RUDN university

5100

Программирование в среде R

Шевцов Василий Викторович,
директор ДИТ РУДН, shevtsov_vv@rudn.university

Визуализация 3D-распределений

ggplot2 & 3D

В пакете ggplot2 3D-графики (т. е. графики с тремя координатными осями) как таковые не реализованы.
Имеется возможность строить графики с изолиниями, отражающими изменение той или иной количественной переменной в зависимости от двух других количественных переменных

Контуры плотности вероятности

geiom_dencity2d()

Аргументы

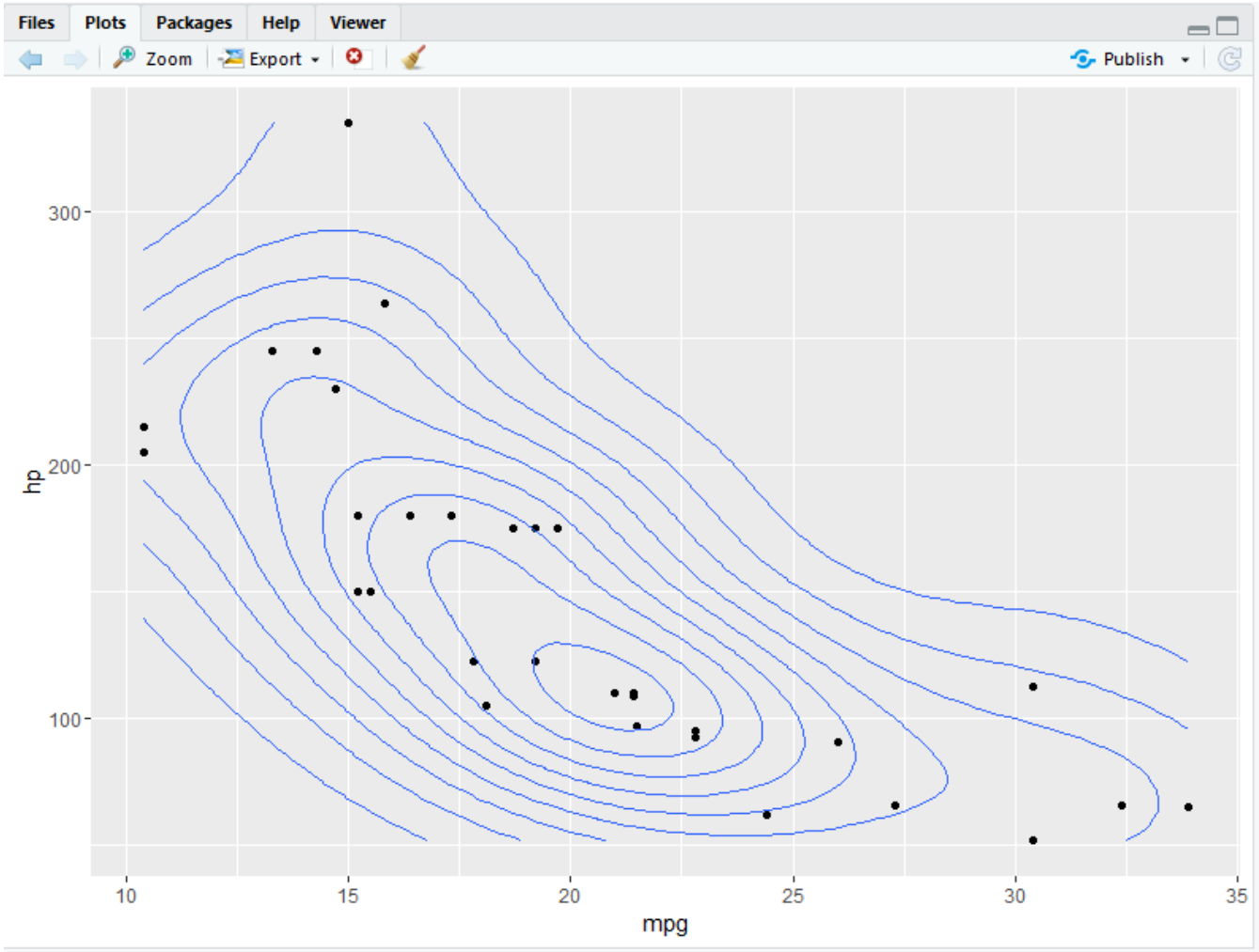
- contour — логический аргумент. Значение contour = TRUE (задано по умолчанию) включает отображение контуров плотности вероятности на графике.
- n - число, определяющее гладкость контурных линий.

Эстетические атрибуты

- x и y - переменные X и Y соответственно.
- alpha - степень прозрачности цвета.
- colour - цвет линии.
- linetype - тип линии.
- size - толщина линии.

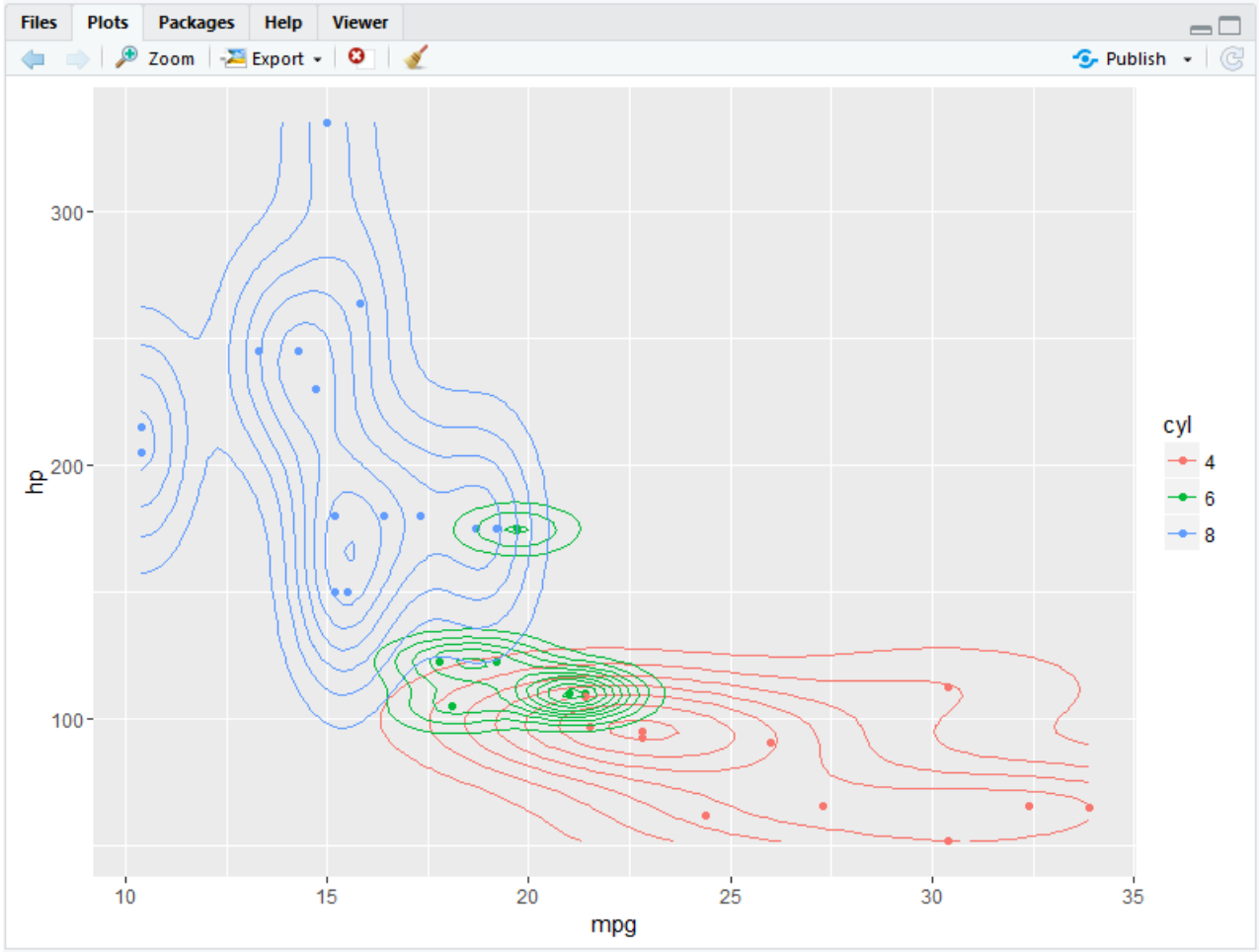
Контуры плотности вероятности

```
library(ggplot2)
df <- mtcars
p <- ggplot(data = df, aes(x=mpg,y=hp))+
  geom_point()
p + stat_density2d()
```



Контуры плотности вероятности

```
df$cyl <- factor(df$cyl)
p <- ggplot(data = df, aes(x=mpg,y=hp))+
  geom_point(aes(colour=cyl))
p + stat_density2d(aes(colour=cyl))
```



Изолинии

Изолиния, или контурная линия, функции двух переменных представляет собой линию, вдоль которой значение этой функции постоянно. Графики с использованием изолиний можно встретить во многих научных областях - в картографии, метеорологии, океанографии, экологии и т. д. В зависимости от контекста изолинии могут иметь разные названия (изобаты, изобары, изотермы, изоклины) В пакете ggplot2 для построения графиков с изолиниями служит функция geom_contour().

Аргументы

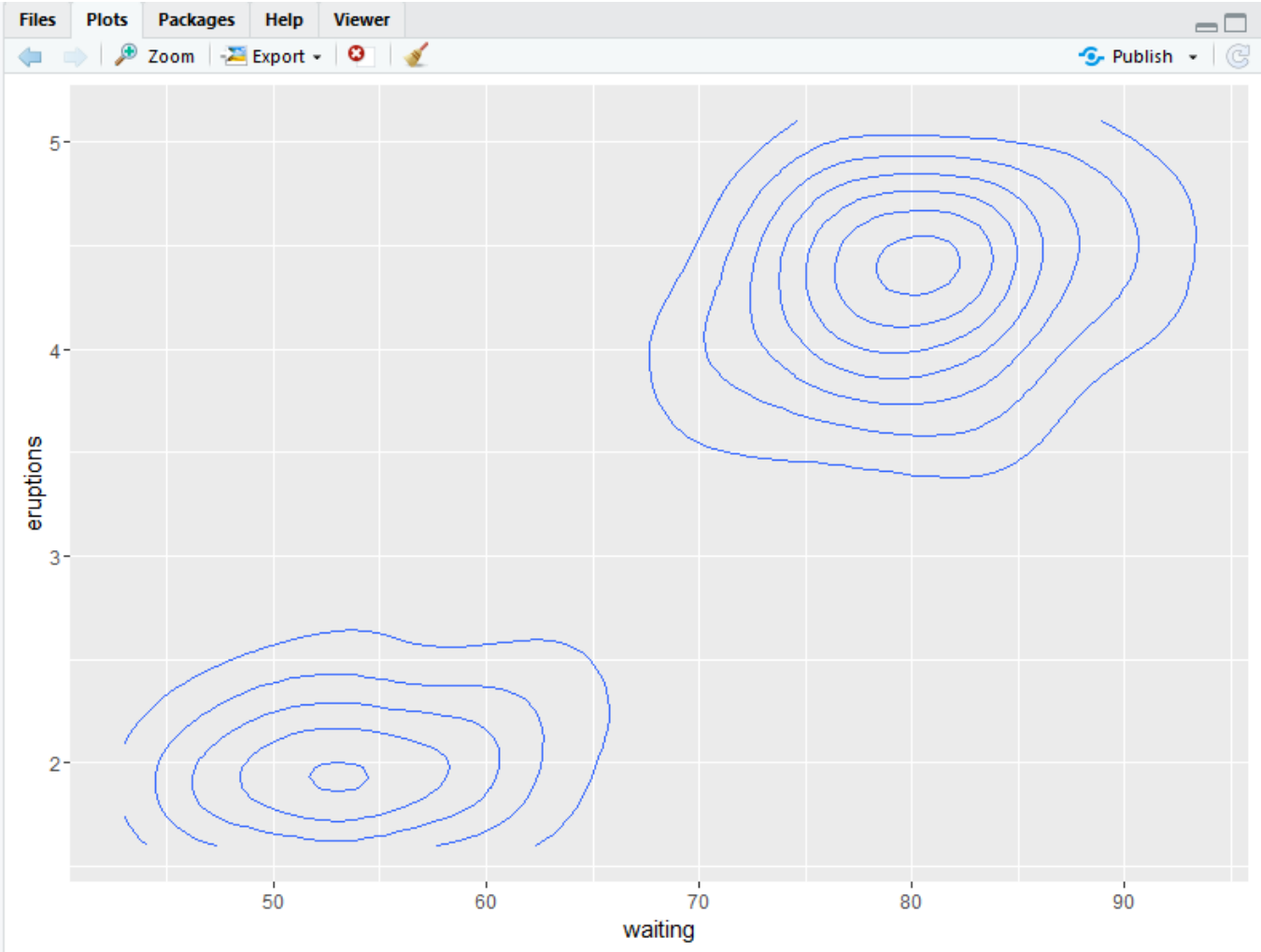
- bins -- задает размер классового промежутка, используемого для объединения нескольких соседних изолиний в один класс.

Эстетические атрибуты

- x и y — переменные X и Y соответственно.
- z — переменная, являющаяся функцией от x и y.
- alpha — степень прозрачности цвета.
- colour — цвет линии.
- linetype — тип линии.
- size — толщина линии.

Изолинии

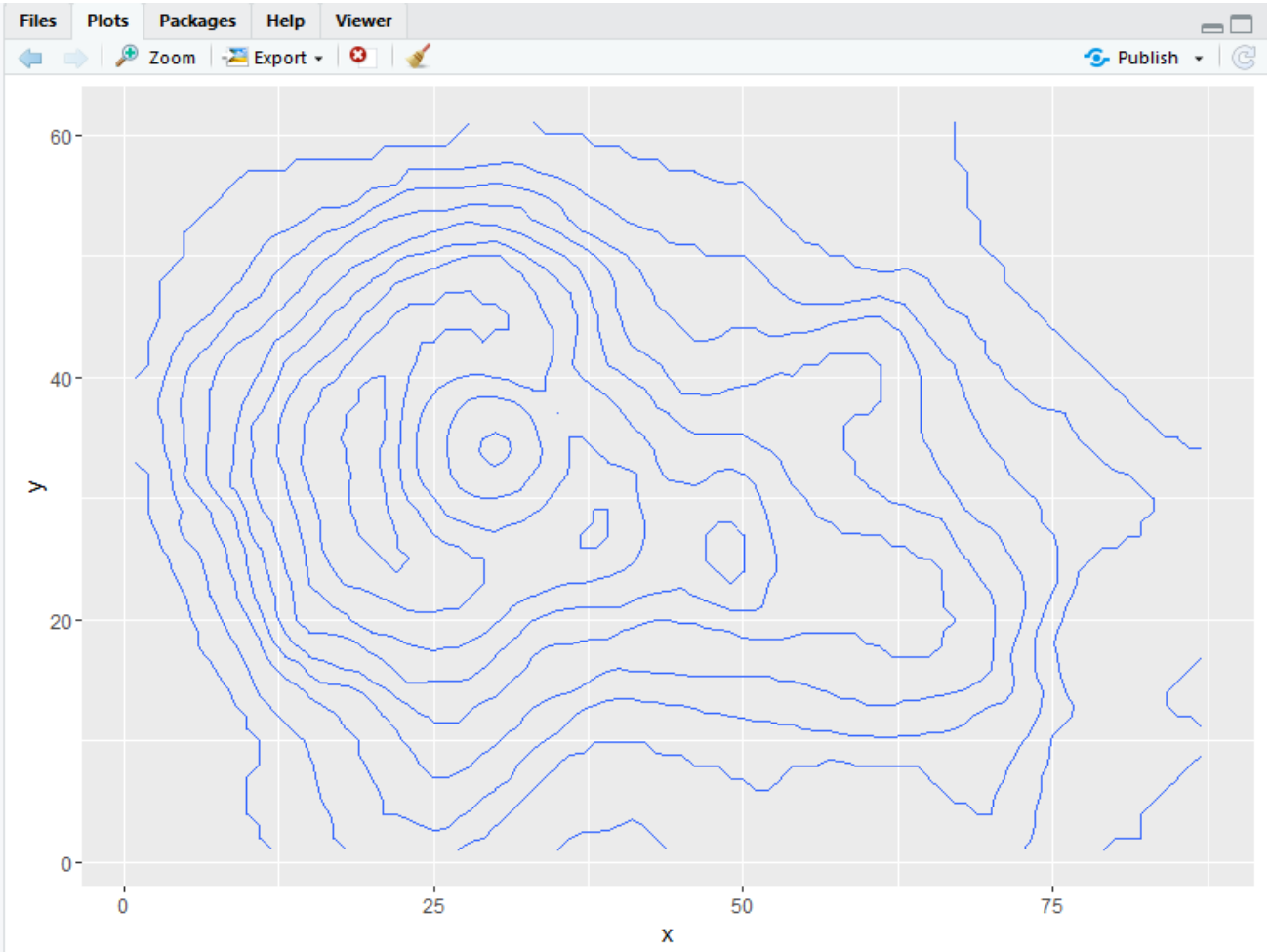
```
df6 <- faithful
v <- ggplot(faithfuld, aes(waiting, eruptions, z = density))
v + geom_contour()
```



RUDN university

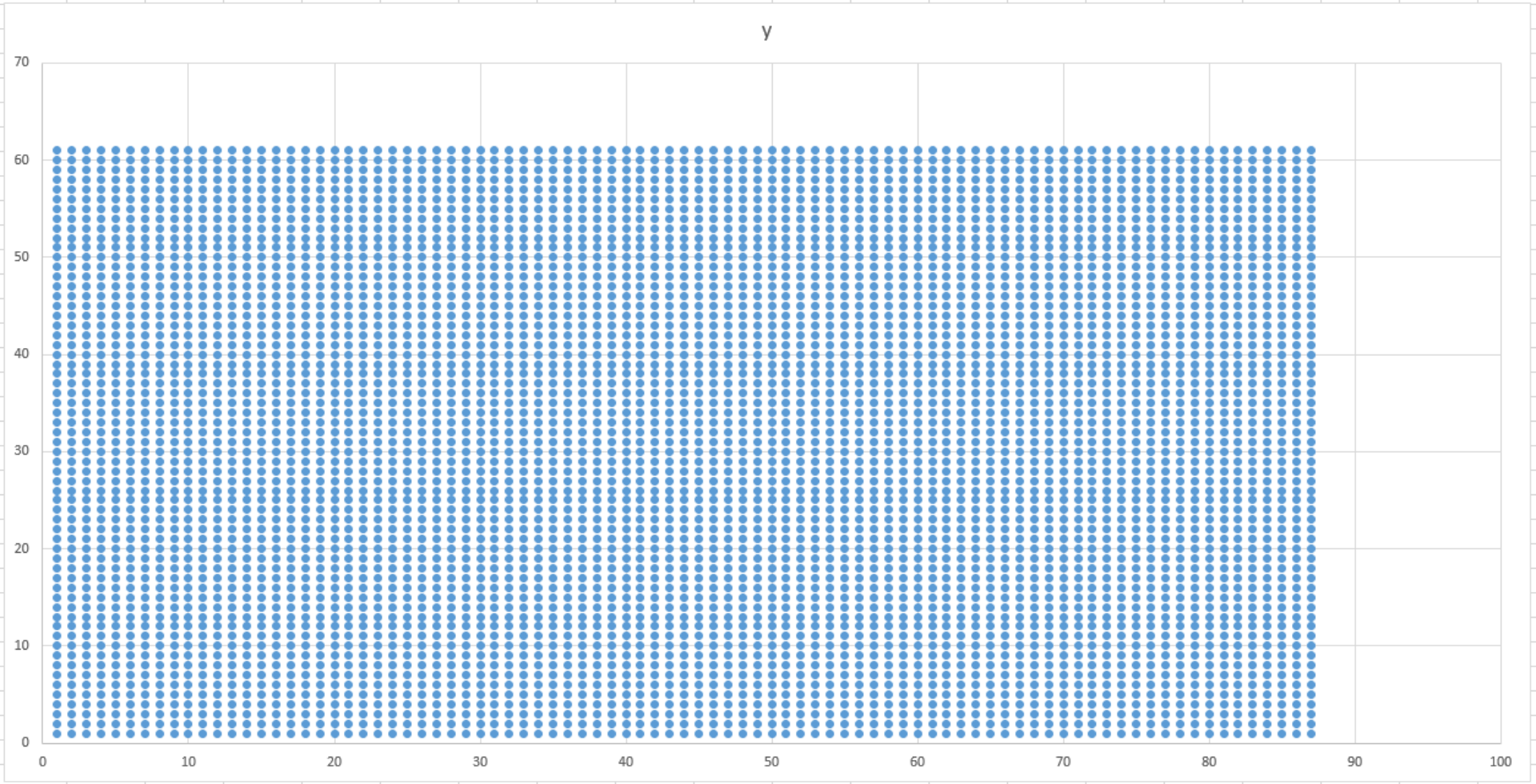
Изолинии

```
df5 <- reshape2::melt(volcano)
names(df5) <- c("x","y","z")
p <- ggplot(df5,aes(x=x,y=y,z=z))
p+geom_contour()
```



Изолинии. Правила построения

- {x,y} - представляет собой координатную сетку
- z – значения, образующие группы (в некотором смысле это фактор)



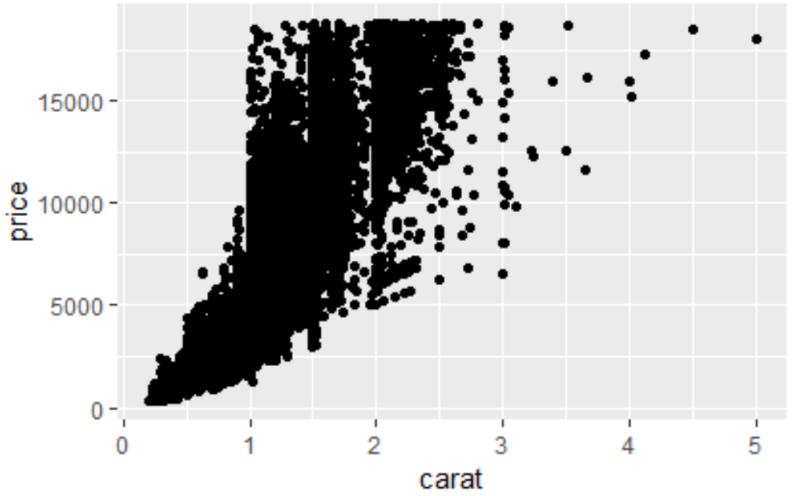
Сотовые диаграммы

Сотовые диаграммы

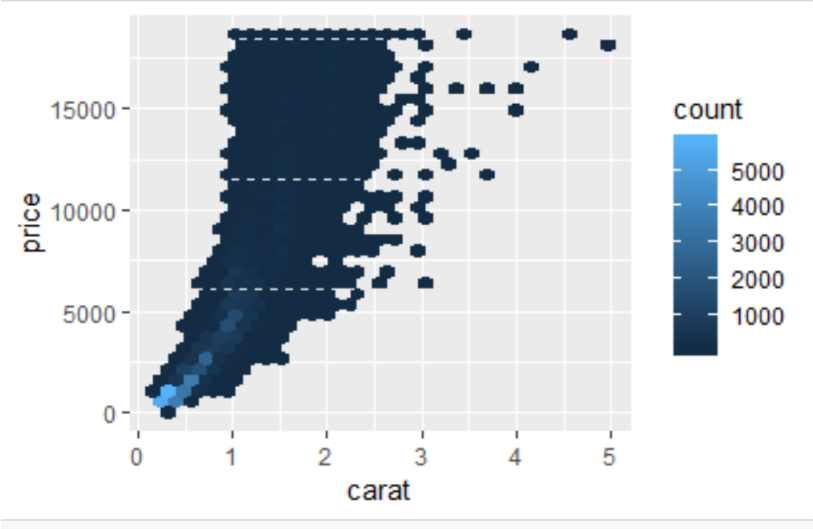
При работе с данными большого объема визуализация зависимости между двумя переменными на диаграмме рассеяния неизбежно сопровождается «наползанием» точек друг на друга, что затрудняет выявления характера этой зависимости. Одним из возможных способов решения этой проблемы является использование «сотовых диаграмм» (англ. hexagon plots). На такой диаграмме координатная плоскость разбивается на **гексагоны**, которые закрашиваются цветом в соответствии с градиентом плотности попавших в них точек (в англоязычной литературе этот процесс имеет название «hexagon binning»). При этом гексагоны с нулевым количеством попавших в них точек на графике не изображаются. Таким образом, сотовые диаграммы представляют собой вариант двухмерной гистограммы.

Сотовые диаграммы

```
p <- ggplot(data=diamonds,
aes(x=carat,y=price))
p+geom_point()
```

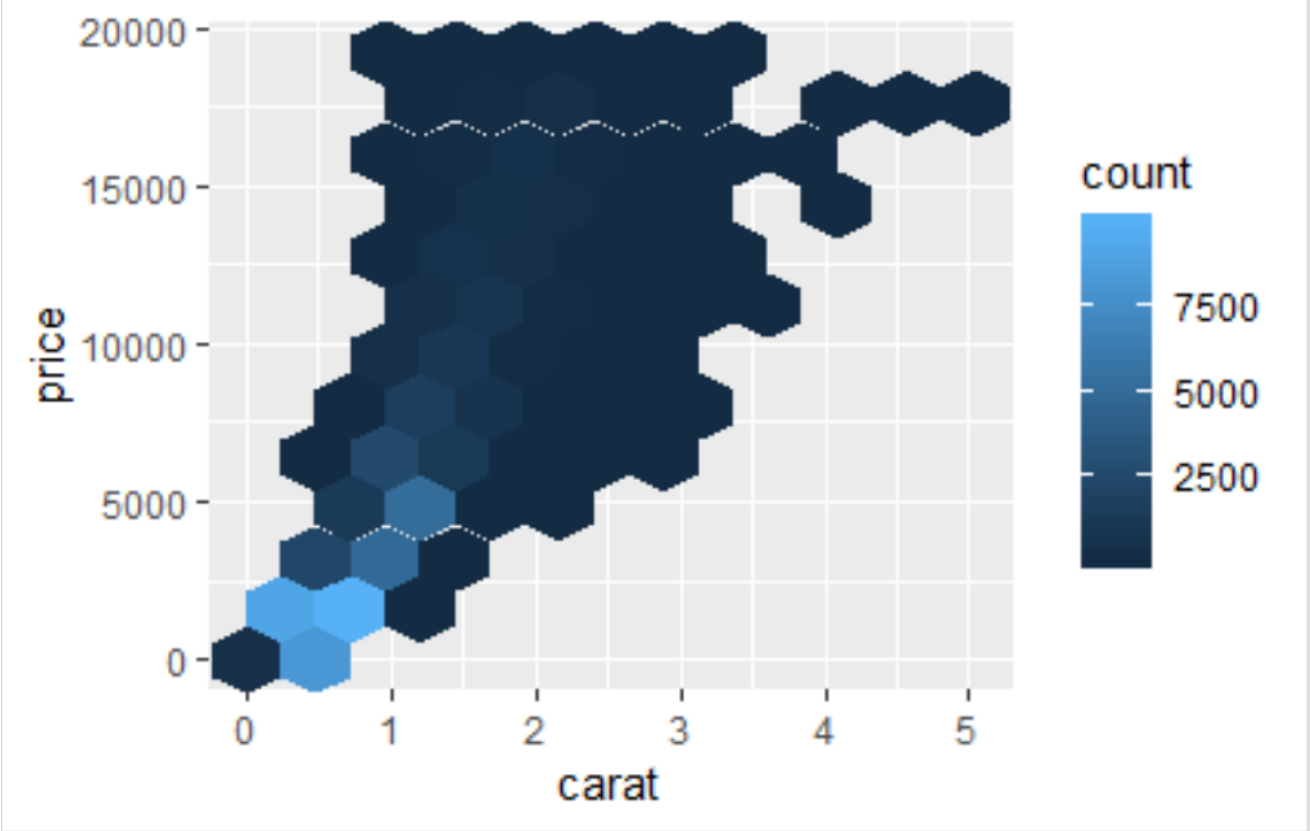



```
p <- ggplot(data=diamonds,
aes(x=carat,y=price))
p+geom_hex()
```



Сотовые диаграммы

```
p <- ggplot(data=diamonds,
aes(x=carat,y=price))
p+geom_hex(bins=10)
```



Диаграммы диапазонов

Диаграммы диапазонов

geom_linerange(),
geom_pointrange(),
geom_errorbar(),
geom_crossbar()

Диаграммы диапазонов служат для визуализации интервалов значений анализируемых переменных (минимум и максимум, нижний и верхний квартили и т.п.)

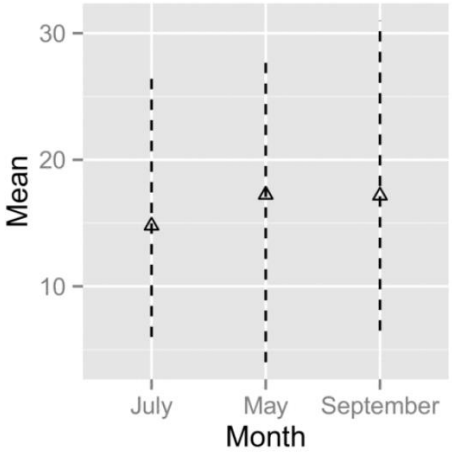
x — переменная X (функция geom_crossbar () требует также обязательного указания У-координат для линий, изображающих выбранную пользователем меру центральной тенденции);

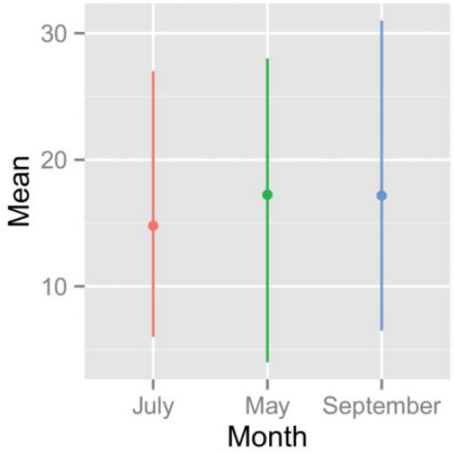
- ymax — вектор с У-координатами, соответствующими верхним границам интервалов;
- ymin — вектор с У-координатами, соответствующими нижним границам интервалов;
- alpha — степень прозрачности цвета;
- colour — цвет линии;
- linetype — тип линии;

Диаграммы диапазонов

- geom_linerange() - создает слой, на котором интервалы значений количественных переменных представлены в виде вертикальных или горизонтальных линий;
- geom_pointrange() — создает слой, на котором интервалы значений количественных переменных представлены в виде вертикальных или горизонтальных линий с точкой посередине (например, среднее значение ± стандартное отклонение);
- geom_errorbar () - создает слой с вертикальными отрезками, обозначающими стандартные ошибки, доверительные интервалы и другие подобные им статистические показатели. Эти отрезки по обоим концам ограничены более короткими перпендикулярными отрезками. Обычно слой, созданный при помощи geom_errorbar(), добавляют к уже существующему ggplot-графику с точками, символизирующими, например, средние значения какой-либо переменной. Для создания слоя с аналогичными горизонтальными отрезками служит функция geom_errorbarh();
- geom_crossbar () - создает слой, на котором небольшие горизонтальные линии, соответствующие какой-либо мере центральной тенденции, окаймлены прямоугольниками, длина которых соответствуют какой-либо мере разброса данных.

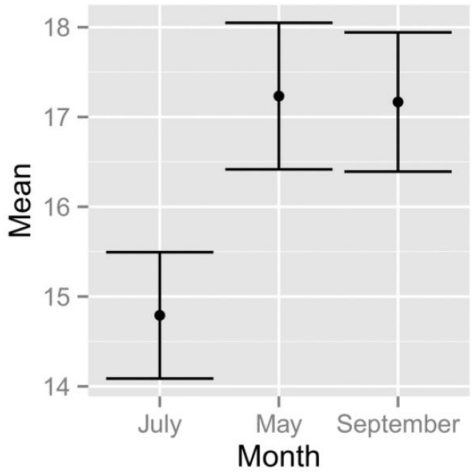
Диаграммы диапазонов

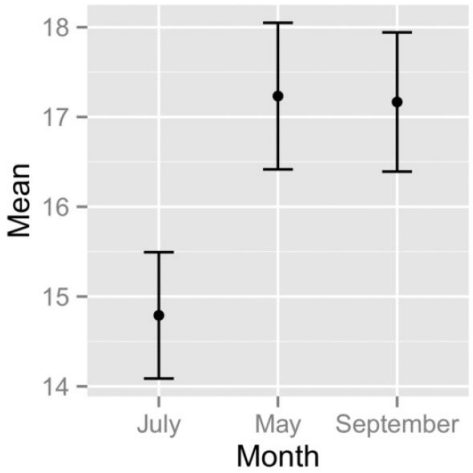




geom_linerange()
geom_pointrange()

geom_point()
geom_errorbar()

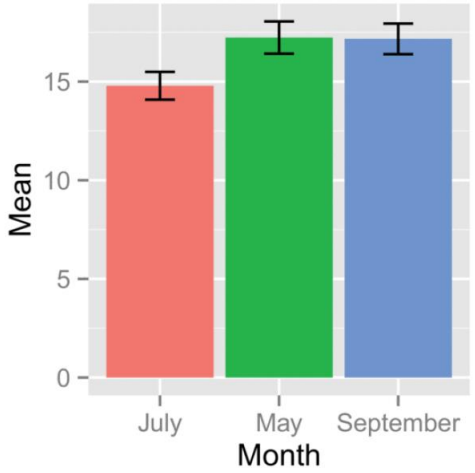


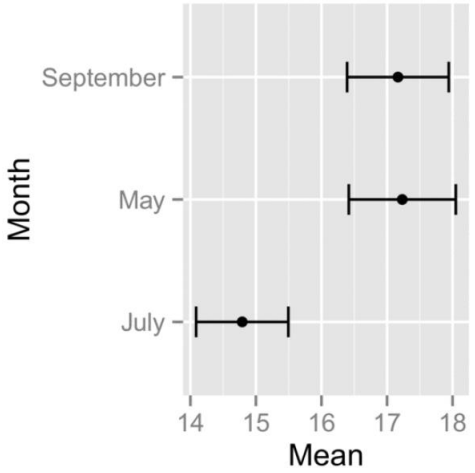


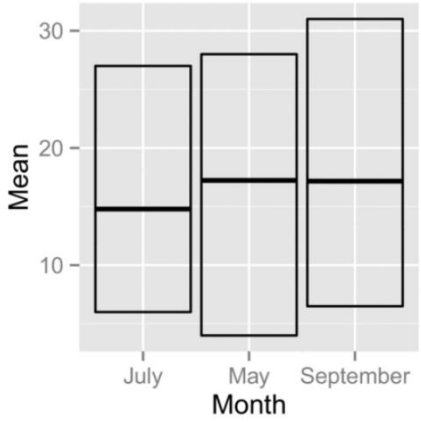
RUDN university

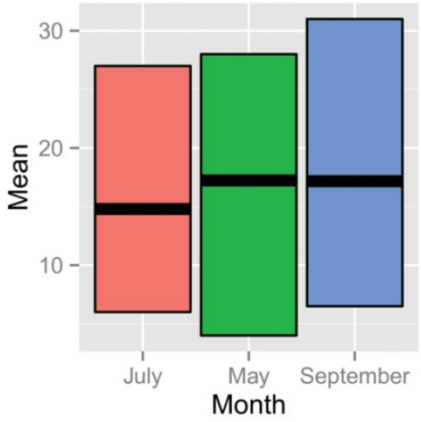
Диаграммы диапазонов

geom_bar()
geom_errorbar()









geom_crossbar()

Диаграммы размахов

Диаграммы размахов

Диаграммы размахов позволяют очень полно представить на одном графике сводную статистическую информацию одновременно для нескольких групп данных, выделенных в соответствии с уровнями той или иной качественной переменной (или переменных). Для каждой группы данных изображаются медиана, нижний и верхний квартили, интервал («усы»), в который попадает подавляющее большинство наблюдений, а также наблюдения-выбросы, оказавшиеся за пределами этого интервала. По бокам «ящиков» могут также присутствовать «насечки» (англ. notches) шириной $1{,}5 \times IQR/\sqrt{n}$. Ширина насечек примерно соответствует 95%-ным доверительным интервалам медиан, что позволят выполнять быструю визуальную оценку различий между группами данных (если эти интервалы перекрываются, то делается заключение об отсутствии различий между генеральными медианами соответствующих групп).

Диаграммы размахов

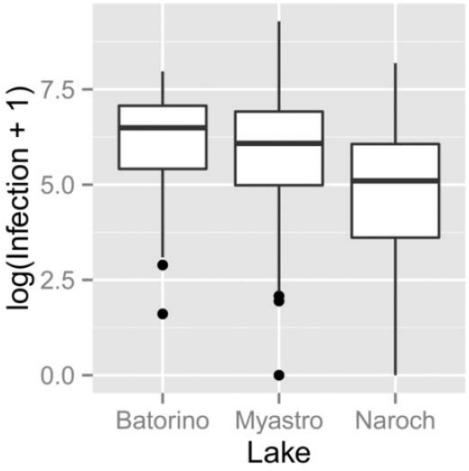
Аргументы

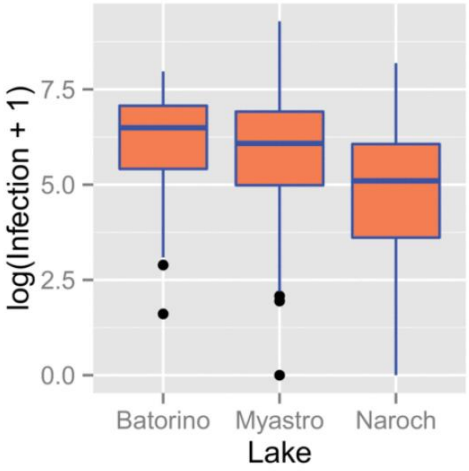
- notch — логический аргумент, включающий отображение «насечек» (по умолчанию notch = FALSE).
- notchwidth — задаст глубину насечки относительно ширины «ящика» (по умолчанию notch = 0.5).

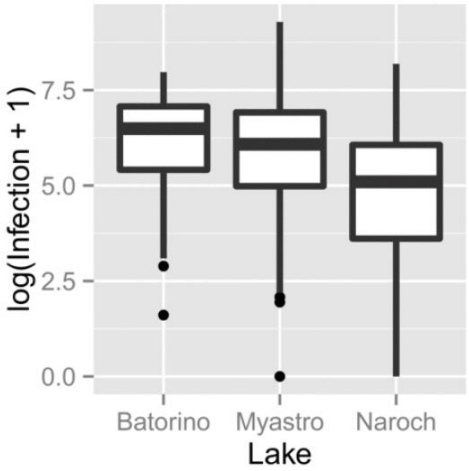
Эстетические атрибуты

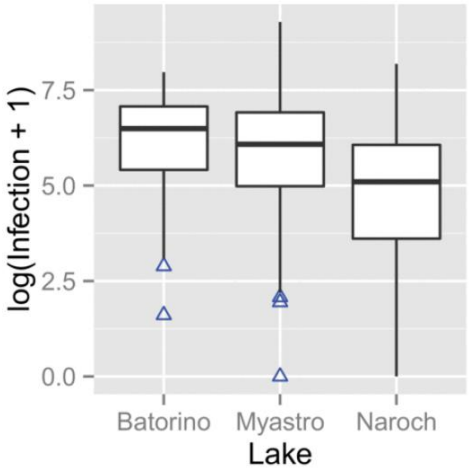
- lower и upper — нижняя и верхняя F-координаты прямоугольников. По умолчанию это квартили, однако при помощи аргументов lower и upper можно задать и другие процентили.
- middle - F-координаты линий, изображающих меру центральной тенденции в каждой группе данных (по умолчанию это медиана).
- x — переменная X.
- ymin и ymax — нижняя и верхняя F-координаты отрезков, отходящих от прямоугольников.
- outlier. color — цвет точек, обозначающих выбросы.
- outlier. shape — форма точек, обозначающих выбросы.
- outlier, size — размер точек, обозначающих выбросы.
- alpha — степень прозрачности цвета.
- colour — цвет линий.
- linetype — тип линий.

Диаграммы размахов

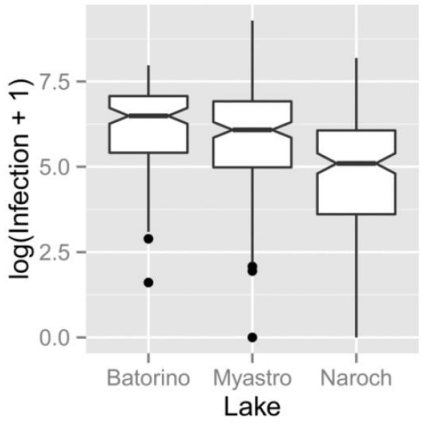


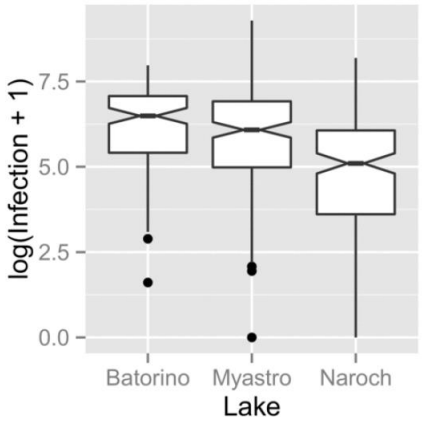






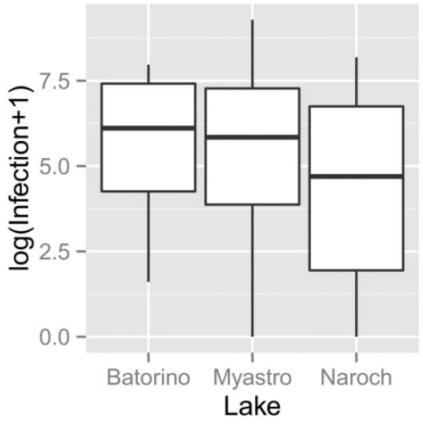
Диаграммы размахов

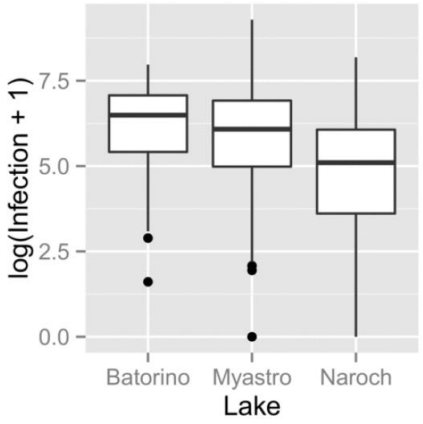




Диаграммы размахов с «насечками».

Диаграммы размахов измененными границами ящиков



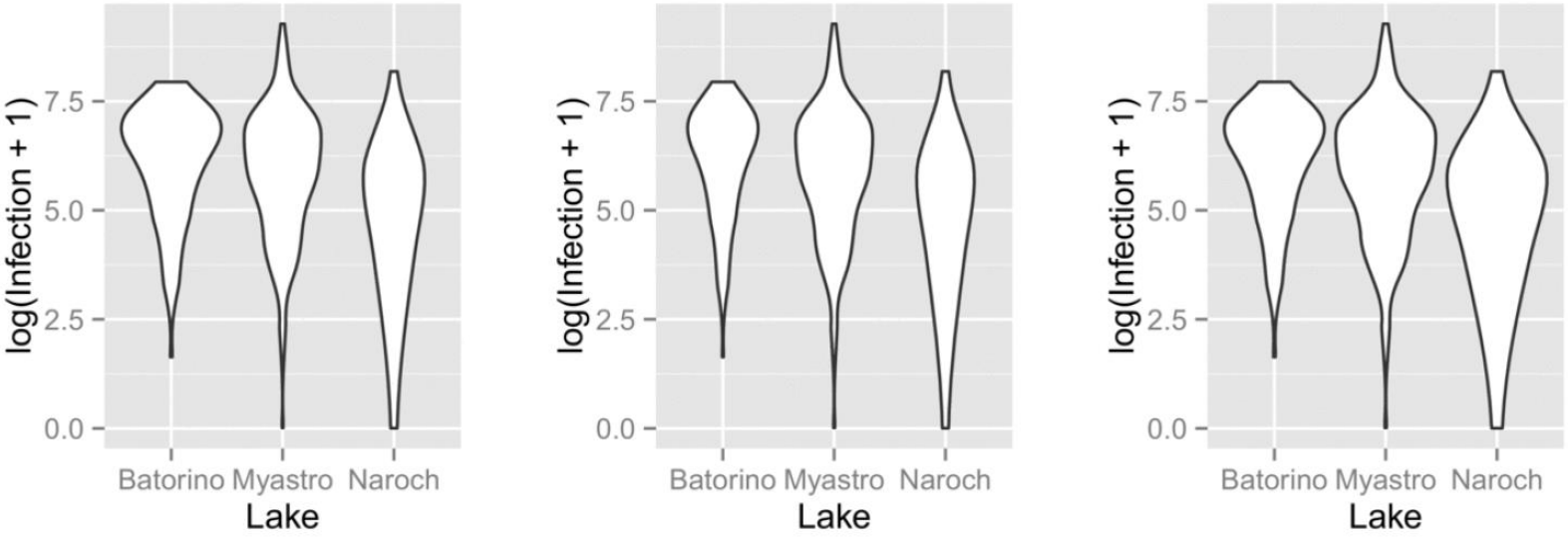


Скрипичные диаграммы

Диаграммы размахов позволяют представить информацию о нескольких статистических параметрах анализируемых групп данных. Однако можно сделать график еще более информативным при помощи «скрипичных диаграмм» (англ. «violin plots»)i , которые объединяют в себе идеи диаграмм размахов и кривых плотности вероятности.
Суть достаточно проста: продольные края «ящиков» на диаграмме размахов замещаются кривыми плотности вероятности. В итоге получаются симметричные фигуры, чьи очертания напоминают очертания скрипки — отсюда название этого типа диаграмм.

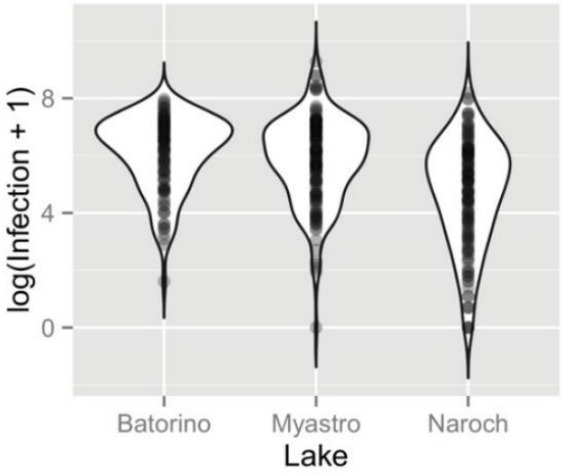
Специфичный аргумент **scale**, который определяет форму «скрипок». При **scale = "area"** (значение по умолчанию) наблюдения в каждой группе нормализуются таким образом, чтобы все «скрипки» имели одинаковую площадь. При **scale = "count"** площади «скрипок» пропорциональны объемам наблюдений в соответствующих группах. Наконец, при **scale = "width"** максимальная ширина у всех «скрипок» будет одинакова.

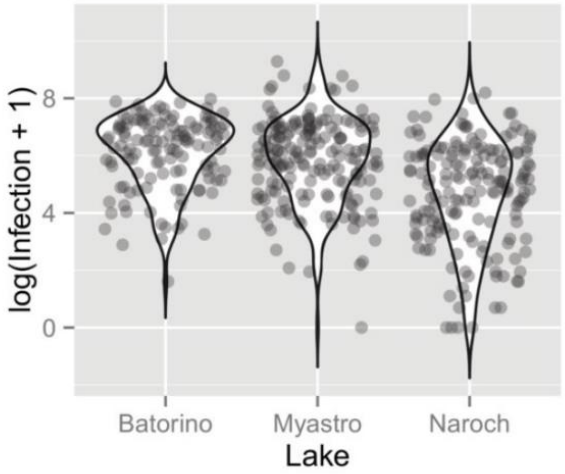
Скрипичные диаграммы



Слева: скрипичная диаграмма, созданная с использованием автоматических настроек функции geom_violin() (площадь всех фигур на диаграмме одинакова). В центре: площадь «скрипок» пропорциональна объемам наблюдений в соответствующих группах (**scale = "count"**). Справа: максимальная ширина у всех «скрипок» одинакова (**scale = "width"**)

Скрипичные диаграммы

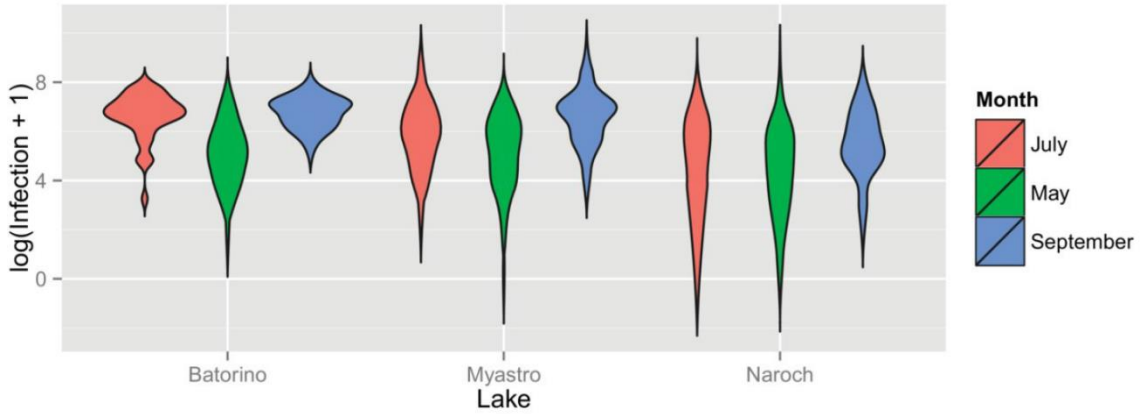




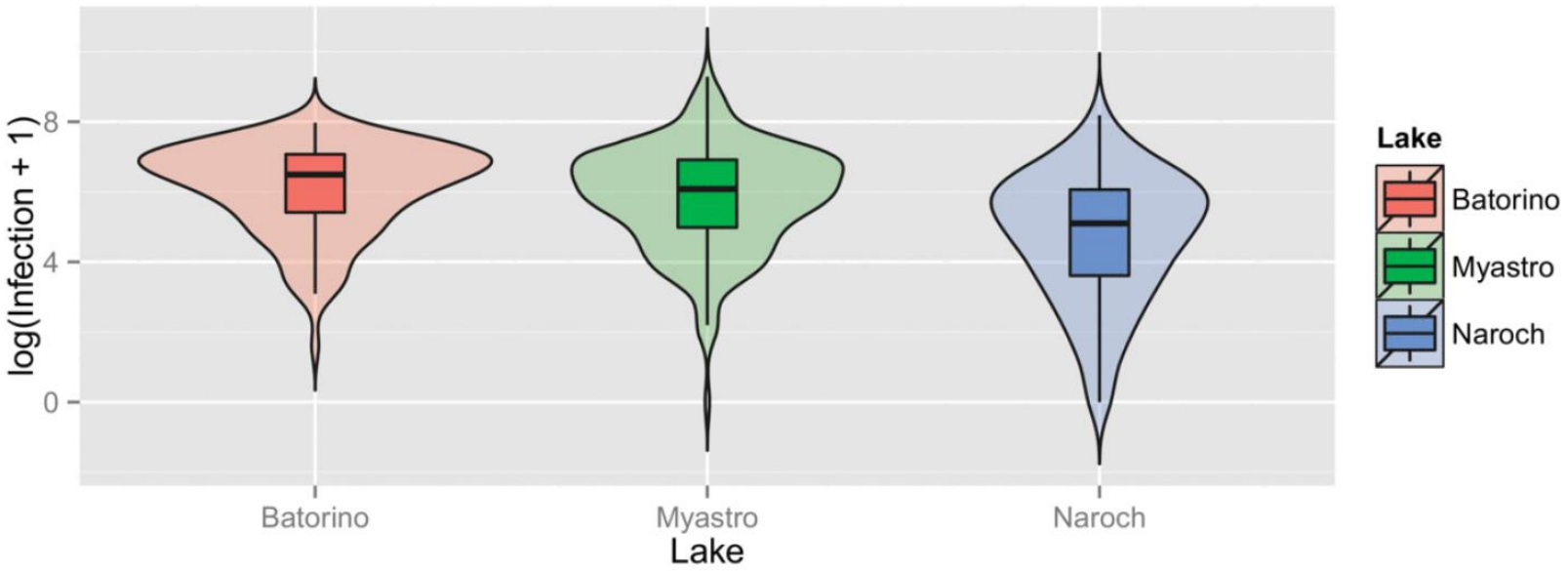
+geom_point()

+geom_jitter()

+geom_violin(aes(fill=Month))



Скрипичные диаграммы



+geom_violin(aes(fill=Lake))
+geom_boxplot(aes(fill=Lake))

Линии тренда

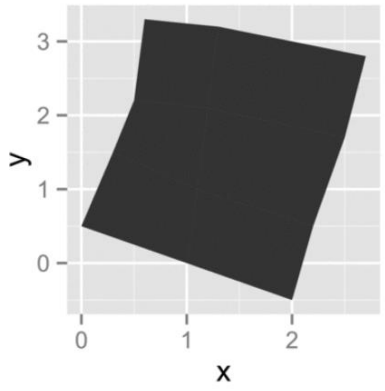
geom_smooth()

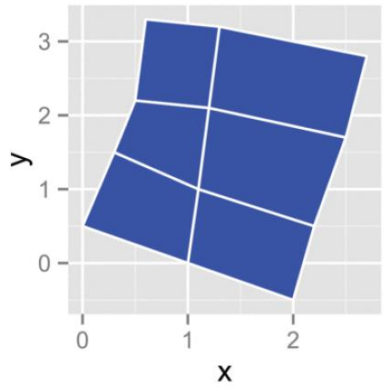
- method = "loess": метод локально взвешенной полиномиальной регрессии10. Степень сглаживания определяется параметром span, который варьирует в пределах от 0 (наиболее низкая степень сглаживания) до 1 (максимально гладкая кривая). При объеме наблюдений n > 1000 loess начинает работать медленно, и поэтому вместо него автоматически включается другой метод сглаживания;
- method = "gam": сглаживание на основе обобщенной аддитивной модели. Выполняется при помощи функции gam () из базового R-пакета mgcv (этот пакет следует предварительно загрузить в рабочую среду программы командой library (mgcv);
- method = "lm": сглаживание на основе линейной модели. По умолчанию подгоняется прямая линия регрессии. Кроме того, имеется возможность подогнать полиномиальную регрессионную модель любой степени
- method = "rim": сглаживание с использованием робастной линейной модели. Работает сходным с "lm" образом, однако использует алгоритм, благодаря которому наблюдения-выбросы не оказывают значительного влияния на параметры получаемой линейной модели. Функция rlm() является частью базового пакета MASS, который перед использованием необходимо загрузить в рабочую среду R командой library(MASS).

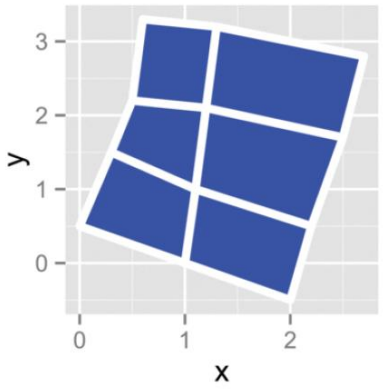
Географические карты

Многоугольники

```
# ids - идентификаторы отдельных частей многоугольника:
ids <- factor(c("1.1", "2.1", "1.2", "2.2", "1.3", "2.3"))
positions <- data.frame( id = rep(ids, each = 4),
  x = c(2, 1, 1.1, 2.2, 1, 0, 0.3, 1.1, 2.2, 1.1, 1.2, 2.5,
        1.1, 0.3, 0.5, 1.2, 2.5, 1.2, 1.3, 2.7, 1.2, 0.5,
        0.6, 1.3),
  y = c(-0.5, 0, 1, 0.5, 0, 0.5, 1.5, 1, 0.5, 1, 2.1,
        1.7, 1, 1.5, 2.2, 2.1, 1.7, 2.1, 3.2, 2.8, 2.1,
        2.2, 3.3, 3.2))
p <- ggplot(data = positions, aes(x = x, y = y, group = id))
p + geom_polygon()
p + geom_polygon(colour = "white", fill = "blue")
p + geom_polygon(colour = "white", fill = "blue", size = 1.6)
```







Географические карты

Функция geom_map() представляет собой модификацию функции geom_polygon() и предназначена для создания географических карт.

Аргументы

- map — таблица с географическими координатами, содержащая столбцы с именами x (или long), y (или lat) и region (или id).

Эстетические атрибуты

- map_id — вектор с идентификаторами территориальных единиц.
- fill — цвет заливки территориальных единиц.
- alpha — степень прозрачности цвета заливки.
- colour, linetype, size — цвет, тип и толщина контурных линий.

Географические карты

Загрузка данных, содержащих географические объекты

http://data.biogeo.ucdavis.edu

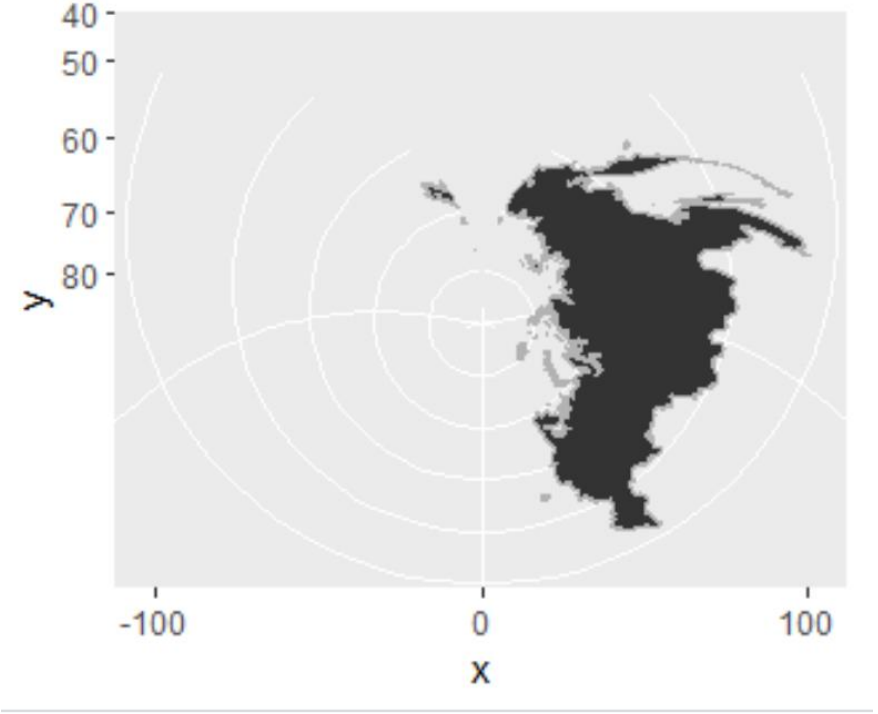
https://gadm.org

```
install.packages("sp")
library("sp")
gadm <- readRDS("gadm36_RUS_0_sp.rds")
str(gadm)
```

```
Formal class 'SpatialPolygonsDataFrame' [package "sp"] with 5 slots
  ..@ data       :'data.frame': 1 obs. of  2 variables:
  .. ..$ GID_0 : chr "RUS"
  .. ..$ NAME_0: chr "Russia"
  ..@ polygons   :List of 1
  .. ..$ :Formal class 'Polygons' [package "sp"] with 5 slots
  .. .. .. ..@ Polygons :List of 5787
  .. .. .. .. ..$ :Formal class 'Polygon' [package "sp"] with 5 slots
  .. .. .. .. .. ..@ labpt  : num [1:2] 131.4 42.7
  .. .. .. .. .. ..@ area   : num 5.15e-05
  .. .. .. .. .. ..@ hole   : logi FALSE
  .. .. .. .. .. ..@ ringDir: int 1
  .. .. .. .. .. ..@ coords : num [1:79, 1:2] 131 131 131 131 131 ...
  .. .. .. .. .. .. .. ..- attr(*, "dimnames")=List of 2
```

```
> slotNames(gadm)
[1] "data"        "polygons"    "plotOrder"   "bbox"        "proj4string"
```

Географические карты




```
install.packages("sp")
library("sp")
gadm <- readRDS("gadm36_RUS_0_sp.rds")
slotNames(gadm)

install.packages("broom")
install.packages("maps")
install.packages("rgeos")
install.packages("maptools")

library(broom)
library(maps)
library(rgeos)
library(maptools)
str(gadm)
counties <- tidy(gadm,region="NAME_0")

install.packages("mapproj")
library(mapproj)
ggplot()+geom_map(data=counties, aes(map_id=id),map=counties,color="gray70")+
  expand_limits(x=counties$long,y=counties$lat)+
  coord_map("polyconic")
```

Animations in ggplot2

library(plotly)

https://plotly.com/ggplot2/animations

Basic Example

Наряду с данными и макетом появляется понятие Frame. Он указывает на список фигур, каждая из которых будет циклически прокручиваться при создании экземпляра сюжета.

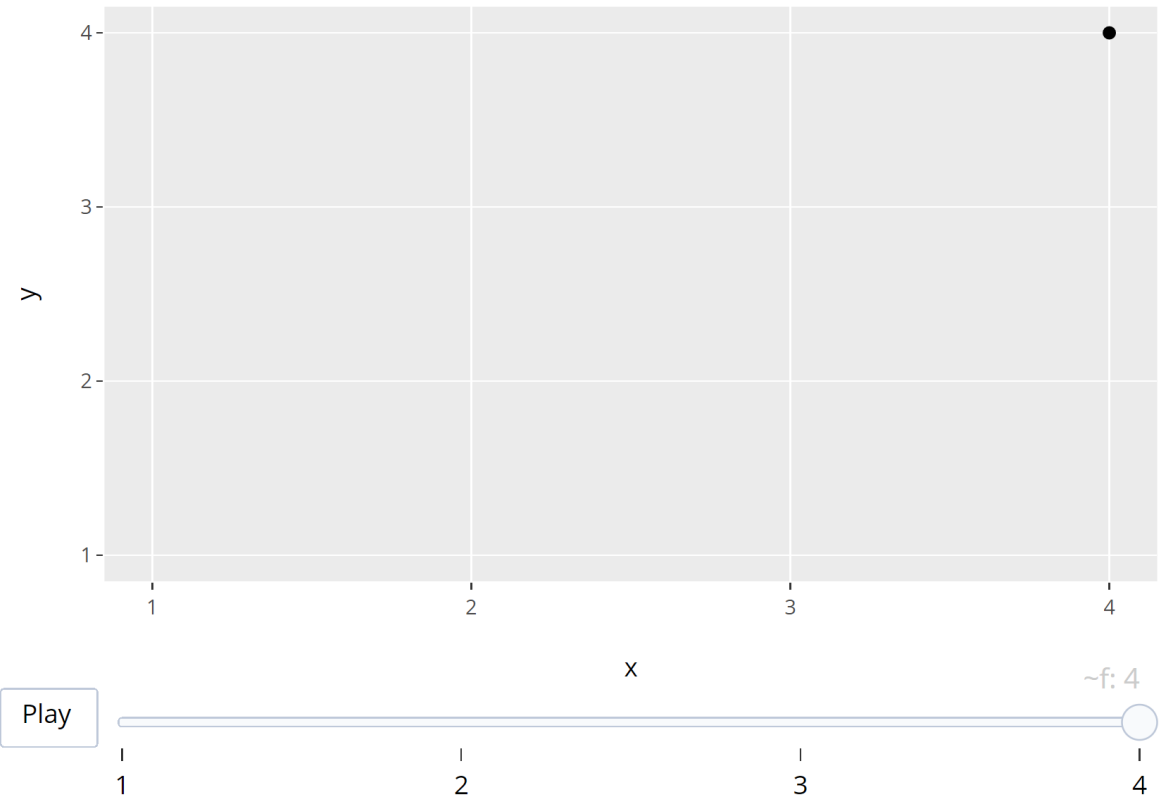
```
library(plotly)

df <- data.frame(
  x = c(1,2,3,4),
  y = c(1,2,3,4),
  f = c(1,2,3,4)
)

p <- ggplot(df, aes(x, y)) +
    geom_point(aes(frame = f))

fig <- ggplotly(p)

fig
```



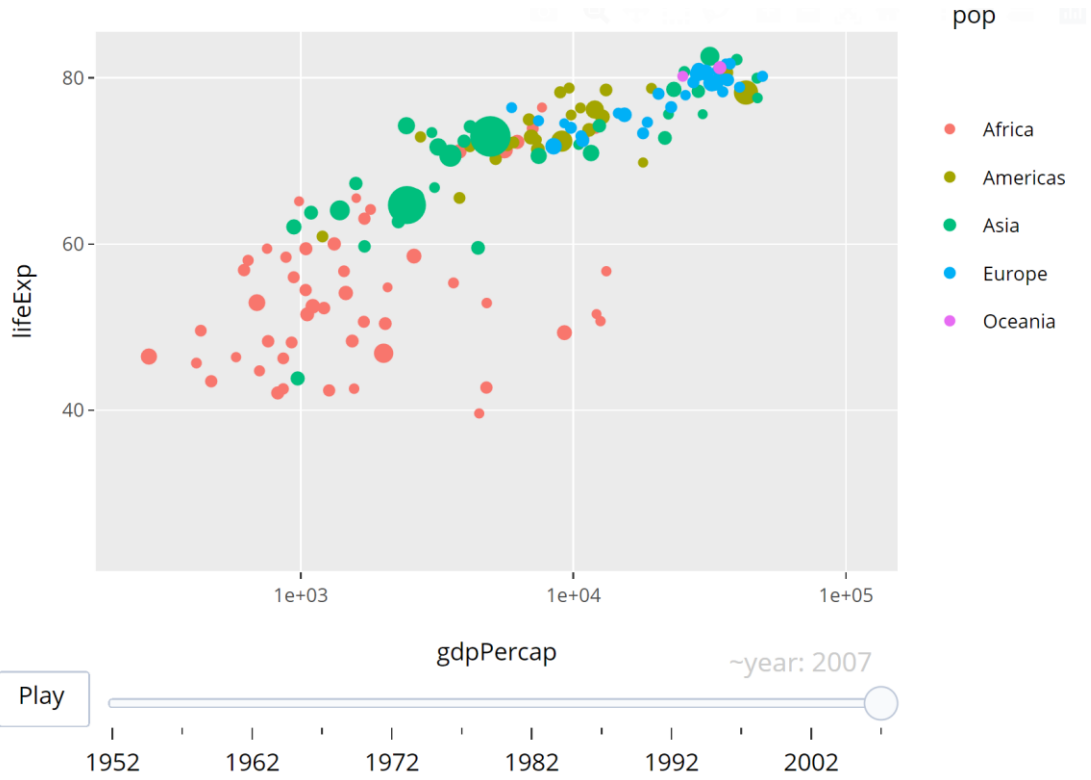
Mulitple Trace Animations

```
library(plotly)
library(gapminder)


p <- ggplot(gapminder, aes(gdpPercap, lifeExp, color = continent)) +
  geom_point(aes(size = pop, frame = year, ids = country)) +
  scale_x_log10()

fig <- ggplotly(p)

fig
```



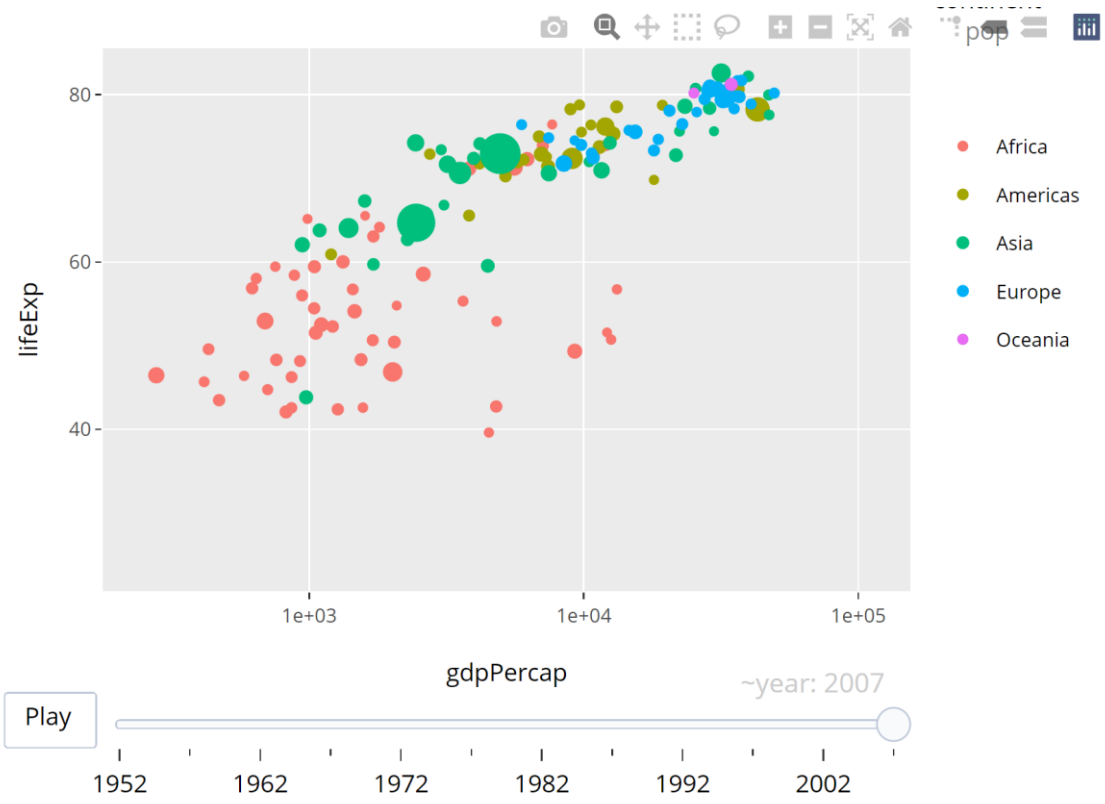
RUDN university

Add Animation Options

```
library(plotly)

fig <- fig %>%
  animation_opts(
    1000, easing = "elastic", redraw = FALSE
  )

fig
```

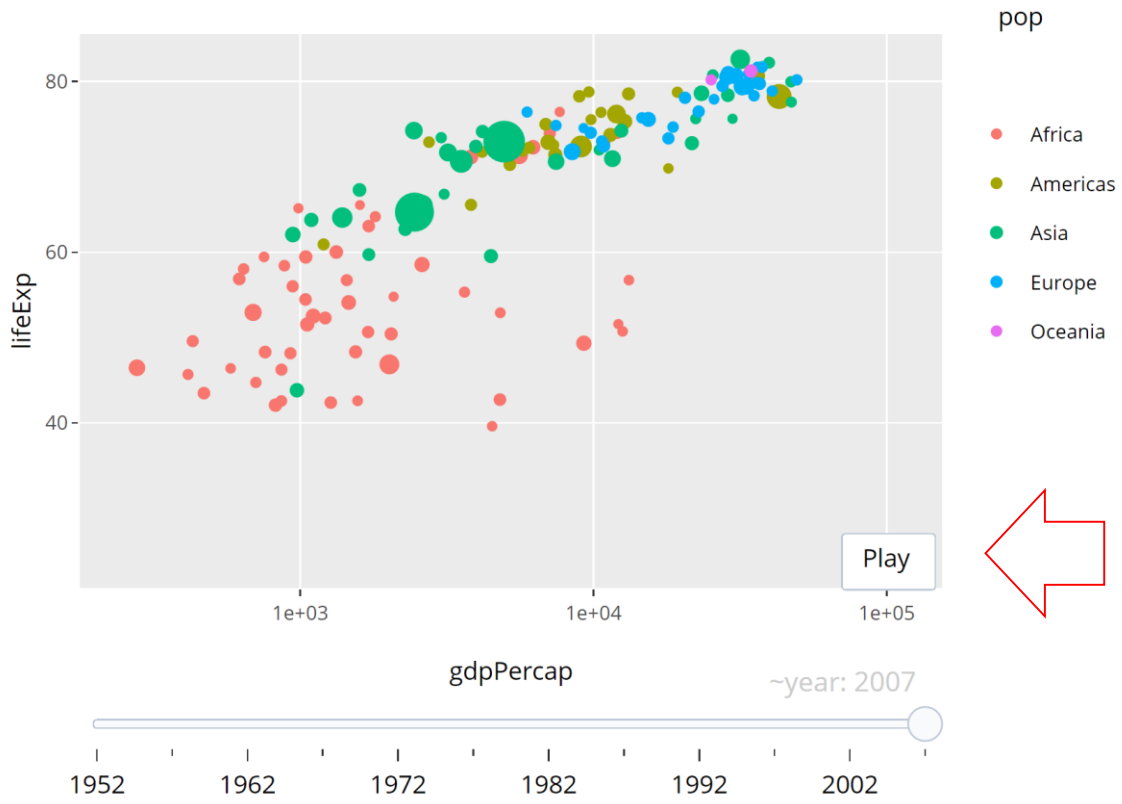


Add Button Options

```
library(plotly)

fig <- fig %>%
  animation_button(
    x = 1, xanchor = "right", y = 0, yanchor = "bottom"
  )

fig
```

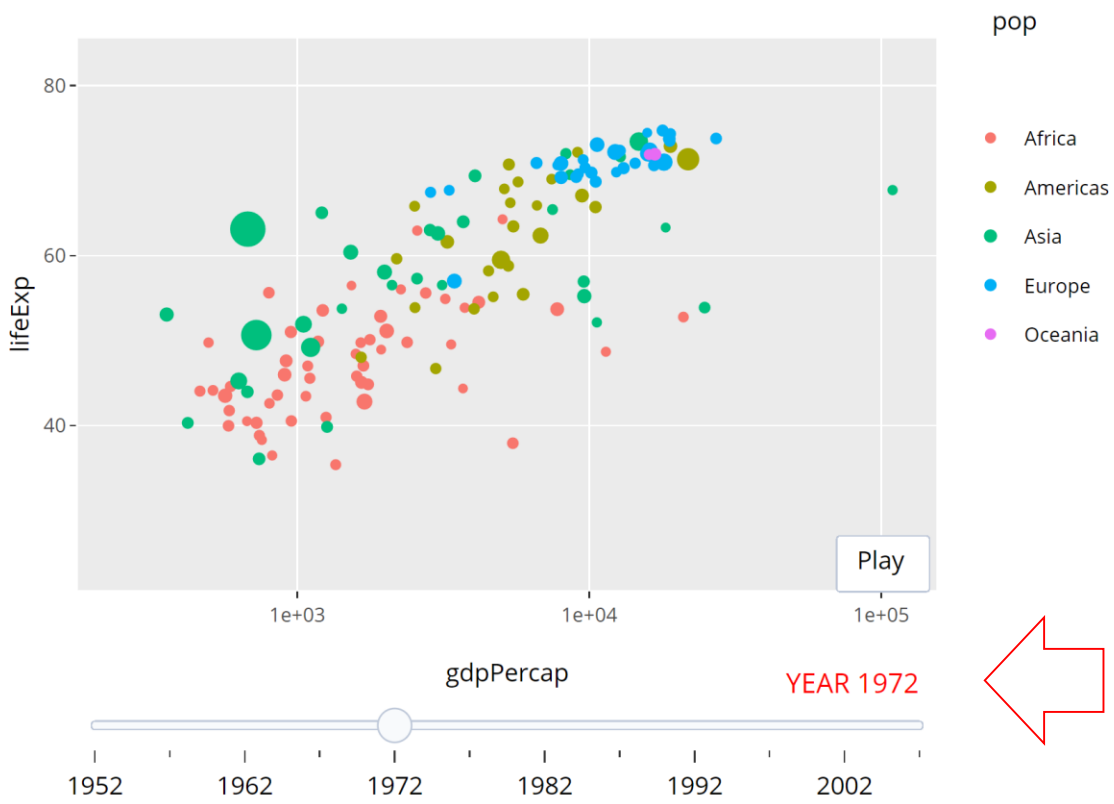


Add Slider Options

```
library(plotly)

fig <- fig %>%
  animation_slider(
    currentvalue = list(prefix = "YEAR ", font = list(color="red"))
  )

fig
```



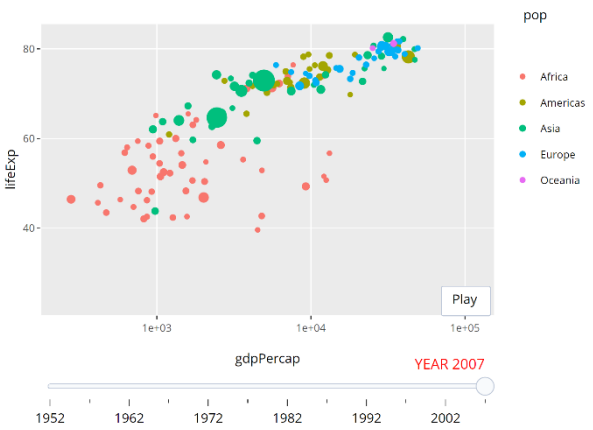
Advanced Example

```
library(plotly)
library(gapminder)

p <- ggplot(gapminder, aes(gdpPercap, lifeExp, color = continent)) +
  geom_point(aes(size = pop, frame = year, ids = country)) +
  scale_x_log10()

fig <- ggplotly(p) %>%
  animation_opts(
    1000, easing = "elastic", redraw = FALSE
  ) %>%
  animation_button(
    x = 1, xanchor = "right", y = 0, yanchor = "bottom"
  ) %>%
  animation_slider(
    currentvalue = list(prefix = "YEAR ", font = list(color="red"))
  )

fig
```



Ссылки

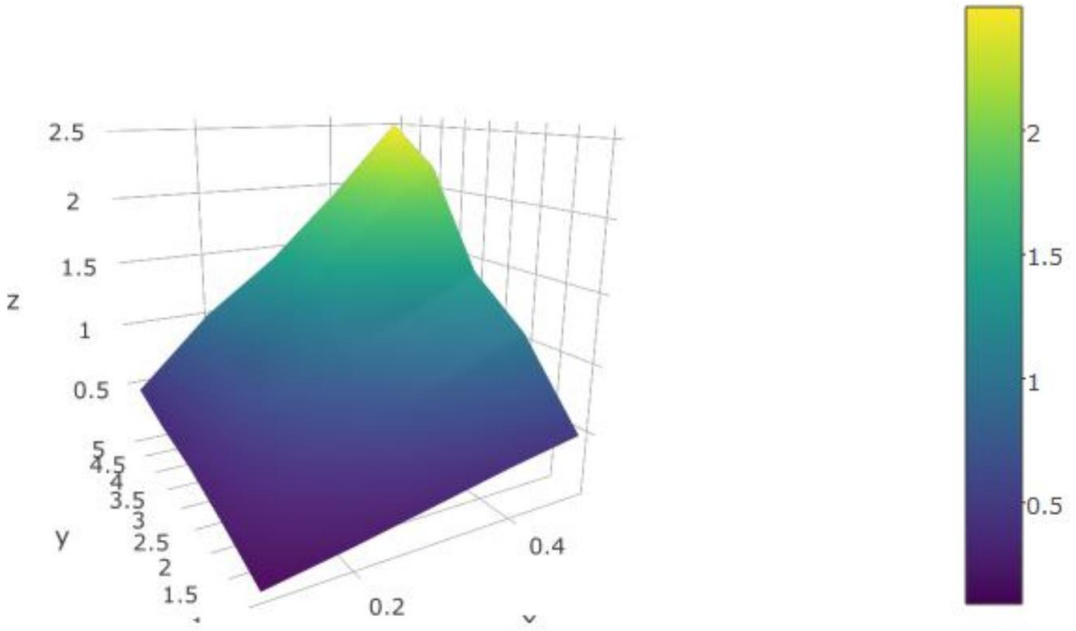
https://gist.github.com/expersso/944f3d4aad15f71b192fff254d4ac5b9

http://curleylab.psych.columbia.edu/nba.html

library(plotly) + 3D

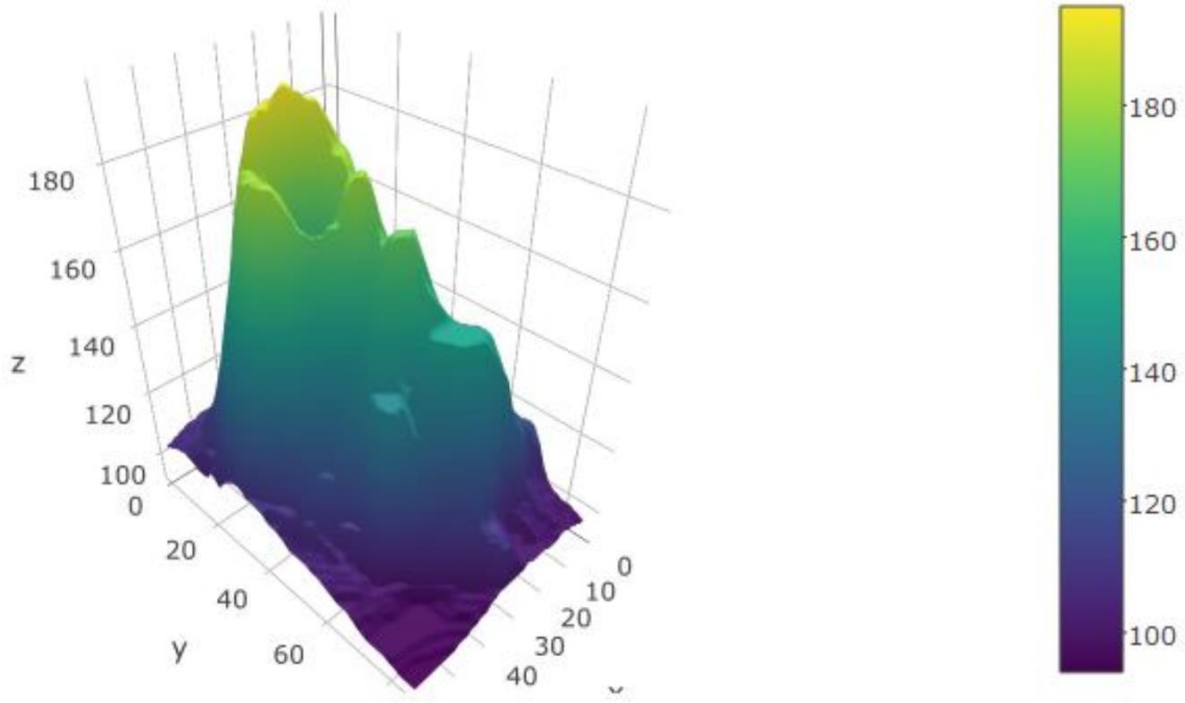
```
x <- 1:5/10
y <- 1:5
z <- x %o% y
z <- z + .2*z*runif(25) - .1*z

library(plotly)
plot_ly(x=x,y=y,z=z, type="surface")
```

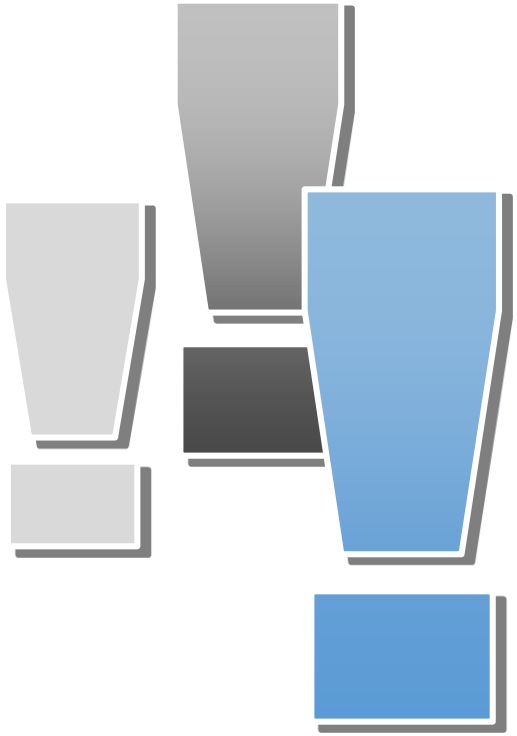


library(plotly) + 3D

```
library(plotly)
plot_ly(z=volcano, type="surface")
```



Спасибо за внимание!



Шевцов Василий Викторович

shevtsov_vv@rudn.university
+7(903)144-53-57