

Анализ данных

## Лекция 5

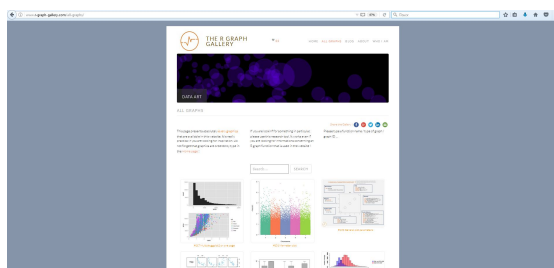
### Визуализация данных в R

Гедранович Ольга Брониславовна,  
старший преподаватель кафедры ИТ, МИУ  
volha.b.k@gmail.com

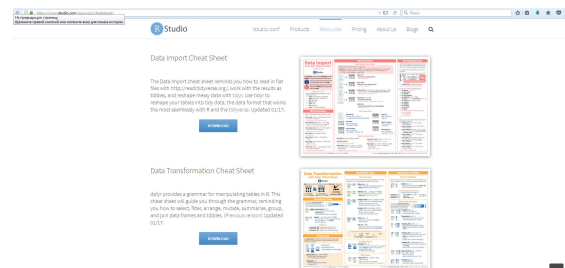
## Вопросы лекции

- Функции `plot()`, `hist()`, `barplot()`, `dotchart()`
- Графические устройства
- Графические опции
- Пакет `ggplot2`
  - Отображение данных
  - Facets
  - Геометрические объекты
  - Системы координат

<http://www.r-graph-gallery.com/all-graphs/>



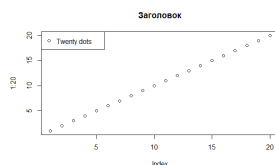
<https://www.rstudio.com/resources/cheatsheets/>



## Функция `plot()`

`plot()` — основная графическая команда, она распознает тип объекта, который подлежит рисованию, и строит соответствующий график.

```
plot(1:20, main = "Заголовок")
legend("topleft", pch = 1, legend = "Twenty dots")
```



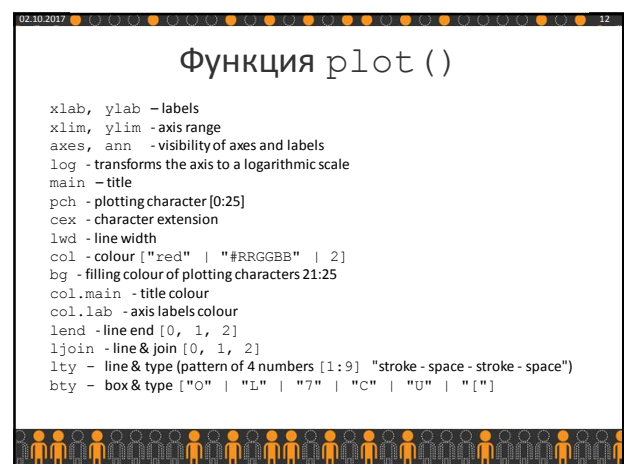
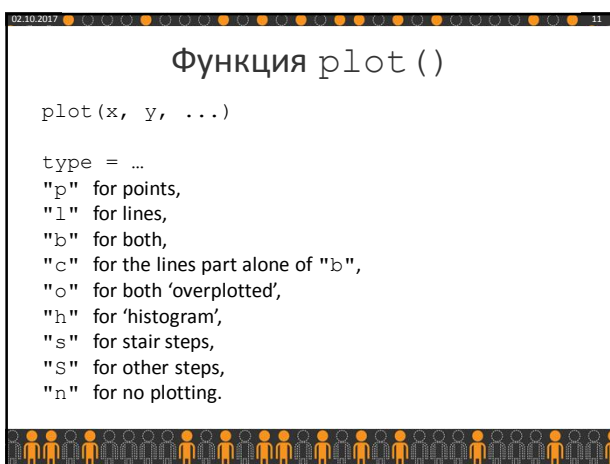
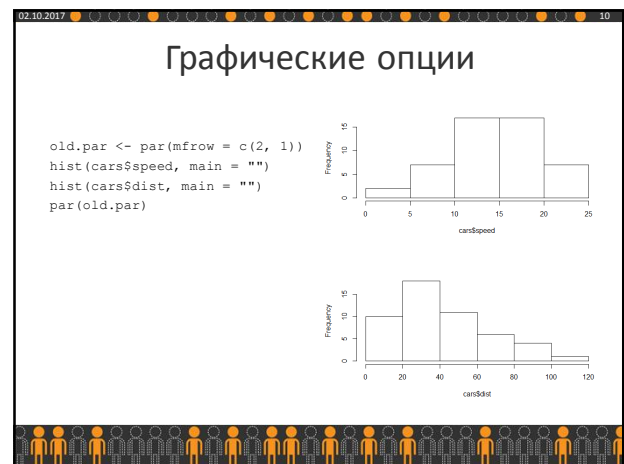
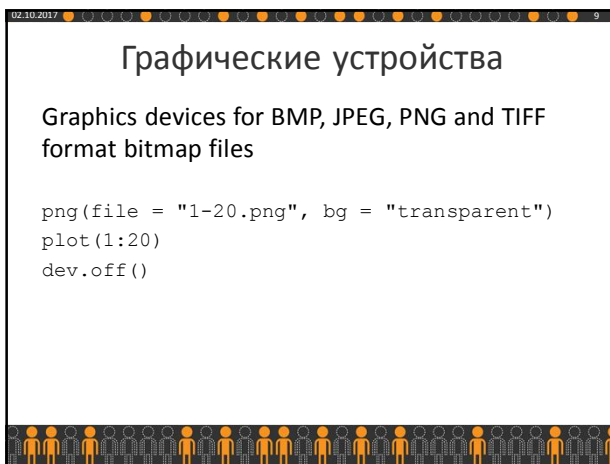
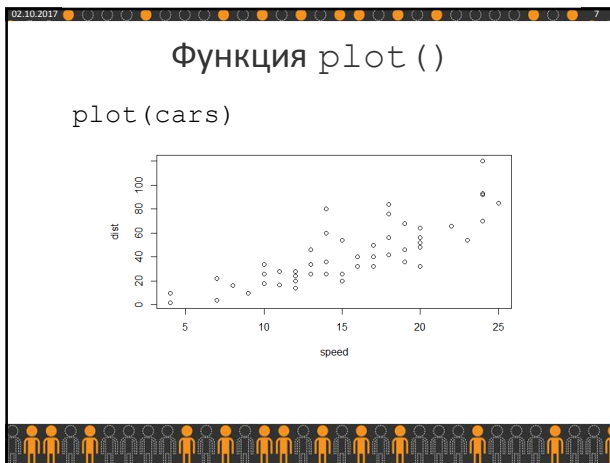
## Функция `plot()`

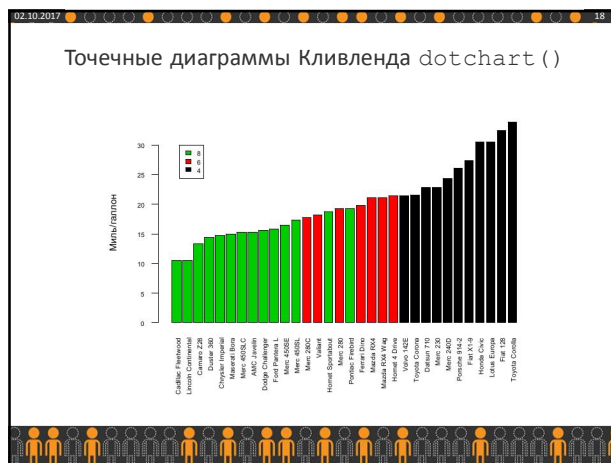
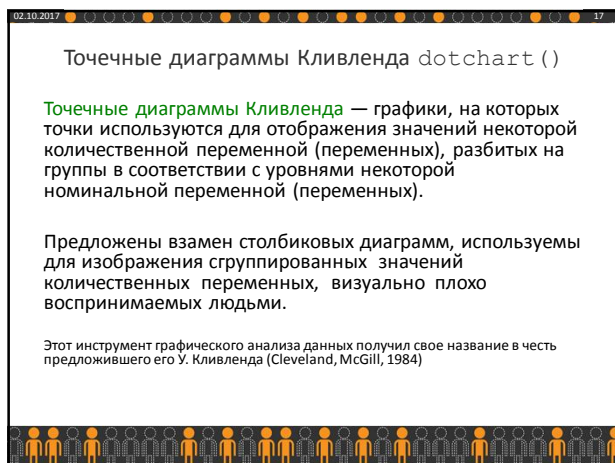
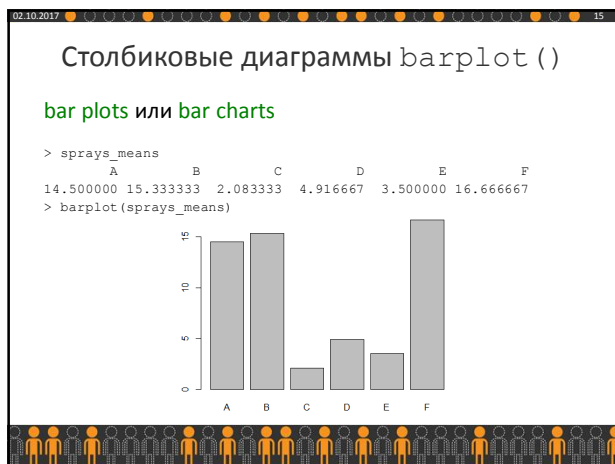
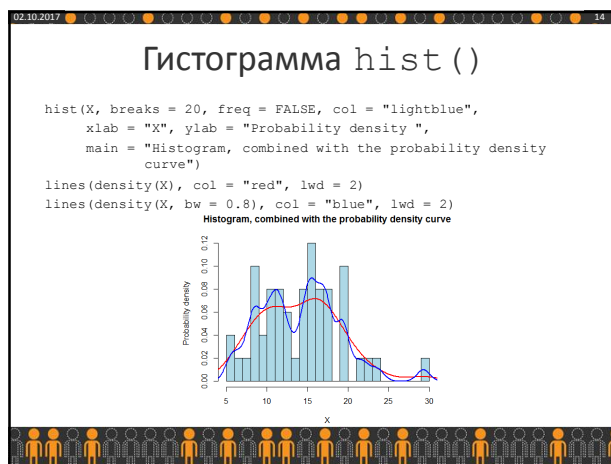
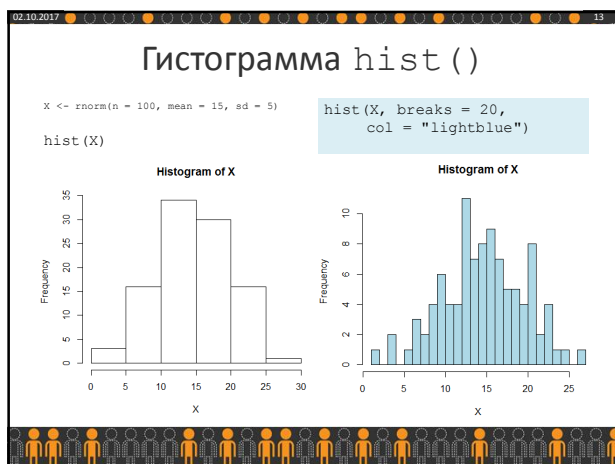
```
> head(cars)
  speed dist
```

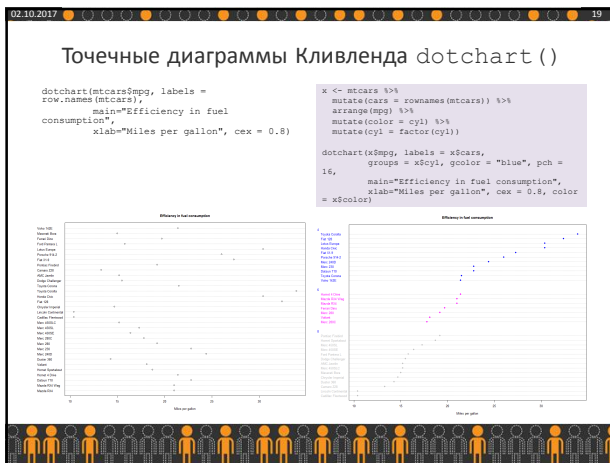
```
1    4     2
2    4    10
3    7     4
4    7    22
5    8    16
6    9    10
```

```
> head(trees)
```

```
  Girth Height Volume
1   8.3     70   10.3
2   8.6     65   10.3
3   8.8     63   10.2
4  10.5     72   16.4
5  10.7     81   18.8
6  10.8     83   19.7
```







02.10.2017 20

### Пакет ggplot2

R has several systems for making graphs, but ggplot2 is one of the most elegant and most versatile. ggplot2 implements the *grammar of graphics*, a coherent system for describing and building graphs.

"The Layered Grammar of Graphics",  
<http://vita.had.co.nz/papers/layered-grammar.pdf>

02.10.2017 21

### Пакет ggplot2

```
library(tidyverse)
#> Loading tidyverse: ggplot2
#> Loading tidyverse: tibble
#> Loading tidyverse: tidyr
#> Loading tidyverse: readr
#> Loading tidyverse: purrr
#> Loading tidyverse: dplyr
#> Conflicts with tidy packages -----
#> filter(): dplyr, stats
#> lag(): dplyr, stats

package::function()
package::function()
```

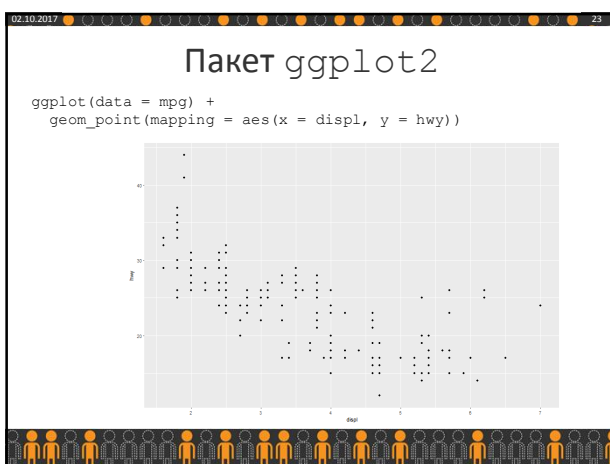
02.10.2017 22

### Пакет ggplot2

```
mpg
#> # A tibble: 234 x 11
#>   manufacturer model displ year   cyl  trans    drv   cty   hwy fl
#>   <chr>    <chr>   <dbl> <int> <int> <chr> <chr> <int> <int> <chr>
#> 1   audi      a4      1.8  1999     4 auto(l5) f    18   29   p
#> 2   audi      a4      1.8  1999     4 manual(m5) f    21   29   p
#> 3   audi      a4      2.0  2008     4 manual(m6) f    20   31   p
#> 4   audi      a4      2.0  2008     4 auto(av) f    21   30   p
#> 5   audi      a4      2.8  1999     6 auto(l5) f    16   26   p
#> 6   audi      a4      2.8  1999     6 manual(m5) f    18   26   p
#> # ... with 228 more rows, and 1 more variables: class <chr>
```

displ — a car's engine size, in litres.

hwy — a car's fuel efficiency on the highway, in miles per gallon (mpg). A car with a low fuel efficiency consumes more fuel than a car with a high fuel efficiency when they travel the same distance.



02.10.2017 24

### Пакет ggplot2

geom\_point() adds a layer of points to the plot, which creates a scatterplot.

ggplot2 comes with many geom functions that each add a different type of layer to a plot.

Each geom function in ggplot2 takes a mapping argument. This defines how variables in the dataset are mapped to visual properties.

The mapping argument is always paired with aes(), and the x and y arguments of aes() specify which variables to map to the x and y axes.

ggplot2 looks for the mapped variable in the data argument, in this case, mpg.

## Пакет ggplot2

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

An **aesthetic** is a visual property of the objects in the plot. Aesthetics include things like the size, the shape, or the color of your points.

```
color / colour  
fill  
alpha (transparency)  
size
```

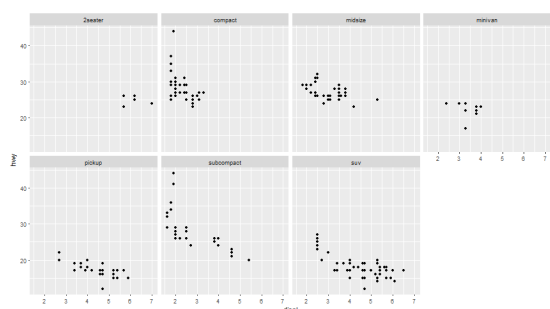
## Пакет ggplot2

**Facets** — subplots that each display one subset of the data.

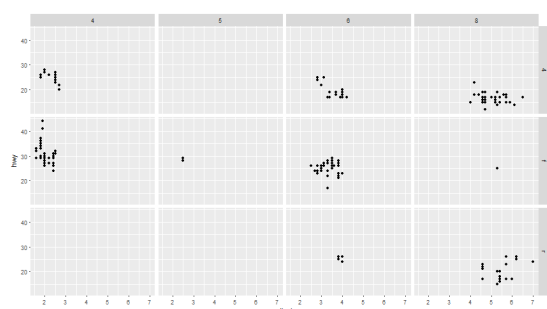
```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy)) +  
  facet_wrap(~ class, nrow = 2)
```

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy)) +  
  facet_grid(drv ~ cyl)
```

## Пакет ggplot2



## Пакет ggplot2



## Пакет ggplot2

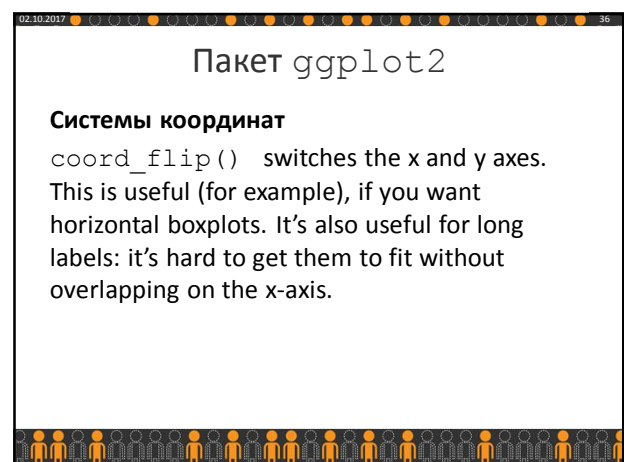
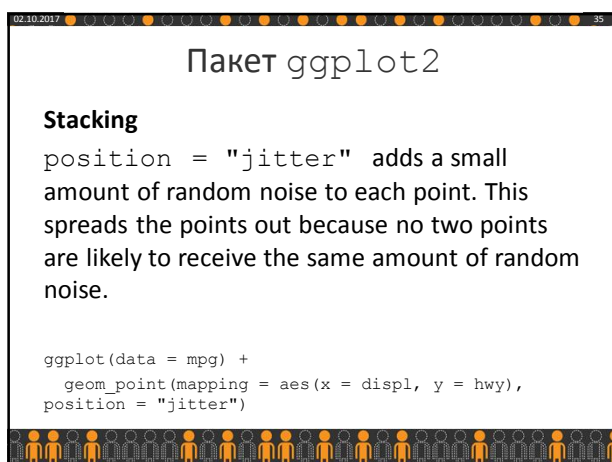
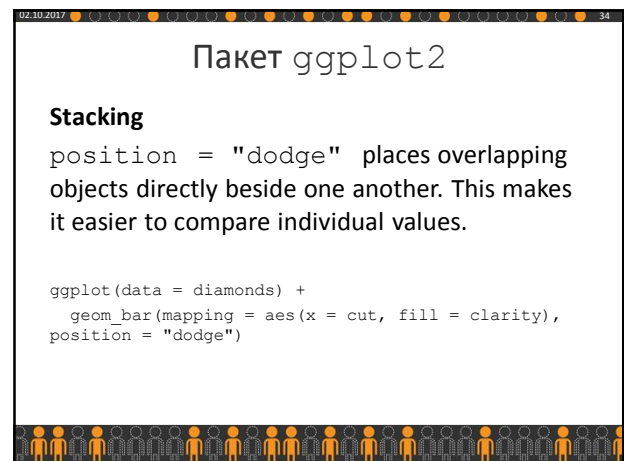
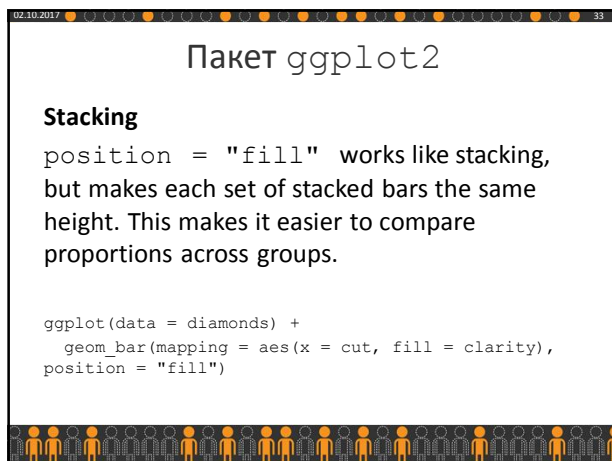
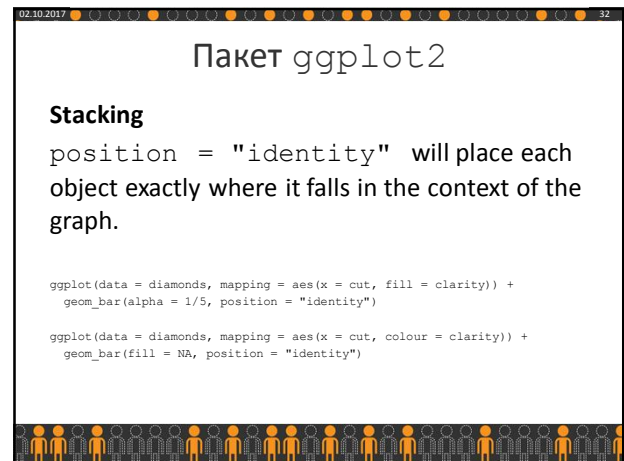
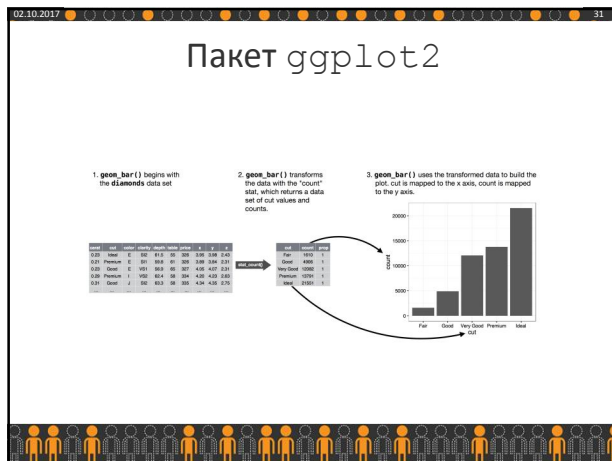
**Geoms**  
(геометрические объекты) –  
вид графика



## Пакет ggplot2

- **bar charts, histograms, and frequency polygons** bin your data and then plot bin counts, the number of points that fall in each bin.
- **smoothers** fit a model to your data and then plot predictions from the model.
- **boxplots** compute a robust summary of the distribution and then display a specially formatted box.





## Пакет ggplot2

### Системы координат

`coord_quickmap()` sets the aspect ratio correctly for maps. This is very important if you're plotting spatial data with ggplot2 (which unfortunately we don't have the space to cover in this book).

## Пакет ggplot2

### Системы координат

`coord_polar()` uses polar coordinates. Polar coordinates reveal an interesting connection between a bar chart and a Coxcomb chart.

## Основная литература

- Мастицкий, С. Э. Статистический анализ и визуализация данных с помощью R [Электронный ресурс] / С. Э. Мастицкий, В. К. Шитиков. – 2014. – Режим доступа: <http://www.ievbras.ru/ecostat/Kiril/R/Mastitsky%20and%20Shitikov%202014.pdf>. – Дата доступа: 01.09.2016.
- Шипунов, А. Б. Наглядная статистика. Используем R! [Электронный ресурс] / А. Б. Шипунов, Е. М. Балдин, П. А. Волкова, А. И. Коробейников, С. А. Назарова, С. В. Петров, В. Г. Суфиянов. – 2014. – Режим доступа: <https://cran.r-project.org/doc/contrib/Shipunov-rbook.pdf>. – Дата доступа: 01.09.2016.
- Grolemund, G. R for Data Science [Electronic resource] / Garrett Grolemund, Hadley Wickham. – 2016. – Mode of access: <http://r4ds.had.co.nz/index.html>. – Date of access: 01.09.2016.