Homework 1

Sungil Kim

September 9, 2017

Problem 2

```
library("MASS")
Z = matrix(c(1, 5, 1, -3, 1, 2, 1, 4), nrow=4, ncol=2, byrow=T)
Y = matrix(c(2, 1, -1, 3), nrow=4, ncol=1, byrow=F)
M = matrix(c(20, 15, 0, 5, 25, 10, 0, 20, 5), nrow=3, ncol=3, byrow=T)
N = matrix(c(-20, 5, 10, 0, -10, 10, 5, 20, -5), nrow=3, ncol=3, byrow=T)
v = matrix(c(1, -1, 3), nrow=3, ncol=1, byrow=F)
w = matrix(c(2, 1, -1), nrow=3, ncol=1, byrow=F)
#dot product is a 1x1 matrix, a.k.a. a scalar
```

a)

```
t(v)%*%w
```

```
##      [,1]
## [1,]   -2
```

b)

```
(-3)*w
```

```
##      [,1]
## [1,]   -6
## [2,]   -3
## [3,]    3
```

c)

```
M%*%v
```

```
##      [,1]
## [1,]    5
## [2,]   10
## [3,]   -5
```

d)

```
M+N
```

```
##      [,1] [,2] [,3]
## [1,]    0   20   10
## [2,]    5   15   20
## [3,]    5   40    0
```

e)

```
M-N
```

```
##      [,1] [,2] [,3]
## [1,]   40   10  -10
## [2,]    5   35    0
## [3,]   -5    0   10
```

f)

```r
t(Z)
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    1    1    1
## [2,]    5   -3    2    4
```

g)

```r
t(Z)%*%Z
```

```
##      [,1] [,2]
## [1,]    4    8
## [2,]    8   54
```

h)

```r
ginv(t(Z)%*%Z)
```

```
##             [,1]        [,2]
## [1,]  0.35526316 -0.05263158
## [2,] -0.05263158  0.02631579
```

i)

```r
t(Z)%*%Y
```

```
##      [,1]
## [1,]    5
## [2,]   17
```

j)

```r
(ginv(t(Z)%*%Z))%*%(t(Z)%*%Y)
```

```
##           [,1]
## [1,] 0.8815789
## [2,] 0.1842105
```

k)

```r
det(t(Z)%*%Z)
```

```
## [1] 152
```

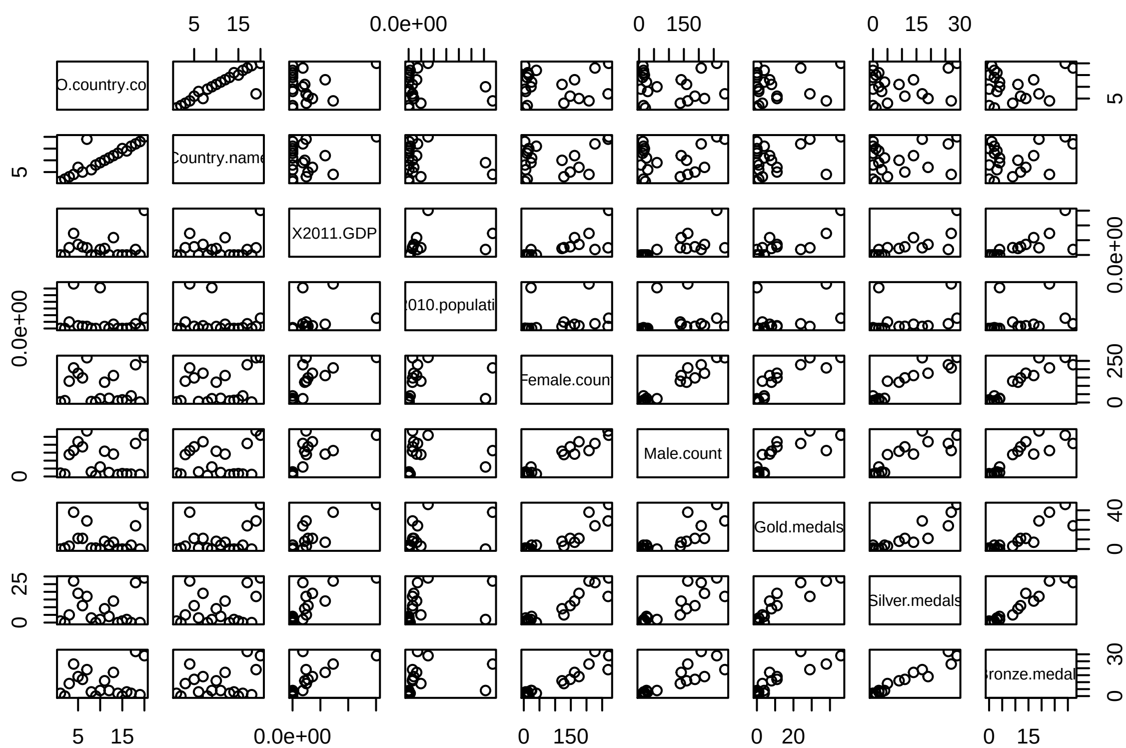
Problem 3

```r
#setting the working directory
setwd("C:/Users/sungi/Documents/CSC424/HW1")

library(corrplot)

#reading the data
olympic = read.table("01 olympics.csv", sep=",", header=T)

#looking at the data itself to first examine
head(olympic)
```

```
##   ISO.country.code Country.name   X2011.GDP X2010.population Female.count
```

```
## 1                USA           US 1.50940e+13          309349000          271
## 2                CHN        China 7.29810e+12         1338300000          208
## 3                JPN        Japan 5.86715e+12          127451000          162
## 4                DEU      Germany 3.57056e+12           81777000          176
## 5                FRA       France 2.77303e+12           64895000          148
## 6                BRA       Brazil 2.47665e+12          194946000          128
##   Male.count Gold.medals Silver.medals Bronze.medals
## 1        260          46            29            29
## 2        163          38            27            23
## 3        141           7            14            17
## 4        219          11            19            14
## 5        187          11            11            12
## 6        138           3             5             9
```

```r
plot(olympic)
```

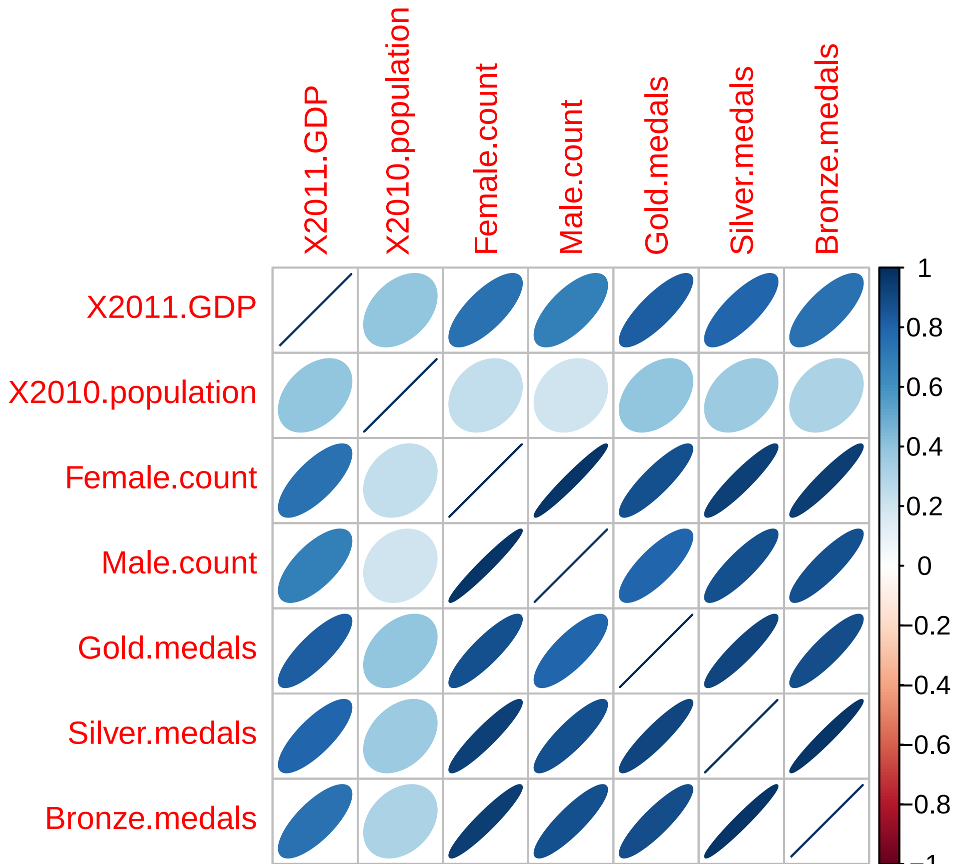



```r
#removing the non-numerical columns
olym = olympic[c(3:9)]
head(olym)
```

```
##     X2011.GDP X2010.population Female.count Male.count Gold.medals
## 1 1.50940e+13        309349000          271        260          46
## 2 7.29810e+12       1338300000          208        163          38
## 3 5.86715e+12        127451000          162        141           7
## 4 3.57056e+12         81777000          176        219          11
## 5 2.77303e+12         64895000          148        187          11
## 6 2.47665e+12        194946000          128        138           3
```

```
##   Silver.medals Bronze.medals
## 1            29            29
## 2            27            23
## 3            14            17
## 4            19            14
## 5            11            12
## 6             5             9
```

```
#correlation
cor.olym = cor(olym)
corrplot(cor.olym, method="ellipse")
```



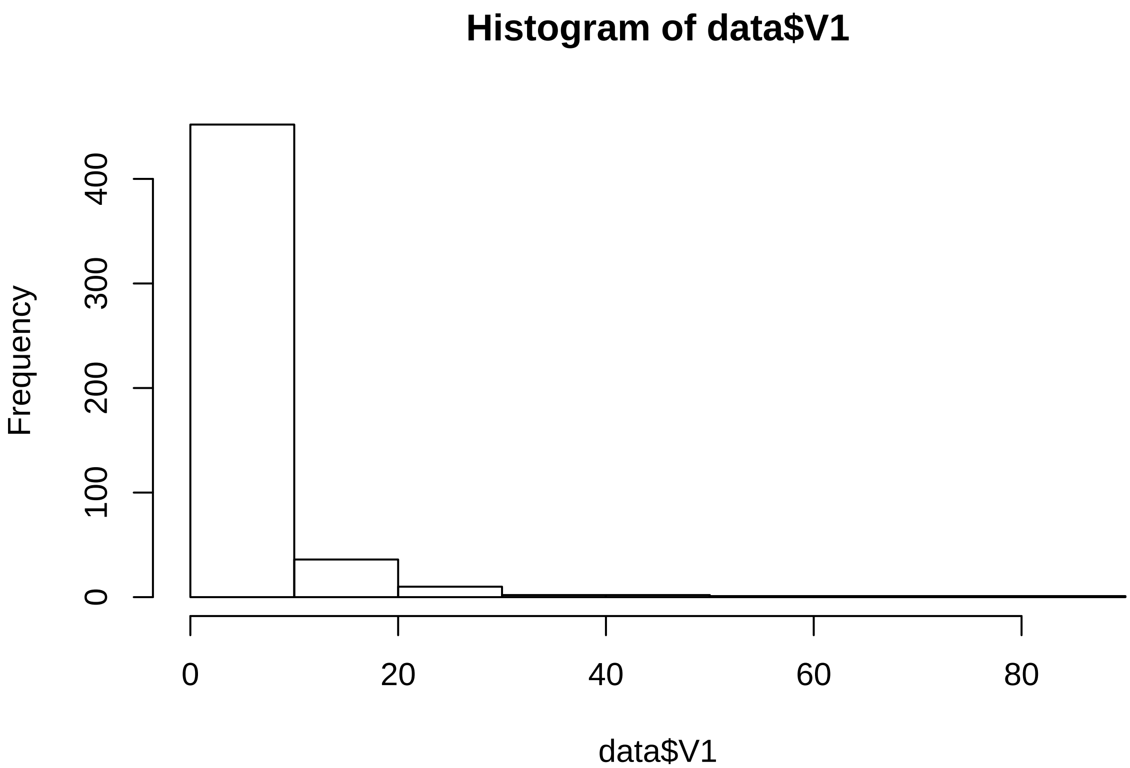
Looking at the correlation plots above, I have noticed that there is a fairly significant relationship between the country's national GDP and the overall medal counts. Although it is obviously not a direct causation, it is interesting to see how the country's GDP (or the wealth of the people in the country) can affect how well their players do in the olympics. It makes sense that if the country is wealthy (high GDP), then the players will be funded more in many different areas such as nutrition, facilities, coaches, etc. In order to further prove this, multiple regression should be used to analyze more. This is just a hypothesis, but if people wanted their country to do well in the olympics, maybe it is time for them to start working to better themselves and their country overall.

Problem 4

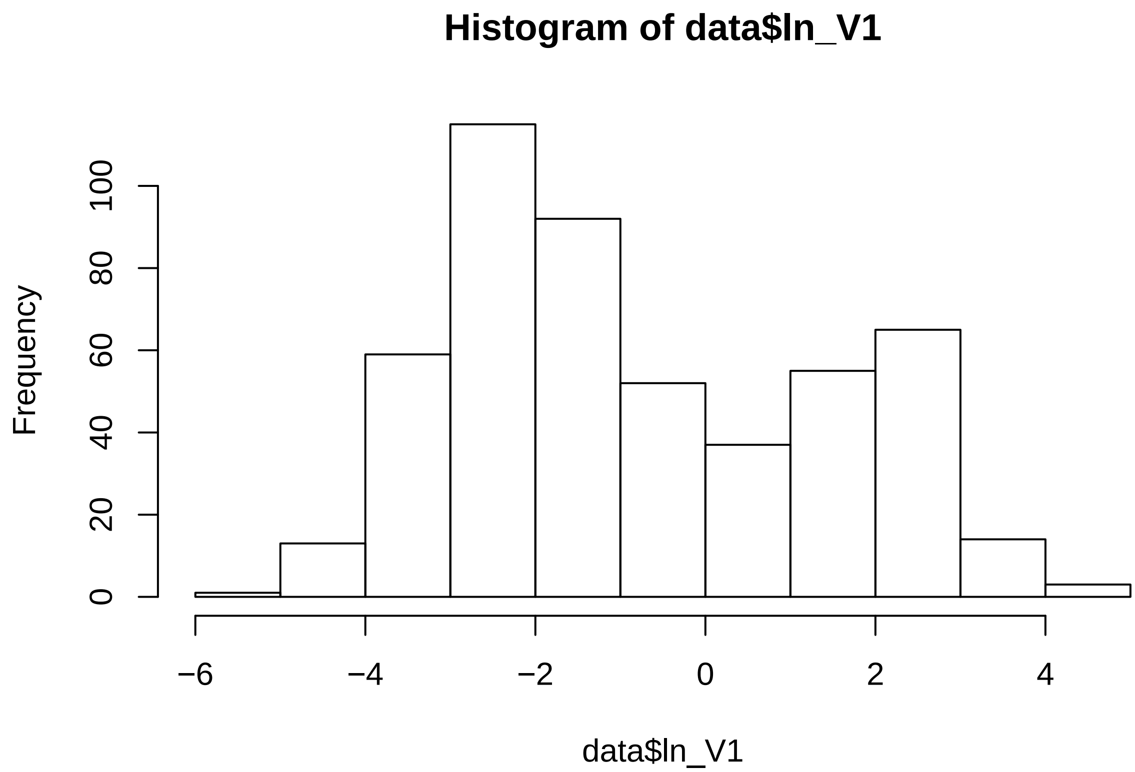
a)

```
library(car)
library(rcompanion)
```

```r
#reading the data
data = read.table("01 housing.data")

#examining variables for the ideal transformations
hist(data$V1)
```




```r
data$ln_V1 = log(data$V1)
hist(data$ln_V1)
```




```
#the histogram examination was performed for all variables; codes are removed for simplicity however.
data$sq_V2 = sqrt(data$V2) #sqrt was used since log transformation gave -inf values
data$ln_V3 = log(data$V3)
data$ln_V5 = log(data$V5)
data$sq_V7 = (data$V7)^2
data$ln_V8 = log(data$V8)
data$ln_V9 = log(data$V9)
data$ln_V10 = log(data$V10)
data$sq_V11 = (data$V11)^2
data$tk_V12 = transformTukey(data$V12,plotit=FALSE)
```

```
##
##     lambda      W Shapiro.p.value
## 800  9.975 0.7817       1.909e-25
##
## if (lambda >  0){TRANS = x ^ lambda}
## if (lambda == 0){TRANS = log(x)}
## if (lambda <  0){TRANS = -1 * x ^ lambda}
```

```
data$ln_V13 = log(data$V13)
data$ln_V14 = log(data$V14)

#modeling after the transformation
M0 = lm(ln_V1 ~ sq_V2 + ln_V3 + as.factor(V4) + ln_V5 + V6 + sq_V7 + ln_V8 + ln_V9 + ln_V10 + sq_V11 +
summary(M0)
```

```
##
```

```
## Call:
## lm(formula = ln_V1 ~ sq_V2 + ln_V3 + as.factor(V4) + ln_V5 +
##     V6 + sq_V7 + ln_V8 + ln_V9 + ln_V10 + sq_V11 + tk_V12 + ln_V13 +
##     ln_V14, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.34484 -0.53589 -0.02572  0.52271  2.31534
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.447e+00  1.467e+00  -1.669 0.095839 .
## sq_V2          -3.863e-02  1.819e-02  -2.124 0.034190 *
## ln_V3           1.756e-01  8.474e-02   2.073 0.038727 *
## as.factor(V4)1 -1.930e-03  1.414e-01  -0.014 0.989112
## ln_V5           1.809e+00  4.085e-01   4.429 1.17e-05 ***
## V6              3.833e-02  7.087e-02   0.541 0.588818
## sq_V7           6.802e-05  1.889e-05   3.602 0.000348 ***
## ln_V8          -3.374e-01  1.445e-01  -2.334 0.019975 *
## ln_V9           1.137e+00  7.166e-02  15.862  < 2e-16 ***
## ln_V10          4.222e-01  1.825e-01   2.313 0.021144 *
## sq_V11         -3.558e-04  6.110e-04  -0.582 0.560623
## tk_V12         -7.111e-27  1.398e-27  -5.085 5.22e-07 ***
## ln_V13         -9.030e-02  1.239e-01  -0.729 0.466595
## ln_V14         -5.345e-01  1.692e-01  -3.158 0.001685 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7702 on 492 degrees of freedom
## Multiple R-squared:  0.8764, Adjusted R-squared:  0.8731
## F-statistic: 268.3 on 13 and 492 DF,  p-value: < 2.2e-16
```

```
#checking vif statistics and correlation for multicollinearity
vif(M0)#all good
```

```
##         sq_V2         ln_V3 as.factor(V4)         ln_V5            V6
##      2.433882      3.690606      1.097318      5.767668      2.110810
##         sq_V7         ln_V8         ln_V9        ln_V10        sq_V11
##      3.613665      5.176782      3.346176      4.456050      1.851019
##        tk_V12        ln_V13        ln_V14
##      1.306453      4.721203      4.074296
```

```
cor(data)#all good
```

```
##                V1          V2          V3            V4          V5
## V1      1.00000000 -0.20046922  0.40658341 -0.0558915822  0.42097171
## V2     -0.20046922  1.00000000 -0.53382819 -0.0426967193 -0.51660371
## V3      0.40658341 -0.53382819  1.00000000  0.0629380275  0.76365145
## V4     -0.05589158 -0.04269672  0.06293803  1.0000000000  0.09120281
## V5      0.42097171 -0.51660371  0.76365145  0.0912028068  1.00000000
## V6     -0.21924670  0.31199059 -0.39167585  0.0912512250 -0.30218819
## V7      0.35273425 -0.56953734  0.64477851  0.0865177743  0.73147010
## V8     -0.37967009  0.66440822 -0.70802699 -0.0991757802 -0.76923011
## V9      0.62550515 -0.31194783  0.59512927 -0.0073682409  0.61144056
## V10     0.58276431 -0.31456332  0.72076018 -0.0355865176  0.66802320
```

```
## V11     0.28994558 -0.39167855  0.38324756 -0.1215151737  0.18893268
## V12    -0.38506394  0.17552032 -0.35697654  0.0487884850 -0.38005064
## V13     0.45562148 -0.41299457  0.60379972 -0.0539292984  0.59087892
## V14    -0.38830461  0.36044534 -0.48372516  0.1752601772 -0.42732077
## ln_V1   0.66648575 -0.51709145  0.73082136  0.0284964804  0.78861573
## sq_V2  -0.23022250  0.96204238 -0.58391710 -0.0426045385 -0.54888926
## ln_V3   0.38051504 -0.65589800  0.94316147  0.0807278208  0.72493114
## ln_V5   0.42964909 -0.56853850  0.78011101  0.0829421041  0.99394305
## sq_V7   0.38293981 -0.54461264  0.68160165  0.0778377493  0.75461814
## ln_V8  -0.46423877  0.59065522 -0.76128612 -0.0870373621 -0.83197732
## ln_V9   0.56866368 -0.35064262  0.56180887  0.0128764949  0.59667211
## ln_V10  0.55129033 -0.30589229  0.70603675 -0.0371919890  0.66028657
## sq_V11  0.29748220 -0.39101545  0.39799522 -0.1255842321  0.21394992
## tk_V12 -0.29018175  0.19376608 -0.40930043 -0.0002093129 -0.41977138
## ln_V13  0.39554286 -0.47770625  0.59742941 -0.0740740491  0.57707225
## ln_V14 -0.52794637  0.36334450 -0.54155616  0.1584119393 -0.51060029
##                 V6          V7          V8           V9         V10
## V1     -0.21924670  0.35273425 -0.37967009  0.625505145  0.58276431
## V2      0.31199059 -0.56953734  0.66440822 -0.311947826 -0.31456332
## V3     -0.39167585  0.64477851 -0.70802699  0.595129275  0.72076018
## V4      0.09125123  0.08651777 -0.09917578 -0.007368241 -0.03558652
## V5     -0.30218819  0.73147010 -0.76923011  0.611440563  0.66802320
## V6      1.00000000 -0.24026493  0.20524621 -0.209846668 -0.29204783
## V7     -0.24026493  1.00000000 -0.74788054  0.456022452  0.50645559
## V8      0.20524621 -0.74788054  1.00000000 -0.494587930 -0.53443158
## V9     -0.20984667  0.45602245 -0.49458793  1.000000000  0.91022819
## V10    -0.29204783  0.50645559 -0.53443158  0.910228189  1.00000000
## V11    -0.35550149  0.26151501 -0.23247054  0.464741179  0.46085304
## V12     0.12806864 -0.27353398  0.29151167 -0.444412816 -0.44180801
## V13    -0.61380827  0.60233853 -0.49699583  0.488676335  0.54399341
## V14     0.69535995 -0.37695457  0.24992873 -0.381626231 -0.46853593
## ln_V1  -0.30694282  0.65828357 -0.68190317  0.853406927  0.82823360
## sq_V2   0.33344389 -0.58839914  0.69717530 -0.344419507 -0.37100313
## ln_V3  -0.43126723  0.62538123 -0.71654970  0.574475834  0.66151988
## ln_V5  -0.30452060  0.76559620 -0.80789127  0.625307830  0.67837774
## sq_V7  -0.25586312  0.98211196 -0.74823427  0.476366904  0.54003886
## ln_V8   0.25658353 -0.77824331  0.96467087 -0.560334669 -0.61900763
## ln_V9  -0.20074161  0.44909766 -0.49008218  0.948265713  0.85053306
## ln_V10 -0.29921397  0.49553439 -0.51399646  0.863531361  0.98858645
## sq_V11 -0.35412219  0.28368188 -0.24977339  0.473568951  0.47516568
## tk_V12  0.17027689 -0.30341252  0.29764210 -0.398969339 -0.43169563
## ln_V13 -0.66452756  0.60680587 -0.48061210  0.460505752  0.52238531
## ln_V14  0.63202122 -0.45342171  0.34278032 -0.481970711 -0.56146566
##               V11         V12        V13        V14       ln_V1
## V1      0.2899456 -0.38506394  0.4556215 -0.3883046  0.66648575
## V2     -0.3916785  0.17552032 -0.4129946  0.3604453 -0.51709145
## V3      0.3832476 -0.35697654  0.6037997 -0.4837252  0.73082136
## V4     -0.1215152  0.04878848 -0.0539293  0.1752602  0.02849648
## V5      0.1889327 -0.38005064  0.5908789 -0.4273208  0.78861573
## V6     -0.3555015  0.12806864 -0.6138083  0.6953599 -0.30694282
## V7      0.2615150 -0.27353398  0.6023385 -0.3769546  0.65828357
## V8     -0.2324705  0.29151167 -0.4969958  0.2499287 -0.68190317
## V9      0.4647412 -0.44441282  0.4886763 -0.3816262  0.85340693
## V10     0.4608530 -0.44180801  0.5439934 -0.4685359  0.82823360
```

```
## V11     1.0000000 -0.17738330  0.3740443 -0.5077867  0.38955367
## V12    -0.1773833  1.00000000 -0.3660869  0.3334608 -0.47875518
## V13     0.3740443 -0.36608690  1.0000000 -0.7376627  0.62661501
## V14    -0.5077867  0.33346082 -0.7376627  1.0000000 -0.45430195
## ln_V1   0.3895537 -0.47875518  0.6266150 -0.4543020  1.00000000
## sq_V2  -0.4408463  0.20389796 -0.4395161  0.3829699 -0.54416901
## ln_V3   0.4300223 -0.33105046  0.5991034 -0.5192702  0.73955264
## ln_V5   0.2302262 -0.37934852  0.6011609 -0.4308060  0.80698806
## sq_V7   0.2633243 -0.28507918  0.6288170 -0.3914658  0.69105598
## ln_V8  -0.2381270  0.32484052 -0.5630306  0.2923157 -0.74392560
## ln_V9   0.4135979 -0.41126646  0.4623540 -0.3426318  0.83894314
## ln_V10  0.4303503 -0.42794690  0.5301311 -0.4747008  0.80997746
## sq_V11  0.9979917 -0.18599772  0.3855399 -0.5106578  0.40967259
## tk_V12 -0.1043693  0.79623085 -0.3413741  0.2643766 -0.49195875
## ln_V13  0.4170223 -0.34127854  0.9440309 -0.8154423  0.59179632
## ln_V14 -0.5017286  0.40238181 -0.8050341  0.9531555 -0.56724216
##              sq_V2       ln_V3      ln_V5       sq_V7       ln_V8
## V1     -0.23022250  0.38051504  0.4296491  0.38293981 -0.46423877
## V2      0.96204238 -0.65589800 -0.5685385 -0.54461264  0.59065522
## V3     -0.58391710  0.94316147  0.7801110  0.68160165 -0.76128612
## V4     -0.04260454  0.08072782  0.0829421  0.07783775 -0.08703736
## V5     -0.54888926  0.72493114  0.9939431  0.75461814 -0.83197732
## V6      0.33344389 -0.43126723 -0.3045206 -0.25586312  0.25658353
## V7     -0.58839914  0.62538123  0.7655962  0.98211196 -0.77824331
## V8      0.69717530 -0.71654970 -0.8078913 -0.74823427  0.96467087
## V9     -0.34441951  0.57447583  0.6253078  0.47636690 -0.56033467
## V10    -0.37100313  0.66151988  0.6783777  0.54003886 -0.61900763
## V11    -0.44084631  0.43002235  0.2302262  0.26332427 -0.23812704
## V12     0.20389796 -0.33105046 -0.3793485 -0.28507918  0.32484052
## V13    -0.43951611  0.59910338  0.6011609  0.62881701 -0.56303055
## V14     0.38296991 -0.51927024 -0.4308060 -0.39146580  0.29231567
## ln_V1  -0.54416901  0.73955264  0.8069881  0.69105598 -0.74392560
## sq_V2   1.00000000 -0.67750503 -0.5987193 -0.56935552  0.63081935
## ln_V3  -0.67750503  1.00000000  0.7504627  0.65459886 -0.73029652
## ln_V5  -0.59871929  0.75046275  1.0000000  0.78351108 -0.86001834
## sq_V7  -0.56935552  0.65459886  0.7835111  1.00000000 -0.79637675
## ln_V8   0.63081935 -0.73029652 -0.8600183 -0.79637675  1.00000000
## ln_V9  -0.36180823  0.58053294  0.6129483  0.46653034 -0.54211160
## ln_V10 -0.36193217  0.65928544  0.6683072  0.53169245 -0.59962147
## sq_V11 -0.44056771  0.44040778  0.2556984  0.28690083 -0.25875802
## tk_V12  0.22635940 -0.37089400 -0.4153167 -0.33038831  0.34643218
## ln_V13 -0.49309392  0.61757844  0.5921363  0.61345685 -0.52434277
## ln_V14  0.39391547 -0.55388663 -0.5152507 -0.47657821  0.40572110
##              ln_V9      ln_V10     sq_V11        tk_V12      ln_V13
## V1      0.56866368  0.55129033  0.2974822 -0.2901817450  0.39554286
## V2     -0.35064262 -0.30589229 -0.3910155  0.1937660791 -0.47770625
## V3      0.56180887  0.70603675  0.3979952 -0.4093004342  0.59742941
## V4      0.01287649 -0.03719199 -0.1255842 -0.0002093129 -0.07407405
## V5      0.59667211  0.66028657  0.2139499 -0.4197713766  0.57707225
## V6     -0.20074161 -0.29921397 -0.3541222  0.1702768946 -0.66452756
## V7      0.44909766  0.49553439  0.2836819 -0.3034125172  0.60680587
## V8     -0.49008218 -0.51399646 -0.2497734  0.2976420950 -0.48061210
## V9      0.94826571  0.86353136  0.4735690 -0.3989693386  0.46050575
## V10     0.85053306  0.98858645  0.4751657 -0.4316956262  0.52238531
```

```
## V11     0.41359791  0.43035027  0.9979917 -0.1043692711  0.41702231
## V12    -0.41126646 -0.42794690 -0.1859977  0.7962308500 -0.34127854
## V13     0.46235402  0.53013110  0.3855399 -0.3413740719  0.94403094
## V14    -0.34263176 -0.47470076 -0.5106578  0.2643766445 -0.81544235
## ln_V1   0.83894314  0.80997746  0.4096726 -0.4919587525  0.59179632
## sq_V2  -0.36180823 -0.36193217 -0.4405677  0.2263594018 -0.49309392
## ln_V3   0.58053294  0.65928544  0.4404078 -0.3708940019  0.61757844
## ln_V5   0.61294826  0.66830724  0.2556984 -0.4153167199  0.59213631
## sq_V7   0.46653034  0.53169245  0.2869008 -0.3303883147  0.61345685
## ln_V8  -0.54211160 -0.59962147 -0.2587580  0.3464321799 -0.52434277
## ln_V9   1.00000000  0.82048710  0.4212032 -0.3812309553  0.43981882
## ln_V10  0.82048710  1.00000000  0.4459480 -0.4298728237  0.51430111
## sq_V11  0.42120316  0.44594802  1.0000000 -0.1163044464  0.42794409
## tk_V12 -0.38123096 -0.42987282 -0.1163044  1.0000000000 -0.32351081
## ln_V13  0.43981882  0.51430111  0.4279441 -0.3235108137  1.00000000
## ln_V14 -0.43451250 -0.55718379 -0.5087360  0.3220093013 -0.82296003
##            ln_V14
## V1     -0.5279464
## V2      0.3633445
## V3     -0.5415562
## V4      0.1584119
## V5     -0.5106003
## V6      0.6320212
## V7     -0.4534217
## V8      0.3427803
## V9     -0.4819707
## V10    -0.5614657
## V11    -0.5017286
## V12     0.4023818
## V13    -0.8050341
## V14     0.9531555
## ln_V1  -0.5672422
## sq_V2   0.3939155
## ln_V3  -0.5538866
## ln_V5  -0.5152507
## sq_V7  -0.4765782
## ln_V8   0.4057211
## ln_V9  -0.4345125
## ln_V10 -0.5571838
## sq_V11 -0.5087360
## tk_V12  0.3220093
## ln_V13 -0.8229600
## ln_V14  1.0000000
```

```
#setting up null and full models for the variable selection
null = lm(ln_V1 ~ 1, data=data)
full = lm(ln_V1 ~ sq_V2 + ln_V3 + as.factor(V4) + ln_V5 + V6 + sq_V7 + ln_V8 + ln_V9 + ln_V10 + sq_V11 +

#forward selection
dataFwd = step(null, scope = list(lower=null, upper=full), direction="forward")
```

```
## Start:  AIC=781.31
## ln_V1 ~ 1
##
##                 Df Sum of Sq     RSS    AIC
```

```
## + ln_V9          1   1661.45  699.15 167.60
## + ln_V10         1   1548.71  811.90 243.26
## + ln_V5          1   1537.29  823.31 250.32
## + ln_V8          1   1306.42 1054.19 375.40
## + ln_V3          1   1291.10 1069.50 382.70
## + sq_V7          1   1127.33 1233.28 454.79
## + ln_V13         1    826.74 1533.87 565.16
## + ln_V14         1    759.56 1601.05 586.85
## + sq_V2          1    699.02 1661.58 605.63
## + tk_V12         1    571.32 1789.28 643.09
## + sq_V11         1    396.18 1964.42 690.35
## + V6             1    222.40 2138.20 733.24
## <none>                       2360.60 781.31
## + as.factor(V4)  1      1.92 2358.69 782.90
##
## Step:  AIC=167.6
## ln_V1 ~ ln_V9
##
##                 Df Sum of Sq    RSS      AIC
## + ln_V5          1    324.08 375.07 -145.511
## + ln_V8          1    279.46 419.69  -88.631
## + sq_V7          1    270.95 428.20  -78.475
## + ln_V3          1    227.04 472.11  -29.081
## + sq_V2          1    157.28 541.87   40.659
## + ln_V13         1    145.30 553.85   51.720
## + ln_V14         1    119.58 579.57   74.692
## + ln_V10         1    106.87 592.28   85.665
## + tk_V12         1     81.83 617.32  106.616
## + V6             1     47.20 651.95  134.233
## + sq_V11         1      9.10 690.05  162.976
## <none>                       699.15  167.604
## + as.factor(V4)  1      0.74 698.41  169.069
##
## Step:  AIC=-145.51
## ln_V1 ~ ln_V9 + ln_V5
##
##                 Df Sum of Sq    RSS     AIC
## + sq_V7          1   26.9273 348.14 -181.21
## + ln_V3          1   25.9787 349.09 -179.83
## + ln_V14         1   24.5129 350.56 -177.71
## + tk_V12         1   22.2570 352.81 -174.47
## + ln_V13         1   18.8881 356.18 -169.66
## + ln_V10         1   16.2057 358.86 -165.86
## + ln_V8          1   15.7478 359.32 -165.22
## + sq_V2          1   15.0088 360.06 -164.18
## + sq_V11         1    9.4779 365.59 -156.46
## + V6             1    7.4303 367.64 -153.64
## <none>                       375.07 -145.51
## + as.factor(V4)  1    0.7297 374.34 -144.50
##
## Step:  AIC=-181.21
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7
##
##                 Df Sum of Sq    RSS     AIC
```

```
## + tk_V12         1   21.5827 326.56 -211.59
## + ln_V3          1   18.1628 329.98 -206.32
## + ln_V14         1   18.0453 330.10 -206.14
## + ln_V10         1   14.1540 333.99 -200.21
## + ln_V13         1    8.4374 339.70 -191.62
## + sq_V2          1    8.3447 339.80 -191.49
## + V6             1    6.6193 341.52 -188.92
## + sq_V11         1    5.1000 343.04 -186.68
## + ln_V8          1    4.5002 343.64 -185.79
## <none>                       348.14 -181.21
## + as.factor(V4)  1    0.9125 347.23 -180.54
##
## Step:  AIC=-211.59
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12
##
##                 Df Sum of Sq    RSS     AIC
## + ln_V3          1   15.9629 310.60 -234.95
## + ln_V14         1   14.1281 312.43 -231.97
## + ln_V10         1    9.7639 316.79 -224.95
## + sq_V2          1    9.2675 317.29 -224.16
## + sq_V11         1    6.3793 320.18 -219.57
## + ln_V13         1    6.3138 320.24 -219.47
## + V6             1    5.5423 321.02 -218.25
## + ln_V8          1    5.3019 321.26 -217.88
## <none>                       326.56 -211.59
## + as.factor(V4)  1    0.6659 325.89 -210.62
##
## Step:  AIC=-234.95
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3
##
##                 Df Sum of Sq    RSS     AIC
## + ln_V14         1    8.2667 302.33 -246.60
## + ln_V10         1    5.3308 305.26 -241.71
## + ln_V8          1    2.3070 308.29 -236.72
## + ln_V13         1    2.2370 308.36 -236.61
## + sq_V2          1    2.1793 308.42 -236.51
## + sq_V11         1    1.6974 308.90 -235.72
## + V6             1    1.2490 309.35 -234.99
## <none>                       310.60 -234.95
## + as.factor(V4)  1    0.9489 309.65 -234.50
##
## Step:  AIC=-246.6
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3 + ln_V14
##
##                 Df Sum of Sq    RSS     AIC
## + ln_V8          1    4.8846 297.44 -252.84
## + ln_V10         1    2.7377 299.59 -249.20
## + sq_V2          1    2.1552 300.17 -248.22
## <none>                       302.33 -246.60
## + ln_V13         1    0.7266 301.60 -245.82
## + V6             1    0.3355 301.99 -245.16
## + as.factor(V4)  1    0.0670 302.26 -244.71
## + sq_V11         1    0.0589 302.27 -244.70
##
```

```
## Step:  AIC=-252.84
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3 + ln_V14 + ln_V8
##
##                  Df Sum of Sq    RSS     AIC
## + ln_V10          1   2.34267 295.10 -254.84
## + sq_V2           1   1.26203 296.18 -253.00
## <none>                        297.44 -252.84
## + ln_V13          1   0.73807 296.71 -252.10
## + V6              1   0.42419 297.02 -251.56
## + sq_V11          1   0.04022 297.40 -250.91
## + as.factor(V4)   1   0.02956 297.41 -250.89
##
## Step:  AIC=-254.84
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3 + ln_V14 + ln_V8 +
##     ln_V10
##
##                  Df Sum of Sq    RSS     AIC
## + sq_V2           1   2.28833 292.81 -256.78
## <none>                        295.10 -254.84
## + ln_V13          1   0.43364 294.67 -253.59
## + V6              1   0.33938 294.76 -253.43
## + sq_V11          1   0.00865 295.09 -252.86
## + as.factor(V4)   1   0.00258 295.10 -252.85
##
## Step:  AIC=-256.78
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3 + ln_V14 + ln_V8 +
##     ln_V10 + sq_V2
##
##                  Df Sum of Sq    RSS     AIC
## <none>                        292.81 -256.78
## + ln_V13          1   0.56907 292.24 -255.77
## + V6              1   0.47410 292.34 -255.60
## + sq_V11          1   0.18019 292.63 -255.09
## + as.factor(V4)   1   0.00052 292.81 -254.78
```

```
summary(dataFwd)
```

```
##
## Call:
## lm(formula = ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3 +
##     ln_V14 + ln_V8 + ln_V10 + sq_V2, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2785 -0.5472 -0.0132  0.5140  2.4074
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.919e+00  1.195e+00  -2.444 0.014885 *
## ln_V9        1.130e+00  6.976e-02  16.200  < 2e-16 ***
## ln_V5        1.880e+00  3.928e-01   4.787 2.23e-06 ***
## sq_V7        6.497e-05  1.771e-05   3.668 0.000271 ***
## tk_V12      -7.210e-27  1.383e-27  -5.214 2.71e-07 ***
## ln_V3        1.555e-01  8.210e-02   1.894 0.058865 .
## ln_V14      -3.803e-01  1.106e-01  -3.438 0.000635 ***
```

```
## ln_V8       -3.396e-01  1.439e-01  -2.359 0.018693 *
## ln_V10       4.262e-01  1.784e-01   2.389 0.017273 *
## sq_V2       -3.364e-02  1.709e-02  -1.969 0.049531 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7683 on 496 degrees of freedom
## Multiple R-squared:  0.876,  Adjusted R-squared:  0.8737
## F-statistic: 389.2 on 9 and 496 DF,  p-value: < 2.2e-16
```

After performing the forward selection method, the independent variables selected were ln_V9, ln_V5, sq_V7, tk_V12, ln_V3, ln_V14, ln_V8, ln_V10 and sq_V2. Looking at the p-value and the t-statistics, the ln_V3 is the only non-significant variable. The multiple R squared is 0.876 and adjusted R squared is 0.8737.

b)

```
#backward elimination
dataBkwd = step(full, direction="backward")
```

```
## Start:  AIC=-250.44
## ln_V1 ~ sq_V2 + ln_V3 + as.factor(V4) + ln_V5 + V6 + sq_V7 +
##     ln_V8 + ln_V9 + ln_V10 + sq_V11 + tk_V12 + ln_V13 + ln_V14
##
##                 Df Sum of Sq    RSS      AIC
## - as.factor(V4)  1     0.000 291.85 -252.444
## - V6             1     0.174 292.03 -252.144
## - sq_V11         1     0.201 292.05 -252.096
## - ln_V13         1     0.315 292.17 -251.899
## <none>                       291.85 -250.445
## - ln_V3          1     2.548 294.40 -248.046
## - sq_V2          1     2.675 294.53 -247.827
## - ln_V10         1     3.173 295.03 -246.973
## - ln_V8          1     3.233 295.09 -246.871
## - ln_V14         1     5.917 297.77 -242.289
## - sq_V7          1     7.695 299.55 -239.277
## - ln_V5          1    11.637 303.49 -232.661
## - tk_V12         1    15.341 307.19 -226.523
## - ln_V9          1   149.248 441.10  -43.454
##
## Step:  AIC=-252.44
## ln_V1 ~ sq_V2 + ln_V3 + ln_V5 + V6 + sq_V7 + ln_V8 + ln_V9 +
##     ln_V10 + sq_V11 + tk_V12 + ln_V13 + ln_V14
##
##          Df Sum of Sq    RSS      AIC
## - V6      1     0.174 292.03 -254.144
## - sq_V11  1     0.202 292.05 -254.095
## - ln_V13  1     0.315 292.17 -253.899
## <none>                291.85 -252.444
## - ln_V3   1     2.589 294.44 -249.975
## - sq_V2   1     2.678 294.53 -249.823
## - ln_V10  1     3.195 295.05 -248.935
## - ln_V8   1     3.236 295.09 -248.864
## - ln_V14  1     6.027 297.88 -244.102
## - sq_V7   1     7.715 299.57 -241.243
## - ln_V5   1    11.661 303.51 -234.621
## - tk_V12  1    15.343 307.20 -228.519
```

```
## - ln_V9   1   149.621 441.47  -45.027
##
## Step:  AIC=-254.14
## ln_V1 ~ sq_V2 + ln_V3 + ln_V5 + sq_V7 + ln_V8 + ln_V9 + ln_V10 +
##     sq_V11 + tk_V12 + ln_V13 + ln_V14
##
##          Df Sum of Sq    RSS      AIC
## - sq_V11  1     0.217 292.24 -255.768
## - ln_V13  1     0.606 292.63 -255.095
## <none>                292.03 -254.144
## - ln_V3   1     2.452 294.48 -251.913
## - sq_V2   1     2.636 294.66 -251.598
## - ln_V10  1     3.178 295.20 -250.667
## - ln_V8   1     3.200 295.23 -250.629
## - ln_V14  1     5.853 297.88 -246.102
## - sq_V7   1     8.654 300.68 -241.366
## - ln_V5   1    11.734 303.76 -236.211
## - tk_V12  1    15.439 307.47 -230.075
## - ln_V9   1   153.316 445.34  -42.613
##
## Step:  AIC=-255.77
## ln_V1 ~ sq_V2 + ln_V3 + ln_V5 + sq_V7 + ln_V8 + ln_V9 + ln_V10 +
##     tk_V12 + ln_V13 + ln_V14
##
##          Df Sum of Sq    RSS      AIC
## - ln_V13  1     0.569 292.81 -256.783
## <none>                292.24 -255.768
## - ln_V3   1     2.341 294.58 -253.731
## - sq_V2   1     2.424 294.67 -253.588
## - ln_V10  1     3.006 295.25 -252.589
## - ln_V8   1     3.285 295.53 -252.112
## - ln_V14  1     5.680 297.92 -248.028
## - sq_V7   1     8.513 300.76 -243.238
## - ln_V5   1    13.511 305.75 -234.899
## - tk_V12  1    16.166 308.41 -230.524
## - ln_V9   1   155.506 447.75  -41.884
##
## Step:  AIC=-256.78
## ln_V1 ~ sq_V2 + ln_V3 + ln_V5 + sq_V7 + ln_V8 + ln_V9 + ln_V10 +
##     tk_V12 + ln_V14
##
##          Df Sum of Sq    RSS      AIC
## <none>                292.81 -256.783
## - ln_V3   1     2.117 294.93 -255.139
## - sq_V2   1     2.288 295.10 -254.844
## - ln_V8   1     3.286 296.10 -253.136
## - ln_V10  1     3.369 296.18 -252.995
## - ln_V14  1     6.979 299.79 -246.865
## - sq_V7   1     7.944 300.76 -245.238
## - ln_V5   1    13.530 306.34 -235.926
## - tk_V12  1    16.051 308.86 -231.780
## - ln_V9   1   154.940 447.75  -43.881
```

```r
summary(dataBkwd)
```

```
##
## Call:
## lm(formula = ln_V1 ~ sq_V2 + ln_V3 + ln_V5 + sq_V7 + ln_V8 +
##     ln_V9 + ln_V10 + tk_V12 + ln_V14, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2785 -0.5472 -0.0132  0.5140  2.4074
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.919e+00  1.195e+00  -2.444 0.014885 *
## sq_V2       -3.364e-02  1.709e-02  -1.969 0.049531 *
## ln_V3        1.555e-01  8.210e-02   1.894 0.058865 .
## ln_V5        1.880e+00  3.928e-01   4.787 2.23e-06 ***
## sq_V7        6.497e-05  1.771e-05   3.668 0.000271 ***
## ln_V8       -3.396e-01  1.439e-01  -2.359 0.018693 *
## ln_V9        1.130e+00  6.976e-02  16.200  < 2e-16 ***
## ln_V10       4.262e-01  1.784e-01   2.389 0.017273 *
## tk_V12      -7.210e-27  1.383e-27  -5.214 2.71e-07 ***
## ln_V14      -3.803e-01  1.106e-01  -3.438 0.000635 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7683 on 496 degrees of freedom
## Multiple R-squared:  0.876,  Adjusted R-squared:  0.8737
## F-statistic: 389.2 on 9 and 496 DF,  p-value: < 2.2e-16
```

```r
#stepwise regression
dataStp = step(null, scope = list(upper=full), direction="both")
```

```
## Start:  AIC=781.31
## ln_V1 ~ 1
##
##                  Df Sum of Sq     RSS    AIC
## + ln_V9           1   1661.45  699.15 167.60
## + ln_V10          1   1548.71  811.90 243.26
## + ln_V5           1   1537.29  823.31 250.32
## + ln_V8           1   1306.42 1054.19 375.40
## + ln_V3           1   1291.10 1069.50 382.70
## + sq_V7           1   1127.33 1233.28 454.79
## + ln_V13          1    826.74 1533.87 565.16
## + ln_V14          1    759.56 1601.05 586.85
## + sq_V2           1    699.02 1661.58 605.63
## + tk_V12          1    571.32 1789.28 643.09
## + sq_V11          1    396.18 1964.42 690.35
## + V6              1    222.40 2138.20 733.24
## <none>                         2360.60 781.31
## + as.factor(V4)   1      1.92 2358.69 782.90
##
## Step:  AIC=167.6
## ln_V1 ~ ln_V9
```

```
##
##                 Df Sum of Sq     RSS     AIC
## + ln_V5          1    324.08  375.07 -145.51
## + ln_V8          1    279.46  419.69  -88.63
## + sq_V7          1    270.95  428.20  -78.48
## + ln_V3          1    227.04  472.11  -29.08
## + sq_V2          1    157.28  541.87   40.66
## + ln_V13         1    145.30  553.85   51.72
## + ln_V14         1    119.58  579.57   74.69
## + ln_V10         1    106.87  592.28   85.67
## + tk_V12         1     81.83  617.32  106.62
## + V6             1     47.20  651.95  134.23
## + sq_V11         1      9.10  690.05  162.98
## <none>                        699.15  167.60
## + as.factor(V4)  1      0.74  698.41  169.07
## - ln_V9          1   1661.45 2360.60  781.31
##
## Step:  AIC=-145.51
## ln_V1 ~ ln_V9 + ln_V5
##
##                 Df Sum of Sq    RSS     AIC
## + sq_V7          1     26.93 348.14 -181.21
## + ln_V3          1     25.98 349.09 -179.83
## + ln_V14         1     24.51 350.56 -177.71
## + tk_V12         1     22.26 352.81 -174.47
## + ln_V13         1     18.89 356.18 -169.66
## + ln_V10         1     16.21 358.86 -165.86
## + ln_V8          1     15.75 359.32 -165.22
## + sq_V2          1     15.01 360.06 -164.18
## + sq_V11         1      9.48 365.59 -156.46
## + V6             1      7.43 367.64 -153.64
## <none>                       375.07 -145.51
## + as.factor(V4)  1      0.73 374.34 -144.50
## - ln_V5          1    324.08 699.15  167.60
## - ln_V9          1    448.24 823.31  250.32
##
## Step:  AIC=-181.21
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7
##
##                 Df Sum of Sq    RSS      AIC
## + tk_V12         1     21.58 326.56 -211.592
## + ln_V3          1     18.16 329.98 -206.320
## + ln_V14         1     18.05 330.10 -206.140
## + ln_V10         1     14.15 333.99 -200.210
## + ln_V13         1      8.44 339.70 -191.623
## + sq_V2          1      8.34 339.80 -191.485
## + V6             1      6.62 341.52 -188.922
## + sq_V11         1      5.10 343.04 -186.676
## + ln_V8          1      4.50 343.64 -185.792
## <none>                       348.14 -181.208
## + as.factor(V4)  1      0.91 347.23 -180.536
## - sq_V7          1     26.93 375.07 -145.511
## - ln_V5          1     80.06 428.20  -78.475
## - ln_V9          1    454.05 802.19  239.169
```

```
##
## Step:  AIC=-211.59
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12
##
##                  Df Sum of Sq    RSS     AIC
## + ln_V3           1     15.96 310.60 -234.95
## + ln_V14          1     14.13 312.43 -231.97
## + ln_V10          1      9.76 316.79 -224.95
## + sq_V2           1      9.27 317.29 -224.16
## + sq_V11          1      6.38 320.18 -219.57
## + ln_V13          1      6.31 320.24 -219.47
## + V6              1      5.54 321.02 -218.25
## + ln_V8           1      5.30 321.26 -217.88
## <none>                        326.56 -211.59
## + as.factor(V4)   1      0.67 325.89 -210.62
## - tk_V12          1     21.58 348.14 -181.21
## - sq_V7           1     26.25 352.81 -174.47
## - ln_V5           1     64.65 391.20 -122.20
## - ln_V9           1    406.16 732.72  195.34
##
## Step:  AIC=-234.95
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3
##
##                  Df Sum of Sq    RSS     AIC
## + ln_V14          1      8.27 302.33 -246.60
## + ln_V10          1      5.33 305.26 -241.71
## + ln_V8           1      2.31 308.29 -236.72
## + ln_V13          1      2.24 308.36 -236.61
## + sq_V2           1      2.18 308.42 -236.51
## + sq_V11          1      1.70 308.90 -235.72
## + V6              1      1.25 309.35 -234.99
## <none>                        310.60 -234.95
## + as.factor(V4)   1      0.95 309.65 -234.50
## - ln_V3           1     15.96 326.56 -211.59
## - sq_V7           1     18.97 329.57 -206.95
## - tk_V12          1     19.38 329.98 -206.32
## - ln_V5           1     34.87 345.46 -183.11
## - ln_V9           1    350.82 661.42  145.53
##
## Step:  AIC=-246.6
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3 + ln_V14
##
##                  Df Sum of Sq    RSS     AIC
## + ln_V8           1      4.88 297.44 -252.84
## + ln_V10          1      2.74 299.59 -249.20
## + sq_V2           1      2.16 300.17 -248.22
## <none>                        302.33 -246.60
## + ln_V13          1      0.73 301.60 -245.82
## + V6              1      0.34 301.99 -245.16
## + as.factor(V4)   1      0.07 302.26 -244.71
## + sq_V11          1      0.06 302.27 -244.70
## - ln_V14          1      8.27 310.60 -234.95
## - ln_V3           1     10.10 312.43 -231.97
## - sq_V7           1     16.29 318.62 -222.05
```

```
## - tk_V12         1     16.78 319.11 -221.27
## - ln_V5          1     33.50 335.83 -195.43
## - ln_V9          1    335.60 637.93  129.24
##
## Step:  AIC=-252.84
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3 + ln_V14 + ln_V8
##
##                 Df Sum of Sq    RSS     AIC
## + ln_V10         1      2.34 295.10 -254.84
## + sq_V2          1      1.26 296.18 -253.00
## <none>                       297.44 -252.84
## + ln_V13         1      0.74 296.71 -252.10
## + V6             1      0.42 297.02 -251.56
## + sq_V11         1      0.04 297.40 -250.91
## + as.factor(V4)  1      0.03 297.41 -250.89
## - ln_V8          1      4.88 302.33 -246.60
## - ln_V3          1      6.31 303.76 -244.21
## - sq_V7          1      8.22 305.66 -241.05
## - ln_V14         1     10.84 308.29 -236.72
## - ln_V5          1     14.64 312.08 -230.53
## - tk_V12         1     17.32 314.76 -226.21
## - ln_V9          1    332.36 629.81  124.75
##
## Step:  AIC=-254.84
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3 + ln_V14 + ln_V8 +
##     ln_V10
##
##                 Df Sum of Sq    RSS      AIC
## + sq_V2          1     2.288 292.81 -256.783
## <none>                       295.10 -254.844
## + ln_V13         1     0.434 294.67 -253.588
## + V6             1     0.339 294.76 -253.427
## + sq_V11         1     0.009 295.09 -252.859
## + as.factor(V4)  1     0.003 295.10 -252.849
## - ln_V10         1     2.343 297.44 -252.843
## - ln_V8          1     4.490 299.59 -249.204
## - ln_V3          1     5.160 300.26 -248.074
## - ln_V14         1     7.855 302.96 -243.551
## - sq_V7          1     8.516 303.62 -242.449
## - ln_V5          1    13.813 308.91 -233.697
## - tk_V12         1    15.656 310.76 -230.688
## - ln_V9          1   158.446 453.55  -39.375
##
## Step:  AIC=-256.78
## ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3 + ln_V14 + ln_V8 +
##     ln_V10 + sq_V2
##
##                 Df Sum of Sq    RSS      AIC
## <none>                       292.81 -256.783
## + ln_V13         1     0.569 292.24 -255.768
## + V6             1     0.474 292.34 -255.603
## - ln_V3          1     2.117 294.93 -255.139
## + sq_V11         1     0.180 292.63 -255.095
## - sq_V2          1     2.288 295.10 -254.844
```

```
## + as.factor(V4)  1     0.001 292.81 -254.784
## - ln_V8          1     3.286 296.10 -253.136
## - ln_V10         1     3.369 296.18 -252.995
## - ln_V14         1     6.979 299.79 -246.865
## - sq_V7          1     7.944 300.76 -245.238
## - ln_V5          1    13.530 306.34 -235.926
## - tk_V12         1    16.051 308.86 -231.780
## - ln_V9          1   154.940 447.75  -43.881
```

```r
summary(dataStp)
```

```
##
## Call:
## lm(formula = ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3 +
##     ln_V14 + ln_V8 + ln_V10 + sq_V2, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2785 -0.5472 -0.0132  0.5140  2.4074
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.919e+00  1.195e+00  -2.444 0.014885 *
## ln_V9        1.130e+00  6.976e-02  16.200  < 2e-16 ***
## ln_V5        1.880e+00  3.928e-01   4.787 2.23e-06 ***
## sq_V7        6.497e-05  1.771e-05   3.668 0.000271 ***
## tk_V12      -7.210e-27  1.383e-27  -5.214 2.71e-07 ***
## ln_V3        1.555e-01  8.210e-02   1.894 0.058865 .
## ln_V14      -3.803e-01  1.106e-01  -3.438 0.000635 ***
## ln_V8       -3.396e-01  1.439e-01  -2.359 0.018693 *
## ln_V10       4.262e-01  1.784e-01   2.389 0.017273 *
## sq_V2       -3.364e-02  1.709e-02  -1.969 0.049531 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7683 on 496 degrees of freedom
## Multiple R-squared:  0.876,  Adjusted R-squared:  0.8737
## F-statistic: 389.2 on 9 and 496 DF,  p-value: < 2.2e-16
```

After performing the backward selection method, the independent variables selected were ln_V9, ln_V5, sq_V7, tk_V12, ln_V3, ln_V14, ln_V8, ln_V10 and sq_V2. Looking at the p-value and the t-statistics, the ln_V3 is the only non-significant variable. The multiple R squared is 0.876 and adjusted R squared is 0.8737. For the stepwise selection, the independent variables selected were ln_V9, ln_V5, sq_V7, tk_V12, ln_V3, ln_V14, ln_V8, ln_V10 and sq_V2. Looking at the p-value and the t-statistics, the ln_V3 is the only non-significant variable. The multiple R squared is 0.876 and adjusted R squared is 0.8737. All the selection method I have performed selected the same independent variables, just in different orders. The multiple R squared and adjusted R Squared values were also the same.

c) Since all three selection method performed gave the same result, this model best represents the data. They all had the same adjusted R squared values. The adjusted and multiple R squared values were high, with all t-statistics highly significant.

d)

```r
#performing ANOVA
M1 = aov(ln_V1 ~ ln_V9 + ln_V5 + sq_V7 + tk_V12 + ln_V3 + ln_V14 + ln_V8 + ln_V10 + sq_V2, data=data)
summary(M1)
```

```
##              Df Sum Sq Mean Sq  F value   Pr(>F)    
## ln_V9         1 1661.5  1661.5 2814.358  < 2e-16 ***
## ln_V5         1  324.1   324.1  548.967  < 2e-16 ***
## sq_V7         1   26.9    26.9   45.613 4.05e-11 ***
## tk_V12        1   21.6    21.6   36.559 2.92e-09 ***
## ln_V3         1   16.0    16.0   27.040 2.92e-07 ***
## ln_V14        1    8.3     8.3   14.003 0.000204 ***
## ln_V8         1    4.9     4.9    8.274 0.004195 ** 
## ln_V10        1    2.3     2.3    3.968 0.046913 *  
## sq_V2         1    2.3     2.3    3.876 0.049531 *  
## Residuals   496  292.8     0.6                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After performing ANOVA for the model that was selected after the selection methods, the F-statistics and the p-value suggests that all covariates included in the ANOVA have significant effect on Y. The linear regression result from the selection methods demonstrates that all t-statistics and estimated coefficients of the covariates except for ln_V3 are significant. Even the p-value of ln_V3 were very close to being 0.05. The multiple and adjusted R- Squared high enough to demonstrate significance. The standard errors also looked not too big. For every 1 increase in ln_V9, the Y increases by 1.130e+00. For every 1 increase in ln_V5, the Y increases by 1.880e+00. For every 1 increase in sq_V7, the Y increases by 6.497e-05. For every 1 increase in tk_V12, the Y decreases by 7.210e-27. For every 1 increase in ln_V3, the Y increases by 1.555e-01.For every 1 increase in ln_V14, the Y decreases by 3.803e-01. For every 1 increase in ln_V8, the Y decreases by 3.396e-01. For every 1 increase in ln_V10, the Y increases by 4.262e-01. For every 1 increase in sq_V2, the Y decreases by 3.364e-02.

Problem 5

It would be very interesting to see if multiple regression is used in the field of streetwear/fashion. The dependent variable would be total sales made by each brands. There would be many possible independent variables, such as different types of clothes made, the materials used, cost of materials used, where the clothes were manufactured, brand media coverage (# of articles or news), fundings received from outside sources, collaborations between other brands, and so on. Obviously it would be extremely hard to gather all these data, and even so, there should be many more independent variables that cause the total sales fluctuate.