**Задание 1**

К частям таблицы можно обращаться так же, как и к матрицам.  
HairEyeColor - таблица с данными, встроенными в R. Посмотрите на неё в R. Команда dimnames(HairEyeColor) позволит нам посмотреть, какие измерения есть в этой таблице и как они называются. Например, чтобы обратиться к части таблицы, в которой хранятся данные только о мужчинах, нам нужно выполнить следующую команду:

HairEyeColor[ , ,'Male']

Ваша задача в переменную **blond\_men** сохранить долю блондинов (Blond) от общего числа кареглазых мужчин.

**Обратите внимание, что нужны не проценты, а просто доля, то есть десятичная дробь  (например, не 10%, а 0.1).**

**Задание 2**

С таблицами, как и с матрицами, можно совершать разные арифметические операции, например, суммировать все элементы таблицы.  
Посчитайте число голубоглазых женщин в наборе данных HairEyeColor.

**Задание 3**

Постройте столбчатую диаграмму распределения цвета глаз по цвету волос только у женщин из таблицы HairEyeColor. По оси X должен идти цвет волос, цвет столбиков должен отражать цвет глаз. По оси Y - количество наблюдений.

Чтобы построить столбчатую диаграмму в ggplot, вам нужно подключить нужный пакет, затем преобразовать таблицу HairEyeColor в data frame:

mydata <- as.data.frame(HairEyeColor)

Постройте график на основе предложенного кода. Укажите, чему равен аргумент data, что должно находиться в aes(). Изучите справку по geom\_bar(), чтобы узнать, чему должен равняться аргумент position для отображения цвета глаз в виде соседних столбиков, также вам может быть полезна эта памятка. Там же вы найдёте ответ на вопрос, за что отвечает аргумент stat. С помощью scale\_fill\_manual мы говорим графику, что мы хотим, чтобы он использовал указанные нами цвета.

Дополните предложенный код:

library("ggplot2")

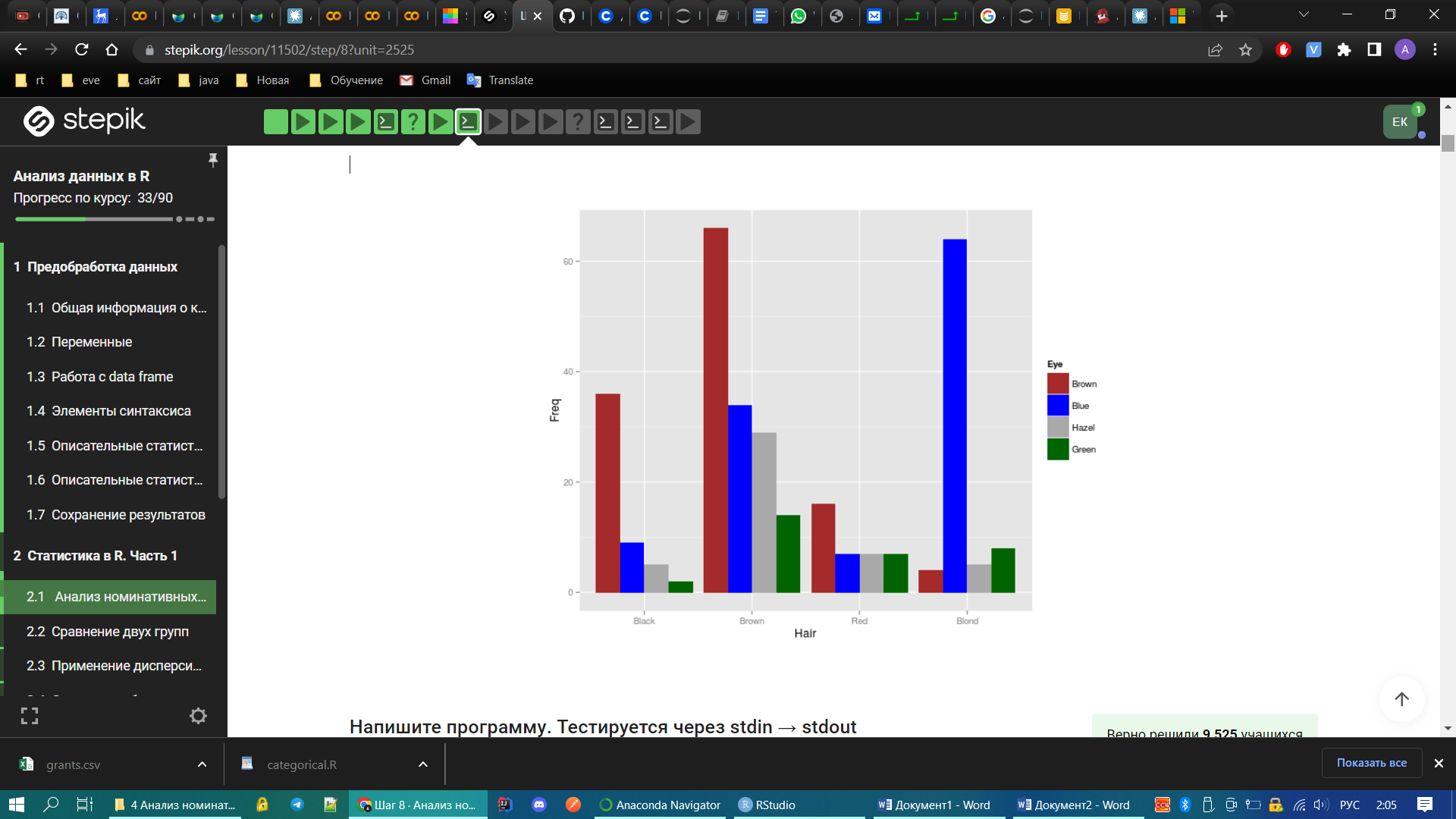
mydata <- as.data.frame(HairEyeColor)

ggplot(data = , aes(x = , y = Freq)) +

geom\_bar(stat="identity", position = ) +

scale\_fill\_manual(values=c("Brown", "Blue", "Darkgrey", "Darkgreen"))

В случае, если все сделано правильно, он будет выглядеть так (обратите внимание на название осей и легенды):



**Задание 4**

На основе таблицы HairEyeColor создайте ещё одну таблицу, в которой хранится информация о распределении цвета глаз у женщин-брюнеток (Hair = 'Black'). Проведите тест равномерности распределения цвета глаз у брюнеток и выведите значение хи-квадрата для этого теста.

**Задание 5**

Воспользуемся данными diamonds из библиотеки ggplot2. При помощи критерия Хи - квадрат проверьте гипотезу о взаимосвязи цены (price) и каратов (carat) бриллиантов. Для этого сначала нужно перевести эти количественные переменные в формат пригодный для Хи - квадрат. Создайте две новые переменные в данных diamonds:

factor\_price - где будет 1, если значение цены больше либо равно чем среднее, и 0, если значение цены ниже среднего цены по выборке.

factor\_carat - где будет 1, если число карат больше либо равно чем среднее, и 0, если ниже среднего числа карат по выборке.

Важный момент - на больших данных цикл for() работает довольно медленно, постарайтесь решить эту задачу без его использования!

Используя эти шкалы при помощи Хи - квадрат проверьте исходную гипотезу. Сохраните в переменную main\_stat значение критерия Хи - квадрат.

**Задание 6**

При помощи точного критерия Фишера проверьте гипотезу о взаимосвязи типа коробки передач (am) и типа двигателя (vs) в данных mtcars. Результат выполнения критерия сохраните в переменную. Получившийся p - уровень значимости сохраните в переменную fisher\_test.

**t-test**

**Задание 7**

Воспользуемся еще одним встроенным набором данных в R - ToothGrowth. Данные позволяют исследовать рост зубов у морских свинок в зависимости от дозировки витамина C и типа потребляемых продуктов. Сравните среднее значение длины зубов свинок, которые потребляли апельсиновый сок (OJ) с дозировкой 0.5 миллиграмм, со средним значением длины зубов свинок, которые потребляли аскорбиновую кислоту (VC) с дозировкой 2 миллиграмма. Значение t - критерия сохраните в переменную t\_stat.

**Задание 8**

В этом задании нужно проверить гипотезу о равенстве средних двух выборок, загрузив набор данных и выполнив все необходимые операции на вашем компьютере. В скачанных данных вы найдете две переменные: количественную переменную, и номинативную переменную с двумя градациями (которая разделяет наблюдения на две группы).

Для того чтобы без труда прочитать скачанные данные воспользуйтесь функцией:

read.table("dataset\_11504\_15.txt")

Сначала с помощью теста Бартлетта проверьте гомогенность дисперсий двух выборок. В случае, если дисперсии значимо не отличаются (с уровнем 0.05), примените тест Стьюдента, иначе - непараметрический тест (Манна-Уитни). Выведите получившийся p-value, с точностью четыре знака после запятой.

Обратите внимание, что по умолчанию в t.test стоит var.equal = FALSE, так как мы будем применять его только в случае гомогенности дисперсий, измените значение этого параметра на var.equal = TRUE.