Домашнее задание

В файле vgsales.csv сохранены данные по продажам видео-игр, собранные с сайта [vgchartz.com](http://www.vgchartz.com/).

## Показатели в файле

* Rank: ранг продажи;
* Name: название игры;
* Platform: платформа, для которой предназначена игра (PC, PS4, и. т. д.);
* Year: год выпуска игры;
* Genre - жанр игры;
* Publisher - компания, выпустившая игру;
* NA\_Sales - число продаж в Северной Америке (в миллионах);
* EU\_Sales - число продаж в Европе (в миллионах);
* JP\_Sales - число продаж в Японии (в миллионах);
* Other\_Sales - число продаж в остальных частях света/странах (в миллионах);
* Global\_Sales - общее число продаж.

Подробнее про данные можно почитать на [странице](https://www.kaggle.com/gregorut/videogamesales) файла на Kaggle.

### Задача 1

Загрузите данные из файла vgsales.csv и сохраните их в датафрейм games.

*Ответ:*

setwd("C:/Users/alexandra\_c/Desktop/Уроки R/Практические задания/10 Визуализация данных в R/10-07 Практическая работа")  
games <- read.csv("vgsales.csv")

### Задача 2

Оставьте в датафрейме games только строки, которые соответствуют видео-играм с общим числом продаж не более 5 миллионов.

*Ответ:*

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --  
## v ggplot2 3.3.6 v purrr 0.3.4   
## v tibble 3.1.8 v dplyr 1.0.10  
## v tidyr 1.2.1 v stringr 1.4.1   
## v readr 2.1.2 v forcats 0.5.2   
## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

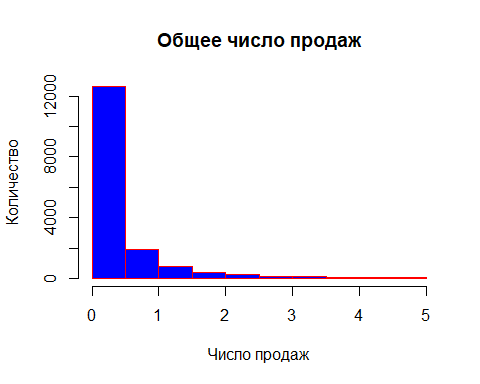
games <- games %>% filter(Global\_Sales <= 5)

### Задача 3

Постройте гистограмму для показателя «общее число продаж», используя обновлённый датафрейм games. Измените цвет заливки гистограммы и цвет границ столбцов. Добавьте содержательные подписи на русском языке по осям x и y. Добавьте заголовок графика.

*Ответ:*

hist(games$Global\_Sales,  
 col = "blue",  
 border = "red",  
 xlab = "Число продаж",  
 ylab = "Количество",  
 main = "Общее число продаж"  
 )



### Задача 4

Проинтерпретируйте полученную в задаче 3 гистограмму. В интерпретации укажите форму распределения (является ли оно симметричным или скошенным, если скошенным, то в какую сторону) и напишите, какие значения встречаются чаще, а какие реже. *Ответ:*

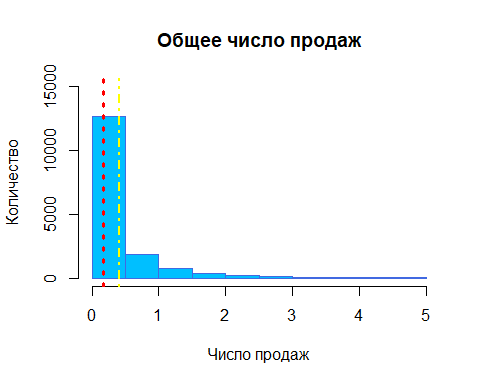
# Форма распредления числа продаж скошенная в право.  
# Чаще всего встречаются значения с маленьким кол-ом продаж, реже с высоким кол-вом продаж

### Задача 5

Добавьте на гистограмму из задачи 3 вертикальные линии, отчерчивающие среднее и медианное значение показателя соответственно. Линии должны быть разного цвета и начертания (сплошная линия, пунктирная, точечная, и так далее). Что можно сказать о различии между средним значением и медианным значением?

*Ответ:*

hist(games$Global\_Sales,  
 col = "deepskyblue",  
 border = "royalblue",  
 xlab = "Число продаж",  
 ylab = "Количество",  
 main = "Общее число продаж",  
 ylim = c(0, 15000)  
 )  
abline(v = mean(games$Global\_Sales), col = "yellow", lty = 4, lwd = 2)  
abline(v = median(games$Global\_Sales), col = "red", lty = 3,lwd = 3)



# Среднее значение выше медианного.

### Задача 6

Приведите ответ на следующий вопрос. В каких случаях использование различных начертаний линий может быть более рациональным, чем использование различных цветов?

*Ответ:*

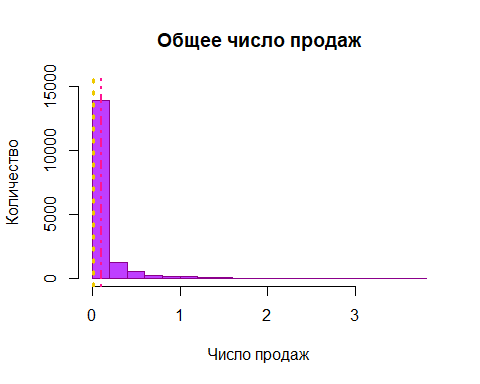
# Использование различных начертаний линий может быть более рациональным, чем использование различных цветов, в том случае если график будет использоваться в черно-белом виде.

### Задача 7

Постройте гистограммму для показателя «число продаж в Европе». Можно ли сказать, что распределение показателей «число продаж в Европе» и «общее число продаж» схожи? Приведите ответ на вопрос и обоснуйте свой ответ.

*Ответ:*

hist(games$EU\_Sales,  
 col = "darkorchid1",  
 border = "darkmagenta",  
 xlab = "Число продаж",  
 ylab = "Количество",  
 main = "Общее число продаж",  
 ylim = c(0, 15000)  
 )  
abline(v = mean(games$EU\_Sales), col = "deeppink", lty = 4, lwd = 2)  
abline(v = median(games$EU\_Sales), col = "gold2", lty = 3,lwd = 3)



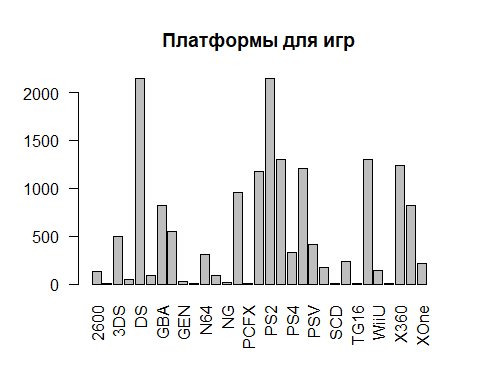
# Форма распредления числа продаж в Европе скошенная в право, как у общего числа продаж.  
# Среднее значение больше медианного, как и у общего числа продаж.  
# Также как и у общего числа продаж, чаще всего встречаются значения с маленьким кол-ом продаж, реже с высоким кол-вом продаж

### Задача 8

Постройте столбчатую диаграмму для показателя «платформа, для которой предназначена игра». При необходимости выполните поворот подписей к столбцам. Какая платформа является наиболее популярной? Приведите ответ на вопрос.

*Ответ:*

barplot(table(games$Platform),  
 main = "Платформы для игр",  
 las = 2)



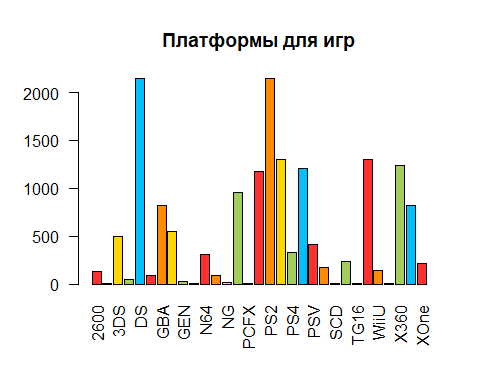
# Наиболее популярными для игр являются платформы DS и PS2

### Задача 9

Измените цвета столбцов диаграммы из задачи 8 таким образом, чтобы последовательность из пяти разных цветов повторялась необходимое число раз (по числу столбцов). Добавьте название графика на русском языке.

*Ответ:*

cols\_pl <- c("firebrick1", "darkorange", "gold", "darkolivegreen3", "deepskyblue1")   
   
barplot(table(games$Platform),  
 main = "Платформы для игр",  
 las = 2,  
 col = cols\_pl)

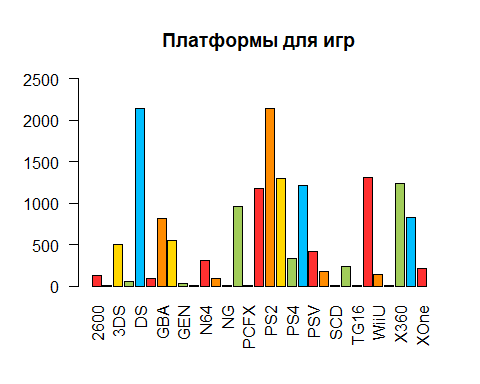


### Задача 10

Скорректируйте вертикальную ось диаграммы из задачи 8 таким образом, чтобы максимальное указанное значение превышало высоту самого высокого столбца.

*Ответ:*

cols\_pl <- c("firebrick1", "darkorange", "gold", "darkolivegreen3", "deepskyblue1")   
   
barplot(table(games$Platform),  
 main = "Платформы для игр",  
 las = 2,  
 col = cols\_pl,  
 ylim = c(0, 2500))



### Задача 11

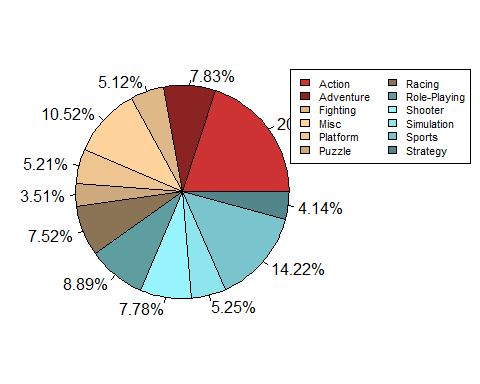
Постройте круговую диаграмму для показателя «жанр игры». Измените цвета диаграммы таким образом, чтобы каждому жанру соответствовал уникальный цвет. Скорректируйте диаграмму таким образом, чтобы на самой диаграмме были указаны частоты в процентах, а в легенде графика — соответствие между цветами и жанрами. При необходимости, если диаграмма по-прежнему выглядит не очень презентабельно, внесите дополнительные изменения по своему усмотрению.

*Ответ:*

unique(games$Genre)

## [1] "Platform" "Shooter" "Misc" "Role-Playing" "Action"   
## [6] "Sports" "Racing" "Puzzle" "Strategy" "Adventure"   
## [11] "Simulation" "Fighting"

cols <- colors()[35:46]  
tab <- table(games$Genre)  
perc <- paste0(round((tab / sum(tab) \* 100), 2), "%")  
location <- "topright"  
  
par(mar = c(2, 2, 2, 8), xpd = TRUE)  
pie(table(games$Genre),  
 col = colors()[35:46],  
 labels = perc)  
legend(location, inset = c(-0.5, 0.1), names(tab), cex = 0.7, fill = cols, ncol = 2, bty = "y", xpd = TRUE)



### Задача 12

Постройте линейный график, который отражает динамику показателя «скорость скачивания файла в торрент-программе в мегабайтах в секунду» в течение случайно выбранной минуты. Значения этого показателя сохранены в векторе torr\_pace.

torr\_pace <- c(0.3, 0.6, 1.5, 1.7, 1.2, 0.8, 1.7, 1.8, 2.0, 2.5,   
 2.2, 2.7, 3.3, 3.1, 3.5, 3.0, 3.2, 3.3, 3.8, 4.0,   
 3.7, 3.9, 3.5, 3.3, 3.2, 2.8, 2.5, 2.2, 1.8, 1.5,  
 1.4, 1.5, 1.2, 1.6, 1.8, 2.0, 2.2, 1.9, 1.8, 1.4,  
 1.2, 1.0, 0.9, 0.8, 1.1, 1.2, 1.3, 1.0, 0.8, 0.7,   
 1.1, 1.3, 1.5, 1.7, 1.6, 1.8, 1.9, 2.0, 1.5, 1.1)

Отредактируйте график таким образом, чтобы по нему было удобно отслеживать изменения значений показателя (добавьте подписи, разлиновку, при необходимости сделайте деления по осям более детальными).

*Ответ:*

torr\_pace <- c(0.3, 0.6, 1.5, 1.7, 1.2, 0.8, 1.7, 1.8, 2.0, 2.5,  
 2.2, 2.7, 3.3, 3.1, 3.5, 3.0, 3.2, 3.3, 3.8, 4.0,  
 3.7, 3.9, 3.5, 3.3, 3.2, 2.8, 2.5, 2.2, 1.8, 1.5,  
 1.4, 1.5, 1.2, 1.6, 1.8, 2.0, 2.2, 1.9, 1.8, 1.4,  
 1.2, 1.0, 0.9, 0.8, 1.1, 1.2, 1.3, 1.0, 0.8, 0.7,  
 1.1, 1.3, 1.5, 1.7, 1.6, 1.8, 1.9, 2.0, 1.5, 1.1)  
seconds <- 1:60  
  
plot(seconds, torr\_pace,  
 xlab = "Секунды",  
 ylab = "Скорость в мб",  
 ylim = c(0, 5),  
 pch = 20,  
 type = "o")  
grid(nx = 60, ny = 10)

