Лабораторная работа № 10

Задание 1. Используя библиотеку dplyr и конвеерный оператор (%>%) для встроенного дата-фрейма warpbreaks определить, для каких комбинаций типа шерсти (wool) и натяжения (tension) справедливо, что среднее число поломок больше 25, либо максимальное количество поломок больше 42.

Задание 2. Установите соответствие между выражениями базового R и выражениями, формируемыми с использованием библиотеки dplyr.

|  |  |
| --- | --- |
| subset(avian, PDB > 0, c("Observer", "PDB")) | mutate(avian, PDB = PDB > 0) |
| aggregate(avian[, "PDB"], list(Observer = avian[, "Observer"]), function(x) sum(x > 0)) | filter(select(avian, Observer, PDB), PDB > 0) |
| avian[, "PDB"] <- avian[, "PDB"] > 0 | summarise(group\_by(avian, Observer), x = sum(PDB > 0)) |

Задание 3.

1. В RStudio подключите библиотеки stringr и dplyr.
2. Загрузите данные из файла avianhabitat.csv.
3. Используя функцию subset, выбрать из дата-фрейма все наблюдения, в которых имеется хотя бы одна карликовая береза, т.е. покрытие этим видом отлично от 0. В подфрейме вывести переменные: Site, Observer, PDB, DBHt.
4. Видоизмените переменную Site и представьте её в виде фактора с помощью следующего выражения:

avian$Site<-factor(str\_replace(avian$Site,"[:digit:]+",""))

1. С помощью функции aggregate осуществите агрегацию значений переменной DBHt для переменных Site и Observer. Из полученного подфрейма выделите фрагмент со значениями DBHt, которые не меньше 5. В результате выводится информация о карликовых березах, высота которых составляет хотя бы 5 см.
2. Выполните операции 3-5 с использованием конвейеров (оператора «%>%»).
3. Выполните операции 3-5 используя библиотеку dplyr и конвейеры.