

Практическая работа №9

Функции `ggplot()` пакета `ggplot2`

Цель: изучить возможности графического пакета `ggplot2` для анализа данных на примере функции `ggplot()`, освоить навыки работы построения графиков по слоям.

Данные для практической работы

1. Набор данных `Pedometer.csv` – данные шагометра.

Steps – общее число шагов за день;

Moderate – число шагов при умеренной скорости;

Min – время ходьбы со средней скоростью (в мин.);

kcal – количество калорий, сгоревших при ходьбе с умеренной скоростью;

Mile – общее количество пройденных миль;

Rain – тип погоды (**cold** – холодно, **rain** – дождь, **shine** – солнечно);

Day – день недели (**U**=воскресенье, **M**=понедельник, **T**=вторник, **W**=среда, **R**=четверг, **F**=пятница, **S**=суббота);

DayType – тип дня (**Weekday** – рабочий день, **Weekend** – выходной).

Варианты заданий

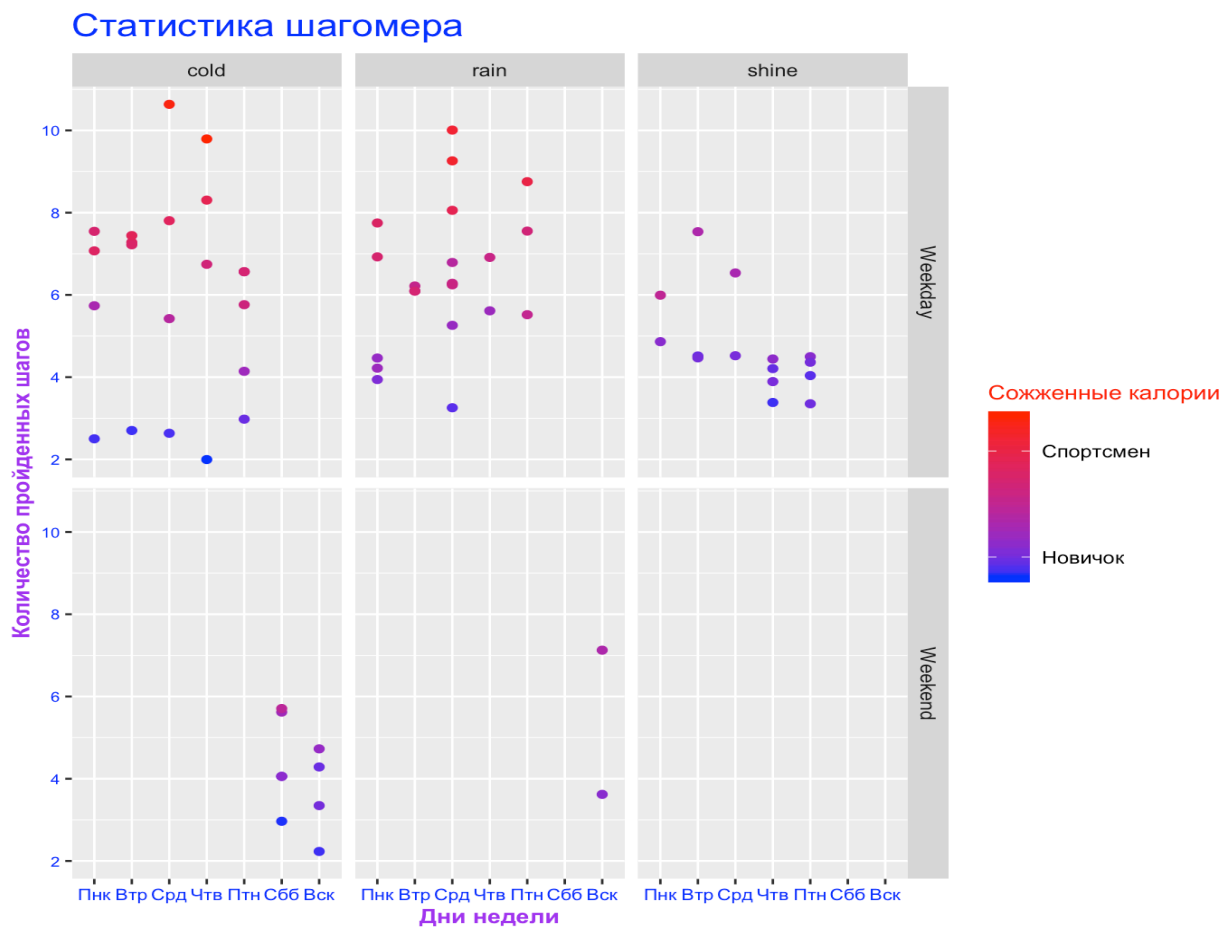
№ студента в журнале	№ варианта	Цвет (точек и линий)	Цвет заполнения (столбиков, области под графиком)	Прозрачность
1,11,21	1	Красный	Желтый	50%
2,12,22	2	Оранжевый	Зеленый	51%
3,13,23	3	Желтый	Голубой	52%
4,14,24	4	Зеленый	Синий	53%
5,15,25	5	Голубой	Фиолетовый	54%
6,16,26	6	Синий	Розовый	49%
7,17,27	7	Фиолетовый	Черный	48%
8,18,28	8	Розовый	Коричневый	47%
9,19,29	9	Черный	Красный	46%
10,20,30	10	Коричневый	Оранжевый	45%

Задания

1. Создать новую переменную **Step**, загрузив в **R** данные **Pedometer.csv**.
2. Сделать переменную день недели факторной и упорядочить уровни переменной по порядку (с понедельника по воскресенье).
3. Используя функцию **ggplot**, инициировать график **StepG**, указав **x** – день недели, **y** – количество калорий, сгоревших при ходьбе.
4. Добавить к графику **StepG** слой с диаграммой рассеяния.
5. С помощью аргумента **aes()** задать визуальные свойства графика: окрасить точки на основе типа дня, изменить размер точек на основе типа погоды.
6. Сделать прозрачность точек на графике в соответствии с вариантом.
7. Инициировать новый график **StepG1**, указав **x** – общее число шагов за день и **y** – количество калорий, сгоревших при ходьбе.
8. Добавить к графику **StepG1** слой с диаграммой рассеяния.
9. Добавить сглаженную линию тренда с 95%-ным доверительным интервалом.
10. Убрать доверительный интервал, оставив на графике только линию тренда.
11. Отрегулировать диапазон значений осей графика, чтобы приблизить график для рассмотрения.
12. Инициировать новый график **StepG2**, состоящий из значений переменной тип дня.
13. Добавить на график **StepG2** слой со столбиковой диаграммой числа шагов в разные типы дня.
14. Поменять цвет столбиков, цвет и тип линий, окаймляющих столбики, в соответствии с заданным вариантом.
15. Инициировать график **StepG3**, состоящий из значений возраста пассажиров.
16. Добавить на график **StepG3** слой с гистограммой числа шагов, сделанных за день. Внешний вид гистограммы задать по варианту, размер классового промежутка соответствует номеру варианта, умноженного на 100.
17. Построить цветную (в соответствии с вариантом) столбиковую диаграмму среднего количества шагов, пройденных за день для разных типов дня.
18. Построить график пропорций набора данных **Step** по переменной день недели тремя способами:
 - расположив каждый сегмент друг на друга;
 - расположив каждый сегмент друг на друга, заполняя пространство от 0 до 1;

– выстраивая каждый сегмент рядом с другим.

19. Построить график **StepG4** распределения пройденных миль в разную погоду. Цвет точек раскрасить в зависимости от сожженных калорий за время тренировки. Добавить на график заголовок, названия осей, задать цвет и размер всем текстовым элементам графика.
20. Изучить функции пакета **ggplot2** в разделе **help**. Найти те функции, которые меняют визуальные свойства легенды. Поменять цвет, размер, фон легенды.
21. Построить следующий график:



22. Построить круговую диаграмму распределения типа погоды. Сохранить график в переменную **StepG5**.
23. Сохранить график **StepG5** в файл.

Домашнее задание

1. Оформить отчет по практической работе, согласно форме для практической работы №8.
2. Набор данных **Titanic.csv** – данные о пассажирах Титаника.

Name – имя пассажира;

PClass – класс пассажира (1st, 2nd, 3rd);

Age – возраст (в годах);

Sex – пол;

Survived – выживший (1 = выживший, 0 = погибший);

SexCode – код пола (1 = женщина, 0 = мужчина).

Построить диаграмму рассеяния возраста всех пассажиров для разного класса. Цвет точек раскрасить в зависимости от возраста. На график добавить заголовок, названия осей, название легенды и меток (20 лет – молодые, 40 лет – среднего возраста, 60 лет – пожилые).

3. Разбить график на два по уровням выживаемости, расположить графики горизонтально.
4. Сохранить график в файл.

Контрольные вопросы

1. Поясните принцип построения графика с помощью функции **ggplot()**.
2. Какой слой отвечает за тип статистического преобразования?
3. Какой аргумент функции **ggplot()** отвечает за визуальную разметку?
4. С помощью какого слоя строится диаграмма рассеяния?
5. Как убрать доверительный интервал у сглаженной линии тренда?
6. К анализу каких данных применяется построение столбиковой диаграммы и с помощью какого слоя она строится?
7. Каким аргументом регулируется прозрачность гистограмм?
8. Перечислите основные отличия трёх способов задания эстетических атрибутов.
9. Какое значение аргумента **position** необходимо использовать, чтобы объекты выстроились в ряд?
10. С помощью какого слоя можно приблизить часть графика?
11. Какой слой отвечает за добавление заголовка графика и названия осей?
12. Назовите способы организации разбиения графика на несколько подграфиков.