

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

гистограмма

фасетизация

анимация

markdown

Визуализация данных: базовые функции R и пакет ggplot2

Г. Мороз

Р для визуализации? Совсем не обязательно...

Взято [отсюда](#). Куча ресурсов, которые скоро устареют.

Matplotlib

Bokeh

- AnyChart
- Chart Tool
- Chart.js
- Chartbuilder
- Chartbuilder 2.0
- ChartGo
- Chiasm
- D3plus
- Datahero
- Datamatic
- Datavisual
- Datawrapper
- Diagrammer

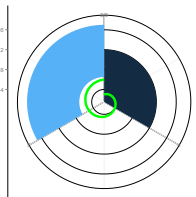
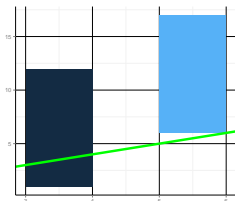
- Diychart
- Dygraphs
- Echarts
- Envision.js
- filtergraph
- Flare
- Google Charts
- Highcharts
- iCharts
- Infogr.am
- JS Charts
- JavaScript InfoVis Toolkit
- Livegap Charts
- Lyra
- Plotly

Processing

- Qlik
- Raw
- Lumira
- Slemma
- Spotfire
- Sprites
- Tableau
- VIDI
- Vega
- Visage
- Vizydrop
- Weave
- Zingchart

элементы визуализации

○ система координат



элементы визуализации

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

гистограмма

фасетизация

анимация

markdown

- позиция
- длина
- форма
- угол
- направление
- размер
- цвет
- ...

Anscombe's quartet

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

гистограмма

фасетизация

анимация

markdown

В работе Anscombe, F. J. (1973). "Graphs in Statistical Analysis" был представлен следующий набор данных:

```
quartet <- read.csv("https://goo.gl/KHQLZI")  
head(quartet)
```

	X1	Y1	X2	Y2	X3	Y3	X4	Y4
1	10	8.04	10	9.14	10	7.46	8	6.58
2	8	6.95	8	8.14	8	6.77	8	5.76
3	13	7.58	13	8.74	13	12.74	8	7.71
4	9	8.81	9	8.77	9	7.11	8	8.84
5	11	8.33	11	9.26	11	7.81	8	8.47
6	14	9.96	14	8.10	14	8.84	8	7.04

Anscombe's quartet: статистика

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

гистограмма

фасетизация

анимация

markdown

```
sapply(quartet[, 1:4*2-1], mean)
```

mean of x

X1	X2	X3	X4
9	9	9	9

```
sapply(quartet[, 1:4*2], mean)
```

mean of y

Y1	Y2	Y3	Y4
7.500909	7.500909	7.500000	7.500909

```
sapply(quartet[, 1:4*2-1], sd)
```

sd of x

X1	X2	X3	X4
3.316625	3.316625	3.316625	3.316625

```
sapply(quartet[, 1:4*2], sd)
```

sd of y

Y1	Y2	Y3	Y4
2.031568	2.031657	2.030424	2.030579

```
mapply(cor, quartet[, 1:4*2-1], quartet[, 1:4*2])
```

correlation

X1	X2	X3	X4
0.8164205	0.8162365	0.8162867	0.8165214

```
mapply(function(x, y) lm(y ~ x)$coefficients,  
quartet[, 1:4*2-1], quartet[, 1:4*2])
```

linear regression

	X1	X2	X3	X4
(Intercept)	3.0000909	3.000909	3.0024545	3.0017273
x	0.5000909	0.500000	0.4997273	0.4999091

презентация доступна: <https://goo.gl/4T1zAr>

Anscombe's quartet: статистика

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

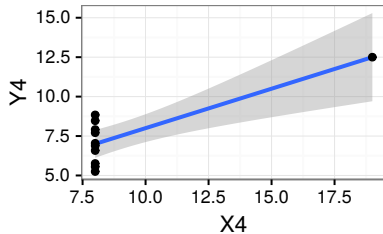
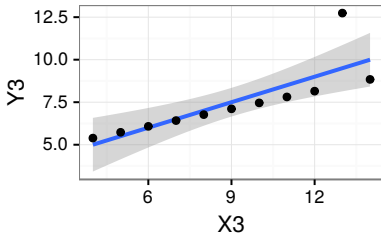
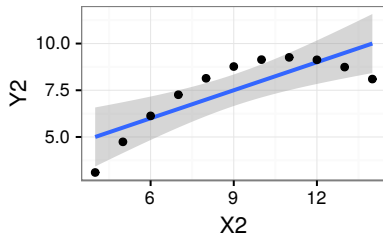
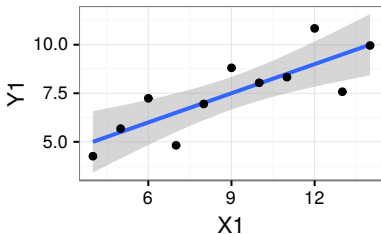
гистограмма

фасетизация

анимация

markdown

Anscombe's quartet



Данные

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

гистограмма

фасетизация

анимация

markdown

В данной презентации все примеры будут приводиться на примере датасета из работы [Chi-kuk 2007] (доступна по ссылке <http://goo.gl/ZjrgaF>). В работе исследовались речь 7 гомосексуальных и 7 гетеросексуальных носителей кантонского диалекта языка юэ. В датасете есть следующие переменные:

- долгота s (`s.duration.ms`)
- долгота гласных (`vowel.duration.ms`)
- среднее значение ЧОТ (`average.f0.Hz`)
- диапазон ЧОТ (`f0.range.Hz`)
- сколько носителей воспринимает говорящего как гомосексуала (`perceived.as.homo`)
- сколько носителей воспринимает говорящего как гетеросексуала (`perceived.as.hetero`)
- ориентация говорящего (`orientation`)
- возраст говорящего (`age`)

презентация доступна: <https://goo.gl/4T1zAr>

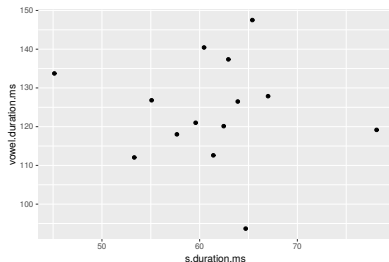
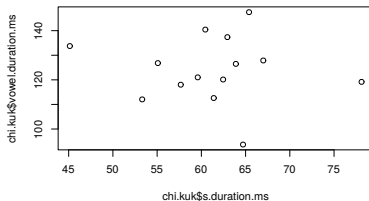
В R визуализация реализована по-разному:

- core R (всегда будет слева)
- библиотекой ggplot2 (всегда будет справа)
- библиотекой lattice (в данной презентации не используется)

Все примеры ggplot2 в данной презентации реализованы в связке с пакетом dplyr. Так что, для того чтобы код из данной презентации работал следует включить библиотеки и скачать датасет:

```
library(tidyverse)
chi.kuk <- read.csv("http://goo.gl/Zjr9aF") # закачивает датасет
```

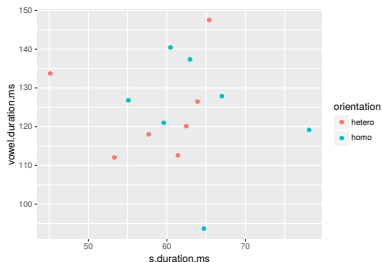
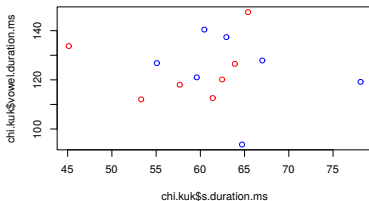
scatterplot



```
# base R
plot(chi.kuk$s.duration.ms, chi.kuk$vowel.duration.ms)
```

```
# dplyr, ggplot2
chi.kuk %>%
  ggplot(aes(s.duration.ms, vowel.duration.ms)) +
  geom_point()
```

scatterplot: цвет



```
# base R
```

```
plot(chi.kuk$s.duration.ms, chi.kuk$vowel.duration.ms,  
     col = c("red", "blue")[chi.kuk$orientation])
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(s.duration.ms, vowel.duration.ms, color = orientation)) +  
  geom_point()
```

scatterplot: форма

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

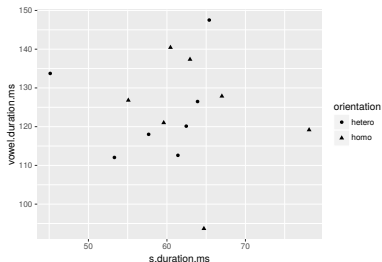
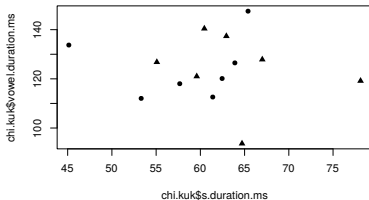
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
plot(chi.kuk$s.duration.ms, chi.kuk$vowel.duration.ms,  
     pch = c(16, 17)[chi.kuk$orientation])
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(s.duration.ms, vowel.duration.ms, shape = orientation)) +  
  geom_point()
```

scatterplot: размер

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

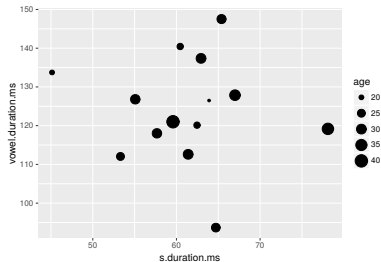
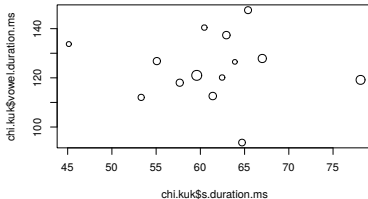
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
plot(chi.kuk$s.duration.ms, chi.kuk$vowel.duration.ms,  
     cex = chi.kuk$age/20)
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(s.duration.ms, vowel.duration.ms, size = orientation)) +  
  geom_point()
```

scatterplot: текст

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

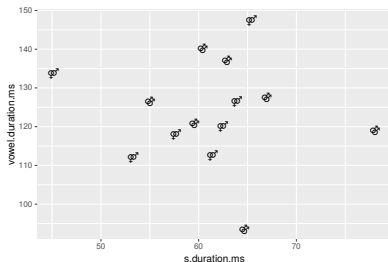
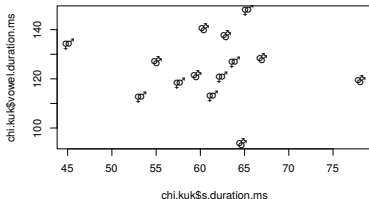
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
plot(chi.kuk$s.duration.ms, chi.kuk$vowel.duration.ms,  
     pch = c("ø", "ø̃")[chi.kuk$orientation])
```

```
# dplyr, ggplot2
```

```
levels(chi.kuk$orientation) <- list("homo"="ø", "hetero"="ø̃")
```

```
chi.kuk %>%
```

```
  ggplot(aes(s.duration.ms, vowel.duration.ms, label = orientation)) +  
  geom_text()
```

scatterplot: заголовок

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

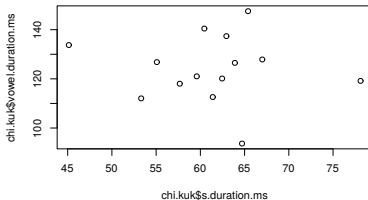
гистограмма

фасетизация

анимация

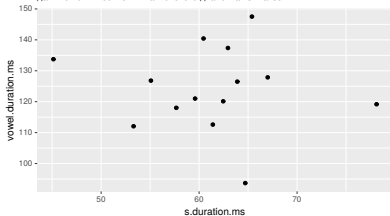
markdown

длина [s] vs. долгота гласного



длина [s] vs. долгота гласного

данные по 14 носителям кантонского диалекта языка юэ



данные из [Chi-Kuk 2007]

```
# base R
```

```
plot(chi.kuk$s.duration.ms, chi.kuk$vowel.duration.ms,  
     main = "длина [s] vs. долгота гласного")
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(s.duration.ms, vowel.duration.ms)) +
```

```
  geom_point() +
```

```
  labs(title = "длина [s] vs. долгота гласного",
```

```
       subtitle = "данные по 14 носителям кантонского диалекта языка юэ",
```

```
       caption = "данные из [Chi-Kuk 2007]")
```

scatterplot: подписи осей

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

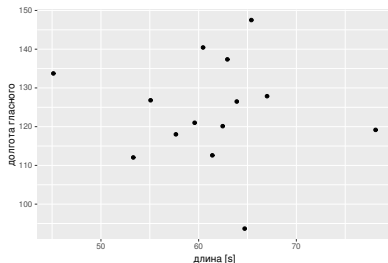
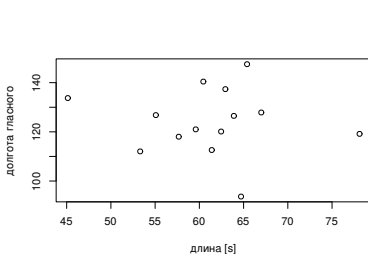
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
plot(chi.kuk$s.duration.ms, chi.kuk$vowel.duration.ms,  
     xlab = "длина [s]", ylab = "долгота гласного")
```

```
# dplyr, ggplot2
```

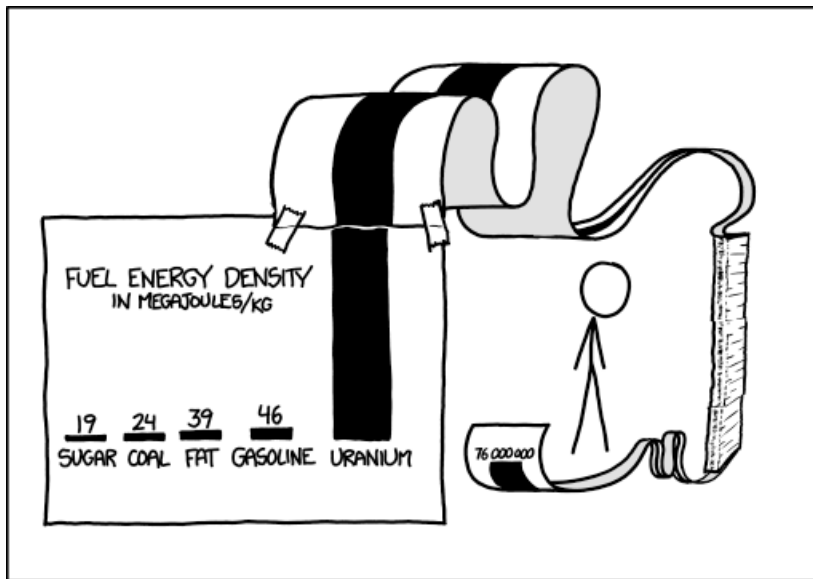
```
chi.kuk %>%
```

```
  ggplot(aes(s.duration.ms, vowel.duration.ms)) +
```

```
  geom_point() +
```

```
  xlab("длина [s]") +
```

```
  ylab("долгота гласного")
```

SCIENCE TIP: LOG SCALES ARE FOR QUITTERS WHO CAN'T
FIND ENOUGH PAPER TO MAKE THEIR POINT *PROPERLY*.

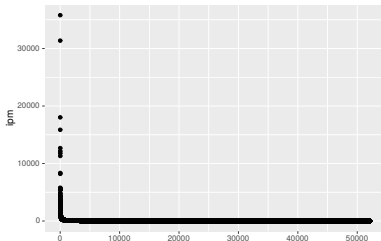
scatterplot: логарифмирование осей

Воспользуемся частотным словарем [Ляшевская, Шаров 2009] (топ-50000 лемм СЛРЯ) и посмотрим на параметр частоты слова (Freq.ipm., среднее на миллион словоупотреблений).

```
freq <- read.csv("https://goo.gl/TIX7xW", sep = "\t")
```

Если оси не логарифмировать, то получится следующее:

```
freq %>%  
  ggplot(aes(1:52138, Freq.ipm.)) +  
  geom_point() +  
  xlab("") +  
  ylab("ipm")
```



scatterplot: логарифмирование осей

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

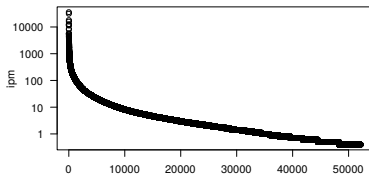
плотность

гистограмма

фасетизация

анимация

markdown



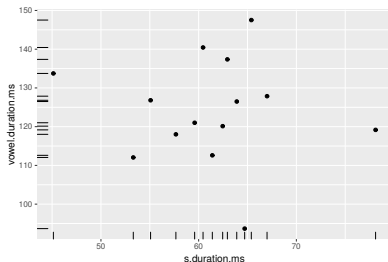
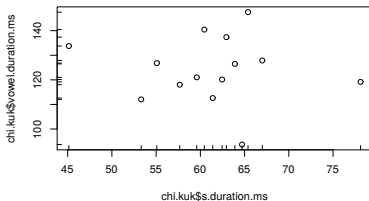
```
# base R
plot(1:52138, freq$Freq.ipm.,
     xlab = NA, ylab = "ipm",
     las = 1,
     log = "y")
```

поворот значений на оси y

```
# dplyr, ggplot2
freq %>%
  ggplot(aes(1:52138, Freq.ipm.)) +
  geom_point() +
  xlab("") +
  ylab("ipm") +
  scale_y_log10()
```

презентация доступна: <https://goo.gl/4T1zAr>

scatterplot: rug



```
# base R
```

```
plot(chi.kuk$s.duration.ms, chi.kuk$vowel.duration.ms)
```

```
rug(chi.kuk$s.duration.ms)
```

```
rug(chi.kuk$vowel.duration.ms, side = 2)
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(s.duration.ms, vowel.duration.ms)) +
```

```
  geom_point() +
```

```
  geom_rug()
```

scatterplot: линии

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

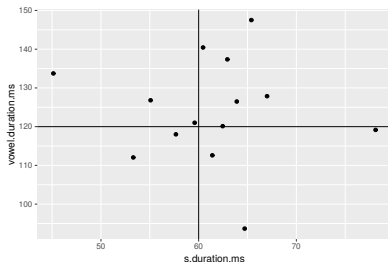
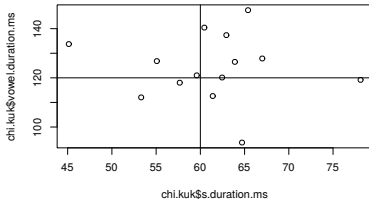
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
plot(chi.kuk$s.duration.ms, chi.kuk$vowel.duration.ms)
```

```
abline(h = 120, v = 60)
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(s.duration.ms, vowel.duration.ms)) +
```

```
  geom_point() +
```

```
  geom_hline(yintercept = 120) +
```

```
  geom_vline(xintercept = 60)
```

scatterplot: линии

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

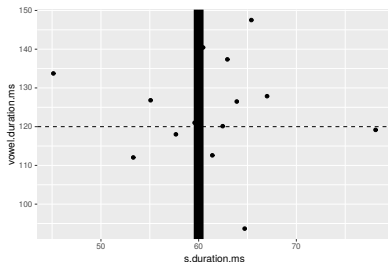
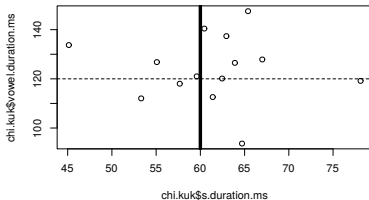
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
plot(chi.kuk$s.duration.ms, chi.kuk$vowel.duration.ms)
```

```
abline(h = 120, lty = 2)
```

```
abline(v = 60, lwd = 5)
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(s.duration.ms, vowel.duration.ms)) +
```

```
  geom_point() +
```

```
  geom_hline(yintercept = 120, linetype = 2) +
```

```
  geom_vline(xintercept = 60, size = 5)
```

scatterplot: линии

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

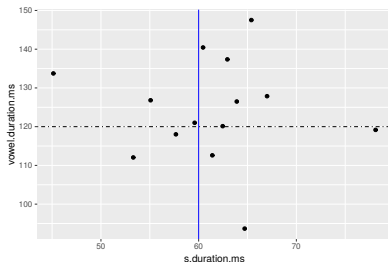
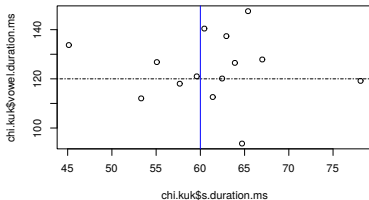
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
plot(chi.kuk$s.duration.ms, chi.kuk$vowel.duration.ms)
```

```
abline(h = 120, lty = 4)
```

```
abline(v = 60, col = "blue")
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(s.duration.ms, vowel.duration.ms)) +
```

```
  geom_point() +
```

```
  geom_hline(yintercept = 120, linetype = 4) +
```

```
  geom_vline(xintercept = 60, color = "blue")
```

barplot

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

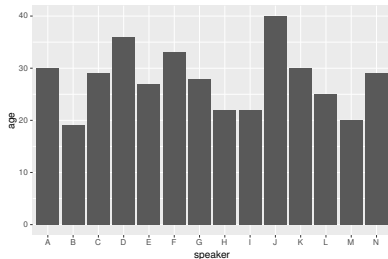
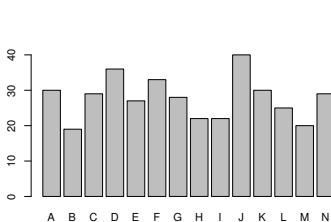
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R  
barplot(chi.kuk$age, names.arg = chi.kuk$speaker)
```

```
# dplyr, ggplot2  
chi.kuk %>%  
  ggplot(aes(speaker, age)) +  
  geom_bar(stat = "identity")
```


barplot: цвет

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

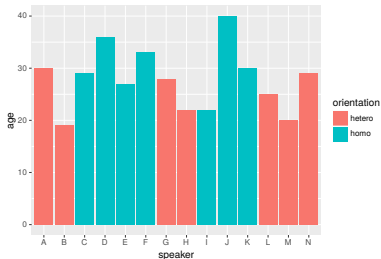
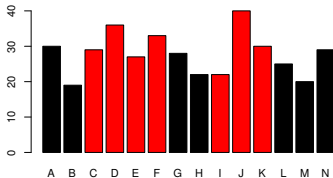
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
barplot(chi.kuk$age, names.arg = chi.kuk$speaker  
        col = chi.kuk$orientation)
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%  
  ggplot(aes(speaker, age, fill = orientation)) +  
  geom_bar(stat = "identity")
```

преобразование данных

Наши данные устроены достаточно неудобно:

```
head(chi.kuk[c(1,6,7,9)], 3)
```

	speaker	perceived.as.homo	perceived.as.hetero	orientation
1	A	7	18	hetero
2	B	20	5	hetero
3	C	9	16	homo

В связи с этим мы создадим новый датафрейм `chi.kuk.2`:

```
library(tidyr)
chi.kuk %>%
  select(speaker, perceived.as.homo, perceived.as.hetero) %>%
  gather(perceive, number, 2:3) ->
  chi.kuk.2
head(chi.kuk.2)
```

	speaker	perceive	number
1	A	perceived.as.homo	7
2	B	perceived.as.homo	20
3	C	perceived.as.homo	9

barplot: цвет

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

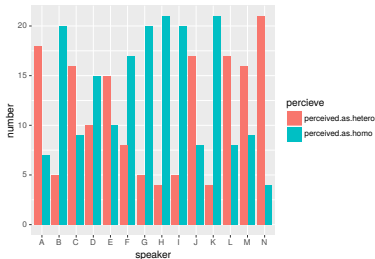
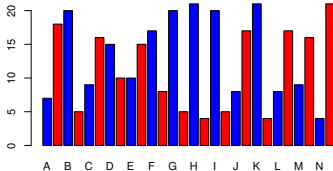
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
barplot(chi.kuk.2$number, names.arg = chi.kuk.2$speaker  
        col = c("blue", "red"))
```

```
# dplyr, ggplot2
```

```
chi.kuk.2 %>%
```

```
  ggplot(aes(speaker, number, fill = percieve)) +  
  geom_bar(stat = "identity", position = "dodge")
```

boxplot

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

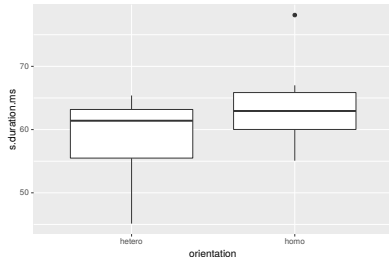
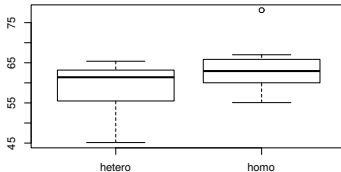
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
boxplot(chi.kuk$s.duration.ms~chi.kuk$orientation)
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(orientation, s.duration.ms)) +
```

```
  geom_boxplot()
```

boxplot: наблюдения

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

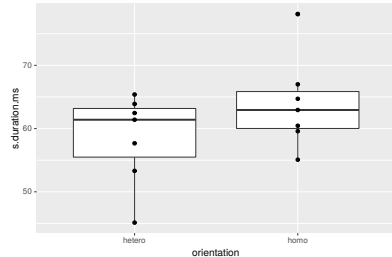
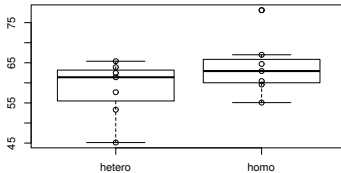
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
boxplot(chi.kuk$s.duration.ms~chi.kuk$orientation)
```

```
stripchart(chi.kuk$s.duration.ms~chi.kuk$orientation,  
  pch = 1, vertical = T, add = T)
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(orientation, s.duration.ms)) +
```

```
  geom_boxplot() +
```

```
  geom_point()
```

boxplot: jitter

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

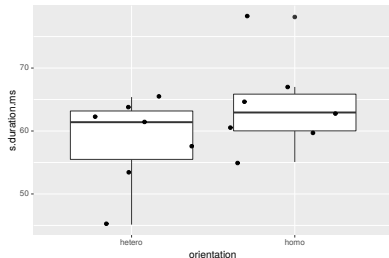
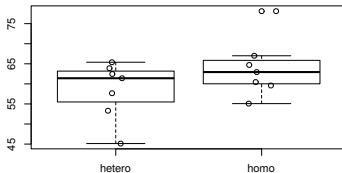
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
boxplot(chi.kuk$s.duration.ms~chi.kuk$orientation)  
stripchart(chi.kuk$s.duration.ms~chi.kuk$orientation,  
  pch = 1, vertical = T, add = T, method = "jitter")
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(orientation, s.duration.ms)) +  
  geom_boxplot() +  
  geom_jitter()
```

boxplot: violin

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

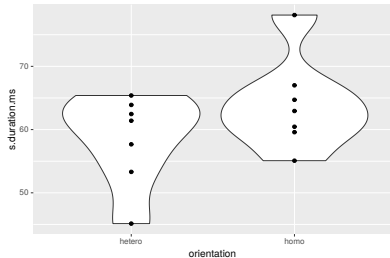
плотность

гистограмма

фасетизация

анимация

markdown



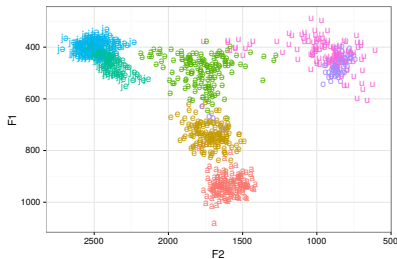
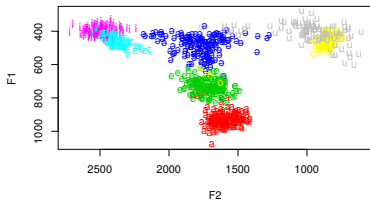
В base R есть пакет `vioplot`, но он плохо сделан.

```
# dplyr, ggplot2
chi.kuk %>%
  ggplot(aes(orientation, s.duration.ms)) +
  geom_violin() +
  geom_point()
```

адыгейские гласные

Для следующих графиков будет использован датасет, состоящий из данных по 7 гласным кабардинского языка, собранных от одного носителя. Анализировалось анкета в 210 слов, каждое слово произносилось 3 раза изолированно, а четвертый раз в предложении. Сбор и разметка осуществлялись Соней Сиговой.

```
adyghe <- read.csv("https://goo.gl/rYZPlj", sep = "\t")
```



countourplot

введение

2 переменные

scatterplot

barplot

boxplot

countourplot

распределение

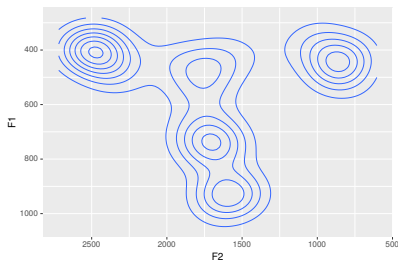
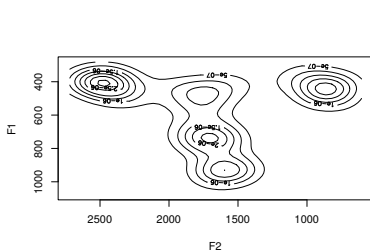
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R, MASS
```

```
library(MASS)
```

```
z <- kde2d(adyghe$F1, adyghe$F2, n = 50) # 2D kernel density estimation
```

```
contour(z, nlevels = 9,
```

```
  xlim = rev(range(adyghe$F2)),
```

```
# переворот осей
```

```
  ylim = rev(range(adyghe$F1)))
```

```
# переворот осей
```

```
# dplyr, ggplot2
```

```
adyghe%>%
```

```
ggplot(aes(F2, F1)) +
```

```
  geom_density_2d() +
```

```
  scale_y_reverse() +
```

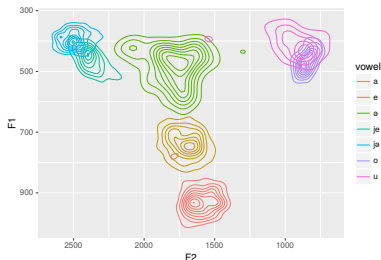
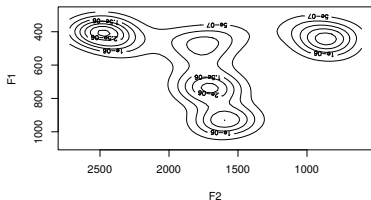
```
  scale_x_reverse()
```

```
# переворот осей
```

```
# переворот осей
```

презентация доступна: <https://goo.gl/4T1zAr>

countourplot



```
# base R, MASS
```

```
library(MASS)
```

```
z <- kde2d(adyghe$F1, adyghe$F2, n = 50) # 2D kernel density estimation
```

```
contour(z, nlevels = 9,  
        xlim = rev(range(adyghe$F2)),  
        ylim = rev(range(adyghe$F1)))
```

```
# переворот осей
```

```
# переворот осей
```

```
# dplyr, ggplot2
```

```
adyghe%>%
```

```
ggplot(aes(F2, F1, color = vowel)) +
```

```
  geom_density_2d() +
```

```
  scale_y_reverse() +
```

```
  scale_x_reverse()
```

```
# переворот осей
```

```
# переворот осей
```

презентация доступна: <https://goo.gl/4T1zAr>

Итоги: декартова система координат

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

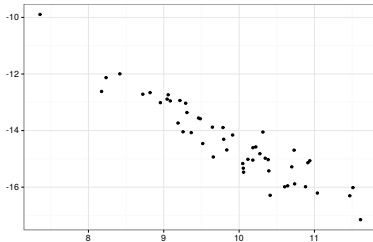
гистограмма

фасетизация

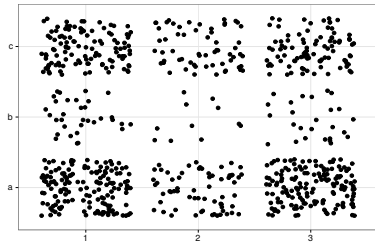
анимация

markdown

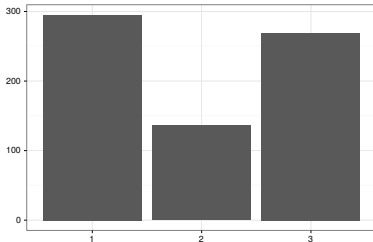
две числовые переменные



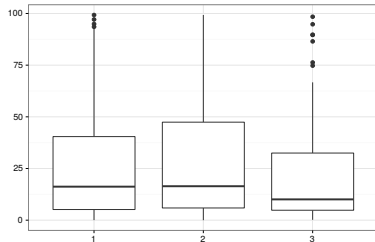
две номинативные пер.



номинативная переменная и число



номинативная и числовая пер.



ПЛОТНОСТЬ

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

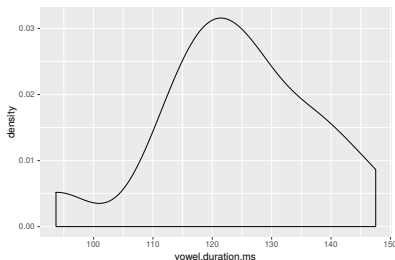
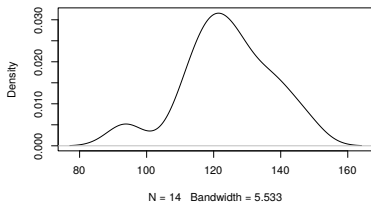
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R  
plot(density(chi.kuk$vowel.duration.ms), main = "")
```

```
# dplyr, ggplot2  
chi.kuk %>%  
  ggplot(aes(vowel.duration.ms)) +  
  geom_density()
```

ПЛОТНОСТЬ: заполнение

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

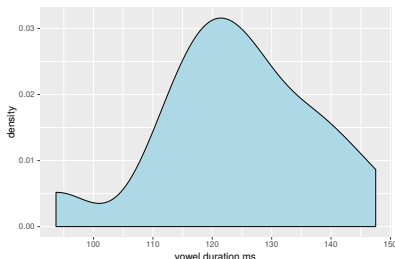
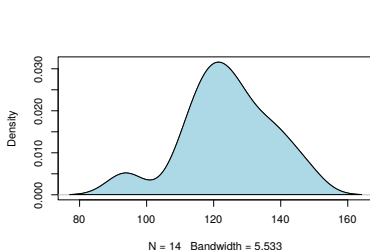
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
```

```
plot(density(chi.kuk$vowel.duration.ms), main = "")
```

```
  polygon(density(chi.kuk$vowel.duration.ms), col = "lightblue")
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(vowel.duration.ms)) +
```

```
  geom_density(fill = "lightblue")
```

плотность: сравнение групп

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

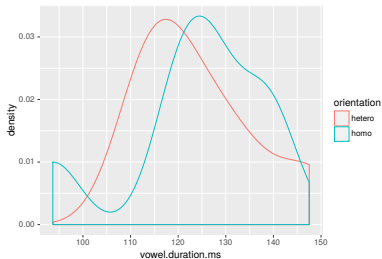
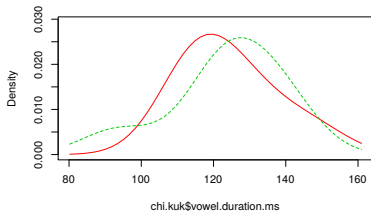
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R, sm
```

```
library(sm)
```

```
sm.density.compare(chi.kuk$vowel.duration.ms, chi.kuk$orientation)
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(vowel.duration.ms, color = orientation)) +  
  geom_density()
```

плотность: сравнение групп

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

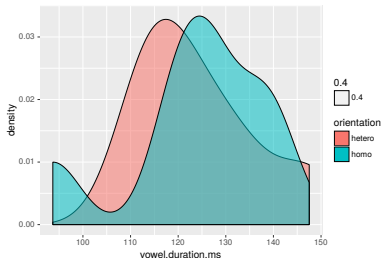
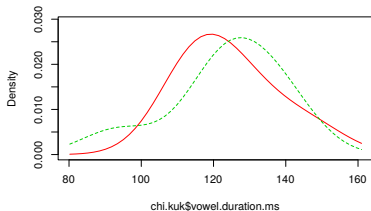
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R, sm
```

```
library(sm)
```

```
sm.density.compare(chi.kuk$vowel.duration.ms, chi.kuk$orientation)
```

```
# dplyr, ggplot2
```

```
chi.kuk %>%
```

```
  ggplot(aes(vowel.duration.ms, fill = orientation, alpha = 0.4)) +  
  geom_density()
```

гистограмма

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

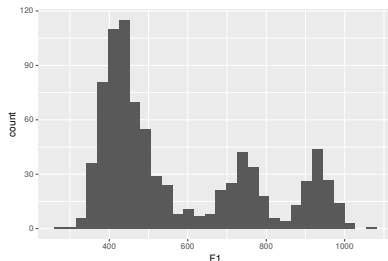
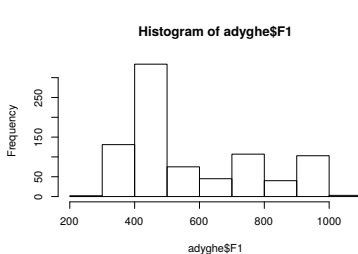
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R  
hist(adyghe$F1)
```

```
# dplyr, ggplot2  
adyghe %>%  
  ggplot(aes(F1)) +  
  geom_histogram()
```


гистограмма: количество ячеек

В википедии [предложено](#) несколько способов выбора количества ячеек гистограммы. В R реализовано несколько из них:

```
nclass.Sturges(adyghe$F1)      # по умолчанию в R base [Sturges 1926]
nclass.scott(adyghe$F1)        # [Scott 1979]
nclass.FD(adyghe$F1)           # [Freedman, Diaconis 1981]
30                             # по умолчанию в ggplot2
```

гистограмма: количество ячеек

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

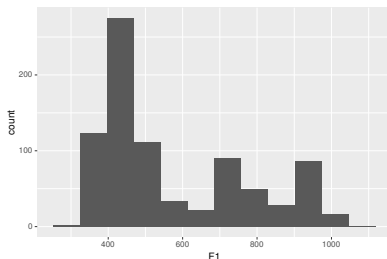
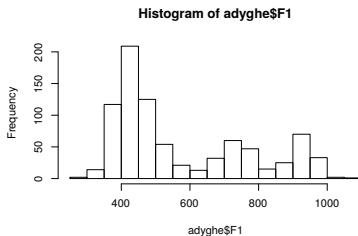
плотность

гистограмма

фасетизация

анимация

markdown



```
# base R
hist(adyghe$F1, breaks = nclass.FD(adyghe$F1))
# Аргумент breaks почему-то работает некорректно

# dplyr, ggplot2
adyghe %>%
  ggplot(aes(F1)) +
  geom_histogram(bins = nclass.FD(adyghe$F1))
```

гистограмма: цвет

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

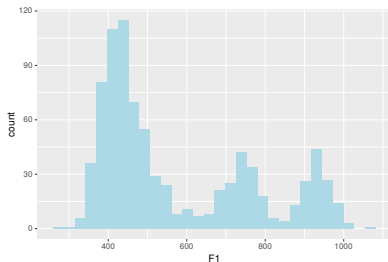
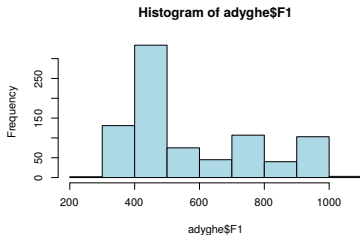
плотность

гистограмма

фасетизация

анимация

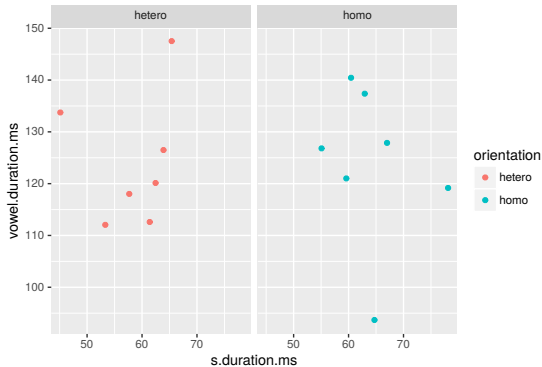
markdown



```
# base R
hist(adyghe$F1, col = "lightblue")

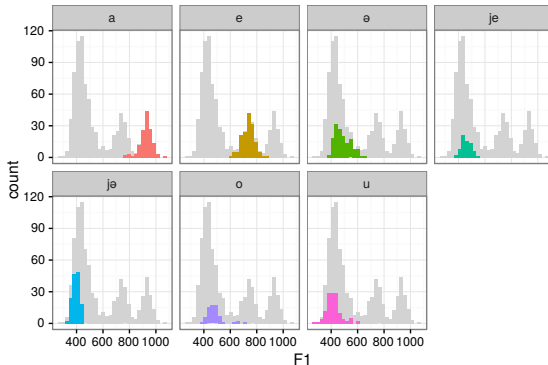
# dplyr, ggplot2
adyghe %>%
  ggplot(aes(F1)) +
  geom_histogram(fill = "lightblue")
```

Фасетизация



```
chi.kuk %>%  
  ggplot(aes(s.duration.ms, vowel.duration.ms, color = orientation)) +  
  geom_point() +  
  facet_wrap(~orientation)
```

Фасетизация: с данными на фоне



```
adyghe2 <- adyghe[,-1]      создаем датафрейм без переменной vowel
adyghe %>%
  ggplot(aes(F1, fill = vowel))+
  geom_histogram(data = adyghe2, fill = "lightgrey") +      # adyghe2
  geom_histogram() +                                         # a поверх adyghe
  facet_wrap(~vowel, nrow = 2) +
  guides(fill = FALSE) +                                     # убирает легенду
  theme_bw()                                                 # черно-белая тема
```

презентация доступна: <https://goo.gl/4T1zAr>

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

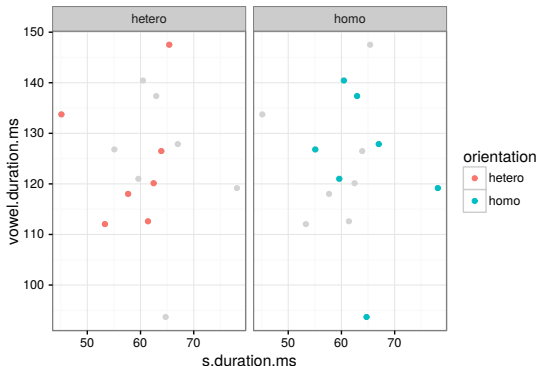
гистограмма

фасетизация

анимация

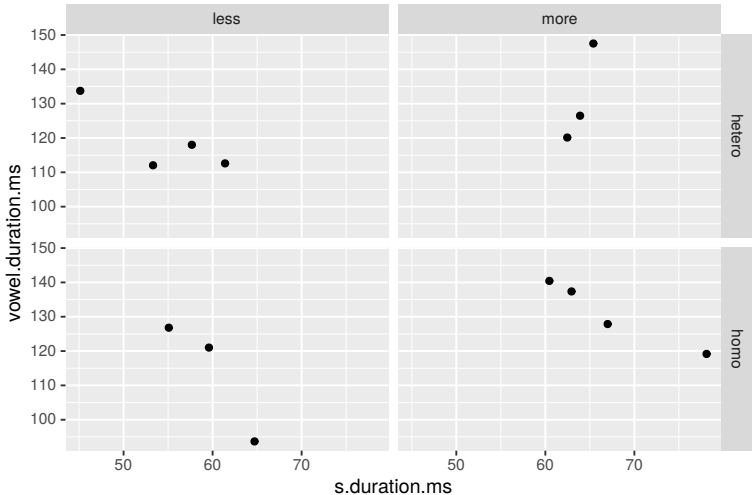
markdown

Фасетизация: с данными на фоне



```
chi.kuk2 <- chi.kuk[, -9] # создаем датафрейм без перемен. orientation
chi.kuk %>%
  ggplot(aes(s.duration.ms, vowel.duration.ms, color = orientation)) +
  geom_point(data = chi.kuk2, color = "lightgrey") + # chi.kuk2
  geom_point() + # и добавляем поверх данные из chi.kuk
  theme_bw() + # черно-белая тема
  facet_wrap(~orientation)
```

Фасетизация



```
chi.kuk %>%  
  ggplot(aes(s.duration.ms, vowel.duration.ms)) +  
  geom_point() +  
  facet_grid(orientation~mean.homo)
```

презентация доступна: <https://goo.gl/4T1zAr>

Анимация: base R

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

гистограмма

фасетизация

анимация

markdown

```
library(animation)
```

```
plot_frame1 <- function(n) {  
  plot(n, n,  
        col = n,  
        pch = 20,  
        cex = 10 - n,  
        xlim = c(0, 10),  
        ylim = c(0, 10),  
        xlab = NA,  
        ylab = NA)  
}
```

цвет
тип
размер

```
saveGIF(  
  # команда lapply применяет к функции значения  
  # первого аргумента и возвращает некоторый список  
  lapply(1:10, function(i) {  
    plot_frame1(i)  
  }),  
  interval = 0.5, # время между слайдами  
  ani.width = 800,  
  ani.height = 600,  
  movie.name = 'myanim.gif')
```


Анимация: ggplot2, gganimation

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

гистограмма

фасетизация

анимация

markdown

```
devtools::install_github("dgrtwo/gganimate")      # установка из github
library(gganimate)
df <- data.frame(n = 1:10)                        # создаем датафрейм
df %>%
  ggplot(aes(n, n,
              size = n,                          # размер
              color = factor(n),                 # цвет
              frame = n)) +                       # параметр, по которому анимировать
  geom_point() ->
  p                                               # записывает график в переменную
gg_animate(p)
gg_animate(p, "output.gif")
gg_animate(p, "output.mp4")
gg_animate(p, "output.swf")
gg_animate(p, "output.html")
```

R Markdown

R Markdown позволяет создавать документы, содержащие отчет или презентацию с прямым вызовом функций R и его выдачей.

- Инсталляция: `install.packages("rmarkdown")`
- жанры: документ, презентация
- доступные форматы: html, pdf (если стоит TeX), docx
- доступно легкое форматирование
 - ***италик***
 - италик
 - ****жирный****
 - жирный
 - ^надписное^
 - ~~перечеркнутое~~
 - # 1. Заголовок
 - ## 1.2 Подзаголовок
 - ~~TEX~~: $\frac{4}{3}$
 - Горизонтальная линия: ***
 - > отдельный блок
 - * булет
 - + подбулет

презентация доступна: <https://goo.gl/4T1zAr>

R Markdown: оформление кода

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

гистограмма

фасетизация

анимация

markdown

- код и результат

```
```{r}  
plot(99:22)
```
```

- только код

```
```{r eval = FALSE}  
plot(99:22)
```
```

- только результат

```
```{r echo = FALSE}  
plot(99:22)
```
```

- в тексте

\$5 \ choose 7\$ будет `r choose(7, 5)`

$\binom{5}{7}$ будет 21

- без предупреждений

```
```{r warning = FALSE}  
plot(99:22)
```
```

- без сообщений

```
```{r message = FALSE}  
plot(99:22)
```
```

введение

2 переменные

scatterplot

barplot

boxplot

contourplot

распределение

плотность

гистограмма

фасетизация

анимация

markdown

Спасибо за внимание!

Пишите письма

agricolamz@gmail.com