Session 11 assignment 1

Submitted By: Ashish Saxena

1. Use the given link and locate the bank marketing dataset. Data Set Link Perform the below operations:

1. Create a visual for representing missing values in the dataset.
2. Show a distribution of clients based on a Job.
3. Check whether is there any relation between Job and Marital Status?
4. Check whether is there any association between Job and Education?

|  |  |
| --- | --- |
| ## The data set can be obtained from http://archive.ics.uci.edu/ml/datasets/Bank+Marketing  ## DATASET UNDERSTANDING **library**(readr)  bank\_full <- **read\_delim**("C:/Users/Seshan/Desktop/Bank/bank-full.csv", ";", escape\_double = FALSE, trim\_ws = TRUE)  ## Parsed with column specification:  ## cols(  ## age = col\_integer(),  ## job = col\_character(),  ## marital = col\_character(),  ## education = col\_character(),  ## default = col\_character(),  ## balance = col\_integer(),  ## housing = col\_character(),  ## loan = col\_character(),  ## contact = col\_character(),  ## day = col\_integer(),  ## month = col\_character(),  ## duration = col\_integer(),  ## campaign = col\_integer(),  ## pdays = col\_integer(),  ## previous = col\_integer(),  ## poutcome = col\_character(),  ## y = col\_character()  ## ) | |
| *#Lets look at dataset and generate initial understanding about the column* |  |
| *types* **str**(bank\_full)  ## Classes 'tbl\_df', 'tbl' and 'data.frame': 45211 obs. of 17 variables:  ## $ age : int 58 44 33 47 33 35 28 42 58 43 ...  ## $ job : chr "management" "technician" "entrepreneur" "bluecollar" ...  ## $ marital : chr "married" "single" "married" "married" ...  ## $ education: chr "tertiary" "secondary" "secondary" "unknown" ...  ## $ default : chr "no" "no" "no" "no" ...  ## $ balance : int 2143 29 2 1506 1 231 447 2 121 593 ...  ## $ housing : chr "yes" "yes" "yes" "yes" ...  ## $ loan : chr "no" "no" "yes" "no" ...  ## $ contact : chr "unknown" "unknown" "unknown" "unknown" ...  ## $ day : int 5 5 5 5 5 5 5 5 5 5 ...  ## $ month : chr "may" "may" "may" "may" ...  ## $ duration : int 261 151 76 92 198 139 217 380 50 55 ...  ## $ campaign : int 1 1 1 1 1 1 1 1 1 1 ...  ## $ pdays : int -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...  ## $ previous : int 0 0 0 0 0 0 0 0 0 0 ...  ## $ poutcome : chr "unknown" "unknown" "unknown" "unknown" ...  ## $ y : chr "no" "no" "no" "no" ...  ## - attr(\*, "spec")=List of 2 ## ..$ cols :List of 17  ## .. ..$ age : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector" ## .. ..$ job : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  ## .. ..$ marital : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  ## .. ..$ education: list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  ## .. ..$ default : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector" ## .. ..$ balance : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector" ## .. ..$ housing : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  ## .. ..$ loan : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  ## .. ..$ contact : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector" ## .. ..$ day : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector" ## .. ..$ month : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector" ## .. ..$ duration : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector" ## .. ..$ campaign : list() |

## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"

|  |
| --- |
| ## .. ..$ pdays : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"  ## .. ..$ previous : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector" ## .. ..$ poutcome : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  ## .. ..$ y : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector" ## ..$ default: list()  ## .. ..- attr(\*, "class")= chr "collector\_guess" "collector"  ## ..- attr(\*, "class")= chr "col\_spec" |

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

|  |
| --- |
| *#A deep check for NA in a particular column let say age* **if**(**length**(**which**(**is.na**(bank\_full**$**age)**==**TRUE)**>**0)){ **print**("Missing Value found in the specified column")  } **else** **print**("All okay: No Missing Value found in the specified column") ## [1] "All okay: No Missing Value found in the specified column"  *# Check another example say*  **if**(**length**(**which**(**is.na**(bank\_full**$**campaign)**==**TRUE)**>**0)){**print**("Missing Value found in the specified column")} **else** **print**("All okay: No Missing Value found in the specified column") ## [1] "All okay: No Missing Value found in the specified column" **head**(bank\_full) ## Displays first 6 rows for each variable  ## # A tibble: 6 x 17  ## age job marital education default balance housing loan contact  ## <int> <chr> <chr> <chr> <chr> <int> <chr> <chr> <chr>  ## 1 58 management married tertiary no 2143 yes no unknown  ## 2 44 technician single secondary no 29 yes no unknown  ## 3 33 entrepren~ married secondary no 2 yes yes unknown  ## 4 47 blue-coll~ married unknown no 1506 yes no unknown  ## 5 33 unknown single unknown no 1 no no unknown  ## 6 35 management married tertiary no 231 yes no unknown  ## # ... with 8 more variables: day <int>, month <chr>, duration <int>, ## # campaign <int>, pdays <int>, previous <int>, poutcome <chr>, y <chr> **str**(bank\_full) ## Describes each variables |

## Classes 'tbl\_df', 'tbl' and 'data.frame': 45211 obs. of 17 variables:

|  |
| --- |
| ## $ age : int 58 44 33 47 33 35 28 42 58 43 ...  ## $ job : chr "management" "technician" "entrepreneur" "blue-collar" ...  ## $ marital : chr "married" "single" "married" "married" ...  ## $ education: chr "tertiary" "secondary" "secondary" "unknown" ...  ## $ default : chr "no" "no" "no" "no" ...  ## $ balance : int 2143 29 2 1506 1 231 447 2 121 593 ...  ## $ housing : chr "yes" "yes" "yes" "yes" ...  ## $ loan : chr "no" "no" "yes" "no" ...  ## $ contact : chr "unknown" "unknown" "unknown" "unknown" ...  ## $ day : int 5 5 5 5 5 5 5 5 5 5 ...  ## $ month : chr "may" "may" "may" "may" ...  ## $ duration : int 261 151 76 92 198 139 217 380 50 55 ...  ## $ campaign : int 1 1 1 1 1 1 1 1 1 1 ...  ## $ pdays : int -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...  ## $ previous : int 0 0 0 0 0 0 0 0 0 0 ...  ## $ poutcome : chr "unknown" "unknown" "unknown" "unknown" ...  ## $ y : chr "no" "no" "no" "no" ...  ## - attr(\*, "spec")=List of 2 ## ..$ cols :List of 17  ## .. ..$ age : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector" ## .. ..$ job : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  ## .. ..$ marital : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  ## .. ..$ education: list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  ## .. ..$ default : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector" ## .. ..$ balance : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector" ## .. ..$ housing : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  ## .. ..$ loan : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  ## .. ..$ contact : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector" ## .. ..$ day : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector" ## .. ..$ month : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector" ## .. ..$ duration : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"  ## .. ..$ campaign : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"  ## .. ..$ pdays : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector" ## .. ..$ previous : list() |

## .. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"

|  |
| --- |
| ## .. ..$ poutcome : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector"  ## .. ..$ y : list()  ## .. .. ..- attr(\*, "class")= chr "collector\_character" "collector" ## ..$ default: list()  ## .. ..- attr(\*, "class")= chr "collector\_guess" "collector"  ## ..- attr(\*, "class")= chr "col\_spec" **summary**(bank\_full) ## Provides basic statistical information of each variable  ## age job marital education  ## Min. :18.00 Length:45211 Length:45211 Length:45211  ## 1st Qu.:33.00 Class :character Class :character Class :character  ## Median :39.00 Mode :character Mode :character Mode :character  ## Mean :40.94  ## 3rd Qu.:48.00  ## Max. :95.00  ## default balance housing loan  ## Length:45211 Min. : -8019 Length:45211 Length:45211  ## Class :character 1st Qu.: 72 Class :character Class :character  ## Mode :character Median : 448 Mode :character Mode :character  ## Mean : 1362  ## 3rd Qu.: 1428  ## Max. :102127  ## contact day month duration  ## Length:45211 Min. : 1.00 Length:45211 Min. : 0.0  ## Class :character 1st Qu.: 8.00 Class :character 1st Qu.: 103.0  ## Mode :character Median :16.00 Mode :character Median : 180.0  ## Mean :15.81 Mean : 258.2  ## 3rd Qu.:21.00 3rd Qu.: 319.0  ## Max. :31.00 Max. :4918.0  ## campaign pdays previous poutcome  ## Min. : 1.000 Min. : -1.0 Min. : 0.0000 Length:45211  ## 1st Qu.: 1.000 1st Qu.: -1.0 1st Qu.: 0.0000 Class :character  ## Median : 2.000 Median : -1.0 Median : 0.0000 Mode :character  ## Mean : 2.764 Mean : 40.2 Mean : 0.5803  ## 3rd Qu.: 3.000 3rd Qu.: -1.0 3rd Qu.: 0.0000 ## Max. :63.000 Max. :871.0 Max. :275.0000  ## y  ## Length:45211  ## Class :character  ## Mode :character  ##  ## ##  ## DATA EXPLORATION - Check for Missing Data  ## Option 1  **is.na**(bank\_full) ## Displays True for a missing value |

## age job marital education default balance housing loan

## [1,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [2,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [3,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [4,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [5,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [6,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [7,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [8,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [9,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [10,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [11,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [12,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [13,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [14,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [15,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [16,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [17,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [18,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [19,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [20,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [21,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [22,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [23,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [24,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [25,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [26,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [27,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [28,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [29,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [30,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [31,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [32,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [33,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [34,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [35,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [36,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [37,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [38,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [39,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [40,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [41,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [42,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [43,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [44,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [45,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [46,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

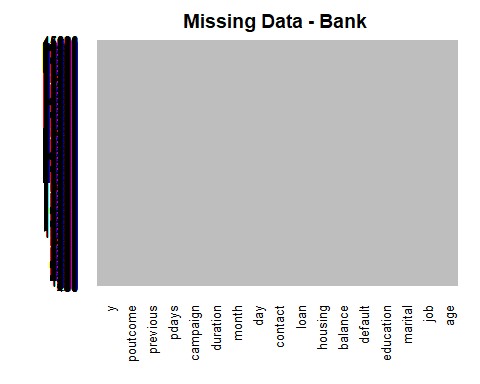
## [47,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [48,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

## [49,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

|  |
| --- |
| Deleted remaining false as it is very lengthy  FALSE FALSE FALSE FALSE FALSE  ## [4653,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4654,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4655,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4656,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4657,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4658,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4659,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4660,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4661,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4662,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4663,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4664,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4665,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4666,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4667,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4668,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4669,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4670,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4671,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4672,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4673,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4674,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4675,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4676,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4677,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4678,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4679,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4680,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4681,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4682,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4683,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4684,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4685,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4686,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4687,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4688,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4689,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4690,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4691,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4692,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4693,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4694,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4695,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4696,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4697,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4698,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  ## [4699,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE |

|  |  |
| --- | --- |
| ## [5867,] FALSE  ## [5868,] FALSE  ## [5869,] FALSE  ## [5870,] FALSE  ## [5871,] FALSE  ## [5872,] FALSE  ## [5873,] FALSE  ## [5874,] FALSE  ## [5875,] FALSE  ## [5876,] FALSE  ## [5877,] FALSE  ## [5878,] FALSE  ## [5879,] FALSE  ## [5880,] FALSE  ## [5881,] FALSE  ## [5882,] FALSE  ## [ reached getOption("max.print") -- omitted 39329 rows ] | |
| ## Since it is a large dataset, graphical display of missing values will | ll |
| prove to be easier  ##Option 2 **require**(Amelia)  ## Loading required package: Amelia  ## Loading required package: Rcpp  ## ##  ## ## Amelia II: Multiple Imputation  ## ## (Version 1.7.5, built: 2018-05-07)  ## ## Copyright (C) 2005-2018 James Honaker, Gary King and Matthew Blackwe  ## ## Refer to http://gking.harvard.edu/amelia/ for more information  ## ##  **missmap**(bank\_full,main="Missing Data - Bank ", col=**c**("red","grey"),legend=FALSE)  ## Warning in if (class(obj) == "amelia") {: the condition has length > 1 and ## only the first element will be used  ## Warning: Unknown or uninitialised column: 'arguments'.    ## Warning: Unknown or uninitialised column: 'arguments'.  ## Warning: Unknown or uninitialised column: 'imputations'. |



|  |  |
| --- | --- |
| ## No red colour stripes are visible. hence no missing values. | ss :character |
| **summary**( bank\_full) ## displays missing values if any under every variable  ## age job marital education  ## Min. :18.00 Length:45211 Length:45211 Length:45211  ## 1st Qu.:33.00 Class :character Class :character Class :character  ## Median :39.00 Mode :character Mode :character Mode :character ## Mean :40.94  ## 3rd Qu.:48.00  ## Max. :95.00  ## default balance housing loan  ## Length:45211 Min. : -8019 Length:45211 Length:45211 ## Class :character 1st Qu.: 72 Class :character Class :character ## Mode :character Median : 448 Mode :character Mode :character  ## Mean : 1362  ## 3rd Qu.: 1428  ## Max. :102127  ## contact day month duration  ## Length:45211 Min. : 1.00 Length:45211 Min. : 0.0 ## Class :character 1st Qu.: 8.00 Class :character 1st Qu.: 103.0  ## Mode :character Median :16.00 Mode :character Median : 180.0  ## Mean :15.81 Mean : 258.2  ## 3rd Qu.:21.00 3rd Qu.: 319.0 ## Max. :31.00 Max. :4918.0 ## campaign pdays previous poutcome  ## Min. : 1.000 Min. : -1.0 Min. : 0.0000 Length:45211  ## 1st Qu.: 1.000 1st Qu.: -1.0 1st Qu.: 0.0000 Cla |

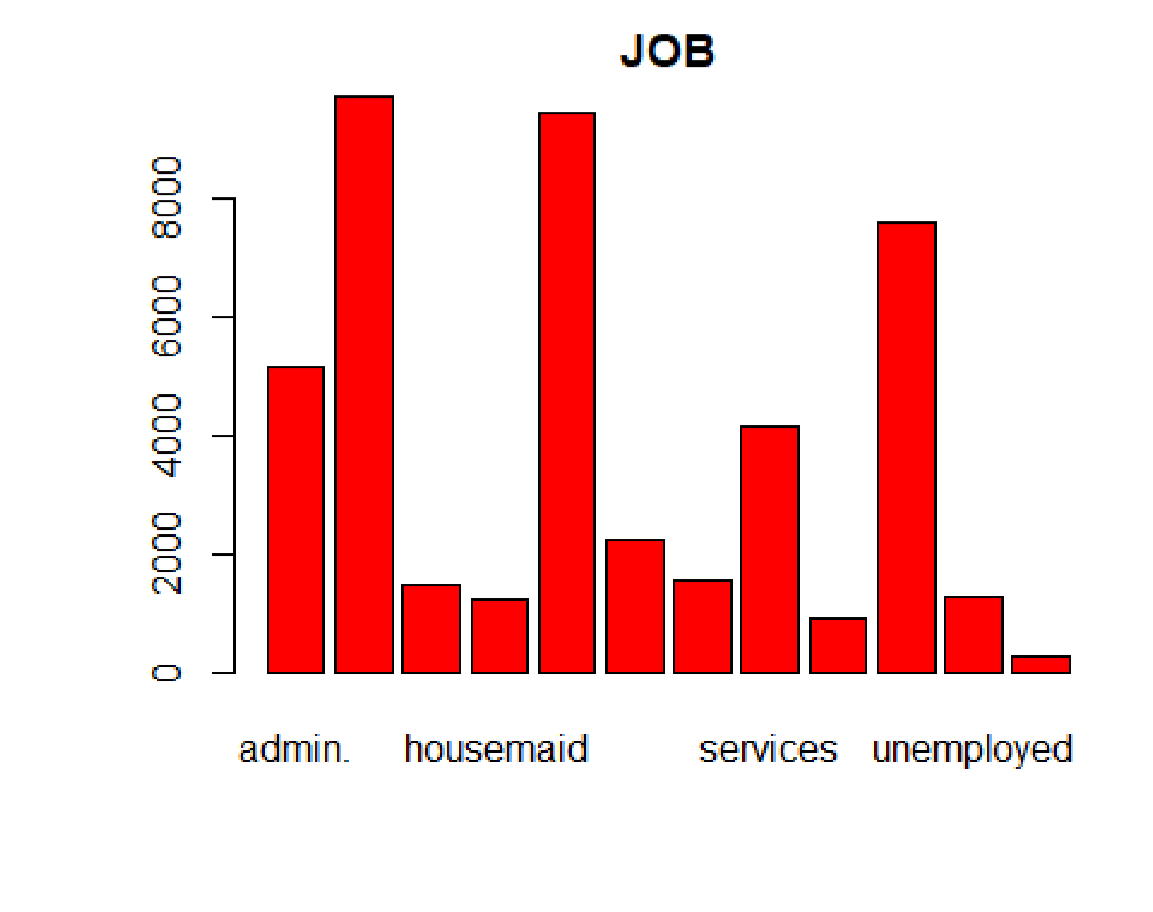
## Median : 2.000 Median : -1.0 Median : 0.0000 Mode :character

|  |
| --- |
| ## Mean : 2.764 Mean : 40.2 Mean : 0.5803  ## 3rd Qu.: 3.000 3rd Qu.: -1.0 3rd Qu.: 0.0000 ## Max. :63.000 Max. :871.0 Max. :275.0000  ## y  ## Length:45211  ## Class :character  ## Mode :character  ##  ## ## |

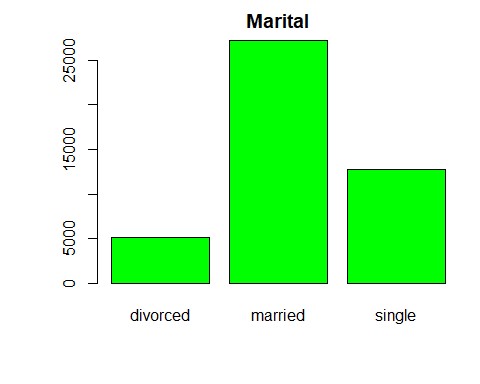
1. Show a distribution of clients based on a Job.
2. Check whether is there any relation between Job and Marital Status?
3. Check whether is there any association between Job and Education?

1. Show a distribution of clients based on a Job.

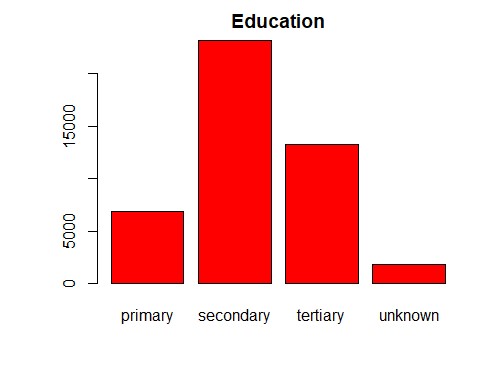
|  |  |
| --- | --- |
| ## Barplotsfor Categorical Variables | ,main="JOB") |
| **barplot**(**table**(bank\_full**$**job),col= "red" |



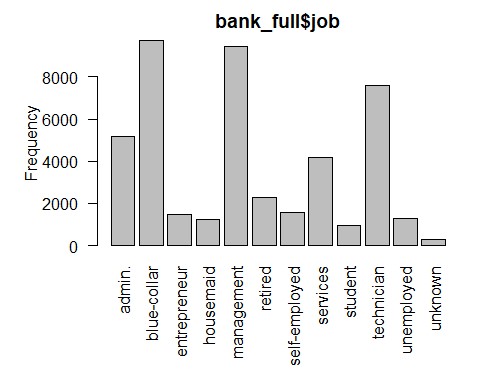
|  |  |
| --- | --- |
| **barplot**(**table**(bank\_full**$**marital),col="green",main="Marital") |  |



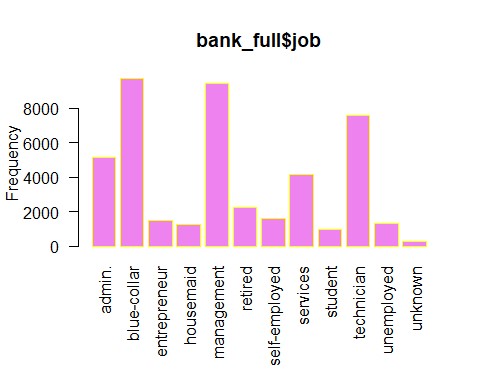
**barplot**(**table**(bank\_full**$**education),col="red",main="Education")



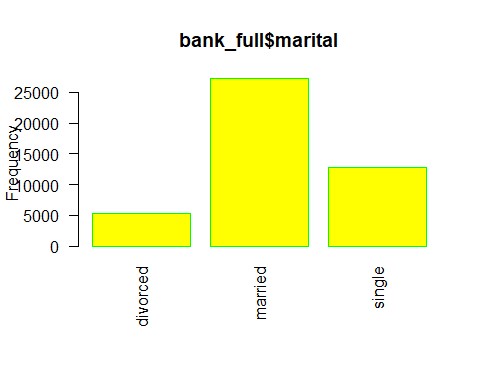
|  |  |
| --- | --- |
| **par**(oma=**c**(2,0,0,0)) *#so labels are not cut off* | , main = "bank\_full$job", |
| **barplot**(**table**(bank\_full**$**job),ylab = "Frequency" border="black", col="grey",las=2) |



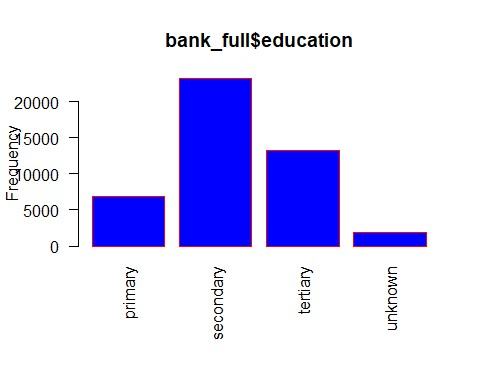
|  |  |
| --- | --- |
| *#Histogram for job,marital and education - three categorical variables* | , |
| **par**(oma=**c**(2,0,0,0)) *#so labels are not cut off*  **barplot**(**table**(bank\_full**$**job),ylab = "Frequency", main = "bank\_full $job" border="yellow", col="violet",las=2) |



|  |  |
| --- | --- |
| **par**(oma=**c**(2,0,0,0)) *#so labels are not cut off* | , main = |
| **barplot**(**table**(bank\_full**$**marital),ylab = "Frequency"  "bank\_full$marital",  border="green", col="yellow",las=2) |



|  |  |
| --- | --- |
| **par**(oma=**c**(2,0,0,0)) *#so labels are not cut off* | "Frequency", main = |
| **barplot**(**table**(bank\_full**$**education),ylab =  "bank\_full$education",  border="red", col="blue",las=2) |



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

As both are a categorical variable this can be checked with chisq.test

|  |  |
| --- | --- |
| **with**(bank\_full, **chisq.test**( job, marital)) | 2.2e-16 |
| ##  ## Pearson's Chi-squared test  ##  ## data: job and marital  ## X-squared = 3837.6, df = 22, p-value < **with**(bank\_full, **table**( job, marital) )  ## marital  ## job divorced married single  ## admin. 750 2693 1728 ## blue-collar 750 6968 2014 ## entrepreneur 179 1070 238 ## housemaid 184 912 144 ## management 1111 5400 2947  ## retired 425 1731 108 ## self-employed 140 993 446 ## services 549 2407 1198  ## student 6 54 878 |

## technician 925 4052 2620

## unemployed 171 731 401

## unknown 17 203 68

|  |  |
| --- | --- |
| *# OR* | (bank\_full, **prop.table**(**table**( job,education)))  ## education  ## job primary secondary tertiary unknown  ## admin. 0.0046227688 0.0933179978 0.0126517883 0.0037822654  ## blue-collar 0.0831213643 0.1187985225 0.0032956581 0.0100418040  ## entrepreneur 0.0040476875 0.0119882330 0.0151732985 0.0016810068  ## housemaid 0.0138683064 0.0087368118 0.0038265024 0.0009953330  ## management 0.0065028422 0.0247948508 0.1725465042 0.0053526797  ## retired 0.0175842162 0.0217646148 0.0080953750 0.0026321028  ## self-employed 0.0028754064 0.0127623808 0.0184247196 0.0008626219  ## services 0.0076308863 0.0764636925 0.0044679392 0.0033177766  ## student 0.0009732145 0.0112362036 0.0049324279 0.0036053173  ## technician 0.0034947247 0.1156576939 0.0435292296 0.0053526797  ## unemployed 0.0056844573 0.0161022760 0.0063922497 0.0006414368  ## unknown 0.0011280441 0.0015704143 0.0008626219 0.0028090509  *-16 means 0.00000000000000022. It is (very much) less than 0.05*  d. Check whether is there any association between Job and Education? |
| **with**  *#<2.2e* |

As both are a categorical variable this can be checked with chisq.test

|  |
| --- |
| **with**(bank\_full, **chisq.test**( job,education))  ##  ## Pearson's Chi-squared test ##  ## data: job and education  ## X-squared = 28483, df = 33, p-value < 2.2e-16 **with**(bank\_full, **table**( job, education) )  ## education  ## job primary secondary tertiary unknown  ## admin. 209 4219 572 171  ## blue-collar 3758 5371 149 454  ## entrepreneur 183 542 686 76  ## housemaid 627 395 173 45  ## management 294 1121 7801 242  ## retired 795 984 366 119 |
| ## self-employed 130 577 833 39  ## services 345 3457 202 150  ## student 44 508 223 163  ## technician 158 5229 1968 242  ## unemployed 257 728 289 29 ## unknown 51 71 39 127  *# OR* **with**(bank\_full, **prop.table**(**table**( job,education)))  ## education  ## job primary secondary tertiary unknown  ## admin. 0.0046227688 0.0933179978 0.0126517883 0.0037822654  ## blue-collar 0.0831213643 0.1187985225 0.0032956581 0.0100418040  ## entrepreneur 0.0040476875 0.0119882330 0.0151732985 0.0016810068  ## housemaid 0.0138683064 0.0087368118 0.0038265024 0.0009953330  ## management 0.0065028422 0.0247948508 0.1725465042 0.0053526797  ## retired 0.0175842162 0.0217646148 0.0080953750 0.0026321028  ## self-employed 0.0028754064 0.0127623808 0.0184247196 0.0008626219  ## services 0.0076308863 0.0764636925 0.0044679392 0.0033177766  ## student 0.0009732145 0.0112362036 0.0049324279 0.0036053173  ## technician 0.0034947247 0.1156576939 0.0435292296 0.0053526797  ## unemployed 0.0056844573 0.0161022760 0.0063922497 0.0006414368  ## unknown 0.0011280441 0.0015704143 0.0008626219 0.0028090509  *#<2.2e-16 means 0.00000000000000022. It is (very much) less than 0.05* |

# R Markdown

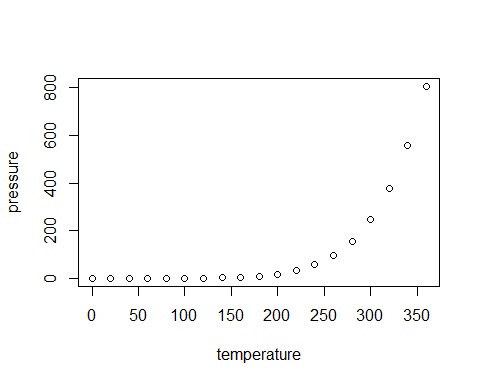
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see [http://rmarkdown.rstudio.com.](http://rmarkdown.rstudio.com/)

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

|  |  |
| --- | --- |
| **summary**(cars) | ## speed dist  ## Min. : 4.0 Min. : 2.00  ## 1st Qu.:12.0 1st Qu.: 26.00  ## Median :15.0 Median : 36.00  ## Mean :15.4 Mean : 42.98  ## 3rd Qu.:19.0 3rd Qu.: 56.00  ## Max. :25.0 Max. :120.00 |
|  |

# Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.