

УДК 311:004.9R
ББК 60.6с515
М32

М32 Мастицкий С. Э.
 Визуализация данных с помощью ggplot2. – М.: ДМК Пресс, 2017. – 222 с.: ил.

ISBN 978-5-97060-470-0

Визуализация данных играет важную роль на всех этапах статистического анализа – от первичного ознакомления со свойствами данных до диагностики качества построенных моделей и представления полученных результатов. Из всего разнообразия статистических программ выделяется R – интенсивно развивающаяся и свободно распространяемая система статистических вычислений, в которой реализовано множество классических и современных методов анализа данных. Программные реализации алгоритмов, входящих в базовую версию R, проверены на практике не одним поколением пользователей и ученых. Кроме того, пользователи R постоянно разрабатывают многочисленные дополнения для этой системы. Настоящая книга посвящена ggplot2 – одному из таких пакетов, который значительно расширяет и без того богатые базовые графические возможности R.

В 2015 г. ggplot2 был установлен более миллиона раз. Такая популярность этого пакета обусловлена несколькими причинами, среди которых можно отметить эстетическую привлекательность и пригодное для публикации качество получаемых с его помощью графиков, возможность создавать пользовательские типы диаграмм, а также большой набор опций для тонкой настройки внешнего вида графиков. В этой книге описаны основы работы с ggplot2 и приведены многочисленные примеры кода, которые читатели легко могут модифицировать для собственных нужд.

Книга окажется полезной для всех пользователей R, желающих освоить новый мощный инструмент анализа данных.

УДК 311:004.9R
ББК 60.6с515

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 978-5-97060-470-0

© Мастицкий С.Э.
 © Оформление, издание, ДМК Пресс, 2017

Оглавление

Предисловие	7
1 Введение	10
1.1 Что представляет собой <code>ggplot2</code> ?	10
1.2 Установка <code>ggplot2</code> и начало работы	10
1.3 Грамматика графических элементов	11
1.4 Данные, используемые в примерах	12
2 Функция <code>qplot()</code>: быстрое решение для задач визуализации	15
2.1 Аргументы функции <code>qplot()</code>	15
2.2 Построение диаграмм рассеяния с помощью <code>qplot()</code>	16
2.3 Другие примеры использования <code>qplot()</code>	19
2.3.1 Линии тренда	20
2.3.2 Одномерные диаграммы рассеяния	21
2.3.3 Диаграммы размахов	22
2.3.4 Гистограммы, кривые плотности вероятности, полигоны частот	25
2.3.5 Столбиковые диаграммы	30
2.4 Категоризованные графики	31
3 Построение графиков слой за слоем	34
3.1 Аргументы функции <code>ggplot()</code>	34
3.2 Слои	35
3.3 Требования к данным	38
3.4 Присваивание эстетических атрибутов	39
3.5 Группирование данных	41
3.6 Геометрические объекты, реализованные в <code>ggplot2</code>	43
3.7 Статистические преобразования	46
4 Основные типы статистических графиков	49
4.1 Общие аргументы <code>geom</code> - и <code>stat</code> -функций	49
4.2 Визуализация одномерных распределений	50
4.2.1 Точечные диаграммы Уилкинсона: <code>geom_dotplot()</code>	50
4.2.2 Столбиковые диаграммы: <code>geom_bar()</code>	54
4.2.3 Гистограммы: <code>geom_histogram()</code>	60
4.2.4 Полигоны частот: <code>geom_freqpoly()</code>	62
4.2.5 Кривые плотности вероятности: <code>geom_density()</code>	64

4.2.6	Кумулятивные функции распределения: <code>geom_step()</code>	67
4.2.7	Квантильные графики: <code>stat_qq()</code>	71
4.3	Визуализация 2D- и 3D-распределений	73
4.3.1	Контуры плотности вероятности: <code>geom_density2d()</code>	74
4.3.2	Изолинии: <code>geom_contour()</code>	76
4.3.3	Сотовые диаграммы: <code>geom_hex()</code>	78
4.4	Визуализация сводной статистической информации о количественных переменных	79
4.4.1	Диаграммы диапазонов: <code>geom_linerange()</code> , <code>geom_pointrange()</code> , <code>geom_errorbar()</code> , <code>geom_crossbar()</code>	79
4.4.2	Диаграммы размахов: <code>geom_boxplot()</code>	85
4.4.3	Скрипичные диаграммы: <code>geom_violin()</code>	89
4.5	Визуализация зависимостей	92
4.5.1	Диаграммы рассеяния: <code>geom_point()</code>	93
4.5.2	Линии тренда: <code>geom_smooth()</code>	95
4.5.3	Линии квантильной регрессии: <code>geom_quantile()</code>	99
4.6	Визуализация временных рядов	101
4.6.1	Функция <code>geom_line()</code>	102
4.6.2	Функция <code>geom_ribbon()</code>	104
4.7	Тепловые карты: <code>geom_tile()</code>	105
4.8	Другие геометрические объекты	107
4.8.1	«График-щетка»: <code>geom_rug()</code>	107
4.8.2	Горизонтальные и вертикальные линии: <code>geom_hline()</code> , <code>geom_vline()</code>	109
4.8.3	Прямоугольные области: <code>geom_rect()</code>	111
4.8.4	Отрезки: <code>geom_segment()</code>	112
4.8.5	Ломанные линии: <code>geom_path()</code>	113
4.8.6	Многоугольники: <code>geom_polygon()</code>	116
4.8.7	Площадь под кривой: <code>geom_area()</code>	118
4.8.8	Текстовые аннотации: <code>geom_text()</code>	119
4.9	Географические карты: <code>geom_map()</code>	124
4.10	Добавление слоев при помощи функций семейства <code>stat</code>	131

5 Шкалы 134

5.1	Шкалы и их основные типы	134
5.2	Аргументы, общие для всех <code>scale</code> -функций	137
5.3	Шкалы положения	140
5.3.1	Шкалы положения для количественных переменных	140
5.3.2	Шкалы положения для дат и времени	145
5.3.3	Шкалы положения для качественных переменных	147
5.4	Цветовые шкалы	149
5.4.1	Цветовые шкалы для количественных переменных	150
5.4.2	Цветовые шкалы для качественных переменных	154
5.5	Пользовательские шкалы для качественных переменных	158
5.6	Тождественные шкалы	160

6	Системы координат	162
6.1	Декартова система и ее разновидности	163
6.2	Полярная система	166
6.3	Картографические проекции	168
7	Подробнее о категоризованных графиках	171
7.1	Два способа организации панелей	171
7.2	Функция <code>facet_grid()</code>	172
7.3	Функция <code>facet_wrap()</code>	180
8	Подготовка графиков к публикации	183
8.1	Стили	183
8.2	Создание составных рисунков	201
8.2.1	Использование окон просмотра	203
8.2.2	Использование пакета <code>gridExtra</code>	206
8.3	Экспорт графиков из среды R	208
9	Дополнительные ресурсы для изучения ggplot2	212
9.1	Литература	212
9.2	Онлайн-ресурсы	213
9.3	Расширения, созданные на основе <code>ggplot2</code>	214
	Предметный указатель	217