University of BRISTOL

Basic plots in R

Osama Mahmoud

Website: http://osmahmoud.com
E-mail: o.mahmoud@bristol.ac.uk

14 November 2019

Goals

In this session, we aim to cover:

- what data visualisation is.
- why we need visualisation.
- what the base graphics and their types are.
- How to produce base plots.
- How to save your generated graphics.

Contents

1. Introduction to data visualisation
2. Overview of base graphics
3. Histograms
4. Bar plots
5. Box plots
6. Scatter plots

Graphics reveal data

"*The greatest value of a picture is when it forces us to notice what we never expected to see.*"

JOHN W. TUKEY (1915 - 2000)

Why do we need visualisations?

I		II		III	
x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46
8.0	6.95	8.0	8.14	8.0	6.77
13.0	7.58	13.0	8.74	13.0	12.74
9.0	8.81	9.0	8.77	9.0	7.11
11.0	8.33	11.0	9.26	11.0	7.81
14.0	9.96	14.0	8.10	14.0	8.84
6.0	7.24	6.0	6.13	6.0	6.08
4.0	4.26	4.0	3.10	4.0	5.39
12.0	10.84	12.0	9.13	12.0	8.15
7.0	4.82	7.0	7.26	7.0	6.42
5.0	5.68	5.0	4.74	5.0	5.73

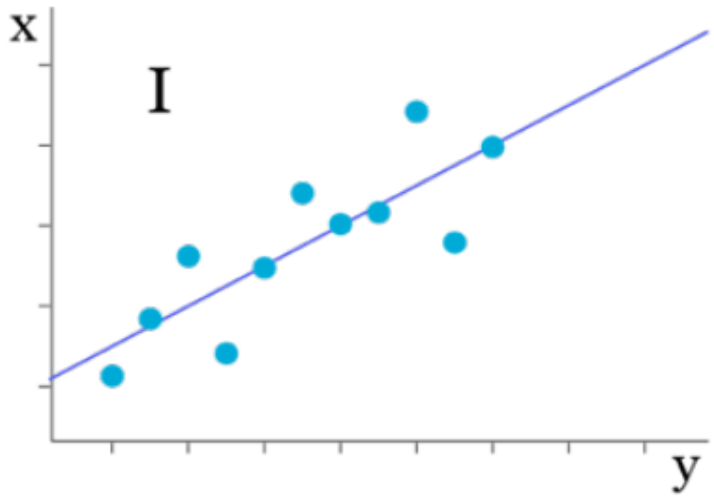
n = 11

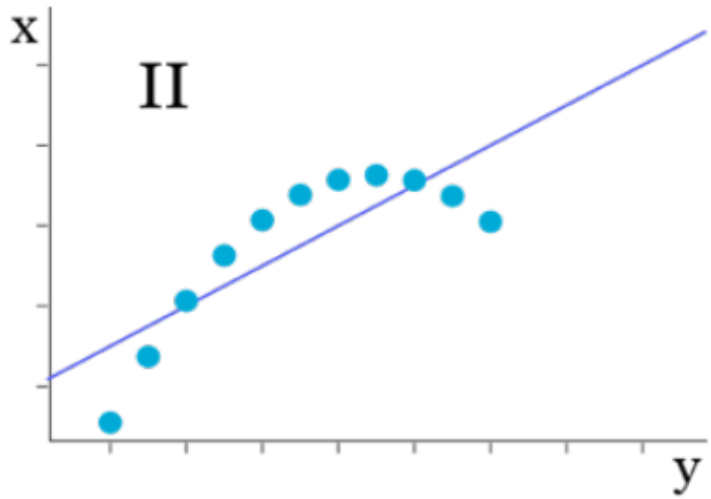
mean x = 9.0
mean y = 7.5

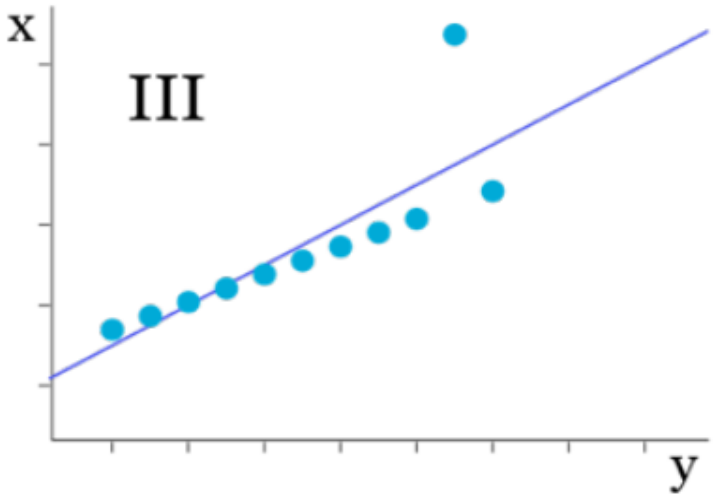
variance x = 11.0
variance y = 4.12

correlation x & y = 0.816
regression line: y = 3+0.5x

Why do we need visualisations?

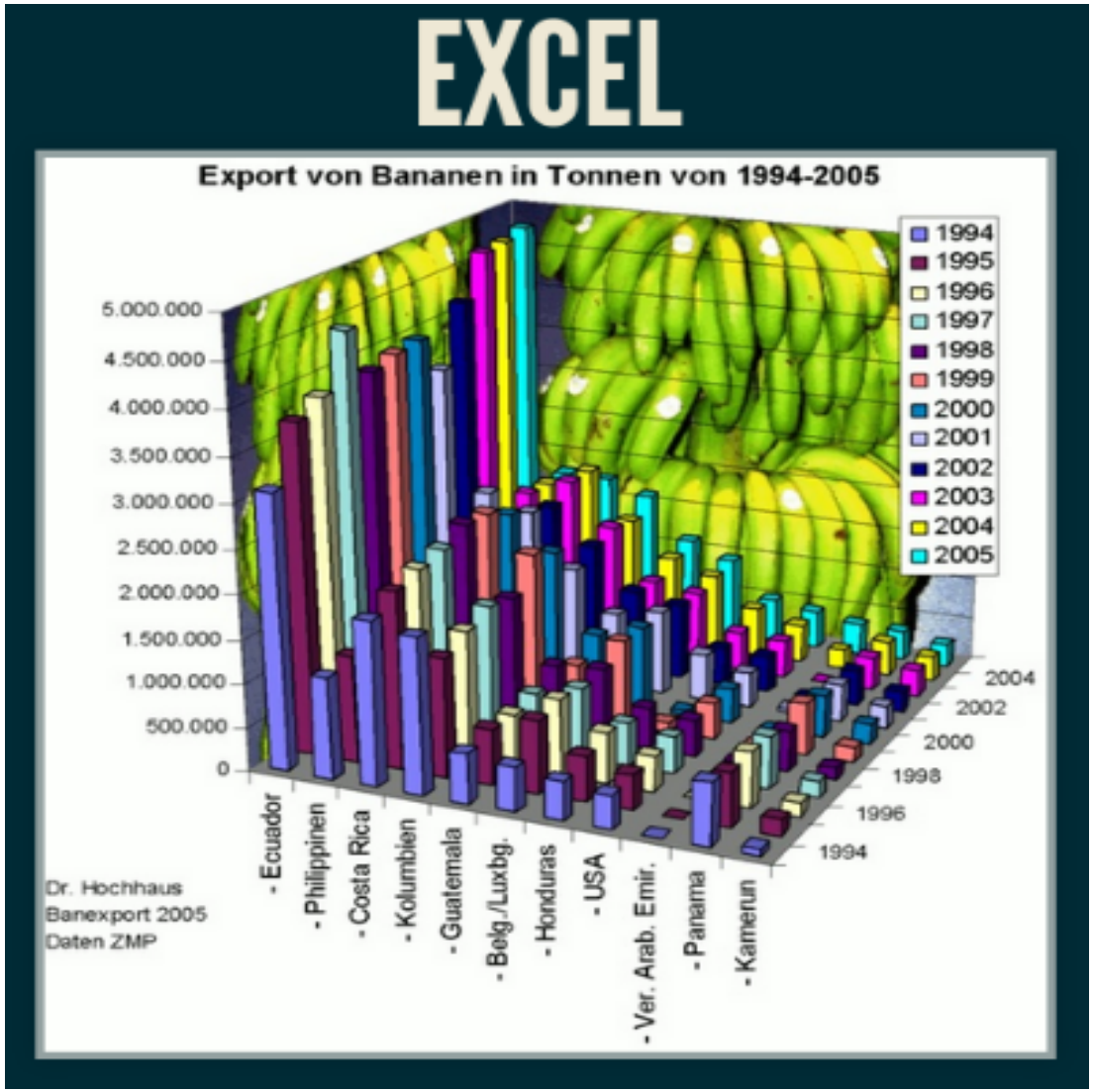




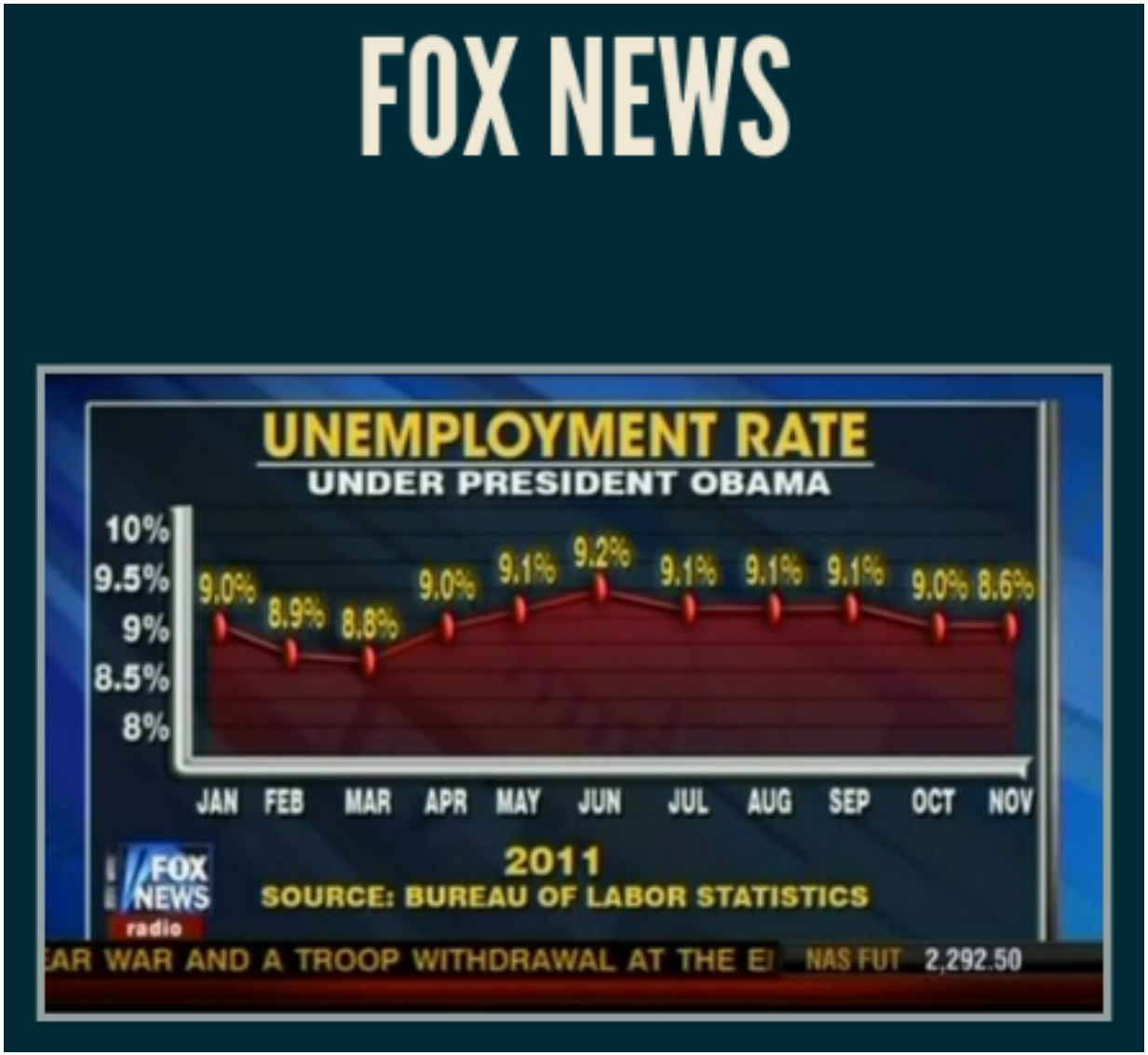


THE UGLY

The ugly


EXCEL
Export von Bananen in Tonnen von 1994-2005
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
5.000.000
4.500.000
4.000.000
3.500.000
3.000.000
2.500.000
2.000.000
1.500.000
1.000.000
500.000
0
2004
2002
2000
1998
1996
1994
- Ecuador
- Philippinen
- Costa Rica
- Kolumbien
- Guatemala
- Belg./Luxbg.
- Honduras
- USA
- Ver. Arab. Emir.
- Panama
- Kamerun
Dr. Hochhaus
Banexport 2005
Daten ZMP

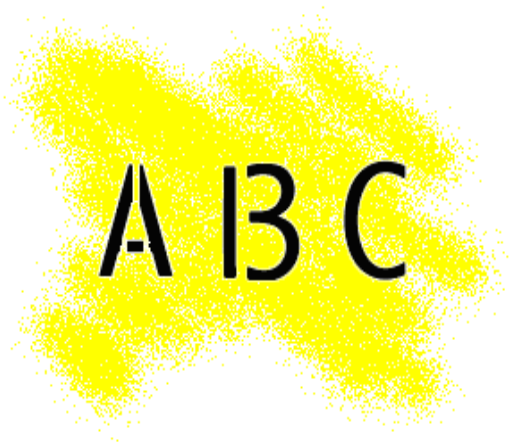
The ugly


FOX NEWS
UNEMPLOYMENT RATE
UNDER PRESIDENT OBAMA
10%
9.5%
9%
8.5%
8%
9.0%
8.9%
8.8%
9.0%
9.1%
9.2%
9.1%
9.1%
9.1%
9.0%
8.6%
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
2011
SOURCE: BUREAU OF LABOR STATISTICS
FOX NEWS radio
AR WAR AND A TROOP WITHDRAWAL AT THE E
NAS FUT 2,292.50

FEW ISSUES

Few issues

- Human perception can be ...

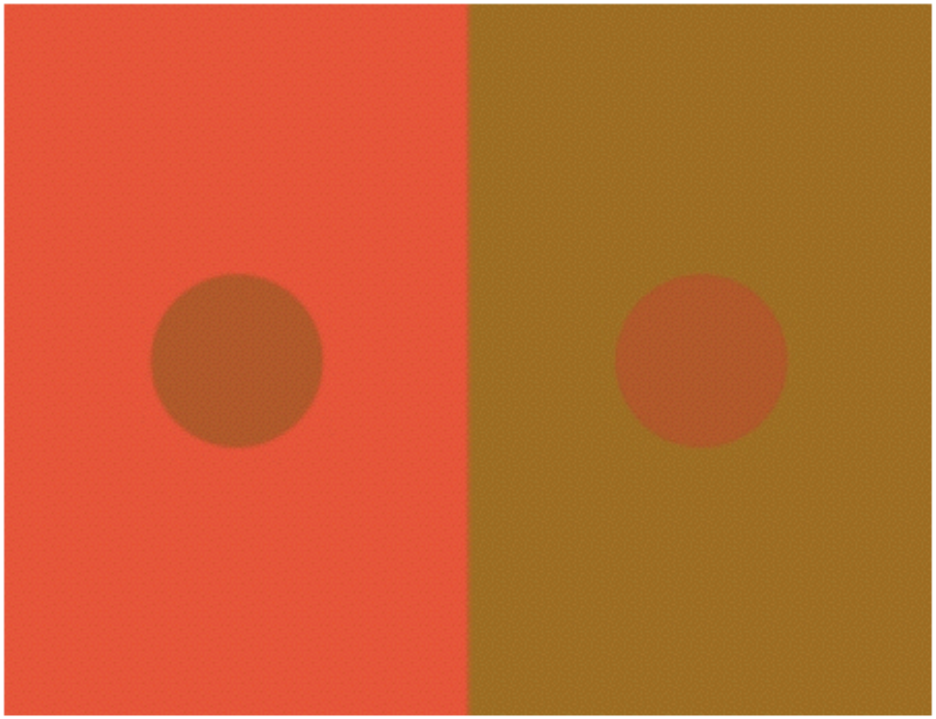


Few issues

- Human perception can be deceiving

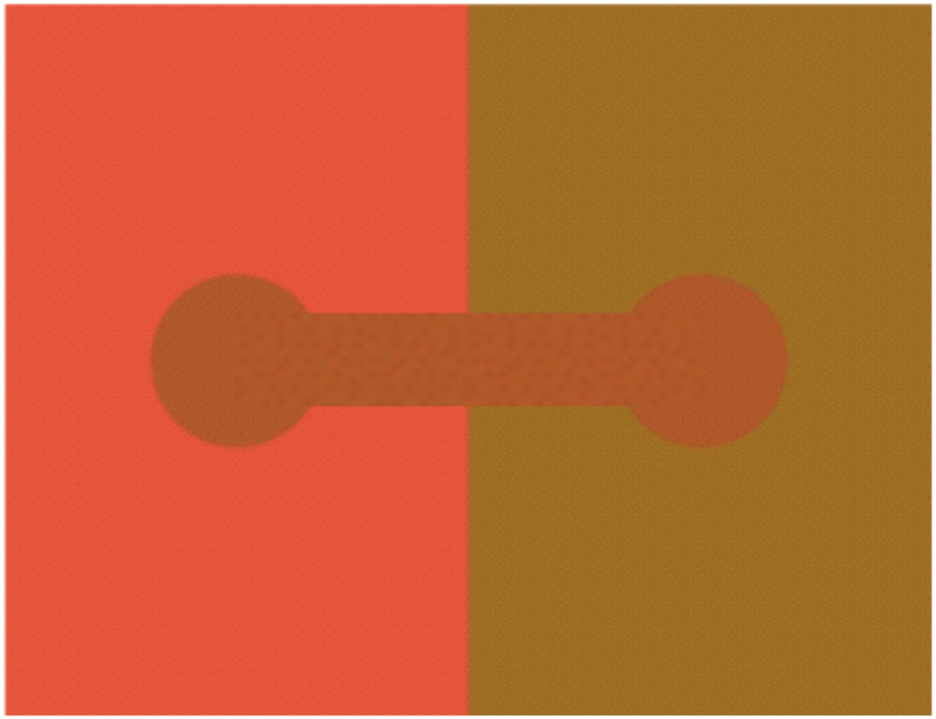
Few issues

- Another example:



Few issues

- Another example:



Few issues

- Selective attention

Selective attention test

Types of R graphics

- Base graphics.
- Grid graphics.
- Lattice graphics.
- ggplot2 graphics.

Base graphics

Installing course R package: BristolVis

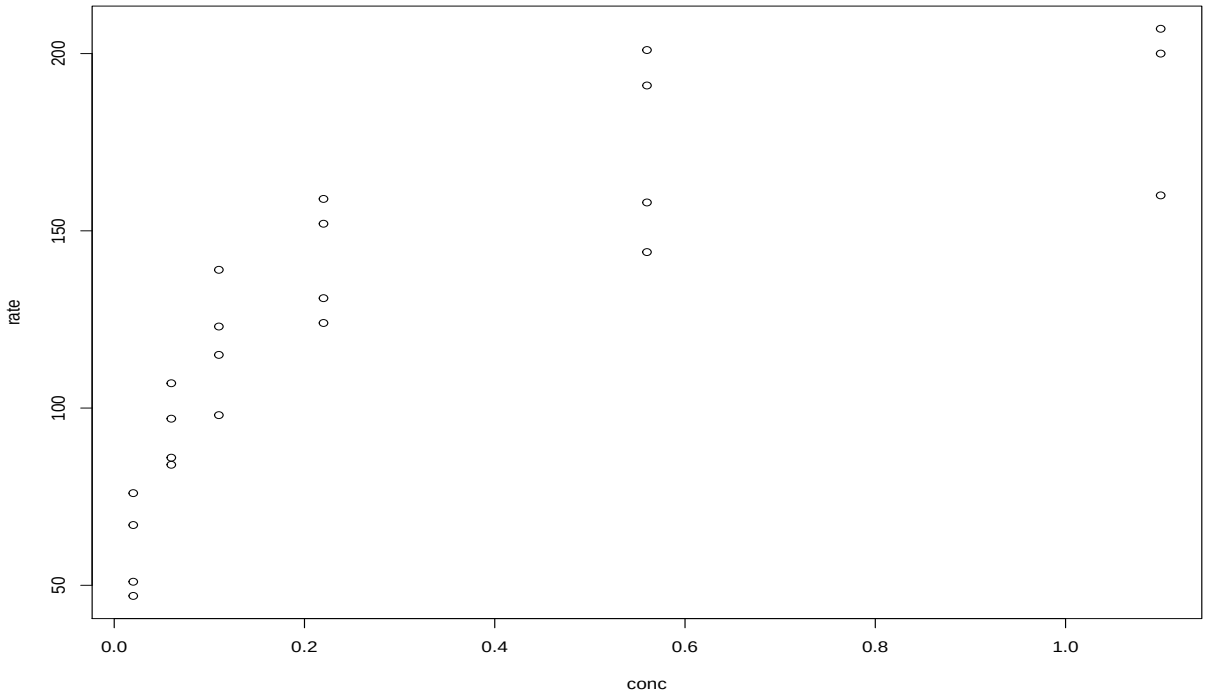
```
> install.packages("drat")

> drat::addRepo("statcourses")

> install.packages("BristolVis")
```

Overview of base graphics

- Graphics can be easily created in R and subsequently included in Word, Power Point, LATEX, etc.
- Let's plot a simple graphic:

```
> plot(rate ~ conc, data = Puromycin)
```

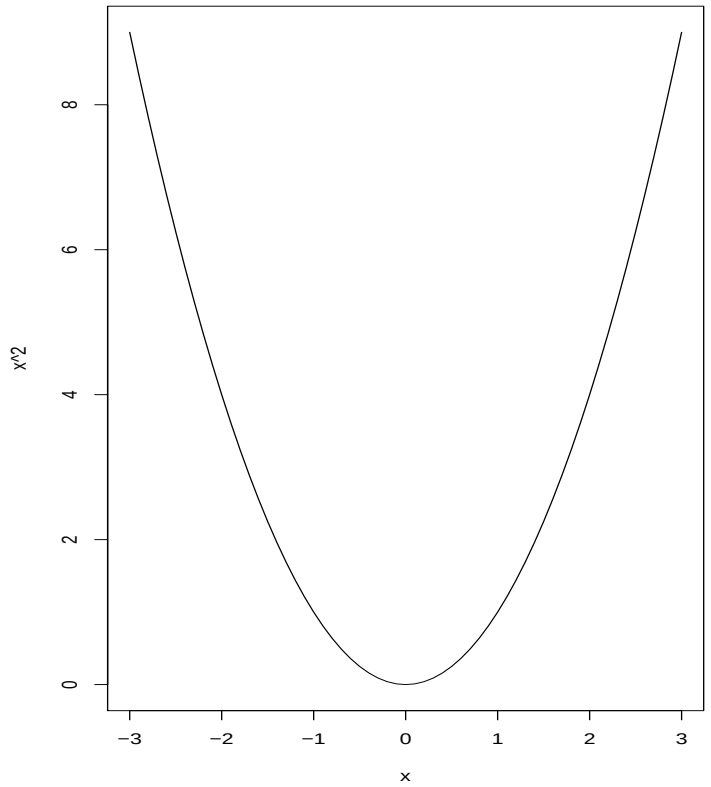
This implicitly opens a new graphics window.

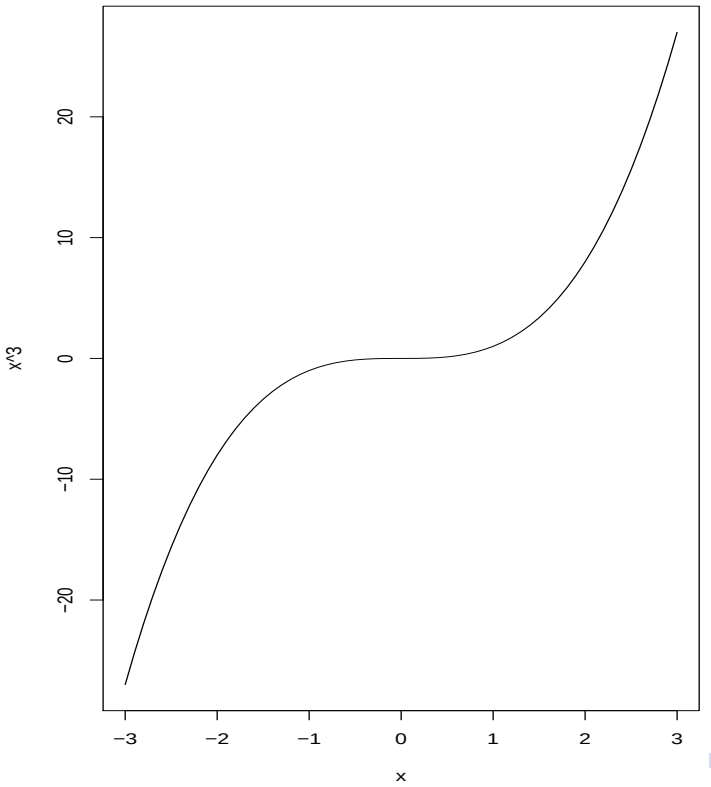


Overview of base graphics

- Change settings of active graphic window

```
> par(mfrow = c(1,2), mar = c(4,4,0,0) + 0.1)
> x <- seq(-3,3, 0.1)
> plot(x, x^2, type = 'l')
> plot(x, x^3, type = 'l')
```





Overview of base graphics

- Close active graphics window:

```
> dev.off()
```

- Close all graphics windows:

```
> graphics.off()
```

Histograms

Histograms

Simple histogram (next slide left)

```
> data(birthweight, package = "BristolVis")
> hist(birthweight$weightgain)
```

- with a fixed set of "bars": (`breaks` give number of cut-points)

```
> hist(birthweight$weightgain, breaks = 3)
```

- with a defined scale for the y-axis:

```
> hist(birthweight$weightgain, breaks = 3, ylim = c(0,
10))
```

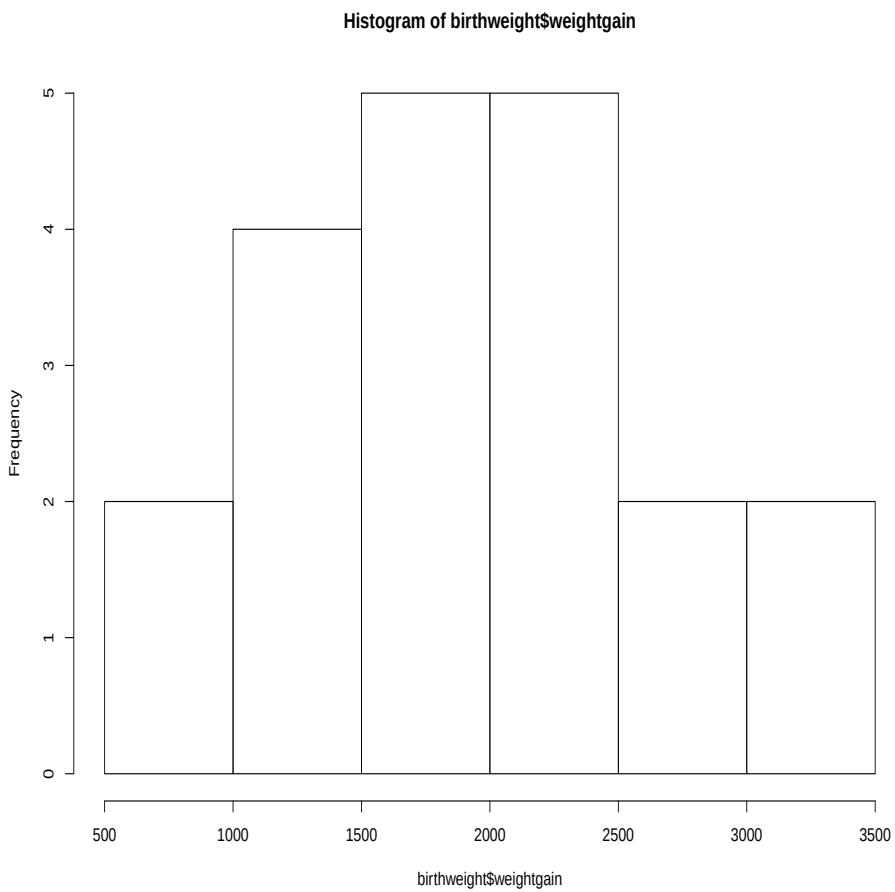
- with colored bars:

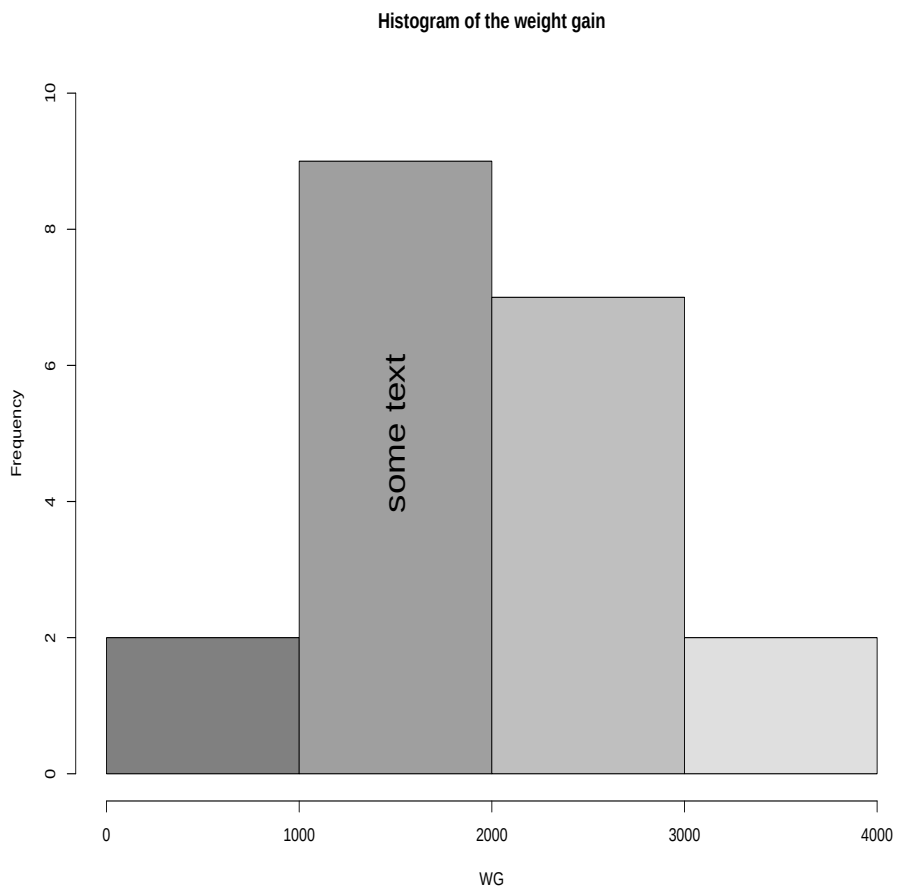
```
> hist(birthweight$weightgain, breaks = 3, ylim = c(0,
10), col = gray(4:7 / 7))
```

with arbitrary text, vertical and double text size (next slide right)

```
> text(1500, 5, "some text", srt = 90, cex = 2)
```

Histograms


Histogram of birthweight$weightgain
Frequency
0
1
2
3
4
5
500
1000
1500
2000
2500
3000
3500
birthweight$weightgain


Histogram of the weight gain
Frequency
0
2
4
6
8
10
some text
0
1000
2000
3000
4000
WG

Bar plots

Bar plots

Simple bar plot (next slide left)

```
> data(med, package = "BristolVis")
> treatment = table(med$treatment)
> barplot(treatment, ylim = c(0, 500))
```

- stacked bar plot:

```
> health_treat = table(med$health, med$treatment)
> barplot(health_treat, ylim = c(0, 500))
```

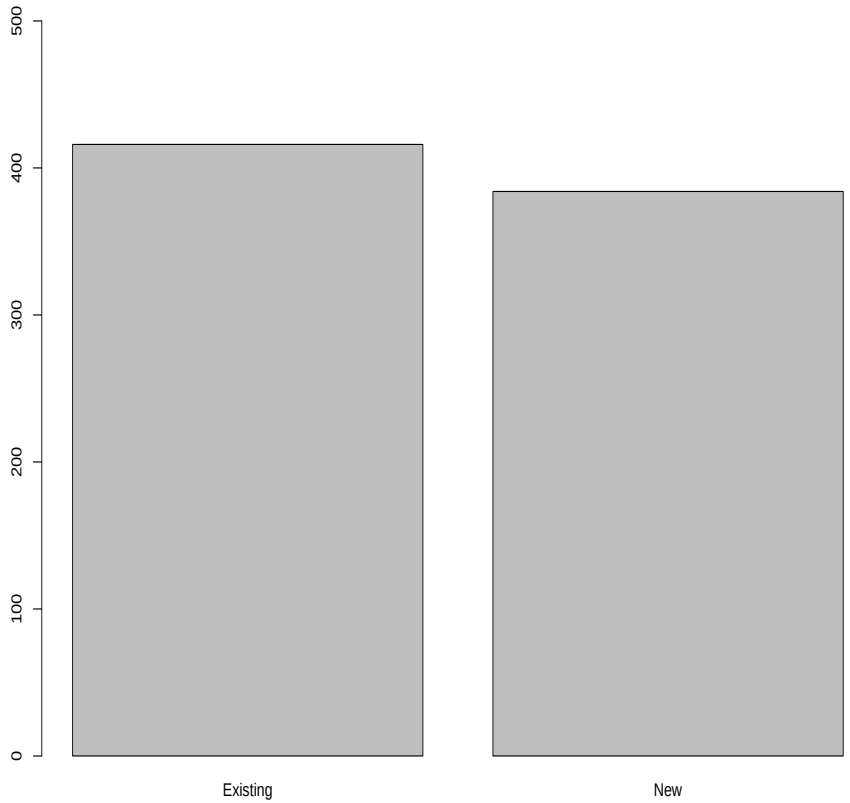
- with juxtaposed bars:

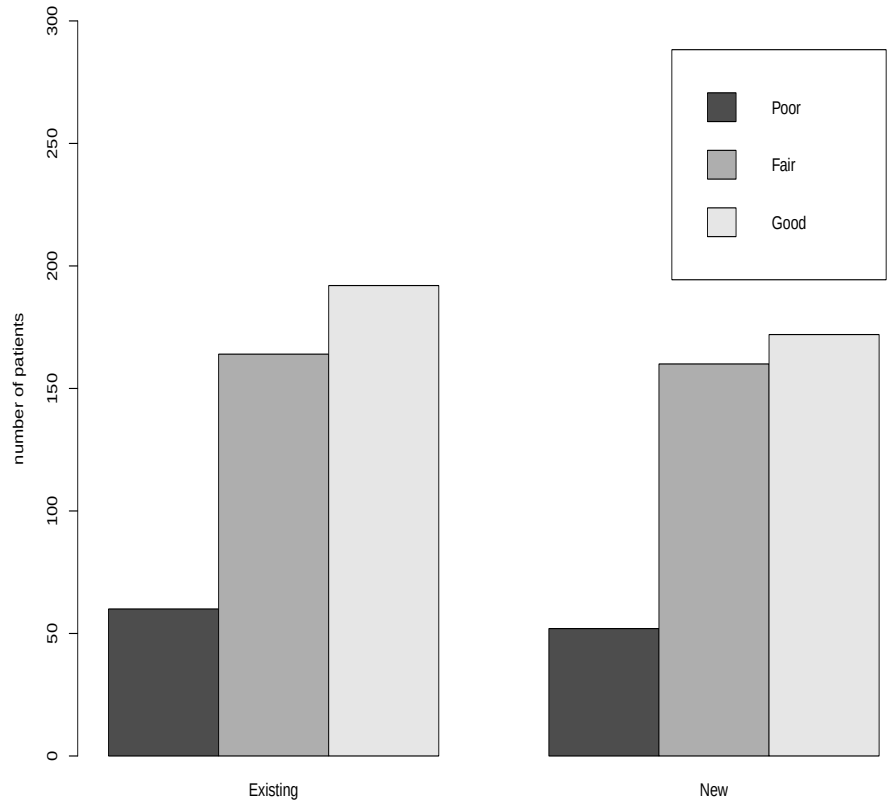
```
barplot(health_treat, ylim = c(0, 300), beside = TRUE)
```

with legend (next slide right)

```
> barplot(health_treat, ylim = c(0, 300), beside = TRUE,
legend.text = rownames(health_treat))
```

Bar plots


500
400
300
200
100
0
Existing
New


number of patients
300
250
200
150
100
50
0
Poor
Fair
Good
Existing
New

Box plots

Box plots

Simple box plot (next slide left)

```
> boxplot(birthweight$weight)
```

- grouped by sex:

```
> boxplot(weight ~ sex, data = birthweight)
```

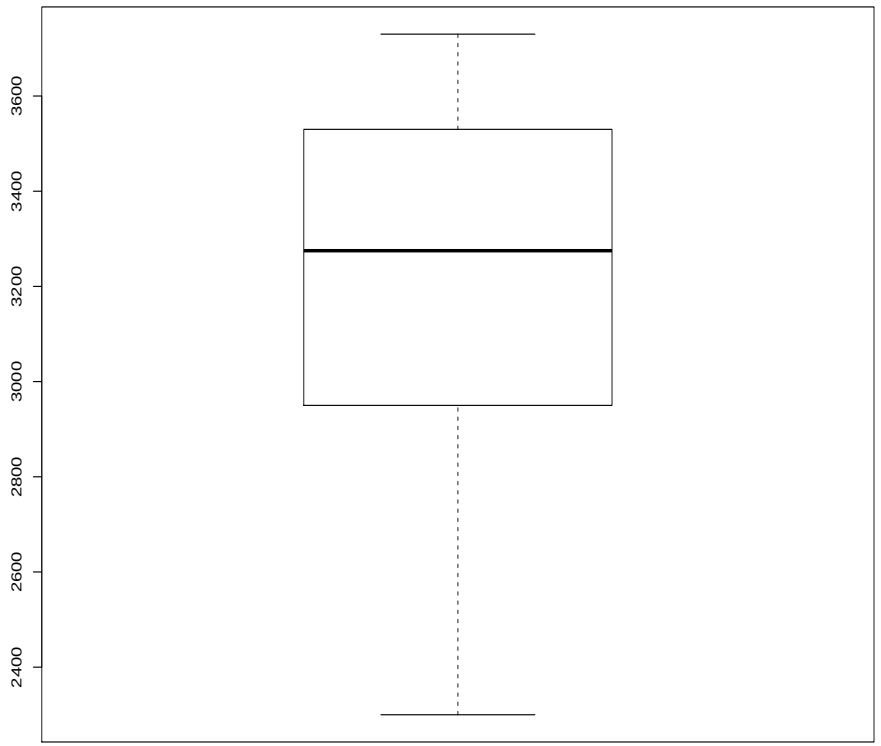
with new labels for sex, main title and a defined y-scale (next slide right)

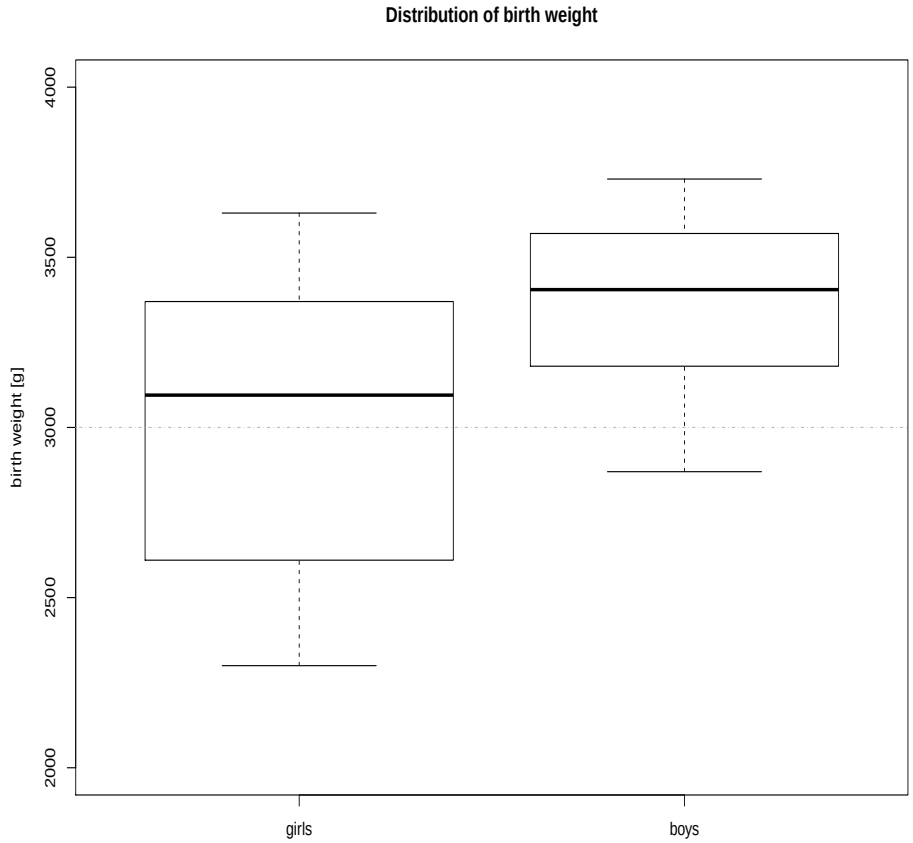
```
> boxplot(weight ~ sex, data = birthweight, names =
c("girls", "boys"), ylab = "birth weight [g]", ylim =
c(2000, 4000), main = "Distribution of birth weight")
```

- with additional line:

```
> abline(h = 3000, col = "gray", lty = 4)
```

Box plots


2400
2600
2800
3000
3200
3400
3600


Distribution of birth weight
birth weight [g]
2000
2500
3000
3500
4000
girls
boys

Scatter plots

Scatter plots

Simple scatter plot (next slide left)

```
> plot(weightgain ~ weight, data = birthweight)
```

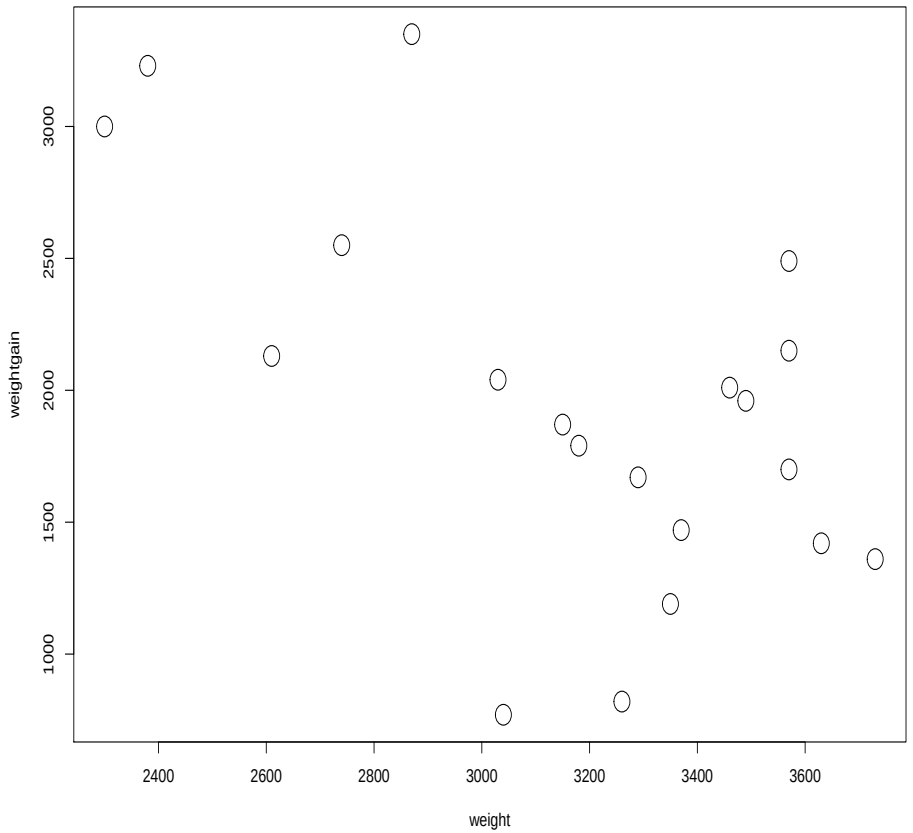
- for a subset:

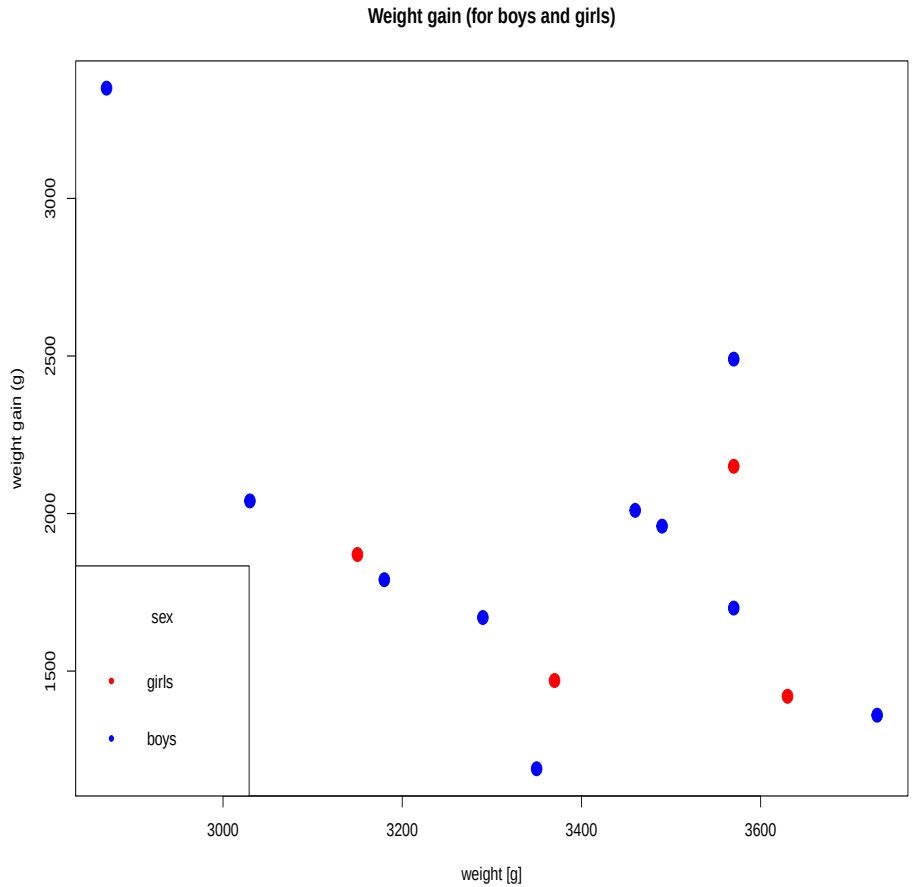
```
> plot(weightgain ~ weight, data = birthweight, subset =
(sex == "F"))
```

with colored solid points by sex and add a legend (next slide right)

```
> plot(weightgain ~ weight, data = birthweight, subset =
(sex == "M"), col = "blue", pch = 20)
> points(weightgain ~ weight, data = birthweight, subset =
(sex == "F"), col = "red", pch = 20) # add girls data
> legend("bottomleft", title = "sex", legend = c("girls",
"boys"), col = c("red","blue"), pch = 20) # add legend
```

Scatter plots





Examples of the graphic library

An overview of available demos for the built-in R library (graphics):

```
> demo(graphics)
```

Saving plots

Saving plots

- Graphics can be saved directly from within R using different devices, e.g. postscript(), pdf(), bmp(), (see `?Devices`)
- For publications vector graphics such as PDFs or postscript files are preferable to pixel graphics such as jpg, bmp, etc.

```
> pdf("Fig1.pdf", width = 10, height = 7)
> boxplot(weight ~ sex, data = birthweight)
> dev.off()
```

Useful links

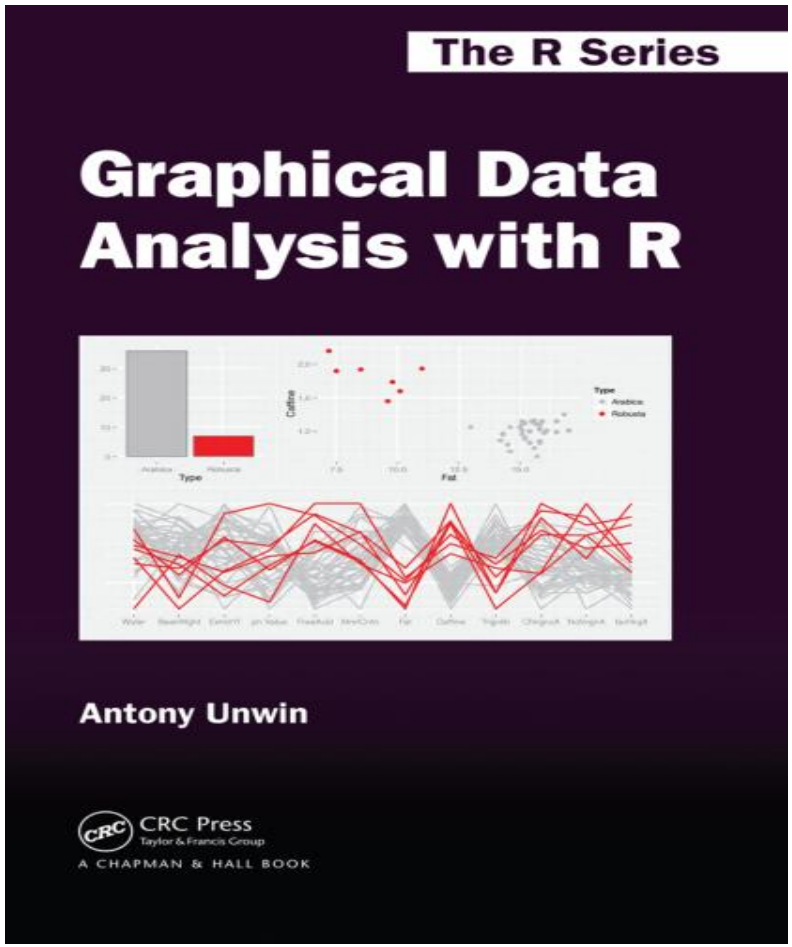
Choice of graphic colours

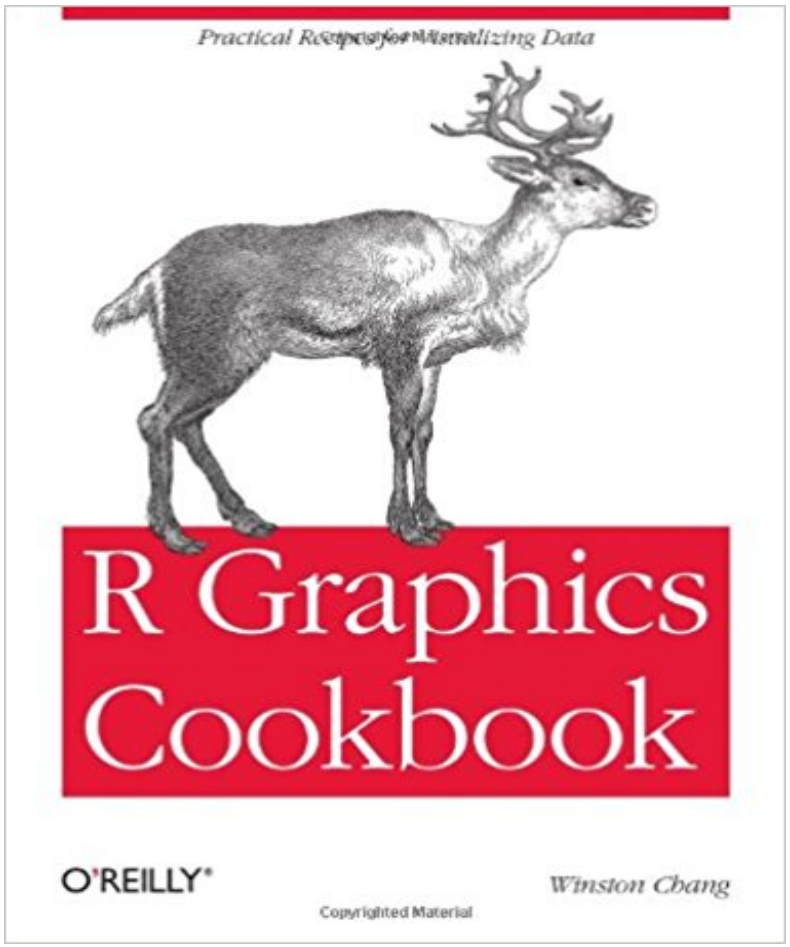
Names of colours in various formats:
http://colorbrewer2.org/

Web-page of the course

The BristolVis tool for learners of data visualisation using R:
https://github.com/statcourses/BristolVis

References


The R Series
Graphical Data Analysis with R
Antony Unwin
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK


Practical Recipes for Visualizing Data
R Graphics Cookbook
O'REILLY
Winston Chang
Copyrighted Material

Thank You