

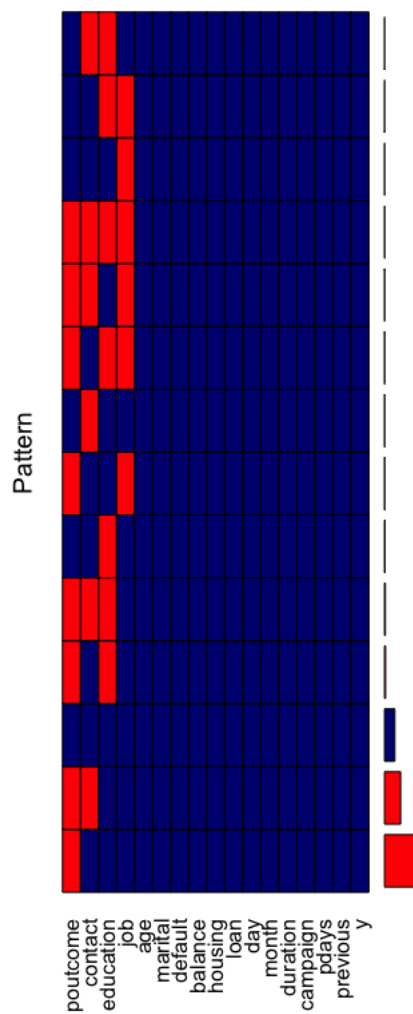
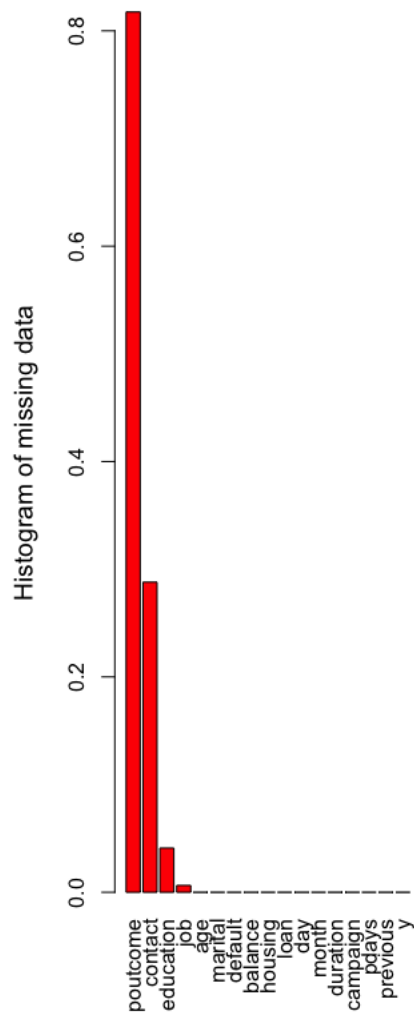
## a. Create a visual for representing missing values in the dataset.

```

Console ~/work/ACADGILD/ <
> # read & view bank data
> setwd("~/Users/gifu/work/ACADGILD")
> getwd()
[1] "~/Users/gifu/work/ACADGILD"
> bank_data <- read.csv("bank-full.csv", header = TRUE, sep = ";")
> View(bank_data)
> summary(bank_data)
  age          job          marital      education  default  balance  housing  loan  contact  day          month  duration  campaign  pdays  previous
Min.   :18.00   blue-collar:9752 divorced: 5207 primary : 6851 no :44396   Min.   : -8019 no :20081 no :37967 cellular:29285 Min.   : 1.00 may :13766 Min.   : 0.0 Min.   : 1.000 Min.   : -1.0 Min.   : 0.0000
1st Qu.:33.00   management:19458 married: 27214 secondary:23202 1st Qu.: 72 yes: 815 1st Qu.: 72 yes:25130 1st Qu.: 8.00 jul : 6895 1st Qu.: 183.0 1st Qu.: 1.000 1st Qu.: -1.0 1st Qu.: 0.0000
Median :39.00   technician:7507 single :12790 tertiary :13381 Median : 444 Median : 444 unknown :13009 Median :15.00 aug : 6247 Median : 180.0 Median : 2.000 Median : -1.0 Median : 0.0000
Mean :40.94   admin.:5171 unknown : 1857 Mean : 1362 3rd Qu.: 1428 3rd Qu.: 21.00 nov : 3979 3rd Qu.: 319.0 3rd Qu.: 3.000 3rd Qu.: -1.0 3rd Qu.: 0.0000
Max.   :95.00   retired :2264 (Other) : 6835 Max.   :182127 Max.   :31.00 apr : 2932 Max.   :4911.0 Max.   :163.000 Max.   :1871.0 Max.   :275.0000
  poutcome
failure:4901 no :39932
other : 1840 yes: 5289
success: 1511
unknown:36959

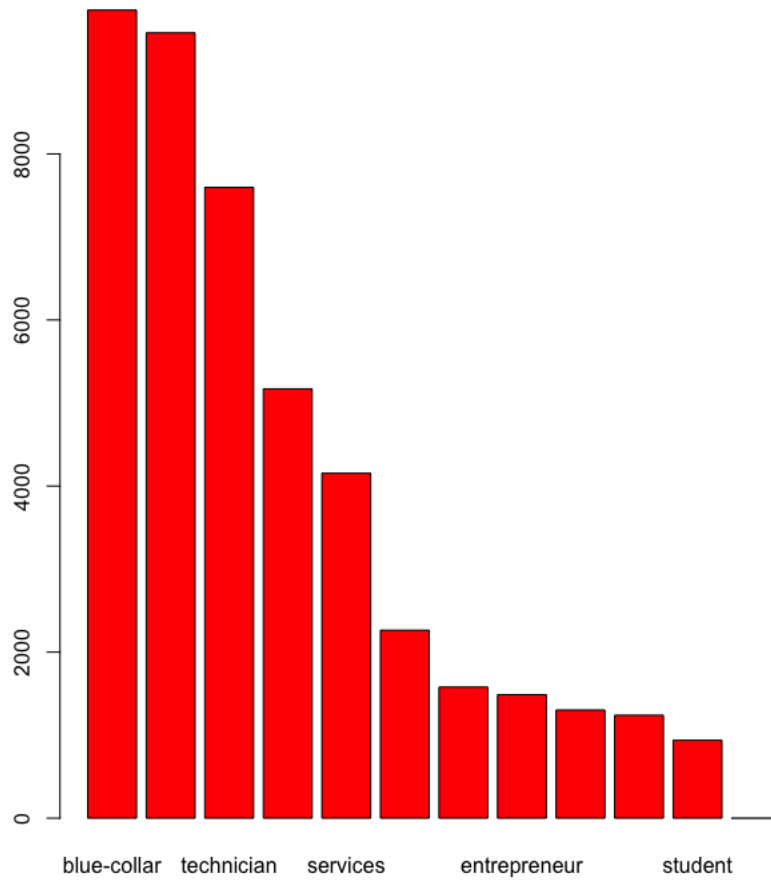
> #a. Create a visual for representing missing values in the dataset.
> bank_data[bank_data=="unknown"] = NA
> #table(bank_data$job)
> summary(is.na(bank_data))
  age          job          marital      education  default  balance  housing  loan  contact  day          month  duration  campaign  pdays  previous
Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical Mode :logical
FALSE:45211 FALSE:44923 FALSE:45211 FALSE:43354 FALSE:45211 FALSE:45211 FALSE:45211 FALSE:45211 FALSE:32191 FALSE:45211 FALSE:45211 FALSE:45211 FALSE:45211 FALSE:45211
TRUE :1288 TRUE :1857 TRUE :1857 TRUE :1857 TRUE :1857 TRUE :1857 TRUE :1857 TRUE :1857 TRUE :1857 TRUE :1857 TRUE :1857 TRUE :1857
poutcome
Mode :logical Mode :logical
FALSE:8252 FALSE:45211
TRUE :36959
> library(VIM)
> vggp <- vggp(bank_data, col=c("navyblue","red"), numbers=FALSE, sortVars=TRUE, labels=names(bank_data), ylab=c("Histogram of Missing data","Pattern"))
Variables sorted by number of missings:
Variable Count
poutcome 0.817478047
contact 0.287963013
education 0.041874075
job 0.006370131
age 0.000000000
marital 0.000000000
default 0.000000000
balance 0.000000000
housing 0.000000000
loan 0.000000000
day 0.000000000
month 0.000000000
duration 0.000000000
campaign 0.000000000
pdays 0.000000000
previous 0.000000000
y 0.000000000
> #b. Show a distribution of clients based on a Job.
> jobcount <- table(bank_data$job)
>

```



## b. Show a distribution of clients based on a Job

```
# Show a distribution of clients based on a Job.  
jobcount <- table(bank_data$job)  
# Show a distribution of clients based on a Job.  
jobcount <- table(bank_data$job)  
barplot(height = jobcount[order(jobcount, decreasing = TRUE)], col = "red")  
|
```

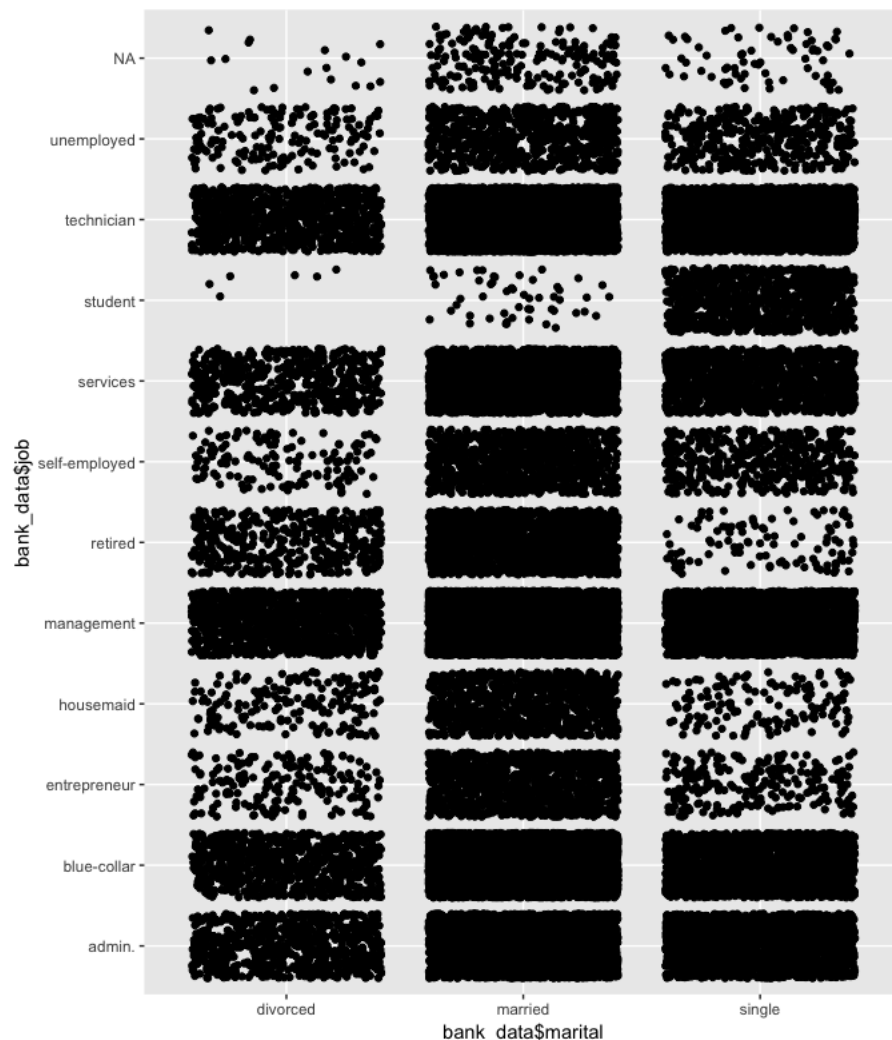


## c. Check whether is there any relation between Job and Marital Status?

```
> #c. Check whether is there any relation between Job and Marital Status/
> library(ggplot2)
> crossTable(bank_data$marital, bank_data$job)
```

Cell Contents												
	N	Chi-square contribution	N / Row Total	N / Col Total	N / Table Total							
Total Observations in Table: 44923												
bank_data\$marital	admin	blue-collar	entrepreneur	housemaid	management	retired	self-employed	services	student	technician	unemployed	Row Total
divorced	750	750	179	184	1111	425	140	549	6	925	171	5190
	38.974	124.638	0.382	11.587	0.307	102.124	9.866	9.945	96.700	2.550	2.782	
	0.145	0.145	0.034	0.035	0.214	0.082	0.027	0.100	0.001	0.178	0.033	0.116
	0.145	0.077	0.120	0.148	0.117	0.188	0.089	0.132	0.006	0.122	0.131	
	0.017		0.004	0.004	0.025	0.009	0.003	0.012	0.000	0.021	0.004	
married	2693	6968	1070	912	5400	1731	993	2407	54	4052	731	27011
	55.709	212.996	34.688	37.147	14.468	100.413	2.001	3.293	461.165	58.260	3.513	
	0.100	0.258	0.040	0.034	0.200	0.064	0.037	0.009	0.002	0.150	0.027	0.601
	0.521	0.716	0.720	0.735	0.571	0.765	0.629	0.579	0.058	0.533	0.561	
	0.060	0.155	0.024	0.020	0.120	0.039	0.022	0.054	0.001	0.090	0.016	
single	1728	2014	238	144	2947	188	446	1198	878	2620	401	12722
	47.448	199.797	79.623	122.212	26.923	443.347	0.003	0.397	1411.652	102.048	2.774	
	0.136	0.158	0.019	0.011	0.232	0.008	0.035	0.094	0.069	0.206	0.032	0.283
	0.334	0.207	0.100	0.116	0.312	0.048	0.282	0.286	0.036	0.345	0.308	
	0.038	0.045	0.005	0.003	0.066	0.002	0.010	0.027	0.020	0.058	0.009	
Column Total	5271	9732	1407	1240	9058	2264	1579	4154	938	7597	1303	44923
	0.115	0.217	0.033	0.028	0.211	0.050	0.035	0.092	0.021	0.169	0.029	

```
> ggplot(bank_data, aes(bank_data$marital, bank_data$job)) + geom_jitter()
>
```



d. Check whether there is any association between Job and Education?

```
> #d. Check whether there is any association between Job and Education?  
> ggplot(bank_data,aes(bank_data$education,bank_data$job)) + geom_jitter()  
>
```

