

**СОФИЙСКИ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“**

**ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА**

**СТАТИСТИКА И ЕМПИРИЧНИ МЕТОДИ – ПРАКТИКУМ ПО R**

**КУРСОВ ПРОЕКТ**

**Тема:**

**Изследване на продажбите на видеоигри за 2016-та**

**Румен Иванов Руменов, ФН 71539**

**25 Май 2020 г.**

**Обща информация**

Проектът изследва продажбите на видеоигри през 2016 г. жанрове, с цел идентифициране на ключови популярни такива, в които хипотетично студио-разработчик на игри би се фокусирало за предстоящите си проекти.

Съгласно заданието, са извършени следните анализи и изчисления:

* Анализ на едномерна променлива - глобални продажби по жанр:
  + обща стойност, средна стойност, обхват и медиана;
  + квантил;
  + дисперсия.
* Сравнение на категорийна величина (жанр) и числова величина (продажби):
  + boxplot;
  + t-test;
  + one-way ANOVA.
* Сравнение на числова с числова величина:
  + dotplot
  + корелация
  + линейна регресия

Даннните са набавени чрез сайта kaggle и са достъпни на следния линк: <https://www.kaggle.com/sidtwr/videogames-sales-dataset/data>. Използван е файлът Video\_Games\_Sales\_as\_at\_22\_Dec\_2016.csv.

Проектът е качен и в GitHub на следния адрес: <https://github.com/RazorDude/fmi-sem-project>.

**Програма**, написана на езика R

library(ggplot2)

library(dplyr)

library(reshape2)

data <- read.csv('data/Video\_Games\_Sales\_as\_at\_22\_Dec\_2016.csv')

View(data)

# Task 2.1 - location and dispersion of sales by genre

dataByGenre <- group\_by(data, Genre)

dataByGenre <- summarise(

dataByGenre,

total = sum(Global\_Sales),

mean = mean(Global\_Sales),

range = paste(range(Global\_Sales)[1], ' - ', range(Global\_Sales)[2]),

median = median(Global\_Sales)

)

View(dataByGenre) # Location data

View(quantile(dataByGenre$total)) # Total sales quantile

var(dataByGenre$total) # Dispersion variability

# Task 2.2

boxplot(data$Global\_Sales ~ data$Genre, xlab='Genre', ylab='Global Sales', main='Global Sales By Genre')

t.test(data$Global\_Sales[data$Genre=='Strategy'], data$Global\_Sales[data$Genre=='Shooter'])

# One-way ANOVA

model<-aov(data$Global\_Sales ~ data$Genre)

summary(model)

anova(model)

plot(TukeyHSD(aov(Global\_Sales ~ as.factor(Genre), data)))

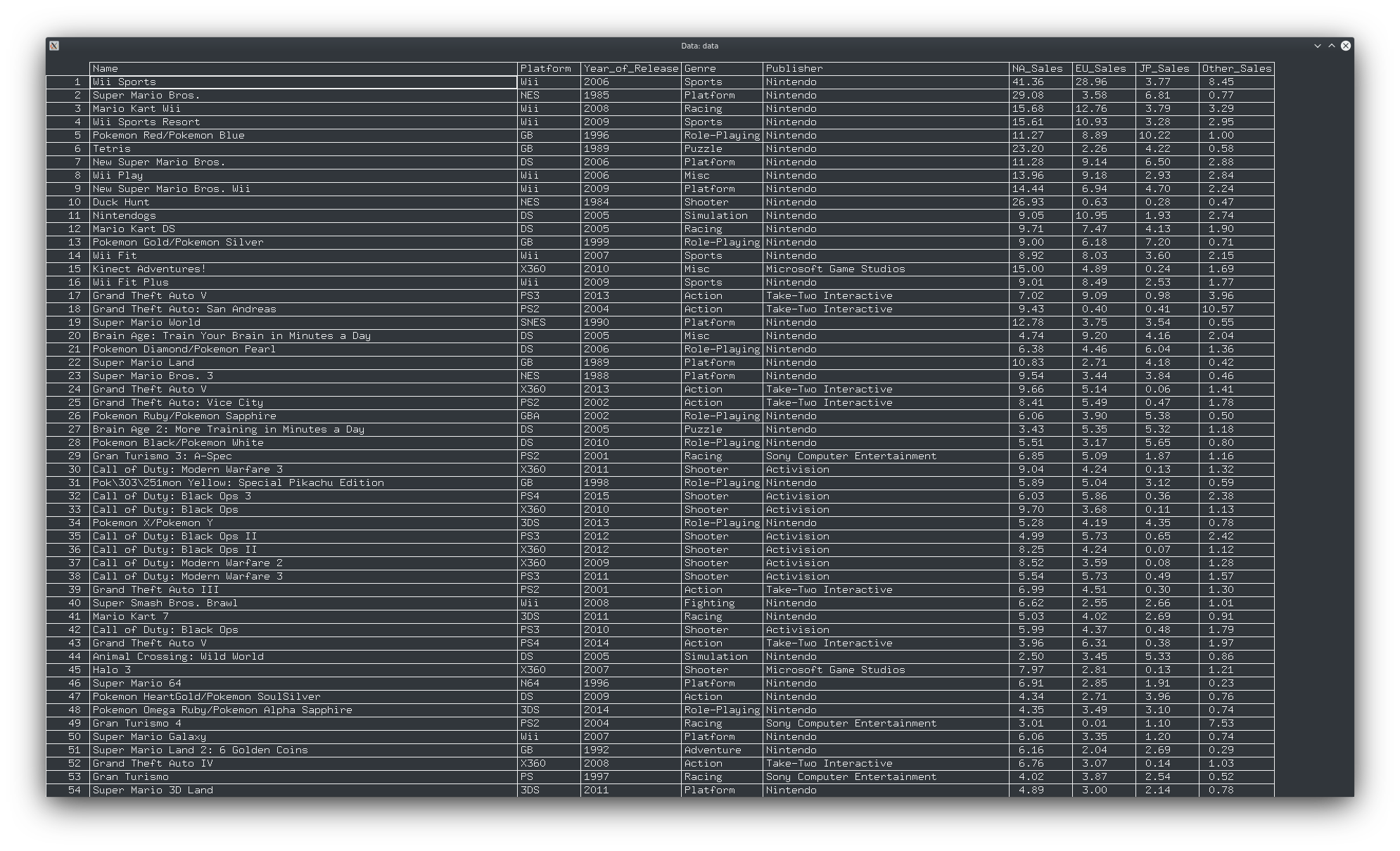
while(1==1) {

Sys.sleep(10000)

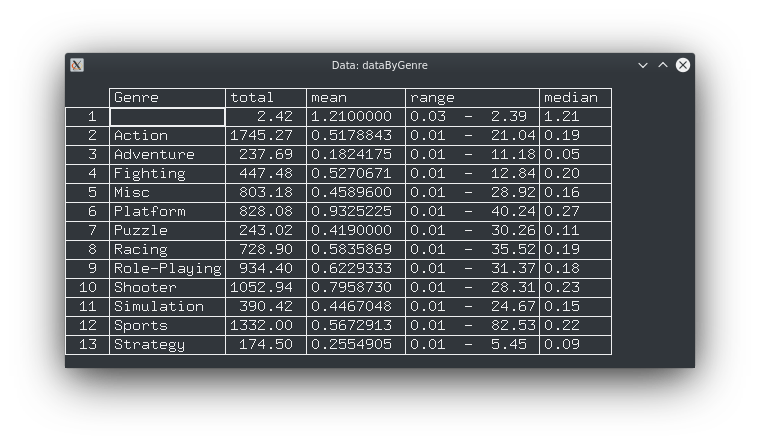
}

Резултати от изпълнението на програмата

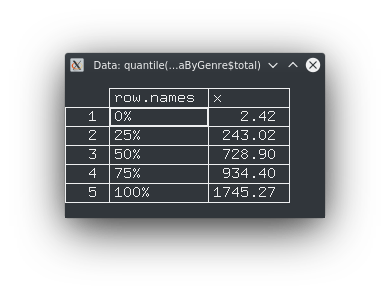
1. Частичен преглед на първоначалните данни ([висока резолюция](https://drive.google.com/open?id=1QapjVVKciohKrk78y9soca15hHUxcSSw))



2. Анализ на данните по жанр – общи продажби, средна стойност на продажбите, обхват (граници) и медиана ([висока резолюция](https://drive.google.com/open?id=1ROEXkvqjlTh7fH-BdtLZe7e96_lJ6rSD))



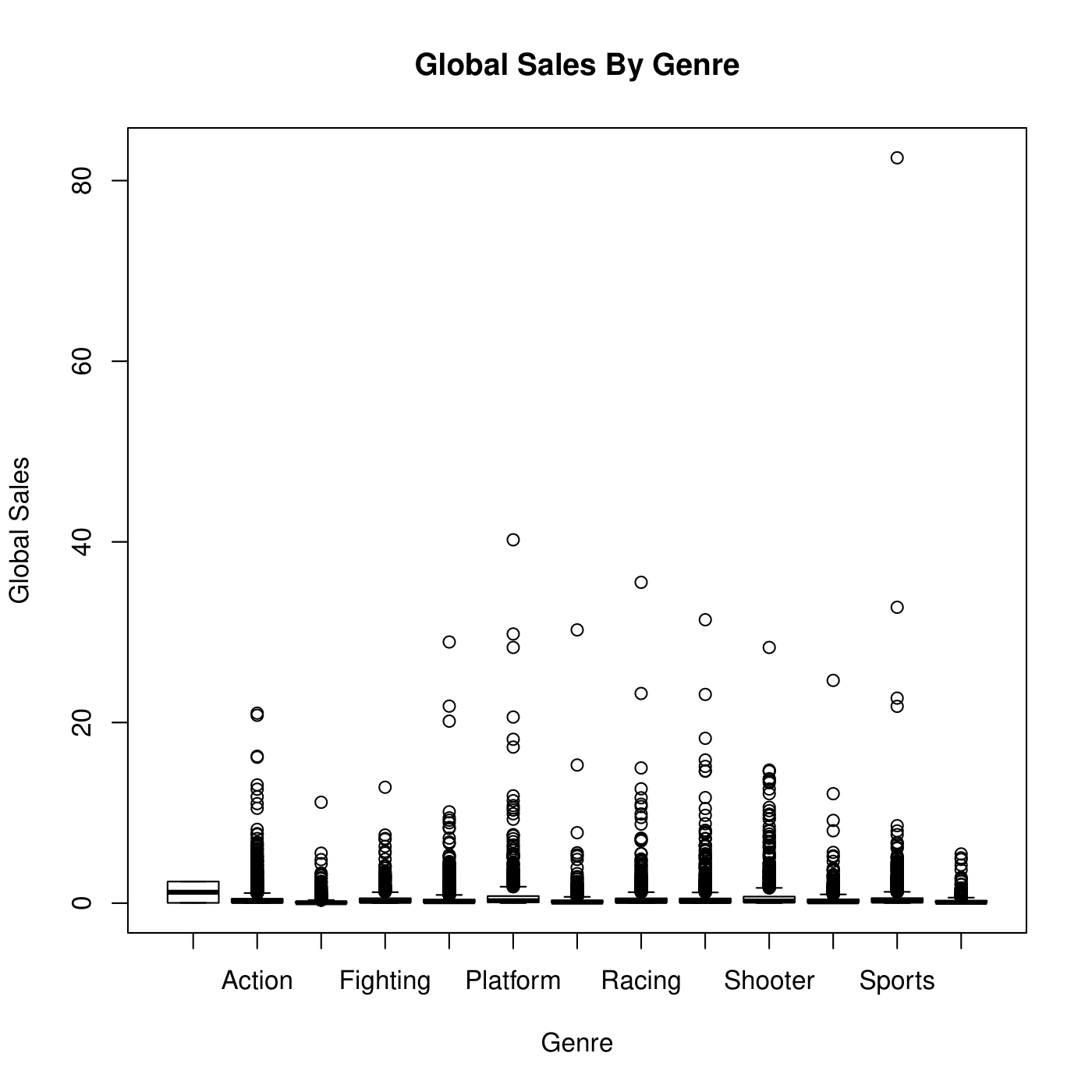
3. Анализ на данните по жанр – общи продажби (квантил) ([висока резолюция](https://drive.google.com/open?id=1OLbxmHuJyyp9bEfZe5AJk8i8S8bys-vG))



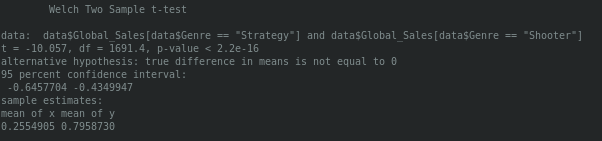
4. Анализ на данните по жанр – дисперсия на общите продажби - var(dataByGenre$total)



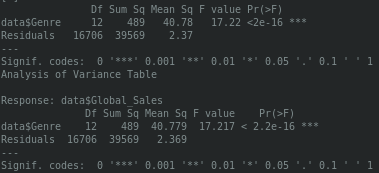
5. Boxplot на общите продажби по жанр ([висока резолюция](https://drive.google.com/open?id=1_jVrSMZuhzMWf2FCkOQtS7odPt6BSi0U))



6. Т-Тест на общите продажби по жанрове „Стратегия“ и „Шутър“ ([висока резолюция](https://drive.google.com/open?id=1YAwxGqgW8rXCtzYWAF7AQSH3xSgGpaW7))



7. One-way ANOVA – общи продажби и жанр ([висока резолюция](https://drive.google.com/open?id=110X5MIeBbCYjSr05bBMGqvJI9oXS6_6i))



8. Dotplot на резултатите от тест на Tukey (TukeyHSD) за изследване на съществени разлики в глобалните продажби ([висока резолюция](https://drive.google.com/open?id=1CnN7G_Wzjyt8sT2aJ4QH6oqEWP5ZiKhu))

