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 больших данных**

Работу выполнил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Сгонник Н.С.

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 работать с графиками.

**Задание:**

1. Собрать данные согласно варианту по указанному виду спорта в заданных странах.
2. Вывести графики динамики олимпийских достижений заданной страны по виду спорта относительно временной шкалы: источник (https://olympteka.ru/olymp/sport/country\_medalists/23/ger.html) 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. http://olympteka.ru/olymp/game/medals/51.html, например,
4. Используя тот же источник, 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), используя изученные приемы оформления, в том числе массив графиков на одном поле.

Вариант: 31.

Страна: Россия

Вид спорта: Спортивная гимнастика.

**Ход работы**:

Соберем статистику достижений Германии по спортивной гимнастике за последние 30 лет.

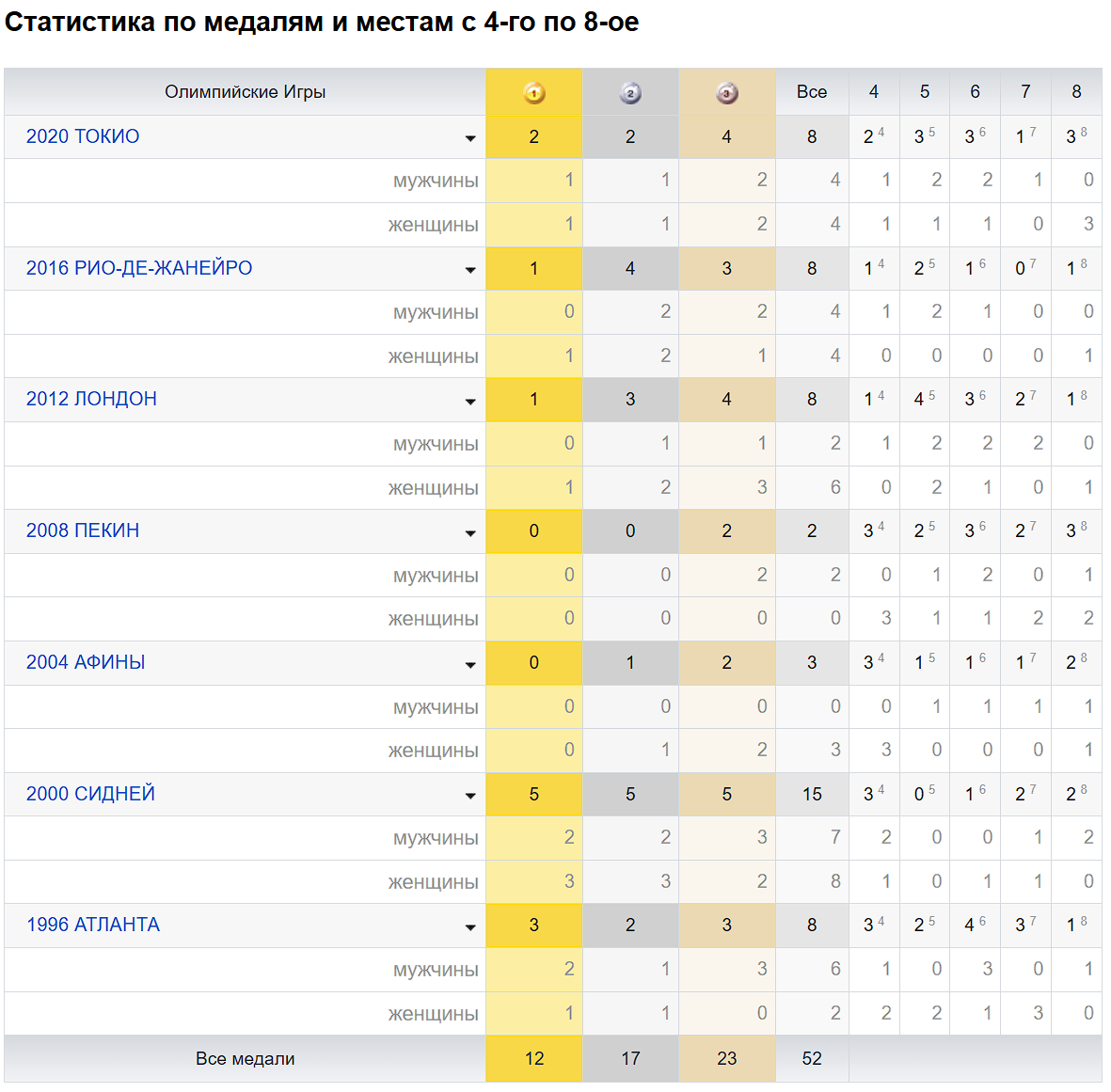


Таблица 1 – Статистика по спортивной гимнастике

Все данные занесем в таблицы эксель, разделяя мужчин, женщин. Спортивная гимнастика проводится только в период летних Олимпийских игр.

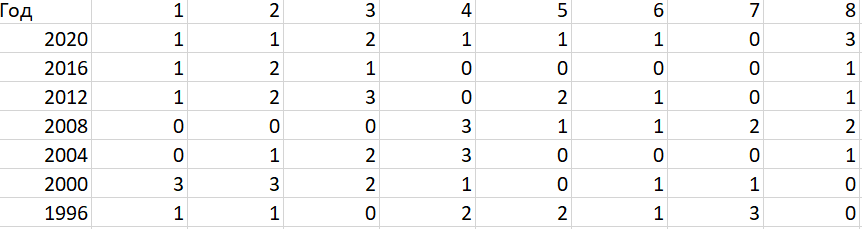


Рисунок 1 – Датасет "Женщины"

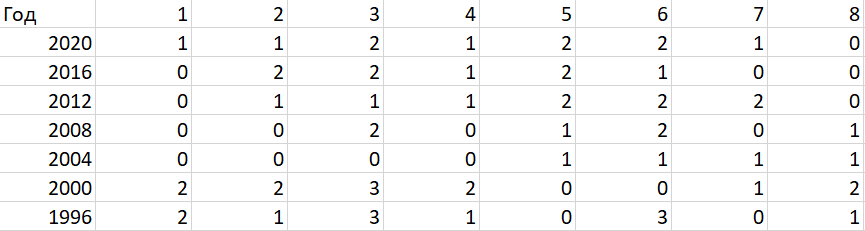


Рисунок 2 – Датасет "Мужчины"

Необходимо построить столбчатые диаграммы по количеству мест 1-8 по каждой Олимпиаде по спортивной гимнастике. Слева выводим статистику по женщинам, справа – по мужчинам, в центре – по миксу, то есть групповым-парным выступлениям.

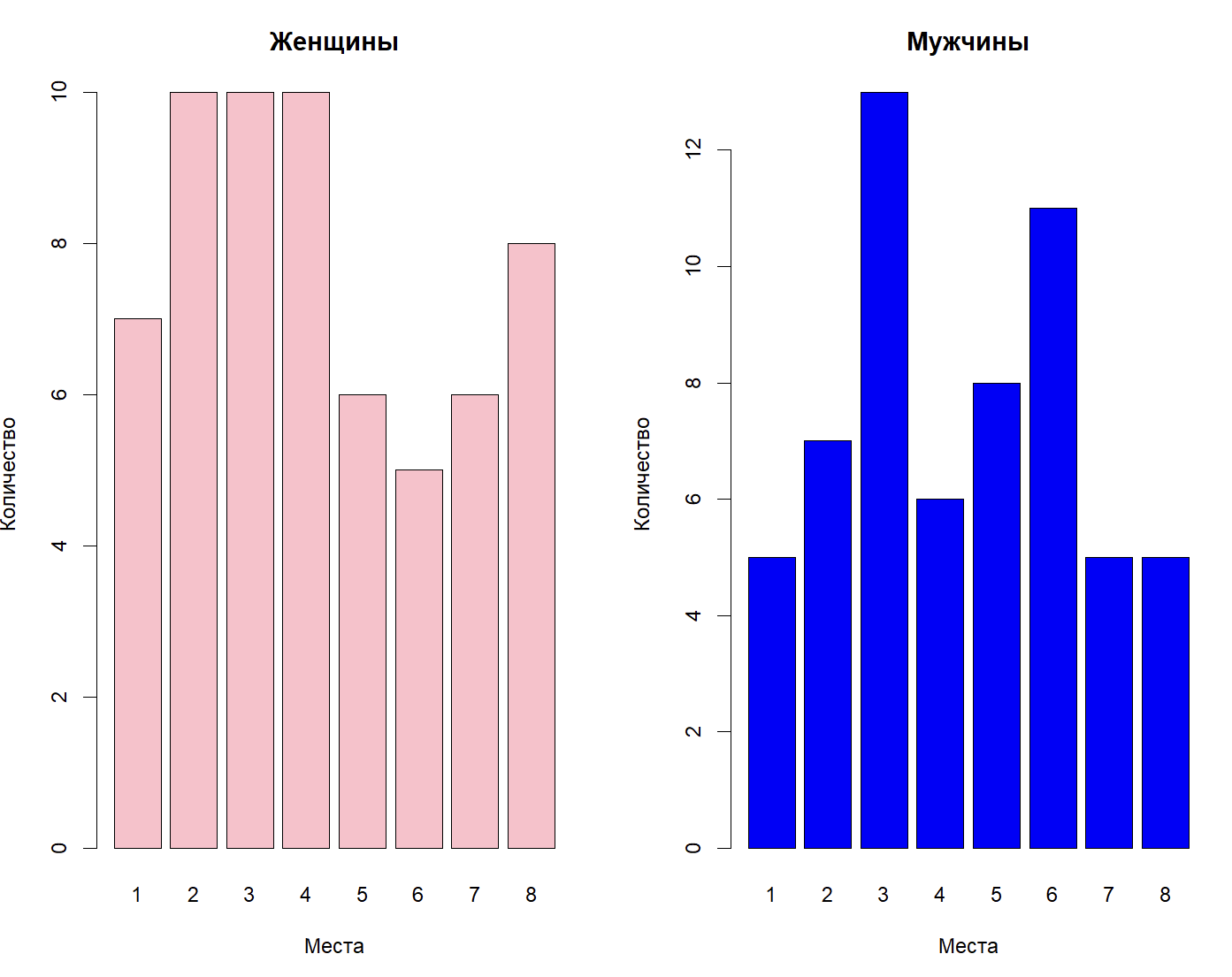


Рисунок 3 – Столбчатые диаграммы количества мест 1-8 по категориям

По построенным диаграммам можно сказать, что и женщины и мужчины стабильно занимали как призовые места так и не призовые. При это женщины чаще занимали первые и вторые места, а мужчины третье.

Далее построим круговые диаграммы по количеству 1-х мест по каждой Олимпиаде для женщин и мужчин.

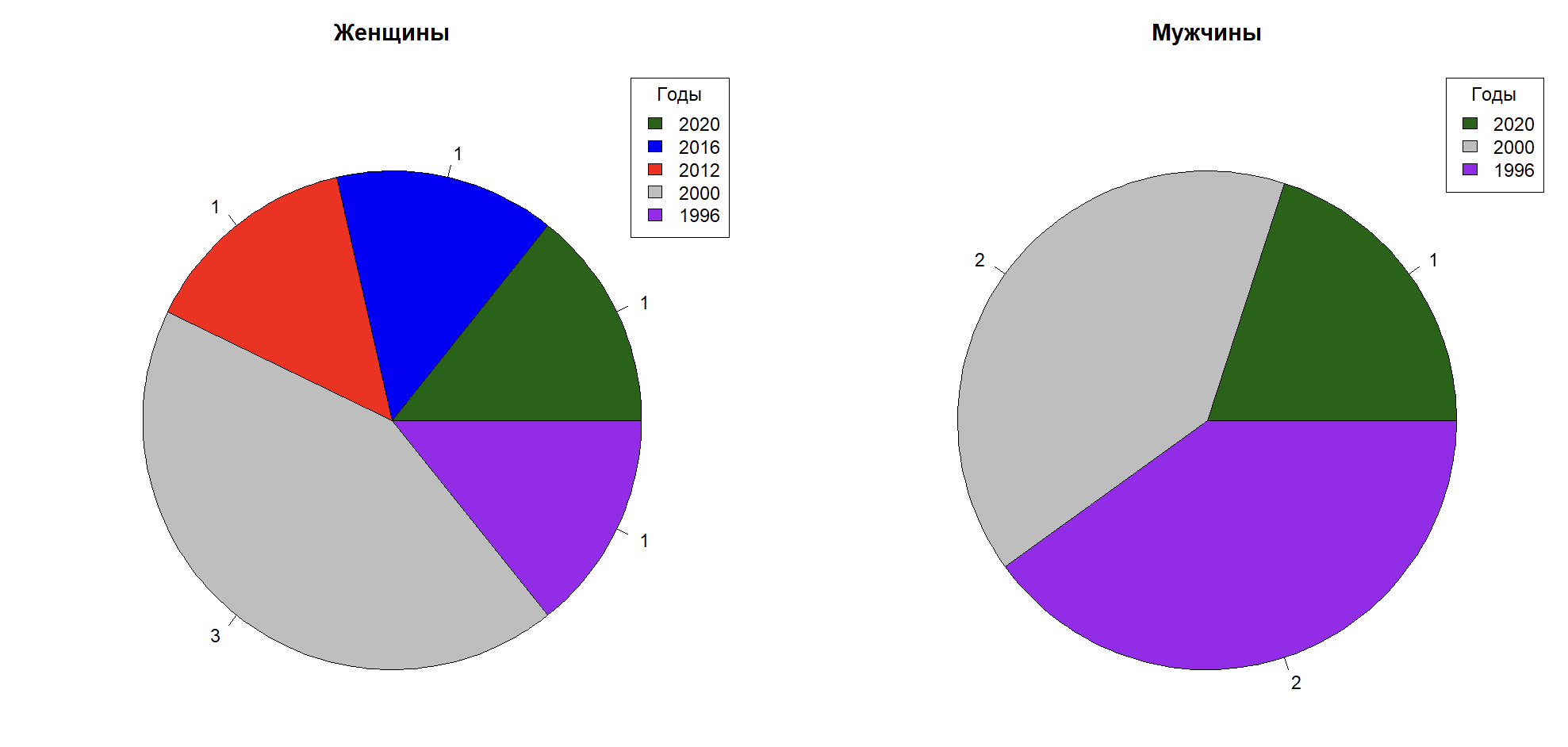


Рисунок 4 – Круговые диаграммы количества 1-х мест

Исходя из диаграмм, видно, что женщины чаще занимали первое место, лишь в двух олимпиадах они не занимали первое место, а у мужчин был период в 20 лет, когда они не занимали первое место.

Выведем также функциональные графики, которые отображают тенденции изменения количества призовых мест по категориям за последние 30 лет.

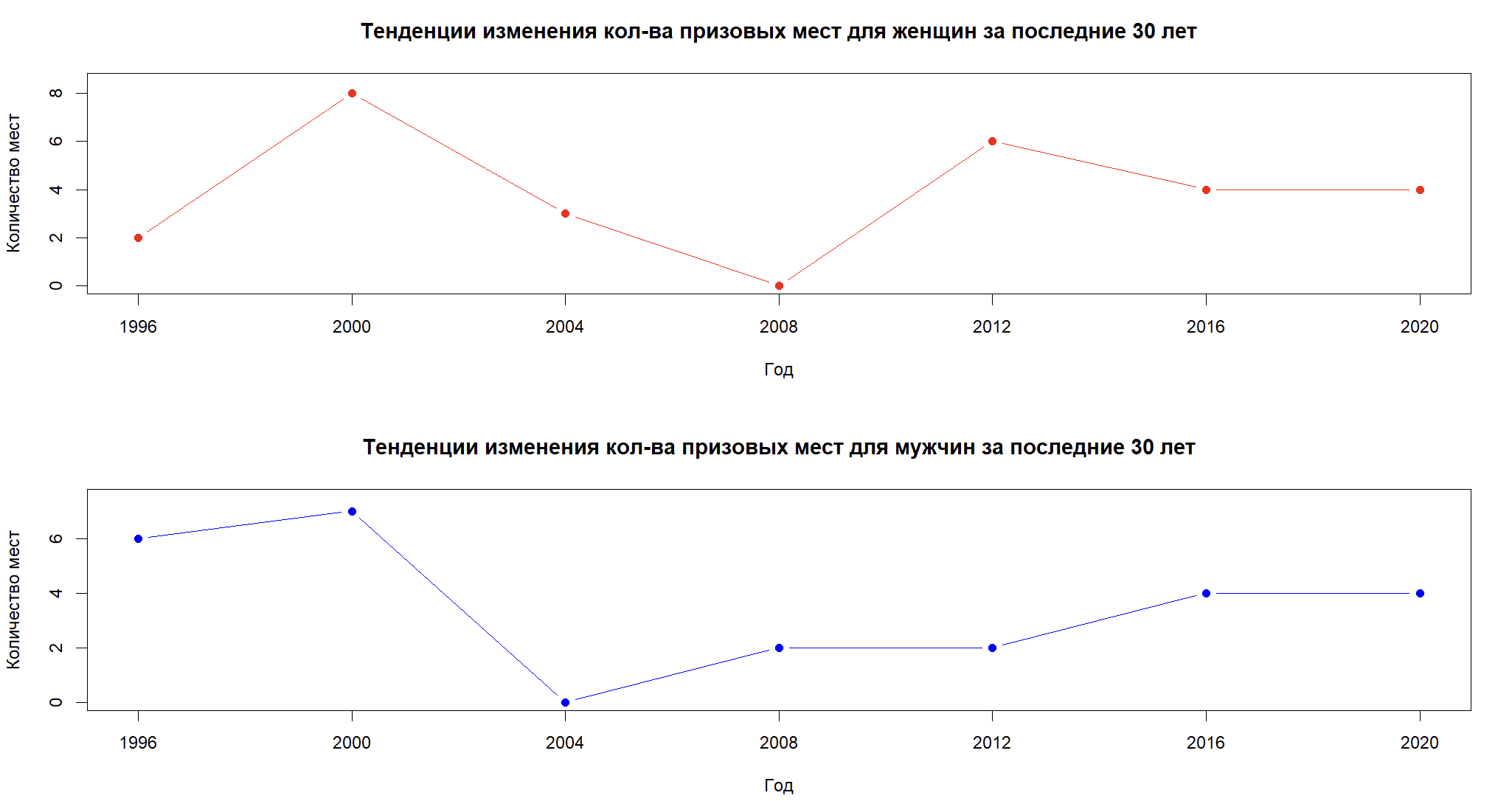


Рисунок 5 – Тенденции изменения количества призовых мест за последние 30 лет по категориям

По графикам видно, что и у женщин и у мужчин пик спортивных достижений на олимпиадах был в 2000 году, а дальше пошёл упадок по количеству призовых мест.

Создадим график изменения спортивных достижений по золотым медалям по 7-ми странам призерам за последние 6 Олимпиад, предварительно собрав данные.



Рисунок 6 – Датасет для выявления динамики количества 1-х мест за последние 6 олимпиад по 7 странам-призерам

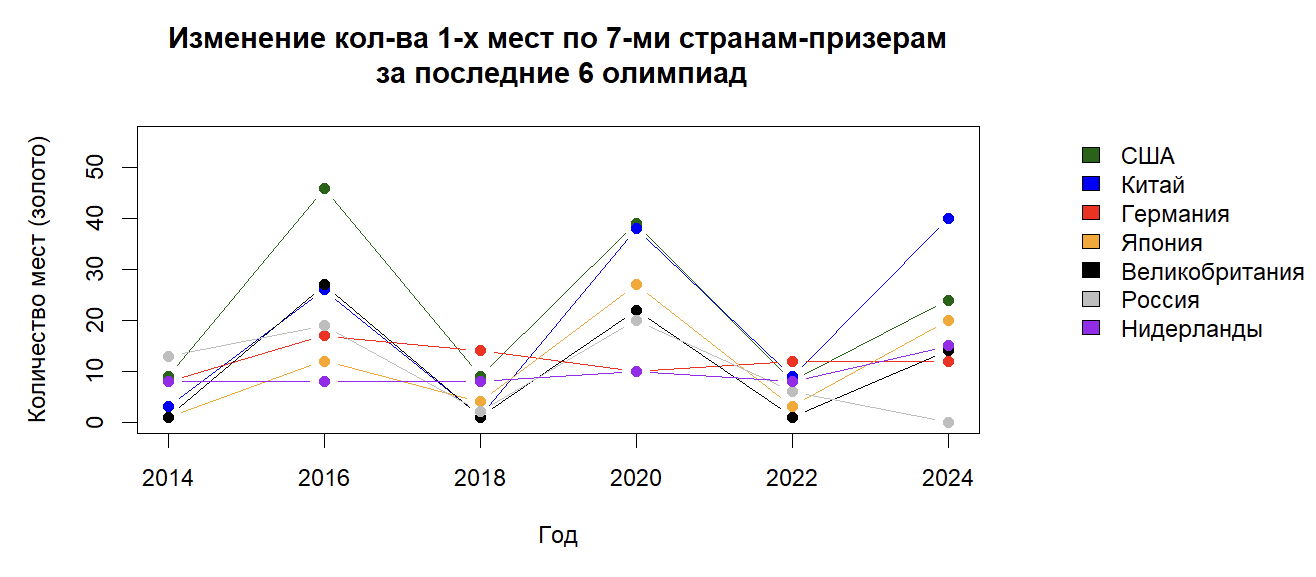


Рисунок 7 – Динамика количества 1-х мест за последние 6 олимпиад по 7 странам-призерам

В основном страны зарабатывали больше всего медалей во время летних Олимпийских игр, где в основном доминировали США и Китай. Россия занимали среднее количество медалей на всех олимпиадах, кроме олимпиады в 2014 и 2024, где её не было по политическим причинам.

Далее сделаем то же самое, но для всех призовых мест.



Рисунок 8 – Датасет для выявления динамики количества призовых мест за последние 6 олимпиад по 7 странам-призерам

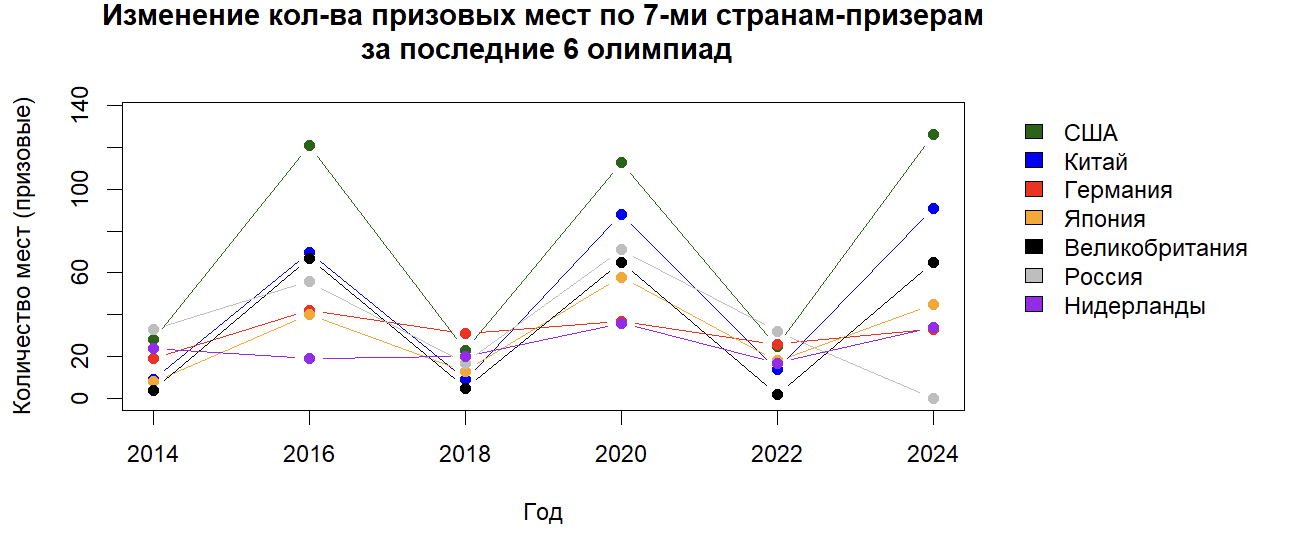


Рисунок 9 – Динамика количества призовых мест за последние 6 олимпиад по 7 странам-призерам

В целом, график не сильно изменился: в основном страны зарабатывали больше всего медалей на летних Олимпийских играх, где США сильно выделяется на фоне побед, хотя можно заметить, что Россия на этом графике находится выше. Это говорит о том, что Россия чаще занимала призовые места, но не первые.

Взяв информацию о призовых местах России по спортивной гимнастике за последние 6 Олимпиад, аккумулируем данные по женщинам и мужчинам и выведем динамику и статистику на графики.

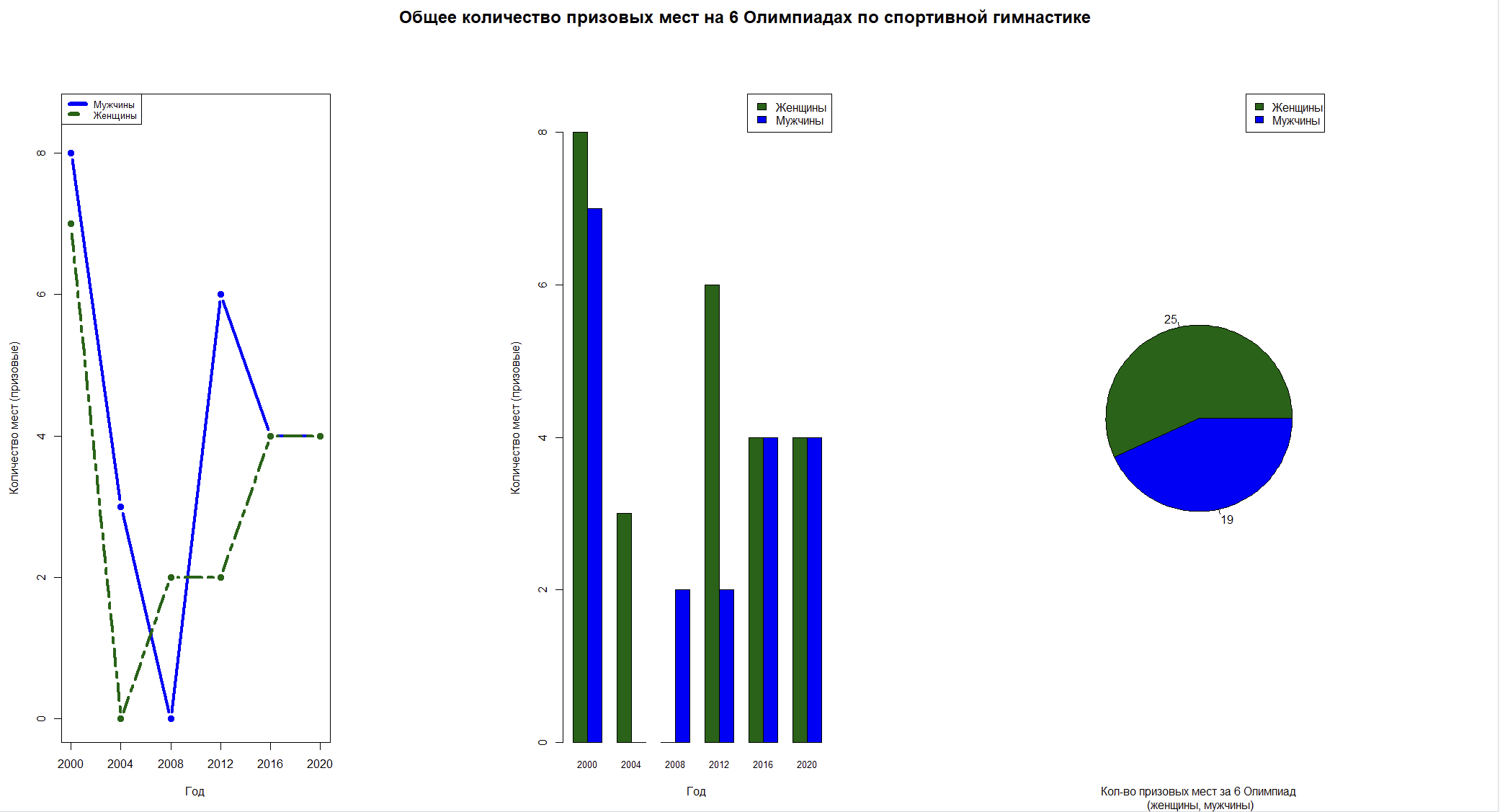


Рисунок 12 – Динамика и статистика по призовым местам за последние 6 Олимпиад по спортивной гимнастике

По данным графиков можно сказать, что Россия достаточно успешна в спортивной гимнастике, хотя были упадки на олимпиадах в 2004 и 2008 годах.

**Вывод:** проделав лабораторную работу, были изучены способы графического представления данных и реализованы на практике.

**Листинг программы:**

library(xlsx)

setwd("C:/university/3\_course/2\_semester/big\_data/labs/lab3")

df\_man <- read.xlsx("dataset.xlsx", 1, check.names = FALSE)

df\_woman <- read.xlsx("dataset.xlsx", 2, check.names = FALSE)

head(df\_woman)

# Столбчатые диаграммы

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

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

par(mfrow = c(1, 2))

barplot(df\_woman\_count,

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

col = "pink",

main = "Женщины",

xlab = "Места",

ylab = "Количество",

)

barplot(df\_man\_count,

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

col = "blue",

main = "Мужчины",

xlab = "Места",

ylab = "Количество")

# Круговая диаграмма

par(mfrow = c(1, 2))

years\_woman <- df\_woman$Год[df\_woman[, 2] != 0]

years\_man <- df\_man$Год[df\_man[, 2] != 0]

# Цвета для каждого года

legend\_colors <- c("darkgreen", "blue", "red", "orange", "black", "grey", "purple")

# Круговая диаграмма для женщин

pie(df\_woman[, 2][df\_woman[,2]!=0], labels = df\_woman[, 2][df\_woman[,2]!=0], main = "Женщины", col = legend\_colors[df\_woman[,2] != 0])

legend("topright",

legend = years\_woman, # Годы из данных

fill = legend\_colors[df\_woman[,2] != 0], # Цвета для каждого года

title = "Годы") # Заголовок легенды

# Круговая диаграмма для мужчин

pie(df\_man[, 2][df\_man[,2]!=0], labels = df\_man[, 2][df\_man[,2]!=0], main = "Мужчины", col = legend\_colors[df\_man[,2] != 0])

# Легенда с годами и цветами

legend("topright",

legend = years\_man, # Годы из данных

fill = legend\_colors[df\_man[,2] != 0], # Цвета для каждого года

title = "Годы") # Заголовок легенды

# Функциональные графики

plot\_graph <- function(df, color, title) {

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

matplot(df$Год, total\_places,main = title, type = "b", pch = 19, lty = 1, col = color,

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

ylim = c(0, max(rowSums(df[, 2:4])+0.5)))

axis(1, at = df$Год, labels = df$Год)

}

par(mfrow = c(2, 1))

plot\_graph(df\_woman, "red", "Тенденции изменения кол-ва призовых мест для женщин за последние 30 лет")

plot\_graph(df\_man, "blue", "Тенденции изменения кол-ва призовых мест для мужчин за последние 30 лет")

# 3 Графики изменения спортивных достижений (золото)

df\_gold <- read.xlsx("dataset.xlsx", 3)

par(mar = c(5, 5, 5, 12))

matplot(df\_gold$Год,

df\_gold[, -1],

main = "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(0.4, 0), fill = legend\_colors)

# 4 Графики изменения спортивных достижений (все призовые)

df\_prize\_places <- read.xlsx("dataset.xlsx", 4)

par(mar = c(5, 5, 5, 12))

matplot(df\_prize\_places$Год,

df\_prize\_places[, -1],

main = "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)

# Динамика cпортивной гимнастики по всем за последние 6 Олимпиад

total\_places1 <- rowSums(df\_woman[-7, 2:4], na.rm = TRUE)

total\_places2 <- rowSums(df\_man[-7, 2:4], na.rm = TRUE)

total\_places\_all <- cbind(total\_places1, total\_places2)

par(mfrow = c(1, 3), oma = c(0, 0, 2, 0))

matplot(df\_woman[-7,]$Год,

total\_places\_all,

type = "b", pch = 19,

lty = c(1, 6, 3),

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

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

xaxt = "n",

lwd = 3,

ylim = c(0, max(total\_places\_all)+0.5))

axis(1, at = df\_woman[-7,]$Год, labels = df\_woman[-7,]$Год)

legend('topleft', c('Мужчины', 'Женщины'), lty = c(1, 6, 3), cex = 0.8, lwd=4, col = c("blue", "darkgreen"))

total\_places\_barplot <- cbind(df\_woman[-7,1],total\_places1, total\_places2)

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

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

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

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

beside = TRUE,

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

xlab = "Год",

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

ylab = "Количество мест (призовые)",

ylim = c(0,max(total\_places\_barplot\_t\_sorted[2]+0.5,total\_places\_barplot\_t\_sorted[3]+0.5)),

cex.names = 0.8)

legend('topright', c('Женщины', 'Мужчины'), fill=c("darkgreen", "blue"))

total\_places\_pie <- rowSums(total\_places\_barplot\_t[-1,])

pie(total\_places\_pie,

labels=c(total\_places\_pie),

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

sub="Кол-во призовых мест за 6 Олимпиад\n (женщины, мужчины)")

legend('topright', c('Женщины', 'Мужчины'), fill=c("darkgreen", "blue"))

mtext("Общее количество призовых мест на 6 Олимпиадах по спортивной гимнастике",

outer = TRUE, cex = 1, font = 2)