Assignment 3.4

1)a. 1. Import the Titanic Dataset from the link Titanic Data Set.

Perform the following:

a. Preprocess the passenger names to come up with a list of titles that represent families

and represent using appropriate visualization graph.

install.packages(Titanic)

library(Titanic)

View(titanic\_train)

library(dplyr)

library(ggplot2)

library(tidyr)

Titanic\_Train\_names<- separate(titanic\_train,Name,c("lastname","First\_name"),sep = ",")

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | 1 | 0 | 3 | Braund | Mr. Owen Harris | male | 22.00 | 1 | 0 | A/5 21171 | 7.2500 |  | S |
| **2** | 2 | 1 | 1 | Cumings | Mrs. John Bradley (Florence Briggs Thayer) | female | 38.00 | 1 | 0 | PC 17599 | 71.2833 | C85 | C |

install.packages("sqldf")

library(sqldf)

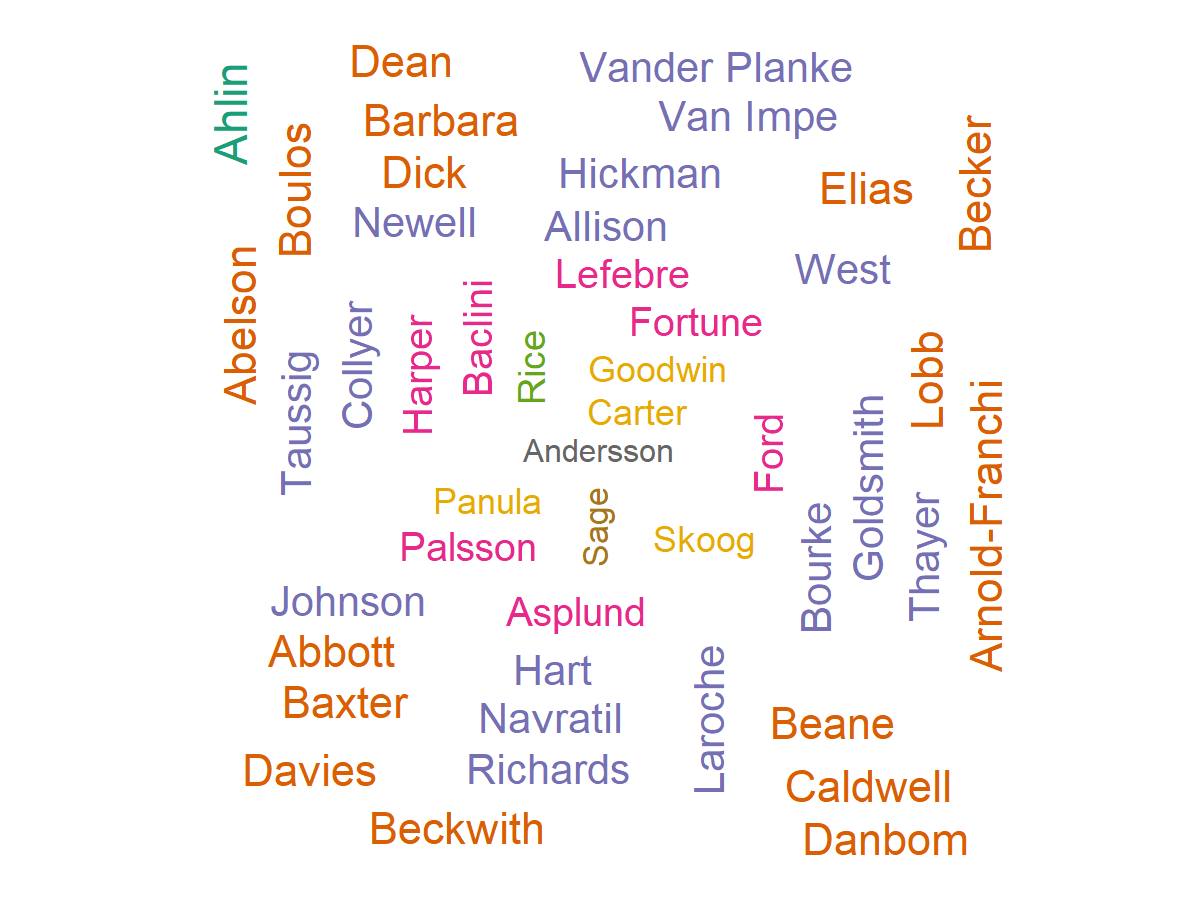
titan\_lastname <- sqldf("select lastname,count(PassengerId)from Titanic\_Train\_names where parch>0 or sibsp>0 group by lastname")

titan\_lastname\_df <- as.data.frame(titan\_lastname)

set.seed(1234)

library(wordcloud)

wordcloud(words = titan\_lastname\_df$lastname,freq = titan\_lastname$`count(PassengerId)`,min.freq = 1,max.words = Inf,scale = c(2,3),random.order = F,rot.per = 0.45,colors = brewer.pal(8,"Dark2"))



1)b. Represent the proportion of people survived from the family size using a graph.

titan\_lastname=sqldf("select lastname, count(PassengerId), sum(Survived) from Titanic\_Train\_names where parch>0 or sibsp>0 group by lastname")

barplot(table(titan\_lastname$`sum(Survived)`, titan\_lastname$`count(PassengerId)`), legend.text = TRUE, beside = TRUE )

