**Задание 4 (стр. 12):**

Используя фрейм diamonds, определите все возможные сочетания качества огранки (cut), цвета (color) и чистоты (clarity) алмазов. Результат представьте в таблице со столбцами: № п/п, cut, color, clarity.

**Ход работы:**

Для начала необходимо создать R script lab1r.R, в котором будут использованы библиотеки ggplot2 и tidyr, которые необходимо предварительно установить командами «install.packages('ggplot2')» и «install.packages('tidyr')». Из ggplot2 нам нужен будет набор данных diamonds, а из tidyr функции unite() и separate() для объединения и разбиения столбцов. После импортирования diamonds мы извлекаем необходимые нам столбцы занося из в переменные, являющиеся факторами с именами cut, color, clarity и предварительно очищая столбцы от дублирующихся значений с помощью функции unique(). Далее с помощью функции expand.grid() создаем таблицу с всеми возможными сочетаниями значений cut и color. Данная функция работает только с двумя векторами или факторами и поэтому, чтобы перемножить полученную таблицу с фактором clarity, необходимо объединить столбцы cut и color в один. Тут мы и применяем функции unite() и separate(). Результат работы скрипта приведен на рисунках 1, 2.

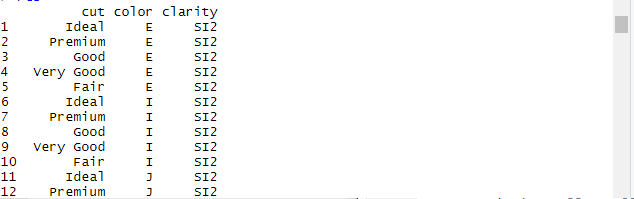


Рисунок 1 – Начало полученной таблицы

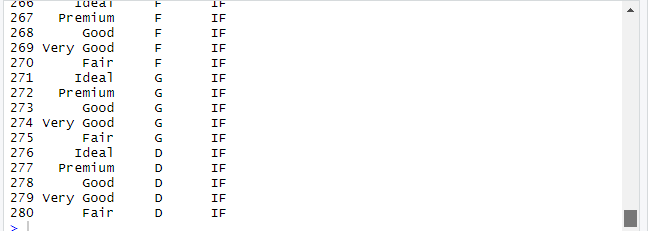


Рисунок 2 – Конец полученной таблицы

Код файла lab1r.R представлен ниже.

require(ggplot2)

require(tidyr)

table = diamonds

cut = unique(table[["cut"]])

color = unique(table[["color"]])

clarity = unique(table[["clarity"]])

res1 = expand.grid(cut, color)

res1 = unite(res1,"United", c("Var1","Var2"),sep="\_")

res1= as.factor(res1[["United"]])

res2 = expand.grid(res1, clarity)

res = separate(res2,"Var1",c("cut","color"),sep="\_")

names(res)[names(res) == "Var2"] <- "clarity"

res