```r
getwd()
```

```
## [1] "C:/Users/uttva/OneDrive/Desktop/Rstuff"
```

```r
setwd("C:\\Users\\uttva\\OneDrive\\Desktop\\Rstuff")
# setting survey data to df
df=read.csv("Psych of Debt.csv")
head(df)[1][1]
```

```
##         StartDate
## 1 5/14/2019 14:56
## 2  5/15/2019 9:45
## 3  5/15/2019 9:48
## 4  5/15/2019 9:45
## 5  5/15/2019 9:45
## 6  5/15/2019 9:48
```

```r
write.csv(df,file="Psych_AG.csv")
df=read.csv("Psych_AG.csv")
head(df)[1]
```

```
##   X
## 1 1
## 2 2
## 3 3
## 4 4
## 5 5
## 6 6
```

```r
library("ggplot2")
```

```r
##INITIAL CLEANING


#Checking for IP Duplicated
table(df$IPAddress)
```

```
##
##                   104.238.59.98   107.11.182.13 107.198.109.163
##               1               1               1               1
##   107.72.178.37  107.77.205.198    107.77.207.9 108.194.139.169
##               1               1               1               1
##   108.195.76.65  108.215.139.10 108.222.113.167 128.180.159.142
##               1               1               1               1
##   128.61.47.142    130.91.31.14  134.114.101.51   143.200.75.46
##               1               1               1               1
##      146.7.4.77   155.201.43.87    162.237.6.96    165.225.0.98
##               1               1               1               1
## 169.237.251.226    171.65.79.59   172.58.17.190  173.169.99.215
##               1               1               1               1
##    173.170.8.31 174.107.190.165 174.195.131.234  174.207.29.202
##               1               1               1               1
## 174.223.129.108  174.255.10.230 181.214.179.111 181.214.179.141
##               1               1               1               1
## 181.214.188.230 181.214.201.141 181.215.121.232  181.215.74.194
##               1               1               1               1
```

```
##  181.215.82.151  184.101.39.152 184.103.171.126  184.174.187.66
##               1               1               1               1
##   198.137.18.59    198.41.62.73 199.116.175.201  199.234.90.149
##               1               1               1               1
##  205.173.36.151   206.130.52.10   208.68.216.10  209.129.119.10
##               1               1               1               1
##  209.251.238.55 216.200.134.101     24.16.99.15  24.243.147.196
##               1               1               1               1
##    24.73.208.98    24.92.114.17   45.56.161.198   47.218.24.123
##               1               1               1               1
##   50.205.167.34  50.227.114.155   50.239.137.83   50.239.137.88
##               1               1               1               1
##   50.249.14.243  50.254.202.241   50.45.224.108   64.179.181.33
##               1               1               1               1
##  64.253.209.180  65.120.235.218   65.190.81.203  65.249.163.137
##               1               1               1               1
##  65.249.163.187   66.241.73.242    66.31.52.115     67.45.33.55
##               1               1               1               1
##      68.0.76.92    68.103.32.54   68.204.27.205  69.138.139.198
##               1               1               1               1
##  69.253.232.224   70.164.96.252    70.92.122.74   71.15.192.196
##               1               1               1               1
##   71.162.192.24   71.196.41.166   71.82.117.167   71.84.252.205
##               1               1               1               1
##    72.176.9.124   72.204.209.28    72.223.74.61   73.106.73.255
##               1               1               1               1
##   73.106.78.225   73.115.93.211    73.15.17.128  73.206.209.217
##               1               1               1               1
##    73.23.83.185  73.234.191.110   73.37.249.176    73.54.245.34
##               1               1               1               1
##      73.6.70.25    74.73.20.127  75.136.132.164    75.64.103.65
##               1               1               1               1
##   75.83.252.102  76.109.133.139   76.251.70.134     76.92.60.53
##               1               1               1               1
##   96.87.203.221  98.125.180.237    98.145.6.160   98.216.192.20
##               1               1               1               1
##   98.221.16.195    99.203.26.99   99.203.81.198
##               1               1               1
```

```
# no duplicates

# Taking out first row(tester)
df<-df[2:nrow(df),]
#Only completed surveys

df<-df[df$Progress==100,]
```

```
# Random Stats
head(df)[1][1]
```

```
##   X
## 2 2
## 3 3
## 4 4
## 5 5
```

```
## 6 6
## 7 7
```

```
# Survey DUration
sort(df$Duration..in.seconds.,decreasing = T)
```

```
##   [1] 1621 1590 1304 1210 1077 1064  948  918  891  880  870  865  816  768
##  [15]  741  737  734  720  720  704  700  692  689  681  668  666  655  642
##  [29]  632  628  621  616  616  610  590  588  571  565  555  549  547  546
##  [43]  536  521  518  516  515  508  506  505  502  498  497  495  482  482
##  [57]  473  472  460  460  459  455  450  438  437  437  436  434  431  425
##  [71]  425  423  421  420  416  413  410  375  370  366  351  349  348  345
##  [85]  342  331  326  326  318  318  315  304  301  294  287  285  283  273
##  [99]  250  233  232  231  227  220  217  208  206
```

```
# Max Survey Time 27 minutes
1621/60
```

```
## [1] 27.01667
```

```
#Min Survey Time
# mhortest surveys 3.433333 3.466667 3.616667 minutes
# Possible that this person did not take it seriously
sort(df$Duration..in.seconds.,decreasing = F)[1:3]/60
```

```
## [1] 3.433333 3.466667 3.616667
```

```
# around 9 minutes on average
mean(df$Duration..in.seconds.)/60
```

```
## [1] 9.020561
```

```
#Median of 8.25
median(df$Duration..in.seconds.)/60
```

```
## [1] 8.25
```

```
# sd of duration =4.37 minutes
sd(df$Duration..in.seconds.)/60
```

```
## [1] 4.372665
```

```
# shortest surveys are  1.277763 1.270139 1.235835 sds from the mean,
#probably nothing to worry about
((mean(df$Duration..in.seconds.)/60)- (sort(df$Duration..in.seconds.,decreasing = F)[1:3]/60))/(sd(df$Du
```

```
## [1] 1.277763 1.270139 1.235835
```

```
# Everyone Consented
sum(df$consent==0)
```

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

```
#
```

```
# only working on CC data so let us subset for only people with cc debt
# first subset is people with debt
df<-df[df$anyDebt==1,]
str(df$debtSources)
```

```
##  Factor w/ 16 levels "","1","1,2","1,2,3",..: 5 8 2 4 15 5 6 9 2 3 ...
```

```r
# Subsetting so that we only get people who are paying off cc debt
df[is.na(df$paying_off_1)==F,][1]
```

```
##        X
## 4      4
## 5      5
## 6      6
## 7      7
## 9      9
## 11    11
## 12    12
## 14    14
## 15    15
## 17    17
## 18    18
## 19    19
## 21    21
## 22    22
## 23    23
## 24    24
## 26    26
## 27    27
## 28    28
## 30    30
## 31    31
## 32    32
## 33    33
## 34    34
## 36    36
## 37    37
## 38    38
## 39    39
## 40    40
## 43    43
## 47    47
## 49    49
## 50    50
## 51    51
## 52    52
## 54    54
## 56    56
## 57    57
## 59    59
## 60    60
## 61    61
## 62    62
## 63    63
## 64    64
## 67    67
## 68    68
## 71    71
## 73    73
## 74    74
## 76    76
```

```
## 77   77
## 78   78
## 80   80
## 81   81
## 82   82
## 83   83
## 85   85
## 86   86
## 87   87
## 88   88
## 89   89
## 90   90
## 91   91
## 92   92
## 95   95
## 96   96
## 100 100
## 101 101
## 102 102
## 103 103
## 105 105
## 106 106
## 107 107
```

```
ccdebt<-df[is.na(df$paying_off_1)==F & df$paying_off_1==1,]
# proof that we only have people who are paying off cc debt

table(ccdebt$paying_off_1)
```

```
##
##  1
## 60
```

```
# TIme for some analysis

head(ccdebt)[1]
```

```
##     X
## 4   4
## 5   5
## 6   6
## 7   7
## 9   9
## 11 11
```

```
# original amounts of cc debt
ccdebt$orig_amt_1
```

```
##  [1] 8000   5000   5000   10000  7000   10,000 4000   20000  2000   20000
## [11] 18000  1500   1,000  15000  10000  10000  3000   1400   100    20000
## [21] 2000   10000  900    0      400    8000   0      5000   700    10000
## [31] 3000   6000   5000   1500   0      356    0      1500   3000   12000
## [41] 10000  17000  500    20000  1000   12000  10000  1800   8000   2000
## [51] 50     28000  20000  7000   5000   500    15000  12000  12000  6000
## 34 Levels:  0 1,000 10,000 100 1000 10000 12000 1300 14,000 1400 ... 900
```

```
# Converting Original amount 1 or original cc debt to numeric
ccdebt$orig_amt_1<-as.numeric(as.character(ccdebt$orig_amt_1))
```

```
## Warning: NAs introduced by coercion
```

```
#Converting Current amount 1 or Current cc debt to numeric
ccdebt$current_amt_1<-as.numeric(as.character(ccdebt$current_amt_1))
```

```
## Warning: NAs introduced by coercion
```

```
ccdebt$current_amt_1-ccdebt$orig_amt_1
```

```
##  [1]  52000  -2000   -700  -3000      0     NA      0      0  -1000 -18000
## [11]  -9000      0     NA      0  -9000  -9000   2000      0      0 -10000
## [21]  13000  -2000      0  20000      0  -5000  11000  -2000   -600   7000
## [31]   -500      0      0   -500  18000   -186  12000   -500  -1500 -10200
## [41]      0   8000      0      0    200  -3000  -5000      0   1000   -800
## [51]    650   3000  -2000  -1000   -500   4500 -13000      0  -4000  -4000
```

```
ccdebt$orig_amt_1
```

```
##  [1]  8000  5000  5000 10000  7000    NA  4000 20000  2000 20000 18000
## [12]  1500    NA 15000 10000 10000  3000  1400   100 20000  2000 10000
## [23]   900     0   400  8000     0  5000   700 10000  3000  6000  5000
## [34]  1500     0   356     0  1500  3000 12000 10000 17000   500 20000
## [45]  1000 12000 10000  1800  8000  2000    50 28000 20000  7000  5000
## [56]   500 15000 12000 12000  6000
```

```
# getting rid of people who said they are paying off cc debt but had no original amount of cc debt
ccdebt<-ccdebt[is.na(ccdebt$orig_amt_1)==F & ccdebt$orig_amt_1 >0,]
nrow(ccdebt)
```

```
## [1] 54
```

```
# average amount of original cc debt  7726.037
mean(ccdebt$orig_amt_1)
```

```
## [1] 7726.037
```

```
#Median Amount of original cc debt  median(ccdebt$orig_amt_1)
median(ccdebt$orig_amt_1)
```

```
## [1] 6000
```

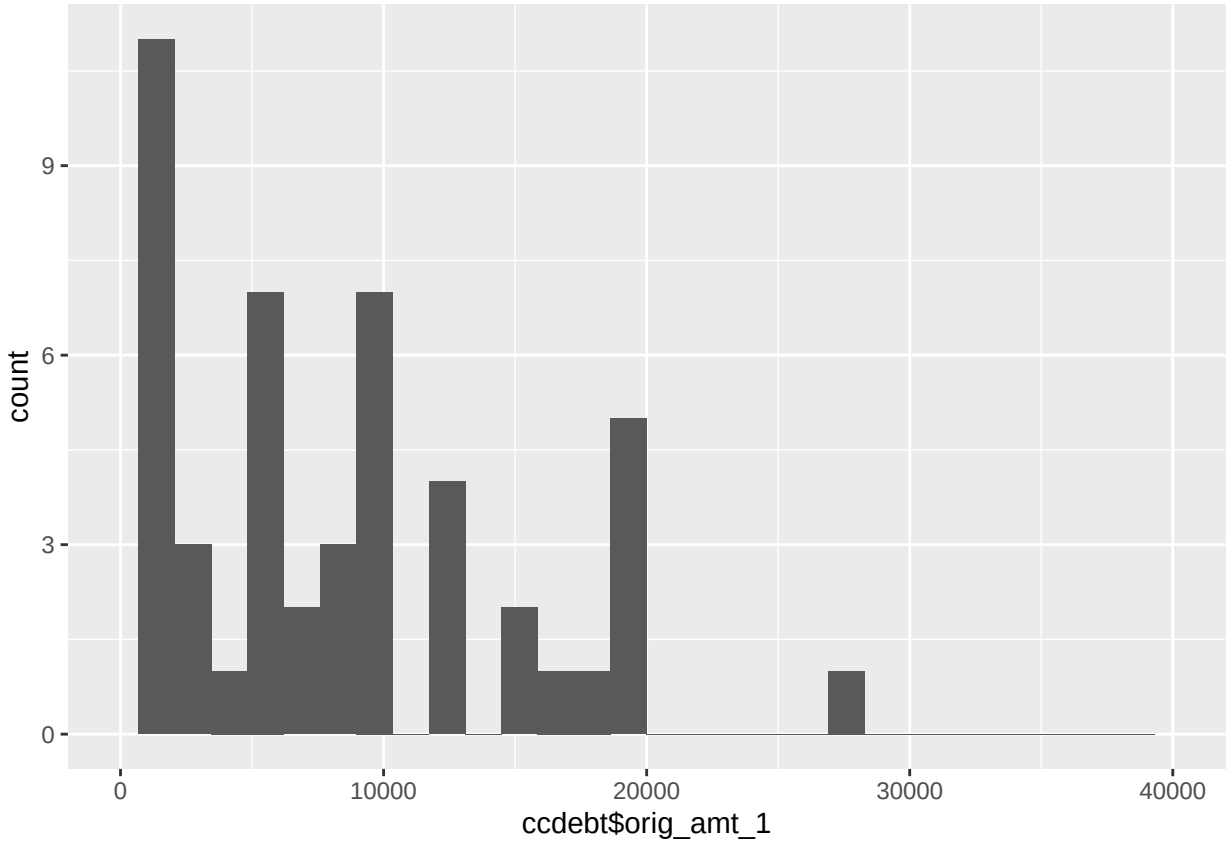
```
# sd original cc debt 6808.292
sd(ccdebt$orig_amt_1)
```

```
## [1] 6808.292
```

```
# bad attmept at plotting it
ggplot(ccdebt$orig_amt_1,data=ccdebt,mapping=aes(x=ccdebt$orig_amt_1))+geom_histogram()+xlim(0,40000)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 2 rows containing missing values (geom_bar).
```




```
# average amount of current debt 7232.778
mean(ccdebt$current_amt_1)
```

```
## [1] 7232.778
```

```
# median amount of current debt 4150
median(ccdebt$current_amt_1)
```

```
## [1] 4150
```

```
# sd curent cc debt even bigger sd  10029.83
sd(ccdebt$current_amt_1)
```

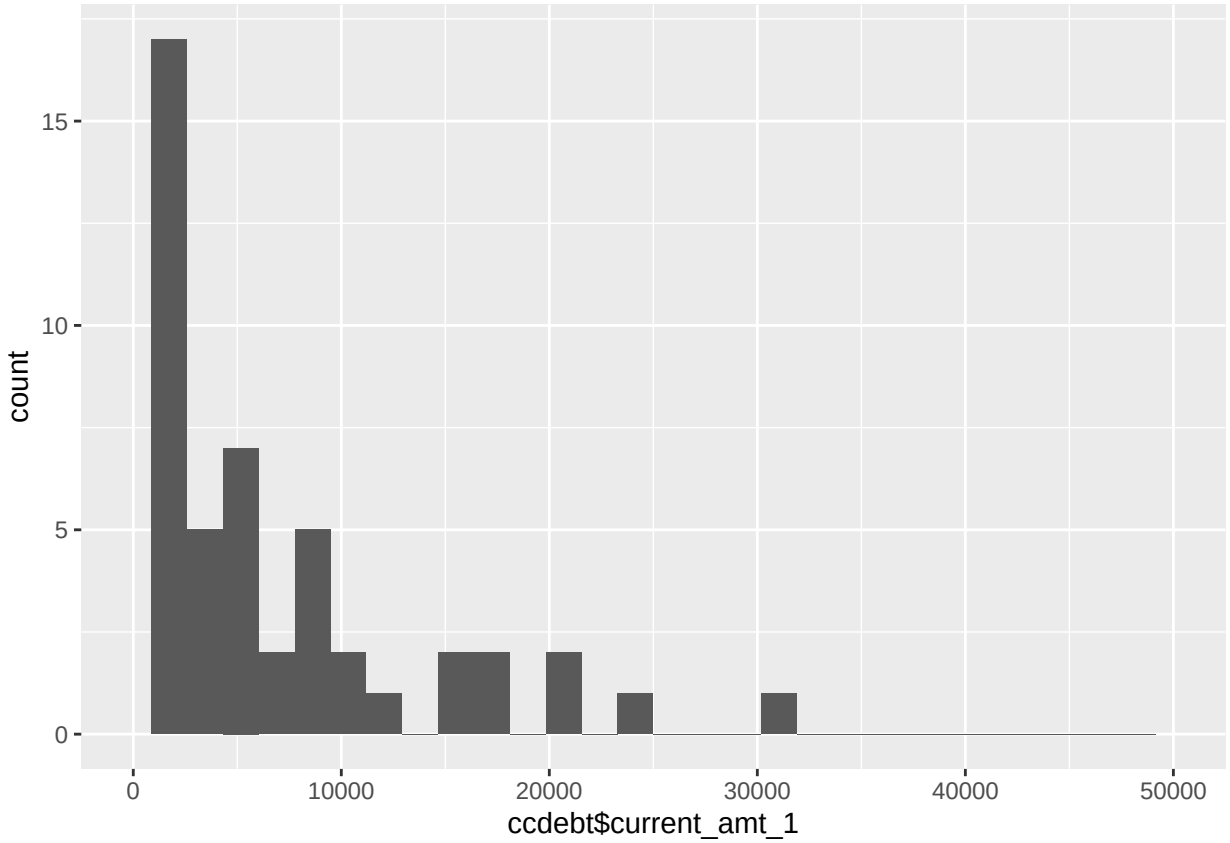
```
## [1] 10029.83
```

```
# Very left skewed
ggplot(ccdebt$current_amt_1,data=ccdebt,mapping=aes(x=ccdebt$current_amt_1))+geom_histogram()+xlim(0,50
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 2 rows containing missing values (geom_bar).
```

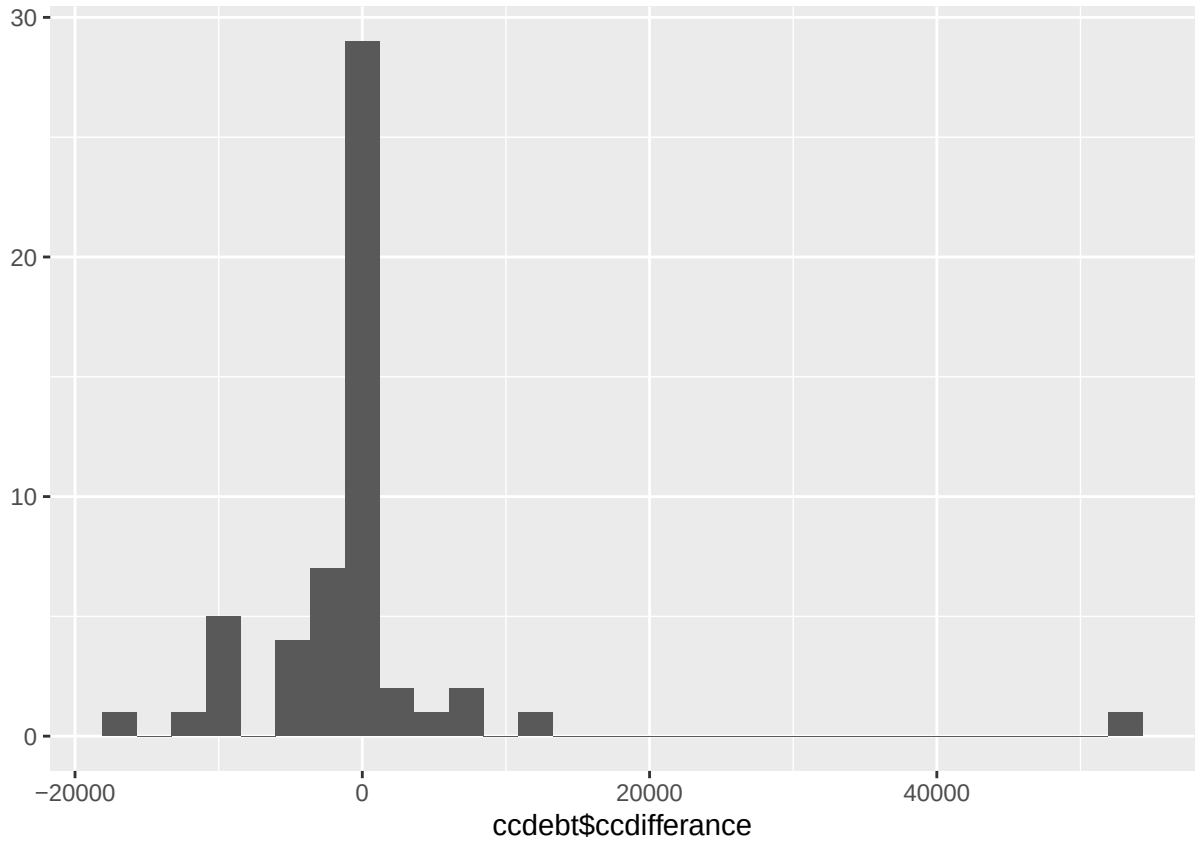



```
# Current amount related to original amount
ccdebt$ccdifferance<-ccdebt$current_amt_1-ccdebt$orig_amt_1
head(ccdebt$ccdifferance)[2]
```

```
## [1] -2000
```

```
qplot(ccdebt$ccdifferance)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# mean difference -493.2593, on average people paid off nearly 500 dollars of debt
mean(ccdebt$ccdifferance)
```

```
## [1] -493.2593
```

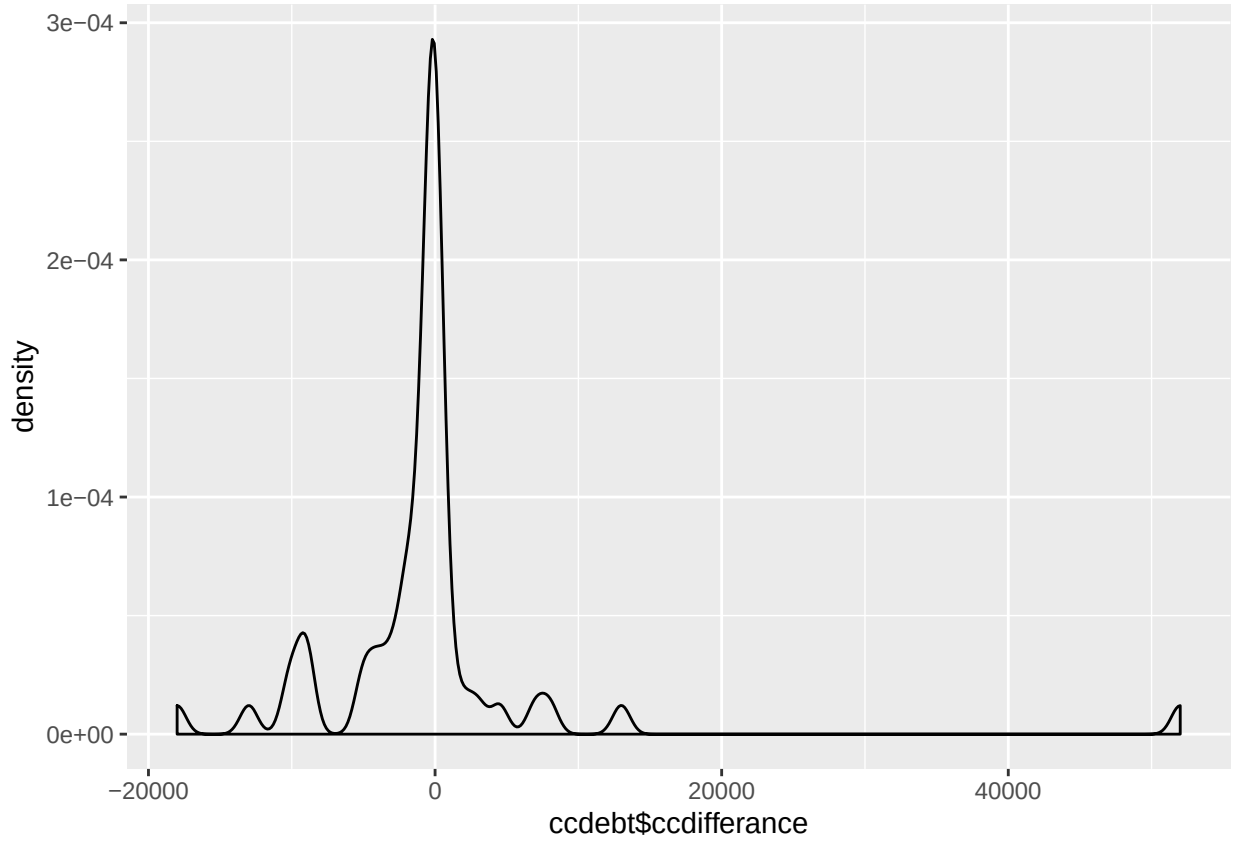
```
# median difference -343
median(ccdebt$ccdifferance)
```

```
## [1] -343
```

```
#sd differance a very big difference, outliers at work  8763.549
 sd(ccdebt$ccdifferance)
```

```
## [1] 8763.549
```

```
ggplot(ccdebt$ccdifferance,data=ccdebt,mapping=aes(x=ccdebt$ccdifferance))+geom_density()
```




```
sort(ccdebt$ccdifferance)
```

```
##  [1] -18000 -13000 -10200 -10000  -9000  -9000  -9000  -5000  -5000  -4000
## [11]  -4000  -3000  -3000  -2000  -2000  -2000  -2000  -1500  -1000  -1000
## [21]   -800   -700   -600   -500   -500   -500   -500   -186      0      0
## [31]      0      0      0      0      0      0      0      0      0      0
## [41]      0      0      0      0    200    650   1000   2000   3000   4500
## [51]   7000   8000  13000  52000
```

```
#Interest rates
ccdebt$int_rate_1
```

```
##  [1] 999   25    26.9  22    999   999   999   999   999   15    26
## [12] 12    999   999   22%   999   999   28.99 12    16    999   999
## [23] 999   26    25    999   999   999   999   5     999   999   29
## [34] 999   23%   11    999   16    999   999   25    999   24.75 999
## [45] 999   999   999   999   29    999   26    999   14.99 11
## 27 Levels:  0 11 12 13 14 14.99 15 16 19 20 21 22 22% 23% 24.75 25 ... 999
```

```
# some interest cc rates had % in them so we took them out and then converted the entire column to nume
ccdebt$int_rate_1<-as.numeric(gsub('%','',ccdebt$int_rate_1))
head(ccdebt$int_rate_1)[2]
```

```
## [1] 25
```

```
# Only 42.59% of people who had cc debt knew their interest rate
(sum(ccdebt$int_rate_1!=999.00)/nrow(ccdebt))*100
```

```
## [1] 42.59259
```

```r
#mean cc debt int rate , omit 999 of people who knew their interest rate mean=20.50565

mean(ccdebt$int_rate_1[ccdebt$int_rate_1!=999.00])
```

```
## [1] 20.50565
```

```r
#median of 23
median(ccdebt$int_rate_1[ccdebt$int_rate_1!=999.00])
```

```
## [1] 23
```

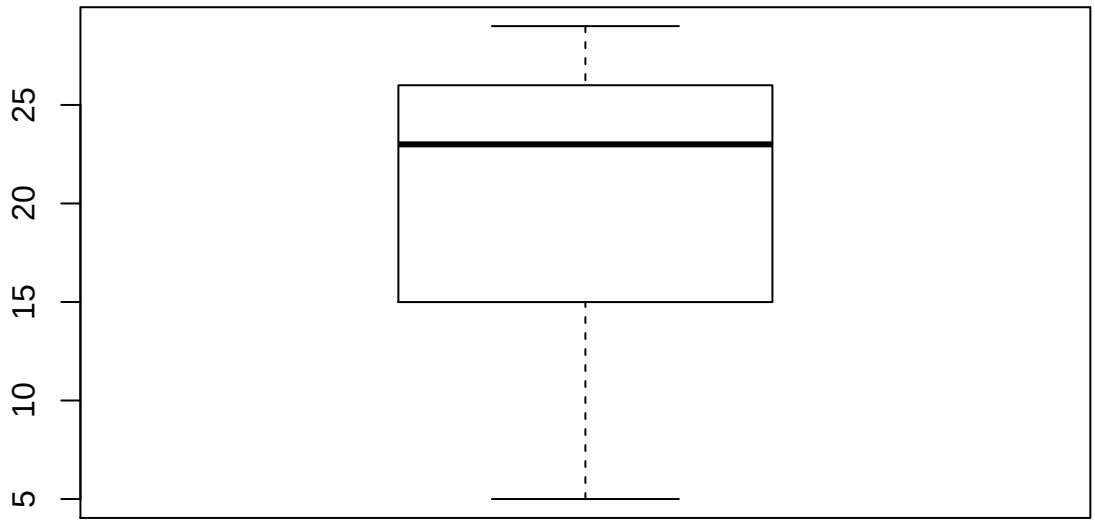
```r
# sd of 7.084%
sd(ccdebt$int_rate_1[ccdebt$int_rate_1!=999.00])
```

```
## [1] 7.084593
```

```r
help(boxplot)
```

```
## starting httpd help server ... done
```

```r
boxplot(ccdebt$int_rate_1[ccdebt$int_rate_1!=999.00],xlab=("Boxplot of ccdebt interest rates"))+ylim(-2
```



Boxplot of ccdebt interest rates

```
## NULL
```

```r
length(ccdebt$int_rate_1[ccdebt$int_rate_1!=999.00])
```

```
## [1] 23
```

```r
sum(ccdebt$int_rate_1[ccdebt$int_rate_1!=999.00]>20.5)
```

```
## [1] 14
```

```r
# The average APR charged in the second quarter of 2019 for credit card accounts that incurred interest
# https://www.nerdwallet.com/blog/credit-cards/what-is-a-good-apr-for-a-credit-card/

# number of people who have a cc int rate higher than national average 14, 23 people knew their int rat
sum(ccdebt$int_rate_1[ccdebt$int_rate_1!=999.00]>17.14)
```

```
## [1] 14
```

```r
length(ccdebt$int_rate_1[ccdebt$int_rate_1!=999.00])
```

```
## [1] 23
```

```r
# percent of people who have a > int rate than national average of people who knew their int rate 60.86
(sum(ccdebt$int_rate_1[ccdebt$int_rate_1!=999.00]>17.14)/length(ccdebt$int_rate_1[ccdebt$int_rate_1!=999
```

```
## [1] 60.86957
```

```r
# mean original debt of people with higher interest rates
# base mean of 7726.037
mean(ccdebt$orig_amt_1)
```

```
## [1] 7726.037
```

```r
#  original amount mean of  7228.571 for people who started off with high interest rate cards
mean(ccdebt$orig_amt_1[ccdebt$int_rate_1 !=999.00 & ccdebt$int_rate_1 >17.14])
```

```
## [1] 7228.571
```

```r
# the peoplle who have higher interest rates actually have 500 dollars lower debt on average
mean(ccdebt$orig_amt_1)-mean(ccdebt$orig_amt_1[ccdebt$int_rate_1 !=999.00 & ccdebt$int_rate_1 >17.14])
```

```
## [1] 497.4656
```

```r
# do people with lower interest rates on average start off with less or more debt

# average original amount of debt for all respondents  7726.037
mean(ccdebt$orig_amt_1)
```

```
## [1] 7726.037
```

```r
# only 9 people have int rates lower than the average
length(ccdebt$orig_amt_1[ccdebt$int_rate_1 !=999.00 & ccdebt$int_rate_1 <17.14])
```

```
## [1] 9
```

```r
head(ccdebt$orig_amt_1[ccdebt$int_rate_1 !=999.00 & ccdebt$int_rate_1 <17.14])[2]
```

```
## [1] 15000
```

```r
# mean of 11277.78
mean(ccdebt$orig_amt_1[ccdebt$int_rate_1 !=999.00 & ccdebt$int_rate_1 <17.14])
```

```
## [1] 11277.78
```

```r
# people who start off with a cc with less int rate have 3551.741 higher starting debt somehow
# only 9 people so probably not statistically signifigant

mean(ccdebt$orig_amt_1)-(mean(ccdebt$orig_amt_1[ccdebt$int_rate_1 !=999.00 & ccdebt$int_rate_1 <17.14]))
```

```
## [1] -3551.741
```

```r
# CURRENT AMOUNTs

# mean current amount of cc debt  7232.778
mean(ccdebt$current_amt_1)
```

```
## [1] 7232.778
```

```r
# difference in means between current and original amounts of cc debt
# people paid off 493 dollars on average
mean(ccdebt$current_amt_1)-mean(ccdebt$orig_amt_1)
```

```
## [1] -493.2593
```

```r
 # people who had higher interest rates had an average of 4707.143
mean(ccdebt$current_amt_1[ccdebt$int_rate_1!=999.00 & ccdebt$int_rate_1>17.14])
```

```
## [1] 4707.143
```

```r
# median of 4400
median(ccdebt$current_amt_1[ccdebt$int_rate_1!=999.00 & ccdebt$int_rate_1>17.14])
```

```
## [1] 4400
```

```r
# sd of 3223.105

sd(ccdebt$current_amt_1[ccdebt$int_rate_1!=999.00 & ccdebt$int_rate_1>17.14])
```

```
## [1] 3223.105
```

```r
#

# the people who had higher than average interest rates paid off -3018.894 of their debt on average
mean(ccdebt$current_amt_1[ccdebt$int_rate_1!=999.00&ccdebt$int_rate_1>17.14])-mean(ccdebt$orig_amt_1)
```

```
## [1] -3018.894
```

```r
# the people who had lower than average interest rates gained 3718.407 of their debt on average , maybe

mean(ccdebt$current_amt_1[ccdebt$int_rate_1!=999.00&ccdebt$int_rate_1<17.14])-mean(ccdebt$orig_amt_1)
```

```
## [1] 3718.407
```

```r
# only 9 respondents with a lower than average interest rate
length((ccdebt$current_amt_1[ccdebt$int_rate_1!=999.00&ccdebt$int_rate_1<17.14]))
```

```
## [1] 9
```

```r
# is it possible that people who do not know their interest rates have more debt, ie not knowing their
# 31 people do not know their interest rate
length(ccdebt$current_amt_1[ccdebt$int_rate_1==999.00])
```

```
## [1] 31
```

```r
# 57.40741 % of respondents with cc debt did not know their interest rates
((length(ccdebt$current_amt_1[ccdebt$int_rate_1==999.00]))/(length(ccdebt$int_rate_1)))*100
```

```
## [1] 57.40741
```

```r
# 6919.548 starting debt on average for people who did not know their interest rate
mean(ccdebt$orig_amt_1[ccdebt$int_rate_1==999.00])
```

```
## [1] 6919.548
```

```r
# 5000 median starting debt for people who did not know their interest rate
median(ccdebt$orig_amt_1[ccdebt$int_rate_1==999.00])
```

```
## [1] 5000
```

```r
# sd of 7215.203 for people who did not know their interest rate
sd(ccdebt$orig_amt_1[ccdebt$int_rate_1==999.00])
```

```
## [1] 7215.203
```

```r
#   7150.645 current ccdebt for people who do not know their interest rate
mean(ccdebt$current_amt_1[ccdebt$int_rate_1==999.00])
```

```
## [1] 7150.645
```

```r
# people who do not know their interest rate gained 231 dollars of cc debt on average
mean(ccdebt$current_amt_1[ccdebt$int_rate_1==999.00])-mean(ccdebt$orig_amt_1[ccdebt$int_rate_1==999.00])
```

```
## [1] 231.0968
```

```r
# Attempt at t tests

# is there a difference in people with higher interest rates compared to people with lower interest rat
# we reject the hypothesis that there is no difference in people with higher/lower interest rates, numb
t.test(ccdebt$current_amt_1[ccdebt$int_rate_1!=999.00&ccdebt$int_rate_1>17.14],ccdebt$current_amt_1[ccd
```

```
##
##  Welch Two Sample t-test
##
## data:  ccdebt$current_amt_1[ccdebt$int_rate_1 != 999 & ccdebt$int_rate_1 >  and ccdebt$current_amt_1
## t = -2.408, df = 9.6977, p-value = 0.03753
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -12997.9383   -476.6649
## sample estimates:
## mean of x mean of y
##  4707.143 11444.444
```

```r
# is there a difference between the people who know their interest rates and the ones who do not

# we fail to reject that there is a difference in current amounts of debt for people who know their int

t.test(ccdebt$current_amt_1[ccdebt$int_rate_1!=999.00],
       ccdebt$current_amt_1[ccdebt$int_rate_1==999.00])
```

```
##
##  Welch Two Sample t-test
##
## data:  ccdebt$current_amt_1[ccdebt$int_rate_1 != 999] and ccdebt$current_amt_1[ccdebt$int_rate_1 == 9
## t = 0.075416, df = 47.457, p-value = 0.9402
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4949.740  5335.406
## sample estimates:
## mean of x mean of y
##  7343.478  7150.645
```

```r
# wealth effect?
# average response of 3/7 on wealth likert scale

mean(ccdebt$wealthy)
```

```
## [1] 2.907407
```

```r
str(ccdebt$wealthy)
```

```
##  int [1:54] 3 2 3 4 2 2 3 4 6 5 ...
```

```r
ccdebt$wealthy<-as.factor(ccdebt$wealthy)
ccdebt$wealthy
```

```
##  [1] 3 2 3 4 2 2 3 4 6 5 2 6 6 6 3 1 1 4 1 4 2 3 1 1 3 1 3 3 1 3 4 4 2 1 1
## [36] 4 2 2 2 2 1 4 3 1 3 3 1 2 5 6 4 2 3 6
## Levels: 1 2 3 4 5 6
```

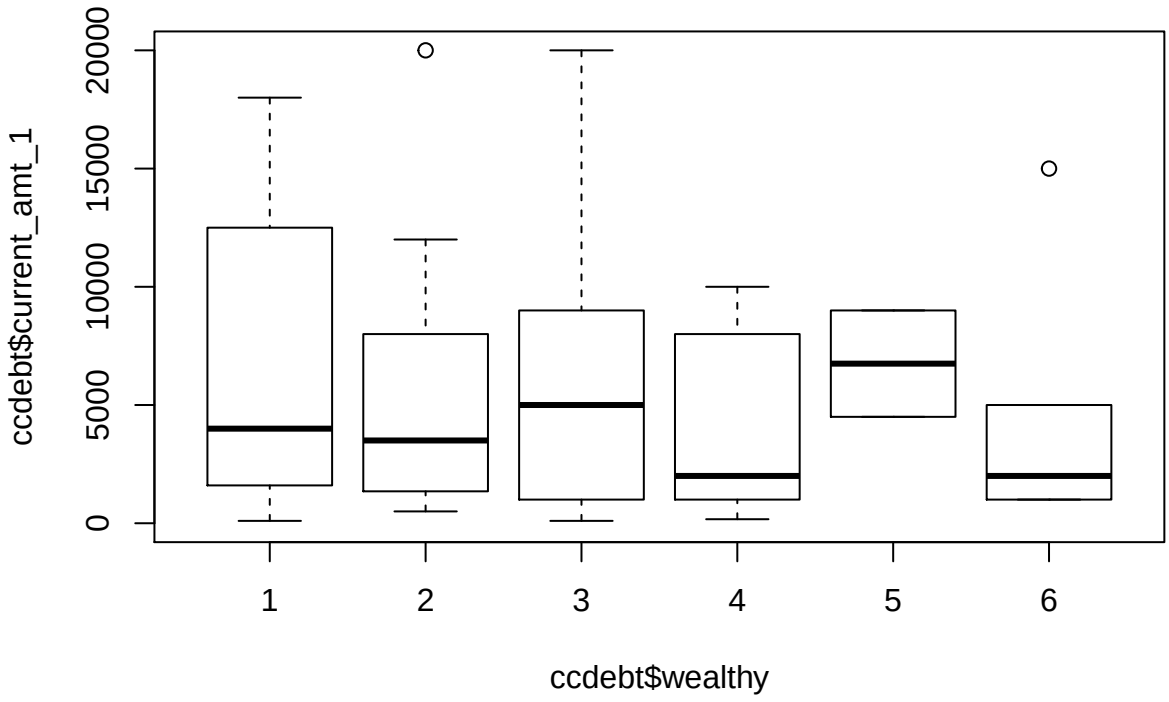
```r
summary(lm(ccdebt$current_amt_1~ccdebt$wealthy))
```

```
##
## Call:
## lm(formula = ccdebt$current_amt_1 ~ ccdebt$wealthy)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -11285  -5177  -2942   2133  48615
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       6708.33    2949.26   2.275   0.0274 *
## ccdebt$wealthy2 -1158.33    4170.89  -0.278   0.7824
## ccdebt$wealthy3  4676.28    4089.89   1.143   0.2586
## ccdebt$wealthy4  -489.44    4505.07  -0.109   0.9139
## ccdebt$wealthy5    41.67    7803.01   0.005   0.9958
## ccdebt$wealthy6 -2375.00    5108.27  -0.465   0.6441
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10220 on 48 degrees of freedom
## Multiple R-squared:  0.06031,    Adjusted R-squared:  -0.03758
## F-statistic: 0.6161 on 5 and 48 DF,  p-value: 0.6881
```

```r
aggregate(ccdebt$current_amt_1~ccdebt$wealthy,FUN = "mean")
```

```
##   ccdebt$wealthy ccdebt$current_amt_1
## 1              1             6708.333
## 2              2             5550.000
## 3              3            11384.615
## 4              4             6218.889
## 5              5             6750.000
## 6              6             4333.333
```

```r
help(boxplot)
# wealth level three seems to have some outliers
boxplot(ccdebt$current_amt_1~ccdebt$wealthy,ylim=c(0,20000))
```




```
# paid off balance in full every month=fullbalance
ccdebt$fullBalance
```

```
##  [1] 2 2 2 2 1 2 1 2 1 1 2 1 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2
## [36] 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1
```

```
nrow(ccdebt)
```

```
## [1] 54
```

```
# changing 2s and 1s to binary values
for (i in 1:nrow(ccdebt)){
  if (ccdebt$fullBalance[i]==2){
    ccdebt$fullBalance[i]<-1
  }
  else{
    ccdebt$fullBalance[i]<-0
  }
}


# wealth , interest rate , and whether or not the respondent pays of their cc are not signifigant predi
summary(lm(ccdebt$current_amt_1~ccdebt$wealthy+ccdebt$int_rate_1 + ccdebt$fullBalance))
```

```
##
## Call:
## lm(formula = ccdebt$current_amt_1 ~ ccdebt$wealthy + ccdebt$int_rate_1 +
##     ccdebt$fullBalance)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -11321  -5237  -2823   2769  48295
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         5490.1104  5237.0386   1.048    0.300
## ccdebt$wealthy2    -1260.7635  4269.4246  -0.295    0.769
## ccdebt$wealthy3     4697.3473  4191.4584   1.121    0.268
## ccdebt$wealthy4     -494.0953  4646.1161  -0.106    0.916
## ccdebt$wealthy5      640.3498  8389.7381   0.076    0.939
## ccdebt$wealthy6    -1760.9050  5587.3340  -0.315    0.754
## ccdebt$int_rate_1      0.2917     3.1199   0.093    0.926
## ccdebt$fullBalance  1226.2455  4018.3470   0.305    0.762
##
## Residual standard error: 10430 on 46 degrees of freedom
## Multiple R-squared:  0.06223,    Adjusted R-squared:  -0.08048
## F-statistic: 0.4361 on 7 and 46 DF,  p-value: 0.8744
```

```
# even when trying to remove outliers non of our variables are signifigant predictiors of current amoun
summary(lm(ccdebt$current_amt_1[ccdebt$current_amt_1<20001]
           ~ccdebt$wealthy[ccdebt$current_amt_1<20001]
           +ccdebt$int_rate_1[ccdebt$current_amt_1<20001]
           +ccdebt$fullBalance[ccdebt$current_amt_1<20001]))
```

```
##
## Call:
## lm(formula = ccdebt$current_amt_1[ccdebt$current_amt_1 < 20001] ~
##     ccdebt$wealthy[ccdebt$current_amt_1 < 20001] + ccdebt$int_rate_1[ccdebt$current_amt_1 <
##         20001] + ccdebt$fullBalance[ccdebt$current_amt_1 < 20001])
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7144  -3961  -1338   1984  14862
##
## Coefficients:
##                                                  Estimate Std. Error
## (Intercept)                                      9897.080   2875.701
## ccdebt$wealthy[ccdebt$current_amt_1 < 20001]2   -1016.045   2304.921
## ccdebt$wealthy[ccdebt$current_amt_1 < 20001]3   -2105.225   2378.581
## ccdebt$wealthy[ccdebt$current_amt_1 < 20001]4   -3197.449   2583.851
## ccdebt$wealthy[ccdebt$current_amt_1 < 20001]5   -2248.199   4545.091
## ccdebt$wealthy[ccdebt$current_amt_1 < 20001]6   -3224.300   3019.935
## ccdebt$int_rate_1[ccdebt$current_amt_1 < 20001]    -2.656      1.745
## ccdebt$fullBalance[ccdebt$current_amt_1 < 20001] -1680.893   2189.380
##                                                 t value Pr(>|t|)
## (Intercept)                                       3.442   0.0013 **
## ccdebt$wealthy[ccdebt$current_amt_1 < 20001]2    -0.441   0.6616
## ccdebt$wealthy[ccdebt$current_amt_1 < 20001]3    -0.885   0.3810
## ccdebt$wealthy[ccdebt$current_amt_1 < 20001]4    -1.237   0.2226
## ccdebt$wealthy[ccdebt$current_amt_1 < 20001]5    -0.495   0.6234
## ccdebt$wealthy[ccdebt$current_amt_1 < 20001]6    -1.068   0.2916
## ccdebt$int_rate_1[ccdebt$current_amt_1 < 20001]  -1.522   0.1353
## ccdebt$fullBalance[ccdebt$current_amt_1 < 20001] -0.768   0.4468
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5628 on 43 degrees of freedom
## Multiple R-squared:  0.08732,    Adjusted R-squared:  -0.06126
## F-statistic: 0.5877 on 7 and 43 DF,  p-value: 0.7621
```