Effective graphical displays and *R*
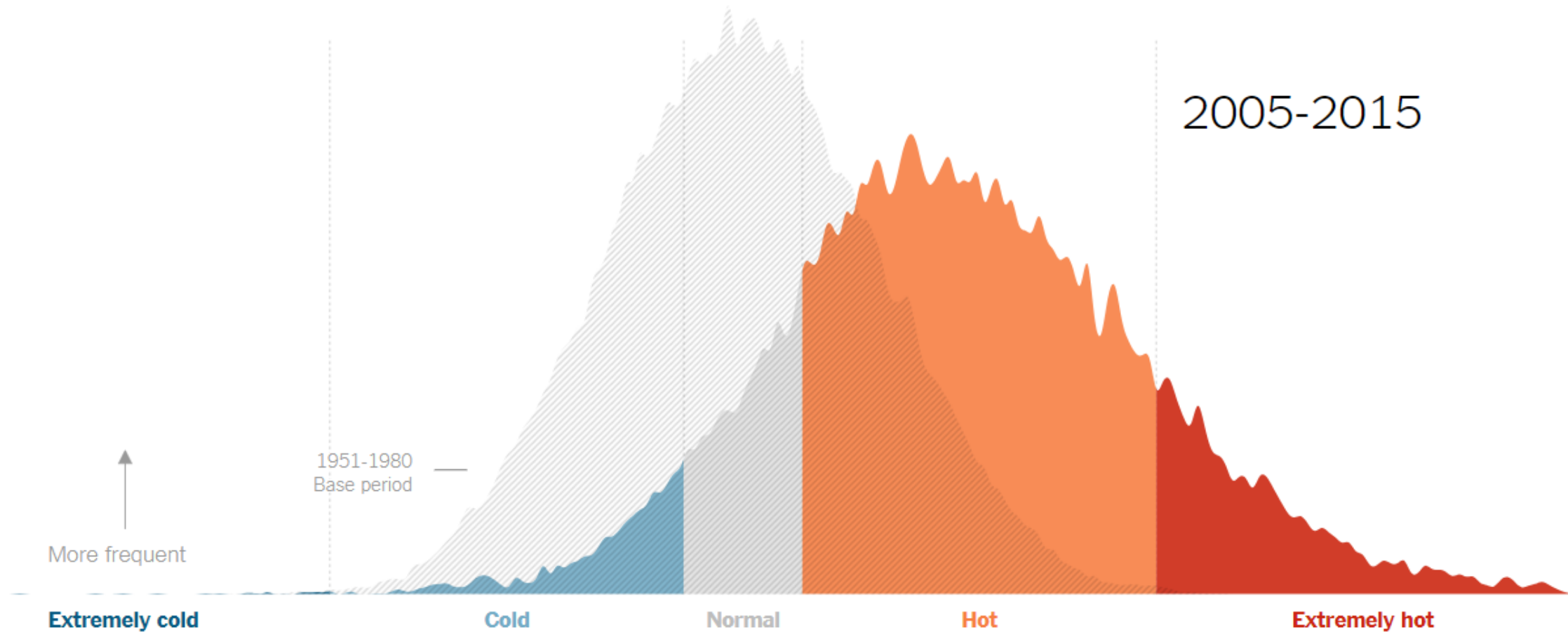
Yishih Chung

National Chiao Tung University

Graphics for communication

- Graphics are used for communicating to both yourself and your audiences.

- Plots should be as self-explanatory as possible.

- References for visualization: [The Truthful Art](https://www.amazon.com/dp/0321934075/ref=cm_sw_su_dp)

- Websites for [ggplot2](https://r4ds.had.co.nz/graphics-for-communication.html) and its [extensions](http://www.ggplot2-exts.org/)

What's Going On in This Graph? | May 8, 2018

By THE LEARNING NETWORK   MAY 3, 2018

Summer temperatures
in the Northern Hemisphere

2005-2015

1951-1980
Base period

More frequent

Extremely cold   **Cold**   Normal   **Hot**   **Extremely hot**

Planning the graph

- Select the type of graph on the basis of
  - The intended message or research question

  (comparison, distribution, correlation, evolution)
  - The structure of the dataset

  (number and nature of variables)

Patient	Gender	Weight [kg]	Conc. [ng/ml]
1	female	75	31.3
2	male	81	16.5
3	female	52	11.8
⋮	⋮	⋮	⋮

Discrete variable

Continuous variables
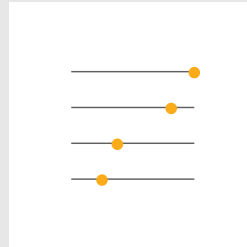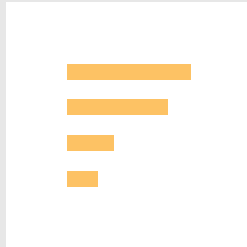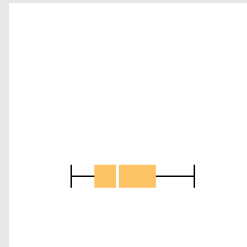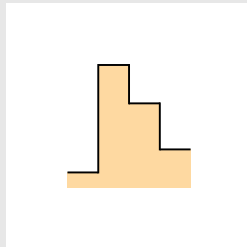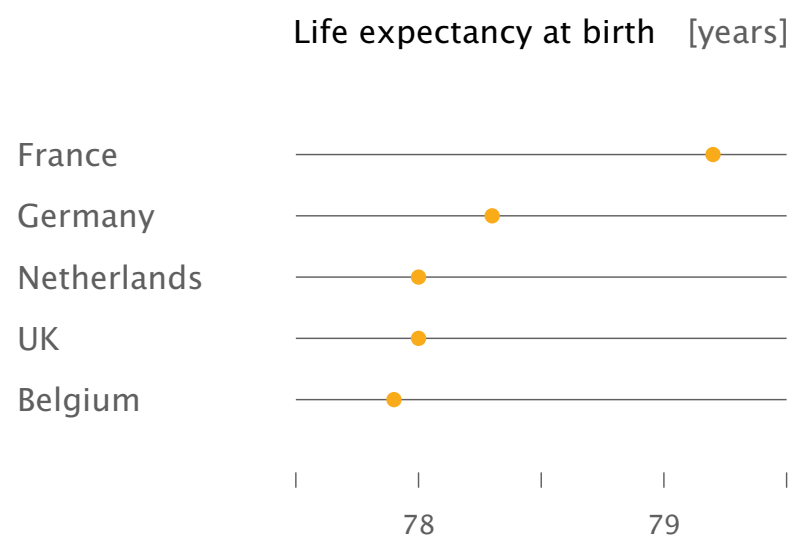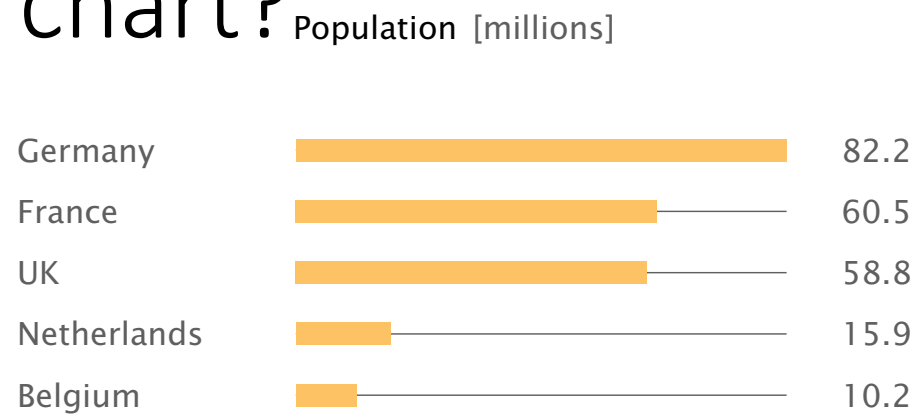
Designing the graph

- Select the basic design on the basis of the intended message or research question and the number of continuous variables.

- Distinguish the subsets within one panel or display them in as many juxtaposed panels (or combine these two approaches to represent several discrete variables).

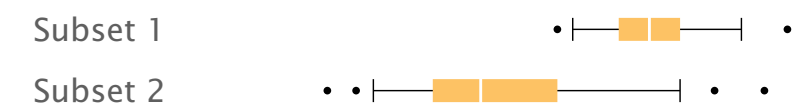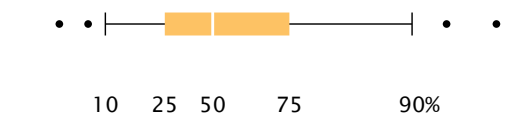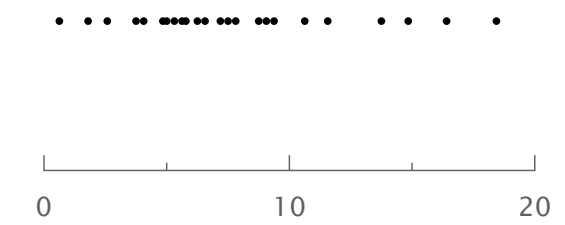# Subsets displayed in juxtaposed panels
(with identical scales)

t_0 Faster diffusion

t_1

t_2

Faster drift
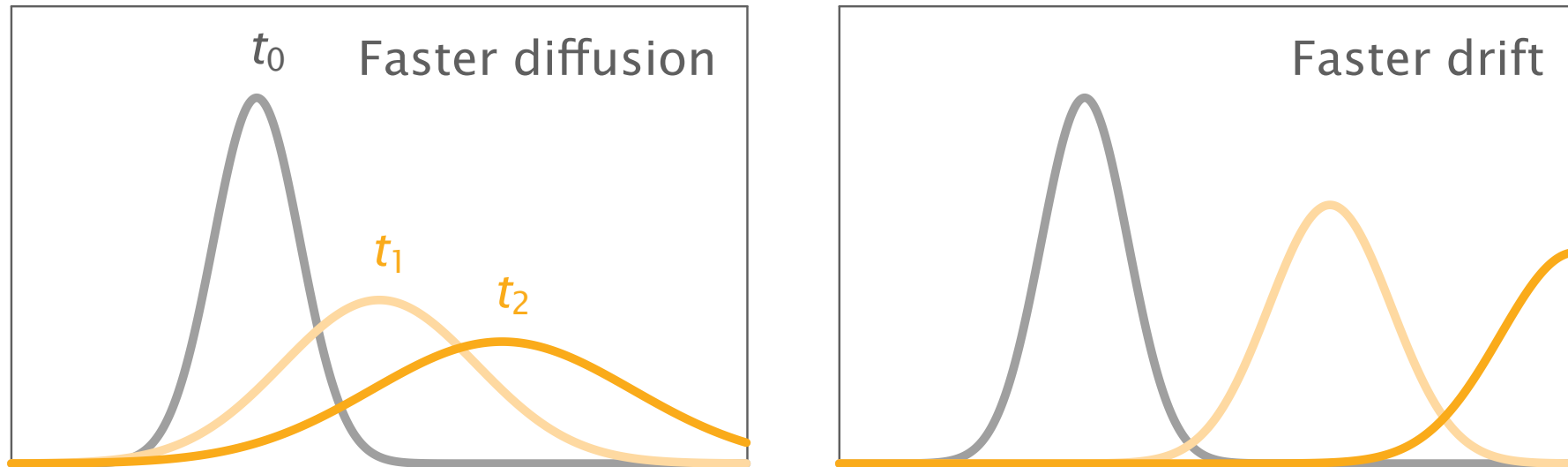
Subsets distinguished
within one panel

5

Compare data in a length presentation or a pie chart?

Showing the entire dataset as points along a scale or a boxplot

Population [millions]

Germany	▬▬▬▬▬▬▬▬	82.2
France	▬▬▬▬▬	60.5
UK	▬▬▬▬▬	58.8
Netherlands	▬	15.9
Belgium	▬	10.2

Life expectancy at birth [years]

France
Germany
Netherlands
UK
Belgium

78   79

0   10   20

10 25 50 75   90%

Subset 1
Subset 2

Scatter plots are powerful to reveal correlation

Unemployment rate [%]
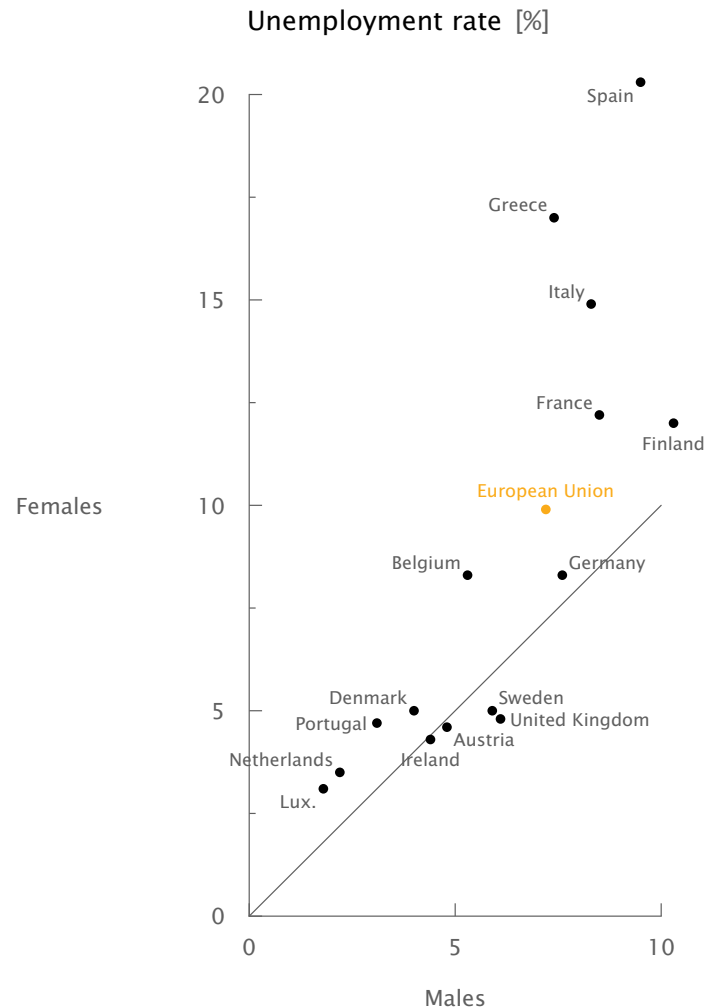
Evolution is best shown by lines. Variables expressed in different units must be drawn in different panels, with a common horizontal scale.
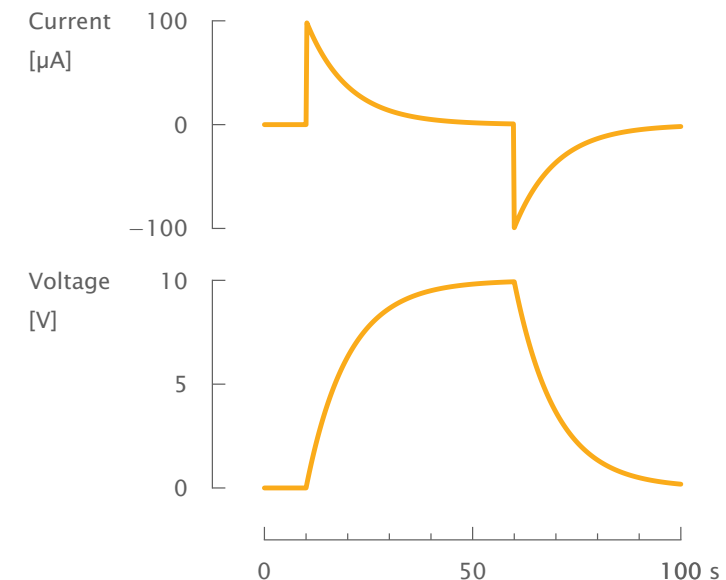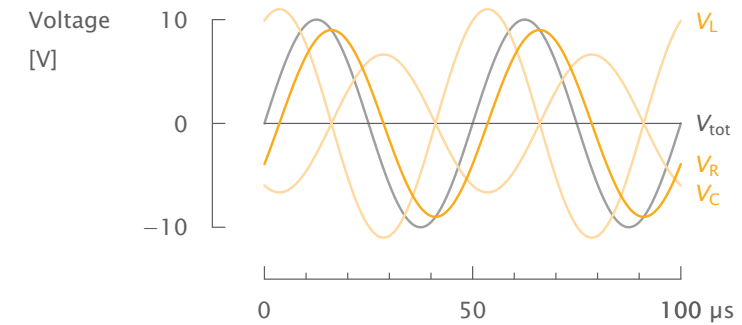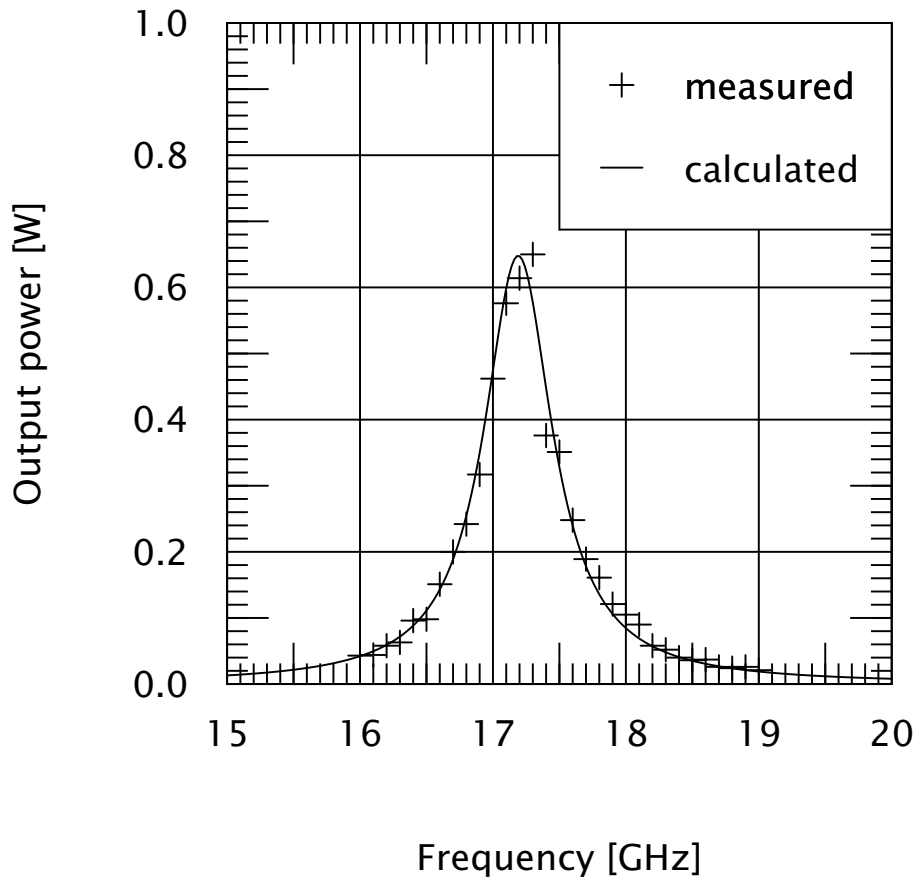
A poor graph

The graph exhibits a very low signal-to-noise ratio, with excessive tick marks and uncalled-for grid lines, and comparatively little ink to represent the data.
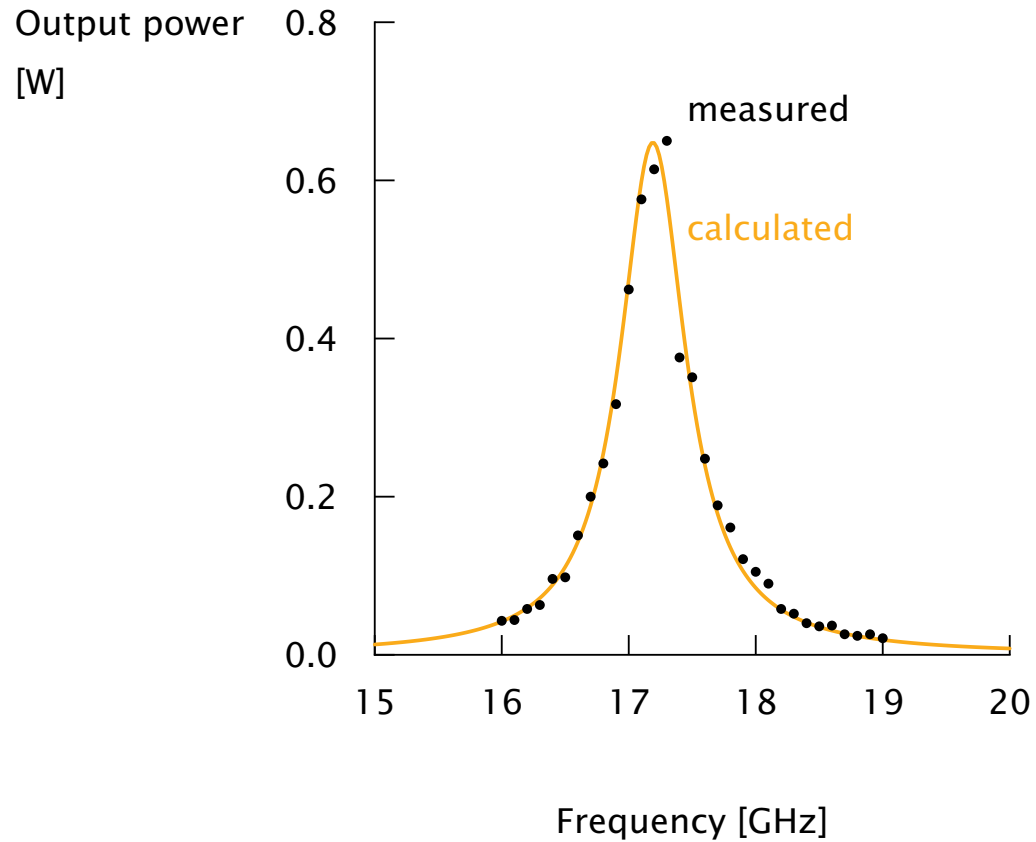
The graph is not intuitive, for the separate legend (a key to the symbols) prevents global processing. In a sense, it is a graph to *read*, not a graph to *view*.

A good graph

The graph is plainer and therefore better contrasted: the background no longer interferes with the data, yet it provides sufficient information about them.

The graph is more intuitive: the labels, positioned next to the data, provide the required clarification where it is needed (when viewers look at the data).

library(ggplot2)

- The grammar of graphics:
  - 由 [Hadley Wickham](#) 於 2005 年發表 (Chief Scientist in RStudio)
  - 當前最多人使用的視覺化 R 套件
  - 取自 "The Grammar of Graphics" (Leland Wilkinson, 2005)
- 設計理念
  - 採用圖層系統 (layer)
  - 用抽象的概念來控制圖形，避免細節繁瑣
  - 圖形美觀

Simple dataset with variables mapped into aesthetic space.

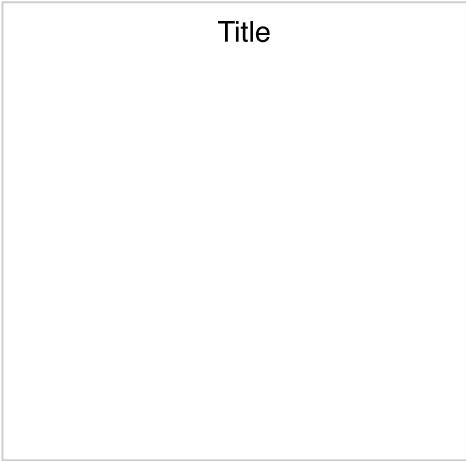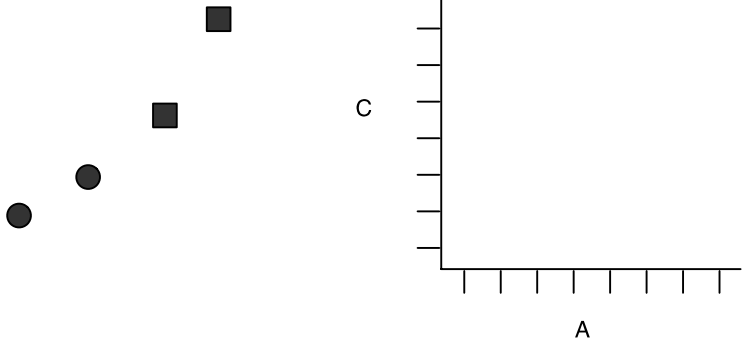| x	y	Shape
25	11	circle
0	0	circle
75	53	square
200	300	square

Figure 1.   Graphics objects produced by (from left to right): geometric objects, scales and coordinate system, plot annotations.
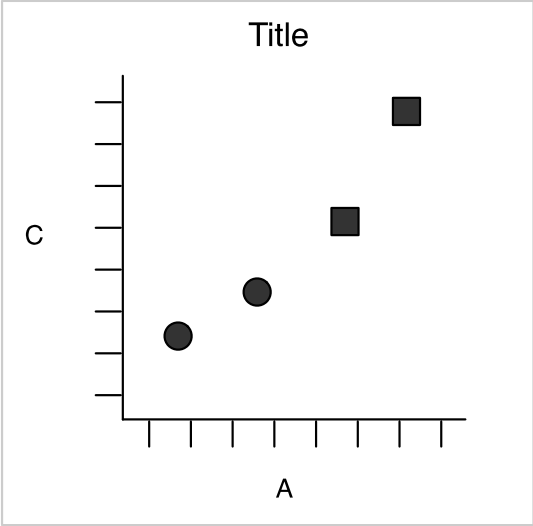

Figure 2.   The final graphic, produced by combining the pieces in Figure 1.

Table 4.   Simple dataset faceted into subsets.

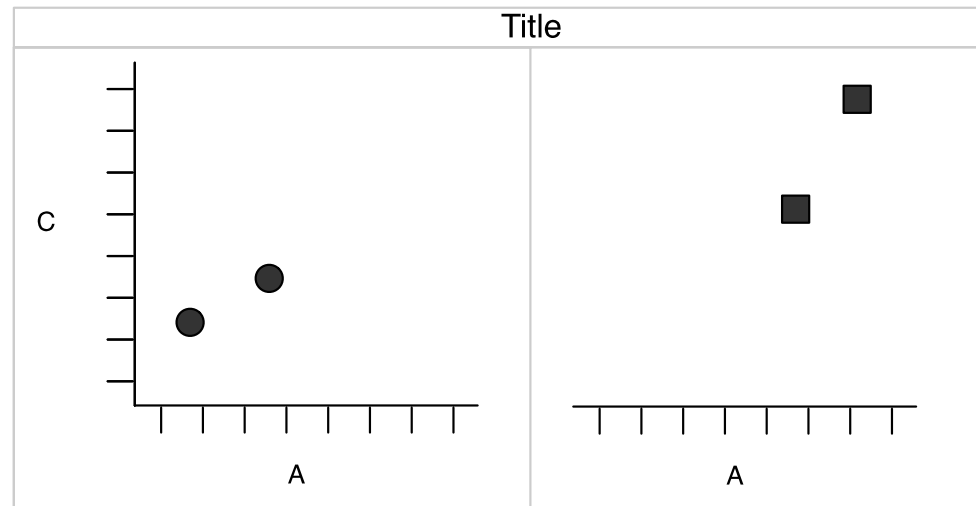|	x	y	Shape
a	2	4	circle
a	1	1	circle
b	4	15	square
b	9	80	square

Figure 3.   A more complicated plot, faceted by variable D. Here the faceting uses the same variable that is mapped to shape so that there is some redundancy in our visual representation. This allows us to easily see how the data have been broken into panels.

基礎語法

 # old school
plot(x = 1:10, y = **seq**(2, 20, 2))

 # ggplot2
library(ggplot2)
ggplot(data = df, **aes**(x = x, y = y)) + **geom_point**()

指定data名稱　　　　　指定x y軸　　　　　　畫點圖
　　　　畫布

- data and aesthetic mappings,
- geometric objects,
- scales, and
- facet specification

Plot components

- a default dataset and set of mappings from variables to aesthetics,
- one or more layers, with each layer having one geometric object, one statistical transformation, one position adjustment, and optionally, one dataset and set of aesthetic mappings,
- one scale for each aesthetic mapping used,
- a coordinate system,
- the facet specification

ggplot() - for fine, granular control of everything

- data layer:
  - data = data.frame
  - long-format (library("reshape2")
  - aes(x = x, y = y)

- geom layer: geom_xxx():
  - multi-layers: +
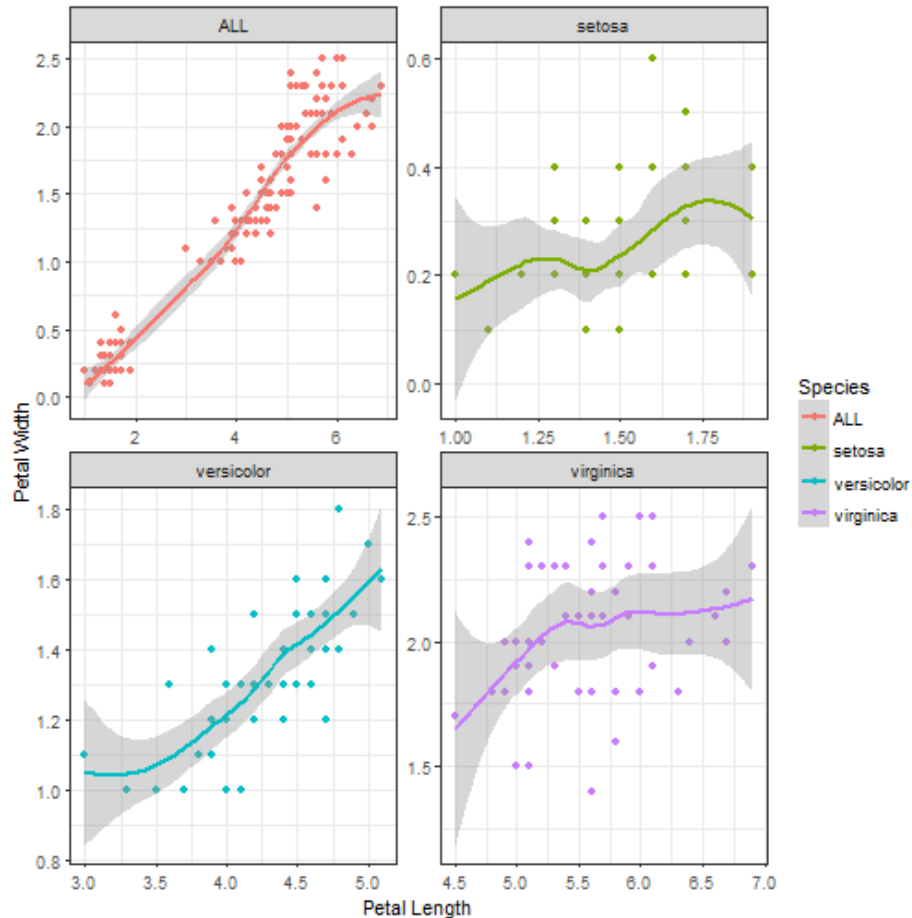 - Geometric objects that define the basic "shape" of the elements on the plot

 # list all geom **ls**(pattern = '^geom_', env = **as.environment**('package:ggplot2'))

- geom_abline, geom_area, geom_bar, …

In-class practice

- Replicate the given graph as close as possible

(a) Facet and smooth with the iris data

(b) Bubble-plot (scatter diagram) with mtcars