library(dplyr)

library(tidyr)

library(stringr)

library(lubridate)

load("C:/Users/sec/Downloads/midterm2020.RData")

Q1.

str(pums.sample)

head(pums.sample)

Q2.

pums.sample$SEX <- as.character(pums.sample$SEX)

pums.sample$SEX <- str\_replace(pums.sample$SEX, "1", "Male")

pums.sample$SEX <- str\_replace(pums.sample$SEX, "2", "Female")

pums.sample$SEX <- as.factor(pums.sample$SEX)

Q3.

pums.sample$MAR <- as.character(pums.sample$MAR)

named\_vector <- c("1" = "Married", "2" = "Widowed", "3"= "Divorced", "4" = "Separated", "5" = "Never married or under 15 years old")

pums.sample$MAR <- sapply(pums.sample$MAR, function(x){named\_vector[x]})

head(pums.sample)

Q4.

colSums(is.na(pums.sample))

nrow(pums.sample)

Q5.

pums.sample.male <- pums.sample[which(pums.sample$SEX == "Male"), ]

nrow(pums.sample.male)

colSums(is.na(pums.sample.male))

pums.sample.female <- pums.sample[which(pums.sample$SEX == "Female"), ]

pums.sample.young\_f <- filter(pums.sample.female, AGEP<15)

pums.sample.female[which(pums.sample.male$AGEP<15), ]

filter(pums.sample, AGEP ==12)

pums.sample.young\_f

pums.sample.female$AGEP < 15

pums.sample.old\_f <- filter(pums.sample.female, AGEP>50)

colSums(is.na(pums.sample.old\_f))

nrow(pums.sample.old\_f)

Q6.

summary(pums.sample)

boxplot(pums.sample$WKHP)

boxplot(pums.sample$PINCP)

Q7.

df <- pums.sample[, c("COW", "SCHL")]

table(df)

nrow(pums.sample)

Q8.

summary(pums.sample$AGEP)

cut\_point <- cut(pums.sample$AGEP, breaks = c(20,30,40,50,60,84), right = F, labels = c("20s", "30s", "40s", "50s", "over 60s"))

cut\_point

cut\_point[is.na(cut\_point)] <- "20s"

pums.sample$age\_group <- cut\_point

tapply(pums.sample$PINCP, pums.sample$age\_group, mean)

Q9.

plot(pums.sample$WKHP, pums.sample$PINCP)

cor(pums.sample$WKHP, pums.sample$PINCP)

Q10.

save(pums.sample, file = "21600685.RData")

colnames(iris)

iris\_1 <- gather(iris, Part, value, 1:4)

name <- ifelse((iris\_1$Part == "Sepal.Length") | (iris\_1$Part == "Sepal.Width"), "Sepal", "Petal")

length(name)

iris\_1$name <- name

iris\_1 <- iris\_1[, 1:3]

spread(iris\_1, Part, value)

iris\_1

iris\_1\_length <- filter(iris\_1, (Part == "Sepal.Length" | Part == "Petal.Length"))

iris\_1\_length$Part <- ifelse((iris\_1\_length$Part == "Sepal.Length" | iris\_1\_length$Part == "Petal.Length"), "Length", iris\_1\_length)

iris\_1\_length <- iris\_1\_length[, 2:3]

spread(iris\_1\_length, Part, value)

iris\_1\_Width <- filter(iris\_1, (Part == "Sepal.Width" | Part == "Petal.Width"))

iris\_1\_Width$Part <- ifelse((iris\_1\_Width$Part == "Sepal.Width" | iris\_1\_Width$Part == "Petal.Width"), "Width", iris\_1\_Width)

iris\_1\_Width <- iris\_1\_Width[, 2:3]

iris\_1\_Width

iris.wide <- iris\_1["Species"]

iris.wide$name <- name

iris.wide <- cbind(iris.wide, iris\_1\_length)

iris.wide <- cbind(iris.wide, iris\_1\_Width)

iris.wide

iris\_1 <- gather(iris, Part, value, 1:4)

iris\_1$name <- ifelse((iris\_1$Part == "Sepal.Length" | iris\_1$Part == "Sepal.Width"), "Sepal", "Petal")

iris\_1

iris\_1$Length <- ifelse((iris\_1$Part == "Sepal.Length" | iris\_1$Part == "Petal.Length"), "Length", "Width")

iris\_1 <- iris\_1[, c(1,3,4,5)]

iris\_1 <- iris\_1[, c(1, 2, 4)]

iris\_1

nrow(iris\_1)

iris\_1$id <- 1:600

iris\_1 <- iris\_1[, 1:4]

iris\_1 <- iris\_1[, 2:4]

iris\_1

spread(iris\_1, Length, value)

iris\_2 <- iris\_1[, c(2, 4)]

colnames(iris\_1)[3] <- "Part"

spread(iris\_1, Char, value)

iris\_2 <- iris\_1[, c(1,2,4)]

spread(iris\_1, Char, value)

iris\_2

iris\_2 <- iris\_1[, c(1, 3, 4, 5)]

iris\_2 <- iris\_1[, c(1, 2, 4)]

iris\_2$id <- 1:600

spread(iris\_2, Length, value)

iris\_2 <- iris\_1[, c(1, 3,4,5)]

colnames(iris\_2)[2] <- "Value"

colnames(iris\_2)[4] <- "Measure"

colnames(iris\_2)[3] <- "Part"

iris.tidy <- iris\_2

save(pums.sample, iris.tidy, file = "21600685.RData")