

Data Visualization with R Language

DENG, Xiaodong (xiaodong_deng@nuhs.edu.sg)

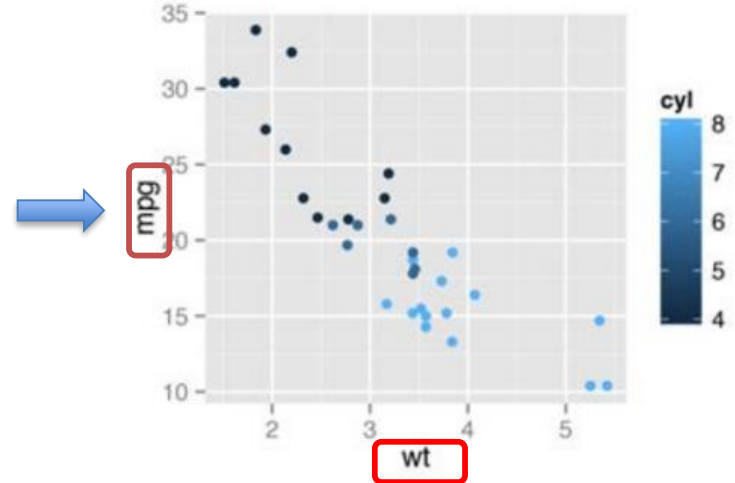
Research Assistant

Saw Swee Hock School of Public Health, National University of Singapore

Why Visualize Data?

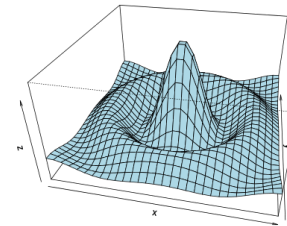
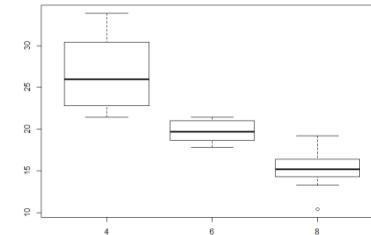
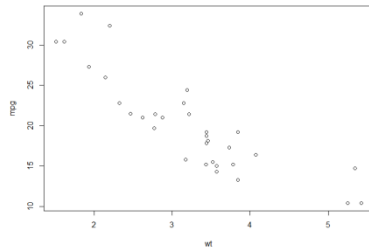
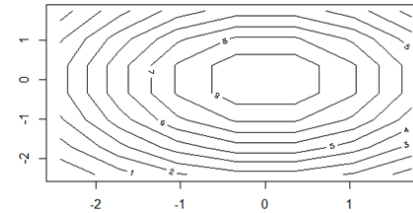
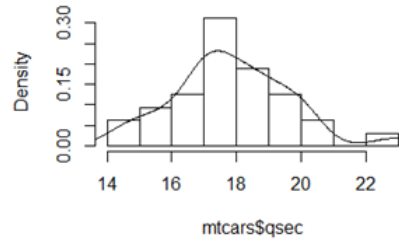
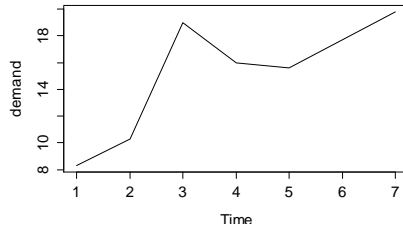
- For better presentation and communication

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2



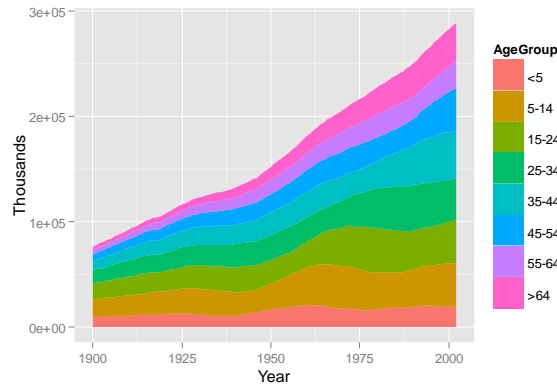
What Can We Do with R?

- Exploratory Visualization

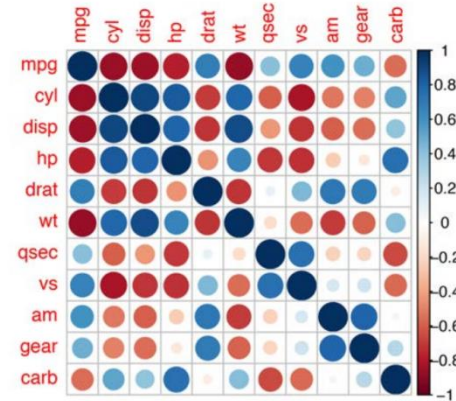


What Can We Do with R?

- Plotting for deeper insights



The change of age structure



Correlation matrix with correlation coefficient

Outline

- Basic Plotting Function in R
- Advanced Plotting in R: “ggplot2” packages

Data Visualization with R Language: Basic Plotting Function in R

Dataset I

- “BOD” data

The biochemical oxygen demand and corresponding time was recorded in an evaluation of water quality.

Time	1	2	3	4	5	6
demand	8.3	10.3	19.0	16.0	15.6	19.8

Dataset I

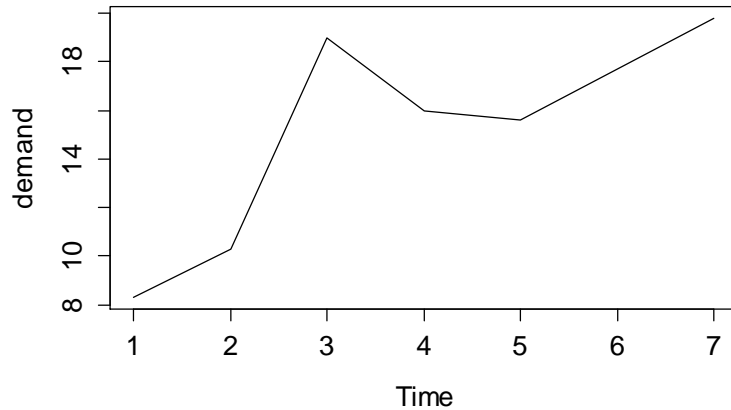
- “BOD” data

R object

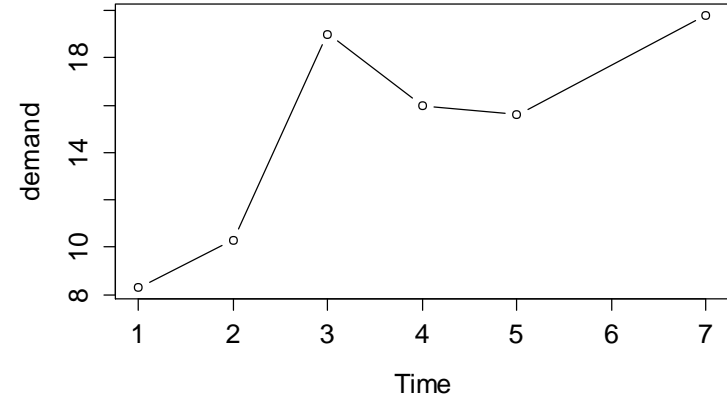
```
> BOD
  Time demand
1     1    8.3
2     2   10.3
3     3   19.0
4     4   16.0
5     5   15.6
6     7   19.8
```


Basic Plotting Functions

Line Plot



Line only



Line and dot

Basic Plotting Functions

Commonly Used Function for Plotting

```
plot(x = BOD$Time, y = BOD$demand, type="l")
```

Function Name
Function we use

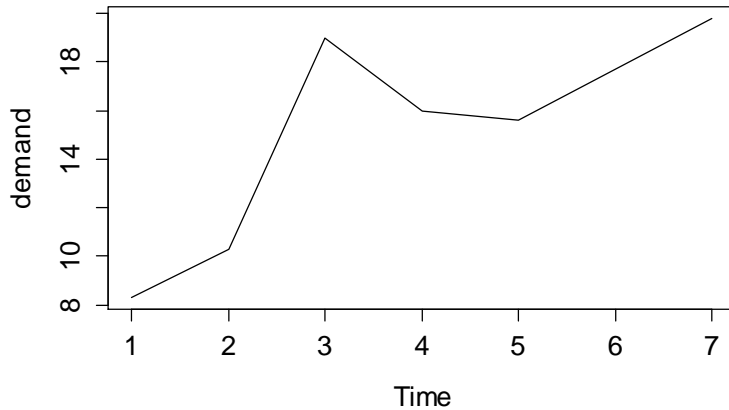
Arguments - 1
Specify the data we
use

Arguments - 2 (optional)
Plotting options

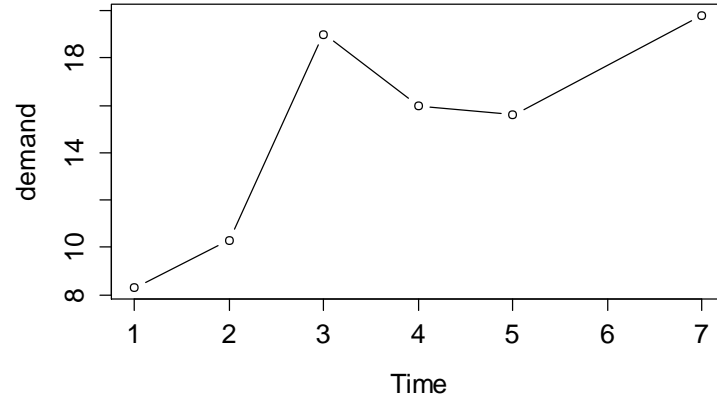
Basic Plotting Functions

Line Plot

```
plot(x=BOD$Time, y=BOD$demand, type="l")
```



```
plot(x=BOD$Time, y=BOD$demand, type="b")
```



Basic Plotting Functions

Line Plot

Previous usage

```
plot(x=BOD$Time, y=BOD$demand, type="l")
```

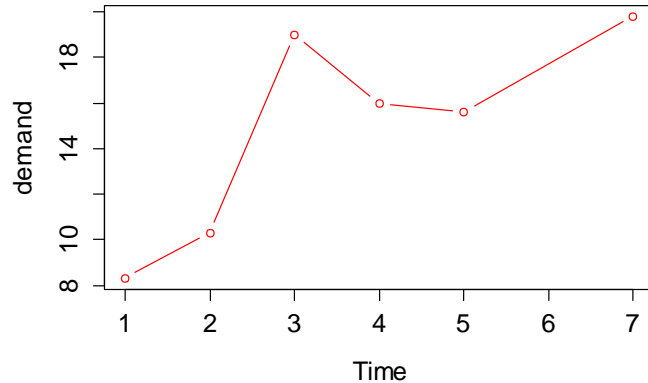
Alternative commands

```
plot(demand ~ Time, data=BOD, type="l")
```

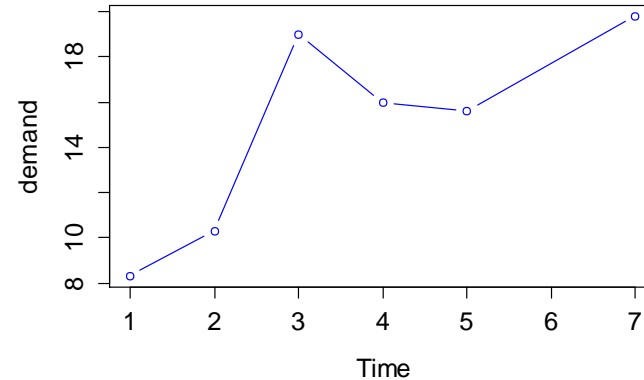
Basic Plotting Functions

More Options - Color

```
plot(demand ~ Time, data=BOD,  
     type="b", col="red")
```



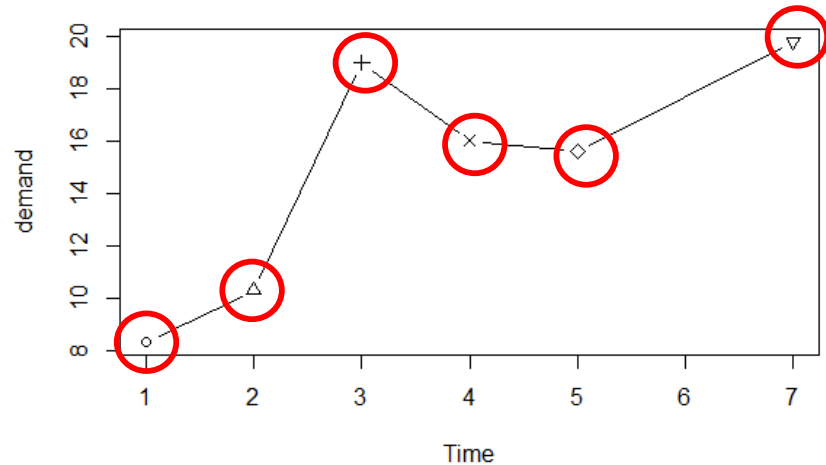
```
plot(demand ~ Time, data=BOD,  
     type="b", col="blue")
```



Basic Plotting Functions

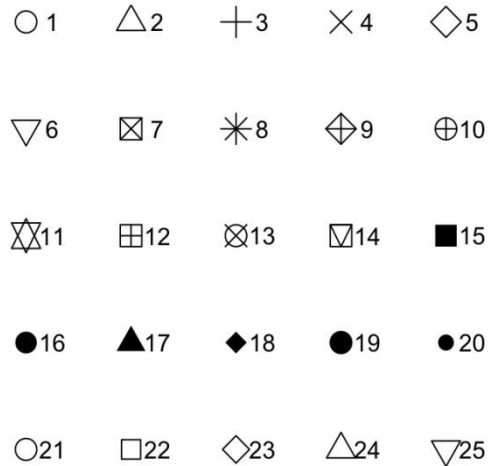
More Options – Shape

```
plot(demand ~ Time, data=BOD,  
     type="b", pch=1:6)
```



Basic Plotting Functions

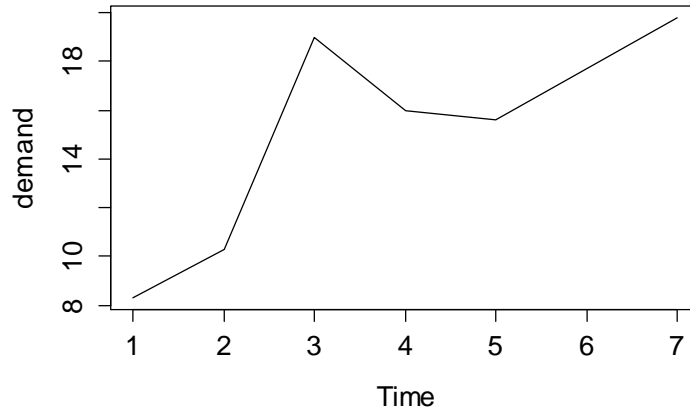
More Options on shape by using “pch” argument



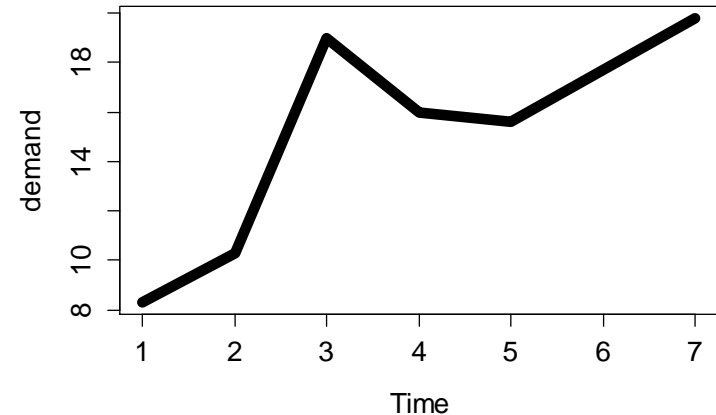
Basic Plotting Functions

More Options – Line Width

```
plot(demand ~ Time, data=BOD, type="l")
```



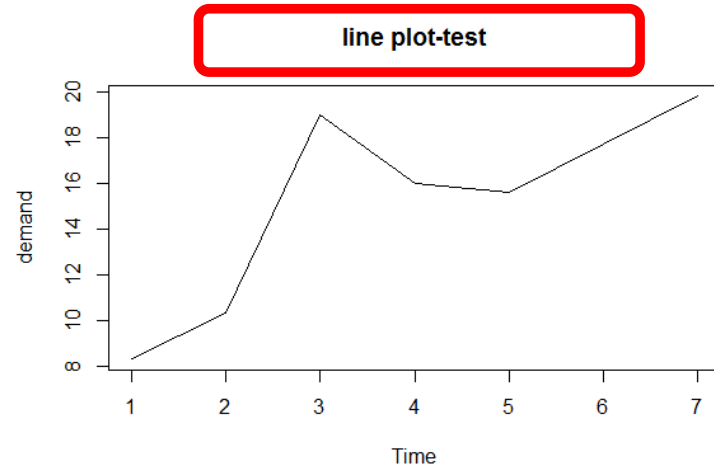
```
plot(demand ~ Time, data=BOD, type="l", lwd=5)
```



Basic Plotting Functions

More Options – Add Title

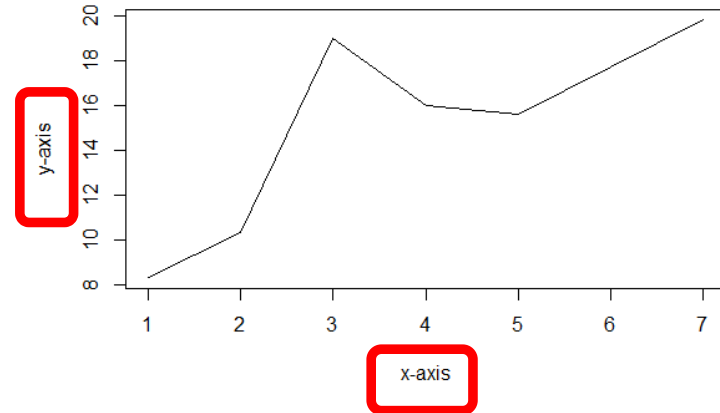
```
plot(demand ~ Time, data=BOD,  
     type="l",  
     main="line plot-test")
```



Basic Plotting Functions

More Options – Change Labels

```
plot(demand ~ Time, data=BOD,  
     type="l",  
     xlab="x-axis",  
     ylab="y-axis")
```



Basic Plotting Functions

More Options

To explore more options in plotting functions,

?par

Graphical Parameters

`adj`

The value of `adj` determines the way in which text strings are justified in `text`, `mtext` and `title`. A value of 0 produces left-justified text, 0.5 (the default) centered text and 1 right-justified text. (Any value in $[0, 1]$ is allowed, and on most devices values outside that interval will also work.)

Note that the `adj` argument of `text` also allows `adj = c(x, y)` for different adjustment in x- and y- directions. Note that whereas for `text` it refers to positioning of text about a point, for `mtext` and `title` it controls placement within the plot or device region.

`ann`

If set to `FALSE`, high-level plotting functions calling `plot.default` do not annotate the plots they produce with axis titles and overall titles. The default is to do annotation.

`ask`

logical. If `TRUE` (and the R session is interactive) the user is asked for input, before a new figure is drawn. As this applies to the device, it also affects output by packages `grid` and `lattice`. It can be set even on non-screen devices but may have no effect there.

This is not really a graphics parameter, and its use is deprecated in favour of `devAskNewPage`.

• • • • •

Dataset II

- “mtcars” data

The data was extracted from the 1974 *Motor Trend* US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21	6	160	110	3.9	2.62	16.46	0	1	4	4
Mazda RX4 Wag	21	6	160	110	3.9	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.44	17.02	0	0	3	2

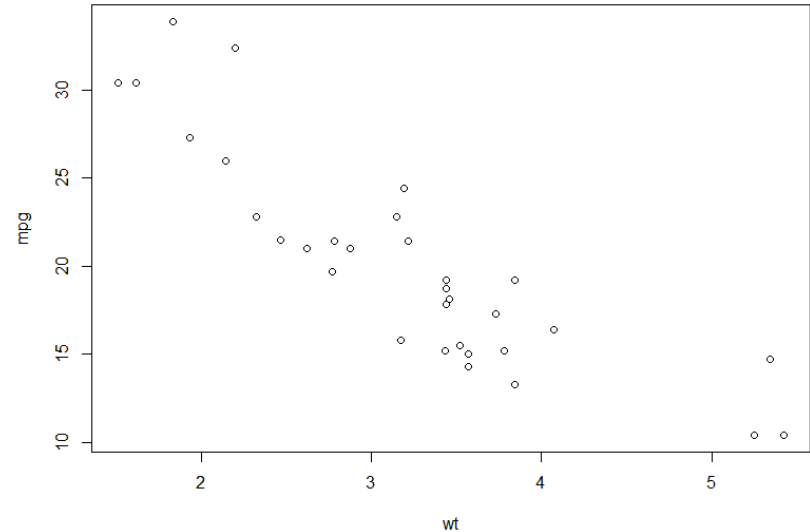
---: numerical
 ---: categorical

.....

Basic Plotting Functions

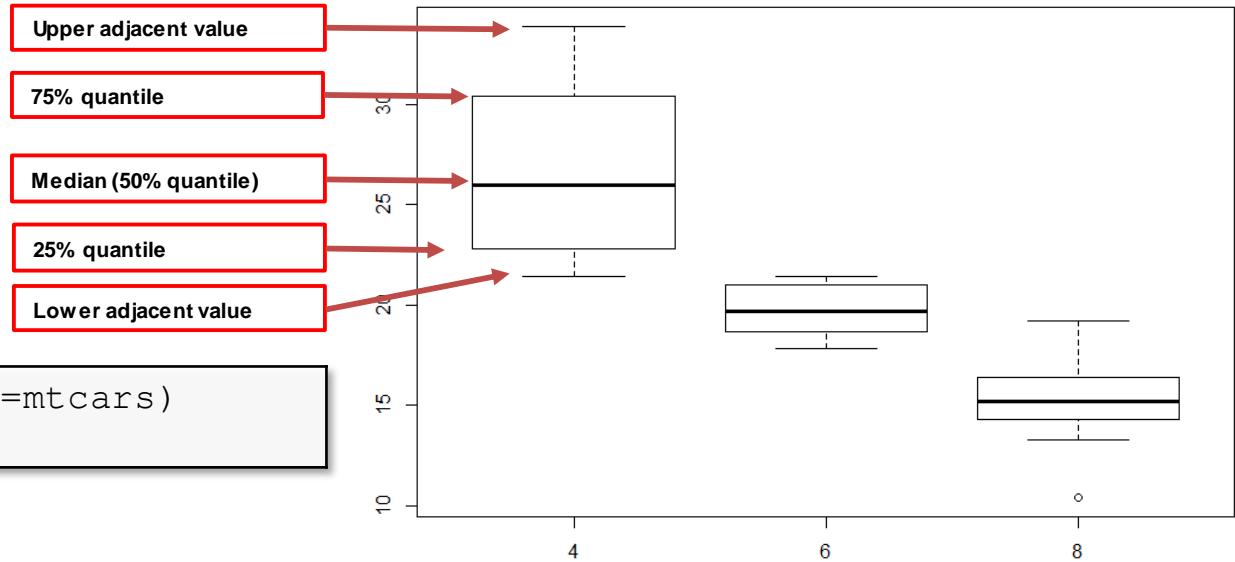
Scatter Plot

```
plot(mpg ~ wt, data = mtcars,  
     type = "p")
```



Basic Plotting Functions

Boxplot

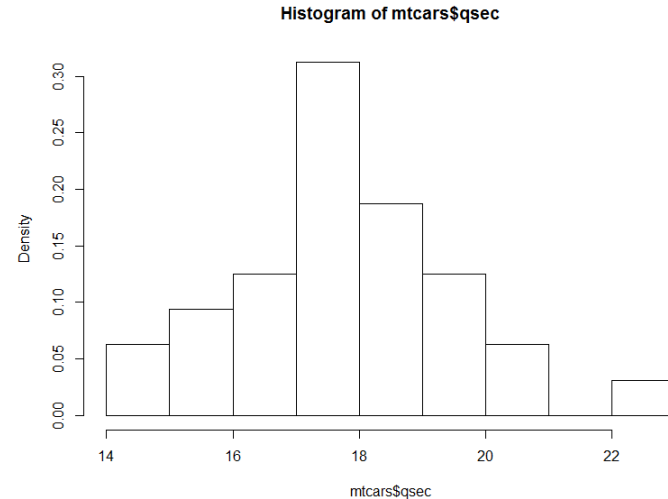


```
boxplot(mpg ~ cyl, data=mtcars)
```

Basic Plotting Functions

Histogram

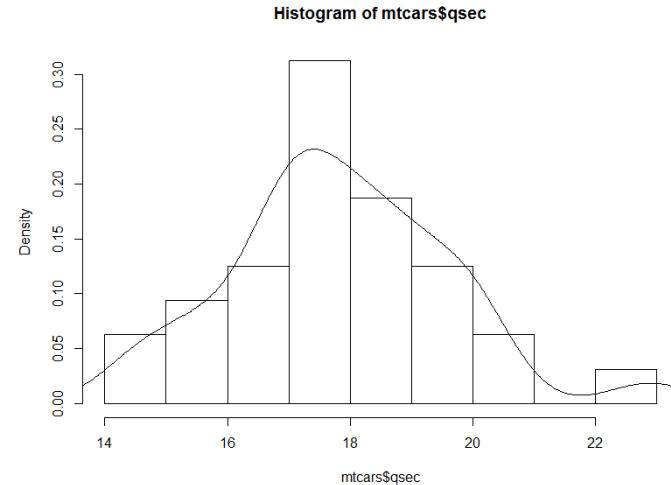
```
hist(mtcars$qsec, freq = FALSE)
```



Basic Plotting Functions

Histogram

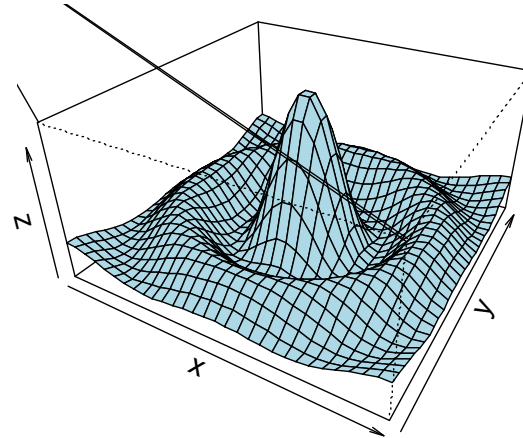
```
hist(mtcars$qsec, freq = FALSE)  
d <- density(mtcars$qsec)  
lines(d)
```



Basic Plotting Functions

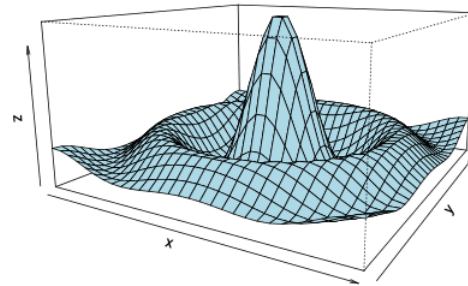
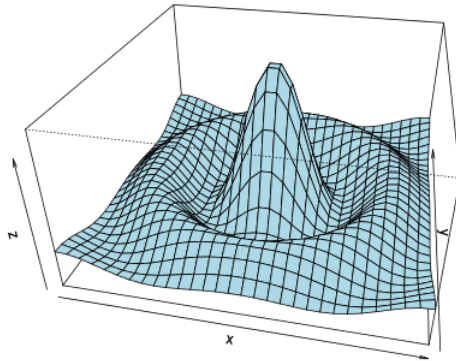
3-D Graphics

```
x <- seq(-10, 10, length= 30)
y <- x
f <- function(x, y) {
  r <- sqrt(x^2+y^2)
  10 * sin(r)/r
}
z <- outer(x, y, f)
persp(x, y, z, theta = 30, phi = 30,
      expand = 0.5,
      col = "lightblue")
```



Basic Plotting Functions

3-D Graphics

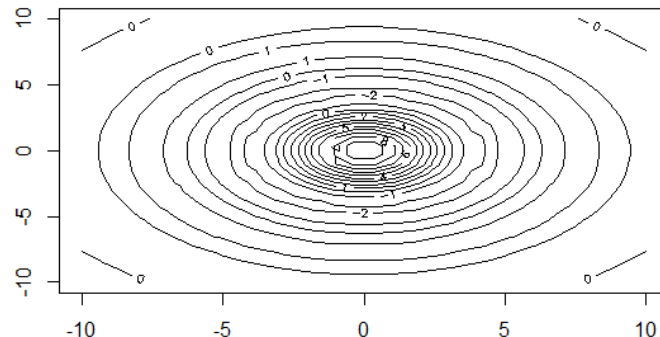


(GIF made with “animation” package)

Basic Plotting Functions

Contour Plot

```
x <- seq(-10, 10, length= 30)
y <- x
f <- function(x, y) {
  r <- sqrt(x^2+y^2)
  10 * sin(r)/r
}
z <- outer(x, y, f)
contour(x, y, z)
```



Data Visualization with R Language: “ggplot2” package

What is “ggplot2”

- A packaged developed by **Hadley Wickham** (Chief scientist at Rstudio);
- A plotting system for R, based on **the grammar of graphics** by Leland Wilkinson
 - An abstraction which makes thinking, reasoning and communicating graphics easier
 - Enables us to concisely describe the components of a graphic
- Produce complex **multi-layered** graphics

An Example

- Call “ggplot2” package

```
library(ggplot2)
```

“ggplot2” is a separate package, so we need to “call” it before we use it.

An Example

- A “wrong” example

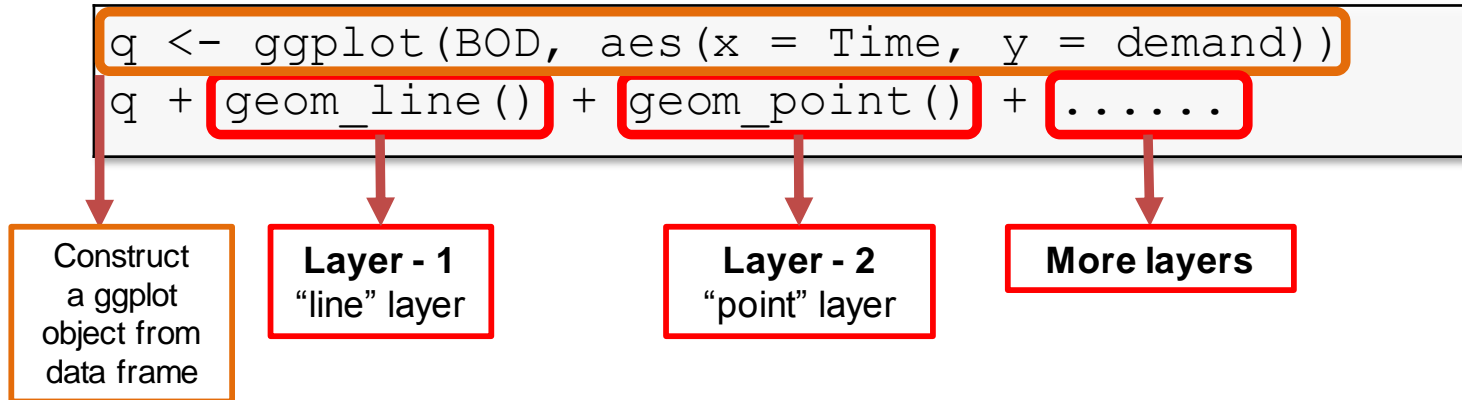
```
library(ggplot2)  
ggplot(BOD, aes(x = Time, y = demand))
```

```
Error: No layers in plot
```

- ggplot() is typically used to construct a plot incrementally
- Use + operator to add layers to the existing ggplot object
- Explicitly decide which layers are added and the order in which they are added

Common Pattern

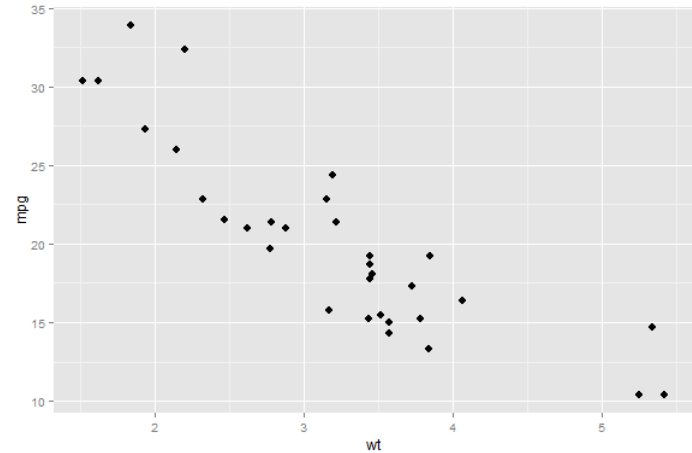
- Common pattern in the usage of “ggplot2”



An Example

- Add a “point” layer

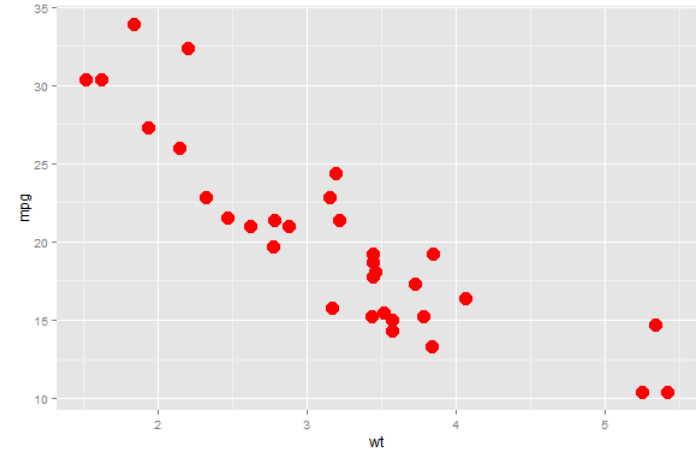
```
q_0 <- ggplot(mtcars, aes(x = wt, y = mpg))  
q_1 <- q_0 + geom_point()  
q_1
```



An Example

- Add a “point” layer

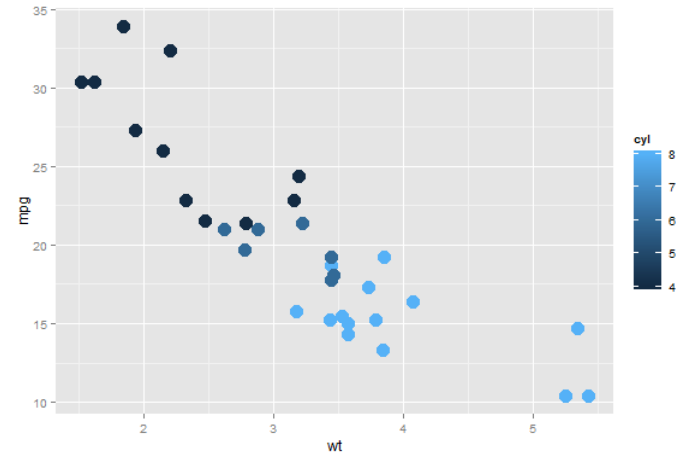
```
q_0 <- ggplot(mtcars, aes(x = wt, y = mpg))  
q_1 <- q_0 + geom_point(col="red", size=5)  
q_1
```



An Example

- Add a “point” layer

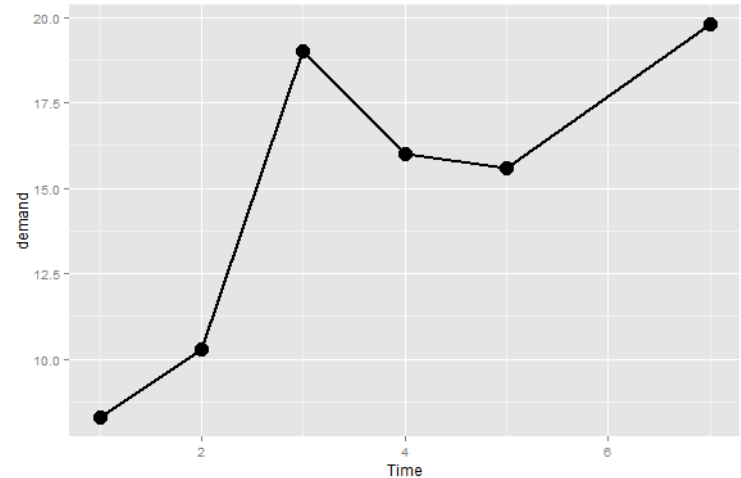
```
q_0 <- ggplot(mtcars, aes(x = wt, y = mpg,  
                           col=cyl))  
q_1 <- q_0 + geom_point(size=5)  
q_1
```



An Example

- Add a multiple layer

```
q_0 <- ggplot(BOD, aes(x = Time, y = demand))  
q_1 <- q_0 + geom_point()  
q_1  
q_2 <- q_1 + geom_line()  
q_2
```



An Example

- Add a multiple layer

```
q_0 <- ggplot(BOD, aes(x = Time, y = demand))  
q_1 <- q_0 + geom_line() + geom_point()  
q_1
```

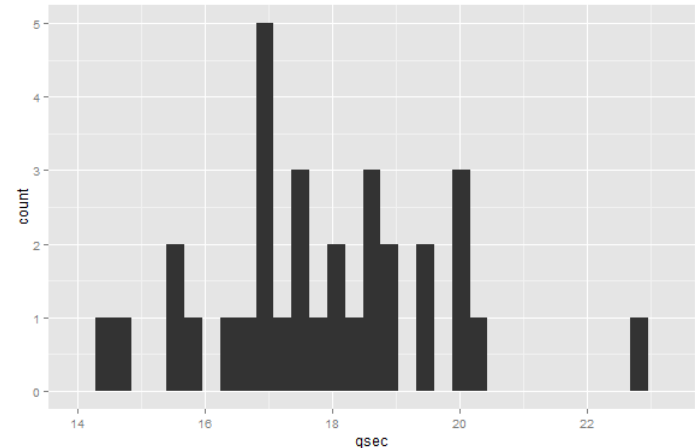
Layer - 1
"line" layer

Layer - 2
"point" layer

“ggplot2” Other Plots

- Histogram

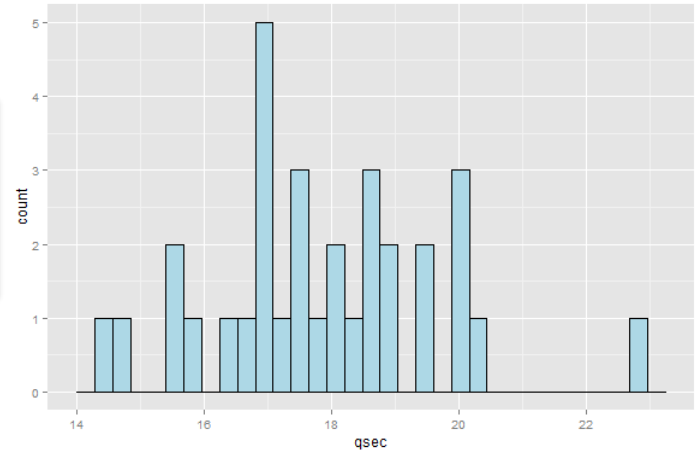
```
q_0 <- ggplot(mtcars, aes(x=qsec))  
q_1 <- q_0 + geom_histogram()  
q_1
```



“ggplot2” Other Plots

- Histogram

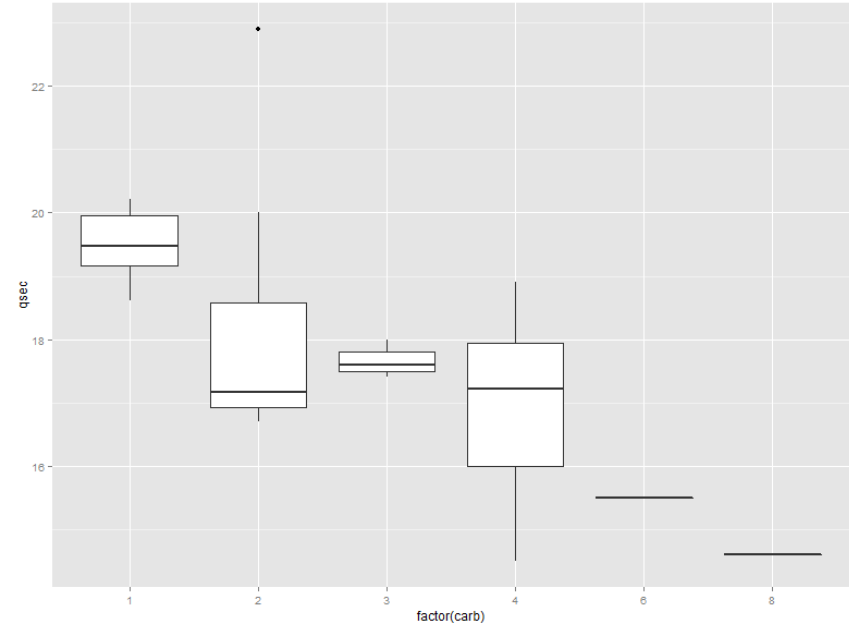
```
q_0 <- ggplot(mtcars, aes(x=qsec))  
q_1 <- q_0 + geom_histogram(col="black",  
                             fill = "lightblue")  
q_1
```



“ggplot2” Other Plots

- Boxplot

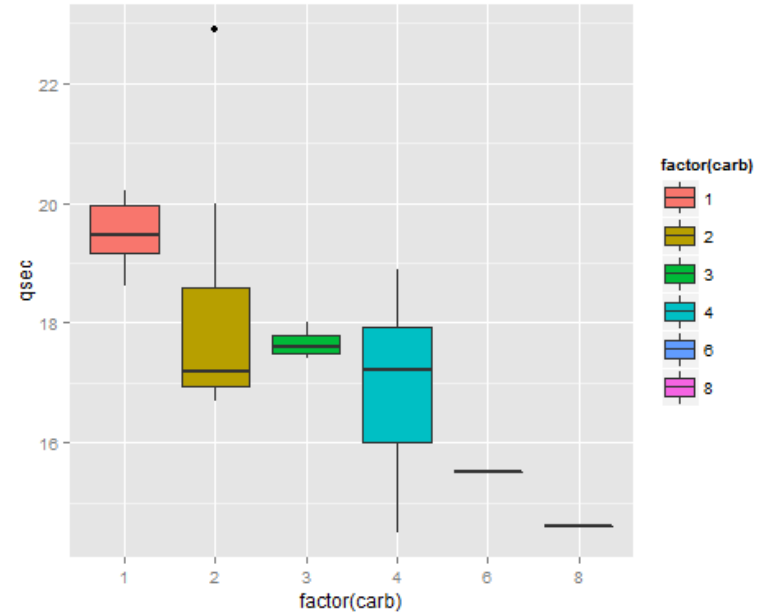
```
q_0 <- ggplot(mtcars, aes(y=qsec,  
                           x=factor(carb)))  
q_1 <- q_0 + geom_boxplot()  
q_1
```



“ggplot2” Other Plots

- Boxplot

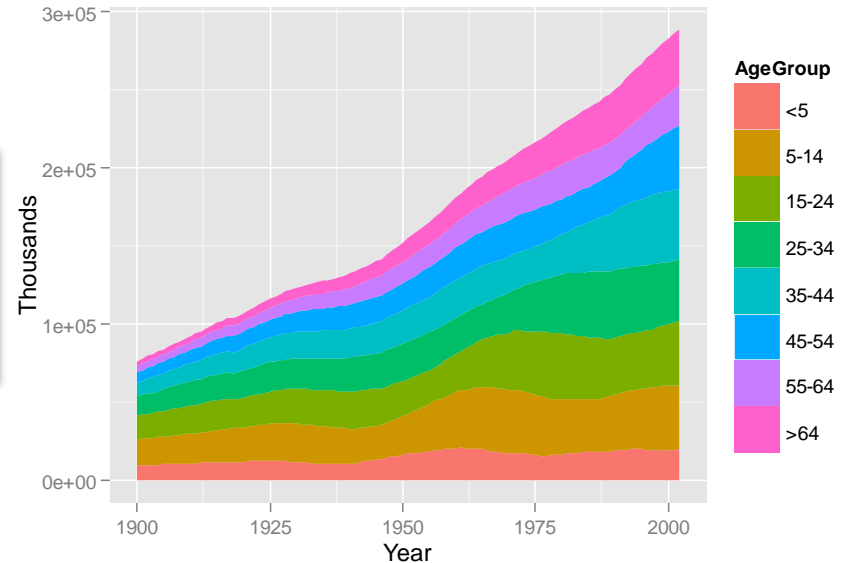
```
q_0 <- ggplot(mtcars, aes(y=qsec,  
                           x=factor(carb),  
                           fill=factor(carb)))  
q_1 <- q_0 + geom_boxplot()  
q_1
```



“ggplot2” Other Plots

- Stacked Area Graph

```
library(gcookbook)
ggplot(uspopage, aes(x=Year, y=Thousands,
  fill=AgeGroup)) + geom_area()
```



“ggplot2” Other Options

- Others Applicable Layers

`geom_abline`

`geom_area`

`geom_bar`

`geom_boxplot`

`geom_contour`

`geom_density`

`geom_dotplot`

`geom_histogram`

`geom_line`

`geom_map`

`geom_path`

`geom_point`

`geom_rug`

`geom_step`

`geom_text`

`geom_violin`

Reference

- Hadley Wickham, A Layered Grammar of Graphics, <http://vita.had.co.nz/papers/layered-grammar.pdf>
- <http://fbmap.bitasthetics.com/>
- <https://support.rstudio.com/hc/en-us/articles/200551906-Interactive-Plotting-with-Manipulate>
- <http://www.cookbook-r.com/Graphs/>
- Book: Winston Chang, R Graphics Cookbook, O'Reilly

Thanks