Exploratory_Earnings

Krishanu Banerjee

Thursday, September 29, 2016

There are total 73 fields in Earnings data. I started analysis for 2013 data but no data. Same for 2012 and same thing again. There are only few fields populated for 2011. Probably for recent years, earnings data (for example, students earning 6 years after entry) not available. So I have to focus on 2010 and 2009 data.

Data loading and NULL fields

```
library(RSQLite)

library('dplyr')

library('ggplot2')

library(mice)

library(VIM)

library(tigerstats)

db <- dbConnect(dbDriver("SQLite"),
"D:/college_score/output/database.sqlite")

tmp_2010<- dbGetQuery(db, "
                  select Year,lower(INSTNM)||'-'||ZIP as INSTNM
,count_ed
,count_nwne_p10
                  ,count_wne_p10
                  ,mn_earn_wne_p10
                  ,md_earn_wne_p10
                  ,pct10_earn_wne_p10
                  ,pct25_earn_wne_p10
                  ,pct75_earn_wne_p10
                  ,pct90_earn_wne_p10
                  ,sd_earn_wne_p10
                  ,count_wne_inc1_p10
                  ,count_wne_inc2_p10
                  ,count_wne_inc3_p10
                  ,count_wne_indep0_inc1_p10
                  ,count_wne_indep0_p10
                  ,count_wne_indep1_p10
                  ,count_wne_male0_p10
                  ,count_wne_male1_p10
                  ,gt_25k_p10
```

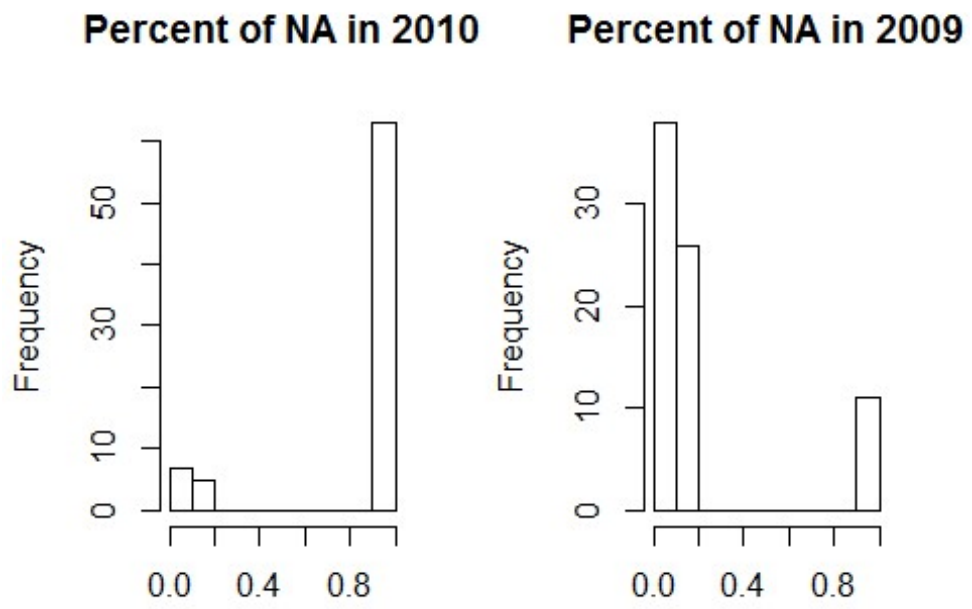
```
                  ,mn_earn_wne_inc1_p10
                  ,mn_earn_wne_inc2_p10
                  ,mn_earn_wne_inc3_p10
,mn_earn_wne_indep0_inc1_p10
,mn_earn_wne_indep0_p10
                  ,mn_earn_wne_indep1_p10
                  ,mn_earn_wne_male0_p10
                  ,mn_earn_wne_male1_p10
                  ,count_nwne_p6
                  ,count_wne_p6
                  ,mn_earn_wne_p6
                  ,md_earn_wne_p6
                  ,pct10_earn_wne_p6
                  ,pct25_earn_wne_p6
                  ,pct75_earn_wne_p6
                  ,pct90_earn_wne_p6
                  ,sd_earn_wne_p6
                  ,count_wne_inc1_p6
                  ,count_wne_inc2_p6
                  ,count_wne_inc3_p6
                  ,count_wne_indep0_inc1_p6
                  ,count_wne_indep0_p6
                  ,count_wne_indep1_p6
                  ,count_wne_male0_p6
                  ,count_wne_male1_p6
                  ,gt_25k_p6
                  ,mn_earn_wne_inc1_p6
                  ,mn_earn_wne_inc2_p6
                  ,mn_earn_wne_inc3_p6
                  ,mn_earn_wne_indep0_inc1_p6
                  ,mn_earn_wne_indep0_p6
                  ,mn_earn_wne_indep1_p6
                  ,mn_earn_wne_male0_p6
                  ,mn_earn_wne_male1_p6
                  ,count_nwne_p7
                  ,count_wne_p7
                  ,mn_earn_wne_p7
                  ,sd_earn_wne_p7
                  ,gt_25k_p7
                  ,count_nwne_p8
                  ,count_wne_p8
                  ,mn_earn_wne_p8
                  ,md_earn_wne_p8
                  ,pct10_earn_wne_p8
                  ,pct25_earn_wne_p8
                  ,pct75_earn_wne_p8
                  ,pct90_earn_wne_p8
                  ,sd_earn_wne_p8
                  ,gt_25k_p8
                  ,count_nwne_p9
```

```
                  ,count_wne_p9
                  ,mn_earn_wne_p9
                  ,sd_earn_wne_p9
                  ,gt_25k_p9
    from Scorecard
where Year in (2010)
                  ")

tmp_2009<- dbGetQuery(db, "
                  select Year,lower(INSTNM)||'-'||ZIP as INSTNM
,count_ed
,count_nwne_p10
                  ,count_wne_p10
                  ,mn_earn_wne_p10
                  ,md_earn_wne_p10
                  ,pct10_earn_wne_p10
                  ,pct25_earn_wne_p10
                  ,pct75_earn_wne_p10
                  ,pct90_earn_wne_p10
                  ,sd_earn_wne_p10
                  ,count_wne_inc1_p10
                  ,count_wne_inc2_p10
                  ,count_wne_inc3_p10
                  ,count_wne_indep0_inc1_p10
                  ,count_wne_indep0_p10
                  ,count_wne_indep1_p10
                  ,count_wne_male0_p10
                  ,count_wne_male1_p10
                  ,gt_25k_p10
                  ,mn_earn_wne_inc1_p10
                  ,mn_earn_wne_inc2_p10
                  ,mn_earn_wne_inc3_p10
,mn_earn_wne_indep0_inc1_p10
,mn_earn_wne_indep0_p10
                  ,mn_earn_wne_indep1_p10
                  ,mn_earn_wne_male0_p10
                  ,mn_earn_wne_male1_p10
                  ,count_nwne_p6
                  ,count_wne_p6
                  ,mn_earn_wne_p6
                  ,md_earn_wne_p6
                  ,pct10_earn_wne_p6
                  ,pct25_earn_wne_p6
                  ,pct75_earn_wne_p6
                  ,pct90_earn_wne_p6
                  ,sd_earn_wne_p6
                  ,count_wne_inc1_p6
                  ,count_wne_inc2_p6
                  ,count_wne_inc3_p6
                  ,count_wne_indep0_inc1_p6
```

```
                  ,count_wne_indep0_p6
                  ,count_wne_indep1_p6
                  ,count_wne_male0_p6
                  ,count_wne_male1_p6
                  ,gt_25k_p6
                  ,mn_earn_wne_inc1_p6
                  ,mn_earn_wne_inc2_p6
                  ,mn_earn_wne_inc3_p6
                  ,mn_earn_wne_indep0_inc1_p6
                  ,mn_earn_wne_indep0_p6
                  ,mn_earn_wne_indep1_p6
                  ,mn_earn_wne_male0_p6
                  ,mn_earn_wne_male1_p6
                  ,count_nwne_p7
                  ,count_wne_p7
                  ,mn_earn_wne_p7
                  ,sd_earn_wne_p7
                  ,gt_25k_p7
                  ,count_nwne_p8
                  ,count_wne_p8
                  ,mn_earn_wne_p8
                  ,md_earn_wne_p8
                  ,pct10_earn_wne_p8
                  ,pct25_earn_wne_p8
                  ,pct75_earn_wne_p8
                  ,pct90_earn_wne_p8
                  ,sd_earn_wne_p8
                  ,gt_25k_p8
                  ,count_nwne_p9
                  ,count_wne_p9
                  ,mn_earn_wne_p9
                  ,sd_earn_wne_p9
                  ,gt_25k_p9
    from Scorecard
where Year in (2009)
                   ")

par(mfrow=c(1,2))
hist(sapply(tmp_2010,function(x)
sum(is.na(x))/nrow(tmp_2010)),breaks=10,main='Percent of NA in 2010')
hist(sapply(tmp_2009,function(x)
sum(is.na(x))/nrow(tmp_2009)),breaks=10,main='Percent of NA in 2009')
```



p_2010, function(x) sum(is.na(x))/nrp_2009, function(x) sum(is.na(x))/nr

We can see only few fields (12) are not entirely null for 2010 but 2009 has more not null fields.

2010 missing value imputation –

For missing value imputation, I am using multivariate chain equation based package and predictive mean matching method (pmm). I have ignored the fields with more than 60% NA. But many fields are completely NA.

```
v=c()
v=sapply(tmp_2010,function(x) sum(is.na(x))/nrow(tmp_2010))
which(v <.6)
```

```
##           Year         INSTNM  count_nwne_p7   count_wne_p7 mn_earn_wne_p7
##              1              2             56             57             58
## sd_earn_wne_p7      gt_25k_p7  count_nwne_p9   count_wne_p9 mn_earn_wne_p9
##             59             60             71             72             73
## sd_earn_wne_p9      gt_25k_p9
##             74             75
```

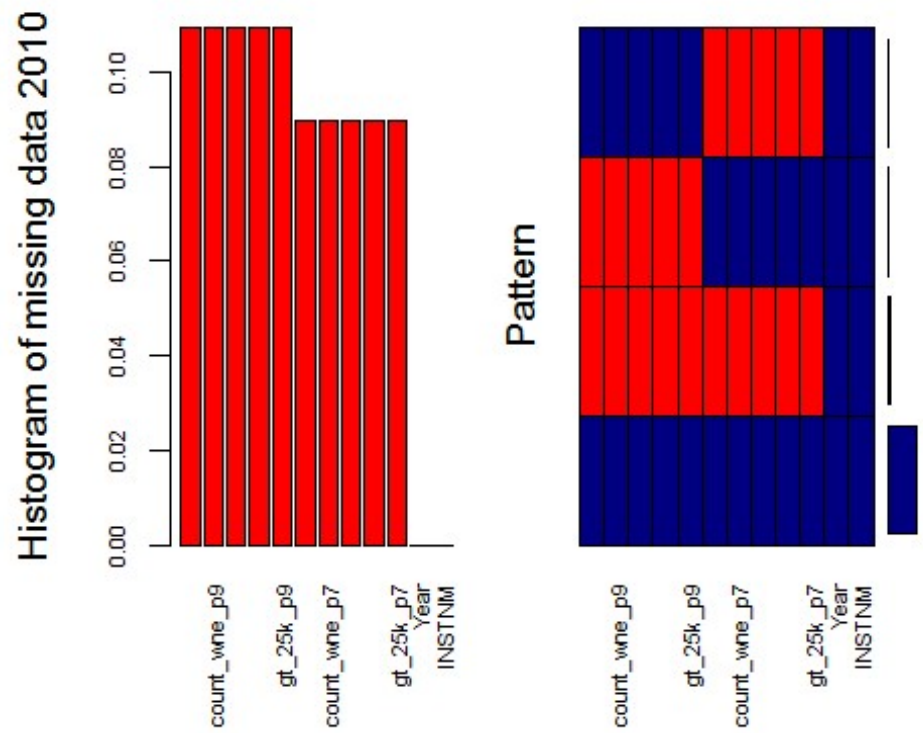
```
EAR_data=tmp_2010[,which(v <.6)]
sapply(EAR_data,function(x) sum(is.na(x))/nrow(EAR_data))
```

```
##           Year         INSTNM  count_nwne_p7   count_wne_p7 mn_earn_wne_p7
##     0.00000000     0.00000000     0.08956029     0.08956029     0.08956029
## sd_earn_wne_p7      gt_25k_p7  count_nwne_p9   count_wne_p9 mn_earn_wne_p9
##     0.08956029     0.08956029     0.10925277     0.10925277     0.10925277
```

```
## sd_earn_wne_p9      gt_25k_p9
##     0.10925277     0.10925277
```

```
aggr_plot <- aggr(EAR_data, col=c('navyblue','red'), numbers=TRUE,
sortVars=TRUE,
                  labels=names(data), cex.axis=.7, gap=3,
                  ylab=c("Histogram of missing data 2010","Pattern"))
```

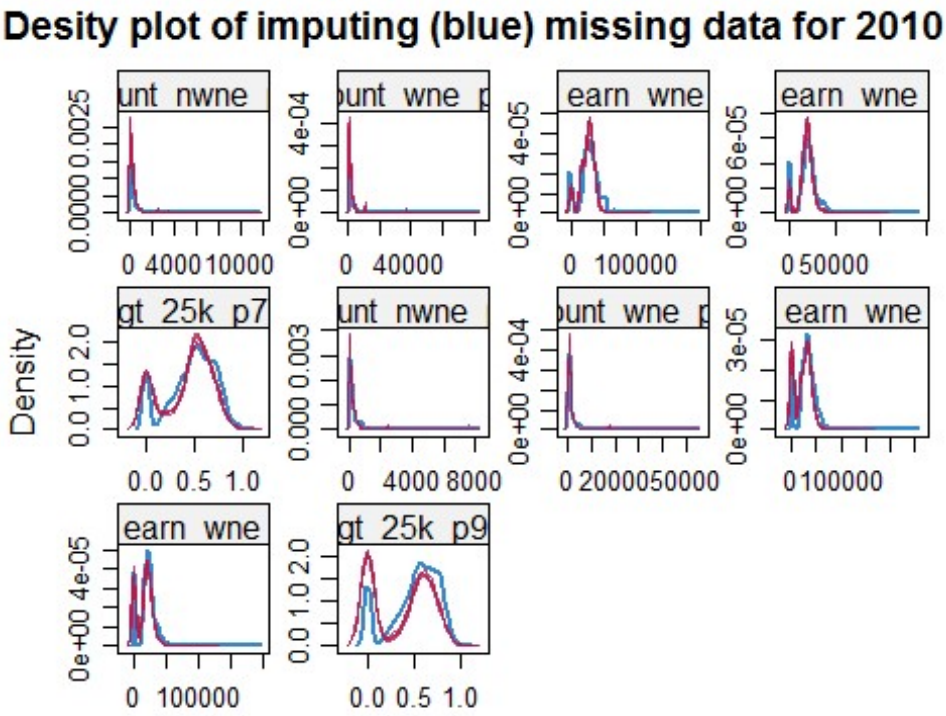
```
## Warning in plot.aggr(res, ...): not enough horizontal space to display
## frequencies
```




```
##
##  Variables sorted by number of missings:
##        Variable      Count
##   count_nwne_p9 0.10925277
##    count_wne_p9 0.10925277
##  mn_earn_wne_p9 0.10925277
##  sd_earn_wne_p9 0.10925277
##       gt_25k_p9 0.10925277
##   count_nwne_p7 0.08956029
##    count_wne_p7 0.08956029
##  mn_earn_wne_p7 0.08956029
##  sd_earn_wne_p7 0.08956029
##       gt_25k_p7 0.08956029
##            Year 0.00000000
##          INSTNM 0.00000000
```

```
tempData <- mice(EAR_data,m=5,maxit=20,meth='pmm',seed=500,verbose=FALSE)
```

```
densityplot(tempData,main='Desity plot of imputing (blue) missing data for
2010')
```




```
EAR=complete(tempData,1)
```

2009 missing value imputation

```
v_2009=c()
v_2009=sapply(tmp_2009,function(x) sum(is.na(x))/nrow(tmp_2009))
which(v_2009 <.6)
```

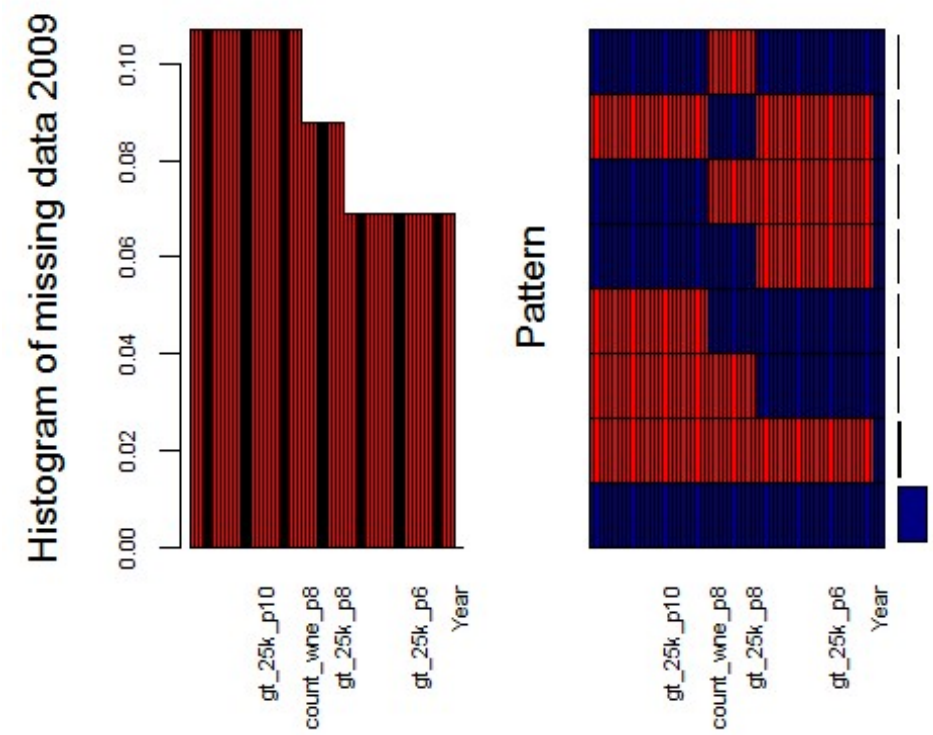
```
##                          Year                        INSTNM
##                             1                             2
##               count_nwne_p10               count_wne_p10
##                             4                             5
##              mn_earn_wne_p10              md_earn_wne_p10
##                             6                             7
##           pct10_earn_wne_p10           pct25_earn_wne_p10
##                             8                             9
##           pct75_earn_wne_p10           pct90_earn_wne_p10
##                            10                            11
##              sd_earn_wne_p10           count_wne_inc1_p10
##                            12                            13
##           count_wne_inc2_p10           count_wne_inc3_p10
##                            14                            15
##    count_wne_indep0_inc1_p10         count_wne_indep0_p10
##                            16                            17
```

```
##        count_wne_indep1_p10         count_wne_male0_p10
##                          18                          19
##         count_wne_male1_p10                 gt_25k_p10
##                          20                          21
##        mn_earn_wne_inc1_p10        mn_earn_wne_inc2_p10
##                          22                          23
##        mn_earn_wne_inc3_p10 mn_earn_wne_indep0_inc1_p10
##                          24                          25
##      mn_earn_wne_indep0_p10      mn_earn_wne_indep1_p10
##                          26                          27
##       mn_earn_wne_male0_p10       mn_earn_wne_male1_p10
##                          28                          29
##               count_nwne_p6               count_wne_p6
##                          30                          31
##             mn_earn_wne_p6             md_earn_wne_p6
##                          32                          33
##           pct10_earn_wne_p6           pct25_earn_wne_p6
##                          34                          35
##           pct75_earn_wne_p6           pct90_earn_wne_p6
##                          36                          37
##              sd_earn_wne_p6           count_wne_inc1_p6
##                          38                          39
##           count_wne_inc2_p6           count_wne_inc3_p6
##                          40                          41
##    count_wne_indep0_inc1_p6         count_wne_indep0_p6
##                          42                          43
##         count_wne_indep1_p6          count_wne_male0_p6
##                          44                          45
##          count_wne_male1_p6                  gt_25k_p6
##                          46                          47
##         mn_earn_wne_inc1_p6         mn_earn_wne_inc2_p6
##                          48                          49
##         mn_earn_wne_inc3_p6  mn_earn_wne_indep0_inc1_p6
##                          50                          51
##       mn_earn_wne_indep0_p6       mn_earn_wne_indep1_p6
##                          52                          53
##        mn_earn_wne_male0_p6        mn_earn_wne_male1_p6
##                          54                          55
##               count_nwne_p8               count_wne_p8
##                          61                          62
##             mn_earn_wne_p8             md_earn_wne_p8
##                          63                          64
##           pct10_earn_wne_p8           pct25_earn_wne_p8
##                          65                          66
##           pct75_earn_wne_p8           pct90_earn_wne_p8
##                          67                          68
##              sd_earn_wne_p8                  gt_25k_p8
##                          69                          70
```

```
EAR_data_2009=tmp_2009[,which(v_2009 <.6)]
sapply(EAR_data_2009,function(x) sum(is.na(x))/nrow(EAR_data_2009))
```

```
aggr_plot_2009 <- aggr(EAR_data_2009, col=c('navyblue','red'), numbers=TRUE,
sortVars=TRUE,
                  labels=names(data), cex.axis=.7, gap=3,
                  ylab=c("Histogram of missing data 2009","Pattern"))
```

```
## Warning in plot.aggr(res, ...): not enough horizontal space to display
## frequencies
```




```
##
##  Variables sorted by number of missings:
##                       Variable      Count
##                count_nwne_p10 0.10700797
##                 count_wne_p10 0.10700797
##               mn_earn_wne_p10 0.10700797
##               md_earn_wne_p10 0.10700797
##            pct10_earn_wne_p10 0.10700797
##            pct25_earn_wne_p10 0.10700797
##            pct75_earn_wne_p10 0.10700797
##            pct90_earn_wne_p10 0.10700797
##               sd_earn_wne_p10 0.10700797
##            count_wne_inc1_p10 0.10700797
##            count_wne_inc2_p10 0.10700797
##            count_wne_inc3_p10 0.10700797
##    count_wne_indep0_inc1_p10 0.10700797
```

```
##         count_wne_indep0_p10 0.10700797
##         count_wne_indep1_p10 0.10700797
##          count_wne_male0_p10 0.10700797
##          count_wne_male1_p10 0.10700797
##                   gt_25k_p10 0.10700797
##         mn_earn_wne_inc1_p10 0.10700797
##         mn_earn_wne_inc2_p10 0.10700797
##         mn_earn_wne_inc3_p10 0.10700797
##  mn_earn_wne_indep0_inc1_p10 0.10700797
##       mn_earn_wne_indep0_p10 0.10700797
##       mn_earn_wne_indep1_p10 0.10700797
##        mn_earn_wne_male0_p10 0.10700797
##        mn_earn_wne_male1_p10 0.10700797
##                count_nwne_p8 0.08798433
##                 count_wne_p8 0.08798433
##               mn_earn_wne_p8 0.08798433
##               md_earn_wne_p8 0.08798433
##            pct10_earn_wne_p8 0.08798433
##            pct25_earn_wne_p8 0.08798433
##            pct75_earn_wne_p8 0.08798433
##            pct90_earn_wne_p8 0.08798433
##               sd_earn_wne_p8 0.08798433
##                    gt_25k_p8 0.08798433
##                count_nwne_p6 0.06882081
##                 count_wne_p6 0.06882081
##               mn_earn_wne_p6 0.06882081
##               md_earn_wne_p6 0.06882081
##            pct10_earn_wne_p6 0.06882081
##            pct25_earn_wne_p6 0.06882081
##            pct75_earn_wne_p6 0.06882081
##            pct90_earn_wne_p6 0.06882081
##               sd_earn_wne_p6 0.06882081
##            count_wne_inc1_p6 0.06882081
##            count_wne_inc2_p6 0.06882081
##            count_wne_inc3_p6 0.06882081
##     count_wne_indep0_inc1_p6 0.06882081
##          count_wne_indep0_p6 0.06882081
##          count_wne_indep1_p6 0.06882081
##           count_wne_male0_p6 0.06882081
##           count_wne_male1_p6 0.06882081
##                    gt_25k_p6 0.06882081
##          mn_earn_wne_inc1_p6 0.06882081
##          mn_earn_wne_inc2_p6 0.06882081
##          mn_earn_wne_inc3_p6 0.06882081
##   mn_earn_wne_indep0_inc1_p6 0.06882081
##        mn_earn_wne_indep0_p6 0.06882081
##        mn_earn_wne_indep1_p6 0.06882081
##         mn_earn_wne_male0_p6 0.06882081
##         mn_earn_wne_male1_p6 0.06882081
```

```
##                          Year 0.00000000
##                        INSTNM 0.00000000
```

```
tempData_2009 <-
mice(EAR_data_2009,m=5,maxit=10,meth='pmm',seed=500,verbose=FALSE)
```

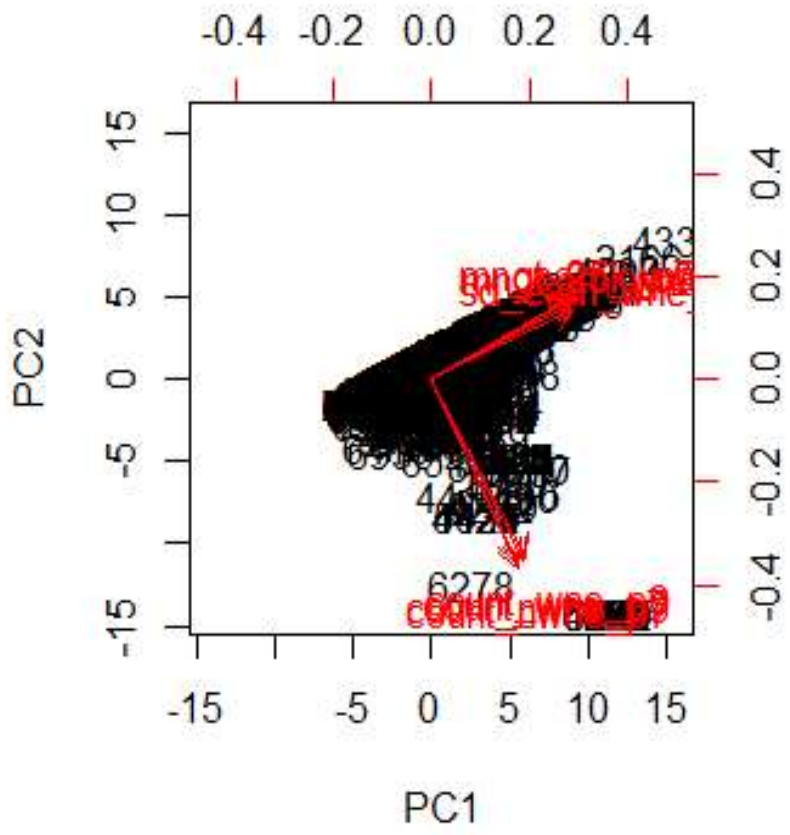
```
EAR_2009=complete(tempData_2009,1)
```

PCA for 2010

```
EAR_PCA=EAR[-c(1,2)]
pr.out=prcomp(EAR_PCA,scale=TRUE)
pr.out$rotation[,1:4]
```

```
##                       PC1        PC2         PC3          PC4
## count_nwne_p7  0.2216374 -0.4555887  0.02815083  0.013486693
## count_wne_p7   0.2358061 -0.4420886 -0.03888262  0.022415555
## mn_earn_wne_p7 0.3695422  0.1994203  0.02792210  0.446001334
## sd_earn_wne_p7 0.3585872  0.1711851  0.45315535  0.458206844
## gt_25k_p7      0.3553853  0.2012610 -0.53189705  0.306501743
## count_nwne_p9  0.2213852 -0.4441474  0.02373958 -0.020912149
## count_wne_p9   0.2423653 -0.4325407 -0.06206705 -0.003806183
## mn_earn_wne_p9 0.3763761  0.1911103  0.08174228 -0.368670850
## sd_earn_wne_p9 0.3584117  0.1715116  0.49372578 -0.453749736
## gt_25k_p9      0.3549044  0.1926656 -0.50377470 -0.392737581
```

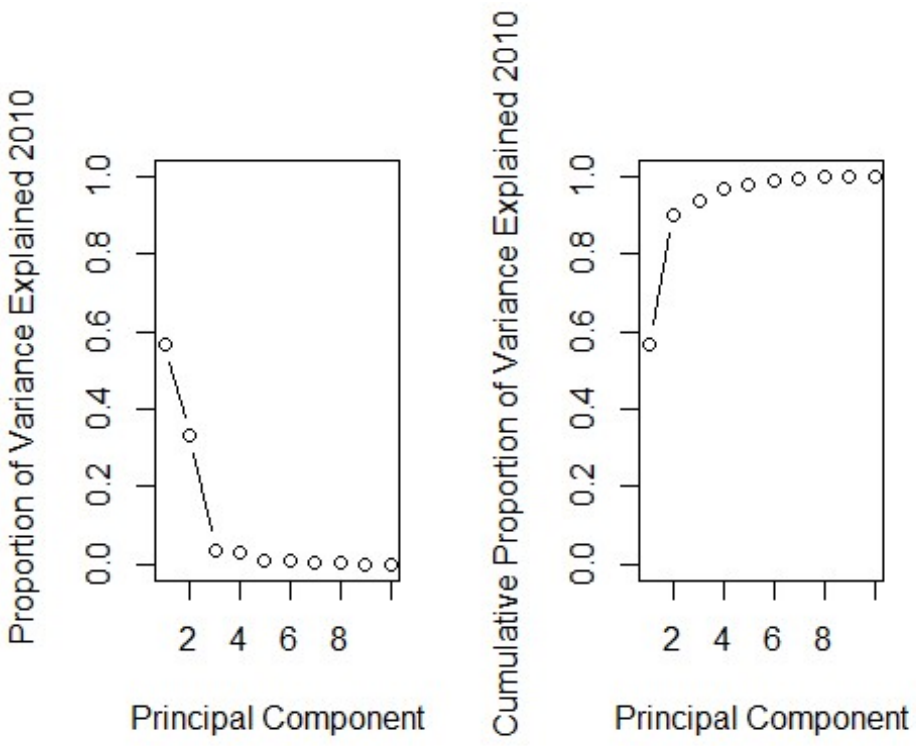
```
par(mfrow=c(1,1))
biplot(pr.out,scale=0)
```




```
pr.var =pr.out$sdev ^2
pve=pr.var/sum(pr.var )
pve
```

```
##  [1] 0.5687941755 0.3327825562 0.0366040311 0.0308227293 0.0117429584
##  [6] 0.0081094334 0.0053260112 0.0048057760 0.0009032316 0.0001090973
```

```
par(mfrow=c(1,2))
plot(pve,xlab=" Principal Component ",ylab=" Proportion of Variance Explained
2010",
     ylim=c(0,1),type='b')
plot(cumsum (pve ), xlab=" Principal Component ", ylab ="
Cumulative Proportion of Variance Explained 2010 ", ylim=c(0,1) ,
     type='b')
```




```
pr.out$rotation[,1:4]
```

```
##                        PC1        PC2         PC3          PC4
## count_nwne_p7  0.2216374 -0.4555887  0.02815083  0.013486693
## count_wne_p7   0.2358061 -0.4420886 -0.03888262  0.022415555
## mn_earn_wne_p7 0.3695422  0.1994203  0.02792210  0.446001334
## sd_earn_wne_p7 0.3585872  0.1711851  0.45315535  0.458206844
## gt_25k_p7      0.3553853  0.2012610 -0.53189705  0.306501743
## count_nwne_p9  0.2213852 -0.4441474  0.02373958 -0.020912149
## count_wne_p9   0.2423653 -0.4325407 -0.06206705 -0.003806183
## mn_earn_wne_p9 0.3763761  0.1911103  0.08174228 -0.368670850
## sd_earn_wne_p9 0.3584117  0.1715116  0.49372578 -0.453749736
## gt_25k_p9      0.3549044  0.1926656 -0.50377470 -0.392737581
```

PCA for 2009

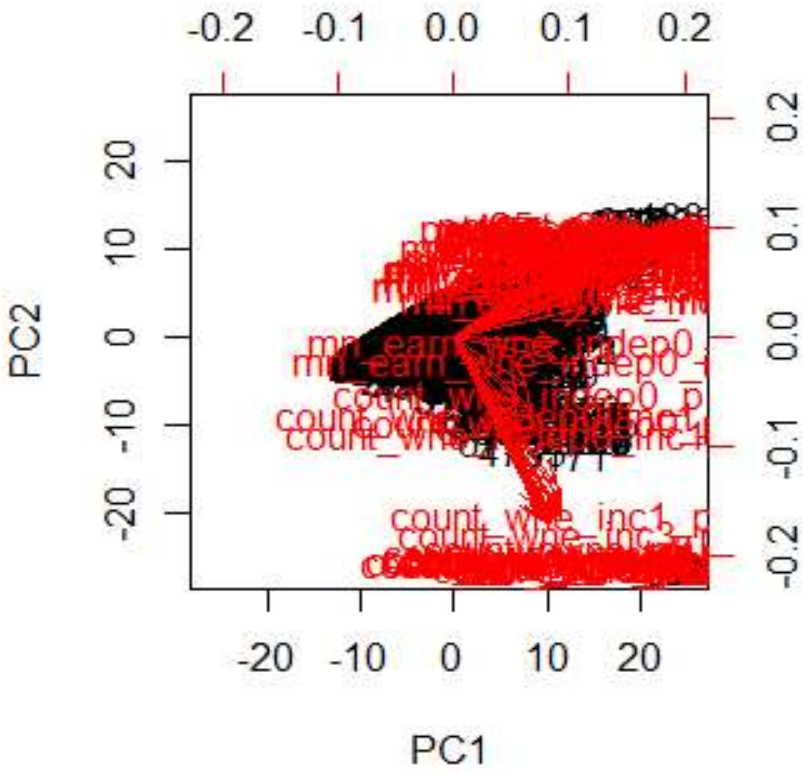
```
EAR_PCA_2009=EAR_2009[-c(1,2)]
pr.out_2009=prcomp(EAR_PCA_2009,scale=TRUE)
sort(pr.out_2009$rotation[,1],decreasing = TRUE)
```

```
##           pct75_earn_wne_p8          pct75_earn_wne_p10
##                  0.15250288                  0.15224099
##           pct90_earn_wne_p8          pct25_earn_wne_p10
##                  0.15155632                  0.15138896
##          pct90_earn_wne_p10           pct25_earn_wne_p8
##                  0.15105569                  0.15098765
##           pct75_earn_wne_p6           pct90_earn_wne_p6
##                  0.14786470                  0.14772617
```

```
##        mn_earn_wne_male1_p6        mn_earn_wne_male0_p6
##                   0.14723623                   0.14633441
##       mn_earn_wne_male0_p10       mn_earn_wne_male1_p10
##                   0.14413001                   0.14337903
##           pct25_earn_wne_p6             mn_earn_wne_p10
##                   0.14259630                   0.14255118
##             md_earn_wne_p10         mn_earn_wne_inc3_p6
##                   0.14212691                   0.14179880
##          pct10_earn_wne_p10         mn_earn_wne_inc2_p6
##                   0.14084087                   0.13990696
##           pct10_earn_wne_p8                    gt_25k_p8
##                   0.13941725                   0.13929862
##        mn_earn_wne_inc3_p10              mn_earn_wne_p8
##                   0.13927494                   0.13849691
##        mn_earn_wne_inc2_p10      mn_earn_wne_indep1_p10
##                   0.13839730                   0.13793578
##                   gt_25k_p10              md_earn_wne_p8
##                   0.13784731                   0.13780216
##      mn_earn_wne_indep0_p10         mn_earn_wne_inc1_p10
##                   0.13711440                   0.13673677
##         mn_earn_wne_inc1_p6       mn_earn_wne_indep1_p6
##                   0.13664315                   0.13594436
##       mn_earn_wne_indep0_p6                    gt_25k_p6
##                   0.13489580                   0.13229741
##              mn_earn_wne_p6             sd_earn_wne_p10
##                   0.13156721                   0.13013051
##              sd_earn_wne_p6              sd_earn_wne_p8
##                   0.12919437                   0.12904740
##              md_earn_wne_p6           pct10_earn_wne_p6
##                   0.12845089                   0.12655856
##         count_wne_male1_p10                count_wne_p10
##                   0.11476716                   0.11468122
##          count_wne_inc2_p10  mn_earn_wne_indep0_inc1_p6
##                   0.11420294                   0.11231911
##          count_wne_inc3_p10         count_wne_male0_p10
##                   0.11171665                   0.11143786
## mn_earn_wne_indep0_inc1_p10                count_wne_p8
##                   0.11019869                   0.10936602
##           count_wne_inc3_p6              count_nwne_p10
##                   0.10834032                   0.10633561
##          count_wne_inc1_p10          count_wne_male1_p6
##                   0.10618378                   0.10600143
##                count_wne_p6           count_wne_inc1_p6
##                   0.10568749                   0.10453785
##          count_wne_male0_p6           count_wne_inc2_p6
##                   0.10287399                   0.10236223
##        count_wne_indep1_p10               count_nwne_p6
##                   0.10030491                   0.10003833
##         count_wne_indep1_p6               count_nwne_p8
##                   0.09450903                   0.09295235
```

```
##         count_wne_indep0_p6        count_wne_indep0_p10
##                  0.08340124                  0.07770769
##   count_wne_indep0_inc1_p10    count_wne_indep0_inc1_p6
##                  0.06881303                  0.06681765
```

```
par(mfrow=c(1,1))
biplot(pr.out_2009,scale=0)
```

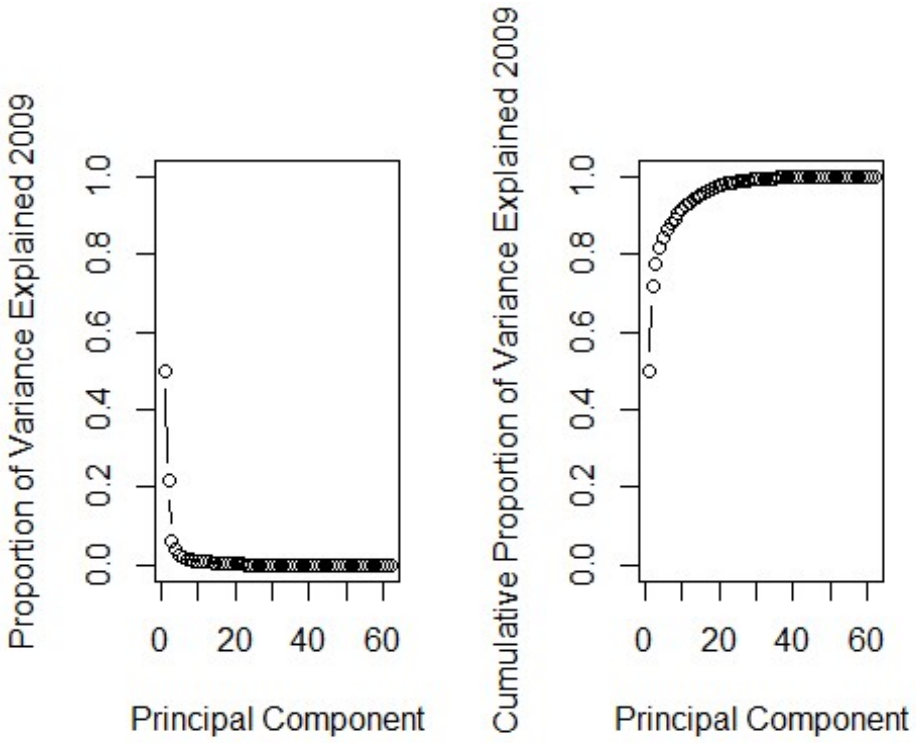



```
pr.var_2009 =pr.out_2009$sdev ^2
pve_2009=pr.var_2009/sum(pr.var_2009 )
pve_2009
```

```
##  [1] 5.004396e-01 2.164103e-01 6.041497e-02 3.945740e-02 2.797072e-02
##  [6] 1.856207e-02 1.525167e-02 1.341764e-02 1.246177e-02 1.107423e-02
## [11] 9.618082e-03 9.173487e-03 7.619334e-03 6.998166e-03 5.562444e-03
## [16] 5.394485e-03 5.067557e-03 4.158106e-03 4.019208e-03 3.413013e-03
## [21] 2.977164e-03 2.571870e-03 1.889368e-03 1.762378e-03 1.530287e-03
## [26] 1.442757e-03 1.392630e-03 1.211291e-03 1.089933e-03 9.662050e-04
## [31] 8.255591e-04 7.060343e-04 5.863750e-04 5.657142e-04 4.681519e-04
## [36] 4.176875e-04 3.898668e-04 3.315492e-04 3.202217e-04 2.328995e-04
## [41] 2.184489e-04 2.150022e-04 2.002616e-04 1.790334e-04 1.678711e-04
## [46] 1.555108e-04 1.288196e-04 9.894422e-05 7.911774e-05 6.992625e-05
## [51] 6.032528e-05 5.597910e-05 4.615474e-05 3.234650e-05 3.027228e-05
```

```
## [56] 2.626582e-05 2.562799e-05 1.567803e-05 1.158015e-05 8.755193e-06
## [61] 6.553364e-06 5.330032e-06
```

```
par(mfrow=c(1,2))
plot(pve_2009,xlab=" Principal Component ",ylab=" Proportion of Variance
Explained 2009 ",
     ylim=c(0,1),type='b')
plot(cumsum (pve_2009 ), xlab=" Principal Component ", ylab ="
Cumulative Proportion of Variance Explained 2009", ylim=c(0,1) ,
     type='b')
```




```
sort(pr.out_2009$rotation[,1],decreasing = TRUE)
```

```
##           pct75_earn_wne_p8          pct75_earn_wne_p10
##                  0.15250288                  0.15224099
##           pct90_earn_wne_p8          pct25_earn_wne_p10
##                  0.15155632                  0.15138896
##          pct90_earn_wne_p10           pct25_earn_wne_p8
##                  0.15105569                  0.15098765
##           pct75_earn_wne_p6           pct90_earn_wne_p6
##                  0.14786470                  0.14772617
##        mn_earn_wne_male1_p6        mn_earn_wne_male0_p6
##                  0.14723623                  0.14633441
##       mn_earn_wne_male0_p10       mn_earn_wne_male1_p10
##                  0.14413001                  0.14337903
##           pct25_earn_wne_p6             mn_earn_wne_p10
##                  0.14259630                  0.14255118
```

```
##             md_earn_wne_p10         mn_earn_wne_inc3_p6
##                   0.14212691                   0.14179880
##          pct10_earn_wne_p10         mn_earn_wne_inc2_p6
##                   0.14084087                   0.13990696
##           pct10_earn_wne_p8                  gt_25k_p8
##                   0.13941725                   0.13929862
##        mn_earn_wne_inc3_p10              mn_earn_wne_p8
##                   0.13927494                   0.13849691
##        mn_earn_wne_inc2_p10      mn_earn_wne_indep1_p10
##                   0.13839730                   0.13793578
##                  gt_25k_p10              md_earn_wne_p8
##                   0.13784731                   0.13780216
##      mn_earn_wne_indep0_p10        mn_earn_wne_inc1_p10
##                   0.13711440                   0.13673677
##         mn_earn_wne_inc1_p6       mn_earn_wne_indep1_p6
##                   0.13664315                   0.13594436
##       mn_earn_wne_indep0_p6                   gt_25k_p6
##                   0.13489580                   0.13229741
##              mn_earn_wne_p6             sd_earn_wne_p10
##                   0.13156721                   0.13013051
##              sd_earn_wne_p6              sd_earn_wne_p8
##                   0.12919437                   0.12904740
##              md_earn_wne_p6           pct10_earn_wne_p6
##                   0.12845089                   0.12655856
##         count_wne_male1_p10               count_wne_p10
##                   0.11476716                   0.11468122
##          count_wne_inc2_p10  mn_earn_wne_indep0_inc1_p6
##                   0.11420294                   0.11231911
##          count_wne_inc3_p10         count_wne_male0_p10
##                   0.11171665                   0.11143786
## mn_earn_wne_indep0_inc1_p10               count_wne_p8
##                   0.11019869                   0.10936602
##           count_wne_inc3_p6              count_nwne_p10
##                   0.10834032                   0.10633561
##          count_wne_inc1_p10          count_wne_male1_p6
##                   0.10618378                   0.10600143
##                count_wne_p6           count_wne_inc1_p6
##                   0.10568749                   0.10453785
##          count_wne_male0_p6           count_wne_inc2_p6
##                   0.10287399                   0.10236223
##        count_wne_indep1_p10              count_nwne_p6
##                   0.10030491                   0.10003833
##         count_wne_indep1_p6              count_nwne_p8
##                   0.09450903                   0.09295235
##         count_wne_indep0_p6         count_wne_indep0_p10
##                   0.08340124                   0.07770769
##   count_wne_indep0_inc1_p10    count_wne_indep0_inc1_p6
##                   0.06881303                   0.06681765
```

Conclusion-

1) For 2010 , we can see 90% of the variability explained by first two principal components and 96% of the variability explained by first four principal components. For 2009 , we can see 80% of the variability explained by first three principal components. For first trial, we can use first three/ four principal components.

2) Important variables can be found from the sorted list above , if we don't want go with principopal components. In summary , important fields are-- # 'Mean/median earnings of students 6 yrs/8 yrs after' # '10th/25th/75th percentile of earnings of students working and not enrolled 10 years after entry' # Share of students earning over $25,000/year after 6/7/8/9 years