趙友誠 H24101060

9/27/24

Table of contents

Brief introduction to the data

This is a complete data with no actual missing value while some might be labeled as missing. Dimension of the Data : **1671 samples × 15 columns**

Variables	Explanation
V1、V2、V3	District and Li
V4_1~V4_8	Popularity
V5	Support level
V6	Age
V7	Education level
V8	Sex

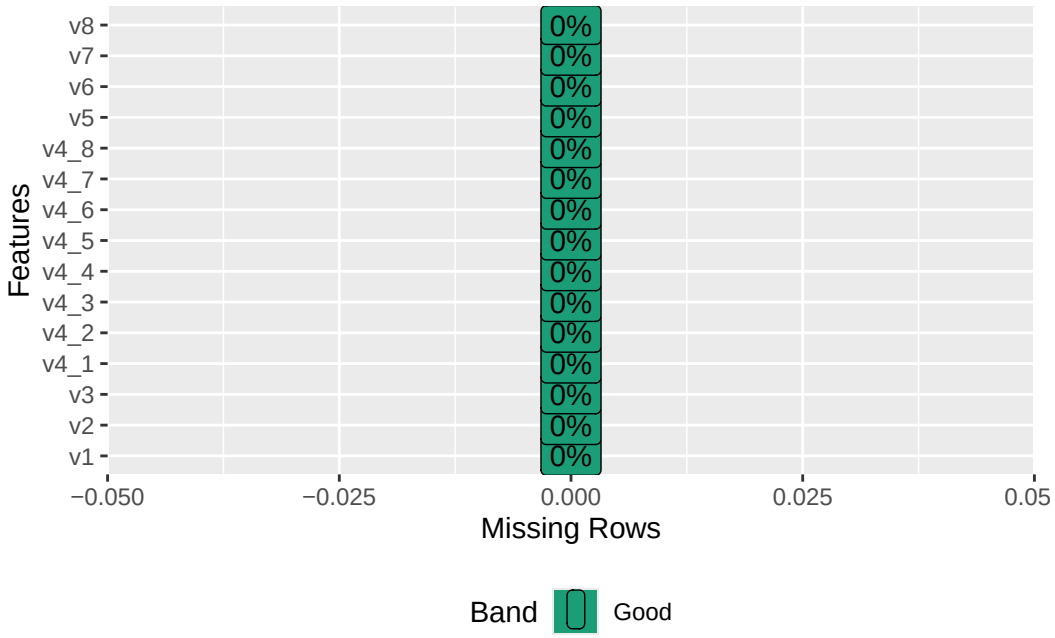
```r
library(haven)
library(Hmisc)
pollsav <- read_sav("poll.sav")
write.csv(pollsav, file = "poll.csv", row.names = FALSE)
pollcsv <- read.csv("poll.csv")
```

Preprocessing

First, we replace 91,95,98,99 by 0 and then converted variables from numeric format to factor one. As the graph shown below, there is no missing value in this data.

```r
pollcsv <- data.frame(
  t(apply(pollcsv,MARGIN = 1, FUN = function(row){
    row[row==99 | row==98 | row==91 | row==95] <- 0
    return(row)
  }))
)
pollcsv[] <- lapply(pollcsv, function(item) return(as.factor(item)))
```

```r
DataExplorer::plot_missing(pollcsv)
```



Descriptive statistic

This chunk is for the convenience of analysis, so it will not be shown here.

```
latex(describe(pollcsv, "Public Opinion"),file = "", size = "normalsize")
```

The definition of the popularity of a candidate in this analysis is the number of appearance of a candidate that a participant answered in the 4th question. And the definition of the support level of a candidate is the number of appearance of a candidate in 5th question divided by the number of the participants who specifically choose a name in 5th question. That is, participants who did not actually answered the question are removed from the calculation.

```
count4 <- unlist(lapply(factor(1:10), function(x){
  return(length(unlist(apply(pollcsv[,4:11], MARGIN = 1, function(row) if(x %in% row) retu
}))
count5 <- unlist(lapply(factor(0:10),function(x){
  return(sum(pollcsv$v5==x))
} ))
p <- data.frame(factor(1:10), popularity=round(count4/1671,3), count4)
```

	Popularity			Support level		
	candidate	rate	count	candidate	rate	count
1	3	0.245	409	3	0.243	205
2	7	0.205	342	7	0.231	195
3	1	0.196	328	1	0.187	158
4	6	0.139	232	6	0.116	98
5	4	0.099	166	4	0.094	79
6	5	0.075	126	10	0.063	53
7	10	0.069	116	5	0.039	33
8	8	0.014	23	2	0.011	9
9	9	0.014	24	9	0.009	8
10	2	0.007	11	8	0.007	6

```
s <- data.frame(factor(1:10), `support level`=round(count5[2:11]/sum(count5[2:11]),3), cou
p <- p[order(p$popularity, decreasing = TRUE ),]
s <- s[order(s$support.level, decreasing = TRUE ),]
latex(data.table::data.table(cbind(p,s)),title="",file = "", booktabs = TRUE, cgroup = c('
```