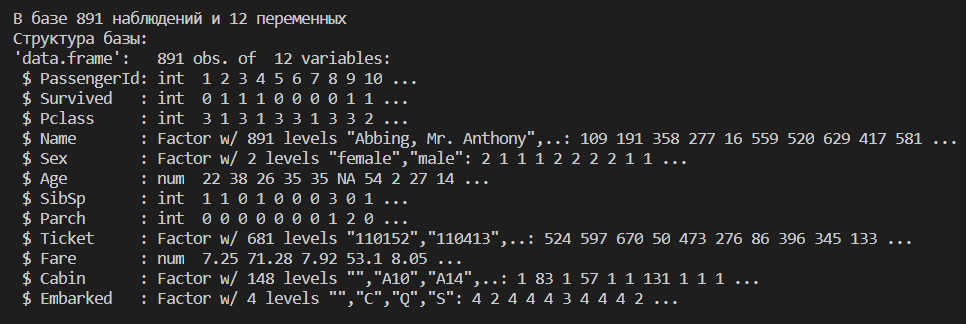
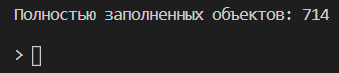
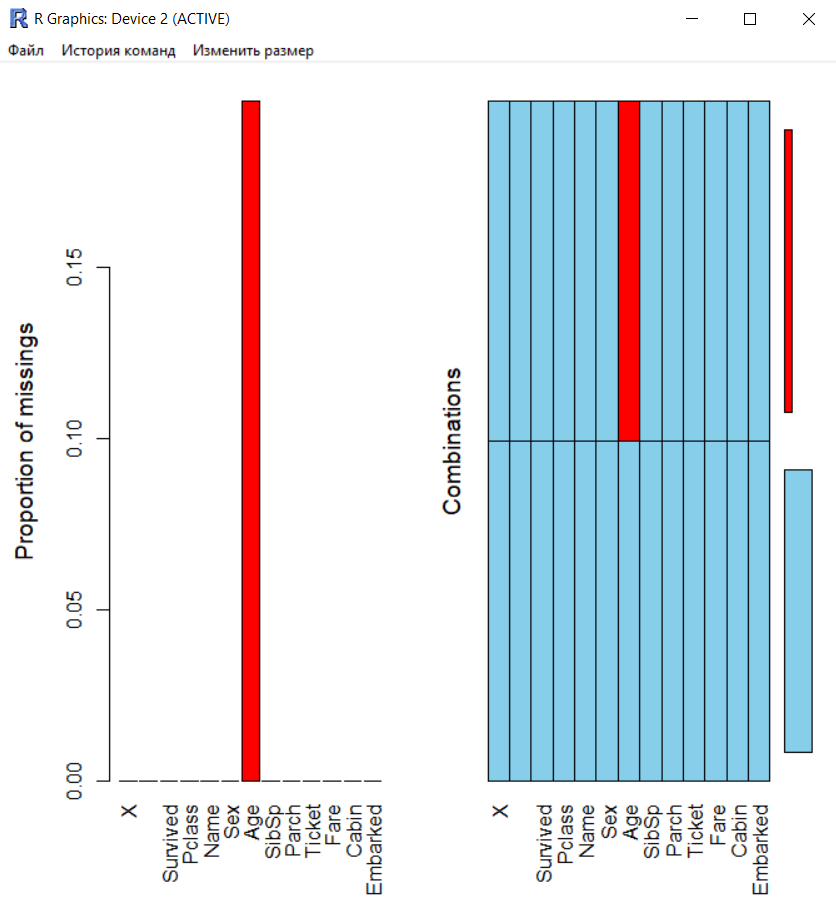
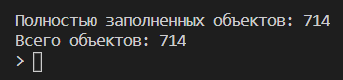
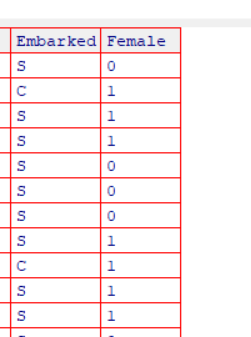
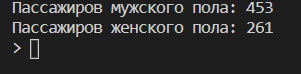
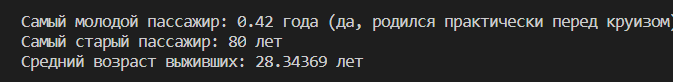
Результаты выполнения:

1. Графики
   1. 
   2. 
   3.   
      Больше всего пропущенных значений было в переменной возраста
   4. Вряд ли это системно, так как прочие поля, такие как место посадки, количество родственников и класс билета, либо пол никак не влияли на данный параметр.
   5. 
2. Наборы данных
   1. 
   2. 
   3. 

Код программы:

# ================================

#            Графики

# ================================

# Задание 1.1

task11 <- function(){

    # Скачиваем и записываем файл

    df <- read.csv("https://raw.githubusercontent.com/agconti/kaggle-titanic/master/data/train.csv")

    write.csv(df, file = "df.csv")

    cat("В базе", nrow(df), "наблюдений и",  ncol(df), "переменных\nСтруктура базы:\n")

    cat(str(df))

}

# Задание 1.2

task12 <- function(){

    # Читаем файл уже с диска

    df <- read.csv("df.csv")

    cat("Полностью заполненных объектов:", sum(complete.cases(df)), "\n")

    # Фильтруем незаполненные объекты

    df\_na <- df[!complete.cases(df),]

    write.csv(df\_na, file = "df\_na.csv")

    View(df\_na)

}

# Задание 1.3

task13 <- function(){

    # install.packages(c("mice", "VIM"))

    library(mice)

    library(VIM)

    df <- read.csv("df.csv");

    aggr(df)

}

# Задание 1.4

task14 <- function(){

    library(mice)

    library(VIM)

    df <- read.csv("df.csv");

    aggr(df)

}

# Задание 1.5

task15 <- function(){

    df <- read.csv("df.csv");

    df <- na.omit(df);

    cat("Полностью заполненных объектов:", sum(complete.cases(df)), "\nВсего объектов:", nrow(df))

}

# ================================

#           Базы данных

# ================================

# Задание 2.1

task21 <- function(){

    # Загружаем теперь уже чистую базу данных

    df <- read.csv("df.csv")

    df <- na.omit(df)

    temp <- sub("female", "1", df$Sex)

    temp <- as.numeric(sub("male", "0", temp))

    df$Female <- factor(temp)

    View(df)

}

# Задание 2.2

task22 <- function(){

    df <- read.csv("df.csv")

    df <- na.omit(df)

    df2 <- subset(df, Age > 25)

    df2 <- subset(df2, Age <= 45)

    View(df2)

}

# Задание 2.3

task23 <- function(){

    df <- read.csv("df.csv")

    df <- na.omit(df)

    cat("Пассажиров мужского пола:",nrow(subset(df, Sex == "male")),'\n')

    cat("Пассажиров женского пола:",nrow(subset(df, Sex == "female")))

}

# Задание 2.4

task24 <- function(){

    df <- read.csv("df.csv")

    df <- na.omit(df)

    ageData <- df$Age

    cat("Самый молодой пассажир:", min(ageData), "года (да, родился практически перед круизом)\n")

    cat("Самый старый пассажир:", max(ageData), "лет\n")

    sdf <- subset(df, Survived == "1")

    survivedAgeData <- sdf$Age

    cat("Средний возраст выживших:", mean(survivedAgeData), "лет\n")

}

# ================================

#          Запуск задания

# ================================

# Метод запуска заданияS

startTask <- function(arg){

    shell("cls");

    switch(

      arg,

      "1.1" = task11(),

      "1.2" = task12(),

      "1.3" = task13(),

      "1.4" = task14(),

      "1.5" = task15(),

      "2.1" = task21(),

      "2.2" = task22(),

      "2.3" = task23(),

      "2.4" = task24(),

    )

    cat('\n');

}

# Непосредственно запуск задания

# startTask("1.1")

# startTask("1.2")

# startTask("1.3")

# startTask("1.4")

# startTask("1.5")

# startTask("2.1")

# startTask("2.2")

# startTask("2.3")

startTask("2.4")