Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Факультет компьютерных технологий и прикладной математики**

**Кафедра вычислительных технологий**

**Отчёт лабораторная №3**

**Дисциплина: ОБРАБОТКА БОЛЬШЫХ ДАННЫХ**

Работу выполнил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Дука В.А.

Направление подготовки 02.03.02 Фундаментальная информатика и

информационные технологии

Преподаватель \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Яхонтов А.А.

Краснодар

2025

**Тема:** работа с графиками и диаграммами в R.

**Цель:** научиться подготавливать наборы данных для анализа. Научиться работать с графиками

**Задания к лабораторной работе:**

1. Собрать данные согласно варианту по указанному виду спорта (лыжные гонки) и в заданной стране (Россия).

2. Вывести графики динамики олимпийских достижений заданной страны по виду спорта относительно временной шкалы: столбчатую диаграмму по количеству мест 1-8 (спортсменов заданных стран) по каждой Олимпиаде по назначенному виду спорта, круговую диаграмму по количеству первых мест в каждой из олимпиад, функциональные графики - тенденции изменения количества призовых мест отдельно по мужчинам и женщинам за последние 30 лет.

3. Вывести графики изменения спортивных достижений 1) по золотым медалям и 2) по призовым 3-местам по 7-и странам-призерам (разными цветами и точками) за последние 6 олимпиад.

4. Выбрать информацию за последние 6 олимпиад по заданному виду спорта, аккумулировать данные по мужчинам и женщинам (отобразить на одном графике разными линиями), вывести динамику и статистику на графики (функциональный, столбчатый, круговой).

**Ход работы:**

1. Для начала собрал данные об олимпиадных результатах в лыжных гонках в стране Россия за последние 30 лет по женщинам, мужчинам, за последние 6 олимпиад для 7-ми стран о первых местах, призовых местах в файл Olimp.xlsx c 4-мя соответствующими листами и с помощью библиотеки readxl загрузил в R:

library(readxl)

df\_woman <- read\_excel("Olimp.xlsx", sheet = 1)

df\_man <- read\_excel("Olimp.xlsx", sheet = 2)

df\_gold <- read\_excel("Olimp.xlsx", sheet = 3)

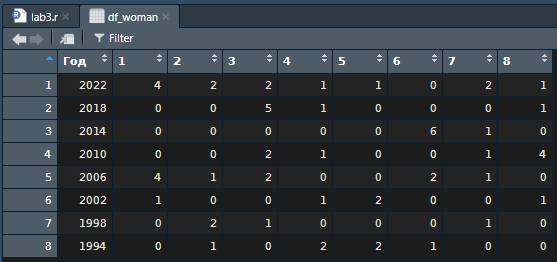
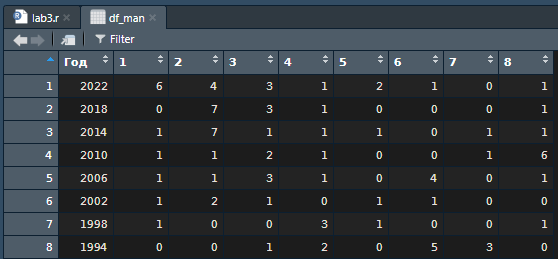
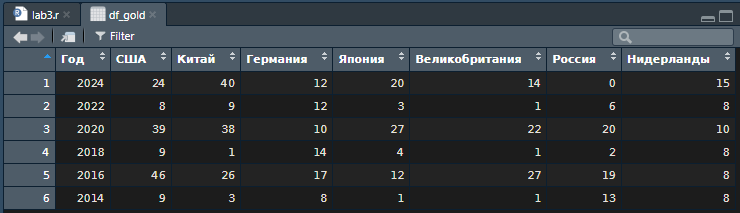
df\_prize\_places <- read\_excel("Olimp.xlsx", sheet = 4)

Рисунок 1 – дата-сет для женщин на олимпиадах за последние 30 лет по лыжным гонкам в России

Рисунок 2 – дата-сет для мужчин на олимпиадах за последние 30 лет по лыжным гонкам в России

Рисунок 3 – дата-сет о количестве первых мест в каждой из 7 стран лидеров за последние 6 олимпиад

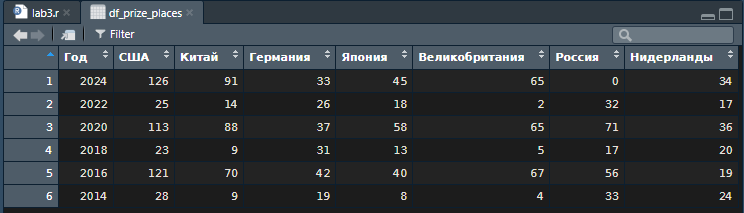
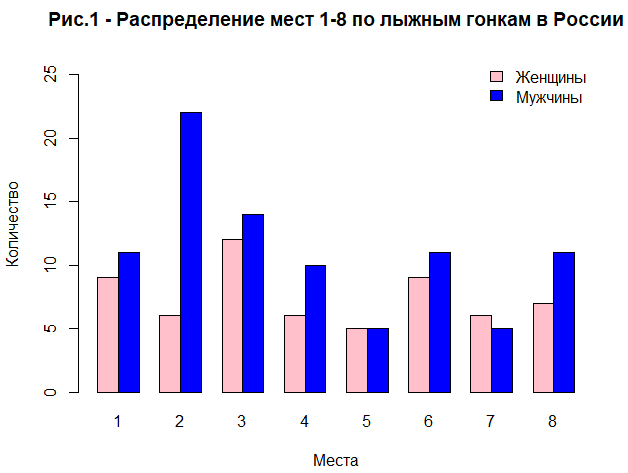
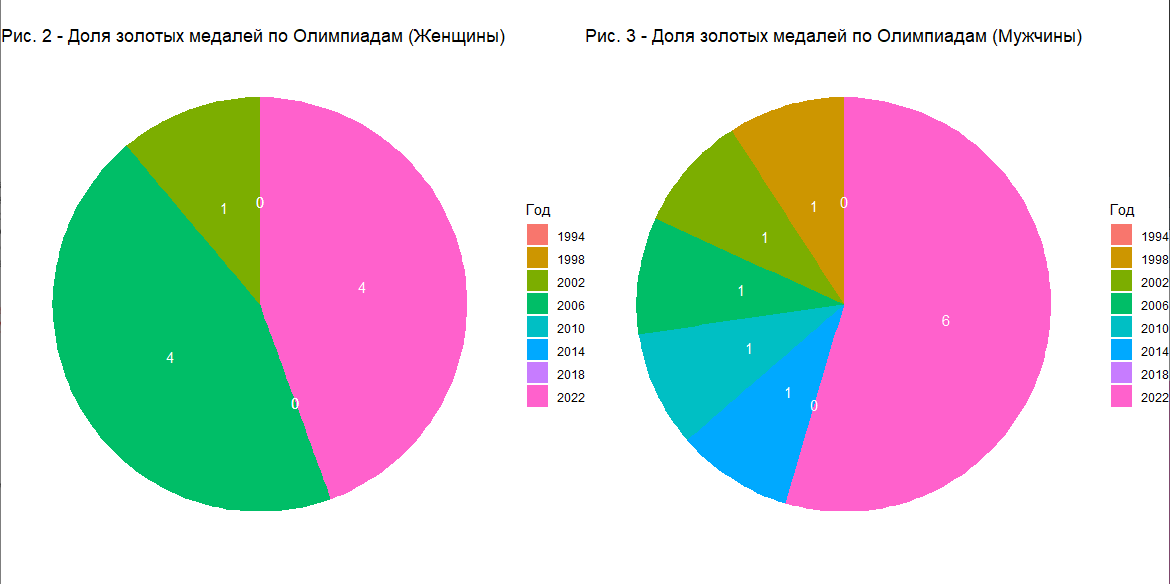


Рисунок 4 – дата-сет о количестве призовых мест в каждой из 7 стран лидеров за последние 6 олимпиад

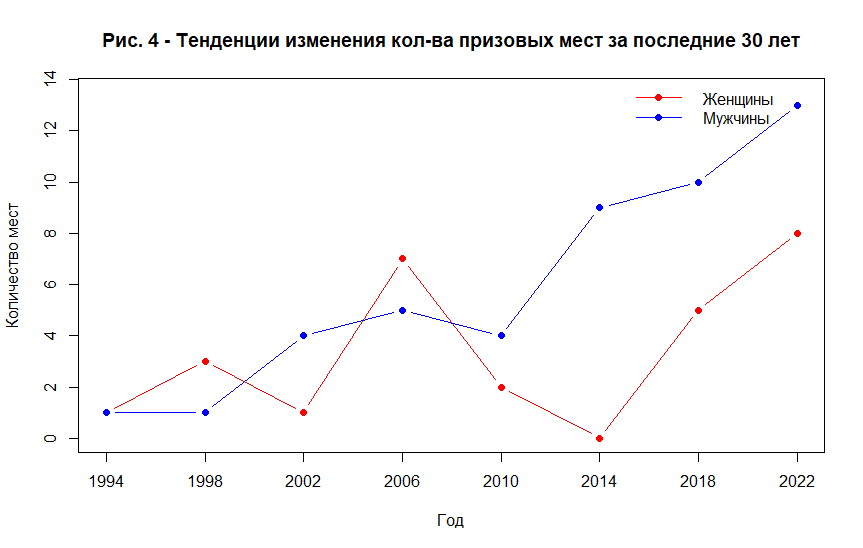
1. После этого разработал графики динамики олимпийских достижений за последние 30 лет по лыжным гонках в России для женщин и мужчин

*1)* *Столбчатая диаграмма:*

На графике видно, что мужчины демонстрируют более высокие результаты, чем женщины, с пиком на 2-м месте (около 20 медалей), в то время как женщины чаще занимают 1-е и 3-е места (около 10 медалей), а в остальных позициях (4–8) обе группы показывают схожие, но менее значительные достижения.

*2) Круговые диаграммы:*

Круговые диаграммы иллюстрируют долю золотых медалей, завоёванных российскими лыжницами и лыжниками на Олимпийских играх с 1994 по 2022 годы, показывая, что женщины чаще выигрывали золото в 1994, 1998 и 2002 годах (по 4 медали), тогда как мужчины доминировали в 2002 году (6 медалей), с меньшим вкладом в другие годы, включая 1994, 1998, 2010, 2014 и 2018 (по 1 медали).

*3) Функциональный график:*

Функциональный график демонстрирует тенденции изменения количества призовых мест, завоёванных российскими лыжницами и лыжниками за последние 30 лет (1994–2022), выявляя пиковые достижения женщин в 2002 году (8 мест) с последующим спадом к 2014 году (1 место) и восстановлением к 2022 году (6 мест), тогда как мужчины показывают более стабильную динамику с ростом с 1994 года (1 место) до пика в 2018 году (10 мест) и незначительным снижением к 2022 году (8 мест).

3. Далее вывести графики изменения спортивных достижений по 7-и странам-призерам (разными цветами и точками) за последние 6 олимпиад по:

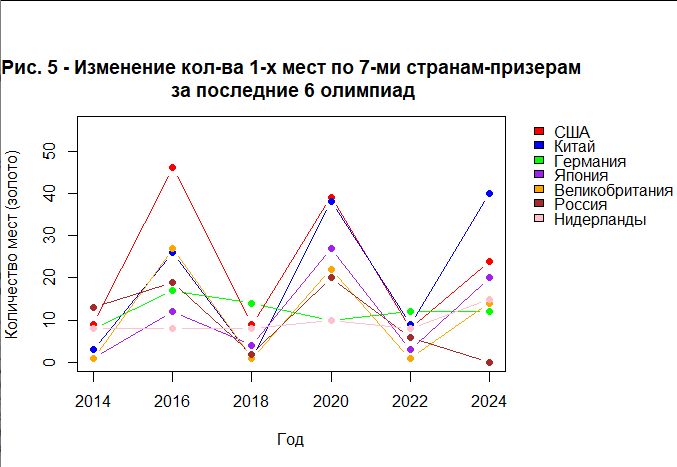
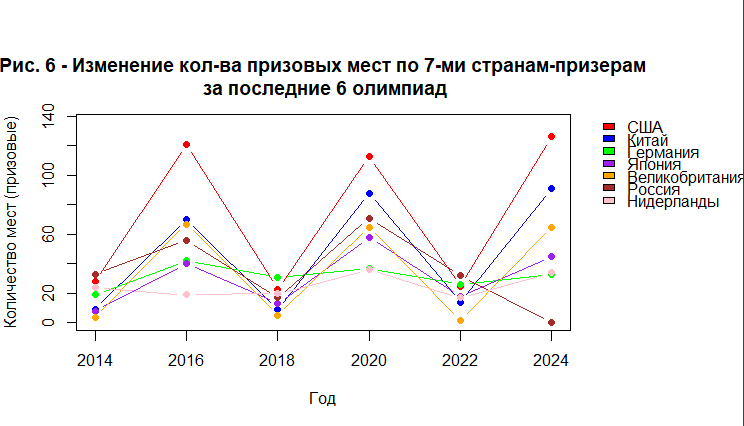
*1) по золотым медалям (первые места):*

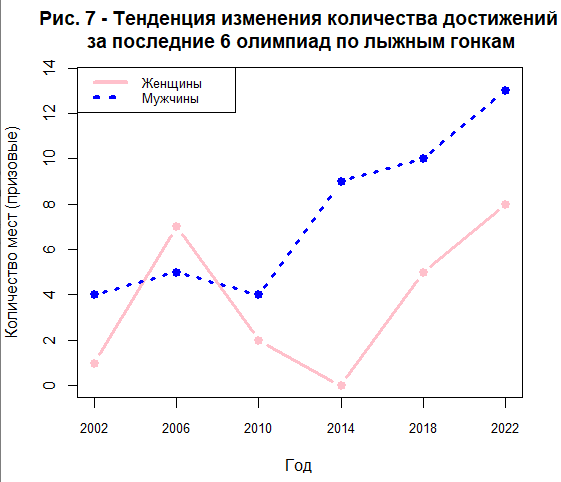
График демонстрирует чередование результатов между летними и зимними играми с примерно одинаковыми показателями, но с общей положительной тенденцией роста, особенно заметной у США и России, достигших пика в 2024 году (около 40 и 30 мест соответственно).

*2) по призовым местам (1-3 места):*

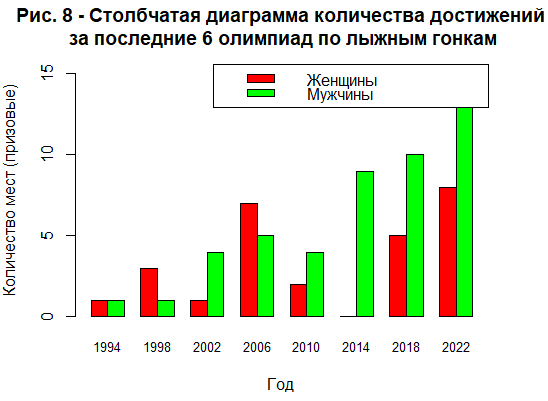
**

Линейный график иллюстрирует изменение количества призовых мест семи стран-призёров на Олимпийских играх с 2014 по 2024 годы, показывая пиковые достижения США и России в 2016 и 2024 годах (до 120 мест), с общей тенденцией роста, несмотря на колебания у других стран, таких как Китай и Германия

4. Выбрать информацию за последние 6 олимпиад по заданному виду спорта, аккумулировать данные по мужчинам и женщинам (отобразить на одном графике разными линиями), вывести динамику и статистику на графики

*1) функциональный график:*

Функциональный график отражает тенденции изменения количества достижений российских лыжников за последние 6 Олимпиад (2002–2022), демонстрируя пиковое значение у женщин в 2002 году (6 мест) с последующим спадом к 2014 году (1 место) и ростом к 2022 году (8 мест), тогда как мужчины показывают устойчивый рост с 2002 года (4 места) до 2022 года (12 мест), указывая на положительную динамику.

*2) Столбчатая диаграмма:*

Столбчатая диаграмма отражает количество достижений российских лыжников за последние 6 Олимпиад (2002–2022), показывая рост у женщин с 2 мест в 2002 году до 8 в 2022 году, тогда как мужчины демонстрируют более стабильные результаты с пиком в 2018 году (10 мест) и снижением к 2022 году (8 мест).

*3) круговая диаграмма:*

Круговая диаграмма демонстрирует распределение общего количества призовых мест, завоёванных российскими лыжниками за последние 6 Олимпиад (2002–2022), выявляя доминирование женщин с 47 местами (63,5% от общего числа) над мужчинами, имеющими 27 мест (36,5%), что подчёркивает более высокую результативность женской сборной в данном виде спорта.

**Код программы:**

library(readxl)

library(gridExtra)

library(ggplot2)

#1. Собираем данные из дата сета

df\_woman <- read\_excel("Olimp.xlsx", sheet = 1)

df\_man <- read\_excel("Olimp.xlsx", sheet = 2)

legend\_colors <- c("red", "blue", "green", "purple", "orange", "brown", "pink")

#2. Столбчатые диаграммы

df\_woman\_count <- apply(df\_woman[,-1], 2, sum)

df\_man\_count <- apply(df\_man[,-1], 2, sum)

combined\_counts <- rbind(df\_woman\_count, df\_man\_count)

rownames(combined\_counts) <- c("Женщины", "Мужчины")

# Строим объединённый график

barplot(combined\_counts,

beside = TRUE,

names.arg = colnames(df\_woman)[-1],

col = c("pink", "blue"),

main = "Рис.1 - Распределение мест 1-8 по лыжным гонкам в России",

xlab = "Места",

ylab = "Количество",

ylim = c(0, max(combined\_counts) \* 1.2))

legend("topright",

legend = c("Женщины", "Мужчины"),

fill = c("pink", "blue"),

bty = "n")

# Подсчитываем золотые медали по годам

gold\_woman\_by\_year <- data.frame(

Year = df\_woman$Год,

Gold = df\_woman[[2]] # Второй столбец — золотые медали

)

gold\_woman\_by\_year <- aggregate(Gold ~ Year, data = gold\_woman\_by\_year, sum)

gold\_man\_by\_year <- data.frame(

Year = df\_man$Год,

Gold = df\_man[[2]] # Второй столбец — золотые медали

)

gold\_man\_by\_year <- aggregate(Gold ~ Year, data = gold\_man\_by\_year, sum)

#2. Создаём круговые диаграммы с подписями

p\_woman <- ggplot(gold\_woman\_by\_year, aes(x = "", y = Gold, fill = as.factor(Year))) +

geom\_bar(stat = "identity", width = 1) +

coord\_polar("y", start = 0) +

geom\_text(aes(label = Gold), position = position\_stack(vjust = 0.5), size = 4, color = "white") +

labs(title = "Рис. 2 - Доля золотых медалей по Олимпиадам (Женщины)", fill = "Год") +

theme\_void() +

theme(legend.position = "right")

p\_man <- ggplot(gold\_man\_by\_year, aes(x = "", y = Gold, fill = as.factor(Year))) +

geom\_bar(stat = "identity", width = 1) +

coord\_polar("y", start = 0) +

geom\_text(aes(label = Gold), position = position\_stack(vjust = 0.5), size = 4, color = "white") +

labs(title = "Рис. 3 - Доля золотых медалей по Олимпиадам (Мужчины)", fill = "Год") +

theme\_void() +

theme(legend.position = "right")

# Выводим две диаграммы рядом

grid.arrange(p\_woman, p\_man, ncol = 2)

#2. Функциональные графики

plot\_graph <- function(df1, df2 = NULL, colors, title) {

total\_places1 <- rowSums(df1[, 2:4], na.rm = TRUE)

if (!is.null(df2)) {

total\_places2 <- rowSums(df2[, 2:4], na.rm = TRUE)

data\_to\_plot <- cbind(total\_places1, total\_places2)

} else {

data\_to\_plot <- total\_places1

}

years <- df1$Год

# Строим график

matplot(years, data\_to\_plot, main = title, type = "b", pch = 19, lty = 1, col = colors[1:2], # Используем только 2 цвета

xlab = "Год", ylab = "Количество мест", xaxt = "n",

ylim = c(0, max(data\_to\_plot, na.rm = TRUE) + 0.5))

axis(1, at = years, labels = years)

if (!is.null(df2)) {

legend("topright",

legend = c("Женщины", "Мужчины"),

col = colors[1:2],

lty = 1,

pch = 19,

bty = "n")

}

}

#2. Графики изменения призовых мест

plot\_graph(df\_woman, df\_man, legend\_colors, "Рис. 4 - Тенденции изменения кол-ва призовых мест за последние 30 лет")

#3. Графики изменения золотых медалей (1 мест) по странам

df\_gold <- read\_excel("Olimp.xlsx", sheet = 3)

par(mar = c(5, 4, 4, 9))

matplot(df\_gold$Год,

df\_gold[, -1],

main = "Рис. 5 - Изменение кол-ва 1-х мест по 7-ми странам-призерам\n за последние 6 олимпиад",

type = "b", pch = 19,

lty = 1,

col = legend\_colors,

xlab = "Год", ylab = "Количество мест (золото)",

xaxt = "n",

ylim = c(0, max(df\_gold[, -1] + 10)))

axis(1, at = df\_gold$Год, labels = df\_gold$Год)

legend("topright", legend = colnames(df\_gold)[-1], xpd = TRUE, bty = "n", inset = -c(0.35, 0), fill = legend\_colors)

#3. Графики изменения всех призовых мест по странам

df\_prize\_places <- read\_excel("Olimp.xlsx", sheet = 4)

par(mar = c(5, 4, 4, 9))

matplot(df\_prize\_places$Год,

df\_prize\_places[, -1],

main = "Рис. 6 - Изменение кол-ва призовых мест по 7-ми странам-призерам\n за последние 6 олимпиад",

type = "b", pch = 19,

lty = 1,

col = legend\_colors,

xlab = "Год", ylab = "Количество мест (призовые)",

xaxt = "n",

ylim = c(0, max(df\_prize\_places[, -1] + 10)))

axis(1, at = df\_prize\_places$Год, labels = df\_prize\_places$Год)

legend("topright", legend = colnames(df\_prize\_places)[-1], xpd = TRUE, bty = "n", inset = -c(0.35, 0), fill = legend\_colors)

#4. Динамика лыжных гонок за последние 6 Олимпиад

# Подсчёт призовых мест

total\_places2 <- rowSums(df\_woman[, 2:4], na.rm = TRUE)

total\_places3 <- rowSums(df\_man[, 2:4], na.rm = TRUE)

total\_places\_all <- cbind(df\_woman[, 1], total\_places2, total\_places3)

colnames(total\_places\_all) <- c("Год", "Женщины", "Мужчины")

total\_places\_all <- total\_places\_all[order(total\_places\_all$Год), ]

n\_rows <- nrow(total\_places\_all)

year <- total\_places\_all[(n\_rows-5):n\_rows, ] # Берем последние 6 строк

par(oma = c(0, 0, 2, 0))

matplot(year$Год,

year[, -1],

type = "b",

pch = 19,

lty = c(1, 3),

col = c("pink", "blue"),

xlab = "Год",

ylab = "Количество мест (призовые)",

xaxt = "n",

lwd = 3,

ylim = c(0, max(year[, -1]) + 0.5),

main = "Рис. 7 - Тенденция изменения количества достижений\n за последние 6 олимпиад по лыжным гонкам")

axis(1, at = year$Год, labels = year$Год, las = 1, cex.axis = 0.8)

legend('topleft', c('Женщины', 'Мужчины'), lty = c(1, 3), cex = 0.8, lwd = 4, col = c("pink", "blue"))

# 4. Столбчатая диаграмма

total\_places\_barplot <- total\_places\_last\_6 # Используем данные только за последние 6 лет

colnames(total\_places\_barplot) <- c("Год", "Женщины", "Мужчины")

# Транспонируем и сортируем (хотя сортировка уже не нужна, так как данные отсортированы)

total\_places\_barplot\_t <- t(total\_places\_barplot)

total\_places\_barplot\_t\_sorted <- total\_places\_barplot\_t[, order(total\_places\_barplot\_t[1, ])]

# Строим столбчатую диаграмму

barplot(as.matrix(total\_places\_barplot\_t\_sorted[-1, ]),

beside = TRUE,

col = c("red", "green"),

xlab = "Год",

names.arg = total\_places\_barplot\_t\_sorted[1, ],

ylab = "Количество мест (призовые)",

ylim = c(0, max(total\_places\_barplot\_t\_sorted[-1, ], na.rm = TRUE) \* 1.2), # Динамический ylim

cex.names = 0.8,

main = "Рис. 8 - Столбчатая диаграмма количества достижений\n за последние 6 олимпиад по лыжным гонкам")

legend('topright', c('Женщины', 'Мужчины'), fill = c("red", "green"))

# 4. Круговая диаграмма

# Подсчитываем общее количество призовых мест за последние 6 Олимпиад

total\_places\_pie <- colSums(total\_places\_barplot[, -1], na.rm = TRUE) # Суммируем по столбцам "Женщины" и "Мужчины"

names(total\_places\_pie) <- c("Женщины", "Мужчины")

# Строим круговую диаграмму

pie(total\_places\_pie,

labels = total\_places\_pie, # Показываем значения на диаграмме

col = c("purple", "orange"),

main = "Рис. 9 - Кол-во призовых мест за последние 6 Олимпиад\n по лыжным гонкам (женщины, мужчины)")

legend('topright', c('Женщины', 'Мужчины'), fill = c("purple", "orange"))

**Вывод:** получилось успешно подготовить и проанализировать набор данных по лыжным гонкам в России, отобразив динамику олимпийских достижений через столбчатые, круговые и функциональные графики, а также сравнительные графики по золотым и призовым местам для семи стран за последние 6 Олимпиад. В результате были выявлены тенденции роста достижений российских лыжников, особенно у женщин, с акцентом на первые места, а также стабильные результаты мужчин, что подчёркивает эффективность применения визуализации данных для анализа спортивных показателей.