

Модели и технологии оперативного анализа данных

## Лекция 8

### Визуализация данных в R

Гедранович Ольга Брониславовна,  
старший преподаватель кафедры ИТ, МИУ  
[volha.b.k@gmail.com](mailto:volha.b.k@gmail.com)

19.02.2017

2

### Вопросы лекции

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

19.02.2017 3

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

The screenshot shows the homepage of 'THE R GRAPH GALLERY'. The website has a dark blue header with the site name and navigation links: HOME, ALL GRAPHS, BLOG, ABOUT, WHO I AM. Below the header is a large banner image with the text 'DATA ART'. The main content area is titled 'ALL GRAPHS' and features a grid of various data visualizations, including bar charts, scatter plots, and heatmaps. A search bar is located in the center of the grid. The website is framed by a decorative border of orange and black circles at the top and bottom.

19.02.2017 4

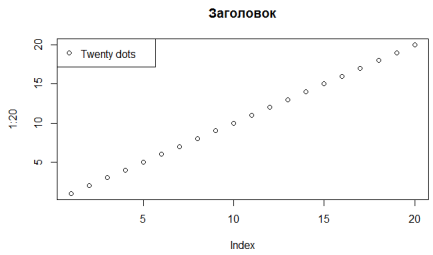
# https://www.rstudio.com/resources/cheatsheets/

The screenshot shows the 'Resources' page of the RStudio website, specifically the 'cheatsheets' section. The page features two main sections: 'Data Import Cheat Sheet' and 'Data Transformation Cheat Sheet'. Each section includes a brief description of the cheat sheet's purpose and a 'Download' button. The 'Data Import Cheat Sheet' is described as a guide for reading data from flat files and working with tibbles. The 'Data Transformation Cheat Sheet' is described as a guide for manipulating tables in R using the dplyr package. The website is framed by a decorative border of orange and black circles at the top and bottom.

19.02.2017 5

## 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")
```

19.02.2017 5

19.02.2017 6

## Функция `plot()`

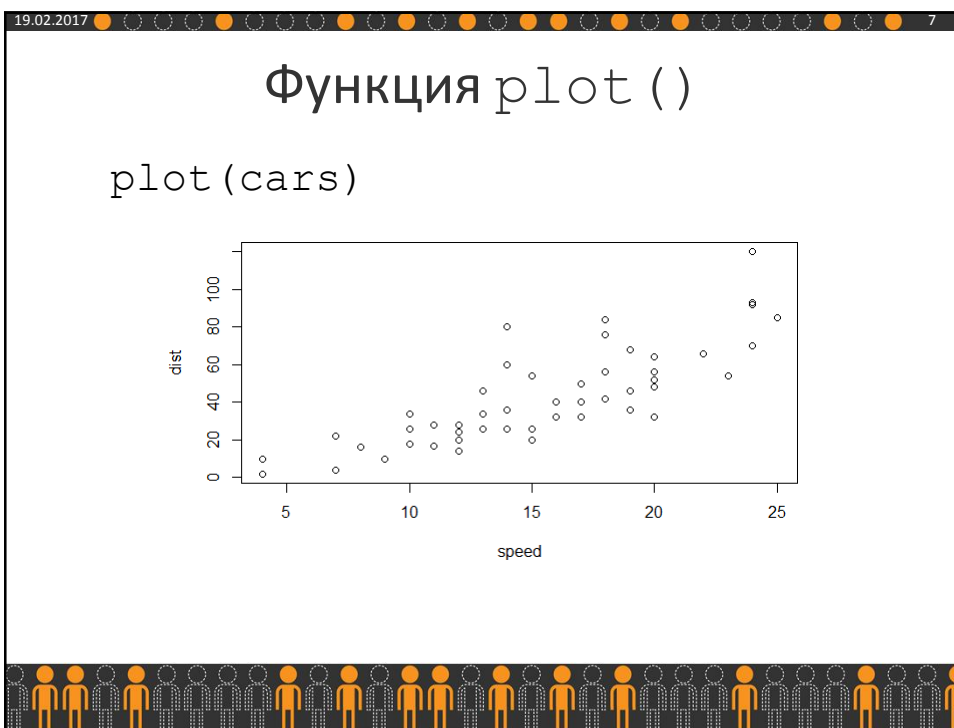
```
> head(cars)
```

	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

19.02.2017 6



19.02.2017 9

## Графические устройства

### Graphics devices for BMP, JPEG, PNG and TIFF format bitmap files

```
png(file = "1-20.png", bg = "transparent")
plot(1:20)
dev.off()
```

19.02.2017 9

19.02.2017 10

## Графические опции

```
old.par <- par(mfrow = c(2, 1))
hist(cars$speed, main = "")
hist(cars$dist, main = "")
par(old.par)
```



The top histogram displays the frequency of car speeds. The x-axis is labeled 'cars\$speed' and ranges from 0 to 25. The y-axis is labeled 'Frequency' and ranges from 0 to 15. The distribution is roughly bell-shaped, centered around 10-15.

The bottom histogram displays the frequency of car distances. The x-axis is labeled 'cars\$dist' and ranges from 0 to 120. The y-axis is labeled 'Frequency' and ranges from 0 to 15. The distribution is roughly bell-shaped, centered around 30-40.

19.02.2017 10

19.02.2017 11

## Функция `plot()`

```
plot(x, y, ...)
```

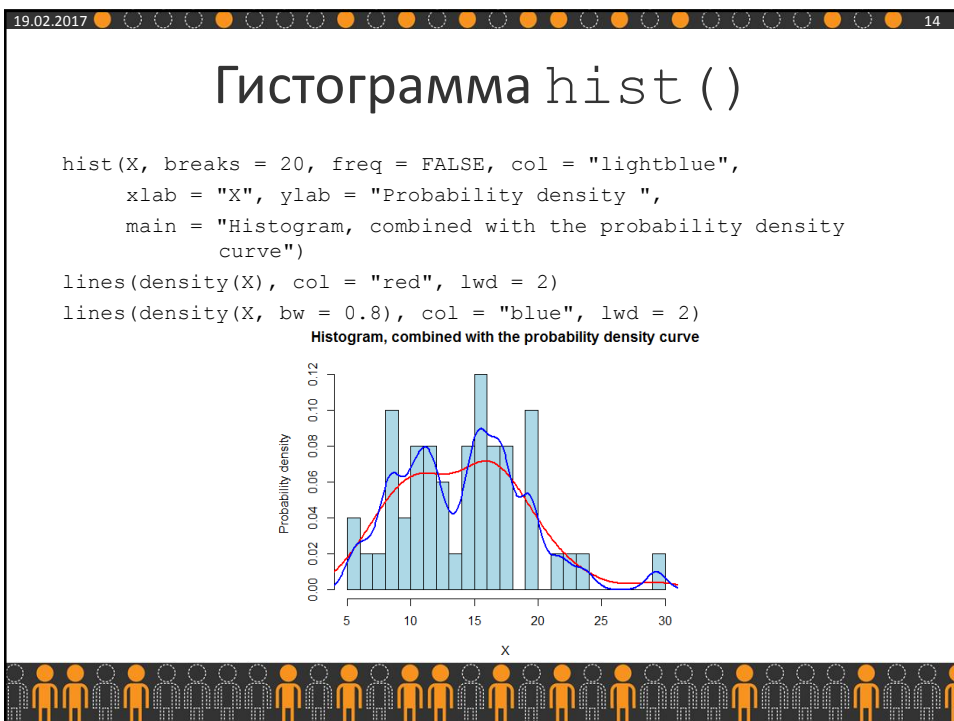
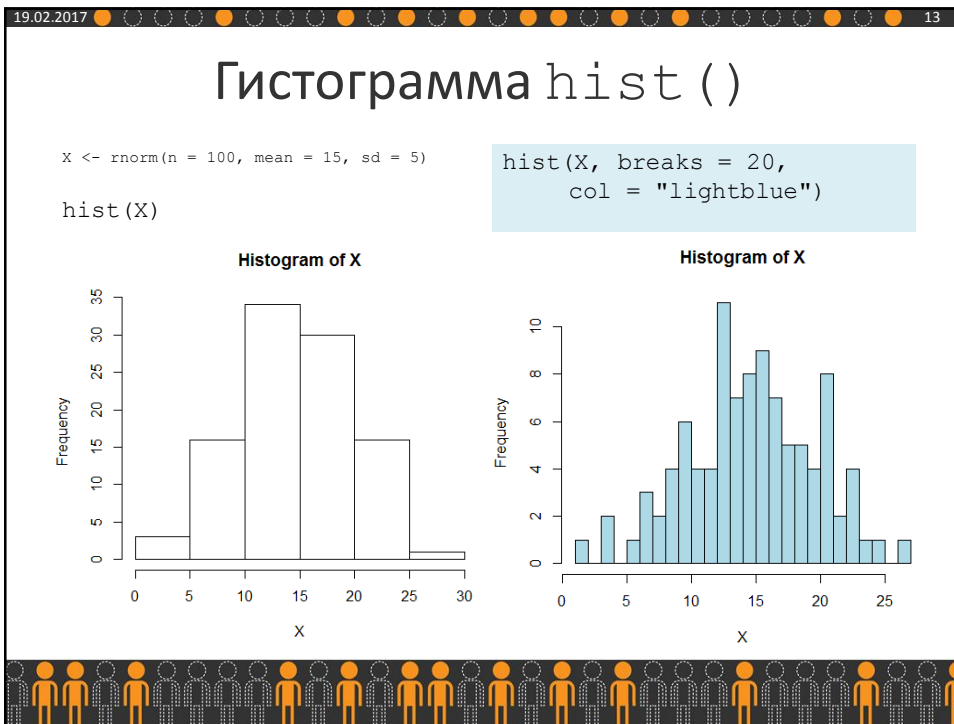
`type = ...`  
`"p"` for points,  
`"l"` for lines,  
`"b"` for both,  
`"c"` for the lines part alone of `"b"`,  
`"o"` for both 'overplotted',  
`"h"` for 'histogram',  
`"s"` for stair steps,  
`"S"` for other steps,  
`"n"` for no plotting.

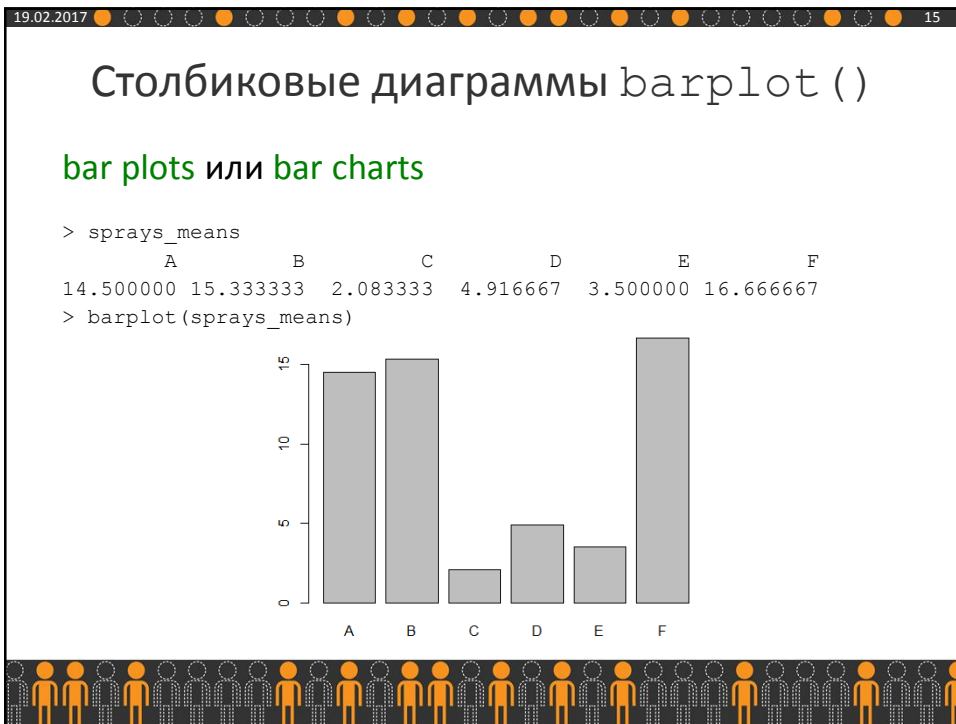
19.02.2017 12

## Функция `plot()`

```

xlab, ylab - labels
xlim, ylim - axis range
axes, ann - visibility of axes and labels
log - transforms the axis to a logarithmic scale
main - title
pch - plotting character [0:25]
cex - character extension
lwd - line width
col - colour ["red" | "#RRGGBB" | 2]
bg - filling colour of plotting characters 21:25
col.main - title colour
col.lab - axis labels colour
lend - line end [0, 1, 2]
ljoin - line & join [0, 1, 2]
lty - line & type (pattern of 4 numbers [1:9] "stroke - space - stroke - space")
bty - box & type ["o" | "L" | "7" | "C" | "U" | "["]
  
```







19.02.2017 17

## Точечные диаграммы Кливленда dotchart ( )

**Точечные диаграммы Кливленда** — графики, на которых точки используются для отображения значений некоторой количественной переменной (переменных), разбитых на группы в соответствии с уровнями некоторой номинальной переменной (переменных).

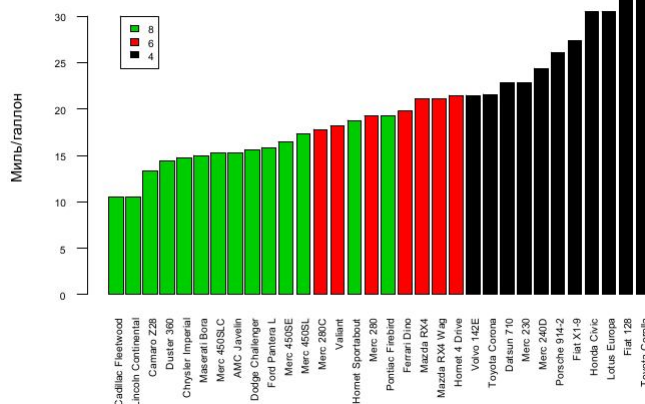
Предложены взамен столбиковых диаграмм, используемы для изображения сгруппированных значений количественных переменных, визуально плохо воспринимаемых людьми.

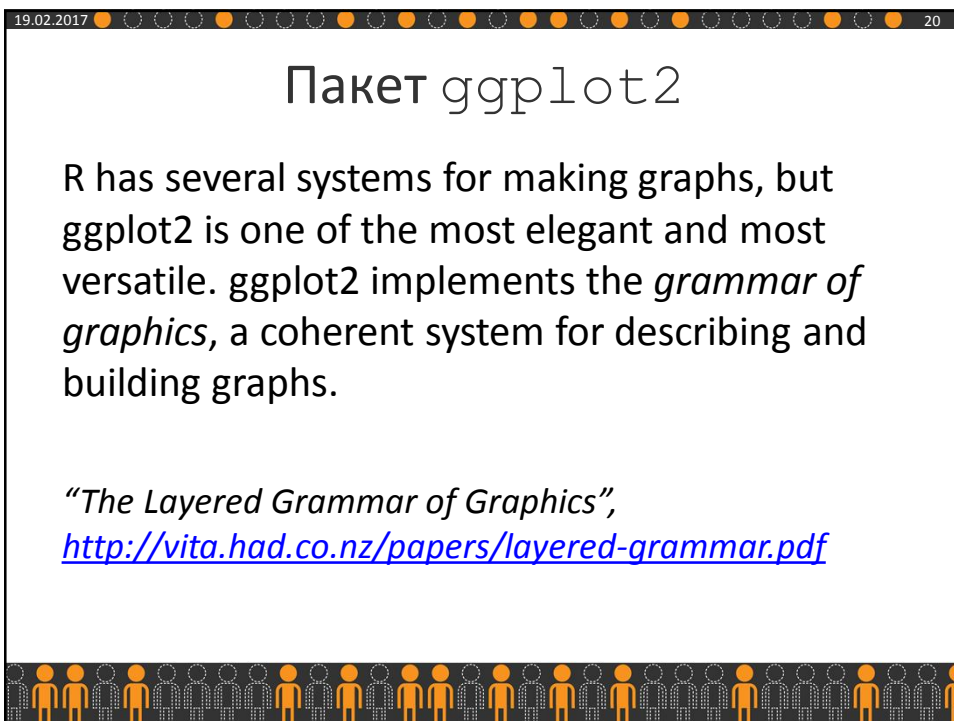
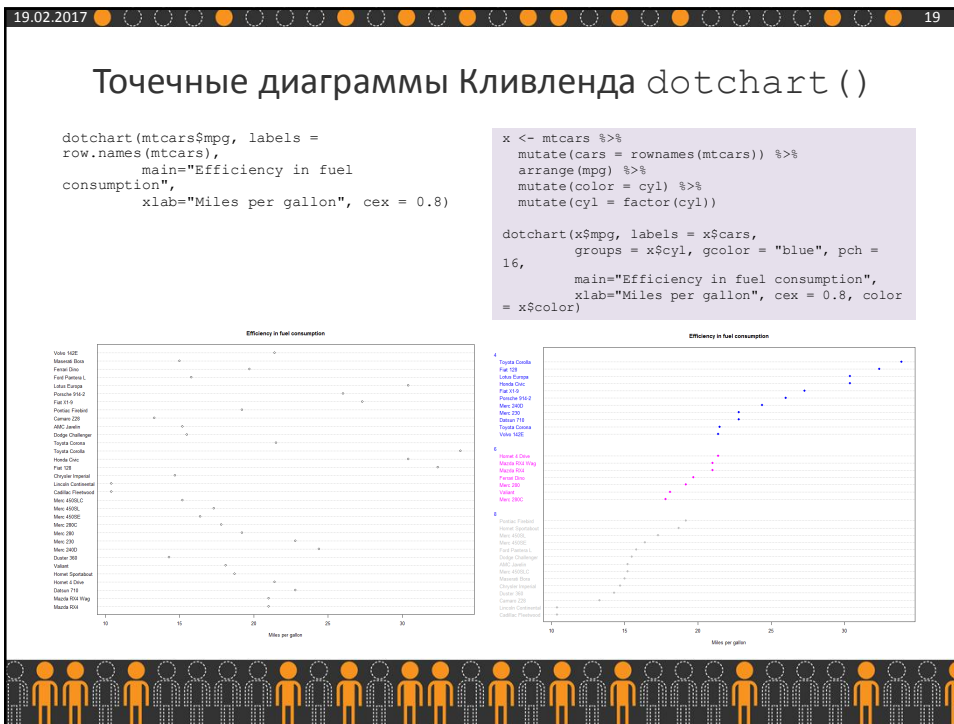
Этот инструмент графического анализа данных получил свое название в честь предложившего его У. Кливленда (Cleveland, McGill, 1984)



19.02.2017 18

## Точечные диаграммы Кливленда dotchart ( )





19.02.2017 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()
```




19.02.2017 22

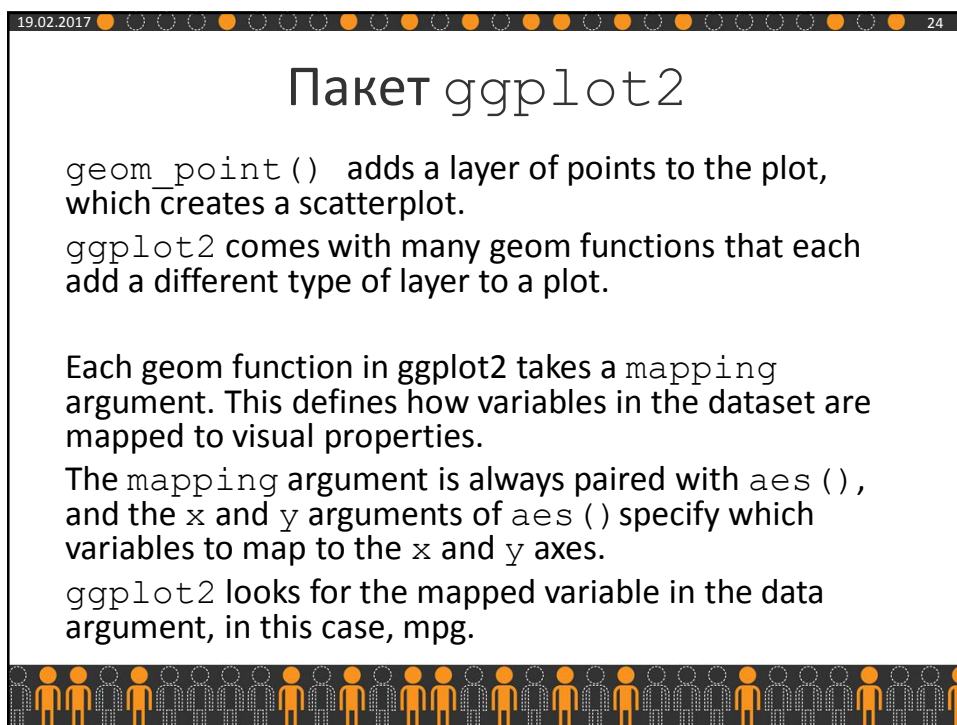
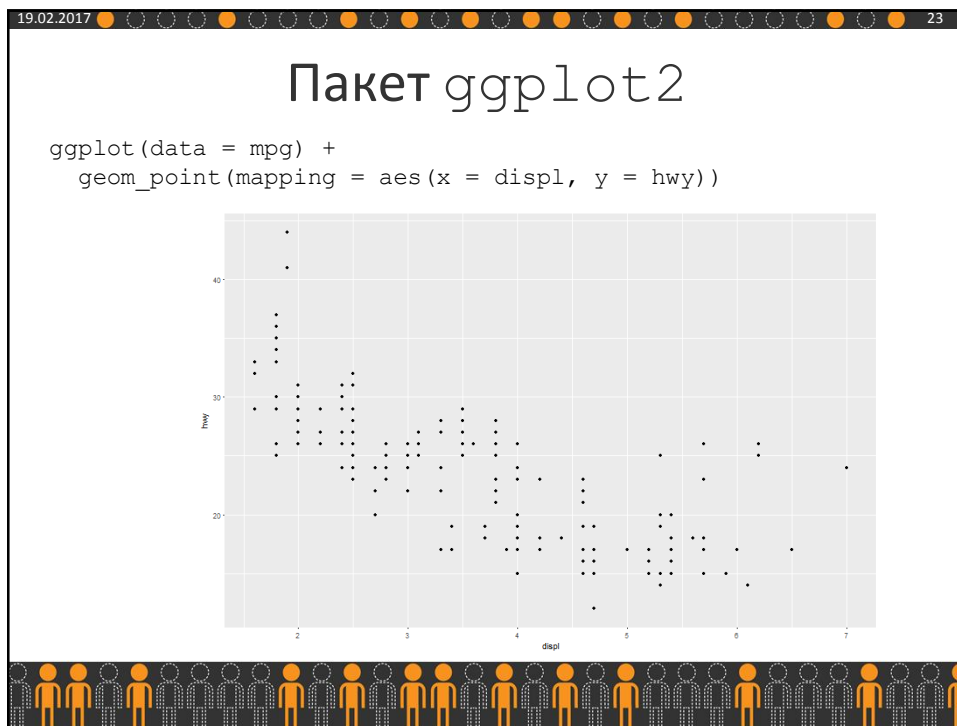
## Пакет ggplot2

```
mpg
#> # A tibble: 234 × 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.






19.02.2017 25

## 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




19.02.2017 26

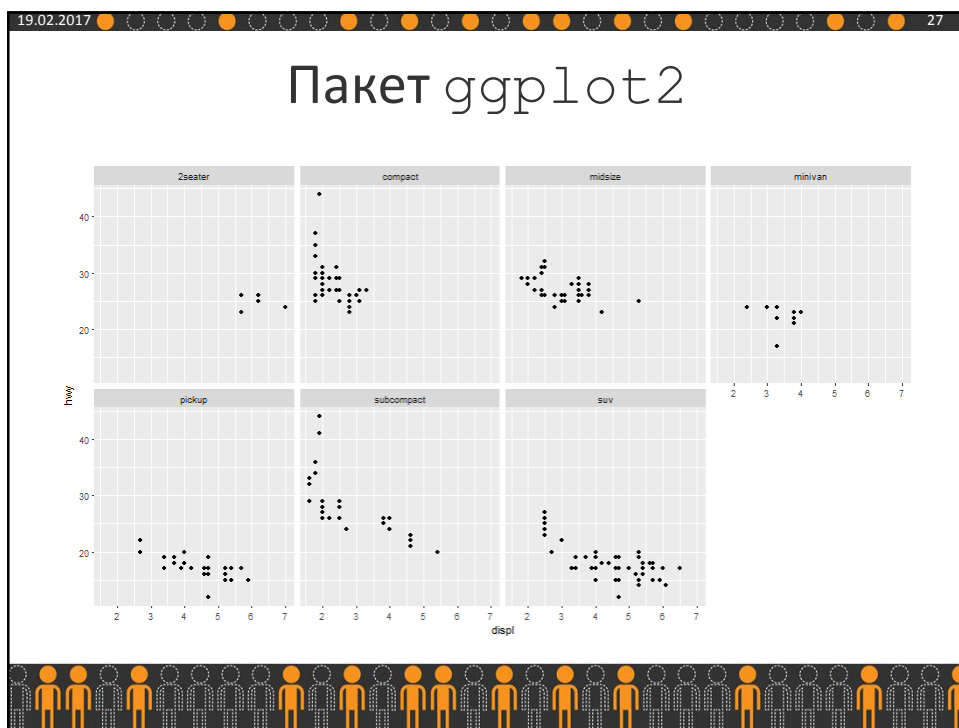
## Пакет ggplot2

**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)
```





19.02.2017 29

# Пакет ggplot2

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

**Graphical Primitives**

- `geom_blank()`
- `geom_curve()`
- `geom_path()`
- `geom_polygon()`
- `geom_rect()`
- `geom_ribbon()`
- `geom_segment()`
- `geom_spoke()`

**One Variable**

- `geom_area()`
- `geom_density()`
- `geom_density2d()`
- `geom_histogram()`
- `geom_jitter()`
- `geom_point()`
- `geom_quantile()`
- `geom_rug()`
- `geom_smooth()`
- `geom_text()`
- `geom_violin()`

**Two Variables**

- `geom_bin2d()`
- `geom_density2d()`
- `geom_hex()`
- `geom_line()`
- `geom_point()`
- `geom_smooth()`
- `geom_text()`
- `geom_violin()`

**Three Variables**

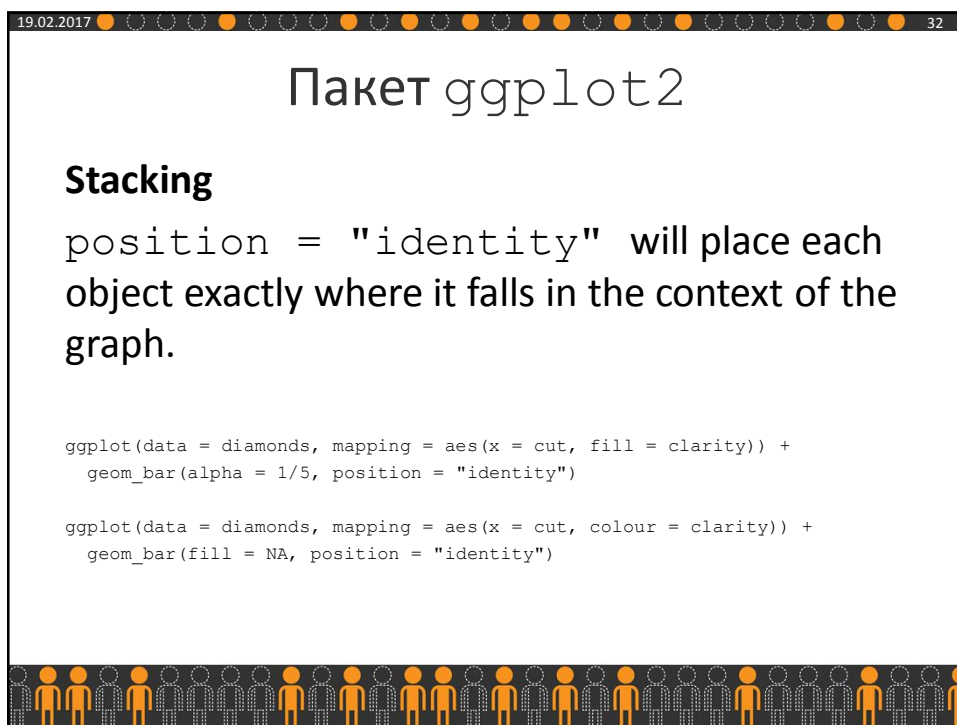
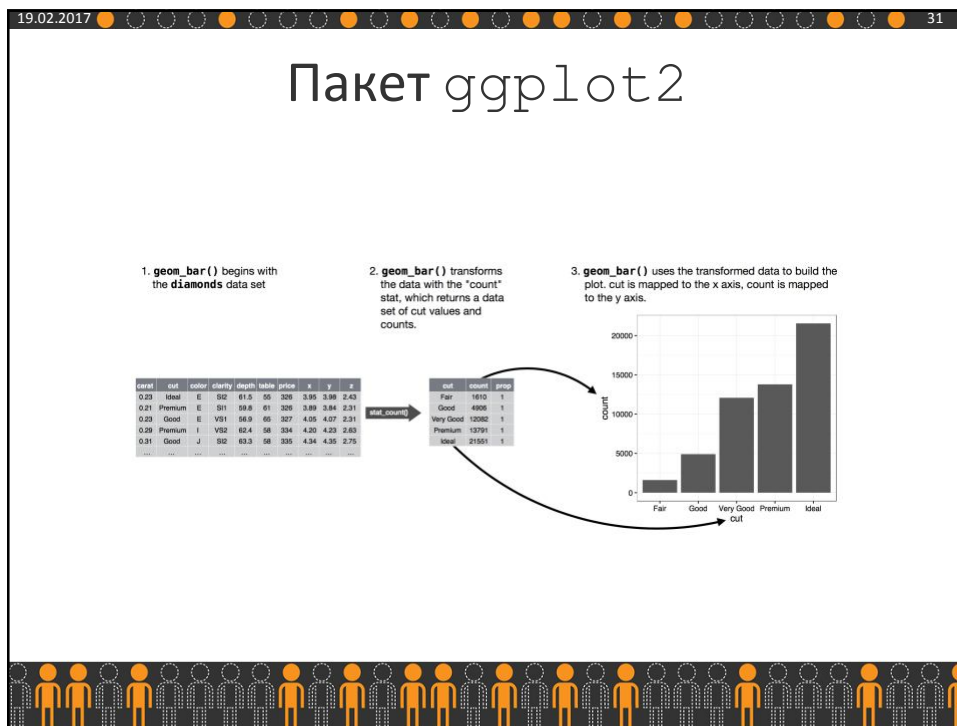
- `geom_bin2d()`
- `geom_density2d()`
- `geom_hex()`
- `geom_line()`
- `geom_point()`
- `geom_smooth()`
- `geom_text()`
- `geom_violin()`

19.02.2017 30

# 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.








19.02.2017 33

## Пакет ggplot2

### Stacking

`position = "fill"` works like stacking, but makes each set of stacked bars the same height. This makes it easier to compare proportions across groups.

```
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, fill = clarity),  
    position = "fill")
```




19.02.2017 34

## Пакет ggplot2

### Stacking

`position = "dodge"` places overlapping objects directly beside one another. This makes it easier to compare individual values.

```
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, fill = clarity),  
    position = "dodge")
```




19.02.2017 35

## Пакет ggplot2

### Stacking

`position = "jitter"` adds a small amount of random noise to each point. This spreads the points out because no two points are likely to receive the same amount of random noise.

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy),  
    position = "jitter")
```




19.02.2017 36

## Пакет ggplot2

### Системы координат

`coord_flip()` switches the x and y axes. This is useful (for example), if you want horizontal boxplots. It's also useful for long labels: it's hard to get them to fit without overlapping on the x-axis.




19.02.2017 37

## 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).




19.02.2017 38

## Пакет `ggplot2`

### Системы координат

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



19.02.2017 39

## 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.
- Grolemond, G. R for Data Science [Electronic resource] / Garrett Grolemond, Hadley Wickham. – 2016. – Mode of access: <http://r4ds.had.co.nz/index.html>. – Date of access: 01.09.2016.

