

# Визуализация данных: ggplot

Анастасия Миллер, июль 2019

# План

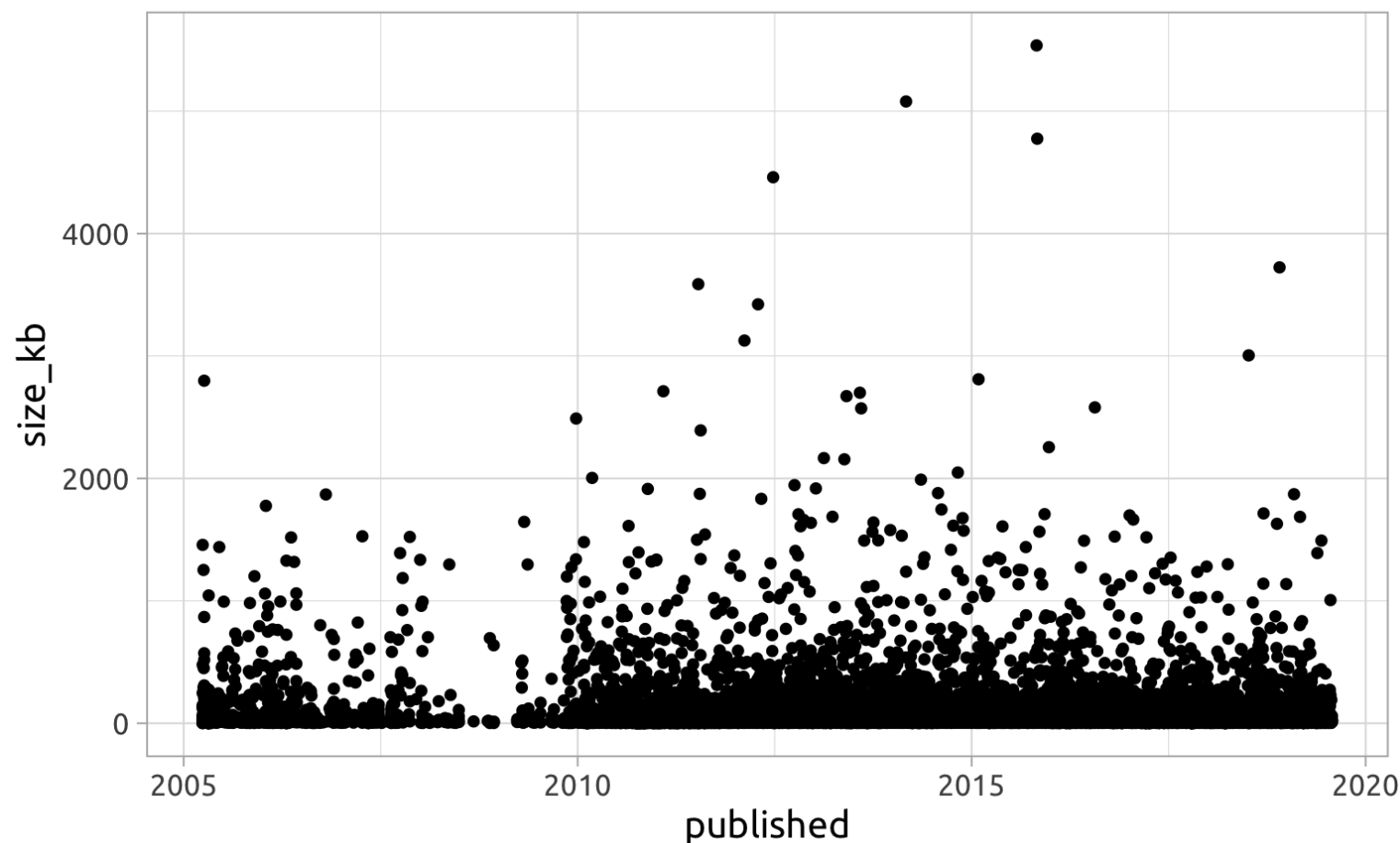
1. Идея послойного представления графика  
What is layered grammar of graphics
2. Aesthetics
3. Трансформации
4. Настройки: легенда, подписи, масштабы, шрифты

# Из чего состоит график

- Набор данных и соответствие между переменными и визуальными элементами

```
ggplot(df, aes(x = published, y = size_kb))
```

- Визуальное представление  
+ `geom_point()`



# Из чего состоит график

## на самом деле

- Набор данных и соответствие между переменными и визуальными элементами  
`ggplot(df, aes(x = published, y = size_kb))`
- [ $\geq 1$ ] Слой, описывающий представление информации  
`+ layer(  
 geom = "point",  
 stat = "identity",  
 position = "identity")`
- Шкалы для каждой использованной переменной  
`+ scale_x_date() + scale_y_continuous()`
- Система координат  
`+ coord_cartesian()`
- Деление на подграфики

# Aesthetics

соответствие между переменными и визуальными представлениями

- **aesthetics** are things that we can perceive on the graphic
- Каждый слой заранее определяет, какие представления он обязан и какие может показать: секция Aesthetics в `?geom_*`
- 👍 `aes(x = published, y = size_kb, colour = rating)`
- 👎 `aes(x = published, y = size_kb, colour = "red")`

# Задачи

- `geom_point`

- Покажите размеры фанфиков в зависимости от даты публикации
- Отметьте цветом рейтинг фанфика
- Используйте разную форму метки (квадрат, треугольник) для переводов и оригинальных произведений

- `geom_jitter`

- Покажите соотношение между категорией размера и фактическим размером
- Отметьте цветом рейтинг фанфика

# Шкалы

- Каждому визуальному представлению соответствует своя шкала
- Настройки шкалы:  
`scale_<aesthetic name>_<scale type>(parameters)`

Пример:

```
scale_x_discrete(  
  labels = c('', 'можно всем', 'осторожно', 'взрослым')  
)
```

# Задачи

- Покажите соотношение между категорией размера и фактическим размером. Отметьте цветом рейтинг фанфика.
- Покажите фактический размер на логарифмической шкале с основанием два  
То есть отсечки должны быть на 2, 4, 8 и так далее
- Переведите оси и легенду графика на русский.

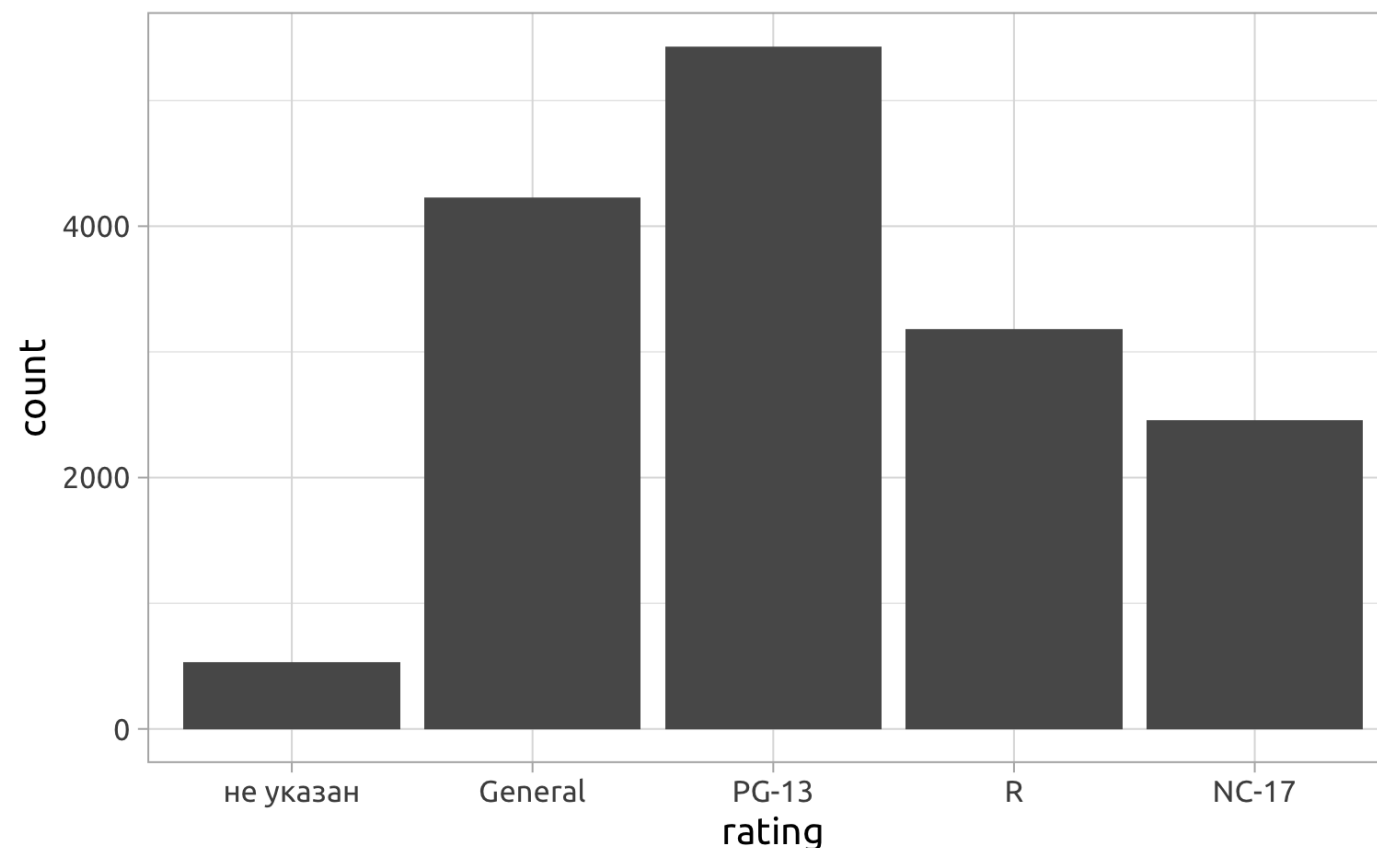


# Трансформации

- $[\geq 1]$  Слой, описывающий представление информации  
+ `layer(  
 geom = "point",  
 stat = "identity",  
 position = "identity")`
- `stat` преобразует данные, обычно резюмирует их
- Шкалы тоже можно трансформировать
- Чтобы написать своё преобразование, нужно определить новый класс, наследующий от `ggplot::Stat`

# Трансформации

- `ggplot(df, aes(x = rating)) +  
 geom_bar(stat = "count")`
- `df_n <- df %>% count(rating) %>% rename(count = n)  
 ggplot(df_n, aes(x = rating, y = count)) +  
 geom_bar(stat = "identity")`



# Задачи

- Покажите, сколько фанфиков публиковалось в каждый месяц, начиная с января 2015 года включительно.
- То же самое, но используйте `geom_bar(stat = "identity")`
- Покажите помесечное количество публикаций за последние три года. Цветом выделите различные рейтинги. Число публикаций за месяц обозначается точкой, точки соединяются линиями.

# Подграфики

- `facet_grid(size_cat ~ rating, switch = 'y')`

	не указан	General	PG-13	R	NC-17
small					
medium					
large					

# Подграфики

- `facet_wrap(~ rating)`



# Задачи

- Покажите ежегодное количество публикаций с 2010 года для каждого сочетания категории размера и рейтинга.  
Один размер — одна строка, один рейтинг — один столбец.
- Каковы самые популярные жанры? Насколько одни популярнее других? Нарисуйте ежегодное количество публикаций в каждом из пяти наиболее популярных жанров, каждый жанр — в отдельном подграфике.

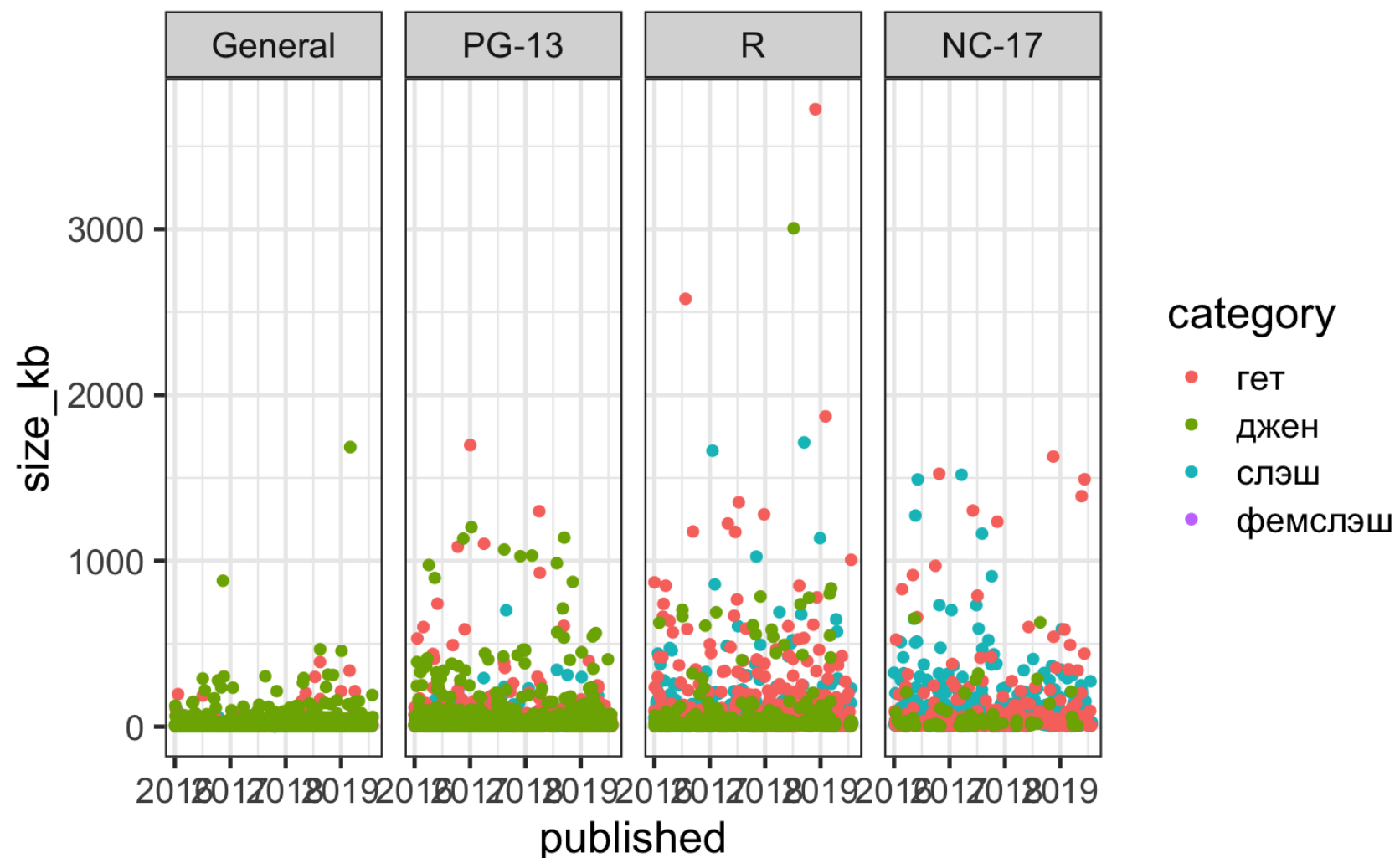
# Тема

- Как сделать график, который можно вставить в статью?
- Как повернуть надписи, сделать розовый фон и шрифт Monotype Corsiva?
- Как сделать одинаковый стиль у всех графиков?

# Тема

Как сделать график, который можно вставить в статью?

Использовать одну из готовых тем; лучше всего `theme_light()` или `theme_bw()`



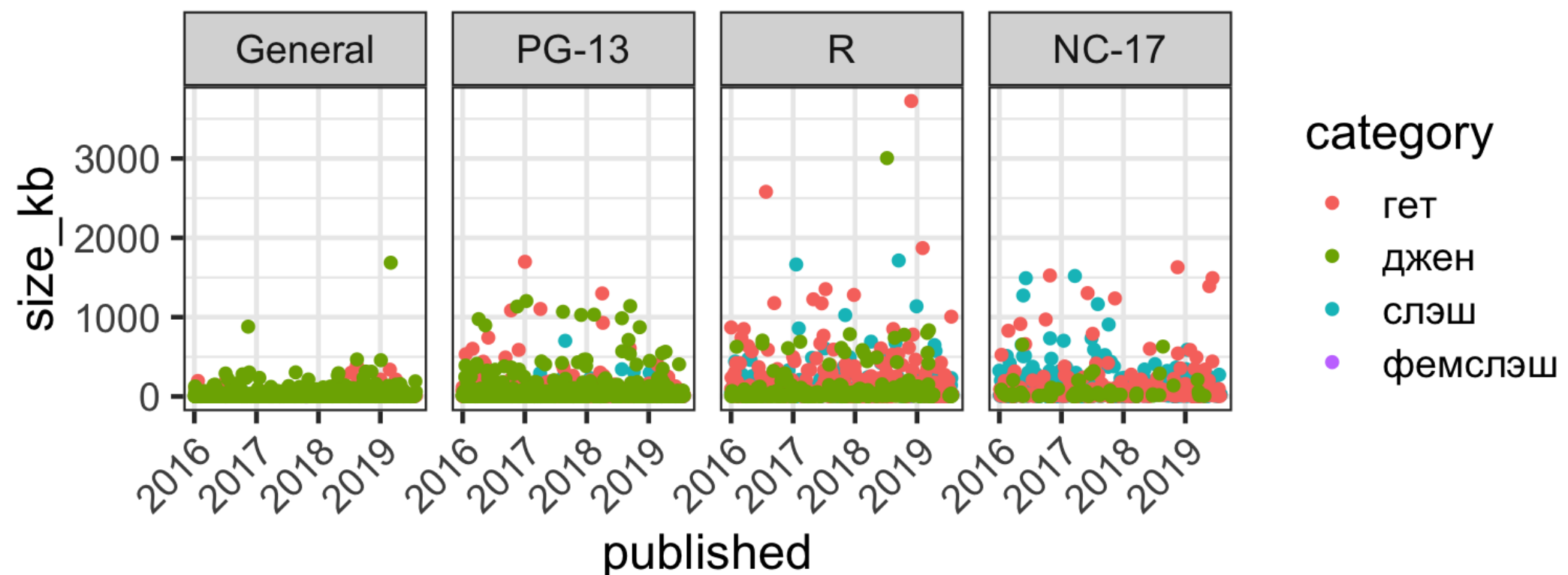


# Тема

Как повернуть надписи, сделать розовый фон и шрифт Monotype Corsiva?

Указать точечные изменения темы:

```
ggplot(...) + theme(  
  axis.text.x = element_text(angle = 45, hjust=1)  
)
```



# Тема

Как сделать одинаковый стиль у всех графиков?

Вызвать `theme_set` в начале скрипта.

# Академические графики

- Ящик с усами: `geom_boxplot()`, по *x* должен быть фактор

Покажите распределение размера фанфика для разных рейтингов.

- Виолончель: `geom_violin()`, по *x* должен быть фактор

Покажите распределение размера фанфика для разных рейтингов.

- Квантильные графики: `geom_qq()`