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.



1. Show a distribution of clients based on a Job.
2. bankFullCsv <- './data/bank/bank-full.csv'
3. bankNamesTxt <- './data/bank/bank-names.txt'
4. bankCsv <- './data/bank/bank.csv'
5. bankFull <- read.csv(bankFullCsv, sep = ';')
6. bankSample <- read.csv(bankCsv, sep = ';')
7. webAnalyticsDemoCsv1 <- './data/webAnalyticsDemo/dataset1.csv'
8. webAnalyticsDemoCsv2 <- './data/webAnalyticsDemo/dataset2.csv'
9. webAnalyticsDemoCsv3 <- './data/webAnalyticsDemo/dataset3.csv'
10. webAnalyticsDemo1 <- read.csv(webAnalyticsDemoCsv1)
11. webAnalyticsDemo2 <- read.csv(webAnalyticsDemoCsv2)
12. webAnalyticsDemo3 <- read.csv(webAnalyticsDemoCsv3)
13. naColumns <- **function**(df) {
14. colnames(df)[unlist(lapply(df, **function**(x) any(is.na(x))))]
15. }
16. naColumns(bankSample)
17. head(bankSample)
18. head(bankFull)
19. dim(bankSample)
20. dim(bankFull)
21. **library**('ggplot2')
22. **library**('dplyr')
23. **library**("gridExtra")
24. bankSampleYYes <-filter(bankSample, y == 'yes')
25. bankSampleYNo <-filter(bankSample, y == 'no')
26. yesTermDepositsByAge <- ggplot(bankSampleYYes, aes(age)) + geom\_histogram(binwidth = 5) + labs(title = "Term Deposits Yes by Age Count", x="age", y="Count of Yes")
27. noTermDepositsByAge <- ggplot(bankSampleYNo, aes(age)) + geom\_histogram(binwidth = 5) + labs(title = "Term Deposits No by Age Count", x="age", y="Count of No")
28. grid.arrange(yesTermDepositsByAge, noTermDepositsByAge)

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

ggplot(bankSample, aes(job, y)) + geom\_jitter()

ggplot(bankSample, aes(marital, y)) + geom\_jitter()

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

ggplot(bankSample, aes(education, y)) + geom\_jitter()