Assignment 1, GR5206

Tushar Ponkshe

September 18, 2018

Part 1

i.

```
titanic <- read.table("C:/Users/tusha/OneDrive/Desktop/Data/titanic.txt", header = TRUE)
is.data.frame(titanic)
```

```
## [1] TRUE
```

ii.

```
nrow(titanic)
```

```
## [1] 891
```

```
ncol(titanic)
```

```
## [1] 12
```

There are 891 rows and 12 columns

iii.

```
titanic$Survived.Word[titanic$Survived == 0] <- "died"
titanic$Survived.Word[titanic$Survived == 1] <- "survived"
mode(titanic$Survived.Word)
```

```
## [1] "character"
```

Part 2

i.

```
subset <- cbind(titanic$Survived, titanic$Age, titanic$Fare)
meanofsubset <- apply(subset, 2, mean)
meanofsubset
```

```
## [1]  0.3838384         NA 32.2042080
```

Mean of Survived is about 0.38, which means that out of 100 people, 38 people survived.

The mean value for Age is NA because some entries in Age are NA or missing, and arithmetic functions on missing values yield NA.

ii.

```
survivedfemaleslogical <- titanic$Sex == "female" & titanic$Survived == 1
survivedfemalesvec <- survivedfemaleslogical[survivedfemaleslogical == TRUE]
survivedfemales <- length(survivedfemalesvec)
totalfemaleslogical <- titanic$Sex == "female"
totalfemalesvec <- totalfemaleslogical[totalfemaleslogical == TRUE]
totalfemales <- length(totalfemalesvec)
proportionsurvived <- round(survivedfemales/totalfemales, digits = 2)
proportionsurvived
```

```
## [1] 0.74
```

The proportion of females passengers who survived is 0.74

iii.

```
totalsurvivors <- length(titanic$Survived[titanic$Survived==1])
prop_FgivenS <- round(survivedfemales/totalsurvivors, digits=2)
prop_FgivenS
```

```
## [1] 0.68
```

Of the survivors, the proportion of females that survived is 0.68

iv.

```
classes <- sort(unique(titanic$Pclass))
Pclass.Survival <- vector("numeric", length = 3)
names(Pclass.Survival) <- classes

for (i in 1:3) {
  Pclass.Survival[i] <- round(mean(titanic$Survived[titanic$Pclass == i]), digits = 2)
}
Pclass.Survival
```

```
##    1    2    3
## 0.63 0.47 0.24
```

v.

```
group <- factor(titanic$Pclass)
survivalrate <- round(tapply(titanic$Survived, group, mean), digits = 2)
survivalrate
```

```
##    1    2    3
## 0.63 0.47 0.24
```

vi. Yes, as the class increases, the survival rate decreases.