**# Считывание данных**

?read.table #знак вопроса впереди вызывает справку по функции, функция read.table считывает произвольную таблицу

?read.csv #считывает файл с расширением .csv

mydata <- read.csv('evals.csv') #присваиваем mydata содержимое таблицы evals (датасет с оценками работы преподавателя)

#(этот файл нужно скопировать в текущую директорию), mydata является датафреймом, т.е. табличной структурой

**# Вывод итогов**

head(mydata, 3) #считываем "шапку" таблицы, т.е. верхние значения. Второй параметр отвечает за количество считываемых строк (по умолчанию 6)

tail(mydata) # "хвост" таблицы

View(mydata) #запускает как в редакторе

str(mydata) #выводит информацию по всем переменным (столбцам) датафрейма

a <- names(mydata) #имена переменных

summary(mydata) #описательные статистики (среднее, медиана, минимум, максимум, 1-й и 3-й квартили)

**# Работа с переменными датафрейма**

b <- mydata$score #обращение через "$" к конкретной переменной датафрейма

mean(mydata$score) #считаем среднее по переменной

summary(mydata$score) # выводим статистику по переменной

mydata$score \* 2 #умножаем все элементы переменной

mydata$ten\_point\_scale <- mydata$score \* 2 #добавляем новую переменную в датафрейм посредством операции присваивания

#(в контексте данной задачи переходим от оценки в 5-бальной шкале к 10-бальной)

summary(mydata$ten\_point\_scale) #статистика по новой переменной

mydata$new\_varible <- 0 #добавляем новую переменную со значением 0

mydata$number <- 1:nrow(mydata) #добавляем ещё одну переменную со значениями от 1 до количества строк в my\_data

summary(mydata$number) #статистика

nrow(mydata) #количество строк

ncol(mydata) #количество столбцов

# выбор подмножества

mydata$score[1:10] #первые 10 значений переменной

mydata[1,1] #значение из первой строки и первого столбца

mydata[c(2,193,225),1] #конкретные строки из первой переменной

mydata[101:200,1] #интервал строк из первой переменной

mydata[5,] #все столбцы пятой строки

mydata[,1] #все строки 1-го столбца

mydata[,2:5] # все строки столбцов 2-5

head(mydata[,2:5]) #выводим первые 6 строк

**# Выбор подмножества по условию (без индексации)**

mydata$gender

mydata$gender == 'female'

head(mydata[mydata$gender == 'female',1:3]) #выводим шапку первых 3 столбцов, для которых выполнено условие, что пол женский

head(subset(mydata, gender == 'female',1:3)) #через subset, без индексации тоже самое. Выводим только данные по женщинам

head(subset(mydata, score > 3.5)) #данные по тем, у кого оценка выше 3.5

**# Объединение датафреймов**

mydata2 <- subset(mydata, gender == 'female') #выбираем женщин

mydata3 <- subset(mydata, gender == 'male') #выбираем мужчин

mydata4 <- rbind(mydata2, mydata3) #соединяем датафреймы при условии одинаковых столбов, в новом датафрейме сначала будут женщины, а потом мужчины

mydata5 <- mydata[,1:10] #с первого по 10-й столбец

mydata6 <- mydata[,11:24] #с 11 по 24-й

mydata7 <- cbind(mydata6, mydata5) #соединяем датафрейм по столбцам, порядок при этом не меняется

library(help="datasets") #информация о встроенных в R датасетах

data(mtcars) #добавляет dataset в рабочую среду

help(mtcars) #информация по датасету

mydata1 <-mtcars #запишем датасет в переменную