440510 10830  7.8661 0.005086 **
## + pctpubliccoverage         1   1516.78 440741 10831  6.8244 0.009060 **
## + avganncount               1   1124.08 441134 10833  5.0530 0.024693 *
## + pctsomecol18_24           1    925.16 441333 10834  4.1569 0.041597 *
## + medianagefemale           1    892.69 441365 10834  4.0108 0.045347 *
## + pctpubliccoveragealone    1    718.71 441539 10835  3.2278 0.072550 .
## <none>                                  442258 10836
## + pctasian                  1    356.95 441901 10836  1.6018 0.205798
## + medianage                 1    320.93 441937 10836  1.4400 0.230279
## + studypercap               1    299.29 441958 10836  1.3429 0.246671
## + popest2015                1    297.87 441960 10836  1.3365 0.247796
## + avgdeathsperyear          1     66.37 442191 10838  0.2976 0.585432
## + pctblack                  1     54.71 442203 10838  0.2453 0.620427
## + pctprivatecoveragealone   1     33.86 442224 10838  0.1518 0.696819
## + pctbachdeg18_24           1     12.33 442245 10838  0.0553 0.814113
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10829.97
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
##     percentmarried + pctemployed16_over + pctempprivcoverage +
```

```
##     pctwhite + pctnohs18_24 + medincome
##
##                            Df Sum of Sq    RSS   AIC F Value   Pr(F)
## + pctpubliccoverage         1   1393.10 439117 10826  6.2879 0.01224 *
## + avganncount               1   1368.03 439142 10826  6.1744 0.01304 *
## + pctsomecol18_24           1   1087.52 439423 10827  4.9052 0.02689 *
## + pctasian                  1    928.57 439582 10828  4.1867 0.04087 *
## + medianagefemale           1    867.48 439643 10828  3.9108 0.04812 *
## + pctpubliccoveragealone    1    804.94 439705 10828  3.6283 0.05695 .
## + popest2015                1    442.10 440068 10830  1.9912 0.15838
## <none>                                  440510 10830
## + medianage                 1    281.48 440229 10831  1.2673 0.26042
## + studypercap               1    196.90 440313 10831  0.8863 0.34659
## + avgdeathsperyear          1    127.91 440382 10831  0.5757 0.44810
## + pctbachdeg18_24           1     36.22 440474 10832  0.1630 0.68647
## + pctblack                  1      2.40 440508 10832  0.0108 0.91733
## + pctprivatecoveragealone   1      2.26 440508 10832  0.0102 0.91960
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10825.63
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
##     percentmarried + pctemployed16_over + pctempprivcoverage +
##     pctwhite + pctnohs18_24 + medincome + pctpubliccoverage
##
##                            Df Sum of Sq    RSS   AIC F Value   Pr(F)
## + pctsomecol18_24           1   1249.13 437868 10822  5.6513 0.01754 *
## + avganncount               1   1175.82 437941 10822  5.3188 0.02120 *
## + pctasian                  1    783.77 438333 10824  3.5422 0.05997 .
## + medianagefemale           1    464.07 438653 10826  2.0958 0.14786
## <none>                                  439117 10826
## + popest2015                1    401.75 438716 10826  1.8141 0.17817
## + medianage                 1    305.54 438812 10826  1.3794 0.24035
## + studypercap               1    140.03 438977 10827  0.6319 0.42675
## + pctprivatecoveragealone   1    110.42 439007 10827  0.4983 0.48035
## + avgdeathsperyear          1     91.77 439025 10827  0.4141 0.51997
## + pctbachdeg18_24           1     47.94 439069 10827  0.2163 0.64193
## + pctpubliccoveragealone    1     46.10 439071 10827  0.2080 0.64838
## + pctblack                  1     29.93 439087 10828  0.1351 0.71329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10821.93
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
##     percentmarried + pctemployed16_over + pctempprivcoverage +
##     pctwhite + pctnohs18_24 + medincome + pctpubliccoverage +
##     pctsomecol18_24
##
##                            Df Sum of Sq    RSS   AIC F Value   Pr(F)
## + avganncount               1    786.82 437081 10820  3.5643 0.05918 .
```

```
## + pctasian                 1    575.97 437292 10821  2.6079 0.10649
## + medianagefemale          1    539.14 437329 10822  2.4409 0.11837
## <none>                                 437868 10822
## + medianage                1    314.18 437554 10822  1.4217 0.23326
## + popest2015               1    178.75 437689 10823  0.8086 0.36864
## + studypercap              1    133.92 437734 10823  0.6058 0.43649
## + pctbachdeg18_24          1    100.52 437768 10824  0.4547 0.50021
## + pctpubliccoveragealone   1     56.55 437812 10824  0.2558 0.61310
## + pctprivatecoveragealone  1     54.73 437813 10824  0.2475 0.61888
## + pctblack                 1     45.10 437823 10824  0.2040 0.65159
## + avgdeathsperyear         1      6.69 437861 10824  0.0302 0.86198
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10820.33
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
##     percentmarried + pctemployed16_over + pctempprivcoverage +
##     pctwhite + pctnohs18_24 + medincome + pctpubliccoverage +
##     pctsomecol18_24 + avganncount
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## + avgdeathsperyear         1    3320.4 433761 10807 15.1489 0.0001026 ***
## + popest2015               1     693.2 436388 10819  3.1435 0.0763820 .
## + medianagefemale          1     465.6 436616 10820  2.1106 0.1464405
## <none>                                 437081 10820
## + medianage                1     333.9 436747 10821  1.5132 0.2188003
## + pctasian                 1     302.5 436779 10821  1.3704 0.2418872
## + studypercap              1     103.8 436977 10822  0.4703 0.4929390
## + pctpubliccoveragealone   1      77.0 437004 10822  0.3485 0.5550108
## + pctblack                 1      57.8 437023 10822  0.2619 0.6088472
## + pctbachdeg18_24          1      49.1 437032 10822  0.2225 0.6372278
## + pctprivatecoveragealone  1      15.9 437065 10822  0.0720 0.7884831
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10807.07
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
##     percentmarried + pctemployed16_over + pctempprivcoverage +
##     pctwhite + pctnohs18_24 + medincome + pctpubliccoverage +
##     pctsomecol18_24 + avganncount + avgdeathsperyear
##
##                           Df Sum of Sq    RSS   AIC F Value   Pr(F)
## + popest2015               1   1444.71 432316 10802  6.6101 0.01021 *
## + pctasian                 1    458.44 433303 10807  2.0928 0.14816
## <none>                                 433761 10807
## + medianagefemale          1    424.23 433337 10807  1.9364 0.16422
## + medianage                1    350.47 433410 10808  1.5995 0.20613
## + pctblack                 1    104.04 433657 10809  0.4745 0.49098
## + pctbachdeg18_24          1     97.61 433663 10809  0.4452 0.50469
## + studypercap              1     77.65 433683 10809  0.3542 0.55183
```

```
## + pctpubliccoveragealone   1     51.39 433710 10809  0.2344 0.62836
## + pctprivatecoveragealone  1     22.37 433739 10809  0.1020 0.74944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10802.4
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
##     percentmarried + pctemployed16_over + pctempprivcoverage +
##     pctwhite + pctnohs18_24 + medincome + pctpubliccoverage +
##     pctsomecol18_24 + avganncount + avgdeathsperyear + popest2015
##
##                            Df Sum of Sq    RSS   AIC F Value  Pr(F)
## + medianagefemale           1    530.09 431786 10802 2.42709 0.1194
## <none>                                  432316 10802
## + medianage                 1    317.58 431999 10803 1.45336 0.2281
## + pctasian                  1    207.12 432109 10803 0.94762 0.3304
## + pctblack                  1    134.33 432182 10804 0.61450 0.4332
## + pctbachdeg18_24           1    107.15 432209 10804 0.49011 0.4840
## + studypercap               1     80.40 432236 10804 0.36773 0.5443
## + pctpubliccoveragealone    1     68.18 432248 10804 0.31183 0.5766
## + pctprivatecoveragealone   1     11.67 432305 10804 0.05336 0.8173
##
## Step:  AIC=10801.94
## target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
##     pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
##     pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
##     percentmarried + pctemployed16_over + pctempprivcoverage +
##     pctwhite + pctnohs18_24 + medincome + pctpubliccoverage +
##     pctsomecol18_24 + avganncount + avgdeathsperyear + popest2015 +
##     medianagefemale
##
##                            Df Sum of Sq    RSS   AIC F Value  Pr(F)
## <none>                                  431786 10802
## + medianage                 1   298.696 431487 10803 1.36788 0.2423
## + pctasian                  1   222.454 431564 10803 1.01855 0.3130
## + pctbachdeg18_24           1    96.173 431690 10804 0.44022 0.5071
## + studypercap               1    64.146 431722 10804 0.29360 0.5880
## + pctblack                  1    61.861 431724 10804 0.28314 0.5947
## + pctpubliccoveragealone    1    22.541 431764 10804 0.10316 0.7481
## + pctprivatecoveragealone   1    20.823 431765 10804 0.09530 0.7576
```

We will perform the step backward to see if it gives us the same final selection of explanatory variables

```
stepbackward=stepAIC(model1b, scope = list(upper=model1b,lower=~1),test='F')
```

```
## Start:  AIC=10811.45
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + studypercap + medianage +
##     medianagemale + medianagefemale + percentmarried + pctnohs18_24 +
##     pcths18_24 + pctsomecol18_24 + pctbachdeg18_24 + pcths25_over +
##     pctbachdeg25_over + pctemployed16_over + pctunemployed16_over +
```

```
##     pctprivatecoverage + pctprivatecoveragealone + pctempprivcoverage +
##     pctpubliccoverage + pctpubliccoveragealone + pctwhite + pctblack +
##     pctasian + pctotherrace + pctmarriedhouseholds + birthrate
##
##                            Df Sum of Sq    RSS   AIC F Value     Pr(F)
## - pctprivatecoveragealone   1        28 430847 10810    0.13 0.7185542
## - pctpubliccoveragealone    1        62 430880 10810    0.28 0.5956525
## - studypercap               1        70 430889 10810    0.32 0.5716005
## - pctbachdeg18_24           1       100 430919 10810    0.46 0.4991689
## - pctblack                  1       220 431039 10810    1.01 0.3159716
## - pctasian                  1       329 431148 10811    1.51 0.2200471
## - medianagefemale           1       329 431148 10811    1.51 0.2198054
## - medianage                 1       339 431158 10811    1.55 0.2130370
## <none>                                  430819 10812
## - pctunemployed16_over      1       433 431251 10812    1.98 0.1596407
## - pctpubliccoverage         1       500 431319 10812    2.29 0.1307139
## - pctsomecol18_24           1      1101 431920 10815    5.04 0.0249488 *
## - pcths18_24                1      1127 431946 10815    5.15 0.0232997 *
## - popest2015                1      1219 432038 10815    5.58 0.0183058 *
## - povertypercent            1      1265 432084 10815    5.79 0.0162499 *
## - pctwhite                  1      1558 432377 10817    7.13 0.0076591 **
## - medianagemale             1      2086 432905 10819    9.54 0.0020376 **
## - medincome                 1      2157 432976 10819    9.87 0.0017095 **
## - pctnohs18_24              1      3180 433999 10824   14.54 0.0001412 ***
## - pctprivatecoverage        1      3358 434177 10825   15.36 9.209e-05 ***
## - avganncount               1      3538 434356 10826   16.18 5.989e-05 ***
## - pctempprivcoverage        1      3860 434678 10827   17.65 2.775e-05 ***
## - avgdeathsperyear          1      3904 434723 10828   17.85 2.497e-05 ***
## - birthrate                 1      7928 438747 10846   36.25 2.063e-09 ***
## - pcths25_over              1      8012 438831 10846   36.64 1.700e-09 ***
## - pctbachdeg25_over         1      9992 440810 10855   45.69 1.820e-11 ***
## - pctotherrace              1     11881 442699 10864   54.33 2.485e-13 ***
## - pctemployed16_over        1     14322 445140 10875   65.49 1.012e-15 ***
## - pctmarriedhouseholds      1     24709 455527 10921  112.99 < 2.2e-16 ***
## - percentmarried            1     24770 455589 10921  113.27 < 2.2e-16 ***
## - incidencerate             1    175423 606242 11493  802.15 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10809.58
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + studypercap + medianage +
##     medianagemale + medianagefemale + percentmarried + pctnohs18_24 +
##     pcths18_24 + pctsomecol18_24 + pctbachdeg18_24 + pcths25_over +
##     pctbachdeg25_over + pctemployed16_over + pctunemployed16_over +
##     pctprivatecoverage + pctempprivcoverage + pctpubliccoverage +
##     pctpubliccoveragealone + pctwhite + pctblack + pctasian +
##     pctotherrace + pctmarriedhouseholds + birthrate
##
##                            Df Sum of Sq    RSS   AIC F Value     Pr(F)
## - pctpubliccoveragealone    1        44 430891 10808    0.20 0.6552894
## - studypercap               1        68 430916 10808    0.31 0.5757543
## - pctbachdeg18_24           1       102 430949 10808    0.46 0.4956850
## - pctblack                  1       222 431069 10809    1.02 0.3135525
```

```
## - medianagefemale          1       329 431176 10809    1.50 0.2203028
## - medianage                1       335 431182 10809    1.53 0.2158985
## - pctasian                 1       345 431192 10809    1.58 0.2091033
## <none>                                 430847 10810
## - pctunemployed16_over     1       443 431290 10810    2.03 0.1546063
## - pctpubliccoverage        1       485 431332 10810    2.22 0.1367006
## + pctprivatecoveragealone  1        28 430819 10812    0.13 0.7185542
## - pctsomecol18_24          1      1126 431973 10813    5.15 0.0233247 *
## - pcths18_24               1      1156 432003 10813    5.29 0.0215762 *
## - popest2015               1      1224 432072 10813    5.60 0.0180411 *
## - povertypercent           1      1257 432104 10813    5.75 0.0165694 *
## - pctwhite                 1      1594 432441 10815    7.29 0.0069787 **
## - medianagemale            1      2097 432944 10817    9.59 0.0019814 **
## - medincome                1      2148 432995 10818    9.82 0.0017470 **
## - pctnohs18_24             1      3173 434020 10822   14.51 0.0001434 ***
## - avganncount              1      3567 434414 10824   16.32 5.557e-05 ***
## - avgdeathsperyear         1      3913 434760 10826   17.90 2.436e-05 ***
## - pctempprivcoverage       1      4010 434857 10826   18.35 1.932e-05 ***
## - pctprivatecoverage       1      5890 436737 10835   26.95 2.308e-07 ***
## - birthrate                1      7925 438772 10844   36.25 2.060e-09 ***
## - pcths25_over             1      8009 438856 10844   36.64 1.696e-09 ***
## - pctbachdeg25_over        1     10220 441067 10854   46.75 1.070e-11 ***
## - pctotherrace             1     11854 442701 10862   54.23 2.611e-13 ***
## - pctemployed16_over       1     14475 445322 10874   66.22 7.062e-16 ***
## - percentmarried           1     24755 455602 10919  113.25 < 2.2e-16 ***
## - pctmarriedhouseholds     1     24808 455655 10920  113.49 < 2.2e-16 ***
## - incidencerate            1    175454 606301 11491  802.65 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10807.79
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + studypercap + medianage +
##     medianagemale + medianagefemale + percentmarried + pctnohs18_24 +
##     pcths18_24 + pctsomecol18_24 + pctbachdeg18_24 + pcths25_over +
##     pctbachdeg25_over + pctemployed16_over + pctunemployed16_over +
##     pctprivatecoverage + pctempprivcoverage + pctpubliccoverage +
##     pctwhite + pctblack + pctasian + pctotherrace + pctmarriedhouseholds +
##     birthrate
##
##                            Df Sum of Sq    RSS   AIC F Value     Pr(F)
## - studypercap               1        66 430957 10806    0.30 0.5815680
## - pctbachdeg18_24           1        89 430980 10806    0.41 0.5224432
## - pctblack                  1       209 431099 10807    0.96 0.3285443
## - medianage                 1       338 431228 10807    1.55 0.2139989
## - pctasian                  1       342 431233 10807    1.57 0.2110217
## - medianagefemale           1       381 431272 10808    1.75 0.1866526
## <none>                                 430891 10808
## - pctunemployed16_over      1       462 431353 10808    2.12 0.1459518
## + pctpubliccoveragealone    1        44 430847 10810    0.20 0.6552894
## + pctprivatecoveragealone   1        10 430880 10810    0.05 0.8274746
## - pctpubliccoverage         1      1002 431892 10810    4.58 0.0323791 *
## - pctsomecol18_24           1      1121 432011 10811    5.13 0.0236377 *
## - pcths18_24                1      1198 432088 10811    5.48 0.0193250 *
```

```
## - popest2015               1      1216 432106 10811    5.56 0.0184265 *
## - povertypercent           1      1308 432198 10812    5.99 0.0145126 *
## - pctwhite                 1      1587 432478 10813    7.26 0.0070961 **
## - medincome                1      2244 433135 10816   10.27 0.0013725 **
## - medianagemale            1      2287 433178 10816   10.47 0.0012349 **
## - pctnohs18_24             1      3213 434104 10821   14.70 0.0001297 ***
## - avganncount              1      3567 434458 10822   16.32 5.541e-05 ***
## - avgdeathsperyear         1      3907 434797 10824   17.88 2.461e-05 ***
## - pctempprivcoverage       1      5201 436091 10830   23.80 1.154e-06 ***
## - birthrate                1      7924 438815 10842   36.27 2.049e-09 ***
## - pcths25_over             1      8104 438995 10843   37.09 1.352e-09 ***
## - pctbachdeg25_over        1     10180 441071 10852   46.59 1.161e-11 ***
## - pctotherrace             1     11822 442713 10860   54.10 2.774e-13 ***
## - pctprivatecoverage       1     11978 442869 10861   54.82 1.947e-13 ***
## - pctemployed16_over       1     14834 445724 10874   67.89 3.116e-16 ***
## - percentmarried           1     24713 455603 10917  113.10 < 2.2e-16 ***
## - pctmarriedhouseholds     1     24890 455781 10918  113.91 < 2.2e-16 ***
## - incidencerate            1    177095 607986 11495  810.49 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10806.1
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianage + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pctbachdeg18_24 + pcths25_over + pctbachdeg25_over +
##     pctemployed16_over + pctunemployed16_over + pctprivatecoverage +
##     pctempprivcoverage + pctpubliccoverage + pctwhite + pctblack +
##     pctasian + pctotherrace + pctmarriedhouseholds + birthrate
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## - pctbachdeg18_24          1        82 431039 10804    0.38 0.5394960
## - pctblack                 1       211 431168 10805    0.96 0.3261389
## - medianage                1       330 431288 10806    1.51 0.2188246
## - pctasian                 1       356 431313 10806    1.63 0.2016811
## - medianagefemale          1       396 431353 10806    1.81 0.1784728
## <none>                                 430957 10806
## - pctunemployed16_over     1       461 431418 10806    2.11 0.1463036
## + studypercap              1        66 430891 10808    0.30 0.5815680
## + pctpubliccoveragealone   1        41 430916 10808    0.19 0.6630862
## + pctprivatecoveragealone  1        10 430947 10808    0.04 0.8322684
## - pctpubliccoverage        1      1023 431980 10809    4.68 0.0305712 *
## - pctsomecol18_24          1      1117 432074 10809    5.11 0.0238328 *
## - pcths18_24               1      1195 432152 10810    5.47 0.0194432 *
## - popest2015               1      1210 432167 10810    5.54 0.0186790 *
## - povertypercent           1      1321 432278 10810    6.05 0.0140257 *
## - pctwhite                 1      1605 432562 10812    7.35 0.0067650 **
## - medianagemale            1      2256 433213 10814   10.33 0.0013312 **
## - medincome                1      2317 433275 10815   10.61 0.0011439 **
## - pctnohs18_24             1      3206 434163 10819   14.68 0.0001316 ***
## - avganncount              1      3610 434567 10821   16.53 4.979e-05 ***
## - avgdeathsperyear         1      3912 434869 10822   17.91 2.423e-05 ***
## - pctempprivcoverage       1      5149 436106 10828   23.57 1.297e-06 ***
## - birthrate                1      8006 438963 10841   36.65 1.684e-09 ***
```

```
## - pcths25_over             1      8181 439138 10842   37.45 1.127e-09 ***
## - pctbachdeg25_over        1     10282 441239 10851   47.07 9.115e-12 ***
## - pctotherrace             1     11805 442762 10858   54.04 2.857e-13 ***
## - pctprivatecoverage       1     11970 442927 10859   54.80 1.966e-13 ***
## - pctemployed16_over       1     14944 445901 10872   68.42 2.404e-16 ***
## - percentmarried           1     24727 455684 10916  113.20 < 2.2e-16 ***
## - pctmarriedhouseholds     1     24863 455820 10916  113.83 < 2.2e-16 ***
## - incidencerate            1    177221 608178 11493  811.35 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10804.48
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianage + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pcths25_over + pctbachdeg25_over + pctemployed16_over +
##     pctunemployed16_over + pctprivatecoverage + pctempprivcoverage +
##     pctpubliccoverage + pctwhite + pctblack + pctasian + pctotherrace +
##     pctmarriedhouseholds + birthrate
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## - pctblack                 1       218 431257 10804    1.00 0.3178682
## - medianage                1       321 431361 10804    1.47 0.2252119
## - pctasian                 1       380 431420 10804    1.74 0.1871265
## - medianagefemale          1       403 431443 10804    1.85 0.1742122
## <none>                                 431039 10804
## - pctunemployed16_over     1       484 431523 10805    2.22 0.1366453
## + pctbachdeg18_24          1        82 430957 10806    0.38 0.5394960
## + studypercap              1        59 430980 10806    0.27 0.6026214
## + pctpubliccoveragealone   1        30 431009 10806    0.14 0.7098714
## + pctprivatecoveragealone  1        12 431027 10806    0.06 0.8115572
## - pctpubliccoverage        1      1000 432039 10807    4.58 0.0324888 *
## - pctsomecol18_24          1      1064 432103 10807    4.87 0.0274106 *
## - popest2015               1      1197 432236 10808    5.48 0.0193280 *
## - povertypercent           1      1263 432303 10808    5.79 0.0162510 *
## - pcths18_24               1      1470 432509 10809    6.73 0.0095461 **
## - pctwhite                 1      1644 432683 10810    7.53 0.0061274 **
## - medincome                1      2236 433275 10813   10.24 0.0013966 **
## - medianagemale            1      2291 433330 10813   10.49 0.0012195 **
## - pctnohs18_24             1      3126 434165 10817   14.32 0.0001592 ***
## - avganncount              1      3608 434647 10819   16.52 4.995e-05 ***
## - avgdeathsperyear         1      3870 434910 10820   17.72 2.668e-05 ***
## - pctempprivcoverage       1      5185 436225 10826   23.75 1.186e-06 ***
## - birthrate                1      7951 438991 10839   36.41 1.900e-09 ***
## - pcths25_over             1      8107 439146 10840   37.13 1.328e-09 ***
## - pctbachdeg25_over        1     11120 442159 10853   50.93 1.342e-12 ***
## - pctotherrace             1     11898 442938 10857   54.49 2.291e-13 ***
## - pctprivatecoverage       1     12017 443056 10858   55.03 1.750e-13 ***
## - pctemployed16_over       1     15120 446159 10872   69.24 < 2.2e-16 ***
## - percentmarried           1     24645 455684 10914  112.86 < 2.2e-16 ***
## - pctmarriedhouseholds     1     24978 456017 10915  114.39 < 2.2e-16 ***
## - incidencerate            1    177438 608477 11492  812.60 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Step:  AIC=10803.49
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianage + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pcths25_over + pctbachdeg25_over + pctemployed16_over +
##     pctunemployed16_over + pctprivatecoverage + pctempprivcoverage +
##     pctpubliccoverage + pctwhite + pctasian + pctotherrace +
##     pctmarriedhouseholds + birthrate
##
##                           Df Sum of Sq    RSS   AIC F Value     Pr(F)
## - pctasian                 1       230 431487 10803    1.05 0.3046643
## - medianage                1       306 431564 10803    1.40 0.2363022
## <none>                                 431257 10804
## - pctunemployed16_over     1       472 431730 10804    2.16 0.1414701
## - medianagefemale          1       526 431784 10804    2.41 0.1207021
## + pctblack                 1       218 431039 10804    1.00 0.3178682
## + pctbachdeg18_24          1        90 431168 10805    0.41 0.5221186
## + studypercap              1        61 431196 10805    0.28 0.5976628
## + pctpubliccoveragealone   1        19 431239 10805    0.09 0.7695767
## + pctprivatecoveragealone  1        16 431241 10805    0.07 0.7844728
## - pctpubliccoverage        1       905 432163 10806    4.15 0.0418736 *
## - pctsomecol18_24          1      1065 432322 10806    4.88 0.0273208 *
## - povertypercent           1      1151 432408 10807    5.27 0.0217929 *
## - popest2015               1      1248 432505 10807    5.72 0.0169061 *
## - pcths18_24               1      1501 432758 10808    6.87 0.0088214 **
## - medianagemale            1      2233 433491 10812   10.23 0.0014047 **
## - medincome                1      2323 433580 10812   10.64 0.0011262 **
## - pctwhite                 1      2362 433619 10812   10.82 0.0010240 **
## - pctnohs18_24             1      3041 434298 10816   13.93 0.0001955 ***
## - avganncount              1      3525 434782 10818   16.14 6.090e-05 ***
## - avgdeathsperyear         1      3861 435119 10819   17.68 2.724e-05 ***
## - pctempprivcoverage       1      5037 436294 10825   23.07 1.681e-06 ***
## - birthrate                1      7759 439016 10837   35.53 2.962e-09 ***
## - pcths25_over             1      8331 439588 10840   38.15 7.921e-10 ***
## - pctbachdeg25_over        1     10905 442162 10852   49.94 2.187e-12 ***
## - pctotherrace             1     11983 443240 10856   54.88 1.890e-13 ***
## - pctprivatecoverage       1     12486 443744 10859   57.18 6.044e-14 ***
## - pctemployed16_over       1     15107 446365 10870   69.19 < 2.2e-16 ***
## - percentmarried           1     24732 455989 10913  113.26 < 2.2e-16 ***
## - pctmarriedhouseholds     1     24943 456200 10914  114.23 < 2.2e-16 ***
## - incidencerate            1    177531 608788 11491  813.03 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10802.56
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianage + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pcths25_over + pctbachdeg25_over + pctemployed16_over +
##     pctunemployed16_over + pctprivatecoverage + pctempprivcoverage +
##     pctpubliccoverage + pctwhite + pctotherrace + pctmarriedhouseholds +
##     birthrate
##
```

```
##                            Df Sum of Sq    RSS   AIC F Value     Pr(F)
## - medianage                 1       299 431786 10802    1.37  0.242318
## <none>                                  431487 10803
## - pctunemployed16_over      1       498 431986 10803    2.28  0.130996
## - medianagefemale           1       511 431999 10803    2.34  0.126163
## + pctasian                  1       230 431257 10804    1.05  0.304664
## + pctbachdeg18_24           1       107 431381 10804    0.49  0.484487
## + studypercap               1        71 431416 10804    0.33  0.568616
## + pctblack                  1        68 431420 10804    0.31  0.577134
## + pctprivatecoveragealone   1        26 431462 10804    0.12  0.732475
## + pctpubliccoveragealone    1        20 431467 10804    0.09  0.762431
## - pctpubliccoverage         1       984 432471 10805    4.51  0.033919 *
## - povertypercent            1      1117 432604 10806    5.11  0.023847 *
## - pctsomecol18_24           1      1133 432621 10806    5.19  0.022831 *
## - pcths18_24                1      1498 432986 10808    6.86  0.008871 **
## - popest2015                1      1515 433003 10808    6.94  0.008495 **
## - medincome                 1      2109 433596 10810    9.66  0.001914 **
## - pctwhite                  1      2182 433669 10811    9.99  0.001597 **
## - medianagemale             1      2224 433711 10811   10.18  0.001438 **
## - pctnohs18_24              1      2917 434404 10814   13.36  0.000264 ***
## - avganncount               1      3579 435067 10817   16.39 5.350e-05 ***
## - avgdeathsperyear          1      4152 435640 10820   19.01 1.364e-05 ***
## - pctempprivcoverage        1      4974 436461 10824   22.78 1.952e-06 ***
## - birthrate                 1      7704 439191 10836   35.28 3.366e-09 ***
## - pcths25_over              1      8418 439906 10839   38.55 6.480e-10 ***
## - pctbachdeg25_over         1     11094 442581 10851   50.80 1.425e-12 ***
## - pctotherrace              1     12109 443596 10856   55.45 1.423e-13 ***
## - pctprivatecoverage        1     12478 443966 10858   57.14 6.160e-14 ***
## - pctemployed16_over        1     14906 446393 10868   68.26 2.592e-16 ***
## - percentmarried            1     24564 456051 10911  112.49 < 2.2e-16 ***
## - pctmarriedhouseholds      1     24761 456248 10912  113.39 < 2.2e-16 ***
## - incidencerate             1    177639 609127 11490  813.50 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=10801.94
## target_deathrate ~ avganncount + avgdeathsperyear + incidencerate +
##     medincome + popest2015 + povertypercent + medianagemale +
##     medianagefemale + percentmarried + pctnohs18_24 + pcths18_24 +
##     pctsomecol18_24 + pcths25_over + pctbachdeg25_over + pctemployed16_over +
##     pctunemployed16_over + pctprivatecoverage + pctempprivcoverage +
##     pctpubliccoverage + pctwhite + pctotherrace + pctmarriedhouseholds +
##     birthrate
##
##                            Df Sum of Sq    RSS   AIC F Value     Pr(F)
## <none>                                  431786 10802
## - pctunemployed16_over      1       455 432241 10802    2.08 0.1489916
## - medianagefemale           1       530 432316 10802    2.43 0.1194141
## + medianage                 1       299 431487 10803    1.37 0.2423176
## + pctasian                  1       222 431564 10803    1.02 0.3129868
## + pctbachdeg18_24           1        96 431690 10804    0.44 0.5070951
## + studypercap               1        64 431722 10804    0.29 0.5879860
## + pctblack                  1        62 431724 10804    0.28 0.5947107
## + pctpubliccoveragealone    1        23 431764 10804    0.10 0.7481042
```

```
## + pctprivatecoveragealone  1        21 431765 10804    0.10 0.7575770
## - pctpubliccoverage        1       963 432749 10804    4.41 0.0358748 *
## - pctsomecol18_24          1      1130 432916 10805    5.18 0.0230179 *
## - povertypercent           1      1154 432940 10805    5.28 0.0216503 *
## - pcths18_24               1      1507 433293 10807    6.90 0.0086914 **
## - popest2015               1      1551 433337 10807    7.10 0.0077730 **
## - pctwhite                 1      2147 433933 10810    9.83 0.0017422 **
## - medincome                1      2154 433940 10810    9.86 0.0017116 **
## - medianagemale            1      2271 434057 10810   10.40 0.0012816 **
## - pctnohs18_24             1      2912 434698 10813   13.33 0.0002675 ***
## - avganncount              1      3544 435330 10816   16.22 5.838e-05 ***
## - avgdeathsperyear         1      4190 435976 10819   19.18 1.250e-05 ***
## - pctempprivcoverage       1      4970 436756 10823   22.75 1.976e-06 ***
## - birthrate                1      7739 439525 10836   35.43 3.113e-09 ***
## - pcths25_over             1      8395 440181 10838   38.44 6.860e-10 ***
## - pctbachdeg25_over        1     11134 442920 10851   50.98 1.308e-12 ***
## - pctotherrace             1     12166 443952 10856   55.70 1.255e-13 ***
## - pctprivatecoverage       1     12478 444264 10857   57.13 6.199e-14 ***
## - pctemployed16_over       1     14964 446750 10868   68.51 2.289e-16 ***
## - percentmarried           1     24606 456392 10911  112.66 < 2.2e-16 ***
## - pctmarriedhouseholds     1     24788 456575 10912  113.50 < 2.2e-16 ***
## - incidencerate            1    177468 609254 11489  812.57 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see that the list of selected variables for both forward and backward are identical.

```
model2c = lm(target_deathrate ~ pctbachdeg25_over + incidencerate + povertypercent +
    pctotherrace + pcths18_24 + birthrate + medianagemale + pctprivatecoverage +
    pcths25_over + pctunemployed16_over + pctmarriedhouseholds +
    percentmarried + pctemployed16_over + pctempprivcoverage +
    pctwhite + pctnohs18_24 + medincome + pctpubliccoverage +
    pctsomecol18_24 + avganncount + avgdeathsperyear + popest2015 +
    medianagefemale,data = learningset5)
p2c= predict(model2c, newdata = testset5)
```

```
cat('Step forward or backward:',sqrt(mean((p2c-testset5$target_deathrate)**2)),'\n')
```

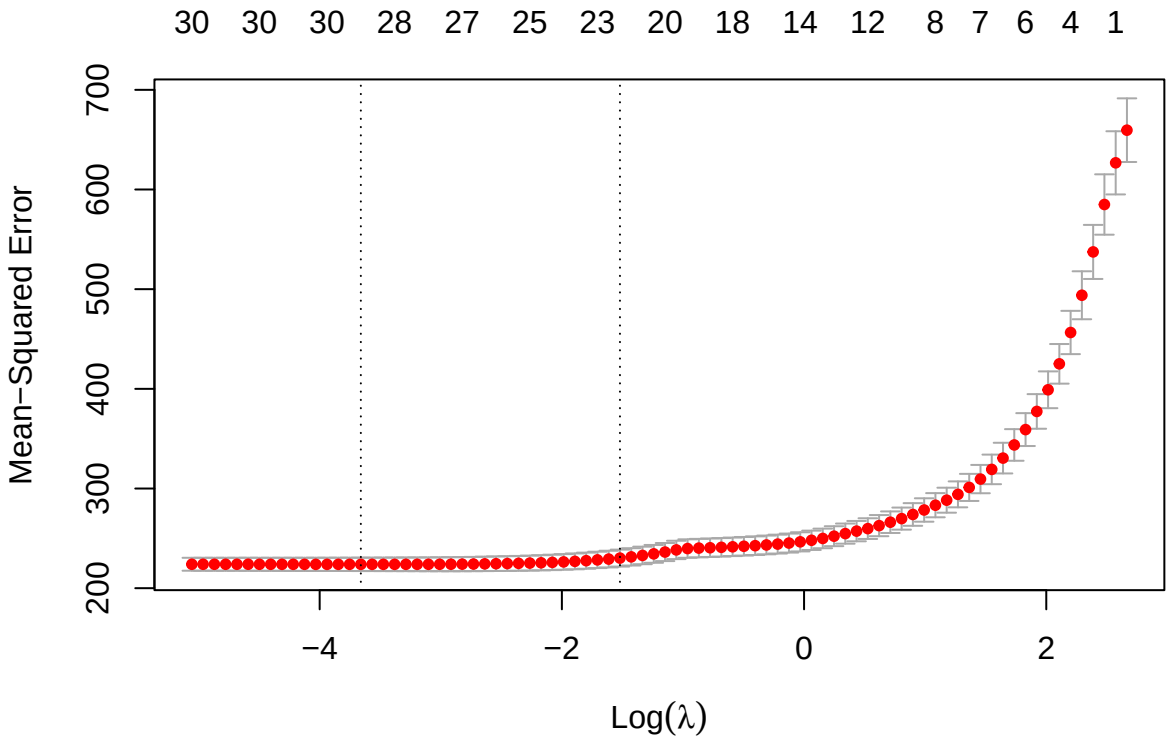
```
## Step forward or backward: 15.16503
```

Lasso

```
# lasso
x=as.matrix(learningset5[-3])
y=as.matrix(learningset5[3])
```

We determine the best value of lambda by cross validation

```
tmplasso=cv.glmnet(x,y)
plot(tmplasso)
```




```
model3=glmnet(x,y,alpha=1,lambda=tmplasso$lambda.1se)
```

```
p3_a=predict(model3, newx = as.matrix(testset5[,-1]), s = tmplasso$lambda.1se)
```

```
cat('LASSO:',sqrt(mean((p3_a-testset5$target_deathrate)^2)),'\n')
```

```
## LASSO: 15.24858
```

LASSO should be used only as a variable selection feature : Once the feature are selected, you have to estimate an OLS model on the selected feature

```
#Selected variables
print(model3$beta)
```

```
## 30 x 1 sparse Matrix of class "dgCMatrix"
##                                      s0
## avganncount             -0.0008659773
## avgdeathsperyear         0.0017240367
## incidencerate            0.1994838302
## medincome                .
## popest2015               .
## povertypercent           0.1368302351
## studypercap             -0.0001715904
## medianage               -0.0054427684
```

```
## medianagemale           -0.4759102794
## medianagefemale         -0.2455017122
## percentmarried           0.6631581246
## pctnohs18_24            -0.1117644413
## pcths18_24               0.1890756511
## pctsomecol18_24         -0.1218472701
## pctbachdeg18_24          .
## pcths25_over             0.5149671289
## pctbachdeg25_over       -0.9149832680
## pctemployed16_over      -0.4872705888
## pctunemployed16_over     0.3733905411
## pctprivatecoverage      -0.5128618173
## pctprivatecoveragealone  .
## pctempprivcoverage       0.2466108959
## pctpubliccoverage        .
## pctpubliccoveragealone   .
## pctwhite                -0.0766508399
## pctblack                 .
## pctasian                 .
## pctotherrace            -0.7877974971
## pctmarriedhouseholds    -0.7411565771
## birthrate               -0.9748226091
```

```
model3 = lm(target_deathrate~ avganncount + avgdeathsperyear + incidencerate +
              povertypercent + studypercap + medianagemale + medianagefemale +
              percentmarried + pctsomecol18_24 + pctbachdeg18_24 + pctwhite +
              pcths18_24 + pcths25_over + pctbachdeg25_over + pctunemployed16_over +
              pctprivatecoverage + pctempprivcoverage + pctotherrace +
              pctmarriedhouseholds + birthrate, data = learningset5)
p3_b = predict(model3, newdata = testset5)
cat('OLS on LASSO selected features:',sqrt(mean((p3_b-testset5$target_deathrate)**2)),'\n')
```

```
## OLS on LASSO selected features: 15.25454
```

We perform also a CART algorithm, a model issue by random forest.

```
# construction of the maximal tree

model4=rpart(target_deathrate~., data=learningset5,
             control = rpart.control(minsplit=2, cp=10^(-15)))
```

We make sure we have the maximal tree

```
Prediction4max=predict(model4)
err=sum((Prediction4max-learningset5$target_deathrate)^2)
err
```

```
## [1] 0
```

Since err=0, we obtain the maximal tree

Lets perform the pruning step and the final selection We get the information about the constructed subtree

```
CP= model4$cptable
CP[1:5,]
```

```
##           CP nsplit rel error    xerror       xstd
## 1 0.23820351      0 1.0000000 1.0026168 0.04201381
## 2 0.06908795      1 0.7617965 0.7734946 0.03433068
## 3 0.06764501      2 0.6927085 0.7269142 0.03169508
## 4 0.03342520      3 0.6250635 0.6703091 0.02972857
## 5 0.02992187      4 0.5916383 0.6374663 0.02941107
```

```
#The first thing is to find the smallest CV error
cvmin=min(CP[,4])
cvmin
```

```
## [1] 0.5078215
```

```
# and to determin which row it corresponds to
r=which(CP[,4]==cvmin)
r
```

```
## 41
## 41
```

Now we will construct the final tree

```
#The threshold for the 1-SE rule
t=CP[r,4]+1*CP[r,5]

z=which(CP[,4]<=t)

z_selected=z[1]
model4_prune=prune(model4, cp= CP[z_selected,1])
```

```
p4_b=predict(model4_prune, newdata = testset5)
cat('CART final tree:',sqrt(mean((p4_b-testset5$target_deathrate)^2)),'\n')
```

```
## CART final tree: 17.59893
```

Let's identify thanks to VSURF the subset of interested variables and use this subset to construct a CART tree.

```
summary(tmp)
```

```
##
##  VSURF computation time: 1.7 hours
##
##  VSURF selected:
##  30 variables at thresholding step (in 34.7 mins)
##  18 variables at interpretation step (in 47.7 mins)
##  15 variables at prediction step (in 16.8 mins)
```

VSURF allows to perform variable selection by using Random Forest.

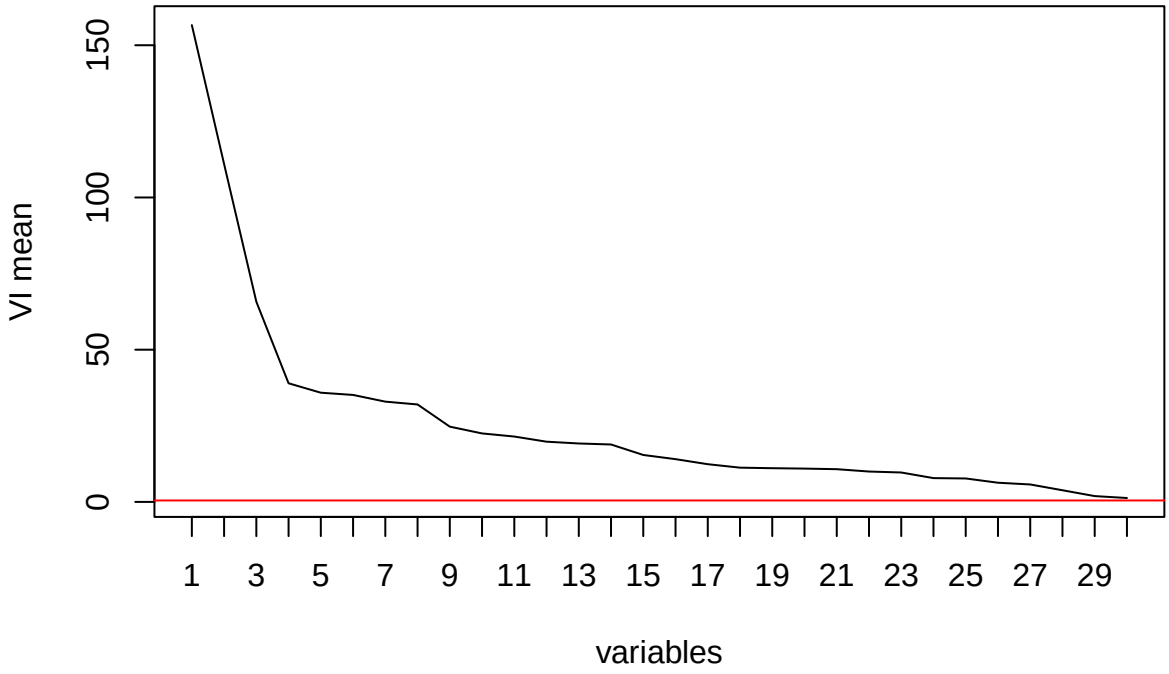
It is a two steps procedures:

Step: Suppression of the noisy variables

-After severals runs of RF, variables are sorted by the RF variable importance (VI).

-Variable with small importance are eliminated, using a threshold given by a CART model where the response variable is the sd associated to the VI variables rank and explanatory variable will by the rank of the VI.

```
plot(tmp, step = "thres", imp.sd = FALSE, var.names = TRUE)
```



The red line corresponds to the threshold value for VI. Only the variables with an averaged VI above this level are kept. At this step, only one variable has been eliminated.

```
number <- c(1:30)
number[tmp$varselect.thres]
```

```
##  [1]  3 17 16  4 18  6 24 21 20 19 23 26  2 28 14 10 29  1 11 27 22 25  5  9  8
## [26] 15 13 12 30  7
```

```
print(colnames(x)[tmp$varselect.thres])
```

```
##  [1] "incidencerate"           "pctbachdeg25_over"
##  [3] "pcths25_over"            "medincome"
##  [5] "pctemployed16_over"      "povertypercent"
```

```
##  [7] "pctpubliccoveragealone"  "pctprivatecoveragealone"
##  [9] "pctprivatecoverage"      "pctunemployed16_over"
## [11] "pctpubliccoverage"       "pctblack"
## [13] "avgdeathsperyear"        "pctotherrace"
## [15] "pctsomecol18_24"         "medianagefemale"
## [17] "pctmarriedhouseholds"    "avganncount"
## [19] "percentmarried"          "pctasian"
## [21] "pctempprivcoverage"      "pctwhite"
## [23] "popest2015"              "medianagemale"
## [25] "medianage"               "pctbachdeg18_24"
## [27] "pcths18_24"              "pctnohs18_24"
## [29] "birthrate"               "studypercap"
```

However it doesn't mean the remaining variables is the correct subset of variables.

Step 2 : variable selection First, We will identify the subset of explanatory variables which gives the model with the lowest OOB (out of bag error). This will the variable selection for interpretation.

```
number[tmp$varselect.interp]
```

```
##  [1]  3 17 16  4 18  6 24 21 20 19 23 26  2 28 14 10 29  1
```

```
print(colnames(x)[tmp$varselect.interp])
```

```
##  [1] "incidencerate"           "pctbachdeg25_over"
##  [3] "pcths25_over"            "medincome"
##  [5] "pctemployed16_over"      "povertypercent"
##  [7] "pctpubliccoveragealone"  "pctprivatecoveragealone"
##  [9] "pctprivatecoverage"      "pctunemployed16_over"
## [11] "pctpubliccoverage"       "pctblack"
## [13] "avgdeathsperyear"        "pctotherrace"
## [15] "pctsomecol18_24"         "medianagefemale"
## [17] "pctmarriedhouseholds"    "avganncount"
```

We have now 18 variables selected

Finally, we have a step of testing, we want to validate that by adding these variables, it is giving us additional information. So we might reduced the number of potential selected variables. This test is based on the decrease of OOB by adding a variable that should by larger than adding the average variation obtained by adding noisy variables.

```
number[tmp$varselect.pred]
```

```
##  [1]  3 17 16  4 18  6 24 19 23 26  2 28 10 29  1
```

```
print(colnames(x)[tmp$varselect.pred])
```

```
##  [1] "incidencerate"          "pctbachdeg25_over"      "pcths25_over"
##  [4] "medincome"              "pctemployed16_over"     "povertypercent"
##  [7] "pctpubliccoveragealone" "pctunemployed16_over"   "pctpubliccoverage"
## [10] "pctblack"               "avgdeathsperyear"       "pctotherrace"
## [13] "medianagefemale"        "pctmarriedhouseholds"   "avganncount"
```

We end up with 15 selected variables.

Let's perfom a second CART tree using this subset of selected variables.

Perform also a CART algorithm, a model issue by random forest.

```
# construction of the maximal tree

model5=rpart(target_deathrate~incidencerate + pctbachdeg25_over + pcths25_over +
               medincome + pctemployed16_over + povertypercent +
               pctpubliccoveragealone + pctunemployed16_over + pctpubliccoverage
             + pctblack + avgdeathsperyear + pctotherrace + medianagefemale +
               pctmarriedhouseholds + percentmarried, data=learningset5,
             control = rpart.control(minsplit=2, cp=10^(-15)))
```

We make sure we have the maximal tree

```
Tree5max=predict(model5)
err=sum((Tree5max-learningset5$target_deathrate)^2)
err
```

```
## [1] 0
```

Since err=0, we obtain the maximal tree

Lets perform the pruning step and the final selection We get the information about the constructed subtree

```
CP= model5$cptable
CP[1:5,]
```

```
##           CP nsplit rel error    xerror       xstd
## 1 0.23820351      0 1.0000000 1.0005786 0.04193924
## 2 0.06908795      1 0.7617965 0.7758218 0.03435207
## 3 0.06764501      2 0.6927085 0.7295394 0.03117863
## 4 0.03342520      3 0.6250635 0.6720870 0.02890792
## 5 0.02992187      4 0.5916383 0.6393725 0.02878077
```

```
#The first thing is to find the smallest CV error
cvmin=min(CP[,4])
cvmin
```

```
## [1] 0.5217984
```

```
# and to determin which row it corresponds to
r=which(CP[,4]==cvmin)
r
```

```
## 50
## 50
```

Now we will construct the final tree

```
#The threshold for the 1-SE rule
t=CP[r,4]+1*CP[r,5]

z=which(CP[,4]<=t)

z_selected=z[1]
model5_pruned=prune(model5, cp= CP[z_selected,1])
```

```
p5=predict(model5_pruned, newdata = testset5)
cat('CART + VSURF final tree:',sqrt(mean((p5-testset5$target_deathrate)^2)),'\n')
```

```
## CART + VSURF final tree: 17.86153
```

RandomForest

```
model6=randomForest(target_deathrate~.,data=learningset5)
p6=predict(model6, newdata = testset5)
```

Summary of the models

```
cat('OLS full dataset (numericals):',sqrt(mean((p1-testset3$target_deathrate)**2)),'\n')
```

```
## OLS full dataset (numericals): 19.25947
```

```
cat('OLS without outliers:',sqrt(mean((p1b-testset5$target_deathrate)**2)),'\n')
```

```
## OLS without outliers: 15.17405
```

```
cat('Step default parameters:',sqrt(mean((p2-testset5$target_deathrate)^2)),'\n')
```

```
## Step default parameters: 15.17311
```

```
cat('Step forward or backward:',sqrt(mean((p2c-testset5$target_deathrate)**2)),'\n')
```

```
## Step forward or backward: 15.16503
```

```
cat('LASSO:',sqrt(mean((p3_a-testset5$target_deathrate)^2)),'\n')
```

```
## LASSO: 15.24858
```

```
cat('OLS on LASSO selected features:',sqrt(mean((p3_b-testset5$target_deathrate)**2)),'\n')
```

```
## OLS on LASSO selected features: 15.25454
```

```
cat('CART final tree:',sqrt(mean((p4_b-testset5$target_deathrate)^2)),'\n')
```

```
## CART final tree: 17.59893
```

```
cat('CART + VSURF final tree:',sqrt(mean((p5-testset5$target_deathrate)^2)),'\n')
```

```
## CART + VSURF final tree: 17.86153
```

```
cat('Random Forest:',sqrt(mean((p6-testset5$target_deathrate)^2)),'\n')
```

```
## Random Forest: 14.73808
```

According the RMSE, the best model is Random Forest followed by step method forward or backward. To improve the analysis k-fold cross-validation could be performed for each model.