Лабораторная работа № 2: Визуализация данных c помощью R

Щербаков М.В Сай Ван К Май Чанг

26 февраля 2018 г

### 1. Цель работы

* Освоить основные распространенные типы графиков в R
* Приобрести основные навыки работы с пакетом ggplot2 в R

### 2. Задачи

* Углубить и закрепить знания по основным типам графикам.
* Научить строить основные типы графики в R
* Совершенствовать навыки самостоятельной работы.

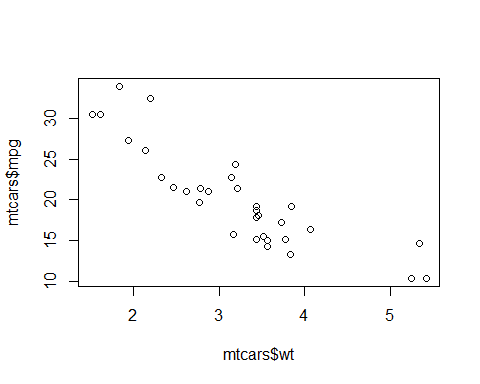
### 3. План выполнения работы

Хотя пакет ggplot2 - пакет к более сложным и красивым графикам для визуализации данных в R, но это не единственный способ сделать графики в R. Для быстрого изучения данных иногда полезно использовать базовые функции в R. они устанавливаются по умолчанию с R и не требуется установить дополнительные пакеты. Эта лаб помогает вам усвоить навыки по построению различных типов графиков в R c использованием базовых функций и пакета ggplot2.

### 3.1 Scatter Plot (диаграмма рассеяния)

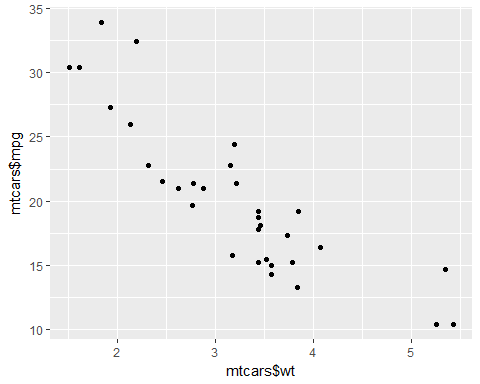
* Для построения Scatter Plot воспользуемся plot():

plot(mtcars$wt, mtcars$mpg)



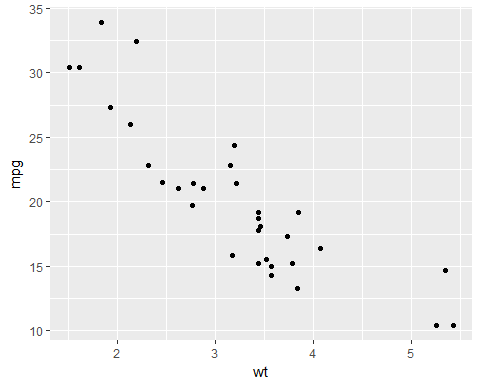
* С пакетом ggplot2, вы также можете получить похожий результат использованием функции qplot():

library(ggplot2)  
qplot(mtcars$wt, mtcars$mpg)

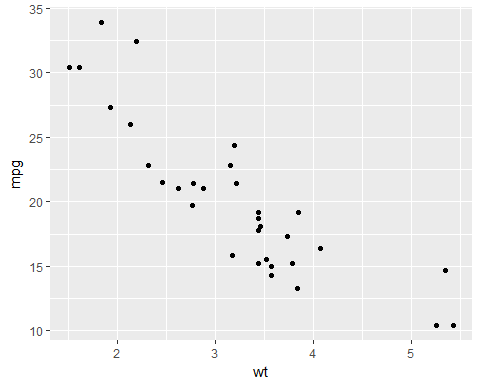


* Если два вектора в одном фрейме данных, вы можете использовать следующий синтаксиc:

qplot(wt, mpg, data=mtcars)



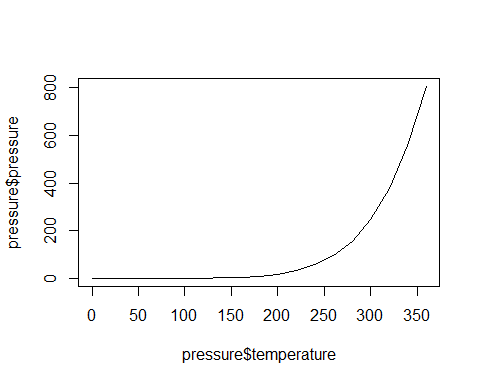
# это эквиваленто:  
ggplot(mtcars, aes(x=wt, y=mpg)) + geom\_point()



### 3.2 Line Graph (Линейный график )

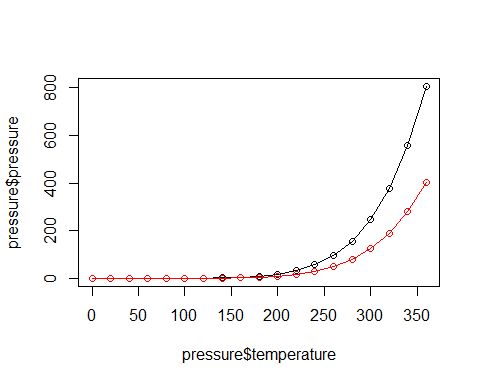
* Для построения Line Graph воспользуемся plot() и указать тип графика type = “l”:

plot(pressure$temperature, pressure$pressure, type="l")



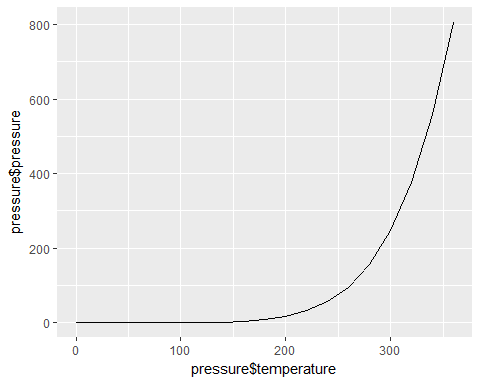
* Для добавления точек и/или линий, сначала вызываем функцию plot() для первой линии, и потом добавляем точки с использованием points() и добавляем другие линии с lines():

plot(pressure$temperature, pressure$pressure, type="l")  
points(pressure$temperature, pressure$pressure)  
lines(pressure$temperature, pressure$pressure/2, col="red")  
points(pressure$temperature, pressure$pressure/2, col="red")



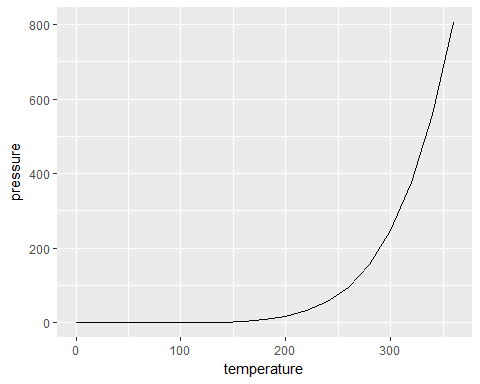
* с ggplot2, вы также можете получить похожий результат использованием функции qplot() c geom = “line” :

library(ggplot2)  
qplot(pressure$temperature, pressure$pressure, geom="line")

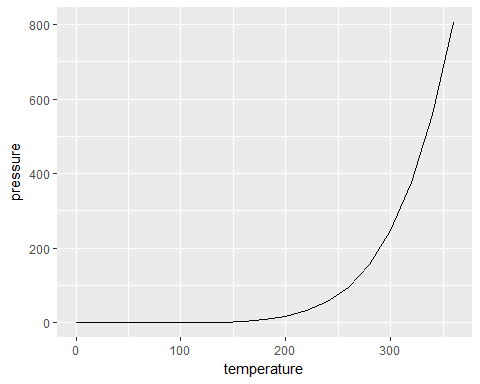


* Если два вектора в одном фрейме данных, вы можете использовать следующий синтаксиc:

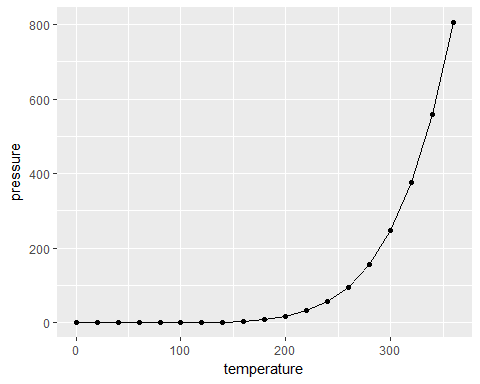
qplot(temperature, pressure, data=pressure, geom="line")



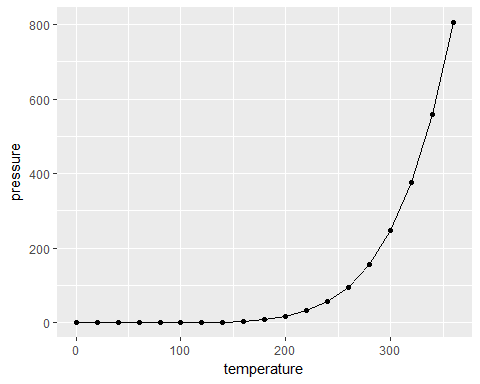
# This is equivalent to:  
ggplot(pressure, aes(x=temperature, y=pressure)) + geom\_line()



# Lines and points together  
qplot(temperature, pressure, data=pressure, geom=c("line", "point"))



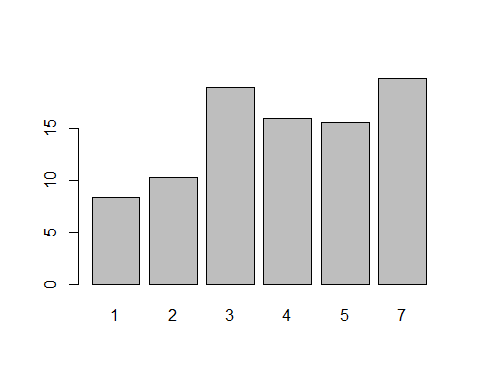
# Equivalent to:  
ggplot(pressure, aes(x=temperature, y=pressure)) + geom\_line() + geom\_point()



### 3.3 Bar Graph (столбчатая диаграмма)

* для построения Bar Graph воспользуемся barplot() и передаем ему вектор значений и вектор надписей:

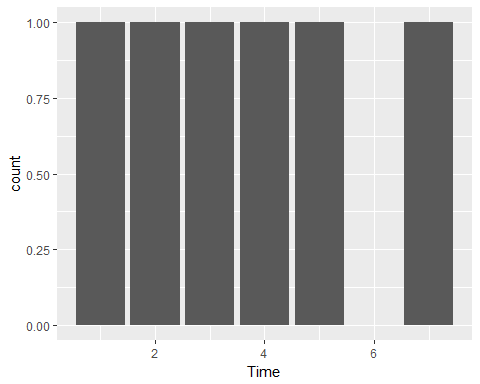
barplot(BOD$demand, names.arg=BOD$Time)



* C пакетом ggplot2 , вы также можете получить похожий результата использованием qplot(). Обратите внимание на разницу в результате, когда переменная x- непрерывная и когда она дискретная :

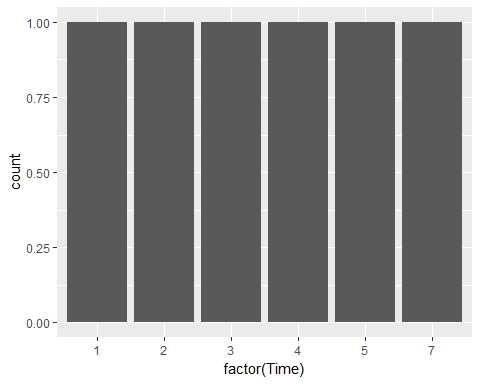
library(ggplot2)  
qplot(Time, data = BOD, geom="bar", stat="identity")

## Warning: `stat` is deprecated



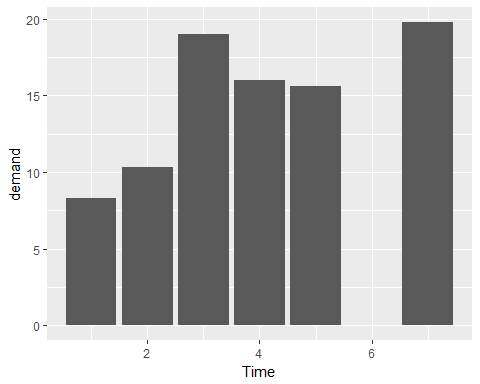
# Convert the x variable to a factor, so that it is treated as discrete  
qplot(factor(Time), data = BOD, geom="bar", stat="identity")

## Warning: `stat` is deprecated



* Если ветор в фрейме данных, вы можете использовать следующий синтаксис:

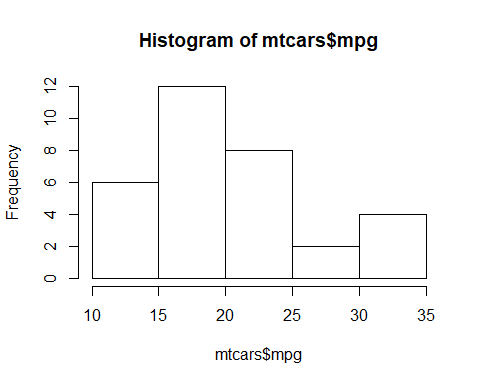
#qplot(Time, demand, data=BOD, geom="bar", stat="identity")  
# This is equivalent to:  
ggplot(BOD, aes(x=Time, y=demand)) + geom\_bar(stat="identity")



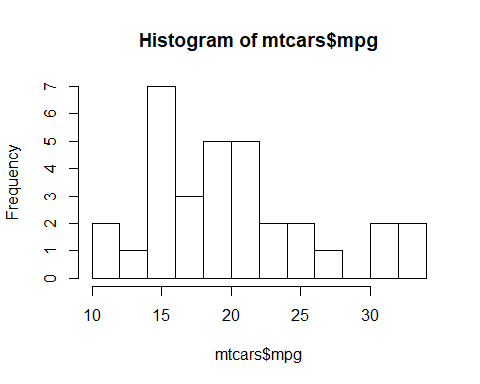
### 3.4 Histogram (гистограмма)

* Для построения Histogram воспользуемся hist():

hist(mtcars$mpg)



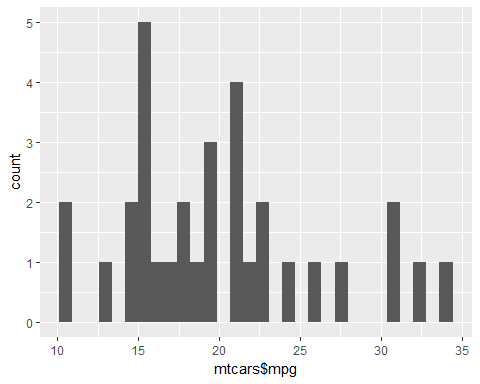
# Specify approximate number of bins with breaks  
hist(mtcars$mpg, breaks=10)



* С пакетом ggplot2 вы также можете получить похожий результат использованием qplot():

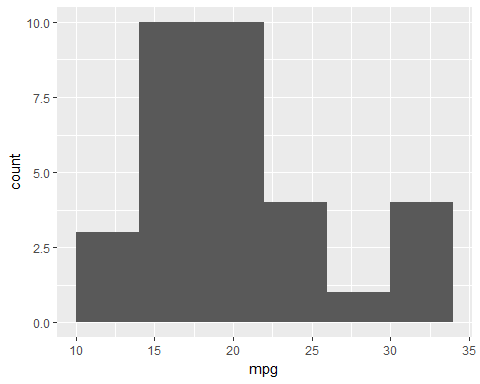
library(ggplot2)  
qplot(mtcars$mpg)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

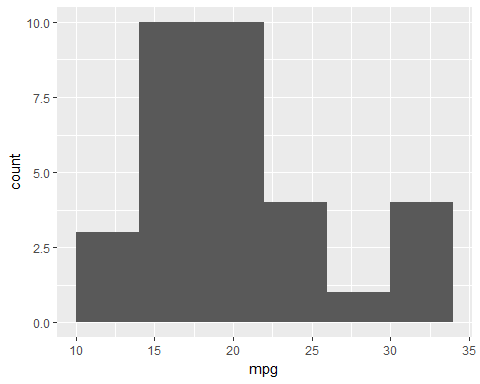


* Если вектор в фрейме данных вы можете использовать следующий синтаксис:

library(ggplot2)  
qplot(mpg, data=mtcars, binwidth=4)



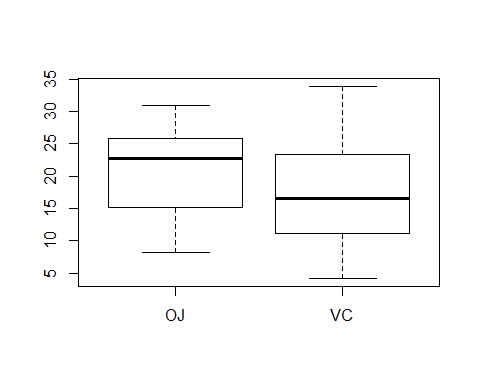
# This is equivalent to:  
ggplot(mtcars, aes(x=mpg)) + geom\_histogram(binwidth=4)



### 3.5 Box Plot (ящики с усами)

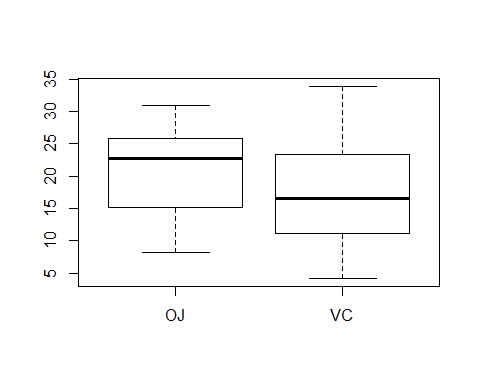
* Для построения box plot воспользуемся plot(). Если x – фактор, то функция plot() автоматически строит box plot:

plot(ToothGrowth$supp, ToothGrowth$len)

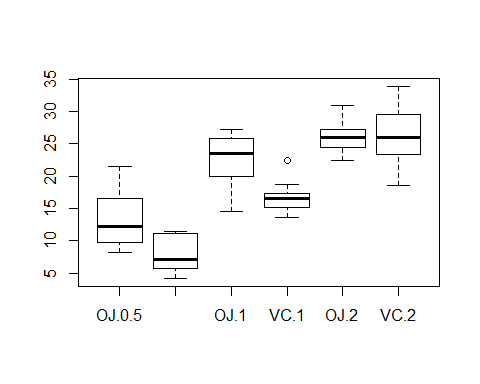


-Если 2 вектора в одном фрейме данных, вы можете использовать следующий синтаксис:

# Formula syntax  
boxplot(len ~ supp, data = ToothGrowth)

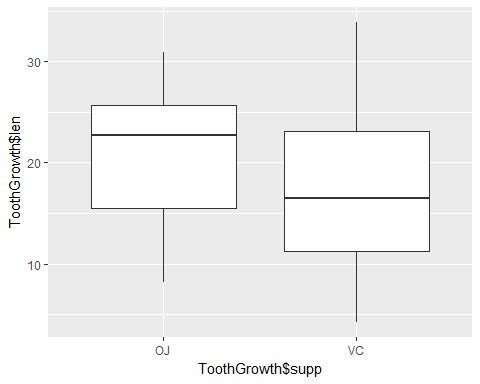


# Put interaction of two variables on x-axis  
boxplot(len ~ supp + dose, data = ToothGrowth)



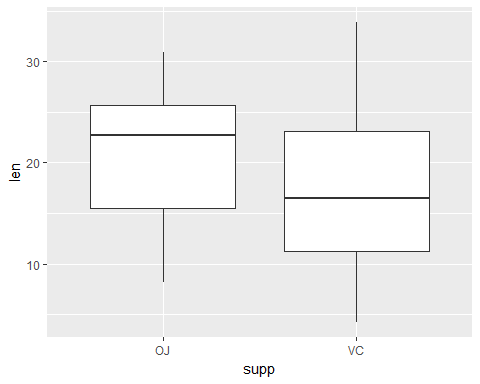
* С пакетом ggplot2 вы также можете получить похожий результат использованием qplot() c geom = “boxplot”:

library(ggplot2)  
qplot(ToothGrowth$supp, ToothGrowth$len, geom="boxplot")

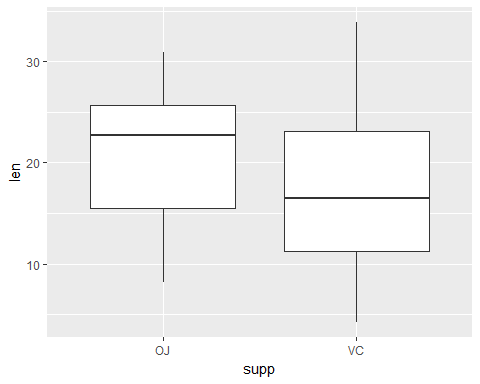


-Если 2 вектора в одном фрейме данных, вы можете использовать следующий синтаксис:

qplot(supp, len, data=ToothGrowth, geom="boxplot")



# This is equivalent to:  
ggplot(ToothGrowth, aes(x=supp, y=len)) + geom\_boxplot()



### 4. Задания:

**Задание № 1**: Загрузка набора данных iris и построение гистограмы по данным из столбца Sepal.Length с использованием базовой функции и пакета ggplot2.

**Задание № 2**: Загрузка набора данных iris и построение Scatter plot по данным из столбцов Sepal.Length и Sepal.Width с использованием базовой функции и пакета ggplot2.

**Задание № 3**: Создание фрейма из двух векторов и построение линейного графика зависимости этих двух векторов с использованием базовой функции и пакета ggplot2.

**Задание № 4**: Создание фрейма из двух векторов и построение Bar Graph для этого фрема данных с использованием базовой функции и пакета ggplot2.

**Задание № 5**: Загрузка набора данных AirPassengers и построение линейного графика для этого набора данных с использованием базовой функции и пакета ggplot2.

**Задание № 6**: Загрузка набора данных Titanic и построение построение гистограмы по данным из столбцов Sex и Freq с использованием базовой функции и пакета ggplot2.

**Задание № 7**: Загрузка набора данных CO2 и построение построение Scatter plot по данным из столбцов conc и uptake с использованием базовой функции и пакета ggplot2. …..