МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КУБГУ»)**

**Факультет компьютерных технологий и прикладной математики**

**Кафедра вычислительных технологий**

**Отчет**

**по индивидуальному заданию №3**

**«Работа с графиками и диаграммами в R»**

**по курсу**

**«Обработка больших данных»**

Работу выполнил

Студент 39 группы

Пшеничнов А.А.

Преподаватель:

Яхонтов А.А.

Краснодар 2025

**Цель работы**

Научиться работать с графиками.

**Задания к лабораторной работе**

В ходе лабораторной работы необходимо поработать с оформлением графиков и диаграмм.

Необходимо учесть, что олимпийские игры по одному и тому же виду спорта бывают летние и зимние, соревнования делятся на мужские и женские.

1. Собрать данные согласно варианту по указанному виду и спорта в заданных странах.

2. Вывести графики динамики олимпийских достижений заданной страны по виду спорта относительно временной шкалы: источник (<http://olympteka.ru/olymp/country/profile/rus.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.

Обязательные требования к отчету: наличие исходных данных (фрагмент таблицы-скриншот из R и olympteka.ru), наличие подписей к осям графиков, заголовки к графику (main), масштабирование текста, наличие легенды, использовать объединение диаграмм и графиков (пункт 3.8), наличие трактовки каждого графика.

В качестве альтернативы можно скачать датасеты по летним и зимним олимпийским играм отсюда: https://www.kaggle.com/the-guardian/olympic-games и выполнить вышеназванные задания по ним, можно также использовать Википедию (по видам спорта, например, https://ru.wikipedia.org/wiki/Бадминтон\_на\_Олимпийских\_играх)

Также обратите внимание на то, что необходимо отделять зимние и летние игры друг от друга – они хронологически перемешаны, но проходят в разных условиях.



**Ход работы**

1. Собрали данные:

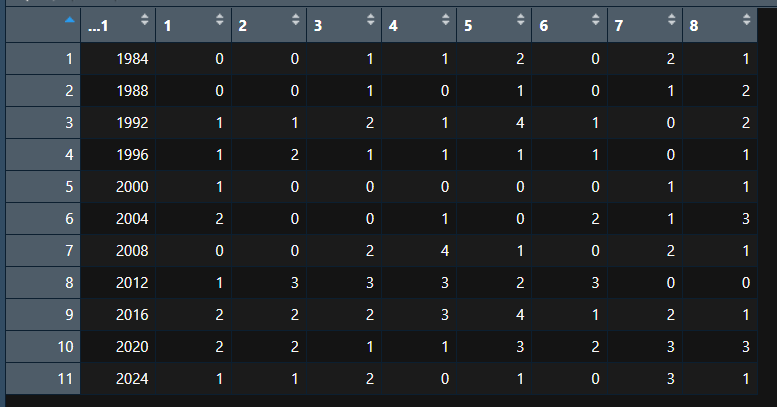


Рис.1 - Собранный датасет.

1. Выводим динамику олимпийских достижений Китая по легкой атлетике:

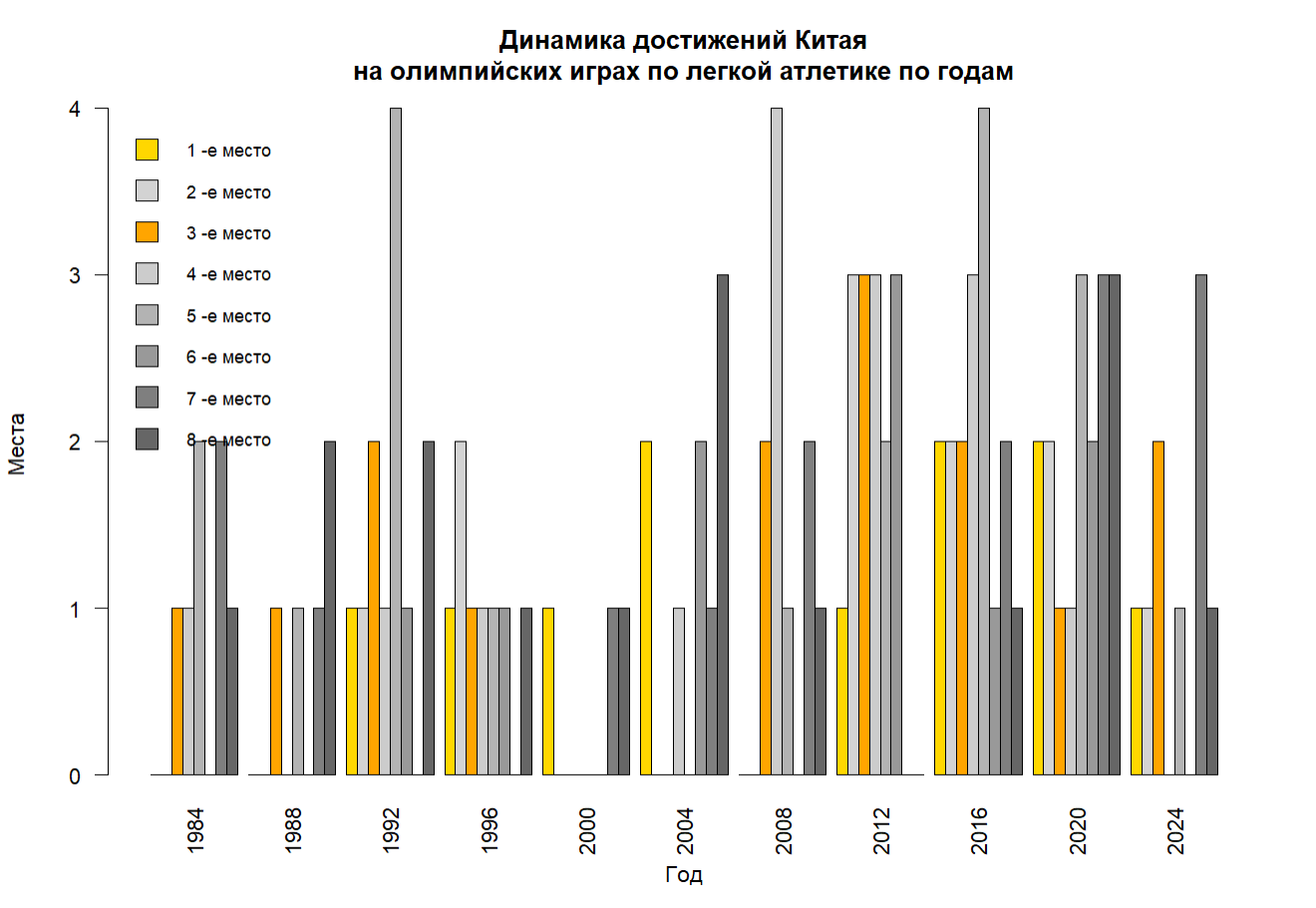


Рис. 2 - Динамика достижений Китая на олимпийских играх по легкой атлетике по годам

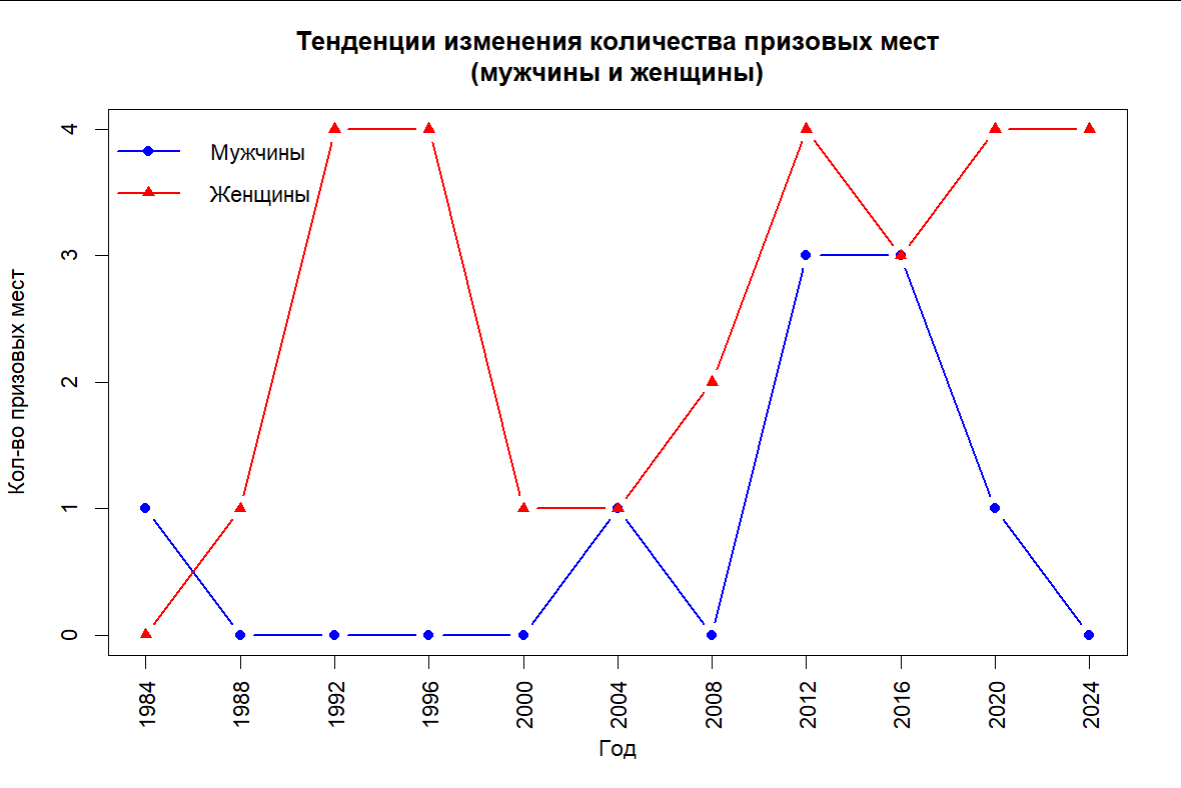


Рис. 3 – Тенденции изменения количества призовых мест (мужчины и женщины.

1. Выводим графики изменения спортивных достижений:

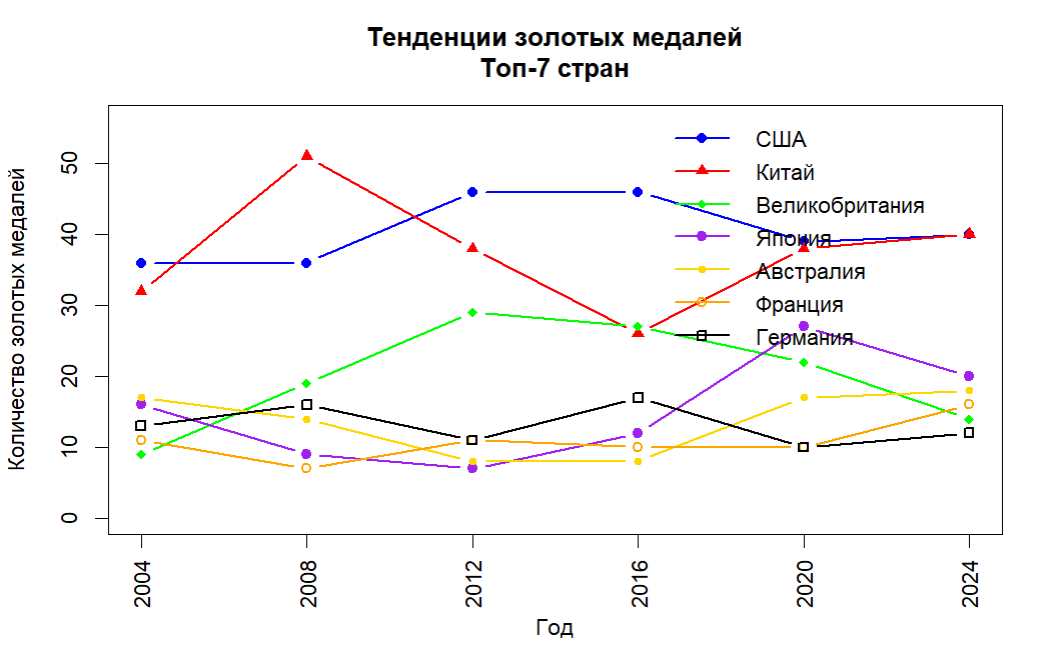


Рис. 4 – Тенденции золотых медалей Топ-7 стран

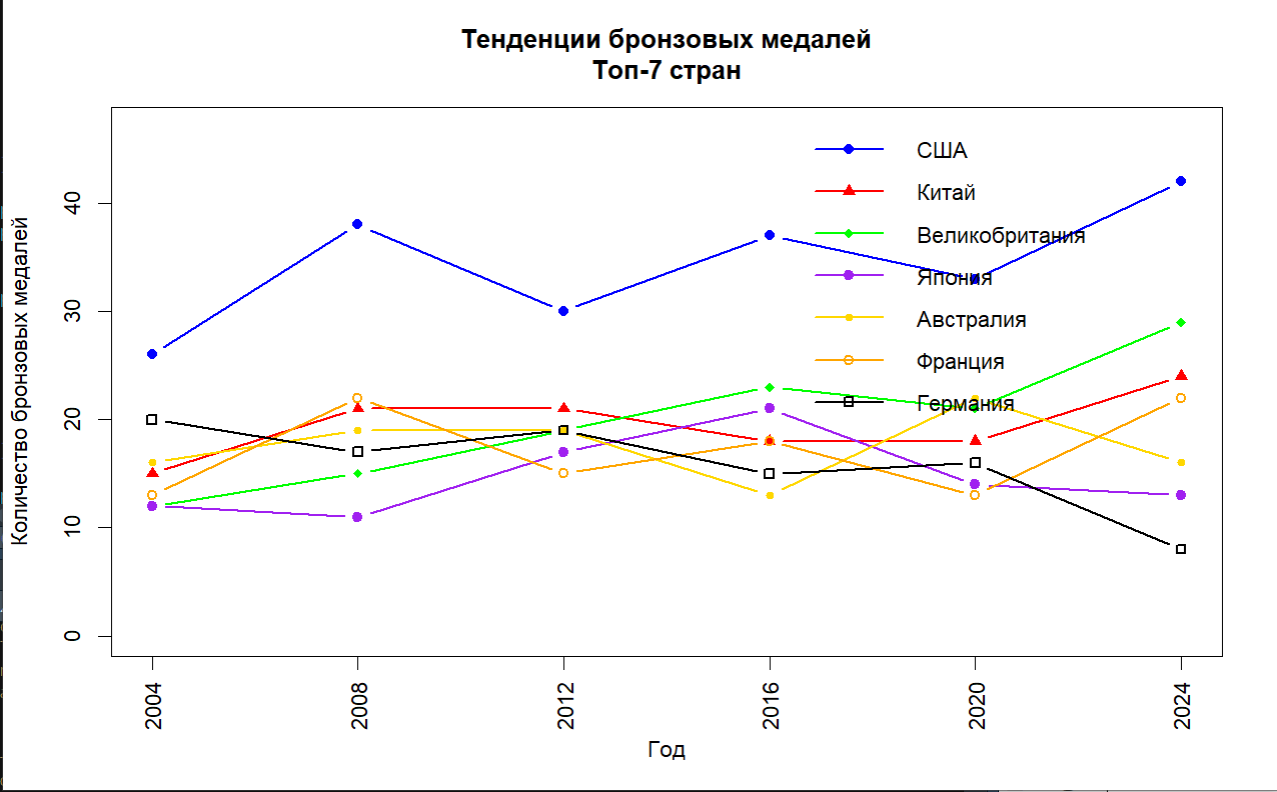


Рис. 5 – Тенденции бронзовых медалей Топ-7 стран.



Рис.6 – Распределение первых мест по годам.

1. Выводим динамику и распределение на одном графике:

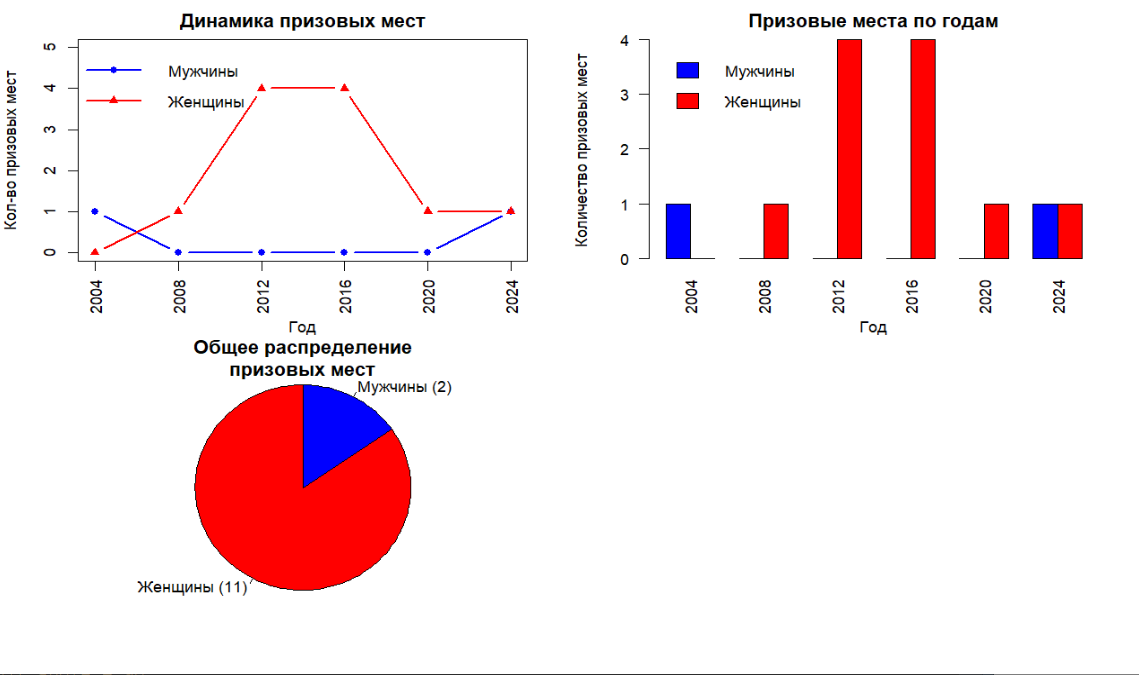


Рис. 7 – Динамика, распределение призовых мест.

**Вывод:** Научились работать с графиками и диаграммами.

**Полный код программы:**

setwd("C:/Users/Kurdicks/Desktop/r yazik/Excel")

library(readxl)

dataset <- read\_excel("datasetmen.xls", sheet = "DataSet")

datasetwoman <- read\_excel("datasetmen.xls", sheet = "WomanDS")

datasetman <- read\_excel("datasetmen.xls", sheet = "ManDS")

years <- seq(from = 1984, to = 2024, by = 4)

menMedals <- rowSums(datasetman[, 2:4])

womenMedals <- rowSums(datasetwoman[, 2:4])

if (!is.numeric(years) || !is.numeric(menMedals) || !is.numeric(womenMedals)) {

years <- as.numeric(years)

menMedals <- as.numeric(menMedals)

womenMedals <- as.numeric(womenMedals)

}

if (any(is.na(years)) || any(is.na(menMedals)) || any(is.na(womenMedals))) {

warning("Данные содержат NA. Удаляем строки с NA.")

valid <- !is.na(years) & !is.na(menMedals) & !is.na(womenMedals)

years <- years[valid]

menMedals <- menMedals[valid]

womenMedals <- womenMedals[valid]

}

if (length(years) == 0 || length(menMedals) == 0 || length(womenMedals) == 0) {

stop("Ошибка: данные пусты после очистки")

}

plot(years, menMedals, type = "n", main = "Тенденции изменения количества призовых мест\n(мужчины и женщины)",

xlab = "Год", ylab = "Кол-во призовых мест", xaxt = "n",

ylim = c(0, max(c(menMedals, womenMedals), na.rm = TRUE)))

axis(1, at = years, labels = years, las = 2)

lines(years, menMedals, type = "b", col = "blue", pch = 16, lwd = 2)

lines(years, womenMedals, type = "b", col = "red", pch = 17, lwd = 2)

legend("topleft", legend = c("Мужчины", "Женщины"), col = c("blue", "red"),

pch = c(16, 17), lwd = 2, bty = "n")

if (nrow(dataset) != length(years)) {

warning(paste("Количество строк в dataset (", nrow(dataset), ") не соответствует количеству лет (", length(years), "). Обрезаем года."))

years <- years[1:nrow(dataset)]

}

if (ncol(dataset) > 1) {

category\_data <- as.matrix(dataset[, -1])

} else {

category\_data <- as.matrix(dataset)

}

if (ncol(category\_data) > 8) {

category\_data <- category\_data[, 1:8]

} else if (ncol(category\_data) < 1) {

stop("Нет данных по категориям")

}

category\_data <- t(category\_data)

colors <- c("gold", "lightgrey", "orange", "grey80", "grey70", "grey60", "grey50", "grey40")[1:nrow(category\_data)]

barplot(category\_data,

beside = TRUE,

names.arg = years,

col = colors,

main = "Динамика достижений Китая\nна олимпийских играх по легкой атлетике по годам",

xlab = "Год",

ylab = "Места",

las = 2,

legend.text = paste(1:nrow(category\_data), "-е место"),

args.legend = list(x = "topleft", bty = "n", cex = 0.8))

GOLD <- dataset[, 2]

years <- dataset[, 1]

positive\_idx <- GOLD > 0

GOLD\_positive <- GOLD[positive\_idx]

years\_positive <- years[positive\_idx]

colors <- rainbow(length(GOLD\_positive))

pie(GOLD\_positive,

labels = GOLD\_positive,

main = "Распределение первых мест по годам",

col = colors)

legend("topright",

legend = paste(years\_positive, "(", GOLD\_positive, ")", sep = ""),

fill = colors)

last6 <- c(2004, 2008, 2012, 2016, 2020, 2024)

usa\_gold <- c(36, 36, 46, 46, 39, 40)

chn\_gold <- c(32, 51, 38, 26, 38, 40)

gbr\_gold <- c(9, 19, 29, 27, 22, 14)

jpn\_gold <- c(16, 9, 7, 12, 27, 20)

aus\_gold <- c(17, 14, 8, 8, 17, 18)

fra\_gold <- c(11, 7, 11, 10, 10, 16)

ger\_gold <- c(13, 16, 11, 17, 10, 12)

usa\_bronze <- c(26, 38, 30, 37, 33, 42)

chn\_bronze <- c(15, 21, 21, 18, 18, 24)

gbr\_bronze <- c(12, 15, 19, 23, 21, 29)

jpn\_bronze <- c(12, 11, 17, 21, 14, 13)

aus\_bronze <- c(16, 19, 19, 13, 22, 16)

fra\_bronze <- c(13, 22, 15, 18, 13, 22)

ger\_bronze <- c(20, 17, 19, 15, 16, 8)

plot(last6, usa\_gold,

type = "n",

ylim = c(0, max(usa\_gold, chn\_gold, gbr\_gold, jpn\_gold, aus\_gold, fra\_gold, ger\_gold) + 5),

main = "Тенденции золотых медалей\nТоп-7 стран",

xlab = "Год",

ylab = "Количество золотых медалей",

xaxt = "n"

)

axis(1, at = last6, labels = last6, las = 2)

lines(last6, usa\_gold, type = "b", col = "blue", pch = 16, lwd = 2) # США

lines(last6, chn\_gold, type = "b", col = "red", pch = 17, lwd = 2) # Китай

lines(last6, gbr\_gold, type = "b", col = "green", pch = 18, lwd = 2) # Великобритания

lines(last6, jpn\_gold, type = "b", col = "purple", pch = 19, lwd = 2) # Япония

lines(last6, aus\_gold, type = "b", col = "gold", pch = 20, lwd = 2) # Австралия

lines(last6, fra\_gold, type = "b", col = "orange", pch = 21, lwd = 2) # Франция

lines(last6, ger\_gold, type = "b", col = "black", pch = 22, lwd = 2) # Германия

legend("topright",

legend = c("США", "Китай", "Великобритания", "Япония", "Австралия", "Франция", "Германия"),

col = c("blue", "red", "green", "purple", "gold", "orange", "black"),

pch = c(16, 17, 18, 19, 20, 21, 22),

lwd = 2,

bty = "n")

plot(last6, usa\_bronze,

type = "n",

ylim = c(0, max(usa\_bronze, chn\_bronze, gbr\_bronze, jpn\_bronze, aus\_bronze, fra\_bronze, ger\_bronze) + 5),

main = "Тенденции бронзовых медалей\nТоп-7 стран",

xlab = "Год",

ylab = "Количество бронзовых медалей",

xaxt = "n"

)

axis(1, at = last6, labels = last6, las = 2)

lines(last6, usa\_bronze, type = "b", col = "blue", pch = 16, lwd = 2) # США

lines(last6, chn\_bronze, type = "b", col = "red", pch = 17, lwd = 2) # Китай

lines(last6, gbr\_bronze, type = "b", col = "green", pch = 18, lwd = 2) # Великобритания

lines(last6, jpn\_bronze, type = "b", col = "purple", pch = 19, lwd = 2) # Япония

lines(last6, aus\_bronze, type = "b", col = "gold", pch = 20, lwd = 2) # Австралия

lines(last6, fra\_bronze, type = "b", col = "orange", pch = 21, lwd = 2) # Франция

lines(last6, ger\_bronze, type = "b", col = "black", pch = 22, lwd = 2) # Германия

legend("topright",

legend = c("США", "Китай", "Великобритания", "Япония", "Австралия", "Франция", "Германия"),

col = c("blue", "red", "green", "purple", "gold", "orange", "black"),

pch = c(16, 17, 18, 19, 20, 21, 22),

lwd = 2,

bty = "n")

menMedals6 = menMedals[1:6]

womenMedals6 = womenMedals[1:6]

menTotal = sum(menMedals6)

womenTotal = sum(womenMedals6)

par(mfrow = c(2,2), mar = c(4,4,2,2) + 0.1)

plot(last6, menMedals6,

type="n", ylim = c(0, max(menMedals6, womenMedals6) + 1),

main = "Динамика призовых мест",

xlab = "Год",

ylab = "Кол-во призовых мест",

xaxt = "n")

axis(1, at = last6, lables = last6, las = 2)

lines(last6, menMedals6, type = "b", col = "blue", pch = 16, lwd = 2)

lines(last6, womenMedals6, type = "b", col = "red", pch = 17, lwd = 2)

legend("topleft",

legend = c("Мужчины", "Женщины"),

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

pch = c(16, 17),

lwd = 2,

bty = "n")

bar\_data <- matrix(c(menMedals6, womenMedals6), nrow = length(last6), byrow = FALSE)

write.csv(bar\_data, file = "bar\_data.csv", row.names = TRUE)

colnames(bar\_data) <- c("Мужчины", "Женщины")

rownames(bar\_data) <- last6

barplot(t(bar\_data),

beside = TRUE,

names.arg = last6,

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

main = "Призовые места по годам",

xlab = "Год",

ylab = "Количество призовых мест",

las = 2,

legend = colnames(bar\_data),

args.legend = list(x = "topleft", bty = "n")

)

pie\_data <- c(menTotal, womenTotal)

pie\_labels <- c(paste("Мужчины (", menTotal, ")", sep = ""),

paste("Женщины (", womenTotal, ")", sep = ""))

pie(pie\_data,

labels = pie\_labels,

main = "Общее распределение\nпризовых мест",

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

clockwise = TRUE,

init.angle = 90,

radius = 1

)

par(mfrow = c(1, 1))