



Санкт-Петербургский
государственный
университет

Графики с использованием ggplot2

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ



Визуализация данных и результатов как инструмент

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

Лучше один раз увидеть...



<https://www.shutterstock.com/ru/image-photo/business-woman-colorful-graphs-charts-concepts-181869041>

При обработке данных и построении моделей, описывающих взаимосвязи между явлениями, ключевым этапом является **визуализация**.

Визуализация — это представление информации в виде наглядного, легкого для восприятия рисунка.

Визуализировать необходимо как исходные данные, так и результаты исследования.

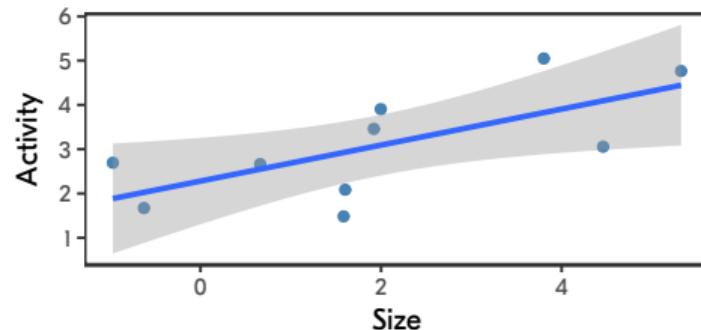
Цифры VS Графики: Всему свое место!

Одни и те же данные, представленные по-разному.

Параметры регрессионной модели, описывающей связь между размером (Size) и активностью (Activity)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.2806	0.4230	5.39	0.0007
Size	0.4066	0.1523	2.67	0.0283

Связь между размером (Size) и активностью (Activity)



Цифры VS Графики: Всему свое место!

Цифровая презентация

- Цифровое (табличное) представление данных требует больших затрат времени на изучение и осмысление. До читателя можно донести всю необходимую информацию.
- Необходимо в тексте (статья, отчет, диплом), реже в постерном докладе.
- Не годится для устного доклада.

Визуальная презентация

- Быстро воспринимается самая главная идея, заложенная в результате, которую необходимо донести до читателя. Менее значимые результаты могут быть опущены.
- Необходимо для устного и постерного доклада.
- Крайне желательно в тексте (помимо таблицы).

Основы грамматики графиков

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

Три системы визуализации в R

- Base Graphics
- Lattice
- ggplot2

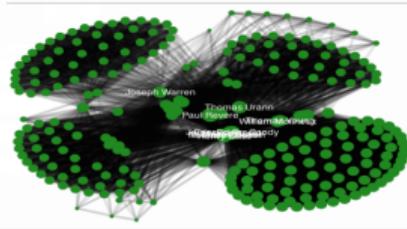
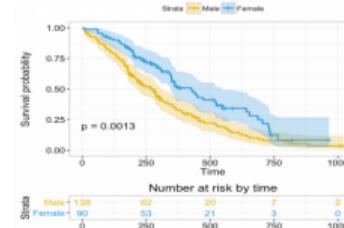
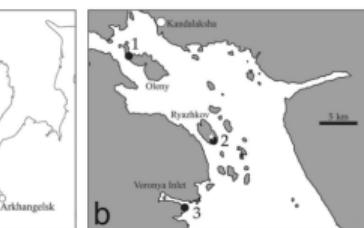
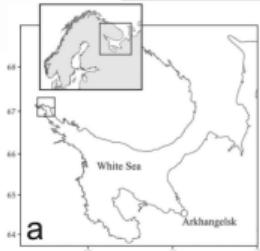
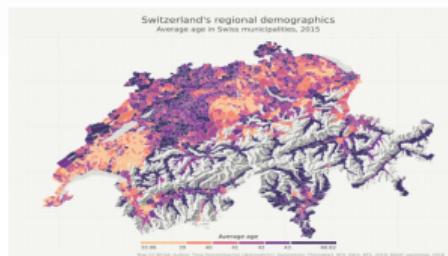
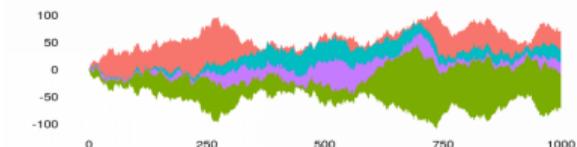
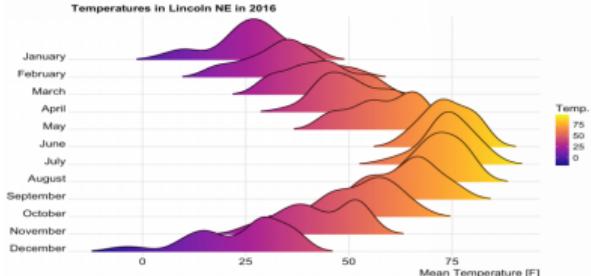
Система ggplot2

Название пакета происходит от слов “Grammar”, “Graphics”, “Plot”.

В основе подхода, реализованного в пакете, лежит идея **грамматики графиков** “Semiology of Graphics” (Bertin, 1983).

Более позднее обобщение “The Grammar of Graphics” (Wilkinson, Anand and Grossman, 2005).

Возможности ggplot2 практически безграничны



[https://timogrossenbacher.ch/2016/12/
beautiful-thematic-maps-with-ggplot2-only/](https://timogrossenbacher.ch/2016/12/beautiful-thematic-maps-with-ggplot2-only/)

<http://www.ggplot2-exts.org/gallery/>

Khaitov et al, 2018

График, как апликация



<https://www.shutterstock.com/ru/image-photo/child-sticks-bow-greeting-card-mum-526386226>

Построение графика в системе ggplot2 напоминает создание апликации.

Берется лист бумаги (базовый слой), на который слой за слоем наносятся новые элементы.

Данные для визуализации. Вес новорожденных



<https://www.shutterstock.com/image-photo/newly-born-child-gazes-camera-low-47693563>

Переменные

- low: имеет ли ребенок экстремально низкий вес
- age: возраст матери
- lwt: вес матери (в фунтах)
- race: раса матери (1 = white, 2 = black, 3 = other)
- smoke: курила ли мать во время беременности
- ptl: количество преждевременных родов в прошлом
- ht: были ли случаи повышенного давления
- ui: присутствие эрозии матки
- ftv: количество посещений врача в первом триместре
- bwt: вес ребенка при рождении

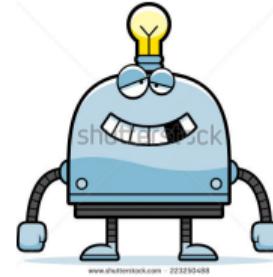
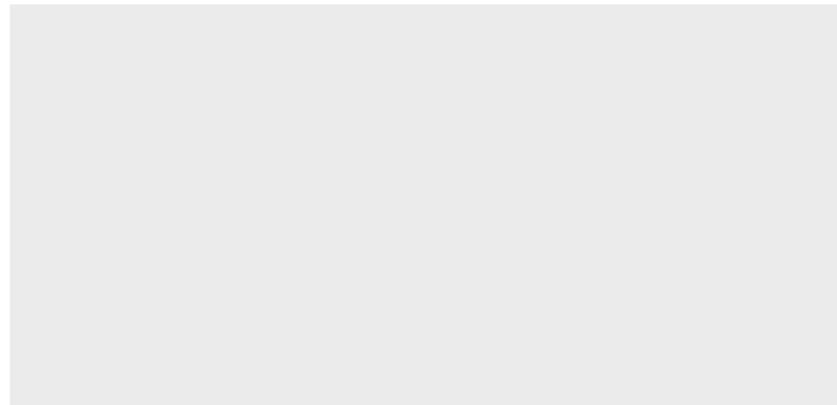
Данные: Hosmer, Lemeshow, 1989

Чтение данных

```
library(MASS)
baby <- birthwt # чтение встроенных данных
```

Робот строит простейший график

```
library(ggplot2)  
ggplot()
```

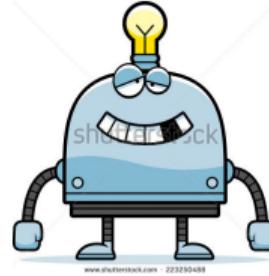
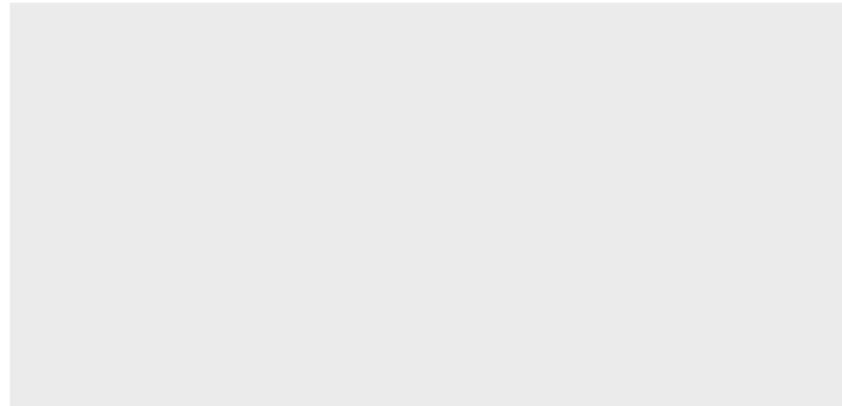


<https://www.shutterstock.com/image-vector/cartoon-illustration-malfunctioning-little-robot-223250488>

Робот, возьми лист бумаги

Робот строит простейший график

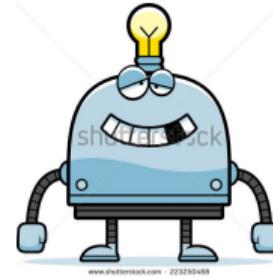
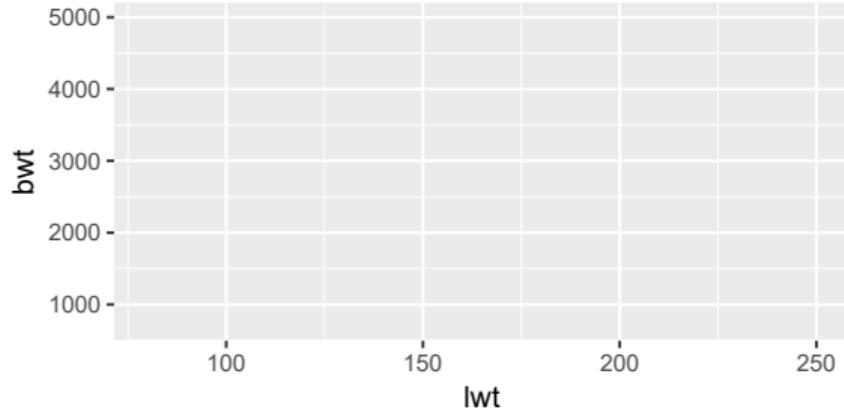
```
| ggplot(data = baby)
```



Прочтай таблицу данных

Робот строит простейший график

```
ggplot(data = baby, aes(x = lwt, y = bwt))
```

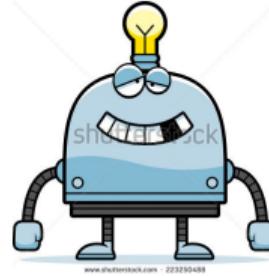
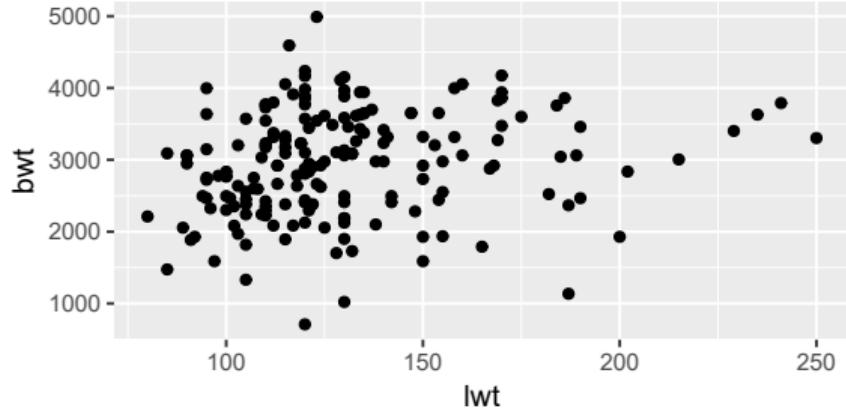


По оси ОХ отложи данные,
записанные в переменной lwt

По оси ОY отложи данные,
записанные в переменной bwt

Робот строит простейший график

```
ggplot(data = baby, aes(x = lwt, y = bwt)) + geom_point()
```



Изобрази в виде точек исходные
данные

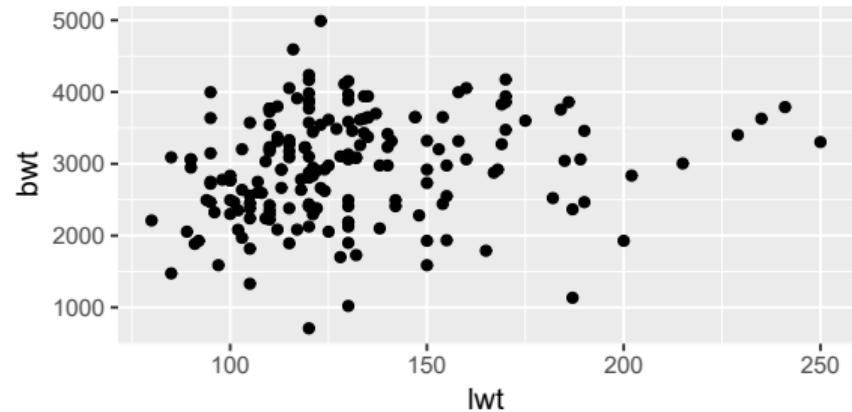
Строим точечный график: `geom_point()`

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

Простейший точечный график

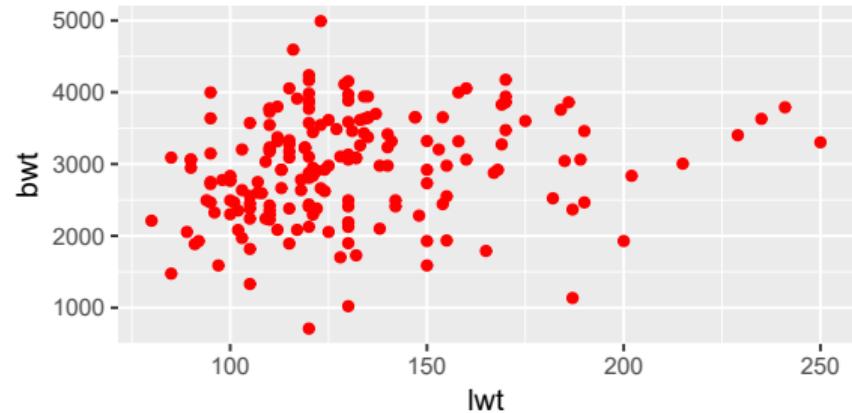
```
ggplot(data = baby, aes(x = lwt, y = bwt)) + geom_point()
```



Этот график можно изменить.

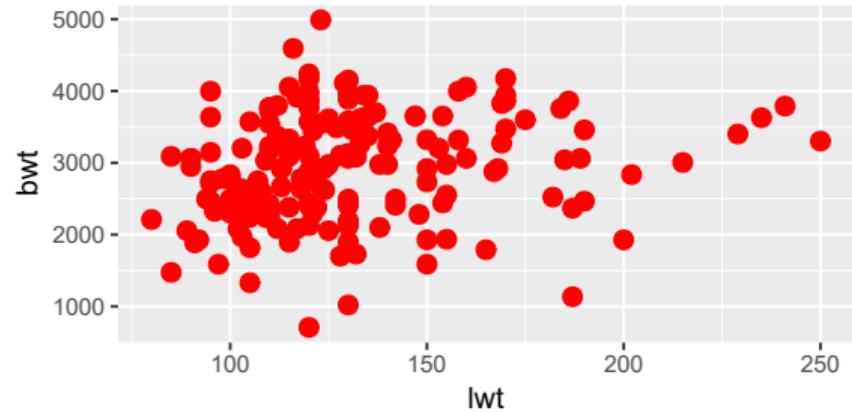
Изменим цвет точек

```
ggplot(baby, aes(x = lwt, y = bwt)) + geom_point(colour = 'red')
```



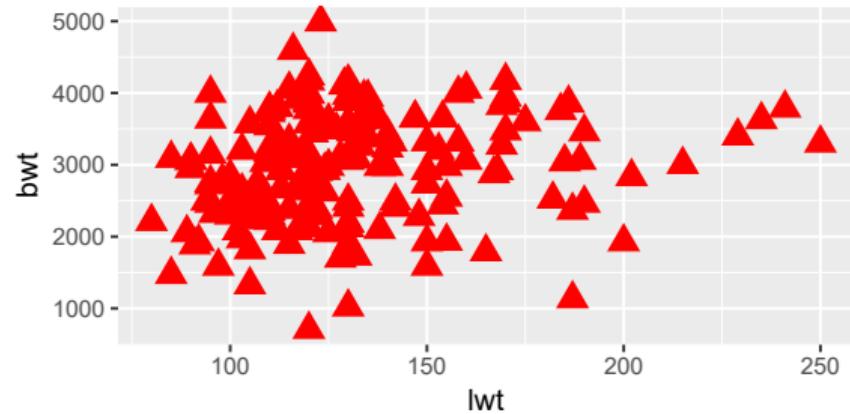
Изменим размер точек

```
ggplot(baby, aes(x = lwt, y = bwt)) + geom_point(colour = 'red', size = 3)
```



Изменим форму точек

```
ggplot(baby, aes(x = lwt, y = bwt)) + geom_point(colour = 'red', size = 4, shape = 17)
```



Три главные части графика в системе `ggplot2`

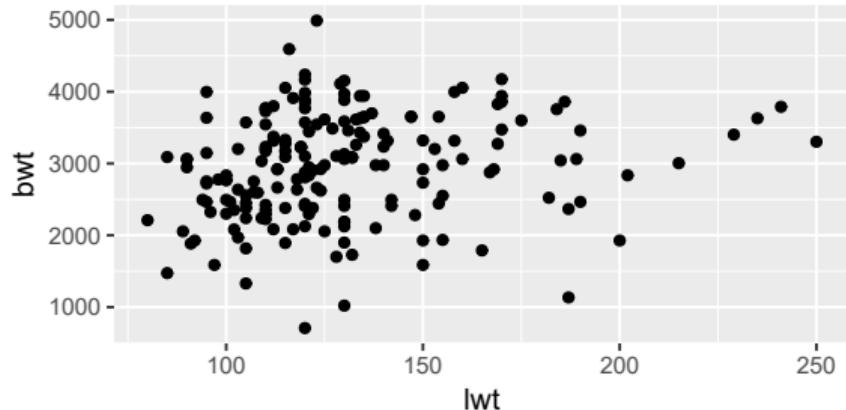
- Data — **данные**, которые необходимо визуализировать. Обязательно должны быть представлены в виде датафрейма (`tidy data!`).
- Aesthetics — эстетики, **что** и из данных мы визуализируем и **чем** отражаем значения (координатами на осях ОХ и ОУ, цветовыми градациями, формой, размером и т.п.).
- Geom - геометрические фигуры, с помощью которых мы визуализируем результаты (точки, столбцы, линии и т.п.).

Эстетики

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

Какая еще информация скрыта в датасете?



На графике приведена лишь
информация о весе матери (ось ОХ) и
весе ребенка (ось ОY).

То, что мы отражаем (эстетики),
описывает функция
`aes(x = lwt, y = bwt)`.

В датасете есть еще информация:

- Статус матери в отношении курения: курящая или некурящая.
- Раса матери
- Количество посещений врача, и т.п.

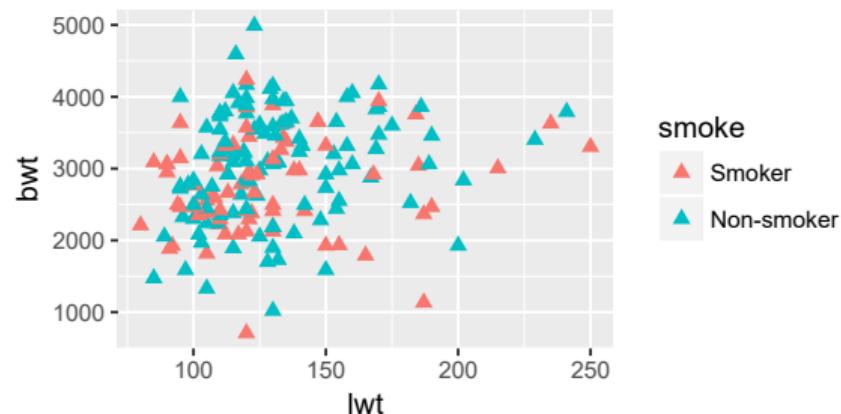
Подготовим данные для более наглядной визуализации

```
# Переименуем градации дискретных факторов
baby$smoke <- factor(baby$smoke,
                       levels = c(1, 0),
                       labels = c('Smoker', 'Non-smoker'))
baby$race <- factor(baby$race,
                      levels = c(1, 2, 3),
                      labels = c('White', 'Black', 'Other'))
```

Вводим в график эстетику colour

Изобразим разным цветом точки для курящих и некурящих матерей (эту информацию маркирует переменная baby\$smoke)

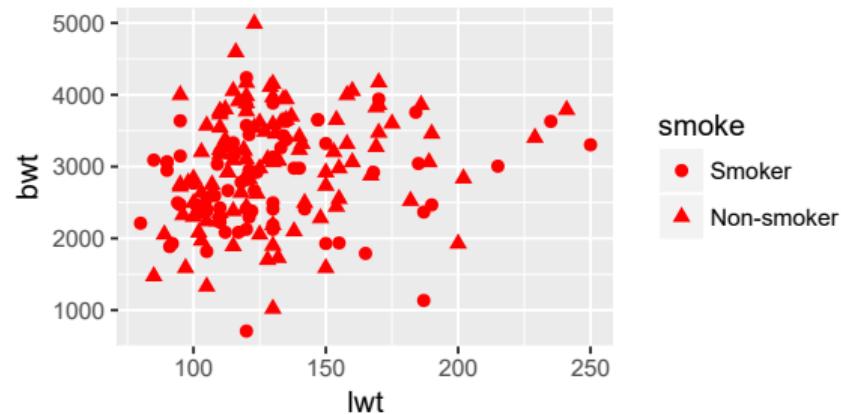
```
ggplot(baby, aes(x = lwt, y = bwt, colour = smoke)) +  
  geom_point(size = 2, shape = 17)  
# Важно! Вся информация о цвете точек теперь должна быть сосредоточена в функции aes()
```



Вводим в график эстетику shape

Отразим ту же информацию, но с помощью разных форм точек для курящих и некурящих матерей.

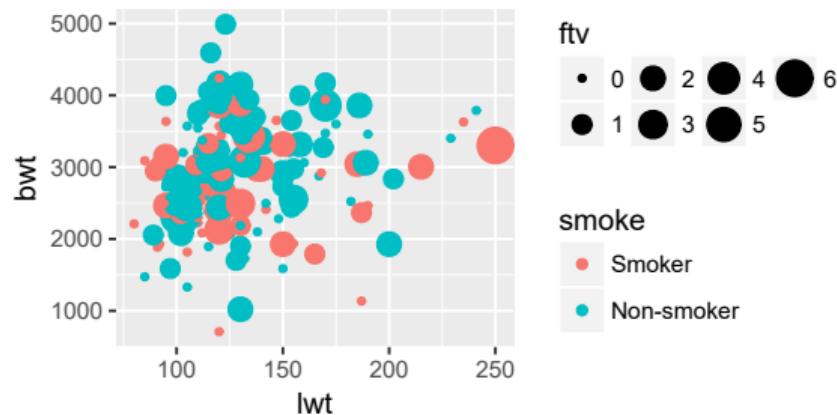
```
ggplot(baby, aes(x = lwt, y = bwt, shape = smoke)) +  
  geom_point(size = 2, colour = 'red')  
# Важно! Вся информация о форме точек теперь должна быть сосредоточена в функции aes()
```



Вводим еще одну эстетику – size

Отразим размером точки информацию о том сколько раз в первом триместре женщина посещала врача (переменная baby\$ftv)

```
ggplot(baby, aes(x = lwt, y = bwt, colour = smoke, size = ftv)) +  
  geom_point() + guides(size = guide_legend(ncol = 4))
```



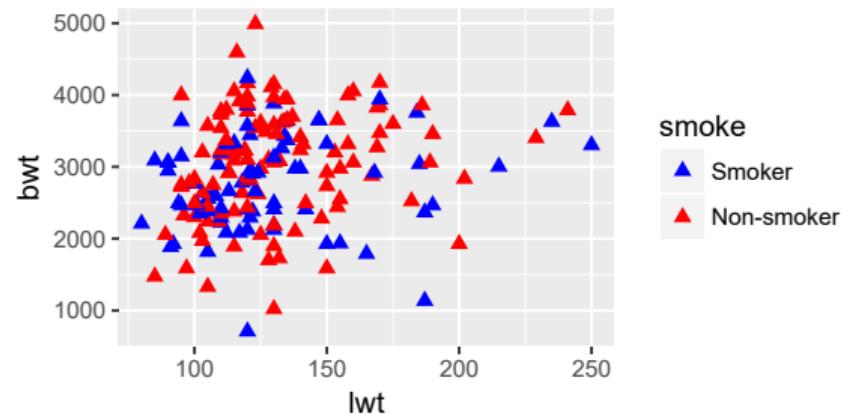
Управление эстетиками. Шкалы

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

Меняем цветовые обозначения эстетики colour

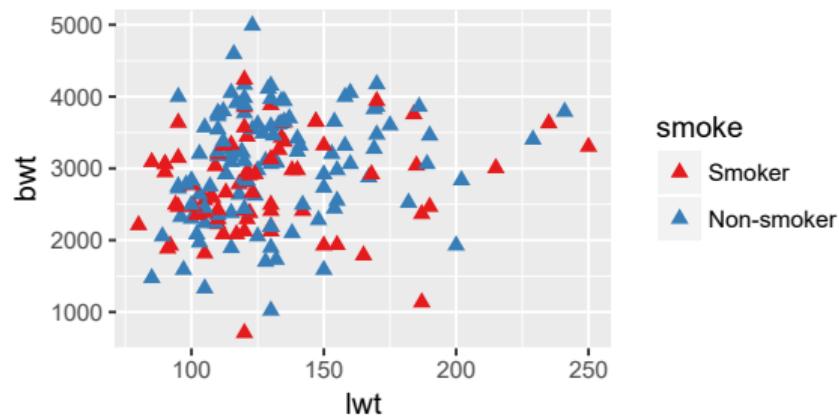
```
ggplot(baby, aes(x = lwt, y = bwt, colour = smoke)) +  
  geom_point(size = 2, shape = 17) +  
  scale_colour_manual(values = c('blue', 'red'))
```



С помощью функций семейства `scale_` можно регулировать форму, цвет заливки, размер и т.д.

Цветовая схема Синтии Брюэр (Cynthia A. Brewer)

```
ggplot(baby, aes(x = lwt, y = bwt, colour = smoke)) +  
  geom_point(size = 2, shape = 17) +  
  scale_colour_brewer(palette = 'Set1')
```



- Эта схема разработана
в ходе исследования,
посвященного сочетаемости
цветов.
- Есть много встроенных
палитр для визуализации
как дискретных, так и
непрерывных данных.
- Учитывает права 8-10%
мужчин, несущих в
Х-хромосоме мутантные
гены

Фасетирование

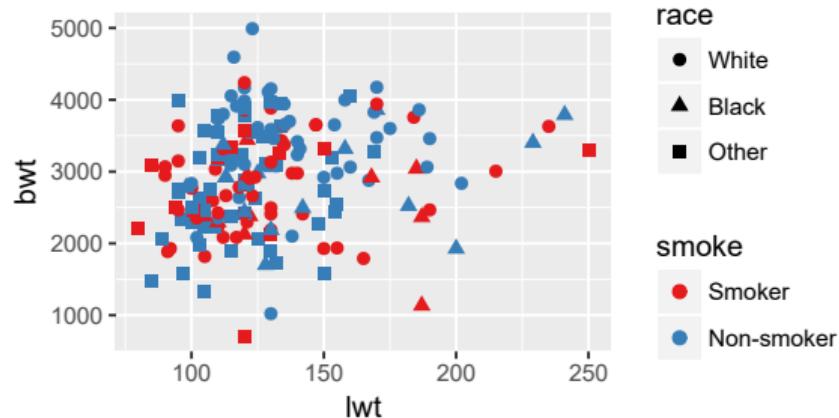
В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

Проблема большого количества эстетик

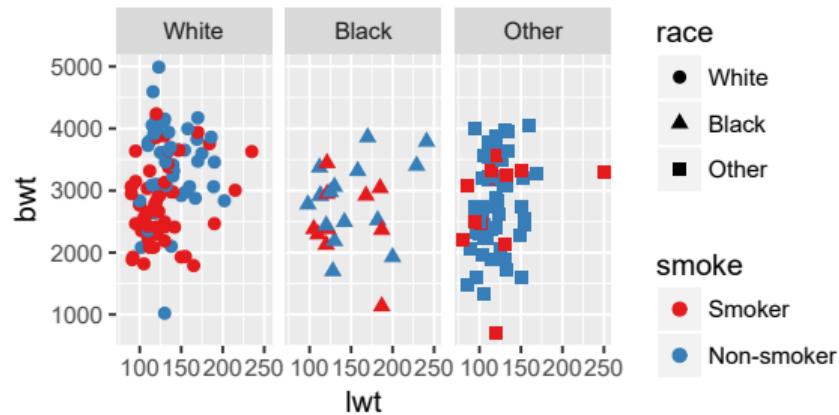
При большом количестве эстетик график может не читаться

```
ggplot(baby, aes(x = lwt, y = bwt, colour = smoke, shape = race)) +  
  geom_point(size = 2) +  
  scale_colour_brewer(palette = 'Set1')
```



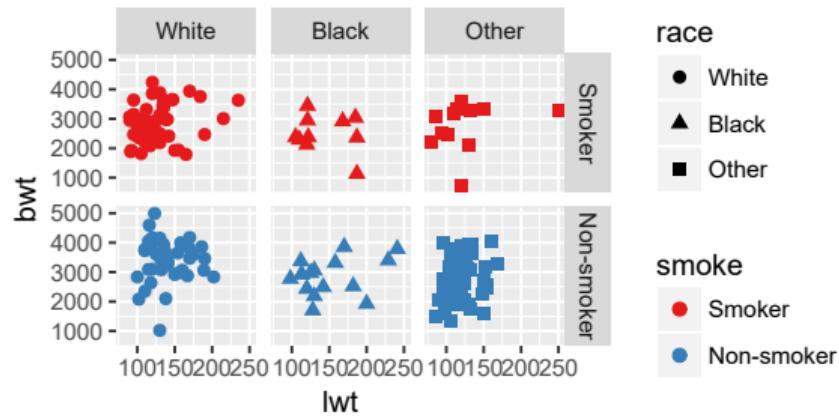
Одна фасетирующая переменная

```
ggplot(baby, aes(x = lwt, y = bwt, colour = smoke, shape = race)) +  
  geom_point(size = 2) +  
  scale_colour_brewer(palette = 'Set1') +  
  facet_wrap(~race, nrow = 1)
```



Две фасетирующие переменные

```
ggplot(baby, aes(x = lwt, y = bwt, colour = smoke, shape = race)) +  
  geom_point(size = 2) +  
  scale_colour_brewer(palette = 'Set1') +  
  facet_grid(smoke~race)
```



Сохранение графиков в переменные

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

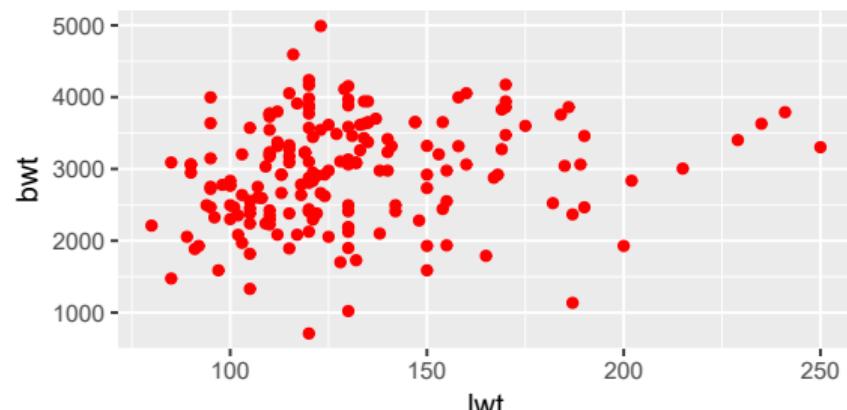
Пошаговое наложение элементов на переменную

Создаем переменную с базовым слоем

```
My_plot <- ggplot(baby, aes(x = lwt, y = bwt))
```

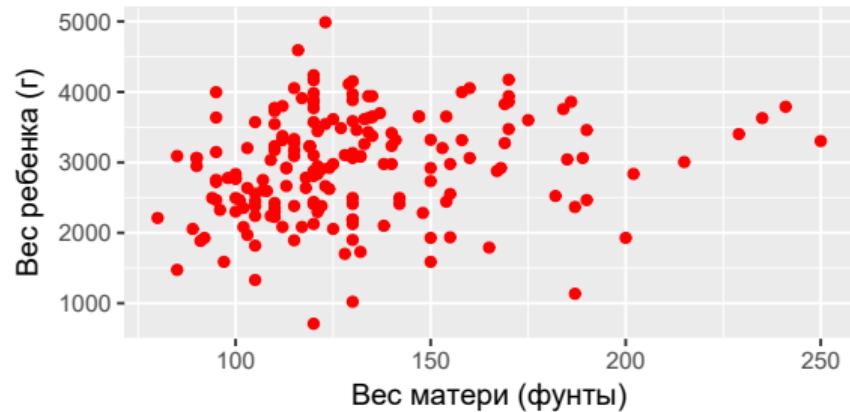
Добавляем новые слои

```
My_plot_2 <- My_plot + geom_point(colour = 'red')  
My_plot_2
```



Меняем подписи осей

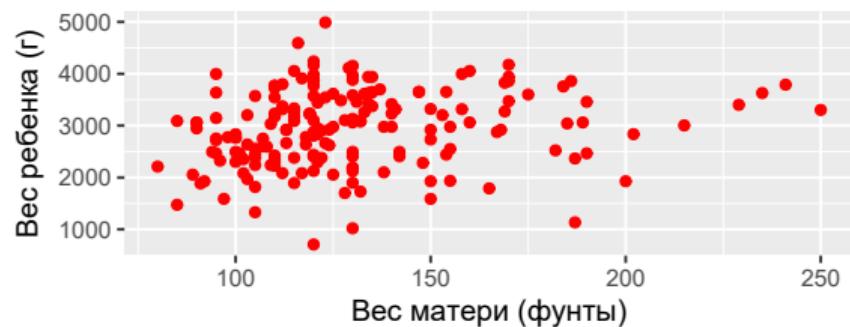
```
My_plot_3 <- My_plot_2 + labs(x = 'Вес матери (фунты)', y = 'Вес ребенка (г)')  
My_plot_3
```



Добавляем заголовок графика

```
My_plot_4 <- My_plot_3 + ggtitle('Соотношение веса матери и \песа новорожденного ребенка')  
My_plot_4
```

Соотношение веса матери и
веса новорожденного ребенка



Темы оформления в ggplot2

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

Требований много, а `ggplot2` один...

В разных журналах, ВУЗах и прочих коллективах требования к оформлению графиков разные.

Вкусы у авторов тоже разные.

Для регуляции внешнего вида графика используется настройка **темы**.

Некоторые встроенные темы в ggplot2

```
Them_1 <- My_plot_2 + theme_gray() + ggtitle('theme_gray')
```

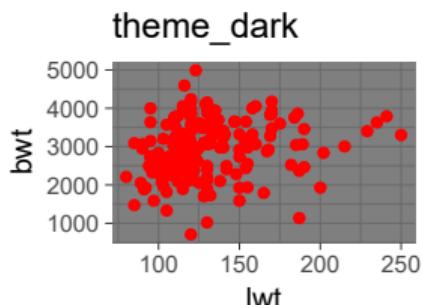
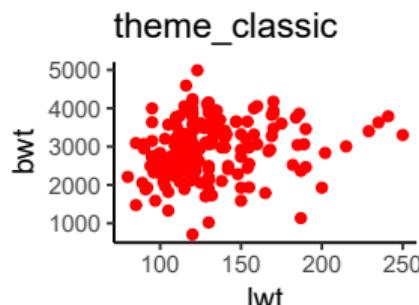
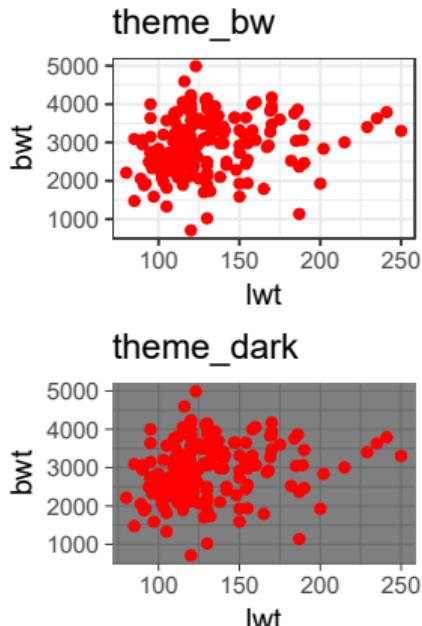
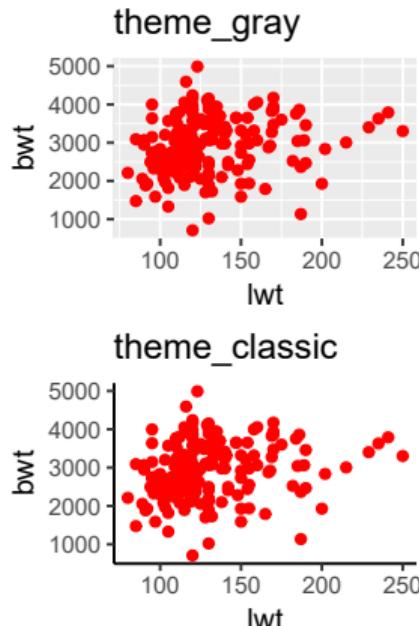
```
Them_2 <- My_plot_2 + theme_bw() + ggtitle('theme_bw')
```

```
Them_3 <- My_plot_2 + theme_classic() + ggtitle('theme_classic')
```

```
Them_4 <- My_plot_2 + theme_dark() + ggtitle('theme_dark')
```

Некоторые встроенные темы в ggplot2

```
library(gridExtra)
grid.arrange(Them_1, Them_2, Them_3, Them_4, ncol = 2)
```



Управление элементами графика

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

Функция theme()

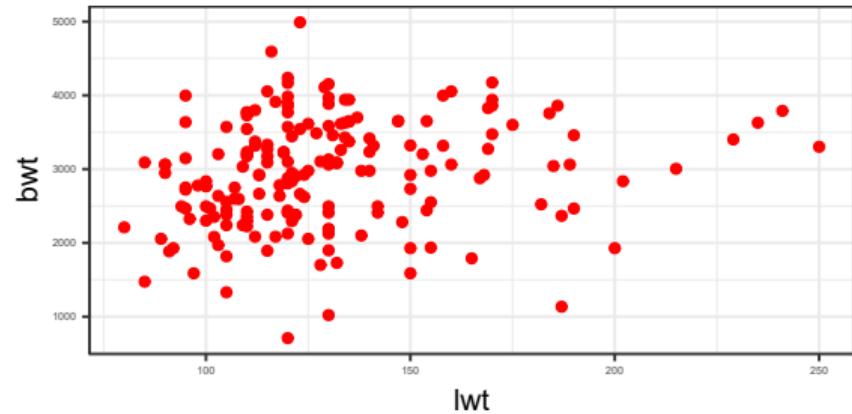
Эта функция регулирует характеристики всех элементов графика.

Наиболее часто приходится регулировать:

- форму осей
- фон графика
- формат подписи осей
- формат цифровой информации на осях.

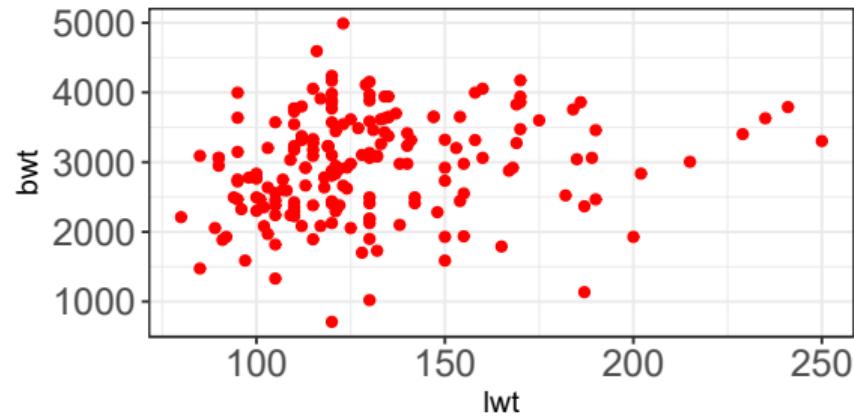
Уменьшаем размер чисел на осях графика

```
My_plot_2 + theme_bw() + theme(axis.text = element_text(size = 4))
```



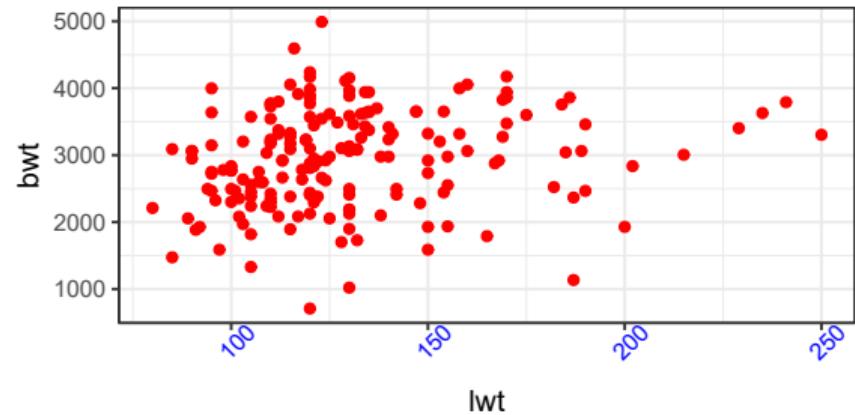
Увеличиваем размер чисел на осях графика

```
My_plot_2 + theme_bw() + theme(axis.text = element_text(size = 14))
```



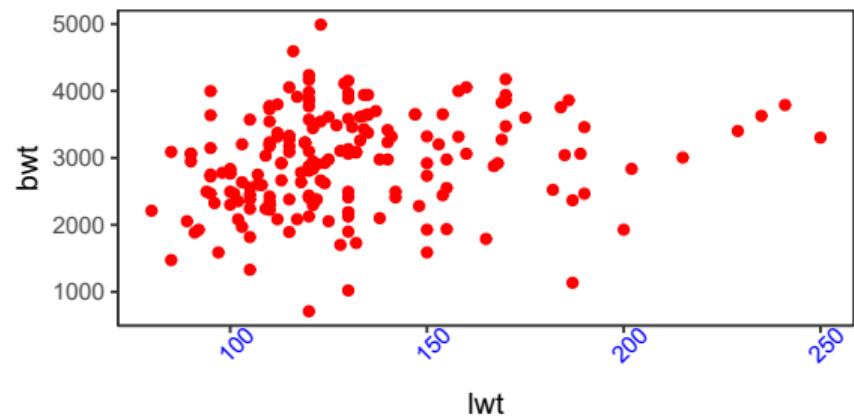
Регулируем угол наклона чисел на осях графика

```
My_plot_2 + theme_bw() + theme(axis.text.x = element_text(angle = 45, colour = 'blue'))
```



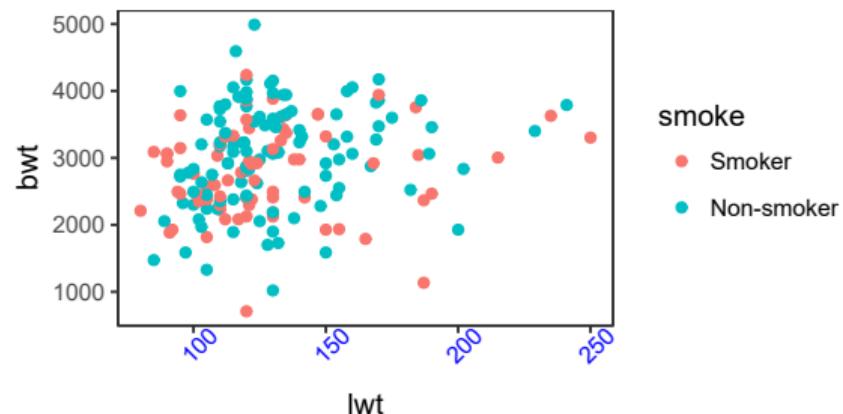
Убираем сетку внутри графика

```
My_plot_2 + theme_bw() +  
  theme(axis.text.x = element_text(angle = 45, colour = 'blue'),  
        panel.grid = element_blank())
```



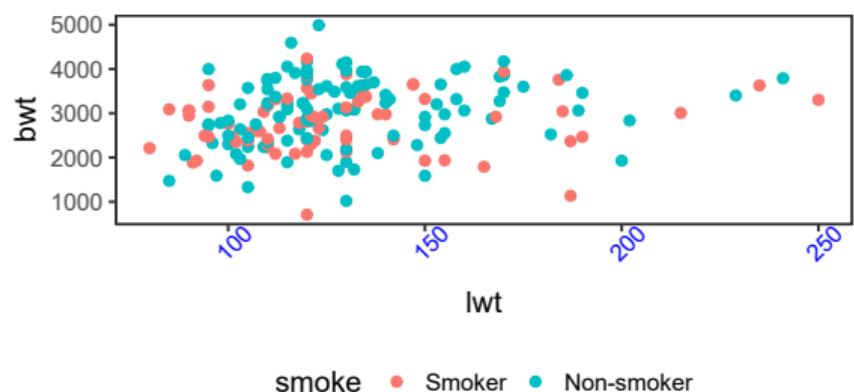
Меняем положение легенды

```
ggplot(baby, aes(x = lwt, y = bwt, colour = smoke)) +  
  geom_point() +  
  theme_bw() +  
  theme(axis.text.x = element_text(angle = 45, colour = 'blue'),  
        panel.grid = element_blank())
```



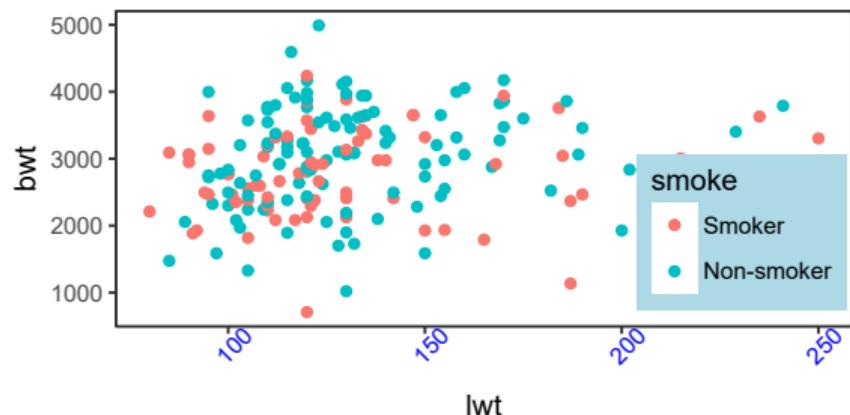
Меняем положение легенды

```
ggplot(baby, aes(x = lwt, y = bwt, colour = smoke)) +  
  geom_point() +  
  theme_bw() +  
  theme(axis.text.x = element_text(angle = 45, colour = 'blue'),  
        panel.grid = element_blank(),  
        legend.position = 'bottom')
```



Меняем положение легенды

```
ggplot(baby, aes(x = lwt, y = bwt, colour = smoke)) +  
  geom_point() +  
  theme_bw() +  
  theme(axis.text.x = element_text(angle = 45, colour = 'blue'),  
        panel.grid = element_blank(),  
        legend.position = c(0.85, 0.3),  
        legend.background = element_rect(fill = 'lightblue'))
```

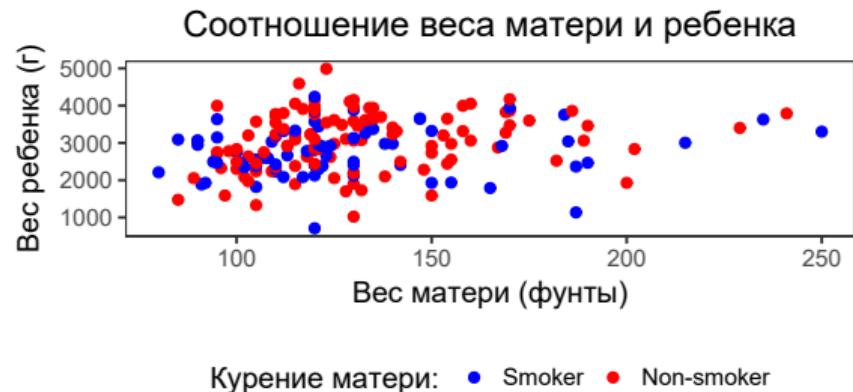


Создаем финальный рисунок

```
My_final_plot <- ggplot(baby, aes(x = lwt, y = bwt)) + # задаем базовый слой
  geom_point(aes(colour = smoke)) + # предсталяем результаты в виде точек
  labs(x = 'Вес матери (фунты)', # подписываем ось ОХ
       y = 'Вес ребенка (г)', # подписываем ось ОY
       colour = 'Курение матери: ') + # подписываем заголовок легенды
  ggtitle('Соотношение веса матери и ребенка') + # заголовок
  scale_colour_manual(values = c('blue', 'red')) + # меняем цвет для эстетики colour
  theme_bw() + # устанавливаем тему
  theme(panel.grid = element_blank(), # убираем сетку
        legend.position = 'bottom', # указываем положение легенды
        plot.title = element_text(hjust = 0.5)) # выравниваем заголовок по центру
```

Итоговый рисунок

My_final_plot



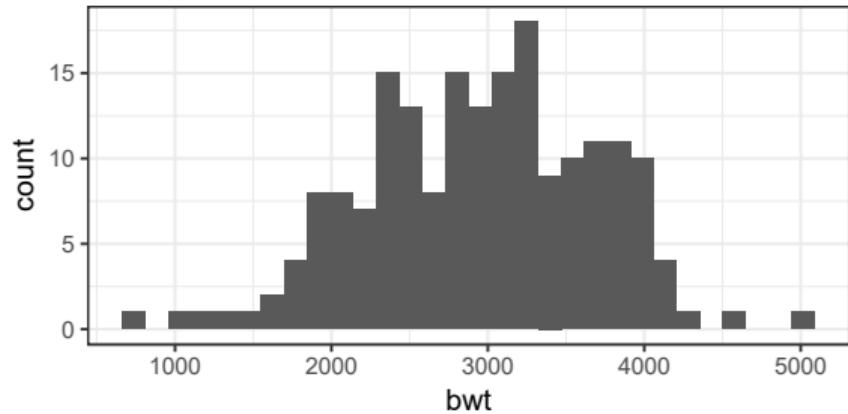
Визуализация частотных распределений

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

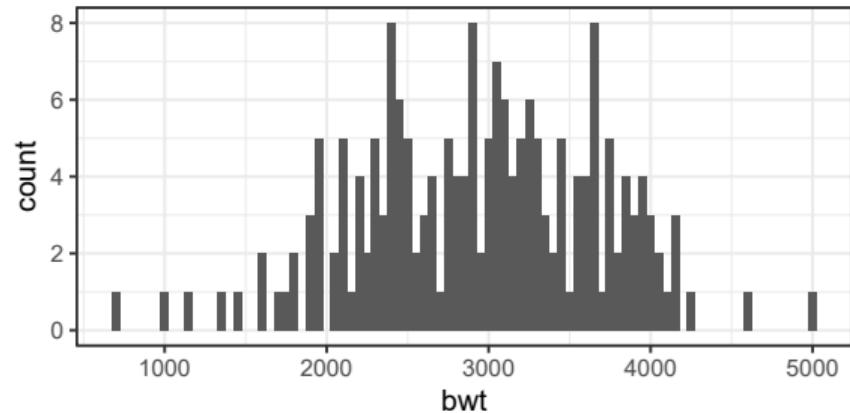
Частотное распределение

```
ggplot(baby, aes(x = bwt)) + geom_histogram() + theme_bw()
```



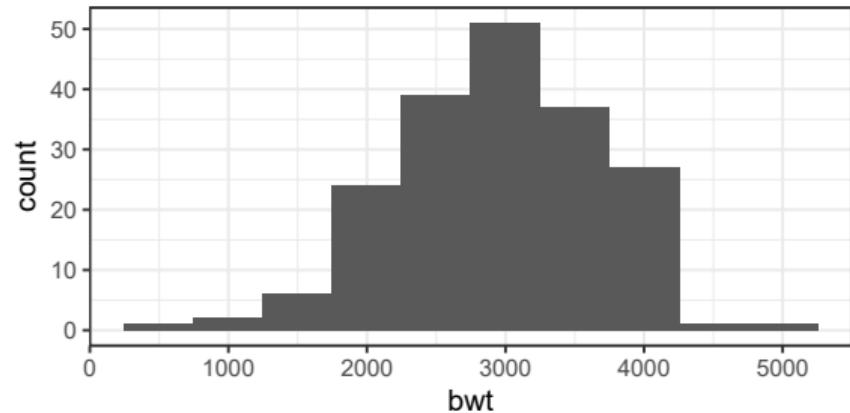
Меняем классовый шаг

```
ggplot(baby, aes(x = bwt)) + geom_histogram(binwidth = 50) + theme_bw()
```



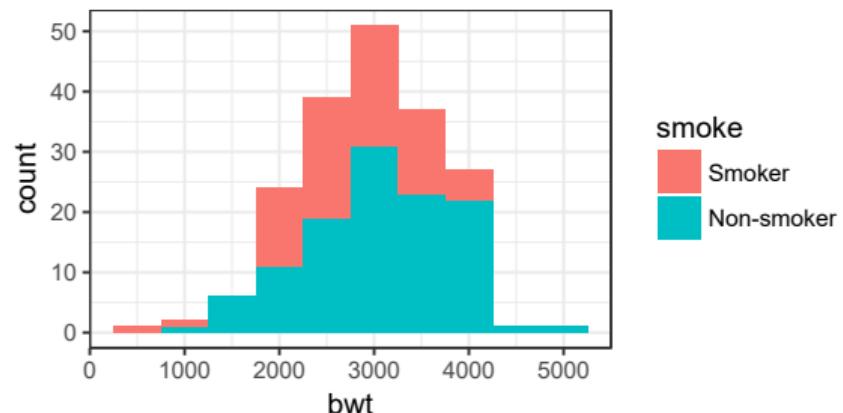
Меняем классовый шаг

```
ggplot(baby, aes(x = bwt)) + geom_histogram(binwidth = 500) + theme_bw()
```



Применяем эстетику fill

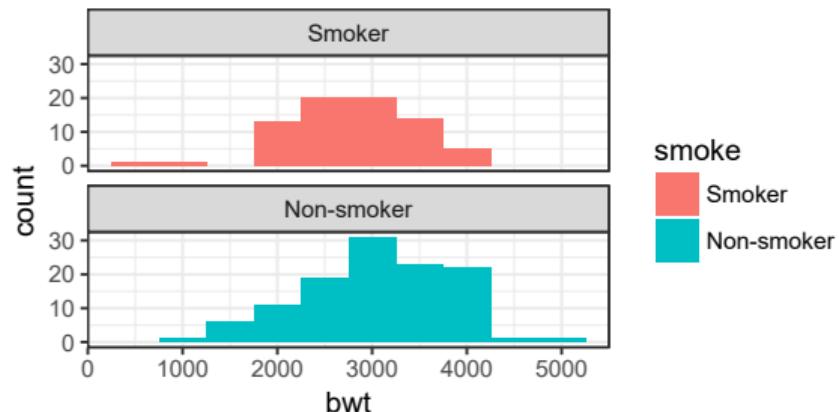
```
ggplot(baby, aes(x = bwt)) + geom_histogram(binwidth = 500, aes(fill = smoke)) +  
  theme_bw()
```



Читать неудобно!

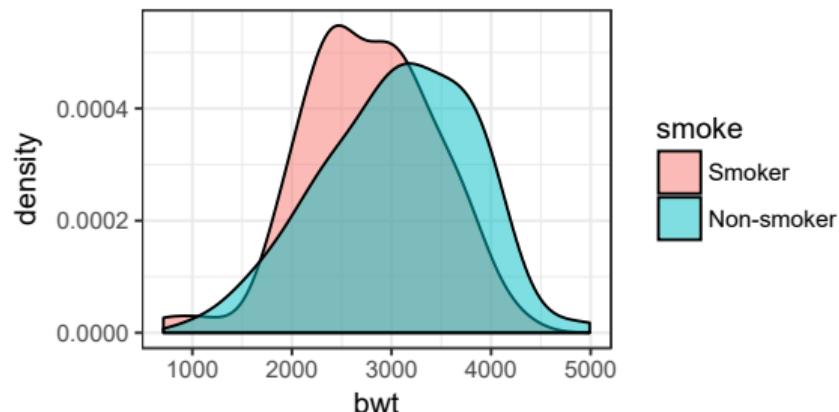
Фасетируем рисунок

```
ggplot(baby, aes(x = bwt)) + geom_histogram(binwidth = 500, aes(fill = smoke)) +  
  theme_bw() +  
  facet_wrap(~smoke, nrow = 2)
```



Использование анализа, основанного на оценке ядерной плотности (kernel density)

```
ggplot(baby, aes(x = bwt)) +  
  geom_density(aes(fill = smoke), alpha = 0.5) +  
  theme_bw()
```



Становится видна некоторая тенденция.

geom_violin()

```
ggplot(baby, aes(x = smoke, y = bwt)) +  
  geom_violin(aes(fill = smoke)) +  
  theme_bw()
```



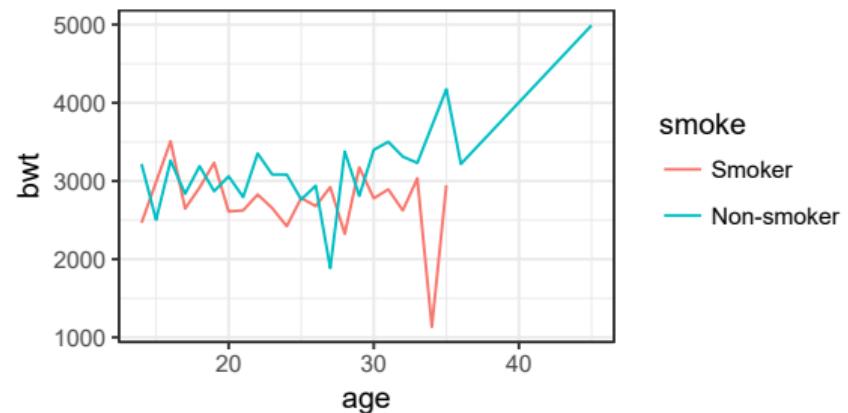
Визуализация данных с простейшей статистической обработкой

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

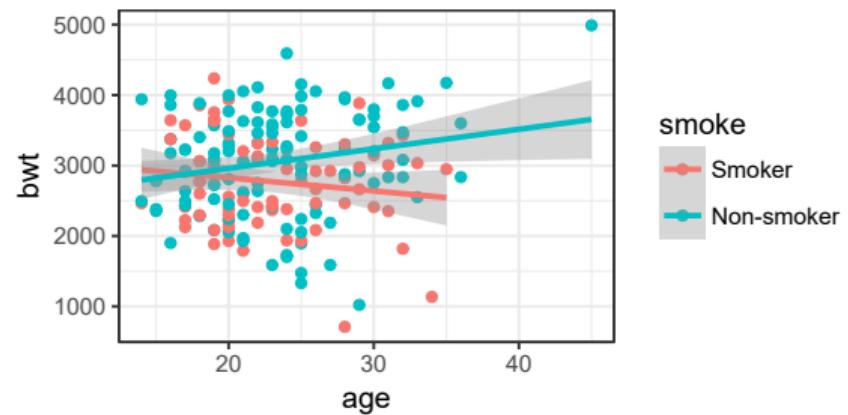
Средний вес ребенка у матерей разного возраста

```
ggplot(baby, aes(x = age, y = bwt, colour = smoke)) +  
  stat_summary(fun.y = mean, geom = 'line') +  
  theme_bw()
```



Тенденции зависимости веса ребенка от возраста матери

```
ggplot(baby, aes(x = age, y = bwt, colour = smoke)) + geom_point() +  
  stat_smooth(method = 'lm') +  
  theme_bw()
```



Усы, отражающие интервалы

```
ggplot(baby, aes(x = smoke, y = bwt)) +  
  stat_summary(fun.y = mean, geom = 'bar', fill = 'gray', colour = 'black') +  
  stat_summary(fun.data = mean_se, geom = 'errorbar', width = 0.2) +  
  theme_bw()
```



Сохранение графиков в виде файлов

В.М. Хайтов, к.б.н.

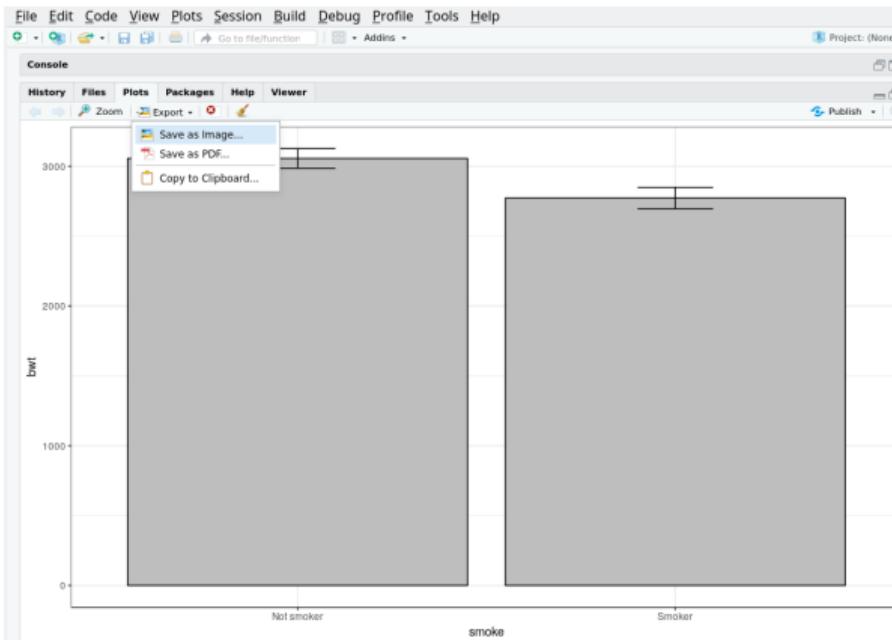
Биологический факультет, СПбГУ

Сохранение графика в файл с помощью RStudio

The screenshot shows the RStudio interface with the following components:

- Code pane:** Displays R code for creating a ggplot. The code defines two facets: "Not smoker" and "Smoker". Each facet contains a bar plot where the mean birth weight (bwt) is represented by a gray bar, and the standard error of the mean is shown as a black error bar.
- Console pane:** Shows the R session history, including the command to load the baby dataset and the ggplot code.
- Plots pane:** Displays the resulting ggplot showing birth weight (bwt) versus smoking status (smoke).
- Environment pane:** Shows the global environment with one dataset named "baby" containing 189 observations and 10 variables.

Сохранение графика в файл с помощью RStudio



Сохранение графика в файл с помощью средств ggplot2

Создаем график

```
Plot_to_save <- ggplot(baby, aes(x = smoke, y = bwt)) +  
  stat_summary(fun.y = 'mean', geom = 'bar', fill = 'gray', colour = 'black') +  
  stat_summary(fun.data = mean_se, geom = 'errorbar', width = 0.2) +  
  theme_bw()
```

Сохранение графика в файл с помощью средств `ggplot2`

Сохраняем график в растровом формате

```
ggsave('c:/My_draw.jpg', Plt_to_save) # сохраняем в растровом формате .jpg  
ggsave('c:/My_draw.png', Plot_to_save) # сохраняем в растровом формате .png
```

Сохранение графика в файл с помощью средств `ggplot2`

Сохраняем график в векторном формате

```
ggsave('c:/My_draw.pdf', Plot_to_save) # сохраняем в формате .pdf  
ggsave('c:/My_draw.wmf', Plot_to_save) # сохраняем в векторном формате .wmf  
ggsave('c:/My_draw.eps', Plot_to_save) # сохраняем в векторном формате .eps
```

Литературное программирование

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

Для исполнителей



Вы создали презентацию, написали текст статьи или отчета.

Начальнику не понравилась форма графиков, цвет заливки, величина шрифта и т. п...

В дополнение коллеги принели вам исправленные файлы с первичными данными...

Все документы надо срочно переделывать...

Выход из этого кошмара — создание RMD-документа.

Для пользователей разных платформ



Ваши коллеги работают под Linux и Windows, а вы предпочитаете Mac OS.

Продукт вашего совместного творчества должен читаться на любом компьютере.

Основа плодотворного сотрудничества — работа с RMD-документами.

Для исследователей



Вы ученый-исследователь.

В статьях, которые вы читаете, видна лишь вершина айсберга: характер данных, процедуры анализа, способ построения таблиц и графиков остаются в тени.

Вас не оставляет чувство сомнения в правоте автора статьи, так как нет возможности воспроизвести те действия с данными, которые произвел автор.

Выход из этого — стать сторонником идеи **Reproducible research**

Reproducible research

Воспроизводимое исследование — это анализ, который может быть воспроизведен на основе тех же данных любым другим исследователем.

Воспроизводимое исследование не надо путать с воспроизводимостью результатов!

Reproducible research

Критерии воспроизводимого исследования

1. Исходные данные (raw data) доступны любому читателю-потребителю (в идеале, публично доступны).
2. Все методы анализа полностью описаны.
3. Весь процесс анализа исходных данных (от чтения данных до финального результата) полностью описан и сохранен в виде некоторого текста (кода анализа).

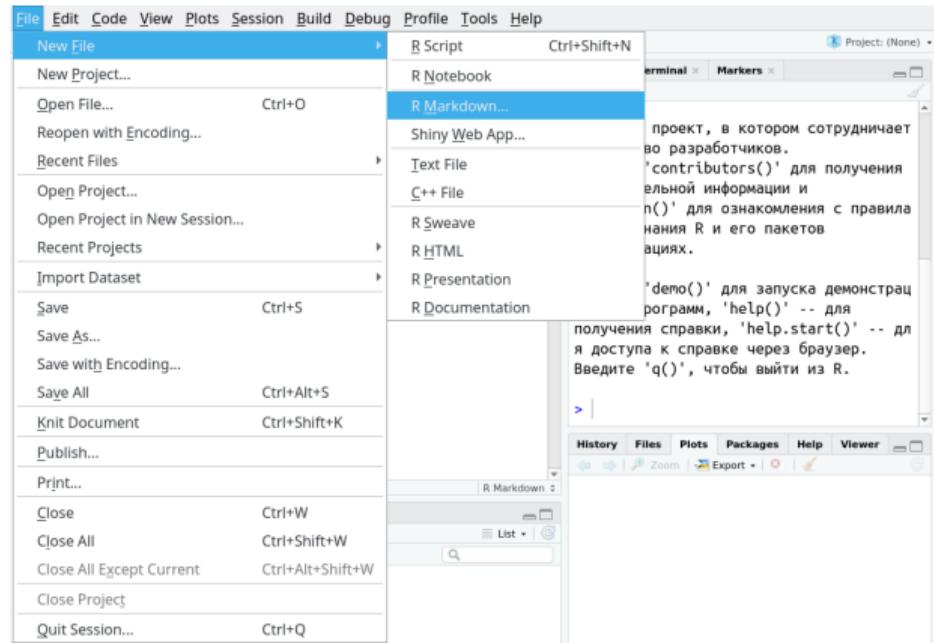
Лучшее средство для представления своих результатов, как воспроизводимого исследования,
— это создание RMD-документов.

Создание RMD-документа

В.М. Хайтов, к.б.н.

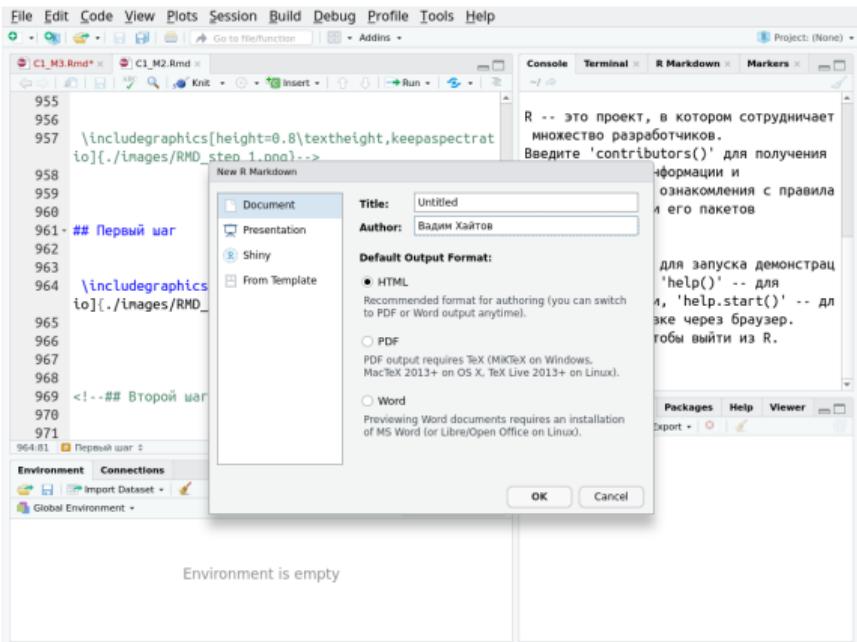
Биологический факультет, СПбГУ

R Markdown



RMD-документы — это документы, созданные с использованием языка разметки R Markdown.

Документ или презентация



Внешний вид RMD-документа

The screenshot shows the RStudio interface with an R Markdown file open. The code editor displays the following R code:

```
1- ---
2 title: "Untitled"
3 author: "Вадим Хайтов"
4 date: '17 апреля 2018 г.'
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple
formatting syntax for authoring HTML, PDF, and MS Word
documents. For more details on using R Markdown see
<http://rmarkdown.rstudio.com>.
15
16 When you click the **Knit** button a document will be
2:1
```

The R Markdown pane shows the generated HTML content:

R -- это проект, в котором сотрудничает множество разработчиков.
Введите 'contributors()' для получения дополнительной информации и 'citation()' для ознакомления с правилами упоминания R и его пакетов в публикациях.

Введите 'demo()' для запуска демонстрационных программ, 'help()' -- для получения справки, 'help.start()' -- для доступа к справке через браузер.
Введите 'q()', чтобы выйти из R.

The Environment pane shows:

Environment Connections

Global Environment

Environment is empty

Создание итогового документа

Текст RMD-документа передается функции `knit()` из пакета `knitr`.

Эта функция создает итоговый документ в том формате, который указан в заголовке RMD-документа (`html`, `pdf`, `docx`).

Чанки

The screenshot shows the RStudio interface with an R Markdown file open. The code editor pane contains the following R code:

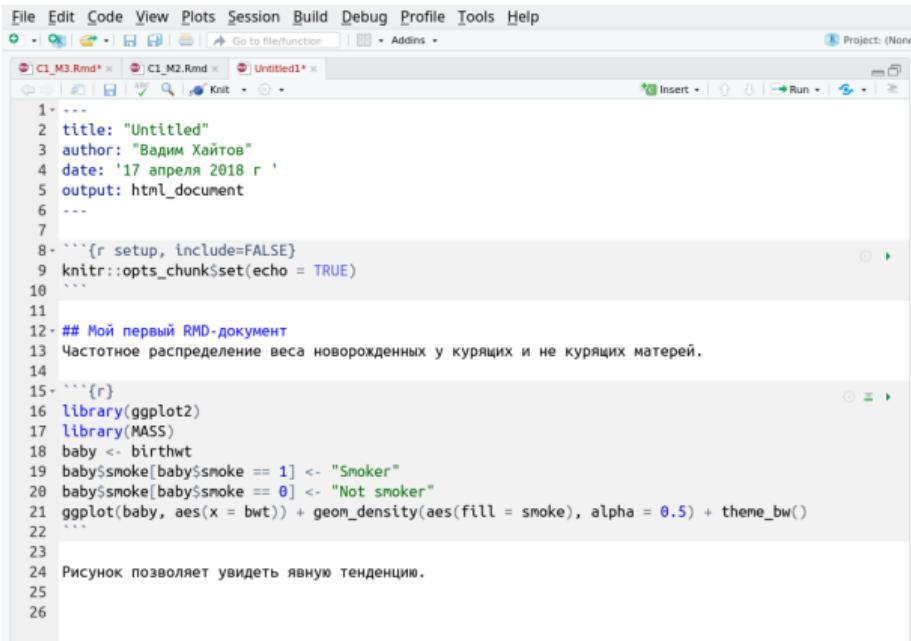
```
1 - ...
2 title: "Untitled"
3 author: "Вадим Хайтов"
4 date: '17 апреля 2018 г.'
5 output: html_document
6 ---
7
8 ````{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple
15 formatting syntax for authoring HTML, PDF, and MS Word
16 documents. For more details on using R Markdown see
<http://rmarkdown.rstudio.com>.
17
18 When you click the **Knit** button a document will be
2:1
```

The right-hand pane displays the R console output, which includes a welcome message for R and its documentation, and the environment pane at the bottom shows that the environment is currently empty.

Чанк — это исполняемый код, встроенный в текст.

Вставка чанка в RStudio: Ctrl+Alt+I

Пример RMD-документа



The screenshot shows the RStudio IDE interface with the following details:

- Menu Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for file operations like Open, Save, Print, and a search bar labeled "Go to file/function".
- Project Bar:** Shows "Project: (None)".
- File List:** Displays three files: C1_M3.Rmd*, C1_M2.Rmd*, and Untitled1*.
- Code Editor:** The main area contains R code for an RMD document. The code includes metadata (title, author, date, output), a header section, a plot section, and a note at the end.
- Run Buttons:** A toolbar at the bottom right includes "Insert", "Run", and other execution options.

```
1 - ---
2 title: "Untitled"
3 author: "Вадим Хайтов"
4 date: '17 апреля 2018 г '
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## Мой первый RMD-документ
13 Частотное распределение веса новорожденных у курящих и не курящих матерей.
14
15 ```{r}
16 library(ggplot2)
17 library(MASS)
18 baby <- birthwt
19 baby$smoke[baby$smoke == 1] <- "Smoker"
20 baby$smoke[baby$smoke == 0] <- "Not smoker"
21 ggplot(baby, aes(x = bw)) + geom_density(aes(fill = smoke), alpha = 0.5) + theme_bw()
22 ```
23
24 Рисунок позволяет увидеть явную тенденцию.
25
26
```

Результат работы knitr в окне браузера

Untitled

Вадим Хайтов

16 апреля 2018 г.

Мой первый RMD-документ

Частотное распределение веса новорожденных у курящих и некурящих матерей.

```
library(ggplot2)
library(MASS)
baby <- birthwt #Чтение встроенных данных
baby$smoke[baby$smoke == 1] <- "Smoker"
baby$smoke[baby$smoke == 0] <- "Non smoker"
ggplot(baby, aes(x = bwt)) + geom_density(aes(fill = smoke), alpha = 0.5) + theme_bw()
```

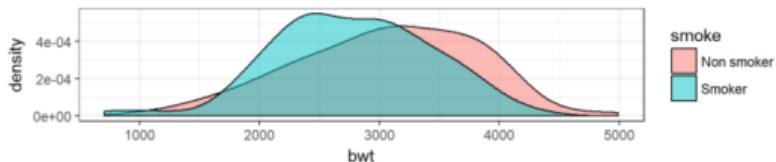


Рисунок позволяет увидеть явную тенденцию.

В результате работы knitr появляется html-файл, который можно просмотреть в обычном браузере.

Параметры чанков

The screenshot shows the RStudio interface with the following details:

- File Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Standard RStudio tools like Open, Save, Run, etc.
- Project:** Project: (None)
- Text Editor:** The code editor contains the following R Markdown code:

```
1 - ---
2 title: "Untitled"
3 author: "Вадим Хайтов"
4 date: '17 апреля 2018 г '
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## Мой первый RMD-документ
13 Частотное распределение веса новорожденных у курящих и не курящих матерей.
14
15 ```{r echo=FALSE, fig.height=1.5, fig.align='center', eval=TRUE}
16 library(ggplot2)
17 library(MASS)
18 baby <- birthwt
19 baby$smoke[baby$smoke == 1] <- "Smoker"
20 baby$smoke[baby$smoke == 0] <- "Not smoker"
21 ggplot(baby, aes(x = bwt)) + geom_density(aes(fill = smoke), alpha = 0.5) + theme_bw()
22 ...
23
24 Рисунок позволяет увидеть явную тенденцию.
25
26
```

Результат работы knitr с новыми параметрами чанка

Untitled

Вадим Хайтов

16 апреля 2018 г

Мой первый RMD-документ

Частотное распределение веса новорожденных у курящих и некурящих матерей.

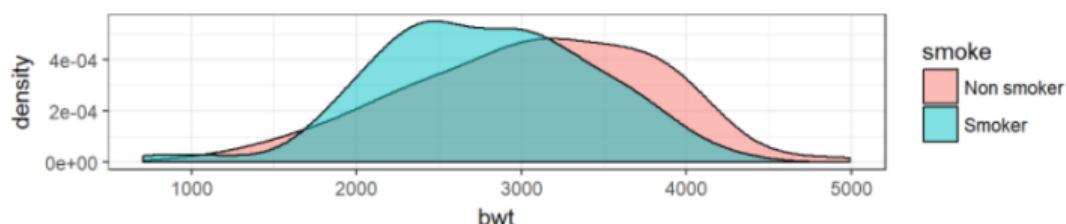


Рисунок позволяет увидеть явную тенденцию.

Наиболее важные параметры чанка

- echo — Приводить ли код из чанка в документе, который получится в результате работы knitr
- eval — Запускать ли код, приведенный в чанке
- fig. — Набор параметров для регуляции положения и размера рисунков
- warning — Выводить ли в итоговый документ предупреждающие сообщения функций

Что мы знаем и что будет дальше

В.М. Хайтов, к.б.н.

Биологический факультет, СПбГУ

Что мы знаем?

При помощи пакета `ggplot2` можно создавать графики практически любой сложности.

Результаты анализа удобно представлять в виде `.Rmd` документов.

Создание отчетов в `.Rmd` поможет сделать результаты воспроизводимыми и легко адаптироваться к изменениям в исходных данных или к новым данным.

Что будет дальше?

Сейчас мы разобрали только самые основы пакета `ggplot2`. В следующих модулях мы еще не раз будем обращаться к его возможностям для визуализации результатов статистического анализа.

Вы сможете попрактиковаться в создании документов в формате `.Rmd` работая над проектами в курсах нашей специализации.

Используемые источники

<https://www.shutterstock.com/ru/image-photo/child-sticks-bow-greeting-card-mum-526386226>

<https://www.shutterstock.com/image-photo/newly-born-child-gazes-camera-low-47693563>

<https://www.shutterstock.com/image-vector/cartoon-illustration-malfunctioning-little-robot-223250488>

<https://timogrossenbacher.ch/2016/12/beautiful-thematic-maps-with-ggplot2-only/>

<http://www.ggplot2-exts.org/gallery/>

Khaitov V. et al. (2018) Biol. Bull. 234

Hosmer, D.W. and Lemeshow, S. (1989) Applied Logistic Regression. New York: Wiley



Санкт-Петербургский
государственный
университет
www.spbu.ru