KJ Morgan

Data Exercise 2

library(tidyr)

setwd("C:/Users/Kenneth.Morgan/Desktop")

titanic <- read.csv("titanic\_original.csv",header = TRUE, na.strings = c("NA", ""), blank.lines.skip = TRUE)

titanic = titanic[1:1309, ]

View(titanic)

class(titanic) ## [1] "data.frame"

head(titanic)

str(titanic)

names(titanic)

summary(titanic$embarked)

titanic$embarked[is.na(titanic$embarked)] <- "S"

summary(titanic$embarked) ## C = 270, Q = 123, S = 916 (S increased by 2)

mean(titanic$age, na.rm = TRUE) ## [1] 29.88113

titanic$age[is.na(titanic$age)] <- mean(titanic$age, na.rm = TRUE) ## populates the NAs with the mean age

library(dplyr)

titanic <- mutate(titanic, has\_cabin\_number = ifelse(cabin == "NA", 0, 1))

titanic$has\_cabin\_number[is.na(titanic$has\_cabin\_number)] <- "0"

write.csv(titanic, "titanic\_clean.csv")