Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования «Кубанский государственный университет»

Кафедра вычислительных технологий

**ОТЧЕТ**

о выполнении лабораторной работы № 3

по дисциплине «Обработка больших данных»

Тема: Работа с графиками и диаграммами в R

Выполнил: ст. гр. 36/2

Агаджанян А. С.

Проверил: преп. кафедры ВТ

Яхонтов А. А.

Краснодар

2025

Вариант: 17

Страна Канада

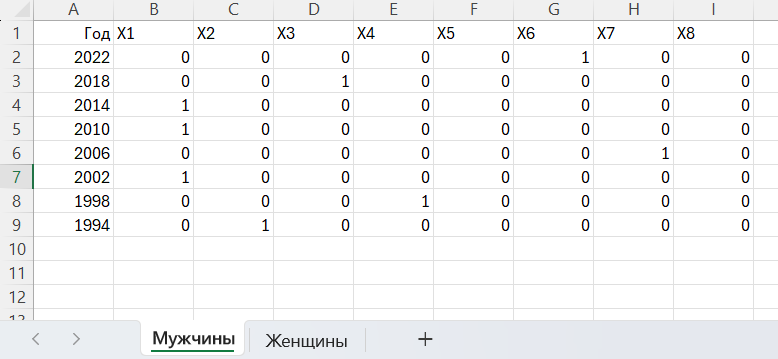
Вид спорта: хоккей на льду

Хоккей на льду проводится только в зимний период.

По ссылке <https://olympteka.ru/olymp/sport/country_medalists/28/can.html> находим статистику достижений Канады в хоккее на льду с 1994 по 2022 год.



Занесём данные в таблицу Excel, разделив их на мужчин и женщин:

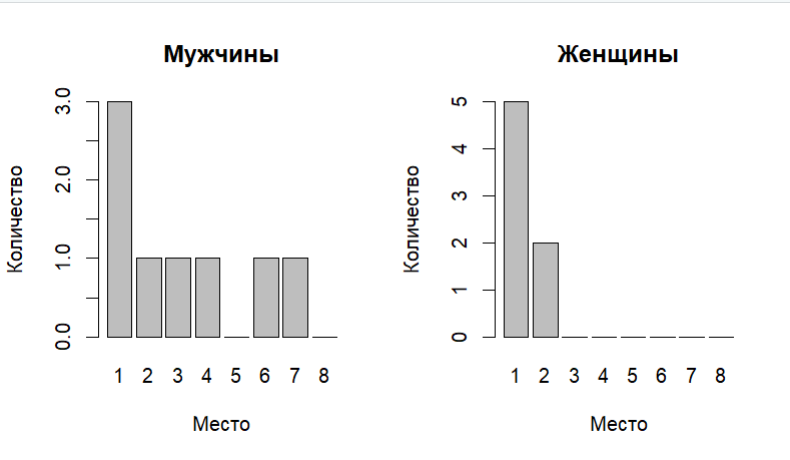


Изображение выглядит как текст, число, снимок экрана, Шрифт

Контент, сгенерированный ИИ, может содержать ошибки.

Женская сборная в 1994 году не участвовала в зимних ОИ. Информация о них в таблице отсутствует.

Построим столбчатые диаграммы по всем местам, просуммировав количество 1–8 мест за все года. Отобразим диаграмму для мужчин слева, а для женщин справа.



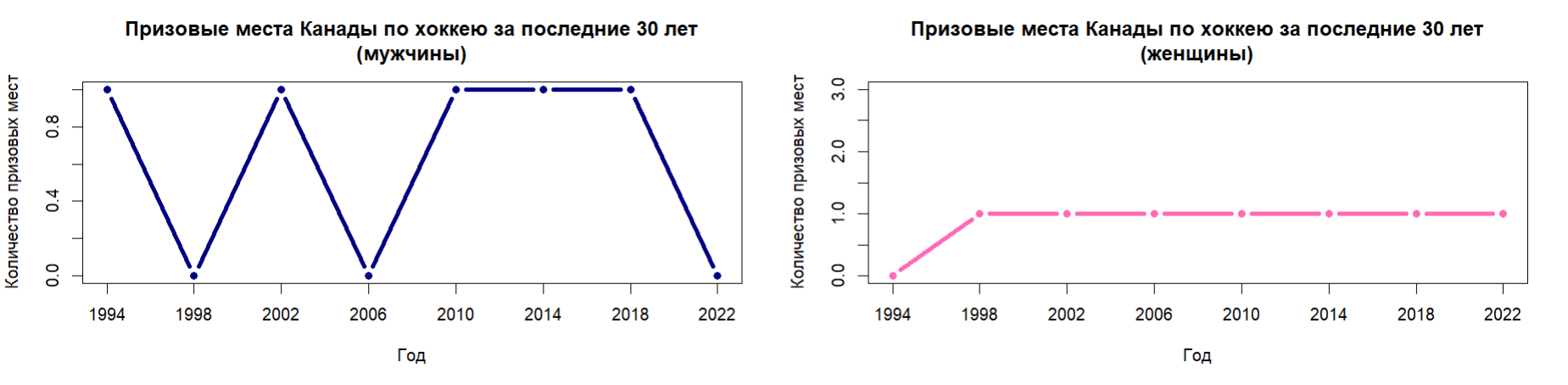
Исходя из полученных данных женщины в Канаде в хоккее на льду занимают больше призовых мест, чем мужчины. Однако мужчины в сумме занимают больше мест, чем женщины.

Построим круговые диаграммы по количеству золотых медалей в олимпиадах.

Изображение выглядит как диаграмма, текст, снимок экрана, дизайн

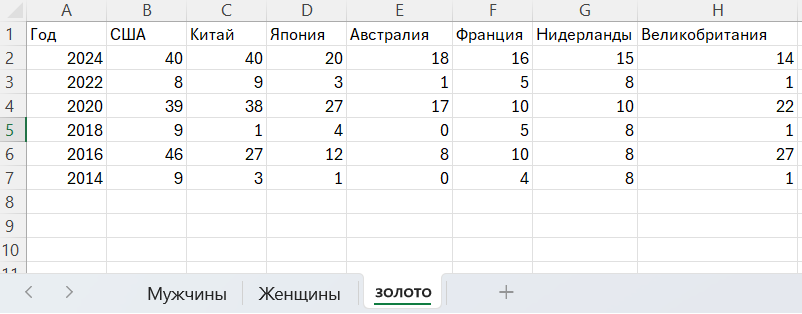
Контент, сгенерированный ИИ, может содержать ошибки.

По этим диаграммам видно, что за всё время женщины получили больше золотых медалей, чем мужчины, хоть далеко не в каждой олимпиаде им удавалось занять первые места. Последняя на текущий момент золотая медаль завоёвана в 2022 году женщинами.



По этим графикам можно сказать, что женщины с 1998 года стабильно занимали призовые места.

Выведем график изменения спортивных достижений по золотым медалям по 7 странам-призёрам за последние 6 олимпиад. Для этого занесём данные в таблицу:

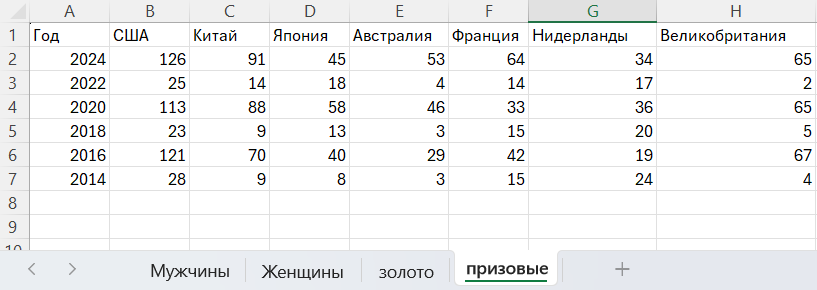


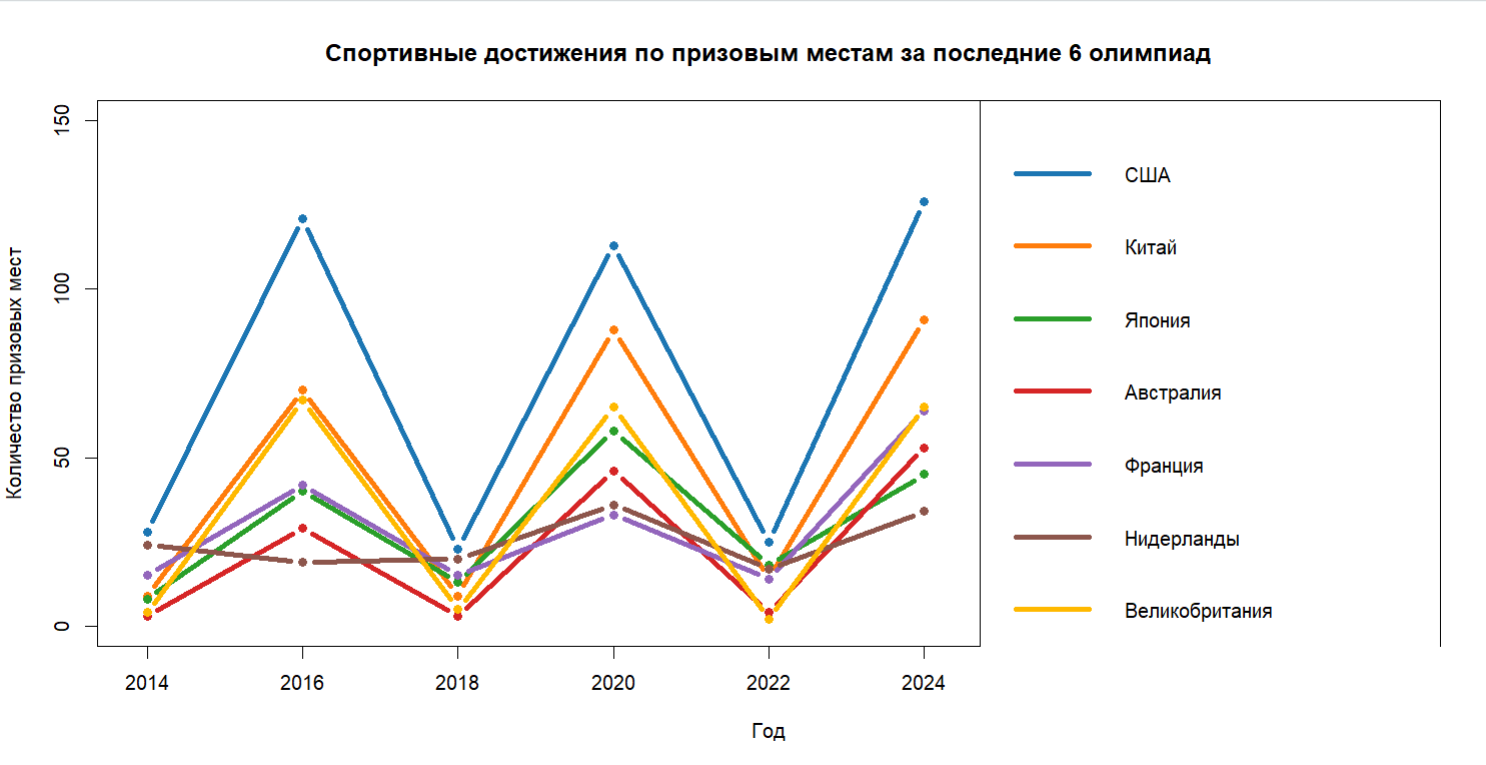
Изображение выглядит как текст, диаграмма, линия, График

Контент, сгенерированный ИИ, может содержать ошибки.

Заметим, что многие страны одерживали больше всего побед во время летних ОИ, кроме Нидерландов (у него более ровная линия). Также заметим, что США одерживали больше всего побед.

Теперь выведем график изменения спортивных достижений по призовым местам по тем же странам за последние 6 олимпиад. Занесём данные в таблицу:





На данном графике тенденция сохраняется.

Используя тот же источник, выберем информацию об успехах канадских спортсменов за последние 6 олимпиад по хоккею на льду, аккумулируем данные по мужчинам и женщинам (отобразим на одном графике разными линиями), выведем динамику и статистику на графики (функциональный, столбчатый, круговой).

Изображение выглядит как текст, диаграмма, График, линия

Контент, сгенерированный ИИ, может содержать ошибки.

По данным графикам можно сказать, что за последние 6 олимпийских игр женщины из Канады всё же больше преуспели в хоккее на льду, чем мужчины. А именно, у женщин на 2 победы больше.

Вывод: в ходе выполнения лабораторной работы были подробно изучены способы графического представления данных. Проделан анализ и визуализация данных об олимпийских играх, построены графики и диаграммы.

Листинг:

setwd('D:/Универ/3 курс/6 семестр/Бигдата/lab3')

library(readxl)

# 3.2

list\_men <- data.frame(read\_excel('статистика.xlsx', sheet='Мужчины'))

list\_women <- data.frame(read\_excel('статистика.xlsx', sheet='Женщины'))

rownames(list\_men) <- list\_men[,1] #для даты в круговой диаграмме

rownames(list\_women) <- list\_women[,1] #для даты в круговой диаграмме

list\_men <- list\_men[,-1]

list\_women <- list\_women[,-1]

sum\_column\_men = sapply(list\_men, sum)

sum\_column\_women = sapply(list\_women, sum, na.rm=TRUE)

barplot(sum\_column\_men, names=c(1:8), xlab="Место", ylab="Количество", main="Мужчины")

barplot(sum\_column\_women, names=c(1:8), xlab="Место", ylab="Количество", main="Женщины")

gold\_men = list\_men[list\_men[,1] > 0, 1]

gold\_women = list\_women[list\_women[,1] > 0 & !is.na(list\_women[,1]), 1]

names(gold\_men) <- rownames(list\_men[list\_men[,1] > 0,])

names(gold\_women) <- rownames(list\_women[list\_women[,1] > 0 & !is.na(list\_women[,1]),])

legend\_colors <- c("#F0F8FF", "#E6E6FA", "#FAF0E6", "#FFF0F5", "#FFFFF0", "#FFFAFA", "#F8F8FF", "#F5F5DC", "#FFF5EE")

pie(gold\_men, labels=gold\_men, main = "Кол-во золотых медалей по годам (мужчины)", col = legend\_colors)

legend("topleft", legend=names(gold\_men), cex=0.5, fill = legend\_colors)

pie(gold\_women, labels=gold\_women, main = "Кол-во золотых медалей по годам (женщины)", col = legend\_colors)

legend("topleft", legend=names(gold\_women), cex=0.5, fill = legend\_colors)

prize\_places\_trend\_men <- data.frame("Количество призовых мест у мужчин"=rowSums(list\_men[rownames(list\_men) >= 1994, 1:3]))

prize\_places\_trend\_women <- data.frame("Количество призовых мест у женщин"=rowSums(list\_women[rownames(list\_men) >= 1994, 1:3]))

prize\_places\_trend\_women[is.na(prize\_places\_trend\_women)] <- 0

prize\_places\_trend <- cbind(prize\_places\_trend\_men, prize\_places\_trend\_women)

color\_men <- "lightblue"

color\_women <- "pink"

matplot(rownames(prize\_places\_trend\_men), prize\_places\_trend\_men, xaxt="n", type="b", pch=16, lwd=4,

main='Призовые места Канады по хоккею за последние 30 лет\n(мужчины)',

xlab='Год', ylab='Количество призовых мест', col=color\_men)

axis(side=1, at=rownames(prize\_places\_trend\_men))

matplot(rownames(prize\_places\_trend\_women), prize\_places\_trend\_women, xaxt="n", type="b", pch=16, lwd=4,

main='Призовые места Канады по хоккею за последние 30 лет\n(женщины)',

xlab='Год', ylab='Количество призовых мест', col=color\_women, ylim=c(0, 3))

axis(side=1, at=rownames(prize\_places\_trend\_women))

# 3.3

colors2 <- c("#1c76b3", "#ff7d0c", "#2a9f2a", "#d62526", "#9366bc", "#8d564c", "#ffb900")

par(mfrow=c(1,1))

gold\_medals <- data.frame(read\_excel('статистика.xlsx', sheet='золото'))

rownames(gold\_medals) <- gold\_medals[,1]

gold\_medals <- gold\_medals[,-1]

gold\_medals\_summer <- gold\_medals[c(1,3,5),]

gold\_medals\_winter <- gold\_medals[c(2,4,6),]

matplot(rownames(gold\_medals), gold\_medals, xaxt="n", type="b", pch=16, lty=1, lwd=4,

xlim=c(2014, 2029), ylim=c(0, 60),

main='Спортивные достижения по золотым медалям за последние 6 олимпиад',

xlab='Год', ylab='Количество золотых медалей', col=colors2)

axis(side=1, at=rownames(gold\_medals))

legend('topright', colnames(gold\_medals), lwd=4, col=colors2)

prize\_places <- data.frame(read\_excel('статистика.xlsx', sheet='призовые'))

rownames(prize\_places) <- prize\_places[,1]

prize\_places <- prize\_places[,-1]

prize\_places\_summer <- prize\_places[c(1,3,5),]

prize\_places\_winter <- prize\_places[c(2,4,6),]

matplot(rownames(prize\_places), prize\_places, xaxt="n", type="b", pch=16, lty=1, lwd=4,

xlim=c(2014, 2030), ylim=c(0, 150),

main='Спортивные достижения по призовым местам за последние 6 олимпиад',

xlab='Год', ylab='Количество призовых мест', col=colors2)

axis(side=1, at=rownames(gold\_medals))

legend('topright', colnames(gold\_medals), lwd=4, col=colors2)

# 3.4

prize\_places\_last\_6 <- head(prize\_places\_trend, 6)

par(mfrow=c(1,3))

matplot(rownames(prize\_places\_last\_6), prize\_places\_last\_6, xaxt="n", type="b", pch=16, lty=c(2,3), lwd=4,

ylim=c(0, 3.5), xlab='Год', ylab='Количество призовых мест', col=c(color\_men, color\_women))

axis(side=1, at=rownames(prize\_places\_last\_6))

legend('topleft', c('Мужчины', 'Женщины'), lty=c(2,3), lwd=4, col=c(color\_men, color\_women))

barplot(t(as.matrix(prize\_places\_last\_6)), beside=TRUE, xlab="Год", ylab="Количество призовых мест",

ylim=c(0, 4), col=c(color\_men, color\_women),

main="Призовые места Канады по хоккею на льду за последние 6 игр")

legend('topleft', c('Мужчины', 'Женщины'), fill=c(color\_men, color\_women))

prize\_places\_last\_6\_sum <- sapply(prize\_places\_last\_6, sum)

pie(prize\_places\_last\_6\_sum, labels=c(prize\_places\_last\_6\_sum[1], prize\_places\_last\_6\_sum[2]),

col=c(color\_men, color\_women), sub="Сумма за 6 игр (мужчины и женщины)")

legend('topleft', c('Мужчины', 'Женщины'), fill=c(color\_men, color\_women))