

Практическая работа №8

Анализ данных с использованием функции `qplot()` пакета `ggplot2`

Цель: изучить возможности графического пакета `ggplot2` для анализа данных на примере функции `qplot()`.

Задания

1. Установить, загрузить пакет `ggplot2` в среду `R`.
2. Создать переменную `salary`, загрузить в неё данные из файла `Salaries.csv`.
3. Построить диаграмму рассеяния: по оси `ох` указать количество лет, отработанных преподавателем (`yrs.service`), по оси `оу` – зарплату преподавателей (`salary`).
4. Изменить цвет точек на графике в зависимости от профессорского звания (`rank`).
5. Построить сглаживающую кривую.
6. Построить на одном графике и диаграмму рассеяния, и линии тренда.
7. Изменить форму точек в зависимости от преподаваемой дисциплины (`discipline`).
8. Изменить размер точек в зависимости от того, сколько лет преподаватель отработал в данном университете до получения звания профессора (`yrs.since.phd`).
9. Сделать график более читаемым, сделав точки прозрачными на 60%.
10. Построить гистограмму для переменной зарплата (`salary`).
11. Задать размер классового промежутка для гистограммы из п.10, добавить длину оси `ох` и цвет гистограммы, зависящий от преподаваемого предмета (`discipline`).
12. Создать кривые плотности вероятности зарплат преподавателей (`salary`), цвет кривых задать уровнями звания преподавателей (`rank`).
13. Заполнить области под кривыми плотности вероятности графика из п.12 в зависимости от звания преподавателей (`rank`).
14. Построить полигон распределения количества лет, отработанных преподавателем в университете (`yrs.service`).
15. Добавить цвет на график из п.14 в зависимости от преподаваемой дисциплины (`discipline`).
16. Построить составную столбиковую диаграмму преподаваемой дисциплины (`discipline`), заполняя столбики в зависимости от звания преподавателя

(**rank**).

17. Построить диаграмму размаха зарплаты преподавателей (**salary**) по группам звания (**rank**), исключив выбросы.
18. На диаграмму размаха из п.17 добавить одномерную диаграмму рассеяния значений зарплаты преподавателей (**salary**) по уровням их звания (**rank**).
19. Построить категоризованные графики плотности вероятности зарплаты преподавателей (**salary**) для разных групп их звания (**rank**). Область под графиками закрасить в зависимости от звания (**rank**).
20. Построить графики линий тренда зависимости зарплаты преподавателей (**salary**) от количества преподаваемых лет в университете (**yrs.service**) для разных групп звания преподавателей (**rank**) и различных дисциплин, которые они преподают (**discipline**).
21. Постройте графики, которые помогут сделать выводы, зависит ли зарплата (**salary**) преподавателей от:
 - количества преподаваемых лет в университете (**yrs.service**);
 - дисциплины (**discipline**);
 - звания (**rank**).

Сделать выводы.

Контрольные вопросы

1. Что такое **ggplot2**?
2. Перечислите основные строительные блоки **ggplot2**?
3. Перечислите основные геометрические объекты в **ggplot2**
4. В чем особенность функции **qplot()**?
5. Какой аргумент функции **qplot()** отвечает за способ разбиения рисунка на подобласти?
6. Как задать заголовок графику функции?
7. Как окрасить точки на диаграмме рассеяния по значениям третьей переменной?
8. Какой аргумент функции **qplot()** используется, чтобы улучшить визуальный вид графика с большим количеством точек?
9. Какое значение должен принимать аргумент **geom** функции **qplot()**, чтобы на графике была изображена линия тренда?
10. Как задать классовый промежуток на гистограмме?
11. Как исключить выбросы на графике диаграммы размаха?

Домашнее задание

1. Оформить отчет по лабораторной работе.
2. В новую переменную **Forbes** загрузить данные **Forbes.csv** (рейтинг крупнейших компаний мира по версии Forbes).
3. Изучить структуру данных.
4. Построить диаграмму рассеяния: по оси *ох* – значения переменной продажи (**sales**), по оси *оу* – доходы (**profits**).
5. Добавить название графика, названия осей, изменить ширину точек, сделать график цветным.
6. Добавить сглаживающую линию к диаграмме рассеяния, задать цвет линии.

Отчет по практической работе №8

1. Титульный лист, включая название работы, ФИО студента, учебную группу.
2. Оглавление.
3. Отчет:
 - 3.1. Номер и описание задания.
 - 3.2. Код решения.
 - 3.3. Результат в консоли.
 - 3.4. Скрин результата из консоли **RStudio**.
4. Список использованных источников литературы.