HW1_v1.r

nhirata

2020-02-19

```
# Using the same crime data set uscrime.txt as in Question 8.2, apply Principal Component Analysis
# and then create a regression model using the first few principal components. Specify your new model in
# terms of the original variables (not the principal components), and compare its quality to that of your
# solution to Question 8.2. You can use the R function prcomp for PCA.
# (Note that to first scale the data, you can include scale. = TRUE to scale as
# part of the PCA function. Don't forget that, to make a prediction for the new city,
# you'll need to unscale the coefficients (i.e., do the scaling calculation in
# reverse)!)


rm(list = ls()) #Clear the list
library(GGally) #used for ggpairs
```

```
## Loading required package: ggplot2
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
set.seed(1) #reproducible values

data <-read.table("C:/Users/nhirata/Desktop/Georgia Tech/OneDrive - Georgia Institute of Technology/Georgia Tech/ISYE_6501/W
eek_6/data 9.1/uscrime.txt", header=TRUE, stringsAsFactors = FALSE)

head(data) #quick check
```

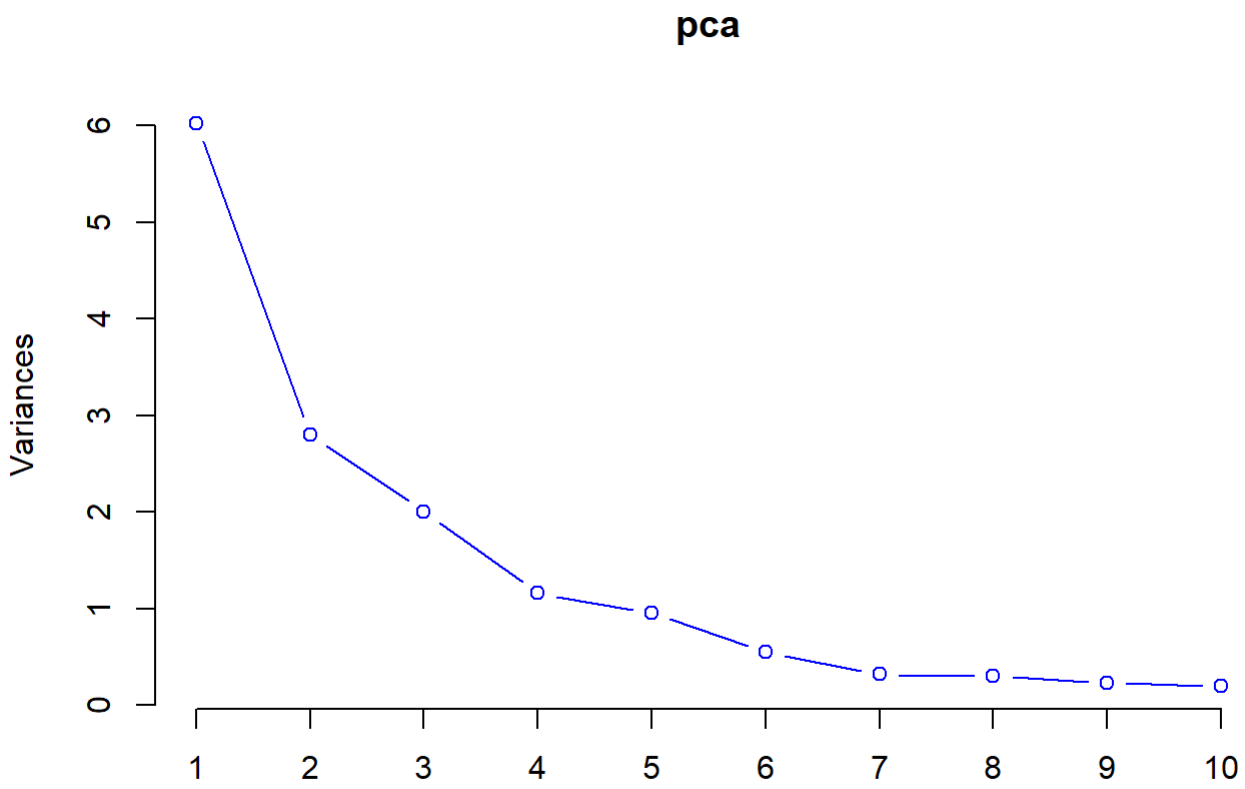
```
##      M So   Ed  Po1  Po2    LF   M.F Pop   NW    U1  U2 Wealth Ineq     Prob
## 1 15.1  1  9.1  5.8  5.6 0.510  95.0  33 30.1 0.108 4.1   3940 26.1 0.084602
## 2 14.3  0 11.3 10.3  9.5 0.583 101.2  13 10.2 0.096 3.6   5570 19.4 0.029599
## 3 14.2  1  8.9  4.5  4.4 0.533  96.9  18 21.9 0.094 3.3   3180 25.0 0.083401
## 4 13.6  0 12.1 14.9 14.1 0.577  99.4 157  8.0 0.102 3.9   6730 16.7 0.015801
## 5 14.1  0 12.1 10.9 10.1 0.591  98.5  18  3.0 0.091 2.0   5780 17.4 0.041399
## 6 12.1  0 11.0 11.8 11.5 0.547  96.4  25  4.4 0.084 2.9   6890 12.6 0.034201
##      Time Crime
## 1 26.2011   791
## 2 25.2999  1635
## 3 24.3006   578
## 4 29.9012  1969
## 5 21.2998  1234
## 6 20.9995   682
```

```
# Perform PCA on scaled attributes
pca <- prcomp(data[,1:15], scale. = TRUE)
summary(pca)
```

```
## Importance of components:
##                           PC1    PC2    PC3     PC4     PC5     PC6     PC7
## Standard deviation     2.4534 1.6739 1.4160 1.07806 0.97893 0.74377 0.56729
## Proportion of Variance 0.4013 0.1868 0.1337 0.07748 0.06389 0.03688 0.02145
## Cumulative Proportion  0.4013 0.5880 0.7217 0.79920 0.86308 0.89996 0.92142
##                            PC8     PC9    PC10    PC11    PC12    PC13   PC14
## Standard deviation     0.55444 0.48493 0.44708 0.41915 0.35804 0.26333 0.2418
## Proportion of Variance 0.02049 0.01568 0.01333 0.01171 0.00855 0.00462 0.0039
## Cumulative Proportion  0.94191 0.95759 0.97091 0.98263 0.99117 0.99579 0.9997
##                           PC15
## Standard deviation     0.06793
## Proportion of Variance 0.00031
## Cumulative Proportion  1.00000
```

```
# Useful visualization when deciding how many principal components to choose.

screeplot(pca, type="lines",col="blue")
```




```
## Retrieve first 4 PCs

# From prcomp output
PCs <- pca$x[,1:4]
attributes(pca$x)
```

```
## $dim
## [1] 47 15
##
## $dimnames
## $dimnames[[1]]
## NULL
##
## $dimnames[[2]]
##  [1] "PC1"  "PC2"  "PC3"  "PC4"  "PC5"  "PC6"  "PC7"  "PC8"  "PC9"  "PC10"
## [11] "PC11" "PC12" "PC13" "PC14" "PC15"
```

```
pca$x
```

```
##              PC1         PC2         PC3         PC4          PC5         PC6
##  [1,] -4.1992835 -1.09383120 -1.11907395  0.67178115  0.055283376  0.30733835
##  [2,]  1.1726630  0.67701360 -0.05244634 -0.08350709 -1.173199821 -0.58323731
##  [3,] -4.1737248  0.27677501 -0.37107658  0.37793995  0.541345246  0.71872230
##  [4,]  3.8349617 -2.57690596  0.22793998  0.38262331 -1.644746496  0.72948841
##  [5,]  1.8392999  1.33098564  1.27882805  0.71814305  0.041590320 -0.39409015
##  [6,]  2.9072336 -0.33054213  0.53288181  1.22140635  1.374360960 -0.69225131
##  [7,]  0.2457752 -0.07362562 -0.90742064  1.13685873  0.718644387 -0.93107472
##  [8,] -0.1301330 -1.35985577  0.59753132  1.44045387 -0.222781388  0.04912052
##  [9,] -3.6103169 -0.68621008  1.28372246  0.55171150 -0.324292990  0.12683417
## [10,]  1.1672376  3.03207033  0.37984502 -0.28887026 -0.646056610  0.33130781
## [11,]  2.5384879 -2.66771358  1.54424656 -0.87671210 -0.324083561  0.44365740
## [12,]  1.0065920 -0.06044849  1.18861346 -1.31261964  0.358087724  0.25696957
## [13,]  0.5161143  0.97485189  1.83351610 -1.59117618  0.599881946  1.04761756
## [14,]  0.4265556  1.85044812  1.02893477 -0.07789173  0.741887592  0.61569775
## [15,] -3.3435299  0.05182823 -1.01358113  0.08840211  0.002969448  0.17074576
## [16,] -3.0310689 -2.10295524 -1.82993161  0.52347187 -0.387454246 -0.20965321
## [17,] -0.2262961  1.44939774 -1.37565975  0.28960865  1.337784608 -0.25633983
## [18,] -0.1127499 -0.39407030 -0.38836278  3.97985093  0.410914404  0.09317136
## [19,]  2.9195668 -1.58646124  0.97612613  0.78629766  1.356288600 -0.89044651
## [20,]  2.2998485 -1.73396487 -2.82423222 -0.23281758 -0.653038858  0.68615337
## [21,]  1.1501667  0.13531015  0.28506743 -2.19770548  0.084621572  0.45958300
## [22,] -5.6594827 -1.09730404  0.10043541 -0.05245484 -0.689327990  0.13338054
## [23,] -0.1011749 -0.57911362  0.71128354 -0.44394773  0.689939865  0.54002731
## [24,]  1.3836281  1.95052341 -2.98485490 -0.35942784 -0.744371276  0.01453851
## [25,]  0.2727756  2.63013778  1.83189535  0.05207518  0.803692524  1.52313508
## [26,]  4.0565577  1.17534729 -0.81690756  1.66990720 -2.895110075 -0.47766314
## [27,]  0.8929694  0.79236692  1.26822542 -0.57575615  1.830793964 -1.11656766
## [28,]  0.1514495  1.44873320  0.10857670 -0.51040146 -1.023229895 -0.74149513
## [29,]  3.5592481 -4.76202163  0.75080576  0.64692974  0.309946510  0.72486153
## [30,] -4.1184576 -0.38073981  1.43463965  0.63330834 -0.254715638 -0.42316550
## [31,] -0.6811731  1.66926027 -2.88645794 -1.30977099 -0.470913997 -0.45866080
## [32,]  1.7157269 -1.30836339 -0.55971313 -0.70557980  0.331277622  1.30802615
## [33,] -1.8860627  0.59058174  1.43570145  0.18239089  0.291863659 -0.13885903
## [34,]  1.9526349  0.52395429 -0.75642216  0.44289927  0.723474420 -0.42036754
## [35,]  1.5888864 -3.12998571 -1.73107199 -1.68604766  0.665406182  0.54144206
## [36,]  1.0709414 -1.65628271  0.79436888 -1.85172698  0.020031154 -2.43356674
## [37,] -4.1101715  0.15766712  2.36296974 -0.56868399 -2.469679496  0.07239996
## [38,] -0.7254706  2.89263339 -0.36348376 -0.50612576  0.028157162  1.06465126
## [39,] -3.3451254 -0.95045293  0.19551398 -0.27716645  0.487259213 -0.20571166
```

```
## [40,] -1.0644466 -1.05265304  0.82886286 -0.12042931 -0.645884788  0.63320546
## [41,]  1.4933989  1.86712106  1.81853582 -1.06112429  0.009855774 -1.03480444
## [42,] -0.6789284  1.83156328 -1.65435992  0.95121379  2.115630145 -0.02332805
## [43,] -2.4164258 -0.46701087  1.42808323  0.41149015 -0.867397522 -1.13982198
## [44,]  2.2978729  0.41865689 -0.64422929 -0.63462770 -0.703116983 -0.65215040
## [45,] -2.9245282 -1.19488555 -3.35139309 -1.48966984  0.806659622 -0.48157983
## [46,]  1.7654525  0.95655926  0.98576138  1.05683769  0.542466034  0.71712602
## [47,]  2.3125056  2.56161119 -1.58223354  0.59863946 -1.140712406  0.39563373
##                PC7          PC8          PC9         PC10          PC11
##  [1,] -0.566408161 -0.007801727  0.223509947  0.452743650 -0.0847454174
##  [2,]  0.195611187  0.154566472  0.436777195  0.212085890 -0.0339166059
##  [3,]  0.103306929  0.351138883  0.062992321 -0.067190215 -0.4814915573
##  [4,]  0.266994985 -1.547460841 -0.379541806  0.229223052  0.1098495110
##  [5,]  0.070507664 -0.543237437  0.224632448  0.477690842 -0.3295818584
##  [6,]  0.226482092  0.562323186  0.417722172  0.091009390  0.0102296864
##  [7,]  0.307507661  1.056861503 -1.160218292  0.791683164  0.2829470570
##  [8,]  0.911404993  0.693339330 -0.421314146  0.613278523 -0.3211719754
##  [9,] -0.417420968 -0.053270500  0.232662026  0.065541569  0.1212937342
## [10,]  0.009579488 -0.329270845 -0.123629746  0.200126861 -0.0005664179
## [11,] -0.182961180  0.587179568 -0.070907596 -0.556615080 -0.1727018439
## [12,] -0.462577031  0.307351101 -0.105197263 -0.132898969  0.2984659116
## [13,] -0.494631320  0.753702337 -0.384056907 -0.340154686 -0.3093005372
## [14,] -0.087093101 -0.046931419 -0.159138488  0.280005792  0.1705829803
## [15,]  1.040213660 -0.139392628 -0.147546022 -1.024276227  0.7966941694
## [16,]  0.262430717  0.641818600  0.526895635  0.828407330 -0.2016395195
## [17,] -0.754882880 -0.959968310  0.351808733 -0.046049514  0.1106976222
## [18,] -1.227238054  0.280226677 -0.412734008 -1.074780984  0.1309449295
## [19,]  0.387161139 -0.002276046  0.555855685  0.598093089  0.3873076362
## [20,] -0.401936004  0.240456772  0.341543809  0.229195572  0.7640552201
## [21,] -0.179283176  0.772072202 -0.344317021 -0.192047623 -0.2491916653
## [22,] -1.337728458  0.261648468  0.225568667  0.361253314 -1.2502555533
## [23,]  0.995827754  0.371597176  1.073655584  0.033997150 -0.0148920689
## [24,]  0.042135169 -0.210603749 -0.111463892  0.570729260 -0.2891751385
## [25,] -0.341012092  0.390172476 -0.015090214 -0.107776581  0.0126408264
## [26,] -0.110906098  0.991890307  0.232407672 -0.727397771 -0.1821057801
## [27,] -0.199196211 -0.044269305 -0.015729946 -0.046457518 -0.2413405035
## [28,]  0.113082804 -0.677219677  0.151930973  0.076617716 -0.4139560352
## [29,]  0.248081636 -0.844089307  0.230269486 -0.342149453 -0.8429456727
## [30,] -0.116127247 -0.891169193 -0.011731985 -0.435636015  0.0144413727
## [31,]  0.704852096 -0.538600585  0.439137868 -0.709658521 -0.5740441221
## [32,] -0.786980332 -0.067086938 -0.169888285  0.072917031  0.6056884273
```

```
## [33,]  0.767856496  0.027448832 -0.773125607  0.126124015  0.1459949892
## [34,]  0.181257930  0.115379461 -0.101718594  0.321007813 -0.4060548228
## [35,] -0.449541256 -0.276891496  0.007657702  0.202491328  0.0936192141
## [36,] -0.333843509  0.384707595  0.642612190 -0.727991803  0.1824929850
## [37,] -0.343611407  0.157984131  0.915881371  0.481641023  1.1919120577
## [38,]  0.863051754 -0.058247210  0.341385143 -0.133649827 -0.5185529852
## [39,]  0.966860079  0.059557654  0.039345212  0.034036490  0.2185933062
## [40,]  0.767470212 -0.704833575 -1.109887730  0.106827471  0.1951224135
## [41,] -0.589160590 -0.468876595 -0.528478950  0.430811630  0.1829897714
## [42,] -0.557413301 -0.963360913  0.485515025  0.007295728  0.4739341401
## [43,]  0.041128192 -0.573696577 -0.773992630 -0.447789368 -0.1172352964
## [44,] -0.442990964 -0.093002011 -0.515838387  0.241578722 -0.1363783451
## [45,]  0.233636019  0.379908278 -0.815127937 -0.541397364  0.2642920144
## [46,]  0.847914876  0.172381544  0.657987377 -0.480124036  0.1175554086
## [47,] -0.171412192  0.327844331 -0.167078790 -0.002371858  0.2888983375
##              PC12         PC13         PC14          PC15
##  [1,]  0.22096639 -0.112616798  0.326964861  0.0233840087
##  [2,]  0.35686524  0.297516509  0.252356741 -0.0607636781
##  [3,] -0.04701948  0.052160542 -0.486551130  0.0421174952
##  [4,]  0.17727101  0.088381306  0.149678420  0.0291749700
##  [5,]  0.41807551 -0.722152235  0.131027187 -0.0751493967
##  [6,] -0.70661980 -0.135172709  0.194925675  0.0155861048
##  [7,] -0.65196573  0.168327740  0.145473719 -0.0654492790
##  [8,]  0.49089082  0.218057687 -0.623230400 -0.0259344691
##  [9,] -0.29249322 -0.242429444  0.026476592  0.0252300906
## [10,] -0.21063943 -0.257769674 -0.276967642  0.0232404560
## [11,] -0.33472808  0.238074383  0.255472039  0.0992321732
## [12,] -0.26641418  0.171319693  0.094123766  0.0190525547
## [13,]  0.59785665 -0.132203906  0.027925309 -0.0148583070
## [14,]  0.18719968  0.571485989  0.250689865  0.0127642083
## [15,]  0.56068471  0.217331625  0.037229143  0.0452385996
## [16,] -0.16367226 -0.082957159  0.137971468 -0.0210413021
## [17,]  0.33986466 -0.128534101 -0.246396571 -0.0073811334
## [18,] -0.16259339 -0.474477655  0.096820598  0.0107830419
## [19,]  0.49141798  0.110318335 -0.185686144  0.1027680411
## [20,]  0.05854928  0.173991982  0.041243802 -0.0108009160
## [21,]  0.03436398 -0.407556122  0.094462966 -0.0062668835
## [22,]  0.15171519  0.319206246  0.003834903 -0.0005073113
## [23,]  0.08607424 -0.037204214  0.545497655  0.0129578778
## [24,] -0.20783571 -0.240516367 -0.122497400 -0.0342080182
## [25,]  0.37619331  0.117057471 -0.105183565 -0.0510978767
```

```
## [26,]  0.30036333  0.137225797 -0.134072192 -0.1184870411
## [27,] -0.51580918  0.066145794 -0.186576416  0.0791823778
## [28,]  0.24306271 -0.140043507  0.629391628 -0.0354269136
## [29,]  0.03561083 -0.229673348 -0.234477116  0.0387679658
## [30,] -0.36730664  0.388569856 -0.025869303 -0.0300544785
## [31,] -0.79220655  0.007892720 -0.201914013  0.0766956405
## [32,] -0.34195913  0.154638372  0.085491563 -0.0800132601
## [33,]  0.25911938 -0.316086918 -0.024206874  0.1045722437
## [34,]  0.25952688  0.166191625  0.152140934  0.0830313640
## [35,] -0.33281300  0.047752123 -0.312239740 -0.1013067365
## [36,]  0.47165172  0.049320737 -0.382422475 -0.0704633747
## [37,] -0.31784996 -0.395326593 -0.238009619  0.0858414347
## [38,] -0.25514910  0.169135060 -0.013058191 -0.0353381517
## [39,]  0.08796506  0.030789317 -0.067516845 -0.1026461875
## [40,] -0.05840207 -0.137544171 -0.177710919 -0.0704026331
## [41,] -0.26187866 -0.058757893 -0.113235908 -0.0939372094
## [42,]  0.33534399  0.291642167  0.013605734 -0.0399895760
## [43,] -0.26398492  0.427157629  0.266115989 -0.0276514754
## [44,]  0.17238472  0.005592707  0.142206916  0.1612571077
## [45,]  0.39144866 -0.508852301  0.223930669  0.0073779464
## [46,] -0.56753437 -0.172018049  0.056680914 -0.0850410458
## [47,]  0.01440895  0.246609753 -0.223916593  0.1659609523
```

```
PCs
```

```
##              PC1         PC2         PC3         PC4
##  [1,] -4.1992835 -1.09383120 -1.11907395  0.67178115
##  [2,]  1.1726630  0.67701360 -0.05244634 -0.08350709
##  [3,] -4.1737248  0.27677501 -0.37107658  0.37793995
##  [4,]  3.8349617 -2.57690596  0.22793998  0.38262331
##  [5,]  1.8392999  1.33098564  1.27882805  0.71814305
##  [6,]  2.9072336 -0.33054213  0.53288181  1.22140635
##  [7,]  0.2457752 -0.07362562 -0.90742064  1.13685873
##  [8,] -0.1301330 -1.35985577  0.59753132  1.44045387
##  [9,] -3.6103169 -0.68621008  1.28372246  0.55171150
## [10,]  1.1672376  3.03207033  0.37984502 -0.28887026
## [11,]  2.5384879 -2.66771358  1.54424656 -0.87671210
## [12,]  1.0065920 -0.06044849  1.18861346 -1.31261964
## [13,]  0.5161143  0.97485189  1.83351610 -1.59117618
## [14,]  0.4265556  1.85044812  1.02893477 -0.07789173
## [15,] -3.3435299  0.05182823 -1.01358113  0.08840211
## [16,] -3.0310689 -2.10295524 -1.82993161  0.52347187
## [17,] -0.2262961  1.44939774 -1.37565975  0.28960865
## [18,] -0.1127499 -0.39407030 -0.38836278  3.97985093
## [19,]  2.9195668 -1.58646124  0.97612613  0.78629766
## [20,]  2.2998485 -1.73396487 -2.82423222 -0.23281758
## [21,]  1.1501667  0.13531015  0.28506743 -2.19770548
## [22,] -5.6594827 -1.09730404  0.10043541 -0.05245484
## [23,] -0.1011749 -0.57911362  0.71128354 -0.44394773
## [24,]  1.3836281  1.95052341 -2.98485490 -0.35942784
## [25,]  0.2727756  2.63013778  1.83189535  0.05207518
## [26,]  4.0565577  1.17534729 -0.81690756  1.66990720
## [27,]  0.8929694  0.79236692  1.26822542 -0.57575615
## [28,]  0.1514495  1.44873320  0.10857670 -0.51040146
## [29,]  3.5592481 -4.76202163  0.75080576  0.64692974
## [30,] -4.1184576 -0.38073981  1.43463965  0.63330834
## [31,] -0.6811731  1.66926027 -2.88645794 -1.30977099
## [32,]  1.7157269 -1.30836339 -0.55971313 -0.70557980
## [33,] -1.8860627  0.59058174  1.43570145  0.18239089
## [34,]  1.9526349  0.52395429 -0.75642216  0.44289927
## [35,]  1.5888864 -3.12998571 -1.73107199 -1.68604766
## [36,]  1.0709414 -1.65628271  0.79436888 -1.85172698
## [37,] -4.1101715  0.15766712  2.36296974 -0.56868399
## [38,] -0.7254706  2.89263339 -0.36348376 -0.50612576
## [39,] -3.3451254 -0.95045293  0.19551398 -0.27716645
```

```
## [40,] -1.0644466 -1.05265304  0.82886286 -0.12042931
## [41,]  1.4933989  1.86712106  1.81853582 -1.06112429
## [42,] -0.6789284  1.83156328 -1.65435992  0.95121379
## [43,] -2.4164258 -0.46701087  1.42808323  0.41149015
## [44,]  2.2978729  0.41865689 -0.64422929 -0.63462770
## [45,] -2.9245282 -1.19488555 -3.35139309 -1.48966984
## [46,]  1.7654525  0.95655926  0.98576138  1.05683769
## [47,]  2.3125056  2.56161119 -1.58223354  0.59863946
```

```
# Build linear regression model with the first 4 principal components

PCcrime <- cbind(PCs, data[,16]) #Create new data matrix with first 4 PCs and crime rate

PCcrime
```

```
##              PC1         PC2         PC3         PC4
##  [1,] -4.1992835 -1.09383120 -1.11907395  0.67178115  791
##  [2,]  1.1726630  0.67701360 -0.05244634 -0.08350709 1635
##  [3,] -4.1737248  0.27677501 -0.37107658  0.37793995  578
##  [4,]  3.8349617 -2.57690596  0.22793998  0.38262331 1969
##  [5,]  1.8392999  1.33098564  1.27882805  0.71814305 1234
##  [6,]  2.9072336 -0.33054213  0.53288181  1.22140635  682
##  [7,]  0.2457752 -0.07362562 -0.90742064  1.13685873  963
##  [8,] -0.1301330 -1.35985577  0.59753132  1.44045387 1555
##  [9,] -3.6103169 -0.68621008  1.28372246  0.55171150  856
## [10,]  1.1672376  3.03207033  0.37984502 -0.28887026  705
## [11,]  2.5384879 -2.66771358  1.54424656 -0.87671210 1674
## [12,]  1.0065920 -0.06044849  1.18861346 -1.31261964  849
## [13,]  0.5161143  0.97485189  1.83351610 -1.59117618  511
## [14,]  0.4265556  1.85044812  1.02893477 -0.07789173  664
## [15,] -3.3435299  0.05182823 -1.01358113  0.08840211  798
## [16,] -3.0310689 -2.10295524 -1.82993161  0.52347187  946
## [17,] -0.2262961  1.44939774 -1.37565975  0.28960865  539
## [18,] -0.1127499 -0.39407030 -0.38836278  3.97985093  929
## [19,]  2.9195668 -1.58646124  0.97612613  0.78629766  750
## [20,]  2.2998485 -1.73396487 -2.82423222 -0.23281758 1225
## [21,]  1.1501667  0.13531015  0.28506743 -2.19770548  742
## [22,] -5.6594827 -1.09730404  0.10043541 -0.05245484  439
## [23,] -0.1011749 -0.57911362  0.71128354 -0.44394773 1216
## [24,]  1.3836281  1.95052341 -2.98485490 -0.35942784  968
## [25,]  0.2727756  2.63013778  1.83189535  0.05207518  523
## [26,]  4.0565577  1.17534729 -0.81690756  1.66990720 1993
## [27,]  0.8929694  0.79236692  1.26822542 -0.57575615  342
## [28,]  0.1514495  1.44873320  0.10857670 -0.51040146 1216
## [29,]  3.5592481 -4.76202163  0.75080576  0.64692974 1043
## [30,] -4.1184576 -0.38073981  1.43463965  0.63330834  696
## [31,] -0.6811731  1.66926027 -2.88645794 -1.30977099  373
## [32,]  1.7157269 -1.30836339 -0.55971313 -0.70557980  754
## [33,] -1.8860627  0.59058174  1.43570145  0.18239089 1072
## [34,]  1.9526349  0.52395429 -0.75642216  0.44289927  923
## [35,]  1.5888864 -3.12998571 -1.73107199 -1.68604766  653
## [36,]  1.0709414 -1.65628271  0.79436888 -1.85172698 1272
## [37,] -4.1101715  0.15766712  2.36296974 -0.56868399  831
## [38,] -0.7254706  2.89263339 -0.36348376 -0.50612576  566
## [39,] -3.3451254 -0.95045293  0.19551398 -0.27716645  826
```

```
## [40,] -1.0644466 -1.05265304  0.82886286 -0.12042931 1151
## [41,]  1.4933989  1.86712106  1.81853582 -1.06112429  880
## [42,] -0.6789284  1.83156328 -1.65435992  0.95121379  542
## [43,] -2.4164258 -0.46701087  1.42808323  0.41149015  823
## [44,]  2.2978729  0.41865689 -0.64422929 -0.63462770 1030
## [45,] -2.9245282 -1.19488555 -3.35139309 -1.48966984  455
## [46,]  1.7654525  0.95655926  0.98576138  1.05683769  508
## [47,]  2.3125056  2.56161119 -1.58223354  0.59863946  849
```

```
model <- lm(V5~., data = as.data.frame(PCcrime)) #Create regression model on PCcrime

summary(model)
```

```
##
## Call:
## lm(formula = V5 ~ ., data = as.data.frame(PCcrime))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -557.76 -210.91  -29.08  197.26  810.35
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   905.09      49.07  18.443  < 2e-16 ***
## PC1            65.22      20.22   3.225  0.00244 **
## PC2           -70.08      29.63  -2.365  0.02273 *
## PC3            25.19      35.03   0.719  0.47602
## PC4            69.45      46.01   1.509  0.13872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 336.4 on 42 degrees of freedom
## Multiple R-squared:  0.3091, Adjusted R-squared:  0.2433
## F-statistic: 4.698 on 4 and 42 DF,  p-value: 0.003178
```

```
## Multiple R-squared:  0.309,  Adjusted R-squared:  0.243


## Get coefficients in terms of original data from PCA coefficients

# PCA Coefficients for this linear regression model

beta0 <- model$coefficients[1]
betas <- model$coefficients[2:5]
beta0
```

```
## (Intercept)
##    905.0851
```

```
# Intercept beta0 is 905

betas
```

```
##       PC1       PC2       PC3       PC4
##  65.21593 -70.08312  25.19408  69.44603
```

```
##  PC1   PC2   PC3   PC4
## 65.2 -70.1  25.2  69.4

# Transform the PC coefficients into coefficients for the original variables
pca$rotation[,1:4]
```

```
##                PC1         PC2           PC3         PC4
## M      -0.30371194  0.06280357  0.1724199946 -0.02035537
## So     -0.33088129 -0.15837219  0.0155433104  0.29247181
## Ed      0.33962148  0.21461152  0.0677396249  0.07974375
## Po1     0.30863412 -0.26981761  0.0506458161  0.33325059
## Po2     0.31099285 -0.26396300  0.0530651173  0.35192809
## LF      0.17617757  0.31943042  0.2715301768 -0.14326529
## M.F     0.11638221  0.39434428 -0.2031621598  0.01048029
## Pop     0.11307836 -0.46723456  0.0770210971 -0.03210513
## NW     -0.29358647 -0.22801119  0.0788156621  0.23925971
## U1      0.04050137  0.00807439 -0.6590290980 -0.18279096
## U2      0.01812228 -0.27971336 -0.5785006293 -0.06889312
## Wealth  0.37970331 -0.07718862  0.0100647664  0.11781752
## Ineq   -0.36579778 -0.02752240 -0.0002944563 -0.08066612
## Prob   -0.25888661  0.15831708 -0.1176726436  0.49303389
## Time   -0.02062867 -0.38014836  0.2235664632 -0.54059002
```

```
alphas <- pca$rotation[,1:4] %*% betas
t(alphas)
```

```
##              M       So       Ed      Po1      Po2        LF       M.F      Pop
## [1,] -21.27796 10.22309 14.35261 63.45643 64.55797 -14.00535 -24.43757 39.83067
##             NW        U1       U2   Wealth      Ineq     Prob      Time
## [1,] 15.43455 -27.22228 1.425902 38.60786 -27.53635 3.295707 -6.612616
```

```
##          M   So   Ed  Po1  Po2  LF   M.F  Pop   NW    U1   U2 Wealth  Ineq Prob  Time
## [1,] -21.3 10.2 14.4 63.5 64.6 -14 -24.4 39.8 15.4 -27.2 1.43   38.6 -27.5  3.3 -6.61

# However, these coefficients listed above are scaled.
# Must convert back to the original data.

# When scaling, this function subtracts the mean and divides by the standard deviation, for each variable.
#
# So, alpha * (x - mean)/sd = originalAlpha * x.
# That means:
# (1) originalAlpha = alpha/sd
# (2) we have to modify the constant term a0 by alpha*mean/sd

originalAlpha <- alphas/sapply(data[,1:15],sd)
originalBeta0 <- beta0 - sum(alphas*sapply(data[,1:15],mean)/sapply(data[,1:15],sd))

# Here are the coefficients for unscaled data:

t(originalAlpha)
```

```
##              M       So       Ed      Po1      Po2        LF       M.F      Pop
## [1,] -16.93076 21.34368 12.82972 21.35216 23.08832 -346.5657 -8.293097 1.046216
##            NW        U1       U2    Wealth      Ineq     Prob       Time
## [1,] 1.500994 -1509.935 1.688367 0.0400119 -6.902022 144.9493 -0.9330765
```

```
originalBeta0
```

```
## (Intercept)
##    1666.485
```

```
## 1667

# Estimates of the model:

estimates <- as.matrix(data[,1:15]) %*% originalAlpha + originalBeta0
estimates
```

```
##           [,1]
##  [1,]  726.3425
##  [2,]  926.9936
##  [3,]  630.3920
##  [4,] 1368.0977
##  [5,] 1013.8482
##  [6,] 1216.0958
##  [7,]  982.3622
##  [8,] 1106.9894
##  [9,]  788.3831
## [10,]  758.2196
## [11,] 1235.6183
## [12,]  913.7572
## [13,]  806.1162
## [14,]  823.7319
## [15,]  664.0043
## [16,]  845.0424
## [17,]  774.2024
## [18,] 1191.9501
## [19,] 1285.8694
## [20,] 1089.2713
## [21,]  825.1714
## [22,]  611.7868
## [23,]  926.1627
## [24,]  758.4594
## [25,]  788.3155
## [26,] 1182.6525
## [27,]  899.7572
## [28,]  780.7204
## [29,] 1534.7847
## [30,]  743.3047
## [31,]  579.9948
## [32,] 1045.5708
## [33,]  789.5315
## [34,] 1007.4079
## [35,] 1067.3629
## [36,]  982.4233
## [37,]  646.0266
## [38,]  610.7420
## [39,]  739.2180
```

```
## [40,]  921.9585
## [41,]  843.7503
## [42,]  756.8244
## [43,]  844.7808
## [44,]  965.2991
## [45,]  610.2137
## [46,] 1051.4106
## [47,]  878.0818
```

```
# Calculate R^2 and Adjusted R^2

SSE = sum((estimates - data[,16])^2)
SStot = sum((data[,16] - mean(data[,16]))^2)
1 - SSE/SStot
```

```
## [1] 0.3091121
```

```
## 0.309

R2 <- 1 - SSE/SStot
R2 - (1 - R2)*4/(nrow(data)-4-1)
```

```
## [1] 0.2433132
```

```
## 0.243

# The R-squared and Adjusted R-squared are eqaul when using the PCA ranges.

# Compare with the regression model from previous homework

model2 <- lm( Crime ~ ., data = data)
summary(model2)
```

```
##
## Call:
## lm(formula = Crime ~ ., data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -395.74  -98.09   -6.69  112.99  512.67
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.984e+03  1.628e+03  -3.675 0.000893 ***
## M            8.783e+01  4.171e+01   2.106 0.043443 *
## So          -3.803e+00  1.488e+02  -0.026 0.979765
## Ed           1.883e+02  6.209e+01   3.033 0.004861 **
## Po1          1.928e+02  1.061e+02   1.817 0.078892 .
## Po2         -1.094e+02  1.175e+02  -0.931 0.358830
## LF          -6.638e+02  1.470e+03  -0.452 0.654654
## M.F          1.741e+01  2.035e+01   0.855 0.398995
## Pop         -7.330e-01  1.290e+00  -0.568 0.573845
## NW           4.204e+00  6.481e+00   0.649 0.521279
## U1          -5.827e+03  4.210e+03  -1.384 0.176238
## U2           1.678e+02  8.234e+01   2.038 0.050161 .
## Wealth       9.617e-02  1.037e-01   0.928 0.360754
## Ineq         7.067e+01  2.272e+01   3.111 0.003983 **
## Prob        -4.855e+03  2.272e+03  -2.137 0.040627 *
## Time        -3.479e+00  7.165e+00  -0.486 0.630708
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 209.1 on 31 degrees of freedom
## Multiple R-squared:  0.8031, Adjusted R-squared:  0.7078
## F-statistic: 8.429 on 15 and 31 DF,  p-value: 3.539e-07
```

```
# This model has R^2 = 0.803 and R^2_adj = 0.708.

# Results suggest that using all the factors peforms better than running the PCA.

# Let's try all possibilities of principle components to double check.

r2 <- numeric(15) # create a vector to store the R-squared values

for (i in 1:15) {
  pclist <- pca$x[,1:i]  # use the first i prinicipal components
  pcc <- cbind(data[,16],pclist)  # create data set
  model <- lm(V1~.,data = as.data.frame(pcc)) # fit model
  r2[i] <- 1 - sum(model$residuals^2)/sum((data$Crime - mean(data$Crime))^2) # calculate R-squared
}

r2
```

```
##  [1] 0.1711351 0.2631339 0.2716416 0.3091121 0.6451941 0.6586023 0.6881819
##  [8] 0.6898765 0.6920491 0.6962873 0.6973865 0.7692656 0.7723664 0.7911447
## [15] 0.8030868
```

```
# This shows that the model is probably overfitted and thus cross validation is needed.
# In the previous homework, cross-validation resulted in a significantly smaller R-squared than the model showed on its trai
ning set.

library(DAAG)
```

```
## Loading required package: lattice
```

```
#
# # do 5-fold cross-validation on PCA Models
#
r2cross <- numeric(15) # create a vector to store the R-squared values

  for (i in 1:15) {
    pclist <- pca$x[,1:i]  # Run all components
    pcc <- cbind(data[,16],pclist)  # generate the data set
    model <- lm(V1~.,data = as.data.frame(pcc)) # fit the model
    c <- cv.lm(as.data.frame(pcc),model,m=5,plotit = FALSE) # Run a cross-validation
    r2cross[i] <- 1 - attr(c,"ms")*nrow(data)/sum((data$Crime - mean(data$Crime))^2) # calculate R-squared for each componen
t
  }
```

```
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value Pr(>F)
## pclist     1 1177568 1177568    9.29 0.0038 **
## Residuals 45 5703359  126741
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##                 1      3       17      18      19     22      36       38
## pclist       -4.2  -4.17   -0.226  -0.113    2.92  -5.66    1.07   -0.725
## cvpred      627.9 629.67  897.725 905.436 1111.35 528.77  985.82  863.827
## V1          791.0 578.00  539.000 929.000  750.00 439.00 1272.00  566.000
## CV residual 163.1 -51.67 -358.725  23.564 -361.35 -89.77  286.18 -297.827
##                 40
## pclist        -1.06
## cvpred       840.81
## V1          1151.00
## CV residual  310.19
##
## Sum of squares = 563955    Mean square = 62662    n = 9
##
## fold 2
## Observations in test set: 10
##                   4       6      12       25       28      32      34      41
## pclist         3.83    2.91    1.01    0.273    0.151    1.72    1.95    1.49
## cvpred      1187.20 1124.76  996.83  947.441  939.275 1044.56 1060.51 1029.60
## V1          1969.00  682.00  849.00  523.000 1216.000  754.00  923.00  880.00
## CV residual  781.80 -442.76 -147.83 -424.441  276.725 -290.56 -137.51 -149.60
##                44      46
## pclist        2.3    1.77
## cvpred     1083.7 1047.91
## V1         1030.0  508.00
## CV residual  -53.7 -539.91
##
## Sum of squares = 1505928    Mean square = 150593    n = 10
##
```

```
## fold 3
## Observations in test set: 10
##                   5       8      9      11     15       23     37     39     43
## pclist         1.84   -0.13  -3.61    2.54  -3.34   -0.101  -4.11  -3.35  -2.42
## cvpred       977.28  831.08 572.72 1029.19 592.53  833.230 535.62 592.41 661.35
## V1          1234.00 1555.00 856.00 1674.00 798.00 1216.000 831.00 826.00 823.00
## CV residual  256.72  723.92 283.28  644.81 205.47  382.770 295.38 233.59 161.65
##                  47
## pclist         2.31
## cvpred      1012.41
## V1           849.00
## CV residual -163.41
##
## Sum of squares = 1469370    Mean square = 146937    n = 10
##
## fold 4
## Observations in test set: 9
##                   7       13       14     20      24       27     30      35
## pclist        0.246    0.516    0.427    2.3    1.38    0.893  -4.12    1.59
## cvpred      966.188  984.485  978.423 1105.2 1043.20 1009.991 670.81 1057.09
## V1          963.000  511.000  664.000 1225.0  968.00  342.000 696.00  653.00
## CV residual  -3.188 -473.485 -314.423  119.8  -75.20 -667.991  25.19 -404.09
##                  45
## pclist        -2.92
## cvpred       751.62
## V1           455.00
## CV residual -296.62
##
## Sum of squares = 1041183    Mean square = 115687    n = 9
##
## fold 5
## Observations in test set: 9
##                   2      10     16      21      26      29       31      33
## pclist         1.17    1.17  -3.03    1.15    4.06    3.56   -0.681   -1.89
## cvpred       951.70  951.41 723.88  950.48 1107.99 1081.04  851.234  785.94
## V1          1635.00  705.00 946.00  742.00 1993.00 1043.00  373.000 1072.00
## CV residual  683.30 -246.41 222.12 -208.48  885.01  -38.04 -478.234  286.06
##                  42
## pclist        -0.679
## cvpred       851.356
## V1           542.000
```

```
## CV residual -309.356
##
## Sum of squares = 1811342    Mean square = 201260    n = 9
##
## Overall (Sum over all 9 folds)
##     ms
## 135995
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value Pr(>F)
## PC1        1 1177568 1177568   10.22 0.0026 **
## PC2        1  633037  633037    5.49 0.0237 *
## Residuals 44 5070322  115235
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##                 1     3   17     18   19   22   36   38   40
## Predicted   707.9 613.5  789 925.35 1207  613 1091  655  909
## cvpred      709.4 619.3  805 936.23 1219  612 1100  675  917
## V1          791.0 578.0  539 929.00  750  439 1272  566 1151
## CV residual  81.6 -41.3 -266  -7.23 -469 -173  172 -109  234
##
## Sum of squares = 425630    Mean square = 47292    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6   12   25   28   32   34    41     44   46
## Predicted   1336 1118  975  739  813 1109  996 871.6 1025.6  953
## cvpred      1331 1141  996  795  853 1115 1028 921.5 1057.7  991
## V1          1969  682  849  523 1216  754  923 880.0 1030.0  508
## CV residual  638 -459 -147 -272  363 -361 -105 -41.5  -27.7 -483
##
## Sum of squares = 1223077    Mean square = 122308    n = 10
##
## fold 3
## Observations in test set: 10
##                5    8   9   11  15   23  37  39  43    47
```

```
## Predicted    932  992 718 1258 683  939 626 754 780 876.4
## cvpred       909  905 620 1165 601  867 541 653 695 880.2
## V1          1234 1555 856 1674 798 1216 831 826 823 849.0
## CV residual  325  650 236  509 197  349 290 173 128 -31.2
##
## Sum of squares = 1134509    Mean square = 113451    n = 10
##
## fold 4
## Observations in test set: 9
##                  7   13   14     20  24   27     30   35   45
## Predicted   926.3  870  803 1176.6 859  908 663.18 1228  798
## cvpred      974.2  906  828 1249.7 884  946 705.36 1317  853
## V1          963.0  511  664 1225.0 968  342 696.00  653  455
## CV residual -11.2 -395 -164  -24.7  84 -604  -9.36 -664 -398
##
## Sum of squares = 1155193    Mean square = 128355    n = 9
##
## fold 5
## Observations in test set: 9
##                2    10    16   21   26   29   31   33   42
## Predicted    934 768.7 854.8  971 1087 1471  744  741  732
## cvpred       882 649.4 919.6  934  995 1552  680  719  664
## V1          1635 705.0 946.0  742 1993 1043  373 1072  542
## CV residual  753  55.6  26.4 -192  998 -509 -307  353 -122
##
## Sum of squares = 2097681    Mean square = 233076    n = 9
##
## Overall (Sum over all 9 folds)
##     ms
## 128427
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value Pr(>F)
## PC1        1 1177568 1177568   10.10 0.0027 **
## PC2        1  633037  633037    5.43 0.0245 *
## PC3        1   58541   58541    0.50 0.4823
## Residuals 43 5011782  116553
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
##
## fold 1
## Observations in test set: 9
##                1     3   17     18   19   22   36    38   40
## Predicted   680 604.1  754 915.57 1231  615 1111 645.9  930
## cvpred      677 603.8  768 925.99 1248  608 1122 660.5  936
## V1          791 578.0  539 929.00  750  439 1272 566.0 1151
## CV residual 114 -25.8 -229   3.01 -498 -169  150 -94.5  215
##
## Sum of squares = 420111    Mean square = 46679    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6   12   25   28   32     34   41     44   46
## Predicted   1342 1131 1005  785  816 1095  976.7  917 1009.4  978
## cvpred      1342 1172 1053  887  869 1095 1007.2 1011 1041.4 1044
## V1          1969  682  849  523 1216  754  923.0  880 1030.0  508
## CV residual  627 -490 -204 -364  347 -341  -84.2 -131  -11.4 -536
##
## Sum of squares = 1355143    Mean square = 135514    n = 10
##
## fold 3
## Observations in test set: 10
##                5    8   9   11  15   23  37  39  43    47
## Predicted    964 1007 750 1297 658  957 686 758 816 836.5
## cvpred       899  898 605 1150 606  859 518 647 680 894.5
## V1          1234 1555 856 1674 798 1216 831 826 823 849.0
## CV residual  335  657 251  524 192  357 313 179 143 -45.5
##
## Sum of squares = 1198019    Mean square = 119802    n = 10
##
## fold 4
## Observations in test set: 9
##                 7   13   14     20  24   27    30   35   45
## Predicted   903.4  917  829 1105.4 783  940 699.3 1184  714
## cvpred      937.1  966  864 1138.2 783  986 751.2 1238  731
## V1          963.0  511  664 1225.0 968  342 696.0  653  455
## CV residual  25.9 -455 -200   86.8 185 -644 -55.2 -585 -276
##
## Sum of squares = 1125976    Mean square = 125108    n = 9
##
```

```
## fold 5
## Observations in test set: 9
##                2    10    16   21   26   29   31   33    42
## Predicted    933 778.3 808.7  978 1067 1490  671  777 690.8
## cvpred       876 650.6 867.7  939  969 1583  591  753 609.7
## V1          1635 705.0 946.0  742 1993 1043  373 1072 542.0
## CV residual  759  54.4  78.3 -197 1024 -540 -218  319 -67.7
##
## Sum of squares = 2118791    Mean square = 235421    n = 9
##
## Overall (Sum over all 9 folds)
##     ms
## 132299
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value Pr(>F)
## PC1        1 1177568 1177568   10.40 0.0024 **
## PC2        1  633037  633037    5.59 0.0227 *
## PC3        1   58541   58541    0.52 0.4760
## PC4        1  257832  257832    2.28 0.1387
## Residuals 42 4753950  113189
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##                 1    3   17   18   19   22   36    38   40
## Predicted   726.3  630  774 1192 1286  612  982 610.7  922
## cvpred      806.4  687  828 1483 1355  638  879 606.3  935
## V1          791.0  578  539  929  750  439 1272 566.0 1151
## CV residual -15.4 -109 -289 -554 -605 -199  393 -40.3  216
##
## Sum of squares = 1010591    Mean square = 112288    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6    12   25   28   32   34    41     44   46
## Predicted   1368 1216 913.8  788  781 1046 1007 843.8  965.3 1051
## cvpred      1381 1282 929.4  881  817 1033 1046 906.3  982.7 1134
```

```
## V1          1969  682 849.0  523 1216  754  923 880.0 1030.0  508
## CV residual  588 -600 -80.4 -358  399 -279 -123 -26.3   47.3 -626
##
## Sum of squares = 1487411    Mean square = 148741    n = 10
##
## fold 3
## Observations in test set: 10
##                 5    8   9   11  15   23  37  39  43    47
## Predicted   1014 1107 788 1236 664  926 646 739 845 878.1
## cvpred       950  992 642 1090 615  831 481 629 707 942.3
## V1          1234 1555 856 1674 798 1216 831 826 823 849.0
## CV residual  284  563 214  584 183  385 350 197 116 -93.3
##
## Sum of squares = 1149649    Mean square = 114965    n = 10
##
## fold 4
## Observations in test set: 9
##                   7   13   14   20  24   27    30   35   45
## Predicted   982.362  806  824 1089 758  900 743.3 1067  610
## cvpred      963.673  923  865 1110 757  971 774.4 1167  665
## V1          963.000  511  664 1225 968  342 696.0  653  455
## CV residual  -0.673 -412 -201  115 211 -629 -78.4 -514 -210
##
## Sum of squares = 977599    Mean square = 108622    n = 9
##
## fold 5
## Observations in test set: 9
##                 2    10    16   21   26   29   31   33   42
## Predicted    927 758.2 845.0  825 1183 1535  580  790  757
## cvpred       873 634.6 889.8  852 1036 1620  535  758  643
## V1          1635 705.0 946.0  742 1993 1043  373 1072  542
## CV residual  762  70.4  56.2 -110  957 -577 -162  314 -101
##
## Sum of squares = 1986093    Mean square = 220677    n = 9
##
## Overall (Sum over all 9 folds)
##     ms
## 140667
## Analysis of Variance Table
##
## Response: V1
```

```
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## PC1        1 1177568 1177568   19.78 6.5e-05 ***
## PC2        1  633037  633037   10.63  0.0022 **
## PC3        1   58541   58541    0.98  0.3272
## PC4        1  257832  257832    4.33  0.0437 *
## PC5        1 2312556 2312556   38.84 2.0e-07 ***
## Residuals 41 2441394   59546
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##                 1     3    17   18   19   22   36  38     40
## Predicted   713.7 506.4 467.8 1098  975  770  978 604 1069.9
## cvpred      779.8 558.9 497.1 1277 1010  809  908 593 1079.3
## V1          791.0 578.0 539.0  929  750  439 1272 566 1151.0
## CV residual  11.2  19.1  41.9 -348 -260 -370  364 -27   71.7
##
## Sum of squares = 466218    Mean square = 51802    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6     12   25   28   32    34     41     44   46
## Predicted   1745  901 831.74  604 1015  970 841.7 841.49 1126.3  927
## cvpred      1716  995 857.44  681 1003  984 895.5 878.69 1121.4  998
## V1          1969  682 849.00  523 1216  754 923.0 880.00 1030.0  508
## CV residual  253 -313  -8.44 -158  213 -230  27.5   1.31  -91.4 -490
##
## Sum of squares = 534491    Mean square = 53449    n = 10
##
## fold 3
## Observations in test set: 10
##                5    8      9   11  15   23   37  39   43   47
## Predicted   1004 1158 862.66 1310 663  768 1212 628 1043 1139
## cvpred       990 1113 860.44 1217 677  699 1329 588 1062 1226
## V1          1234 1555 856.00 1674 798 1216  831 826  823  849
## CV residual  244  442  -4.44  457 121  517 -498 238 -239 -377
##
## Sum of squares = 1250340    Mean square = 125034    n = 10
##
```

```
## fold 4
## Observations in test set: 9
##                7   13     14     20    24   27   30   35    45
## Predicted    818  669 653.81 1238.8 929.0  480  802  915 425.5
## cvpred       816  727 672.63 1253.8 918.5  534  804  992 476.8
## V1           963  511 664.00 1225.0 968.0  342  696  653 455.0
## CV residual 147 -216  -8.63  -28.8  49.5 -192 -108 -339 -21.8
##
## Sum of squares = 235697    Mean square = 26189    n = 9
##
## fold 5
## Observations in test set: 9
##                 2   10     16   21   26   29   31   33  42
## Predicted    1196  906 933.79  806 1846 1464  688  723 272
## cvpred       1173  834 951.87  842 1785 1577  665  687 219
## V1           1635  705 946.00  742 1993 1043  373 1072 542
## CV residual  462 -129  -5.87 -100  208 -534 -292  385 323
##
## Sum of squares = 906598    Mean square = 100733    n = 9
##
## Overall (Sum over all 9 folds)
##    ms
## 72199
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## PC1        1 1177568 1177568   20.05 6.1e-05 ***
## PC2        1  633037  633037   10.78  0.0021 **
## PC3        1   58541   58541    1.00  0.3241
## PC4        1  257832  257832    4.39  0.0425 *
## PC5        1 2312556 2312556   39.38 1.9e-07 ***
## PC6        1   92261   92261    1.57  0.2173
## Residuals 40 2349133   58728
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##                 1     3    17   18   19   22   36  38   40
```

```
## Predicted   695.2 463.1 483.2 1092 1029  762 1124 540 1032
## cvpred      742.7 498.9 501.5 1243 1076  783 1070 522 1045
## V1          791.0 578.0 539.0  929  750  439 1272 566 1151
## CV residual  48.3  79.1  37.5 -314 -326 -344  202  44  106
##
## Sum of squares = 387320    Mean square = 43036    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6     12    25   28   32     34  41   44   46
## Predicted   1701  943 816.27 512.5 1060  891 867.01 904 1166  884
## cvpred      1674 1025 840.99 597.3 1044  912 915.34 932 1156  956
## V1          1969  682 849.00 523.0 1216  754 923.00 880 1030  508
## CV residual  295 -343   8.01 -74.3  172 -158   7.66 -52 -126 -448
##
## Sum of squares = 484414    Mean square = 48441    n = 10
##
## fold 3
## Observations in test set: 10
##                5    8      9   11  15   23   37  39   43   47
## Predicted   1028 1155 855.02 1283 653  736 1207 640 1112 1116
## cvpred      1026 1114 861.14 1178 667  653 1340 611 1172 1189
## V1          1234 1555 856.00 1674 798 1216  831 826  823  849
## CV residual  208  441  -5.14  496 131  563 -509 215 -349 -340
##
## Sum of squares = 1360587    Mean square = 136059    n = 10
##
## fold 4
## Observations in test set: 9
##                 7   13    14     20    24   27   30   35    45
## Predicted   873.8  606 616.7 1197.5 928.1  548  827  882 454.5
## cvpred      877.7  661 638.6 1201.9 910.3  605  837  949 502.4
## V1          963.0  511 664.0 1225.0 968.0  342  696  653 455.0
## CV residual  85.3 -150  25.4   23.1  57.7 -263 -141 -296 -47.4
##
## Sum of squares = 213008    Mean square = 23668    n = 9
##
## fold 5
## Observations in test set: 9
##                2   10    16    21   26   29   31   33  42
## Predicted   1231  886 946.4 778.1 1875 1420  715  731 274
```

```
## cvpred      1198  830 960.7 825.8 1814 1544  689  691 221
## V1          1635  705 946.0 742.0 1993 1043  373 1072 542
## CV residual  437 -125 -14.7 -83.8  179 -501 -316  381 321
##
## Sum of squares = 845402    Mean square = 93934    n = 9
##
## Overall (Sum over all 9 folds)
##    ms
## 70016
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## PC1        1 1177568 1177568   21.40 4.0e-05 ***
## PC2        1  633037  633037   11.51  0.0016 **
## PC3        1   58541   58541    1.06  0.3086
## PC4        1  257832  257832    4.69  0.0366 *
## PC5        1 2312556 2312556   42.03 1.1e-07 ***
## PC6        1   92261   92261    1.68  0.2029
## PC7        1  203535  203535    3.70  0.0617 .
## Residuals 39 2145598   55015
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##               1   3  17    18   19    22   36  38     40
## Predicted   629 475 395 948.3 1074 604.8 1085 641 1121.8
## cvpred      591 460 366 953.5 1108 538.3 1047 654 1139.9
## V1          791 578 539 929.0  750 439.0 1272 566 1151.0
## CV residual 200 118 173 -24.5 -358 -99.3  225 -88   11.1
##
## Sum of squares = 281103    Mean square = 31234    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6  12     25   28    32    34    41     44   46
## Predicted   1732  970 762 472.54 1073 798.7 888.3 834.7 1113.7  983
## cvpred      1714 1069 747 529.07 1049 777.7 943.7 819.8 1076.8 1087
## V1          1969  682 849 523.00 1216 754.0 923.0 880.0 1030.0  508
```

```
## CV residual  255 -387 102  -6.07  167 -23.7 -20.7  60.2  -46.8 -579
##
## Sum of squares = 594267    Mean square = 59427    n = 10
##
## fold 3
## Observations in test set: 10
##                5    8      9   11    15   23   37  39   43   47
## Predicted   1036 1262 806.08 1262 775.0  852 1167 753 1117 1095
## cvpred      1028 1144 854.81 1174 701.4  685 1332 644 1176 1182
## V1          1234 1555 856.00 1674 798.0 1216  831 826  823  849
## CV residual  206  411   1.19  500  96.6  531 -501 182 -353 -333
##
## Sum of squares = 1272182    Mean square = 127218    n = 10
##
## fold 4
## Observations in test set: 9
##                 7    13    14     20    24   27   30   35    45
## Predicted   909.9 547.6 606.5 1150.4 933.0  524  814  830 481.8
## cvpred      917.3 588.3 621.7 1144.8 910.8  582  832  877 519.1
## V1          963.0 511.0 664.0 1225.0 968.0  342  696  653 455.0
## CV residual  45.7 -77.3  42.3   80.2  57.2 -240 -136 -224 -64.1
##
## Sum of squares = 150126    Mean square = 16681    n = 9
##
## fold 5
## Observations in test set: 9
##                2   10   16    21   26   29   31   33  42
## Predicted   1254  887  977 757.1 1862 1449  798  821 208
## cvpred      1240  837 1013 807.2 1816 1582  806  806 130
## V1          1635  705  946 742.0 1993 1043  373 1072 542
## CV residual  395 -132  -67 -65.2  177 -539 -433  266 412
##
## Sum of squares = 932064    Mean square = 103563    n = 9
##
## Overall (Sum over all 9 folds)
##    ms
## 68718
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value  Pr(>F)
```

```
## PC1        1 1177568 1177568   20.97 4.9e-05 ***
## PC2        1  633037  633037   11.27  0.0018 **
## PC3        1   58541   58541    1.04  0.3137
## PC4        1  257832  257832    4.59  0.0386 *
## PC5        1 2312556 2312556   41.18 1.5e-07 ***
## PC6        1   92261   92261    1.64  0.2077
## PC7        1  203535  203535    3.62  0.0645 .
## PC8        1   11661   11661    0.21  0.6512
## Residuals 38 2133937   56156
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##                1   3  17    18   19   22   36    38     40
## Predicted    629 485 367 956.4 1074  612 1096 639.7 1101.5
## cvpred       595 472 329 972.9 1109  558 1070 642.2 1108.9
## V1           791 578 539 929.0  750  439 1272 566.0 1151.0
## CV residual 196 106 210 -43.9 -359 -119  202 -76.2   42.1
##
## Sum of squares = 286849    Mean square = 31872    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6    12    25   28     32    34  41      44   46
## Predicted   1688  986 770.9 483.7 1054 796.73 891.6 821 1111.00  988
## cvpred      1461 1153 782.1 603.8  943 758.09 962.7 756 1039.38 1117
## V1          1969  682 849.0 523.0 1216 754.00 923.0 880 1030.00  508
## CV residual  508 -471  66.9 -80.8  273  -4.09 -39.7 124   -9.38 -609
##
## Sum of squares = 954140    Mean square = 95414    n = 10
##
## fold 3
## Observations in test set: 10
##                5    8        9   11    15   23   37  39   43   47
## Predicted   1021 1282 804.5481 1279 771.0  863 1172 755 1100 1105
## cvpred      1040 1126 856.0483 1158 702.6  674 1329 641 1190 1174
## V1          1234 1555 856.0000 1674 798.0 1216  831 826  823  849
## CV residual  194  429  -0.0483  516  95.4  542 -498 185 -367 -325
##
```

```
## Sum of squares = 1314007    Mean square = 131401    n = 10
##
## fold 4
## Observations in test set: 9
##                 7    13    14     20    24   27   30   35  45
## Predicted   940.2 569.3 605.2 1157.3 927.0  523  788  822 493
## cvpred      941.9 605.6 621.4 1153.8 909.7  583  811  876 532
## V1          963.0 511.0 664.0 1225.0 968.0  342  696  653 455
## CV residual  21.1 -94.6  42.6   71.2  58.3 -241 -115 -223 -77
##
## Sum of squares = 146408    Mean square = 16268    n = 9
##
## fold 5
## Observations in test set: 9
##                2   10     16    21   26   29   31   33  42
## Predicted   1258  878  995.6 779.3 1890 1425  783  822 181
## cvpred      1236  838 1007.2 800.7 1802 1590  809  806 142
## V1          1635  705  946.0 742.0 1993 1043  373 1072 542
## CV residual  399 -133  -61.2 -58.7  191 -547 -436  266 400
##
## Sum of squares = 941607    Mean square = 104623    n = 9
##
## Overall (Sum over all 9 folds)
##    ms
## 77511
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## PC1        1 1177568 1177568   20.56 5.9e-05 ***
## PC2        1  633037  633037   11.05   0.002 **
## PC3        1   58541   58541    1.02   0.319
## PC4        1  257832  257832    4.50   0.041 *
## PC5        1 2312556 2312556   40.38 2.1e-07 ***
## PC6        1   92261   92261    1.61   0.212
## PC7        1  203535  203535    3.55   0.067 .
## PC8        1   11661   11661    0.20   0.654
## PC9        1   14950   14950    0.26   0.612
## Residuals 37 2118988   57270
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##
## fold 1
## Observations in test set: 9
##                1   3  17   18   19   22   36    38     40
## Predicted   620 483 354  972 1054  604 1072 627.0 1142.8
## cvpred      588 477 307 1000 1070  545 1003 638.2 1175.6
## V1          791 578 539  929  750  439 1272 566.0 1151.0
## CV residual 203 101 232  -71 -320 -106  269 -72.2  -24.6
##
## Sum of squares = 302263    Mean square = 33585    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6    12    25   28    32    34  41     44   46
## Predicted   1702  970 774.8 484.3 1048 803.0 895.4 841 1130.2  964
## cvpred      1426 1181 772.9 608.6  940 743.3 960.6 720 1005.2 1158
## V1          1969  682 849.0 523.0 1216 754.0 923.0 880 1030.0  508
## CV residual  543 -499  76.1 -85.6  276  10.7 -37.6 160   24.8 -650
##
## Sum of squares = 1083268    Mean square = 108327    n = 10
##
## fold 3
## Observations in test set: 10
##                5    8     9   11    15   23   37  39   43   47
## Predicted   1012 1297 795.9 1281 776.5  823 1138 754 1129 1111
## cvpred      1022 1155 821.3 1149 707.6  562 1217 626 1254 1192
## V1          1234 1555 856.0 1674 798.0 1216  831 826  823  849
## CV residual  212  400  34.7  525  90.4  654 -386 200 -431 -343
##
## Sum of squares = 1410746    Mean square = 141075    n = 10
##
## fold 4
## Observations in test set: 9
##                  7   13    14     20    24   27   30   35   45
## Predicted    983.4  584 611.1 1144.6 931.1  524  788  821  523
## cvpred      1043.2  637 632.8 1143.2 927.3  597  803  886  617
## V1           963.0  511 664.0 1225.0 968.0  342  696  653  455
## CV residual  -80.2 -126  31.2   81.8  40.7 -255 -107 -233 -162
##
## Sum of squares = 189062    Mean square = 21007    n = 9
```

```
##
## fold 5
## Observations in test set: 9
##                2   10    16    21   26   29   31   33  42
## Predicted   1242  883 976.0 792.1 1881 1416  766  851 163
## cvpred      1218  841 983.8 815.7 1785 1580  787  835 120
## V1          1635  705 946.0 742.0 1993 1043  373 1072 542
## CV residual  417 -136 -37.8 -73.7  208 -537 -414  237 422
##
## Sum of squares = 937526    Mean square = 104170    n = 9
##
## Overall (Sum over all 9 folds)
##    ms
## 83465
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## PC1        1 1177568 1177568   20.29 6.8e-05 ***
## PC2        1  633037  633037   10.90  0.0022 **
## PC3        1   58541   58541    1.01  0.3220
## PC4        1  257832  257832    4.44  0.0421 *
## PC5        1 2312556 2312556   39.84 2.7e-07 ***
## PC6        1   92261   92261    1.59  0.2155
## PC7        1  203535  203535    3.51  0.0693 .
## PC8        1   11661   11661    0.20  0.6567
## PC9        1   14950   14950    0.26  0.6149
## PC10       1   29162   29162    0.50  0.4830
## Residuals 36 2089825   58051
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##                1   3  17  18   19   22   36    38     40
## Predicted   646 479 351 911 1087  624 1031 619.5 1148.8
## cvpred      623 468 277 780 1129  557  893 657.9 1217.5
## V1          791 578 539 929  750  439 1272 566.0 1151.0
## CV residual 168 110 262 149 -379 -118  379 -91.9  -66.5
##
```

```
## Sum of squares = 445720    Mean square = 49524    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6    12    25   28     32    34  41     44   46
## Predicted   1715  975 767.3 478.2 1052 807.15 913.4 865 1143.8  937
## cvpred      1429 1181 772.5 607.2  941 744.03 962.5 724 1007.9 1154
## V1          1969  682 849.0 523.0 1216 754.00 923.0 880 1030.0  508
## CV residual  540 -499  76.5 -84.2  275   9.97 -39.5 156   22.1 -646
##
## Sum of squares = 1073200    Mean square = 107320    n = 10
##
## fold 3
## Observations in test set: 10
##                5    8     9   11  15   23   37  39   43   47
## Predicted   1039 1332 799.6 1250 719  825 1165 756 1104 1111
## cvpred      1066 1199 832.3 1108 616  569 1274 620 1214 1196
## V1          1234 1555 856.0 1674 798 1216  831 826  823  849
## CV residual  168  356  23.7  566 182  647 -443 206 -391 -347
##
## Sum of squares = 1438676    Mean square = 143868    n = 10
##
## fold 4
## Observations in test set: 9
##                7   13    14     20     24   27    30   35   45
## Predicted   1028  564 626.9 1157.5 963.26  521 763.8  833  493
## cvpred      1104  617 647.3 1168.6 972.09  593 767.2  903  601
## V1           963  511 664.0 1225.0 968.00  342 696.0  653  455
## CV residual -141 -106  16.7   56.4  -4.09 -251 -71.2 -250 -146
##
## Sum of squares = 186254    Mean square = 20695    n = 9
##
## fold 5
## Observations in test set: 9
##                2   10     16    21   26   29   31   33  42
## Predicted   1254  894 1022.7 781.3 1840 1397  726  858 163
## cvpred      1218  842  997.8 811.7 1766 1575  773  838 122
## V1          1635  705  946.0 742.0 1993 1043  373 1072 542
## CV residual  417 -137  -51.8 -69.7  227 -532 -400  234 420
##
## Sum of squares = 925990    Mean square = 102888    n = 9
```

```
##
## Overall (Sum over all 9 folds)
##    ms
## 86592
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## PC1        1 1177568 1177568   19.79 8.4e-05 ***
## PC2        1  633037  633037   10.64  0.0025 **
## PC3        1   58541   58541    0.98  0.3280
## PC4        1  257832  257832    4.33  0.0447 *
## PC5        1 2312556 2312556   38.87 3.8e-07 ***
## PC6        1   92261   92261    1.55  0.2213
## PC7        1  203535  203535    3.42  0.0728 .
## PC8        1   11661   11661    0.20  0.6607
## PC9        1   14950   14950    0.25  0.6193
## PC10       1   29162   29162    0.49  0.4885
## PC11       1    7564    7564    0.13  0.7236
## Residuals 35 2082261   59493
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##               1   3  17  18   19    22   36    38     40
## Predicted   643 464 355 915 1099 586.0 1037 603.6 1154.8
## cvpred      591 427 273 764 1161 469.9  920 632.5 1224.4
## V1          791 578 539 929  750 439.0 1272 566.0 1151.0
## CV residual 200 151 266 165 -411 -30.9  352 -66.5  -73.4
##
## Sum of squares = 464975    Mean square = 51664    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6    12    25   28    32    34  41     44   46
## Predicted   1718  976 776.4 478.6 1040 825.7 901.0 871 1139.6  940
## cvpred      1438 1180 791.7 615.8  919 781.2 942.6 732 1001.8 1160
## V1          1969  682 849.0 523.0 1216 754.0 923.0 880 1030.0  508
## CV residual  531 -498  57.3 -92.8  297 -27.2 -19.6 148   28.2 -652
```

```
##
## Sum of squares = 1078958    Mean square = 107896    n = 10
##
## fold 3
## Observations in test set: 10
##                5    8     9   11    15   23   37    39   43   47
## Predicted   1029 1322 803.3 1245 743.2  825 1201 762.2 1100 1120
## cvpred      1001 1144 943.4 1088 851.4  638 1715 729.7 1234 1254
## V1          1234 1555 856.0 1674 798.0 1216  831 826.0  823  849
## CV residual  233  411 -87.4  586 -53.4  578 -884  96.3 -411 -405
##
## Sum of squares = 2034335    Mean square = 203434    n = 10
##
## fold 4
## Observations in test set: 9
##                7  13    14     20     24   27    30   35   45
## Predicted   1037 555 632.1 1180.9 954.41  514 764.3  836  501
## cvpred      1122 605 650.2 1205.4 972.15  582 761.4  916  622
## V1           963 511 664.0 1225.0 968.00  342 696.0  653  455
## CV residual -159 -94  13.8   19.6  -4.15 -240 -65.4 -263 -167
##
## Sum of squares = 193784    Mean square = 21532    n = 9
##
## fold 5
## Observations in test set: 9
##                2   10     16   21   26   29   31   33    42
## Predicted   1253  894 1016.5  774 1835 1371  709  862 177.5
## cvpred      1240  838 1027.9  849 1838 1705  859  813  45.2
## V1          1635  705  946.0  742 1993 1043  373 1072 542.0
## CV residual  395 -133  -81.9 -107  155 -662 -486  259 496.8
##
## Sum of squares = 1204103    Mean square = 133789    n = 9
##
## Overall (Sum over all 9 folds)
##     ms
## 105876
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## PC1        1 1177568 1177568   25.22 1.6e-05 ***
```

```
## PC2        1  633037  633037   13.56  0.0008 ***
## PC3        1   58541   58541    1.25  0.2707
## PC4        1  257832  257832    5.52  0.0247 *
## PC5        1 2312556 2312556   49.52 4.0e-08 ***
## PC6        1   92261   92261    1.98  0.1689
## PC7        1  203535  203535    4.36  0.0444 *
## PC8        1   11661   11661    0.25  0.6205
## PC9        1   14950   14950    0.32  0.5752
## PC10       1   29162   29162    0.62  0.4349
## PC11       1    7564    7564    0.16  0.6899
## PC12       1  494595  494595   10.59  0.0026 **
## Residuals 34 1587667   46696
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##               1   3    17  18   19    22   36  38     40
## Predicted   707 451 453.3 868 1241 629.9 1174 530 1137.8
## cvpred      666 316 475.8 678 1532 504.2 1435 397 1067.5
## V1          791 578 539.0 929  750 439.0 1272 566 1151.0
## CV residual 125 262  63.2 251 -782 -65.2 -163 169   83.5
##
## Sum of squares = 829767    Mean square = 92196    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6  12   25   28    32    34  41     44   46
## Predicted   1769  771 699  588 1110 726.6 976.2 795 1189.5  776
## cvpred      1526  961 710  658 1008 689.9 991.2 693 1068.5  974
## V1          1969  682 849  523 1216 754.0 923.0 880 1030.0  508
## CV residual  443 -279 139 -135  208  64.1 -68.2 187  -38.5 -466
##
## Sum of squares = 617992    Mean square = 61799    n = 10
##
## fold 3
## Observations in test set: 10
##                5    8     9   11   15   23   37  39   43   47
## Predicted   1150 1464 718.6 1148  906  850 1109 788 1024 1124
## cvpred      1121 1328 819.3  973 1036  685 1496 782 1151 1261
```

```
## V1          1234 1555 856.0 1674  798 1216  831 826  823  849
## CV residual  113  227  36.7  701 -238  531 -665  44 -328 -412
##
## Sum of squares = 1616893    Mean square = 161689    n = 10
##
## fold 4
## Observations in test set: 9
##                7   13  14     20  24    27    30   35   45
## Predicted    848  728 686 1197.8 894 364.2 657.9  739  614
## cvpred       733  912 720 1218.6 846 345.9 620.9  791  740
## V1           963  511 664 1225.0 968 342.0 696.0  653  455
## CV residual 230 -401 -56    6.4 122  -3.9  75.1 -138 -285
##
## Sum of squares = 337813    Mean square = 37535    n = 9
##
## fold 5
## Observations in test set: 9
##                 2    10    16   21     26   29   31   33  42
## Predicted    1356 832.8 969.1  784 1921.8 1382  479  937 275
## cvpred       1304 793.3 986.7  843 1903.2 1698  630  874 110
## V1           1635 705.0 946.0  742 1993.0 1043  373 1072 542
## CV residual  331 -88.3 -40.7 -101   89.8 -655 -257  198 432
##
## Sum of squares = 857884    Mean square = 95320    n = 9
##
## Overall (Sum over all 9 folds)
##    ms
## 90646
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## PC1        1 1177568 1177568   24.81 2.0e-05 ***
## PC2        1  633037  633037   13.34 0.00089 ***
## PC3        1   58541   58541    1.23 0.27478
## PC4        1  257832  257832    5.43 0.02603 *
## PC5        1 2312556 2312556   48.72 5.6e-08 ***
## PC6        1   92261   92261    1.94 0.17257
## PC7        1  203535  203535    4.29 0.04628 *
## PC8        1   11661   11661    0.25 0.62342
## PC9        1   14950   14950    0.31 0.57844
```

```
## PC10       1   29162   29162    0.61 0.43873
## PC11       1    7564    7564    0.16 0.69231
## PC12       1  494595  494595   10.42 0.00282 **
## PC13       1   21336   21336    0.45 0.50723
## Residuals 33 1566331   47465
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##               1   3  17  18   19   22   36    38   40
## Predicted   698 455 443 829 1250  656 1178 543.6 1127
## cvpred      637 330 421 420 1561  599 1471 469.3 1031
## V1          791 578 539 929  750  439 1272 566.0 1151
## CV residual 154 248 118 509 -811 -160 -199  96.7  120
##
## Sum of squares = 1104867    Mean square = 122763    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6  12   25   28    32    34  41     44   46
## Predicted   1777  760 713  597 1098 739.3 989.8 790 1190.0  762
## cvpred      1535  950 720  667 1002 702.4 998.6 692 1071.5  962
## V1          1969  682 849  523 1216 754.0 923.0 880 1030.0  508
## CV residual  434 -268 129 -144  214  51.6 -75.6 188  -41.5 -454
##
## Sum of squares = 595285    Mean square = 59529    n = 10
##
## fold 3
## Observations in test set: 10
##                5    8     9   11   15   23   37    39   43   47
## Predicted   1091 1482 698.8 1167  923  846 1077 790.2 1059 1144
## cvpred      1014 1366 761.7 1013 1087  661 1361 785.9 1217 1291
## V1          1234 1555 856.0 1674  798 1216  831 826.0  823  849
## CV residual  220  189  94.3  661 -289  555 -530  40.1 -394 -442
##
## Sum of squares = 1554455    Mean square = 155446    n = 10
##
## fold 4
## Observations in test set: 9
```

```
##               7   13  14     20  24     27  30   35   45
## Predicted   862  717 733 1212.1 875 369.64 690  743  572
## cvpred      737  905 740 1217.8 835 351.99 639  788  711
## V1          963  511 664 1225.0 968 342.00 696  653  455
## CV residual 226 -394 -76    7.2 133  -9.99  57 -135 -256
##
## Sum of squares = 316718    Mean square = 35191    n = 9
##
## fold 5
## Observations in test set: 9
##                2   10    16   21   26   29   31   33    42
## Predicted   1381  812 962.3  750 1933 1363  480  911 298.5
## cvpred      1285  807 990.3  874 1889 1719  630  896  82.1
## V1          1635  705 946.0  742 1993 1043  373 1072 542.0
## CV residual  350 -102 -44.3 -132  104 -676 -257  176 459.9
##
## Sum of squares = 928410    Mean square = 103157    n = 9
##
## Overall (Sum over all 9 folds)
##    ms
## 95739
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## PC1        1 1177568 1177568   26.22 1.4e-05 ***
## PC2        1  633037  633037   14.10 0.00069 ***
## PC3        1   58541   58541    1.30 0.26204
## PC4        1  257832  257832    5.74 0.02259 *
## PC5        1 2312556 2312556   51.49 3.8e-08 ***
## PC6        1   92261   92261    2.05 0.16147
## PC7        1  203535  203535    4.53 0.04106 *
## PC8        1   11661   11661    0.26 0.61386
## PC9        1   14950   14950    0.33 0.56801
## PC10       1   29162   29162    0.65 0.42629
## PC11       1    7564    7564    0.17 0.68425
## PC12       1  494595  494595   11.01 0.00226 **
## PC13       1   21336   21336    0.48 0.49562
## PC14       1  129212  129212    2.88 0.09955 .
## Residuals 32 1437119   44910
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##                 1   3  17  18   19   22     36    38   40
## Predicted   769.6 348 389 851 1210  657 1093.8 540.7 1088
## cvpred      743.5 248 341 525 1431  634 1227.9 490.3 1022
## V1          791.0 578 539 929  750  439 1272.0 566.0 1151
## CV residual  47.5 330 198 404 -681 -195   44.1  75.7  129
##
## Sum of squares = 839315    Mean square = 93257    n = 9
##
## fold 2
## Observations in test set: 10
##                4    6  12    25     28    32   34  41   44   46
## Predicted   1810  803 734 574.1 1236.4 758.0 1023 765 1221  774
## cvpred      1631  997 738 622.6 1196.7 717.9 1055 697 1141  968
## V1          1969  682 849 523.0 1216.0 754.0  923 880 1030  508
## CV residual  338 -315 111 -99.6   19.3  36.1 -132 183 -111 -460
##
## Sum of squares = 512847    Mean square = 51285    n = 10
##
## fold 3
## Observations in test set: 10
##                5    8     9   11   15   23   37    39   43   47
## Predicted   1120 1346 704.6 1223  932  966 1025 775.4 1117 1095
## cvpred      1018 1316 762.2 1038 1090  700 1336 781.4 1235 1269
## V1          1234 1555 856.0 1674  798 1216  831 826.0  823  849
## CV residual  216  239  93.8  636 -292  516 -505  44.6 -412 -420
##
## Sum of squares = 1472664    Mean square = 147266    n = 10
##
## fold 4
## Observations in test set: 9
##                7   13   14      20  24    27    30    35   45
## Predicted   893  723  788 1221.12 848 328.7 684.0 674.8  621
## cvpred      803  932  815 1228.73 799 290.4 619.5 686.1  805
## V1          963  511  664 1225.00 968 342.0 696.0 653.0  455
## CV residual 160 -421 -151   -3.73 169  51.6  76.5 -33.1 -350
##
```

```
## Sum of squares = 385805    Mean square = 42867    n = 9
##
## fold 5
## Observations in test set: 9
##                2    10     16   21   26   29   31   33    42
## Predicted   1436 751.0  992.6  771 1904 1312  436  906 301.5
## cvpred      1319 761.4 1015.1  878 1856 1662  573  896  93.6
## V1          1635 705.0  946.0  742 1993 1043  373 1072 542.0
## CV residual  316 -56.4  -69.1 -136  137 -619 -200  176 448.4
##
## Sum of squares = 8e+05    Mean square = 88866    n = 9
##
## Overall (Sum over all 9 folds)
##    ms
## 85328
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## PC1        1 1177568 1177568   26.94 1.2e-05 ***
## PC2        1  633037  633037   14.48 0.00062 ***
## PC3        1   58541   58541    1.34 0.25599
## PC4        1  257832  257832    5.90 0.02114 *
## PC5        1 2312556 2312556   52.91 3.5e-08 ***
## PC6        1   92261   92261    2.11 0.15631
## PC7        1  203535  203535    4.66 0.03879 *
## PC8        1   11661   11661    0.27 0.60916
## PC9        1   14950   14950    0.34 0.56289
## PC10       1   29162   29162    0.67 0.42026
## PC11       1    7564    7564    0.17 0.68027
## PC12       1  494595  494595   11.32 0.00206 **
## PC13       1   21336   21336    0.49 0.48996
## PC14       1  129212  129212    2.96 0.09552 .
## PC15       1   82173   82173    1.88 0.18017
## Residuals 31 1354946   43708
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
```

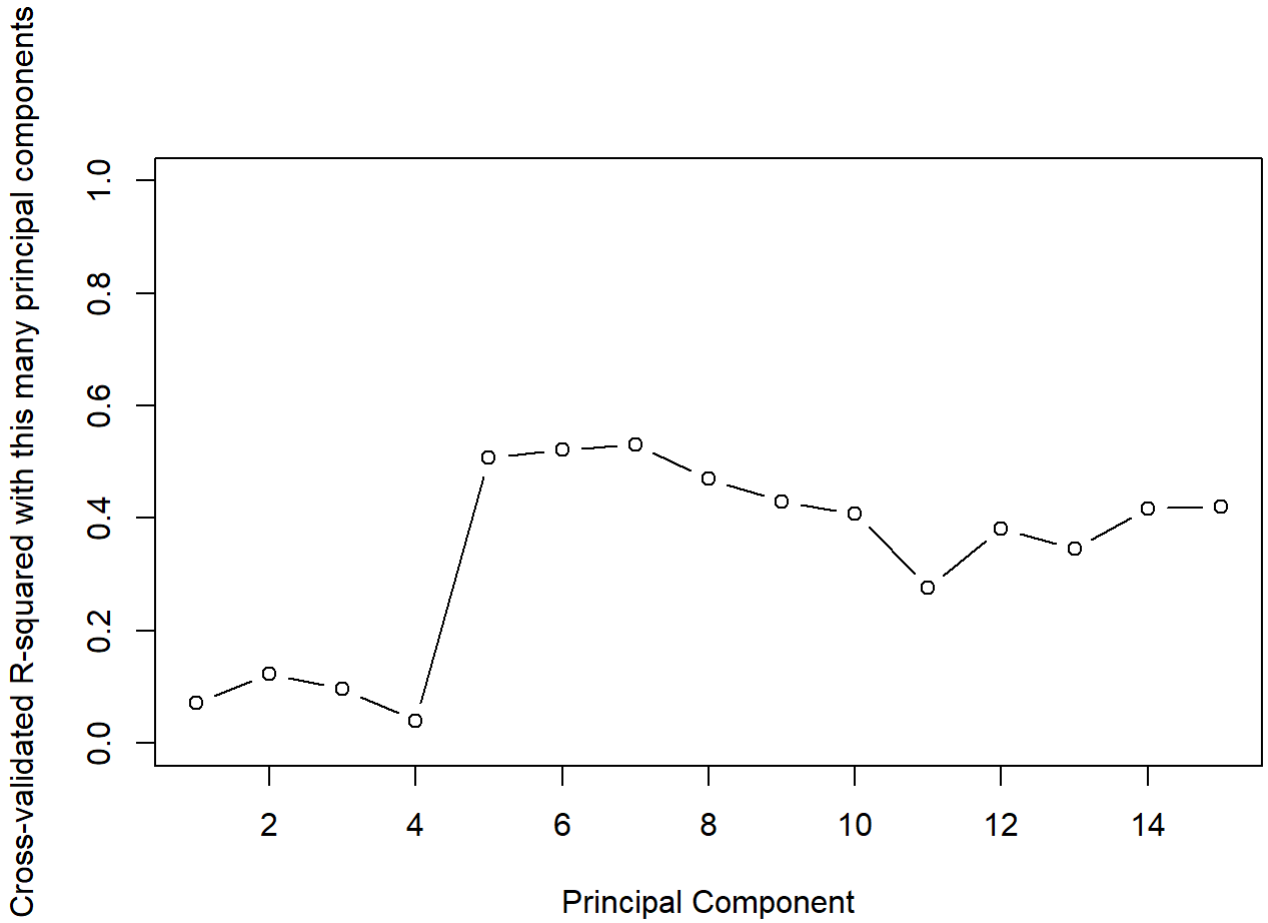
```
##                 1   3  17  18   19   22     36    38     40
## Predicted   755.0 322 393 844 1146  657 1137.6 562.7 1131.5
## cvpred      719.5 227 334 497 1385  620 1261.6 509.1 1057.1
## V1          791.0 578 539 929  750  439 1272.0 566.0 1151.0
## CV residual  71.5 351 205 432 -635 -181   10.4  56.9   93.9
##
## Sum of squares = 804291    Mean square = 89366    n = 9
##
## fold 2
## Observations in test set: 10
##                 4    6    12   25     28    32    34    41   44   46
## Predicted   1791  793 722.0  606 1258.5 807.8 971.5 823.7 1121  827
## cvpred      1543 1026 752.8  733 1170.1 836.6 934.6 786.7  919 1138
## V1          1969  682 849.0  523 1216.0 754.0 923.0 880.0 1030  508
## CV residual  426 -344  96.2 -210   45.9 -82.6 -11.6  93.3  111 -630
##
## Sum of squares = 779686    Mean square = 77969    n = 10
##
## fold 3
## Observations in test set: 10
##                 5    8   9   11   15   23   37    39   43   47
## Predicted   1167 1362 689 1161  903  958  971 839.3 1134  992
## cvpred      1092 1350 717  958 1040  690 1174 838.2 1247 1138
## V1          1234 1555 856 1674  798 1216  831 826.0  823  849
## CV residual  142  205 139  716 -242  526 -343 -12.2 -424 -289
##
## Sum of squares = 1310071    Mean square = 131007    n = 10
##
## fold 4
## Observations in test set: 9
##                 7   13   14     20  24  27    30   35   45
## Predicted   934.2  733  780 1227.8 869 279 702.7  738  617
## cvpred      898.5  929  797 1290.4 864 227 618.7  808  849
## V1          963.0  511  664 1225.0 968 342 696.0  653  455
## CV residual  64.5 -418 -133  -65.4 104 115  77.3 -155 -394
##
## Sum of squares = 410147    Mean square = 45572    n = 9
##
## fold 5
## Observations in test set: 9
##                 2    10     16   21     26   29   31   33  42
```

```
## Predicted   1474 736.5 1005.7  775 1977.4 1287  388  841 326
## cvpred      1380 743.3 1031.4  868 1975.1 1620  525  831 113
## V1          1635 705.0  946.0  742 1993.0 1043  373 1072 542
## CV residual  255 -38.3  -85.4 -126   17.9 -577 -152  241 429
##
## Sum of squares = 688401    Mean square = 76489    n = 9
##
## Overall (Sum over all 9 folds)
##    ms
## 84949
```

```
r2cross #results
```

```
##  [1] 0.0711 0.1228 0.0963 0.0392 0.5068 0.5218 0.5306 0.4706 0.4299 0.4085
## [11] 0.2768 0.3808 0.3461 0.4172 0.4198
```

```
plot(r2cross, xlab = "Principal Component", ylab = "Cross-validated R-squared with this many principal components",
     ylim = c(0,1), type = "b")
```




```
#5th principal component looks like a substantial jump so let's try running lm on this one.
pcc <- cbind(data[,16],pca$x[,5])
model <- lm(V1~.,data = as.data.frame(pcc))
summary(model)
```

```
##
## Call:
## lm(formula = V1 ~ ., data = as.data.frame(pcc))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -639.9 -200.5  -47.8  177.0  694.7
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    905.1       46.5   19.47   <2e-16 ***
## V2            -229.0       48.0   -4.77    2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 319 on 45 degrees of freedom
## Multiple R-squared:  0.336,  Adjusted R-squared:  0.321
## F-statistic: 22.8 on 1 and 45 DF,  p-value: 1.95e-05
```

```
# ## Multiple R-squared:  0.336,     Adjusted R-squared:  0.321

c <- cv.lm(as.data.frame(pcc),model,m=5, plotit = FALSE) # cross-validate
```

```
## Analysis of Variance Table
##
## Response: V1
##           Df  Sum Sq Mean Sq F value Pr(>F)
## V2         1 2312556 2312556    22.8  2e-05 ***
## Residuals 45 4568372  101519
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## fold 1
## Observations in test set: 9
##                     1        3     17      18     19       22      36        38
## V2             0.0553    0.541   1.34   0.411   1.36   -0.689    0.02    0.0282
## cvpred       906.4289  790.203 599.76 821.391 595.34 1084.478  914.86  912.9153
## V1           791.0000  578.000 539.00 929.000 750.00  439.000 1272.00  566.0000
## CV residual -115.4289 -212.203 -60.76 107.609 154.66 -645.478  357.14 -346.9153
##                   40
## V2            -0.646
## cvpred      1074.090
## V1          1151.000
## CV residual   76.910
##
## Sum of squares = 768004    Mean square = 85334    n = 9
##
## fold 2
## Observations in test set: 10
##                    4      6      12       25      28      32      34        41
## V2             -1.64   1.37   0.358    0.804   -1.02   0.331   0.723   0.00986
## cvpred       1213.89 625.53 823.582  736.743 1092.77 828.807 752.376 891.44499
## V1           1969.00 682.00 849.000  523.000 1216.00 754.000 923.000 880.00000
## CV residual  755.11  56.47  25.418 -213.743  123.23 -74.807 170.624 -11.44499
##                   44       46
## V2            -0.703    0.542
## cvpred      1030.388  787.650
## V1          1030.000  508.000
## CV residual   -0.388 -279.650
##
## Sum of squares = 747938    Mean square = 74794    n = 10
##
```

```
## fold 3
## Observations in test set: 10
##                    5        8        9       11        15      23      37
## V2          4.16e-02   -0.223   -0.324   -0.324   0.00297    0.69   -2.47
## cvpred      8.82e+02  958.914  988.480  988.419 893.16365  693.08 1613.33
## V1          1.23e+03 1555.000  856.000 1674.000 798.00000 1216.00  831.00
## CV residual 3.52e+02  596.086 -132.480  685.581 -95.16365  522.92 -782.33
##                   39       43      47
## V2            0.487   -0.867   -1.14
## cvpred      752.113 1146.660 1226.26
## V1          826.000  823.000  849.00
## CV residual  73.887 -323.660 -377.26
##
## Sum of squares = 2113937    Mean square = 211394    n = 10
##
## fold 4
## Observations in test set: 9
##                    7     13       14       20       24      27       30       35
## V2             0.719    0.6    0.742   -0.653   -0.744    1.83   -0.255    0.665
## cvpred       776.081  801.2  771.157 1066.646 1085.993  540.49  982.269  787.358
## V1           963.000  511.0  664.000 1225.000  968.000  342.00  696.000  653.000
## CV residual 186.919 -290.2 -107.157  158.354 -117.993 -198.49 -286.269 -134.358
##                   45
## V2              0.807
## cvpred        757.436
## V1            455.000
## CV residual -302.436
##
## Sum of squares = 400527    Mean square = 44503    n = 9
##
## fold 5
## Observations in test set: 9
##                    2       10      16        21     26      29       31       33
## V2             -1.17   -0.646  -0.387    0.0846   -2.9    0.31   -0.471    0.292
## cvpred       1128.94 1024.167 972.766  878.9358 1471.2  834.15  989.355  837.744
## V1           1635.00  705.000 946.000  742.0000 1993.0 1043.00  373.000 1072.000
## CV residual  506.06 -319.167 -26.766 -136.9358  521.8  208.85 -616.355  234.256
##                 42
## V2            2.12
## cvpred      475.25
## V1          542.00
```

```
## CV residual  66.75
##
## Sum of squares = 1132556    Mean square = 125840    n = 9
##
## Overall (Sum over all 9 folds)
##     ms
## 109850
```

```
1 - attr(c,"ms")*nrow(data)/sum((data$Crime - mean(data$Crime))^2) # calculate R-squared
```

```
## [1] 0.25
```