

Визуализация данных: R & ggplot2

АНТОН АНТОНОВ
tonytonov@gmail.com

Базовые графические средства

- Контр-интуитивны
- Избыточны
- Ужасно выглядят

Например:

```
pch, lwd, cex (?)
```

```
par(mar = c(5, 4, 4, 2) + 0.1) (??)
```

Ключевые особенности ggplot2

- Использование “grammar of graphics” (Lee Wilkinson): любой график составляется из базовых компонент, разделённых семантически
- Читабельный код
- Исключительная гибкость
- Довольно высокая сложность
- Publication quality
- Растущая популярность, обширное сообщество
- Hadley Wickham: поддержка, документация, книги

- **qplot** – “quick plot”, аналог базового plot
- **ggplot** – основная функция, указывающая на набор данных и задающая связь между переменными в данных и графическими переменными (variable mapping)
- **geoms** – геометрические объекты
 - `geom_point()`, `geom_line()`, `geom_density()`, `geom_bar()`, `geom_area()`, `geom_polygon()`, ...
- **aes** – aesthetics, графические переменные
 - `size`, `colour` (color), `fill`, `shape`, `linetype`, `alpha` (transparency)
- **scales** – правила отображения единиц, в которых измеряются aesthetics
 - `continuous`, `discrete`, `log`, ...
- **stats** – статистические преобразования данных
 - `stat_bin()`, `stat_density()`, `stat_ecdf()`, `stat_function()`, ...

Набор данных iris (Anderson, Edgar, 1935; Fisher, 1936)

```
head(iris)
```

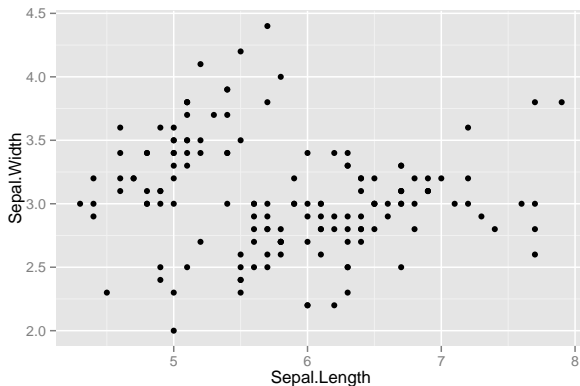
```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1           5.1           3.5           1.4           0.2  setosa
## 2           4.9           3.0           1.4           0.2  setosa
## 3           4.7           3.2           1.3           0.2  setosa
## 4           4.6           3.1           1.5           0.2  setosa
## 5           5.0           3.6           1.4           0.2  setosa
## 6           5.4           3.9           1.7           0.4  setosa
```

```
str(iris)
```

```
## 'data.frame': 150 obs. of  5 variables:
## $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1
```

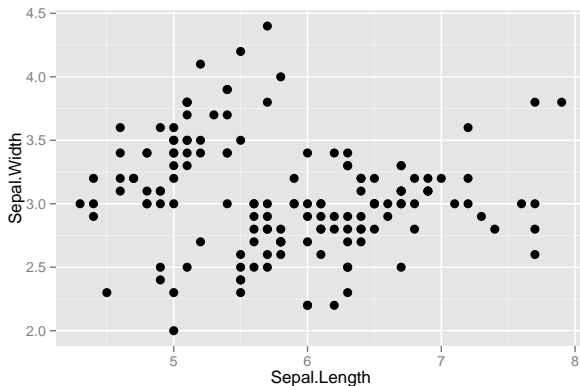
Рисуем наш первый ggplot

```
ggplot(data = iris, aes(x = Sepal.Length, y = Sepal.Width)) +  
  geom_point()
```



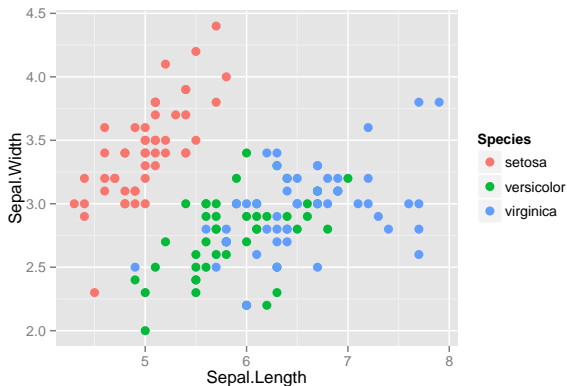
Увеличиваем размер точек

```
ggplot(data = iris, aes(x = Sepal.Length, y = Sepal.Width)) +  
  geom_point(size = 3)
```



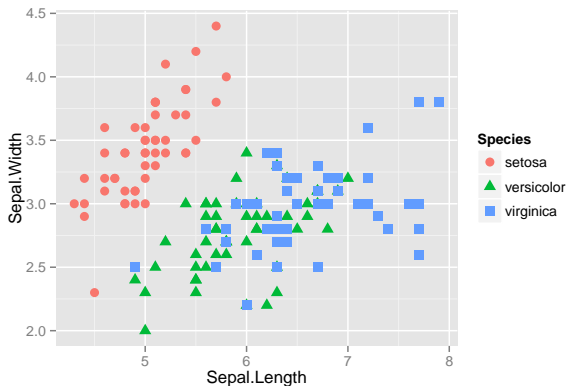
Добавляем цвет, зависящий от переменной в данных

```
ggplot(iris, aes(Sepal.Length, Sepal.Width, colour = Species)) +  
  geom_point(size = 3)
```



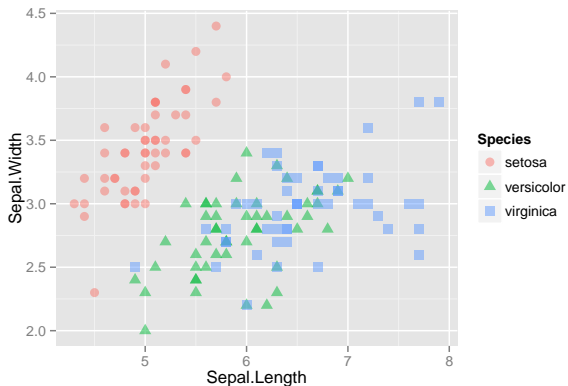
Разделяем точки при помощи формы

```
ggplot(iris, aes(Sepal.Length, Sepal.Width, colour = Species)) +  
  geom_point(aes(shape = Species), size = 3)
```



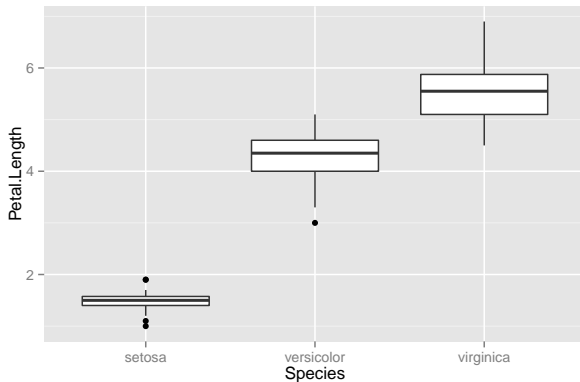
Добавляем полупрозрачность

```
ggplot(iris, aes(Sepal.Length, Sepal.Width, colour = Species)) +  
  geom_point(aes(shape = Species), size = 3, alpha = 1/2)
```



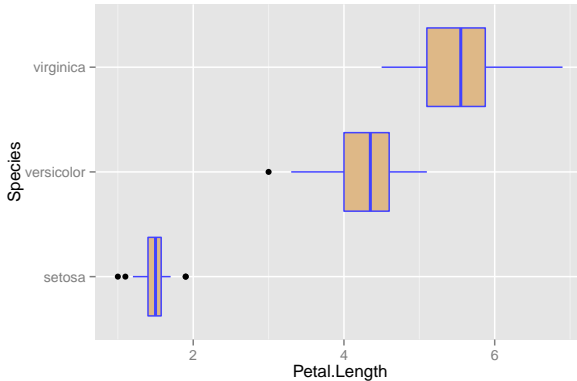
Boxplot

```
ggplot(iris, aes(x = Species, y = Petal.Length)) + geom_boxplot()
```



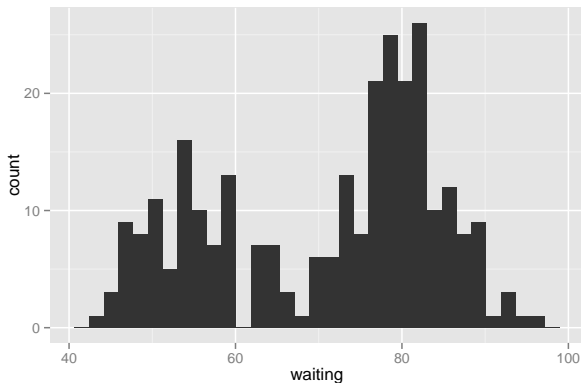
Кастомизируем boxplot

```
ggplot(iris, aes(x = Species, y = Petal.Length)) +  
  geom_boxplot(fill = "burlywood", colour = "#4242FF") +  
  coord_flip()
```



Гистограмма

```
ff.plot <- ggplot(faithful, aes(x = waiting))  
ff.plot + geom_histogram()
```



Кастомизируем гистограмму

```
ff.plot +  
  geom_histogram(binwidth = 6, fill = "coral", colour = "red")
```

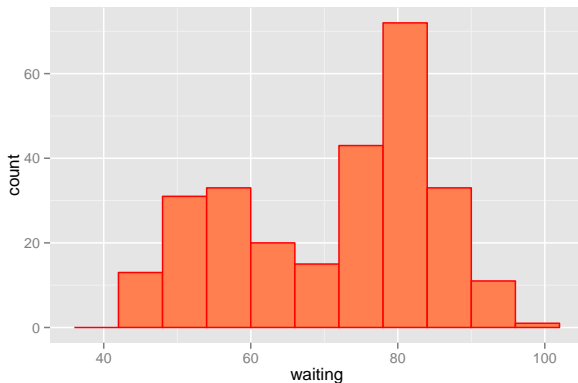
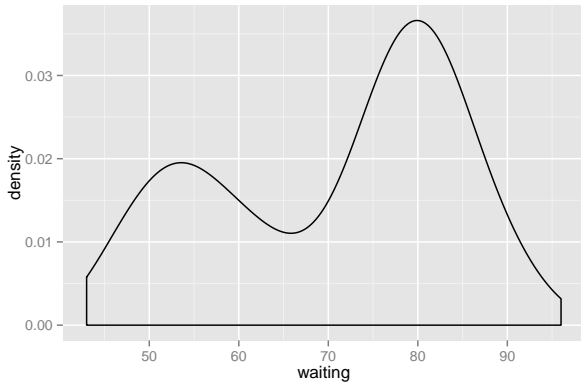


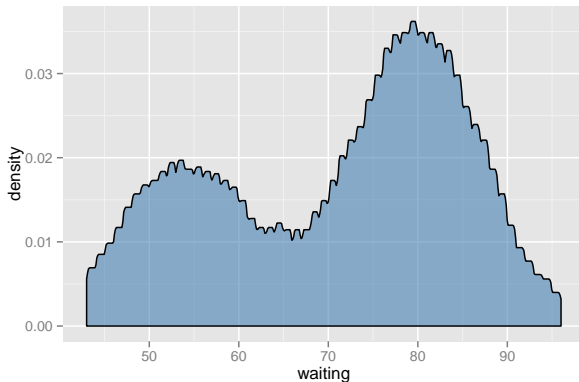
График оценки плотности

```
ff.plot + geom_density()
```



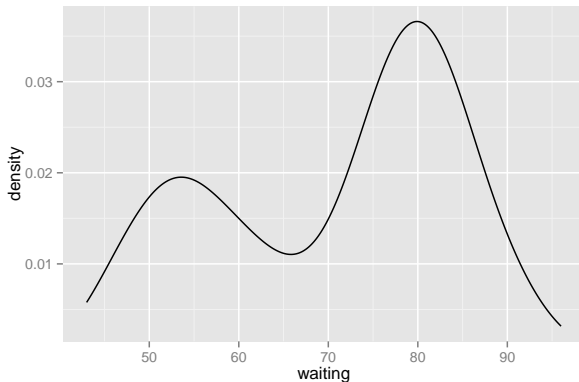
Кастомизируем график оценки плотности

```
ff.plot +  
  geom_density(kernel = "rectangular",  
              fill = "steelblue", alpha = 0.6)
```



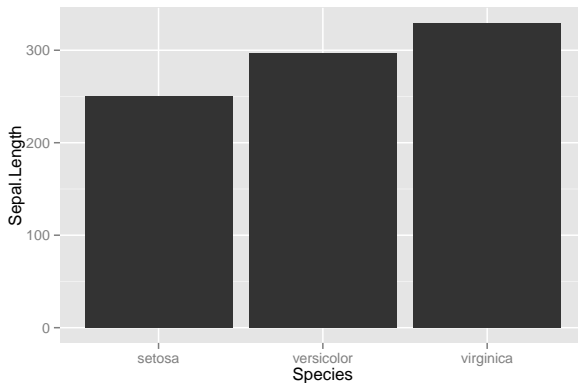
Подход с использованием stat

```
ff.plot + geom_line(stat = "density")
```



Barplot

```
ggplot(iris, aes(Species, Sepal.Length)) +  
  geom_bar(stat = "identity")
```



Из широкого формата – в длинный

plyr & reshape2 – “swiss army knives of R”

```
library(reshape2)
```

```
head(iris, 4)
```

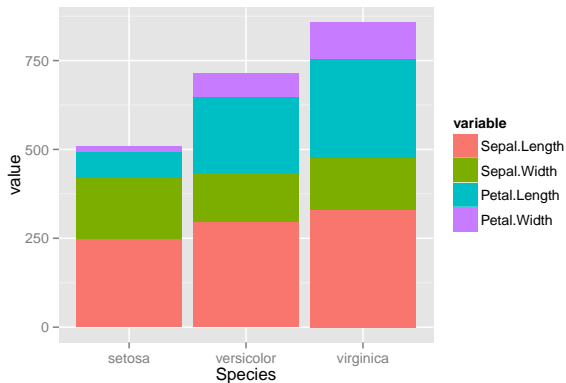
```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1           5.1           3.5           1.4           0.2   setosa
## 2           4.9           3.0           1.4           0.2   setosa
## 3           4.7           3.2           1.3           0.2   setosa
## 4           4.6           3.1           1.5           0.2   setosa
```

```
iris.melt <- melt(iris, id.vars = "Species")
```

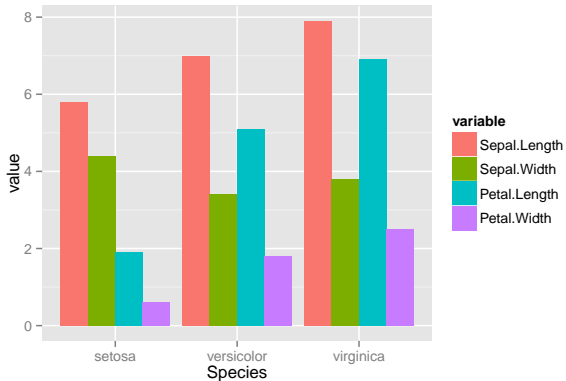
```
head(iris.melt, 4)
```

```
##      Species      variable value
## 1   setosa Sepal.Length    5.1
## 2   setosa Sepal.Length    4.9
## 3   setosa Sepal.Length    4.7
## 4   setosa Sepal.Length    4.6
```

```
iris.melt <- melt(iris, id.vars = "Species")  
ggplot(iris.melt, aes(Species, value, fill = variable)) +  
  geom_bar(stat = "identity")
```

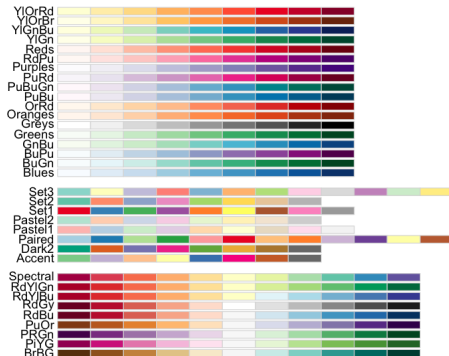


```
ggplot(iris.melt, aes(Species, value, fill = variable)) +  
  geom_bar(stat = "identity", position = "dodge")
```



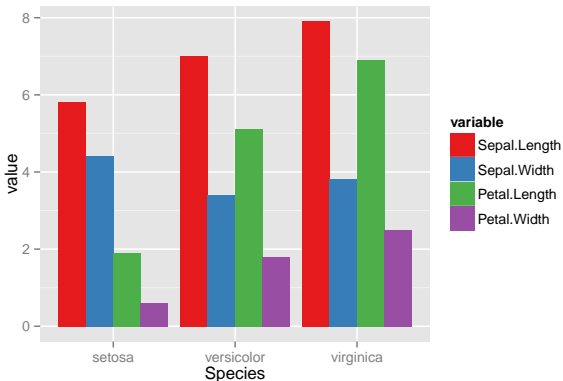
RColorBrewer

```
library(RColorBrewer)
display.brewer.all()
```



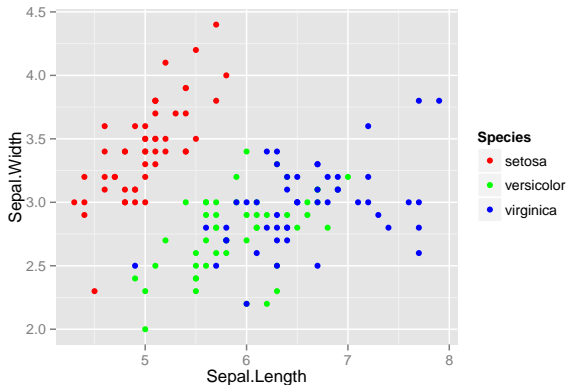
Задание палитры

```
ggplot(iris.melt, aes(Species, value, fill = variable)) +  
  geom_bar(stat = "identity", position = "dodge") +  
  scale_fill_brewer(palette = "Set1")
```



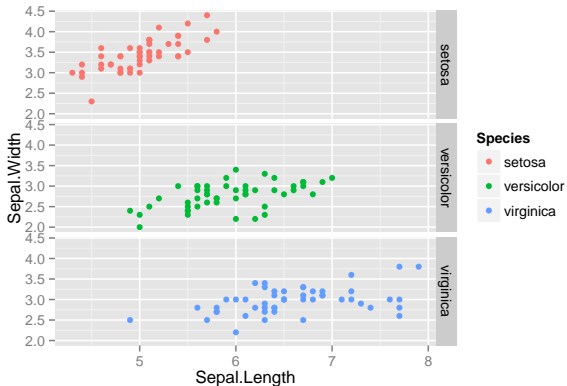
Задание палитры вручную

```
ggplot(iris, aes(Sepal.Length, Sepal.Width, colour = Species)) +  
  geom_point() +  
  scale_colour_manual(values = c("red", "green", "blue"))
```



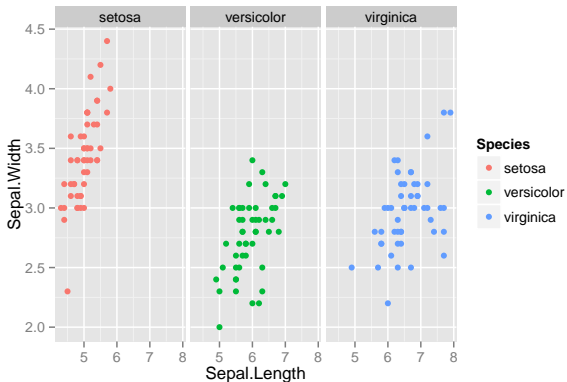
Faceting по колонкам

```
ggplot(iris, aes(Sepal.Length, Sepal.Width, colour = Species)) +  
  geom_point() +  
  facet_grid(Species ~ .)
```



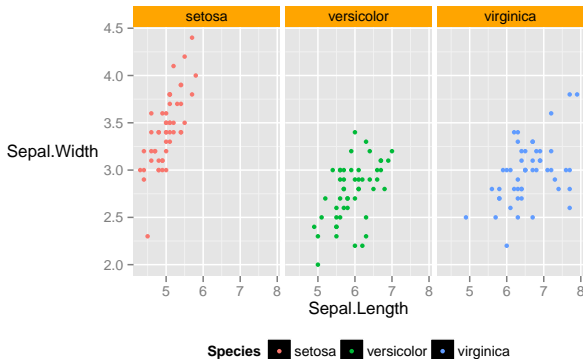
Faceting по рядам

```
ggplot(iris, aes(Sepal.Length, Sepal.Width, colour = Species)) +  
  geom_point() +  
  facet_grid(. ~ Species)
```



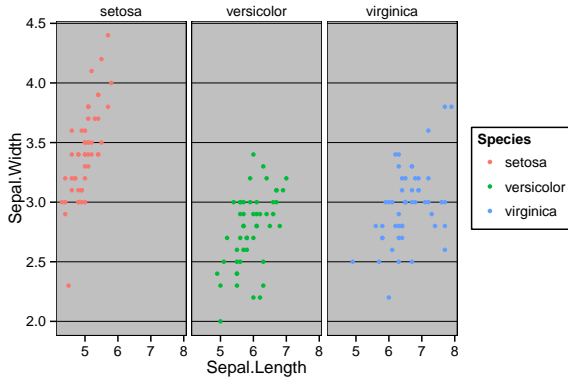
Тотальная кастомизация: theme

```
ggplot(iris, aes(Sepal.Length, Sepal.Width, colour = Species)) +  
  geom_point(size = 1.5) + facet_wrap(~ Species) +  
  theme(legend.key = element_rect(fill = "black"),  
        legend.position = "bottom",  
        strip.background = element_rect(fill = "orange"),  
        axis.title.y = element_text(angle = 0))
```



Готовые темы: ggthemes

```
library(ggthemes)
last_plot() +
  theme_excel()
```

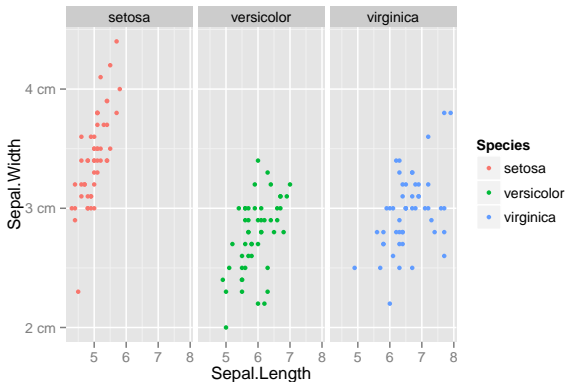


Для отрисовки “ежедневных” графиков стоит пользоваться шаблонами:

```
custom.plot <- function(df, title = "", ...) {  
  df.mod <- do.something.with(df)  
  ggplot(df.mod, aes(x, y, colour), ...) +  
    ggtitle(title) +  
    some.geoms(...)  
}  
plot1 <- custom.plot(dataset1, title = "Figure 1")  
plot1 + some.theme()
```

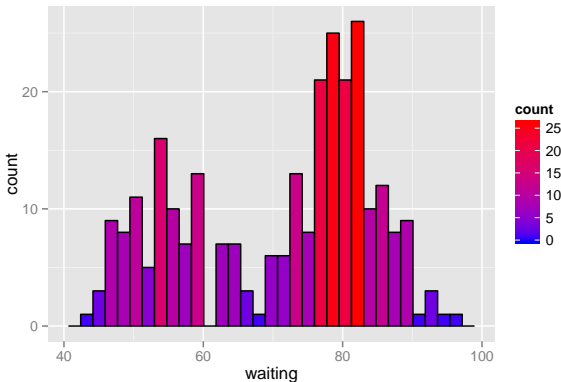
Модифицирование шкал

```
ggplot(iris, aes(Sepal.Length, Sepal.Width, colour = Species)) +  
  geom_point(size = 1.5) + facet_grid(. ~ Species) +  
  scale_y_continuous(  
    breaks = seq(2, 8, by = 1),  
    labels = paste0(2:8, " cm"))
```



Шкалы для графических переменных

```
ff.plot +  
  geom_histogram(aes(fill = ..count..), colour = "black") +  
  scale_fill_gradient(low = "blue", high = "red")
```



Сохранение изображений

- ... с экрана

```
ggsave('~ /path/to/figure/filename.png')
```

- ... из объекта

```
ggsave(plot1, file = "~/path/to/figure/filename.png")
```

- ... с заданными размерами

```
ggsave(plot1, file = "~/path/to/figure/filename.png",  
        width = 6, height = 4)
```

- ... в любом разумном формате (pdf, png, eps, svg, jpg)

```
ggsave(file = "~/path/to/figure/filename.jpg")  
ggsave(file = "~/path/to/figure/filename.pdf")
```


Что дальше?

Чтобы полноценно освоить и использовать ggplot2, необходимо:

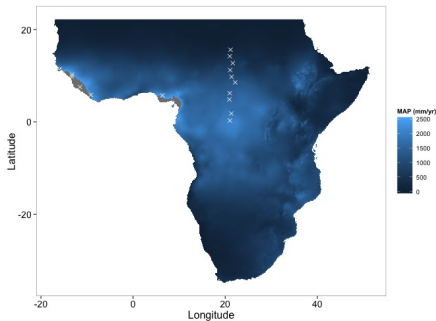
- Практиковаться
- Читать документацию
- Задавать вопросы
- Делиться результатами

Acknowledgments

- DMLabs, DMTrack
- R team
- Hadley Wickham (ggplot2, reshape2, etc.)
- Karthik Ram (github source)
- Yihui Xie (knitr)
- RStudio team
- StackOverflow
- Demo gallery authors

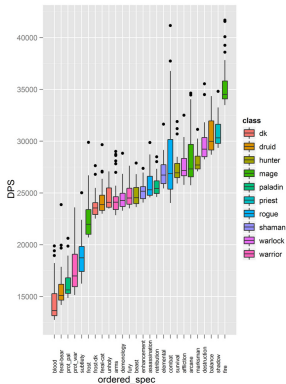
Demo gallery

Andrew Tredennick, <http://nrelscience.org/2013/05/30/this-is-how-i-did-it-mapping-in-r-with-ggplot2/>



Demo gallery

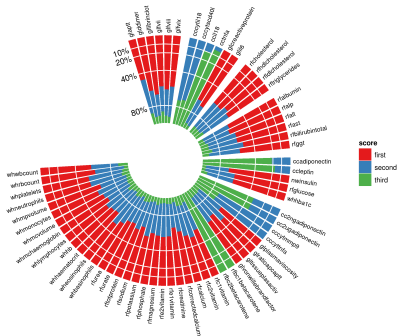
Hayward Godwin, <http://www.psychwire.co.uk/2011/04/further-adventures-in-visualisation-with-ggplot2/>



Demo gallery

Christophe Ladroue, <http://www.cristophe-ladroue.com>

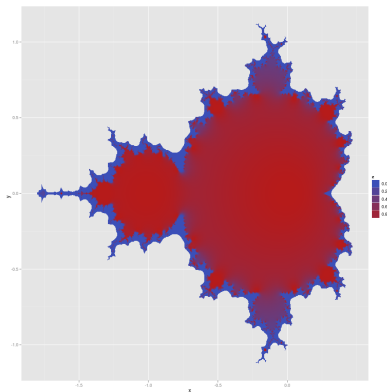
[//chrisladroue.com/2012/02/polar-histogram-pretty-and-useful/](http://chrisladroue.com/2012/02/polar-histogram-pretty-and-useful/)



Demo gallery

Siguniang's blog,

<http://siguniang.wordpress.com/2011/04/17/mandelbrot-set-in-r/>



Demo gallery

Tony Hirst, <http://blog.ouseful.info/2012/03/14/plotting-latitude-and-longitude-with-ggplot-map-projections-in-r/>

